Revisiting the Synthetic Control Estimator

Ferman, Bruno and Pinto, Cristine

Sao Paulo School of Economics - FGV

23 September 2016

Revisiting the Synthetic Control Estimator*

Bruno Ferman[†] Cristine Pinto[‡]

Sao Paulo School of Economics - FGV

First Draft: June, 2016

This Draft: September, 2016

Abstract

The synthetic control (SC) method has been recently proposed as an alternative to estimate treatment effects in comparative case studies. The SC relies on the assumption that there is a weighted average of the control units that reconstructs the potential outcome of the treated unit in the absence of treatment. If these weights were known, then constructing the counterfactual for the treated unit using a weighted average of the control units would provide an unbiased estimator for the treatment effect, even if selection into treatment is correlated with the unobserved heterogeneity. In this paper, we revisit the SC method in a linear factor model where the SC weights are considered nuisance parameters that are estimated to construct the SC estimator. We show that, when the number of control units is fixed, the estimated SC weights will generally not converge to the weights that reconstruct the factor loadings of the treated unit, even when the number of pre-intervention periods goes to infinity. As a consequence, the SC estimator will be asymptotically biased if treatment assignment is correlated with the unobserved heterogeneity. The asymptotic bias only vanishes when the variance of the idiosyncratic error goes to zero. We suggest a slight modification in the SC method that guarantees that the SC estimator is asymptotically unbiased and has a lower asymptotic variance than the difference-in-differences (DID) estimator when the DID identification assumption is satisfied. We also propose an alternative way to estimate the SC weights that provides an asymptotically unbiased estimator under additional assumptions on the error structure. Finally, we consider the implications of our findings to the permutation test suggested in Abadie et al. (2010).

Keywords: synthetic control, difference-in-differences; linear factor model, inference, permutation test
JEL Codes: C12; C13; C21; C23

*We would like to thank Aureo de Paula, Tong Li, Ricardo Masini, Rodrigo Soares and seminar participants at PUC-Chile and of the Bristol Econometric Study Group for comments and suggestions. We also thank Deivis Angeli for excellent research assistance.

[†]bruno.ferman@fgv.br

[‡]cristine.pinto@fgv.br

1 Introduction

In a series of influential papers, Abadie and Gardeazabal (2003), Abadie et al. (2010), and Abadie et al. (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 pre-treatment periods to estimate weights such that a weighted average of the control units reconstructs the treated unit in the absence of treatment. Then they use these weights to compute the counterfactual of the treated unit in case it were not treated. Since then, this method has been used in a wide range of applications, including the evaluation of the impact of terrorism, civil wars and political risk, natural resources and disasters, international finance, education and research policy, health policy, economic and trade liberalization, political reforms, labor, taxation, crime, social connections, and local development.[1] According to Athey and Imbens (2016), *"the simplicity of the 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"*.

In this paper, we revisit the SC method in a linear factor model where the SC weights are considered as nuisance parameters that are estimated to construct the SC estimator. We consider the asymptotic distribution of the SC estimator when the number of pre-intervention periods goes to infinity. With the number of control units fixed, we show that the SC weights will generally not converge in probability to weights that reconstruct the factor loadings of the treated unit even if such weights exist. This implies that the SC estimator will be asymptotically biased if treatment assignment is correlated with the unobserved heterogeneity.[2] The relevance of such bias depends on the variance of the idiosyncratic error and only vanishes when this variance goes to zero.[3] We also show that the specification that uses only the average of pre-intervention outcomes as economic predictor can be particularly problematic. These results corroborate

[1]SC has been used in the evaluation of the impact of terrorism, civil wars and political risk (Abadie and Gardeazabal (2003), Bove et al. (2014), Li (2012), Montalvo (2011), Yu and Wang (2013)), natural resources and disasters (Barone and Mocetti (2014), Cavallo et al. (2013), Coffman and Noy (2011), DuPont and Noy (2012), Mideksa (2013), Sills et al. (2015), Smith (2015)), international finance (Jinjarak et al. (2013), Sanso-Navarro (2011)), education and research policy (Belot and Vandenberghe (2014), Chan et al. (2014), Hinrichs (2012)), health policy (Bauhoff (2014), Kreif et al. (2015)), economic and trade liberalization (Billmeier and Nannicini (2013), Gathani et al. (n.d.), Hosny (2012)), political reforms (Billmeier and Nannicini (2009), Carrasco et al. (2014), Dhungana (2011) Ribeiro et al. (2013)), labor (Bohn et al. (2014), Calderon (2014)), taxation (Kleven et al. (2013), de Souza (2014)), crime (Pinotti (2012b), Pinotti (2012a), Saunders et al. (2014)), social connections (Acemoglu et al. (2013)), and local development (Ando (2015), Gobillon and Magnac (2016), Kirkpatrick and Bennear (2014), Liu (2015), Severnini (2014)).

[2]We define the asymptotic bias as the difference between the expected value of the asymptotic distribution and the parameter of interest. We also show that, in the context of the SC estimator, the limit of the expected value converges to the expected value of the asymptotic distribution. Wong (2015) and Powell (2016) also consider the SC weights as nuisance parameters that must be estimated to construct the SC estimator. They argue that the SC weights would converge in probability to weights that satisfy the SC assumption. However, it is possible to show that, in their setting, the SC estimator will also be asymptotically biased under the same conditions we find in our paper. Details in Appendix A.4.

[3]In their proof that the SC estimator is asymptotically unbiased, Abadie et al. (2010) make an assumption that can only be satisfied if variance of the transitory shock is zero. Therefore, their result is consistent with our findings.

the findings in Ferman et al. (2016) that the SC estimator can misallocate a significant proportion of weights in Monte Carlo simulations, even when the number of pre-treatment periods is large.

Gobillon and Magnac (2013) show that the condition required in Abadie et al. (2010) for asymptotic unbiasedness can be satisfied if the number of control units goes to infinity and the matching variables (factor loadings and exogenous covariates) of the treated units belong to the support of the matching variables of control units. In this case, the SC estimator would be equivalent to the interactive effect methods they recommend. Xu (2016) proposes an alternative to the SC method in which in a first step he estimates the factor loadings, and then in a second step he constructs the SC unit to match the estimated factor loadings of the treated unit. This method would also require a large number of both control units and pre-treatment units, so that the factor loadings estimators are consistent. Differently from Gobillon and Magnac (2013) and Xu (2016), we consider the case with a finite number of control units and let the number of pre-intervention periods go to infinity, which is the same asymptotics considered in Abadie et al. (2010). We show that, in this case, the conditions under which the SC estimator is asymptotically unbiased are unlikely to be satisfied.

We propose two alternatives to the original SC method. First, we recommend a slight modification in the SC method where we demean the data using the pre-intervention period, and then construct the SC estimator using the demeaned data. We show that, if selection into treatment is only correlated with individual fixed effects (which is essentially the identification assumption of the DID model), then this demeaned SC estimator will be asymptotically unbiased. Also, in this case we can guarantee that the asymptotic variance of this demeaned SC estimator will be lower than the asymptotic variance of the DID estimator. If selection into treatment is correlated with time-varying common factors, then both the demeaned SC and the DID estimators would be asymptotically biased. We show that the asymptotic bias of the demeaned SC estimator will be lower than the bias of DID for a particular class of linear factor models. However, we show a specific example in which the asymptotic bias of the SC can be larger. This might happen when selection into treatment depends on common factors with low variance. We also provide an instrumental variables estimator for the SC weights that would be consistent under strong assumptions on the error structure, which would be valid if, for example, the idiosyncratic error is serially uncorrelated and all the common factors are serially correlated.

Finally, we also consider the implication of our findings to the permutation test proposed in Abadie et al. (2010). We evaluate whether the test statistic proposed in Abadie et al. (2010) has, asymptotically, the same distribution for all permutations. If this is the case, then, based on the results from Canay et al. (2014) on randomization tests under an approximate symmetry assumption, this permutation test would be asymptotically valid.[4] We show that this will be the case if the SC weights converge in probability to weights that reconstruct the factor loadings of the treated unit in all permutations. If the SC weights do not converge in probability to such weights, then the distribution of the test statistic might not be asymptotically symmetric for at least two reasons. First, if the SC estimator is asymptotically biased, then the test statistic might have a higher expected value for the treated unit. Also, even if the SC estimator is unbiased, then the variance of the test statistic might depend on characteristics of treated unit if the common factors are serially correlated.[5] We show that, in this case, this distortion would only be (asymptotically) relevant in very particular cases, as it depends at the same time on both the SC weights being different from the weights that reconstruct the factor loadings of the treated unit and on the serial correlation of the common factors being high relative to the variance of the transitory shocks. However, when the variance of the transitory shocks is small, then the SC weights would be close to the weights that reconstruct the factor loadings of the treated unit. Finally, we show in Monte Carlo simulations that distortions in the permutation test can be important if the number of pre-treatment periods is small. This happens because, in this case, the model might overfit the pre-treatment mean squared prediction error (MSPE), so it might not provide a valid correction for the post-treatment mean squared prediction error.

The remainder of this paper proceeds as follows. We start Section 2 with a brief review of the SC estimator. Then we show that the SC estimator that uses all pre-treatment outcome lags as economic predictors is, in general, asymptotically biased. We also consider in Section 2 the asymptotic properties of alternative specifications of the SC estimator. In Section 3 we propose two alternatives to the original SC estimator. In Section 4 we consider the implications of our results for to the permutation test proposed in Abadie et al. (2010). In Section 5, we consider the asymptotic properties and show results from Monte Carlo simulations for a particular class of linear factor models. We conclude in Section 6.

[4]There are two recent papers that analyze in detail the permutation test proposed in Abadie et al. (2010). Firpo and Possebom (2016) formalize the permutation test for the case where treatment is randomly assigned. In this case, the inference method suggested in Abadie et al. (2010) would provide valid inference for *unconditional* tests, provided that one is careful to exclude permutations in which the SC would be discarded due to poor pre-intervention fit if this were the treated unit. Our paper considers the asymptotic properties of the permutation test when we relax the hypothesis of random assignment. Also, even under random assignment, we consider hypothesis testing conditional on the data on hand. See Ferman and Pinto (2016) for details on why conditional tests should be preferable when there are few treated groups. In another recent paper, Ando and Sävje (2013) argue that the permutation test proposed by Abadie et al. (2010) is generally not valid and derive an alternative inference method. Differently from Ando and Sävje (2013), we consider whether Abadie et al. (2010) permutation test can be valid asymptotically when the number of pre-intervention is large.

[5]The SC estimator is asymptotically unbiased if treatment assignment is uncorrelated with the unobserved heterogeneity, even if the SC weights do not converge to weights that reconstruct the factor loadings of the treated unit.

2 Revisiting the Synthetic Control Model

2.1 The Synthetic Control Model

Suppose we have a balanced panel of $J+1$ units indexed by i observed on $t=1,...,T$ periods. We want to estimate the treatment effect of a policy change that affected only unit $j=1$ from period $T_0+1 \leq T$ to T. The potential outcomes are given by:

$$\begin{cases} y_{it}^C = \delta_t + \lambda_t \mu_i + \epsilon_{it} \\ y_{it}^T = \alpha_{it} + y_{it}^C \end{cases} \tag{1}$$

where δ_t is an unknown common factor with constant factor loadings across units, λ_t is a $(1 \times F)$ vector of common factors, μ_i is a $(F \times 1)$ vector of unknown factor loadings, and the error terms ϵ_{it} are unobserved transitory shocks. We only observe $y_{it} = d_{it} y_{it}^T + (1-d_{it}) y_{it}^C$, where $d_{it} = 1$ if unit i is treated at time t. We assume ϵ_{it} independent across units and in time. Note that the unobserved error $u_{it} = \lambda_t \mu_i + \epsilon_{it}$ might be correlated across unit and in time due to the presence of $\lambda_t \mu_i$. As in Abadie et al. (2010), Gobillon and Magnac (2013) and Powell (2016), we allow for correlation between $\lambda_t \mu_i$ and the treatment assignment. Since we hold the number of units $(J+1)$ fixed and look at asymptotics when the number of pre-treatment periods goes to infinity, we treat the vector of unknown factor loads (μ_i) as fixed and the common factors (λ_t) as random variables. In order to simplify the exposition of our main results, we consider the model without observed covariates Z_i until Section 2.3.2.

The main assumption of the Synthetic Control method (SC) is that there is a stable linear combination of the control units that absorbs all time correlated shocks $\lambda_t \mu_i$.

Assumption 1 (existence of weights):

$$\exists \; \mathbf{w}^* = \{w_1^{*j}\}_{j \neq 1} \mid \mu_1 = \sum_{j \neq 1} w_1^{*j} \mu_j, \; \sum_{j \neq 1} w_1^{*j} = 1, \text{ and } w_1^{*j} \geq 0$$

Note that we consider the existence of weights that reconstruct the unobserved factors loadings μ_1, following the structure of Ando and Sävje (2013) and Powell (2016).[6] This assumption is slightly different from the assumption in Abadie et al. (2010). They define the SC weights so that $y_{1t} = \sum_{j \neq 1} w_1^{*j} y_{jt}$ for all $t \leq T_0$. Note, however, that this condition cannot be satisfied in general since ϵ_{it} are independent variables across i.

[6] Powell (2016) treats μ_i as random variables, so he considers that assumption 1 is valid in expectation. Wong (2015) considers weights that reconstruct the expected value of the potential outcome if the observation is not treated, without imposing a linear factor model structure. As we show in Appendix A.4, our main results remain valid in the setting considered in Wong (2015).

We treat the SC weights $\mathbf{w}^*$ as nuisance parameters that we need to estimate in order to construct our SC estimator. Note that there is no guarantee that there is a unique set of weights that satisfies assumption 1, so we define $\Phi_1 = (\mathbf{w} = \{w_1^j\}_{j\neq 1} \mid \mu_1 = \sum_{j\neq 1} w_1^j \mu_j, \ \sum_{j\neq 1} w_1^j = 1$, and $w_1^j \geq 0)$ as the set of weights that satisfy this condition.

The SC method consists of estimating the SC weights $\widehat{\mathbf{w}}_1 = \{\hat{w}_1^j\}_{j\neq 1}$ using information on the pre-treatment period and then constructing the SC estimator $\hat{\alpha}_{1t} = y_{1t} - \sum_{j\neq 1} \hat{w}_1^j y_{jt}$ for $t > T_0$. Abadie et al. (2010) suggest a minimization problem to estimate these weights using the pre-intervention data. They define a set of K economic predictors where X_1 is a $(K \times 1)$ vector containing the economic predictors for the treated unit and X_0 is a $(K \times J)$ matrix of economic predictors for the control units.[7] The SC weights are estimated by minimizing $||X_1 - X_0\mathbf{w}||_V$ subject to $\sum_{i=2}^{J+1} w_1^j = 1$ and $w_1^j \geq 0$, where V is a $(K \times K)$ positive semidefinite matrix. They discuss different possibilities for choosing the matrix V, including an iterative process where V is chosen such that the solution to the $||X_1 - X_0\mathbf{w}||_V$ optimization problem minimizes the pre-intervention prediction error. In other words, let $\mathbf{Y}_1^P$ be a $(T_0 \times 1)$ vector of pre-intervention outcomes for the treated unit, while $\mathbf{Y}_0^P$ be a $(T_0 \times J)$ matrix of pre-intervention outcomes for the control units. Then the SC weights would be chosen as $\widehat{\mathbf{w}}(V^*)$ such that V^* minimizes $||\mathbf{Y}_1^P - \mathbf{Y}_0^P \widehat{\mathbf{w}}(V)||$.

As argued in Ferman et al. (2016), one limitation of the SC estimator is that the theory behind the SC method does not provide a clear guidance on how one should choose the economic predictors in matrices X_1 and X_0. This reflects in a wide range of different specification choices in SC applications. We consider here 3 common specifications: (1) the use of all pre-intervention outcome values, (2) the use of the average of the pre-intervention outcomes, and (3) the use of other time invariant covariates in addition to the average of the pre-intervention outcomes.[8]

2.2 The asymptotic bias of the SC estimator

We focus first on the case where one includes all pre-intervention outcome values as economic predictors. In this case, the matrix V that minimizes the second step of the nested optimization problem would be the identity matrix (see Kaul et al. (2015)), so the optimization problem suggested by Abadie et al. (2010) to

[7] Economic predictors can be, for example, linear combinations of the pre-intervention values of the outcome variable or other covariates not affected by the treatment.

[8] Kaul et al. (2015) show that the weights allocated to time-invariant covariates would be zero if one uses all pre-treatment intervention outcome values as economic predictors. Therefore, we do not consider this case.

estimate the weights simplifies to an M-estimator given by:

$$
\begin{aligned}
\{\hat{w}_1^j\}_{j\neq 1} &= \text{argmin}_{\mathbf{w}\in W} \frac{1}{T_0}\sum_{t=1}^{T_0}\left[y_{1t} - \sum_{j\neq 1} w_1^j y_{jt}\right]^2 \\
&= \text{argmin}_{\mathbf{w}\in W} \frac{1}{T_0}\sum_{t=1}^{T_0}\left[\epsilon_{1t} - \sum_{j\neq 1} w_1^j \epsilon_{jt} + \lambda_t\left(\mu_1 - \sum_{j\neq 1} w_1^j \mu_j\right)\right]^2 \qquad (2)
\end{aligned}
$$

where $W = \{\{w_1^j\}_{j\neq 1} \in \mathbb{R}^J | w_1^j \geq 0 \text{ and } \sum_{j\neq 1} w_1^j = 1\}$.

We impose conditions such that this objective function converges uniformly in probability to its population average.

Assumption 2 (stationary process): $(\epsilon_{jt}, \lambda_t)'$ is weakly stationary and second moment ergodic.

Under assumption 2, we have that:

$$
\frac{1}{T_0}\sum_{t=1}^{T_0}\left[\epsilon_{1t} - \sum_{j\neq 1} w_1^j \epsilon_{jt} + \lambda_t\left(\mu_1 - \sum_{j\neq 1} w_1^j \mu_j\right)\right]^2 \xrightarrow{p} E\left[\epsilon_{1t} - \sum_{j\neq 1} w_1^j \epsilon_{jt} + \lambda_t\left(\mu_1 - \sum_{j\neq 1} w_1^j \mu_j\right)\right]^2 \qquad (3)
$$

Let $\bar{\mathbf{w}} = \{\bar{w}_1^j\}_{j\neq 1}$ be the weights that minimize this expectation and treat $\widehat{\mathbf{w}} = \{\hat{w}_1^j\}_{j\neq 1}$ as an M-estimator. We show now that $\bar{\mathbf{w}} \notin \Phi_1$, which implies that the SC weights will converge in probability to weights that do not satisfy the condition stated in assumption 1, even under the assumption of existence of such weights. We consider a simple case where $var(\epsilon_{it}) = \sigma_\epsilon^2$ for all i and ϵ_{it} is uncorrelated with λ_t. Let $E[\lambda_t'\lambda_t] = \Omega$ be the matrix of second moments of λ_t. Therefore, the objective function simplifies to:

$$
\Gamma(\{w_1^j\}_{j\neq 1}) = \sigma_\epsilon^2\left(1 + \sum_{j\neq 1}(w_1^j)^2\right) + \left(\mu_1 - \sum_{j\neq 1} w_1^j \mu_j\right)' \Omega \left(\mu_1 - \sum_{j\neq 1} w_1^j \mu_j\right) \qquad (4)
$$

Note that the objective function has two parts. The first one reflects that different choices of weights will generate different weighted averages of the idiosyncratic shocks ϵ_{it}. In this simpler case, this part would be minimized when we set all weights equal to $\frac{1}{J}$. The second part reflects the presence of common factors λ_t that would remain after we choose the weights to construct the SC unit. If assumption 1 is satisfied, then we can set this part equal to zero by choosing $\mathbf{w}^* \in \Phi_1$.

Consider that we start at $\{w_1^{*j}\}_{j\neq 1} \in \Phi_1$ and move in the direction of $w_1^j = \frac{1}{J}$ for all $j = 2, .., J+1$, with $w_1^j = w_1^{*j} + \Delta(\frac{1}{J} - w_1^{*j})$. Note that, for all $\Delta \in [0, 1]$, these weights will continue to satisfy the constraints of the minimization problem. If we consider the derivative of function 4 with respect to Δ at $\Delta = 0$, we have that:

$$\Gamma'(\{w_1^{*j}\}_{j\neq 1}) = 2\sigma_\epsilon^2 \left(\frac{1}{J} - \sum_{j=2}^{J+1} (w_1^{*j})^2 \right) < 0 \text{ unless } w_1^{*j} = \frac{1}{J} \tag{5}$$

Therefore, $\mathbf{w}^* \in \Phi_1$ cannot be, in general, a solution of the objective function of the M-estimator. This implies that, when $T_0 \to \infty$, the SC weights will converge in probability to weights $\bar{\mathbf{w}}$ that does not satisfy assumption 1, unless it turns out that $\mathbf{w}^*$ also minimizes the variance of the idiosyncratic errors. The SC estimator will be given by:

$$\hat{\alpha}_{1t} = y_{1t} - \sum_{j\neq 1} \hat{w}_1^j y_{it} \xrightarrow{d} \alpha_{1t} + \left(\epsilon_{1t} - \sum_{j\neq 1} \bar{w}_1^j \epsilon_{jt} \right) + \lambda_t \left(\mu_1 - \sum_{j\neq 1} \bar{w}_1^j \mu_j \right) \tag{6}$$

The SC estimator will only be asymptotically unbiased if we have that $E\left[\epsilon_{1t} - \sum_{j\neq 1} \bar{w}_1^j \epsilon_{jt} | d_{1t}\right] = 0$ and $E\left[\lambda_t \left(\mu_1 - \sum_{j\neq 1} \bar{w}_1^j \mu_j\right) | d_{1t}\right] = 0$.[9] Since $\left(\mu_1 - \sum_{j\neq 1} \bar{w}_1^j \mu_j\right) \neq 0$, this implies that we cannot have selection on unobservables, even if selection is based on the common factors. Abadie et al. (2010) argue that, in contrast to the usual DID model, the SC model would allow the effects of confounding unobserved characteristics to vary with time. However, the SC estimator would not be asymptotically unbiased under selection on unobservable heterogeneity because the SC weights will not satisfy the condition required by the method (even when $T_0 \to \infty$ and under all assumption of the SC model). The asymptotic bias would only converge to zero when we also have that $\sigma_\epsilon^2 \to 0$. In their proof, Abadie et al. (2010) assume the existence of weights $\{w_2^*, ..., w_{J+1}^*\}$ that satisfy the condition $y_{1t} = \sum_{i=2}^{J+1} w_i^* y_{it}$ for all $t \leq T_0$. However, if $T_0 \to \infty$, then the probability that such weights exist converges in probability to zero, unless the variance of ϵ_{it} is equal to zero, in which case we would also find unbiasedness in our setting.[10]

In order to provide a better intuition for this result, it is worth considering a simple example. Suppose there are only 2 factors, so $\lambda_t = (\lambda_t^1, \lambda_t^2)$ and $\mu_i = (\mu_i^1, \mu_i^2) \in \{(1,0),(0,1)\}$. Intuitively, this means that we have two groups, one that follows parallel trend given by λ_t^1 and another one that follows parallel trend given by λ_t^2.[11] Assume that $\mu_1 = (1,0)$, so the treated unit belongs to the first group, and that half of the units in the donor pool belongs to group 1 while the other half belongs to group 2. If we knew μ_i, then we could

[9] We consider the definition of asymptotic unbiasedness as the expected value of the asymptotic distribution of $\hat{\alpha}_{1t} - \alpha_{1t}$ equal to zero. An alternative definition is that $E[\hat{\alpha}_{1t} - \alpha_{1t}] \to 0$. We show in the Appendix that these two definitions are equivalent in our setting under standard assumptions.

[10] Gobillon and Magnac (2013) show that this condition can be satisfied if $J \to \infty$ and the matching variables of the treated units belong to the support of the matching variables of control units. In this case, the SC estimator would be asymptotically unbiased.

[11] This is the data generating process considered in Ferman et al. (2016).

construct our SC estimator by setting equal weights to all units in the donor pool that belong to group 1 and weight equal to zero for all other units. In this case, we would have $\hat{\alpha}^*_{1t} - \alpha_{1t} \overset{d}{\rightarrow} \epsilon_{1t} - \sum_{j\neq 1|\mu_j=(1,0)} \frac{2}{J}\epsilon_{jt}$ and the SC estimator would be asymptotically unbiased even if the treatment is correlated with the common shock that affects unit 1, λ^1_t. Intuitively, the SC estimator would only compare the treated unit to units in the donor pool that experienced the same common shock but were not treated. Since we assume that treatment is randomly assigned conditional on $(\lambda^1_t, \lambda^2_t)$, the estimator would be unbiased. The problem, however, is that we need to estimate the SC weights. Moreover, the SC will not assign the correct weights even when $T_0 \rightarrow \infty$, because there will always be a first-order gain in the optimization problem of moving away from weights that set $w^j_1 = 0$ for j such that $\mu_j = (0,1)$. Let p be the proportion of weights allocated to the correct units. In this case, we have that $\hat{\alpha}_{1t} - \alpha_{1t} \overset{d}{\rightarrow} \epsilon_{1t} - \sum_{j\neq 1} \bar{w}^j_1 \epsilon_{jt} + (1-p)\lambda^1_t - (1-p)\lambda^2_t$. Therefore, the SC estimator would be asymptotically biased if treatment assignment is correlated with the common factors $(\lambda^1_t, \lambda^2_t)$.

2.3 Alternative SC specifications

2.3.1 Average of pre-intervention outcome as predictor

We consider now another very common specification in SC applications, which is to use the average pre-treatment outcome as the economic predictor. Note that if one uses only the average pre-treatment outcome as the economic predictor then the choice of matrix V would be irrelevant. In this case, the minimization problem would be given by:

$$\begin{aligned} \{\hat{w}^j_1\}_{j\neq 1} &= \text{argmin}_{w\in W} \left[\frac{1}{T_0} \sum_{t=1}^{T_0} \left(y_{1t} - \sum_{j\neq 1} w^j_1 y_{jt} \right) \right]^2 \\ &= \text{argmin}_{w\in W} \left[\frac{1}{T_0} \sum_{t=1}^{T_0} \left(\epsilon_{1t} - \sum_{j\neq 1} w^j_1 \epsilon_{jt} + \lambda_t \left(\mu_1 - \sum_{j\neq 1} w^j_1 \mu_j \right) \right) \right]^2 \end{aligned} \tag{7}$$

where $W = \{\{w^j_1\}_{j\neq 1} \in \mathbb{R}^J | w^j_1 \geq 0 \text{ and } \sum_{j\neq 1} w^j_1 = 1\}$.

Therefore, assuming weakly dependence of λ_t, the objective function converges in probability to:

$$\Gamma(\{w^j_1\}_{j\neq 1}) = \left[E\left[\lambda_t\right] \left(\mu_1 - \sum_{j\neq 1} w^j_1 \mu_j \right) \right]^2 \tag{8}$$

Assuming that there is a time-invariant common factor (that is, $\lambda^1_t = 1$ for all t) and that λ_t is weakly stationary, we have that, without loss of generality, $E[\lambda_t^k] = 0$ for $k > 1$. In this case, the objective function collapses to:

$$\Gamma(\{w_1^j\}_{j \neq 1}) = \left[\left(\mu_1^1 - \sum_{j \neq 1} w_1^j \mu_j^1\right)\right]^2 \tag{9}$$

Therefore, while we assume that there exists at least one set of weights that reproduces all factor loadings (assumption 1), the objective function will only look for weights that approximate the first factor loading. This is problematic because it might be that assumption 1 is satisfied, but there are weights $\{\tilde{w}_1^j\}_{j \neq 1} \notin \Phi_1$ that satisfy $\mu_1^1 = \sum_{j \neq 1} \tilde{w}_1^j \mu_j^1$. In this case, there is no guarantee that the SC control method will choose weights that are close to the correct ones. This result is consistent with the Monte Carlo simulations in Ferman et al. (2016) who show that this specification performs particularly bad in allocating the weights correctly.

2.3.2 Adding other covariates as predictors

Most applications that use the average pre-intervention outcome value as economic predictor also consider other time invariant covariates as economic predictors. Let Z_i be a $(R \times 1)$ vector of observed covariates (not affected by the intervention). Model 1 changes to:

$$\begin{cases} y_{it}^C = \delta_t + \theta_t Z_i + \lambda_t \mu_i + \epsilon_{it} \\ y_{it}^T = \alpha_{it} + y_{it}^C \end{cases} \tag{10}$$

We also modify assumption 1 so that the weights reproduce both μ_1 and Z_1.

Assumption 1′ (existence of weights):

$$\exists \ \{w_1^{*j}\}_{j \neq 1} \mid \mu_1 = \sum_{j \neq 1} w_1^{*j} \mu_j, \ Z_1 = \sum_{j \neq 1} w_1^{*j} Z_j, \ \sum_{j \neq 1} w_1^{*j} = 1, \text{ and } w_1^{*j} \geq 0$$

Let X_1 be an $(R + 1 \times 1)$ vector that contains the average pre-intervention outcome and all covariates for unit 1, while X_0 is a $(R + 1 \times J)$ matrix that contains the same information for the control units. For a given V, the first step of the nested optimization problem suggested in Abadie et al. (2010) would be given by:

$$\widehat{\mathbf{w}}(V) \in \text{argmin}_{\mathbf{w}\in W} ||X_1 - X_0\mathbf{w}||_V \tag{11}$$

where $W = \{\{w_1^j\}_{j\neq 1} \in \mathbb{R}^J | w_1^j \geq 0 \text{ and } \sum_{j\neq 1} w_1^j = 1\}$. Note that the objective function of this minimization problem converges to $||\bar{X}_1 - \bar{X}_0\mathbf{w}||_V$, where:

$$\bar{X}_1 - \bar{X}_0\mathbf{w} = \begin{bmatrix} E[\theta_t]\left(Z_1 - \sum_{j\neq 1} w_1^j Z_j\right) + \left(\mu_1^1 - \sum_{j\neq 1} w_1^j \mu_j^1\right) \\ \left(Z_1^1 - \sum_{j\neq 1} w_1^j Z_j^1\right) \\ \vdots \\ \left(Z_1^R - \sum_{j\neq 1} w_1^j Z_j^R\right) \end{bmatrix} \tag{12}$$

Similarly to the case with only the average pre-intervention outcome value as economic predictor, it might be that assumption $1'$ is satisfied, but there are weights $\{\tilde{w}_1^j\}_{j\neq 1}$ that satisfy $\mu_1^1 = \sum_{j\neq 1} \tilde{w}_1^j \mu_j^1$ and $Z_1 = \sum_{j\neq 1} \tilde{w}_1^j Z_j$, although $\mu_1^k \neq \sum_{j\neq 1} \tilde{w}_1^j \mu_j^k$ for some $k > 1$. Therefore, there is no guarantee that an estimator based on this minimization problem would converge to weights that satisfy assumption $1'$ for any given matrix V.

The second step in the nested optimization problem is to choose V such that $\widehat{\mathbf{w}}(V)$ minimizes the pre-intervention prediction error. Note that this problem is essentially given by:

$$\widehat{\mathbf{w}} \quad = \quad \text{argmin}_{w\in\widetilde{W}} \left[\frac{1}{T_0}\sum_{t=1}^{T_0}\left(y_{1t} - \sum_{j\neq 1} w_1^j y_{jt}\right)\right]^2 \tag{13}$$

where $\widetilde{W} \subseteq W$ is the set of $\mathbf{w}$ such that $\mathbf{w}$ is the solution to problem 11 for some positive semidefinite matrix V. Similarly to the SC estimator that includes all pre-treatment outcomes, there is no guarantee that this minimization problem will choose weights that satisfy assumption $1'$ even when $T_0 \to \infty$. More specifically, if the variance of ϵ_{it} is large, then the SC estimator would tend to choose weights that are uniform across the control units in detriment of weights that satisfy assumption $1'$. Moreover, since we might have multiple solutions to problem 11, there might be no V such that $\widehat{\mathbf{w}}(V)$ converges in probability to weights in Φ_1. Therefore, it is not possible to guarantee that this SC estimator would be asymptotically unbiased.

3 Alternatives

3.1 Demeaned SC Estimator

In contrast to the SC estimator, the DID estimator for the treatment effect in a given post-intervention period $t > T_0$ would be given by:

$$
\begin{aligned}
\hat{\alpha}_{1t}^{DID} &= y_{1t} - \frac{1}{J}\sum_{j\neq 1} y_{jt} - \frac{1}{T_0}\sum_{\tau=1}^{T_0}\left[y_{1\tau} - \frac{1}{J}\sum_{j\neq 1} y_{j\tau}\right] \\
&= \epsilon_{1t} - \frac{1}{J}\sum_{j\neq 1}\epsilon_{jt} + \lambda_t\left(\mu_1 - \frac{1}{J}\sum_{j\neq 1}\mu_j\right) - \frac{1}{T_0}\sum_{\tau=1}^{T_0}\left[\epsilon_{1\tau} - \frac{1}{J}\sum_{j\neq 1}\epsilon_{j\tau} + \lambda_\tau\left(\mu_1 - \frac{1}{J}\sum_{j\neq 1}\mu_j\right)\right] \\
&\overset{d}{\to} \epsilon_{1t} - \frac{1}{J}\sum_{j\neq 1}\epsilon_{jt} + (\lambda_t - E[\lambda_\tau])\left(\mu_1 - \frac{1}{J}\sum_{j\neq 1}\mu_j\right)
\end{aligned} \tag{14}
$$

where we assumed that the pre-intervention average for the common factors converges in probability to their unconditional means. Implicitly we assume that λ_t is weakly dependent, so even if some pre-treatment common factors are correlated with the treatment assignment to unit 1 after T_0, when $T_0 \to \infty$ the pre-treatment average would converge to its unconditional expectation.

Therefore, the DID estimator would only be asymptotically unbiased if common factors that are not constant over time are uncorrelated with treatment assignment. In this case, these common factors would enter the error term and would not cause bias because their expectation conditional on treatment status would be equal to zero. The DID model allows for selection on common factors that are constant over time. In this case, the characteristics that are correlated with treatment assignment would be captured by the unit fixed effects. Therefore, if the DID assumptions are satisfied, then the DID estimator would be asymptotically unbiased while the SC estimator would be, in general, asymptotically biased.

As an alternative to the standard SC estimator, we suggest a modification in which we calculate the pre-treatment average for all units and demean the data. If common factors are stationary, this implies a model with no time-invariant common factor. We show in the Appendix that the only difference relative to the original model is that the common factors $\tilde{\lambda}_t$ and factor loadings $\tilde{\mu}_i$ would not include the time-invariant common factor. Also, we can assume without loss of generality that $E[\tilde{\lambda}_t] = 0$. In this case, we guarantee that the SC estimator will be asymptotically unbiased when the DID assumptions are satisfied. Note that we also make assumption 1 weaker, since there might be weights that reconstruct all common factors $\tilde{\lambda}_t$ that are not constant over time, but does not match the level of the treated unit.[12] We can show that, if the DID assumption is valid, then both this demeaned SC estimator and the DID estimator will be asymptotically unbiased, but the variance of the asymptotic distribution of the demeaned SC estimator will always be weakly lower relative to the DID estimator. Let $\hat{\alpha}_{1t}^{\text{SC}'}$ be the demeaned SC estimator. Under the DID assumption, $\tilde{\lambda}_t$ and ϵ_{ij} will be independent of the fact that unit 1 was treated after T_0. Therefore, for a given for $t > T_0$, the variance of the asymptotic distribution of the SC estimator would be given by:

$$a.var(\hat{\alpha}_{1t}^{\text{SC}'} - \alpha_{1t}) = E\left[\left(\tilde{\epsilon}_{1t} - \sum_{j\neq 1} \bar{w}_1^j \tilde{\epsilon}_{jt}\right) + \tilde{\lambda}_t \left(\tilde{\mu}_1 - \sum_{j\neq 1} \bar{w}_1^j \tilde{\mu}_j\right)\right]^2 \tag{15}$$

while:

$$a.var(\hat{\alpha}_{1t}^{\text{DID}} - \alpha_{1t}) = E\left[\left(\tilde{\epsilon}_{1t} - \sum_{j\neq 1} \frac{1}{J} \tilde{\epsilon}_{jt}\right) + \tilde{\lambda}_t \left(\tilde{\mu}_1 - \sum_{j\neq 1} \frac{1}{J} \tilde{\mu}_j\right)\right]^2 \tag{16}$$

Since the DID weights belong to W and the demeaned SC weights converge in probability to weights that minimize the function $E\left[\left(\epsilon_{1t} - \sum_{j\neq 1} w_1^j \epsilon_{jt}\right) + \tilde{\lambda}_t \left(\mu_1 - \sum_{j\neq 1} w_1^j \tilde{\mu}_j\right)\right]^2$, it must be that $a.var(\hat{\alpha}_{1t}^{\text{SC}'} - \alpha_{1t}) \leq a.var(\hat{\alpha}_{1t}^{\text{DID}} - \alpha_{1t})$. Note that this result is valid even if assumption 1 does not hold.

If the correlation comes from common factors that are not constant over time and assumption 1 is satisfied, then the bias of the SC estimator would usually be lower than the bias of the DID estimator. We show in Section 5 a particular class of linear factor models in which the asymptotic bias of the demeaned SC estimator will always be lower. However, we show a very specific example in Appendix A.1 in which the DID bias can be smaller than the bias of the SC. This might happen when selection into treatment depends on common factors with low variance.

3.2 IV-Like SC Estimator

We also propose an alternative way of estimating the SC weights that provide consistent estimators if we impose additional assumptions on the common factors and transitory shocks. Note that the asymptotic bias of the SC estimator derived in Section 2.2 comes from the first step of the SC method in which one estimates the SC weights using the pre-treatment information. As noted by Wong (2015), the minimization problem when one includes all pre-intervention lags is equivalente to a restricted OLS estimator of y_{1t} on

[12] Note that if assumption 1 is valid for the original model, then it will also be valid for the demeaned model.

$y_{2,t}, ..., y_{J+1,t}$. For weights $\{w_1^{*j}\}_{j\neq 1} \in \Phi_1$, we can write:

$$y_{1t} = \sum_{j=1}^{J+1} w_1^{*j} y_{jt} + \eta_t, \text{ for } t \leq T_0$$

where:

$$\eta_t = \epsilon_{1t} - \sum_{j=1}^{J+1} w_1^{*j} \epsilon_{jt}$$

The key problem is that η_t is correlated with y_{jt}, which implies that this restricted OLS regression would be biased. Imposing strong assumptions on the structure of the idiosyncratic error and the common factors, we show that it is possible to consider moment equations that will be equal to zero if, and only if, $\{w_1{}^j\}_{j\neq 1} \in \Phi_1$.

Let $\mathbf{y}_t = (y_{2,t}, ..., y_{J+1,t})'$, $\mu_{\mathbf{0}}$ be a $(F \times J)$ matrix with columns μ_j, $\epsilon_t = (\epsilon_{2,t}, ..., \epsilon_{J+1,t})$, and $\mathbf{w} = (w_1^2, ..., w_1^{J+1})'$. In this case, we can look at:

$$\begin{aligned}
\mathbf{y}_{t-1}(y_{1t} - \mathbf{y}_t'\mathbf{w}) &= (\mu_0'\lambda_{t-1}' + \epsilon_{t-1})\lambda_t\left(\mu_1 - \mu_{\mathbf{0}}\mathbf{w}\right) + (\mu_0'\lambda_{t-1}' + \epsilon_{t-1})(\epsilon_{1t} - \epsilon_t'\mathbf{w}) \\
&= \mu_0'\lambda_{t-1}'\lambda_t\left(\mu_1 - \mu_{\mathbf{0}}\mathbf{w}\right) + \epsilon_{t-1}\lambda_t\left(\mu_1 - \mu_{\mathbf{0}}\mathbf{w}\right) + \mu_0'\lambda_{t-1}'(\epsilon_{1t} - \epsilon_t'\mathbf{w}) + \epsilon_{t-1}(\epsilon_{1t} - \epsilon_t'\mathbf{w})
\end{aligned}$$

If we assume that ϵ_{it} is independent across t and independent of λ_t, then:

$$E\left[\mathbf{y}_{t-1}(y_{1t} - \mathbf{y}_t'\mathbf{w})\right] = \mu_0' E\left[\lambda_{t-1}'\lambda_t\right]\left(\mu_1 - \mu_{\mathbf{0}}\mathbf{w}\right)$$

Therefore, if the $(J \times F)$ matrix $\mu_0' E\left[\lambda_{t-1}'\lambda_t\right]$ has full rank, then the moment conditions equal to zero if, and only if, $\mathbf{w} \in \Phi_1$. One particular case in which this assumption is valid is if λ_t^f and $\lambda_t^{f'}$ are uncorrelated and λ_t^f is serially correlated for all $f = 1, ..., F$. Intuitively, under these assumptions, we can use the lagged outcome values of the control units as instrumental variables for the control units' outcomes.[13] One challenge to analyze this method is that there might be multiple solutions to the moment condition. Based on the results in Chernozhukov et al. (2007), it is possible to consistently estimate this set. Therefore, it is possible to generate an IV-like SC estimator that is asymptotically unbiased. A possible limitation of this method is

[13] The idea of SC-IV is very similar to the IV estimator used in dynamic panel data. In the dynamic panel models, lags of the outcome are used to deal with the endogeneity that comes from the fact the idiosyncratic errors are correlated with the lagged depend variable included in the model as covariates. The number of lags that can be used as instruments depends on the serial correlation of the error terms.

that it might rely on a very large number of pre-treatment periods so that weights are close to weights that satisfy assumption 1. Results from MC simulations available upon request.

4 Permutation Tests

Abadie et al. (2010) argue that large sample inferential techniques are not well suited to comparative case studies when the number of units in the comparison group is small.[14] They propose a permutation test where they apply the SC method to every potential control in the sample. First, they consider a graphical analysis where they compare the prediction error of their SC estimator across time with the prediction error for each of the placebo estimates. Then they consider whether the prediction error when one considers the actual treated unit is "unusually" large relative to the distribution of prediction errors for the units in the donor pool.

Note that this procedure does not provide a clear decision rule on whether the null hypothesis should be rejects. Still, this analysis would implicitly reject the null when the post-intervention mean squared prediction error (MSPE) for the SC estimate is greater than the post-intervention MSPE for the placebo estimates. We check whether this approach provides valid inference asymptotically when $T_0 \to \infty$ applying Canay et al. (2014) results for randomization tests under approximate symmetry. The main idea of Canay et al. (2014) framework is that we can look at test statistics that have asymptotically the same distributions under the null for all permutations. Therefore, the graphical analysis suggested in Abadie et al. (2010) would be valid if the distribution of the following test statistic has the same distribution for all permutations:

$$t_i^{\text{post}} = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} \left[y_{it} - \sum_{j \neq i} \hat{w}_i^j y_{jt} \right]^2 \tag{17}$$

We start assuming that assumption 1 is valid for all i. In particular, we assume that:

$$\exists \ \{w^{*j}_i\}_{j \neq i} \mid \mu_i = \sum_{j \neq i} w^{*j}_i \mu_j, \ \sum_{j \neq 1} w^{*j}_{i\,i} = 1, \text{ and } w^{*j}_i \geq 0 \ \forall i = 1, ..., J+1 \tag{18}$$

Ando and Sävje (2013) argue that in most applications it would not be reasonable to assume that this

[14] Carvalho et al. (2015) and Powell (2016) rely on large sample inferential techniques. Instead of testing the null hypothesis of no effect for all post-treatment periods, they test whether the average effect across time is equal to zero. If both the number of pre- and post-intervention periods is large, then they are able to derive the asymptotic distribution of the estimator. This method would not work if one want to test the null of no effect for all post-treatment periods or if the number of finite post-intervention periods is finite.

assumption is valid for all i. We believe that this condition might be reasonable in some applications. For example, this condition is satisfied if we have different groups of units where time trends are different across groups but parallel within groups, as considered in Ferman et al. (2016). We analyze this case in detail in Section 5. In this case, the main idea of the SC estimator would be to select the control units that follow the same time trend as the treated unit. We consider below the implications in case assumption 1 is not valid for all i.

Therefore, if we assume that the estimator of the SC weights $\hat{\mathbf{w}}_i \xrightarrow{p} \mathbf{w}_i^* \in \Phi_i$, then for all i we will have that:

$$t_i^{\text{post}} \xrightarrow{d} \frac{1}{T-T_0} \sum_{t=T_0+1}^{T} \left[\epsilon_{it} - \sum_{j \neq i} {w^*}_i^j \epsilon_{jt} \right]^2 \tag{19}$$

where $\Phi_i = (\{w_i^j\}_{j \neq i} \mid \mu_i = \sum_{j \neq i} w_1^j \mu_j,\ \sum_{j \neq i} {w^*}_i^j = 1, \text{ and } {w^*}_i^j \geq 0)$.

There are at least three reasons why this test statistic might not be asymptotically symmetric. First, the idiosyncratic shock might be heteroskedastic. Ferman and Pinto (2016) show that this would usually be true if we have unit x time aggregate values when there is variation in the number of observations per unit.[15] This would be the case, for example, if one uses the Current Population Survey (CPS). Note that, in this case, t_i^{post} would tend to attain higher values when the treated unit is small relative to the units in the donor pool. Second, even if the idiosyncratic error is homoskedastic, the variance of $\epsilon_{it} - \sum_{j \neq i} {w^*}_i^j \epsilon_{jt}$ will depend on the weights $\{{w^*}_i^j\}_{j \neq i}$. If the weights for unit i are more concentrated around a few units in the donor pool, then the variance of t_i^{post} should be higher than if the weights were more evenly distributed. Finally, t_i^{post} would not have the same distribution as t_1^{post} if assumption 1 is not valid for unit i or if the SC weights converge in probability to weights that do not satisfy assumption 1. In this case, we would have that $y_{it} - \sum_{j \neq i} {w^*}_i^j y_{jt} \xrightarrow{d} \epsilon_{it} - \sum_{j \neq i} \tilde{w}_i^j \epsilon_{jt} + \lambda_t \left(\mu_i - \sum_{j \neq i} \tilde{w}_i^j \mu_j \right)$.

Abadie et al. (2010) correctly noticed that the outcome variable may not be well reproduced for some units by a convex combination of the other units for the pre-intervention periods, and that the post-intervention MSPE for these units should be high as well. For this reason, they exclude permutations in which the the pre-intervention MSPE was 20 times (or 5 times) larger than the pre-intervention MSPE for the treated unit. Note that, if assumption 1 is satisfied and $\hat{\mathbf{w}}_i \xrightarrow{p} \mathbf{w}_i^* \in \Phi_i$ for all i, then the prediction error would converge to $\epsilon_{it} - \sum_{j \neq i} {w^*}_i^j \epsilon_{jt}$ as $T_0 \to \infty$ whether time t is pre- or post-intervention. Therefore, assuming that ϵ_{it} is stationary, then it would be likely that t_i^{post} has the same distribution as t_1^{post} if the pre-intervention MSPE

[15] Note that Xu (2016) assumes that the transitory shock is homoscedastic in his generalized SC method.

for unit i and unit 1 are similar. In this case, if we could consider only permutations with the same pre-intervention MSPE, then this permutation test would be asymptotically symmetric. The problem is that this procedure would usually lead to few permutations to construct the test.[16] Note, however, that Abadie et al. (2010) procedure only excludes permutations with pre-intervention MSPE *higher* than the pre-intervention MSPE for the treated unit. Therefore, if there are many permutations with lower pre-intervention MSPE, then the test would over-reject the null since t_1^{post} would tend to attain larger values. In this case, Abadie et al. (2010) graphical analysis could be misleading, even if the SC weights converge in probability to weights that satisfy assumption 1 for all units.

A second inference procedure suggested by Abadie et al. (2010) is a permutation test using the ratio of post/pre-intervention MSPE. According to them, *"the main advantage of looking at ratios is that it obviates choosing a cut-off for the exclusion of ill-fitting placebo runs"*. Ando and Sävje (2013) argue that the distribution of this test statistic would not have the same distribution for all permutations. However, they do not consider the asymptotic distribution when $T_0 \to \infty$. We again consider whether this test statistic is asymptotically symmetric, so that this procedure would provide a valid hypothesis testing in Canay et al. (2014) setting. Assuming that, for all i, the SC weights converge in probability to weights that satisfy assumption 1, then with $T_0 \to \infty$ and $T - T_0$ fixed:

$$t_i^{\text{ratio}} = \frac{\frac{1}{T-T_0}\sum_{t=T_0+1}^{T}\left[y_{it} - \sum_{j\neq i}\hat{w}_i^j y_{jt}\right]^2}{\frac{1}{T_0}\sum_{t=1}^{T_0}\left[y_{it} - \sum_{j\neq i}\hat{w}_i^j y_{jt}\right]^2} \xrightarrow{d} \frac{1}{T-T_0}\sum_{t=T_0+1}^{T}\left[\frac{\epsilon_{it} - \sum_{j\neq i} {w^*}_i^j \epsilon_{jt}}{\sqrt{var(\epsilon_{it} - \sum_{j\neq i} {w^*}_i^j \epsilon_{jt})}}\right]^2 \quad (20)$$

Therefore, assuming that the transitory shocks are independent of the treatment assignment, that they are serially uncorrelated, and that linear combinations of the transitory shocks are i.i.d. up to a scale parameter, then the test statistic t_i^{ratio} would be asymptotically symmetric.

An important limitation of this result is that we need to assume that, for all i, the SC weights converge in probability to weights that satisfy assumption 1. However, as we show in Section 2.2, the SC weighs will not converge, in general, to weights that satisfy assumption 1 even if there exist weights such that this assumption is satisfied for all i. In this case, we have that:

$$t_i^{\text{ratio}} \xrightarrow{d} \frac{1}{T-T_0}\sum_{t=T_0+1}^{T}\left[\frac{\epsilon_{it} - \sum_{j\neq i}\bar{w}_i^j\epsilon_{jt} + \lambda_t(\mu_i - \sum_{j\neq i}\bar{w}_i^j\mu_j)}{\sqrt{var(\epsilon_{it} - \sum_{j\neq i}\bar{w}_i^j\epsilon_{jt} + \lambda_t(\mu_i - \sum_{j\neq i}\bar{w}_i^j\mu_j))}}\right]^2 \quad (21)$$

[16] Note that the test would remain valid. However, the test would likely have poor power since we would have to rely on randomization in case of ties.

where $\widehat{w}_i^j \rightarrow_p \bar{w}_i^j$.

There are at least two reasons why this test statistic might not be asymptotically symmetric. First, if treatment assignment is correlated with the unobserved heterogeneity, then the distribution of unit 1 would be differ from the distribution of the control units. Also, even if treatment assignment is uncorrelated with the unobserved heterogeneity, the test statistic t_i^{ratio} might still depend on i. While, in this case, $E[t_i^{\text{ratio}}] = 1$ for all i, if the common factors are serially correlated, the variance of the test statistic might depend on characteristics of the unit i. The reason is that the variance of t_i^{ratio} would depend on the serial correlation of λ_t, so dividing all terms in the numerator by the variance of the prediction error would not necessarily adjust so that all test statistics have the same asymptotic variance. More specifically, if the variance of the transitory shock of unit i is higher, then the variance of the t-statistic should be lower. Therefore, one would (over-) under-reject the null if the variance of the transitory shock of the treated unit is (lower) higher.

Note, however, that for the distribution of t_i^{ratio} to be significantly different depending on i when the SC estimator is asymptotically unbiased, it must be that, at the same time, the SC weights are different from weights that reconstruct the factor loadings of the treated unit and the variance of the transitory shocks are not much higher than the serial correlation of the common factors. However, the SC weights will be (asymptotically) close to weights that reconstruct the factor loadings of the treated unit if the variance of the transitory shocks are small.

5 A Particular Class of a Linear Factor Models

5.1 Asymptotic Results

We consider in detail the implications of our results for a particular class of linear factor models in which all units are divided into groups that follow different times trends.[17] More specifically, we consider that the $J+1$ units are divided into K groups, where for each j we have that:

$$y_{jt}^C = \delta_t + \lambda_t^k + \epsilon_{jt} \tag{22}$$

for some $k = 1, ..., K$.

Consider first an extreme case in which $K = 2$, so the first half of the $J+1$ units follow the parallel trend given by λ_t^1, while the other half follow the parallel trend given by λ_t^2. Assume that $var(\lambda_t^k) = 1$

[17]Monte Carlo simulations using this model was studied in detail in Ferman et al. (2016).

and $var(\epsilon_{jt}) = \sigma_\epsilon^2$. We calculate, for this particular class of linear factor models, the asymptotic proportion of misallocated weights of the SC estimator using all pre-treatment lags as economic predictors. From the minimization problem 4, we have that, when $T_0 \to \infty$, the proportion of misallocated weights converges to:

$$\gamma_2(\sigma_\epsilon^2, J) = \sum_{j=\frac{J+1}{2}+1}^{J+1} \bar{w}_1^j = \frac{J+1}{J^2 + 2 \times J \times \sigma_\epsilon^2 - 1} \times \sigma_\epsilon^2 \tag{23}$$

where $\gamma_K(\sigma_\epsilon^2, J)$ is the proportion of misallocated weights when the $J+1$ groups are divided in K groups.

We present in Figure 1.A the relationship between misallocation of weights, variance of the transitory shocks, and number of control units. Note that, for a fixed J, the proportion of misallocated weights converges to zero when $\sigma_\epsilon^2 = 0$, while this proportion converges to $\frac{J+1}{2J}$ (the proportion of misallocated weights of DID) when $\sigma_\epsilon^2 \to \infty$. This is consistent with the results we have in Section 2.2. Moreover, note that, for a given σ_ϵ^2, the proportion of misallocated weights converges to zero when the number of control units goes to infinity. This is consistent with Gobillon and Magnac (2013), who derive support conditions so that the assumptions in Abadie et al. (2010) for unbiasedness are satisfied.

Note that, in this example, the SC estimator converges to:

$$\hat{\alpha}_{1t} \xrightarrow{d} \alpha_{1t} + \left(\epsilon_{1t} - \sum_{j \neq 1} \bar{w}_1^j \epsilon_{jt} \right) + \lambda_t^1 \times \gamma_2(\sigma_\epsilon^2, J) - \lambda_t^2 \times \gamma_2(\sigma_\epsilon^2, J) \tag{24}$$

Therefore, if $E[\lambda_t^1 | d_{1t} = 1] = 1$ (that is, the expected value of the common factor associated to the treated unit is one standard deviation higher), then the bias of the SC estimators in terms of the standard deviation of y_{1t} would be given by $\frac{\gamma_2(\sigma_\epsilon^2, J)}{\sqrt{1+\sigma_\epsilon^2}}$. Therefore, while a higher σ_ϵ^2 increases the misallocation of weights, the importance of this misallocation in terms of bias of the SC estimator is limited by the fact that the common factor (which we allow to be correlated with treatment assignment) becomes less relevant. We present the asymptotic bias of the SC estimator as a function of σ_ϵ^2 and J in Figure 1.B. Note that, if $J+1 \geq 20$, then the bias of the SC estimator will always be lower than 0.1 standard deviations of y_{1t} when treatment assignment is associated with a one standard deviation increase in λ_t^1. This happens because, in this model, the misallocation of weights diminishes when the number of control groups increases.

We consider now the another extreme case in which the $J+1$ units are divided into $K = \frac{J+1}{2}$ groups that follow the same parallel trend. In other words, in this case each unit has a pair that follows its same parallel trend, while all other units follow different parallel trends. The proportion of misallocated weights converges to:

$$\gamma_{\frac{J+1}{2}}(\sigma_\epsilon^2, J) = \sum_{j=2}^{J+1} \bar{w}_1^j = \frac{J-1}{2+\sigma_\epsilon^2+(1+\sigma_\epsilon^2)(J-1)} \times \sigma_\epsilon^2 \tag{25}$$

We present the relationship between misallocation of weights, variance of the transitory shocks, and number of control units in Figure 1.C. Note that, again, the proportion of misallocated weights converges to zero when $\sigma_\epsilon^2 = 0$ and to the proportion of misallocated weights of DID when $\sigma_\epsilon^2 \to \infty$ (in this case, $\frac{J-1}{J}$). Differently from the previous case, however, for a given σ_ϵ^2, the proportion of misallocated weights converges to $\frac{\sigma_\epsilon^2}{1+\sigma_\epsilon^2}$ when $J \to \infty$. Therefore, the SC estimator would remain asymptotically biased even when the number of control units is large. This happens because, in this model, the number of common factors increases with J, so the conditions derived in Gobillon and Magnac (2013) are not satisfied. As presented in Figure 1.D, in this case, the asymptotic bias can be substantially higher, and it does not vanishes when the number of control units increases. Therefore, the asymptotic bias of the SC estimator can be relevant even when the number of control units increases.

Finally, note that, in both cases, the proportion of misallocated weights is always lower than the proportion of misallocated weights of DID. Therefore, in this particular class of linear factor models, the asymptotic bias of the SC estimator will always be lower than the asymptotic bias of DID. However, this is not a general result, as we show in Appendix A.1.

5.2 Monte Carlo Simulations

We present Monte Carlo (MC) simulation results using a data generating process (DGP) based on 22. We consider in our MC simulations $J+1 = 20$, λ_t^k is normally distributed following an AR(1) process with 0.5 serial correlation parameter, $\epsilon_{jt} \sim N(0, \sigma_\epsilon^2)$, and $T - T_0 = 10$. We also impose that there is no treatment effect, i.e., $y_{jt} = y_{jt}^C = y_{jt}^T$ for each time period $t \in \{1, ..., T\}$. We consider variations DGP in the following dimensions:

- The number of pre-intervention periods: $T_0 \in \{20, 100, 1000\}$.
- The variance of the transitory shocks: $\sigma_\epsilon^2 \in \{0.5, 1, 2\}$.
- The number of groups with different common factors: $K = 2$ (2 groups of 10) or $K = 10$ (10 groups of 2)

We present in Table 1 the proportion of misallocated weights of the SC estimator for different values of T_0, σ_ϵ^2, and K. The MC results corroborate our theoretical results presented in Section 5.1 that the proportion of misallocated weights will be higher when σ_ϵ^2 is higher and when K is higher. With a smaller T_0 the proportion of misallocated weights of the SC estimator is slightly higher.[18]

We consider in Table 2 the permutation test proposed in Abadie et al. (2010) when the SC estimator is asymptotically biased. We modify the DGP so that $\tilde{\lambda}_t^1 = \lambda_t^1 + 1$ if $t > T_0$. The SC estimator will be asymptotically biased because the SC unit will assign positive weight to units that follow different parallel trends. Therefore, in this case, the asymptotic bias would be given by $\gamma_K(\sigma_\epsilon^2, 19)$, which is the proportion of misallocated weights. When $K = 10$, the permutation test over-rejects with 8-9% probability. This was expected, because the test statistic would have a higher expected value for the treated unit (and also for the control unit that follows parallel trend λ_t^1). Rejection rates are close to 5% when $K = 2$. However, this is a very particular case in which the asymptotic bias when we consider a unit that follows parallel trend λ_t^1 is $\gamma_K(\sigma_\epsilon^2, 19)$ while the asymptotic bias when we consider a unit that follows parallel trend λ_t^2 is $-\gamma_K(\sigma_\epsilon^2, 19)$. Therefore, the expected value of the test statistic for all units will be the same.

As seen in Section 4, the permutation test proposed in Abadie et al. (2010) might not be asymptotically symmetric even if the SC estimator is asymptotically unbiased. This would be the case if the SC weights do not converge in probability to weights that satisfy assumption 1 and common factors are serially correlated. We now explore in Monte Carlo simulations the extent to which heteroskedasticity might generate invalid *conditional* hypothesis testing. We modify our DGP so that we have heteroskedasticity. Now all units will be randomly allocated to have either $\sigma_\epsilon^2 = 0.5$ or $\sigma_\epsilon^2 = 2$.[19] Note that, unconditionally, rejection rates will be exactly equal to 5% for a 5% test. However, given our discussion in Section 4, the test might not have the correct size conditional on the variance of the treated unit.

We present in Table 3 the difference in rejection rates when the treated unit has $\sigma_\epsilon^2 = 2$ versus when it has $\sigma_\epsilon^2 = 0.5$. In column 1 we present simulation results when λ_t^k is serially uncorrelated, while in column 2 we present results when λ_t^k follows an AR(1) process with serial correlation equals to 0.9. With large T_0, $K = 10$, and λ_t^k serially correlated, the test slightly (over-) under-rejects the null when the variance of the treated unit is (lower) higher. Still, the rejection rates are close to 5% (4.5% when $\sigma_\epsilon^2 = 2$ and 5.5% when

[18] Ferman et al. (2016) show that the misallocation of weights is even higher with the specification that uses the average of the pre-treatment outcome as economic predictor.

[19] One example would be if we have a common shock that affects all individuals equally (λ_t^1 or λ_t^2) and transitory shocks that are averages of many (in this case, $\sigma_\epsilon^2 = 0.5$) or few ($\sigma_\epsilon^2 = 2$) individual observations, depending on the size of unit i, as analyzed in Ferman and Pinto (2016) in the context of differences-in-differences. Note that the proportion of misallocated weights should not depend (asymptotically) on whether the variance of the transitory shocks of the treated unit is higher or lower.

$\sigma_\epsilon^2 = 0.5$). This is consistent with the discussion in Section 4 that, asymptotically, the distortions of the permutation test would be limited by the fact that it depends on both a high proportion of misallocated weights and the serial correlation of λ_t^k being relevant relative to the variance of ϵ_{jt}.

With small T_0, however, our simulation results suggest that the size distortion can actually be relevant even if the common factors are serially uncorrelated. With a finite T_0, the distribution of t_i^{ratio} is given by:

$$t_i^{\text{ratio}} = \frac{\frac{1}{T-T_0}\sum_{t=T_0+1}^{T}\left[\epsilon_{it} - \sum_{j\neq i}\widehat{w}_i^j\epsilon_{jt} + \lambda_t(\mu_i - \sum_{j\neq i}\widehat{w}_i^j\mu_j)\right]^2}{\frac{1}{T_0}\sum_{t=1}^{T_0}\left[\epsilon_{it} - \sum_{j\neq i}\widehat{w}_i^j\epsilon_{jt} + \lambda_t(\mu_i - \sum_{j\neq i}\widehat{w}_i^j\mu_j)\right]^2} \tag{26}$$

While both numerator and denominator of the test statistic depend on a linear combination of common and transitory shocks, the weights $\widehat{w}_i^j$ are chosen as to minimize the denominator. If T_0 is not large enough relative to J, we might “over-fit” the model. As a consequence, the denominator (in-sample prediction error) would not provide an adequate correction for the variance of the numerator (out-of-sample prediction error), so the conditional distribution of the test statistic would depend on the variance of the treated unit.

One possible solution to this problem is to use pre-treatment periods not used in the estimation of the SC weights in the denominator. We show rejection rates using this modified test statistic on Table 4. We leave out the last 10 periods prior to T_0 from the minimization problem that estimates the SC weights, and calculate the test statistic using only these periods not used in the minimization problem to calculate the pre-intervention MSPE.[20] With $T_0 = 100$, the rejection rates become much less sensitive to the heteroskedasticity. However, since the variance of the denominator will be higher, it is likely that this modified test statistic might provide a test with low power. Another possible solution might be to avoid over-fitting using a different method to estimate the SC weights that takes into account the fact that the number of parameters might be large relative to the number of pre-treatment periods. Doudchenko and Imbens (2016) consider the use of regularization methods such as best subset regression or LASSO to estimate the SC weights. Note that, while these solutions might circumvent the over-fitting problem, they might not solve the problem that the SC weights generally converge to weights that do not reconstruct the factor loadings of the treated unit.

[20] We do not include the case with $T_0 = 20$ because it was not possible to estimate the 19 SC weights using only 10 pre-treatment periods as economic predictors.

6 Conclusion

In this paper, we revisit the theory behind the SC model. We show that, in general, the SC estimator will be asymptotically biased if selection into treatment depends on unobserved heterogeneity. This happens because the SC weights used to construct the SC unit will generically not converge to weights that satisfy the identification assumptions of the method. The magnitude of the bias only vanishes when the variance of the idiosyncratic errors goes to zero. We also show that this can be particularly problematic when one considers the specification that uses the average of the pre-treatment outcome values as economic predictor instead of all pre-intervention outcome lags to estimate the weights. Overall, we show that there are significant subtleties in the application of the SC method that are usually overlooked in SC applications. Therefore, researchers should be more careful about the relevant assumptions when using this method. In particular, researchers should consider that different SC specifications rely on widely different assumptions that might not be adequate depending on the setting.

We recommend a slight modification in the SC method which is to demean the data using the pre-intervention period. In this case, if selection into treatment is only correlated with a time-invariant common factor (which is essentially the identification assumption of the DID model), then this demeaned SC estimator will be asymptotically unbiased. Also, in this case we can guarantee that the demeaned SC estimator will have an (asymptotically) lower variance than the DID estimator. If treatment selection is correlated with time-varying common factors, then both the demeaned SC and the DID estimators would be asymptotically biased. In this case, it is likely that the demeaned SC estimator would be less biased than the DID estimator. However, it is possible to provide examples in which the demeaned SC estimator will be more biased. We also propose an alternative way to estimate the SC weights in which we use lags of the control units as instrumental variables. However, this approach requires additional assumptions on the common factors and transitory shocks.

Finally, we consider the implications of our findings to the permutation test proposed in Abadie et al. (2010). We show that, if the SC weights do not converge to weights that satisfy the SC assumptions, then it is not possible to guarantee that the test statistic will have asymptotically the same distribution in all permutations. However, the permutation test should (asymptotically) provide reasonable inference if the SC estimator is asymptotically unbiased.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program,” *Journal of the American Statiscal Association*, 2010, *105* (490), 493–505.

_ , _ , and _ , “Comparative Politics and the Synthetic Control Method,” *American Journal of Political Science*, 2015, *59* (2), 495–510.

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

Acemoglu, Daron, Simon Johnson, Amir Kermani, James Kwak, and Todd Mitton, “The Value of Connections in Turbulent Times: Evidence from the United States,” December 2013. NBER Working Paper 19701.

Ando, Michihito, “Dreams of Urbanization: Quantitative Case Studies on the Local Impacts of Nuclear Power Facilities using the Synthetic Control Method,” *Journal of Urban Economics*, June 2015, *85*, 68–85.

_ and Fredrik Sävje, “Hypothesis Testing with the Synthetic Control Method,” 2013. Working Paper.

Athey, Susan and Guido Imbens, “The State of Applied Econometrics-Causality and Policy Evaluation,” *arXiv preprint arXiv:1607.00699*, 2016.

Barone, Guglielmo and Sauro Mocetti, “Natural Disasters, Growth and Institutions: a Tale of Two Earthquakes,” *Journal of Urban Economics*, 2014, pp. 52–66.

Bauhoff, Sebastian, “The Effect of School Nutrition Policies on Dietary Intake and Overweight: a Synthetic Control Approach,” *Economics and Human Biology*, 2014, pp. 45–55.

Belot, Michele and Vincent Vandenberghe, “Evaluating the Threat Effects of Grade Repetition: Exploiting the 2001 Reform by the French-Speaking Community of Belgium,” *Education Economics*, 2014, *22* (1), 73–89.

Billmeier, Andreas and Tommaso Nannicini, “Trade Openness and Growth: Pursuing Empirical Glasnost,” *IMF Staff Papers*, 2009, *56* (3), 447–475.

_ and _ , “Assessing Economic Liberalization Episodes: A Synthetic Control Approach,” *The Review of Economics and Statistics*, 2013, *95* (3), 983–1001.

Bohn, Sarah, Magnus Lofstrom, and Steven Raphael, "Did the 2007 Legal Arizona Workers Act Reduce the State's Unauthorized Immigrant Population?," *The Review of Economics and Statistics*, 2014, *96* (2), 258–269.

Bove, Vincenzo, Leandro Elia, and Ron P. Smith, "The Relationship between Panel and Synthetic Control Estimators on the Effect of Civil War," October 2014. Working Paper.

Calderon, Gabriela, "The Effects of Child Care Provision in Mexico," July 2014. Working paper.

Canay, Ivan A., Joseph P. Romano, and Azeem M. Shaikh, "Randomization Tests under an Approximate Symmetry Assumption?," 2014.

Carrasco, Vinicius, Joao M. P. de Mello, and Isabel Duarte, "A Década Perdida: 2003 – 2012," 2014. Texto para Discussão,.

Carvalho, Carlos V., Ricardo Mansini, and Marcelo C. Medeiros, "ArCo: An Artificial Counterfactual Approach for Aggregate Data," February 2015. Working Paper.

Cavallo, Eduardo, Sebastian Galiani, Ilan Noy, and Juan Pantano, "Catastrophic Natural Disasters and Economic Growth," *The Review of Economics and Statistics*, 2013, *95* (5), 1549–1561.

Chan, Ho Fai, Bruno S. Frey, Jana Gallus, and Benno Torgler, "Academic Honors and Performance," *Labour Economics*, 2014, *31*, 188–204.

Chernozhukov, Victor, Han Hong, and Elie Tamer, "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 2007, *75* (5), 1243–1284.

Coffman, Makena and Ilan Noy, "Hurricane Iniki: Measuring the Long-Term Economic Impact of Natural Disaster Using Synthetic Control," *Environment and Development Economics*, 2011, *17*, 187–205.

de Souza, Fernando Friaça Asmar, "Tax Evasion and Inflation: Evidence from the Nota Fiscal Paulista Program," Master's thesis, Pontifícia Universidade Católica March 2014.

Dhungana, Sandesh, "Identifying and Evaluating Large Scale Policy Interventions: What Questions Can We Answer?," December 2011.

Doudchenko, Nikolay and Guido Imbens, "Balancing, regression, difference-in-differences and synthetic control methods: A synthesis," 2016.

DuPont, William and Ilan Noy, "What Happened to Kobe? A Reassessment of the Impact of the 1995 Earthquake in Japan," March 2012.

Ferman, Bruno and Cristine Pinto, "Inference in Differences-in-Differences with Different Group Sizes," 2016. Working Paper.

_ , _ , and Vitor Possebom, "Cherry Picking with Synthetic Controls," 2016. Working Paper.

Firpo, Sergio and Vitor Possebom, "Synthetic Control Estimator: A Generalized Inference Procedure and Confidence Sets," April 2016.

Gathani, Sachin, Massimiliano Santini, and Dimitri Stoelinga, "Innovative Techniques to Evaluate the Impacts of Private Sector Developments Reforms: An Application to Rwanda and 11 other Countries." Working Paper.

Gobillon, Laurent and Thierry Magnac, "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," PSE Working Papers halshs-00849071, HAL July 2013.

_ and _ , "Regional Policy Evaluation: Interative Fixed Effects and Synthetic Controls," *Review of Economics and Statistics*, 2016. Forthcoming.

Hinrichs, Peter, "The Effects of Affirmative Action Bans on College Enrollment, Educational Attainment, and the Demographic Composition of Universities," *Review of Economics and Statistics*, March 2012, *94* (3), 712–722.

Hosny, Amr Sadek, "Algeria's Trade with GAFTA Countries: A Synthetic Control Approach," *Transition Studies Review*, 2012, *19*, 35–42.

Jinjarak, Yothin, Ilan Noy, and Huanhuan Zheng, "Capital Controls in Brazil — Stemming a Tide with a Signal?," *Journal of Banking & Finance*, 2013, *37*, 2938–2952.

Kaul, Ashok, Stefan Klöbner, Gregor Pfeifer, and Manuel Schieler, "Synthetic Control Methods: Never Use All Pre-Intervention Outcomes as Economic Predictors," May 2015. Working Paper.

Kirkpatrick, A. Justin and Lori S. Bennear, "Promoting Clean Enery Investment: an Empirical Analysis of Property Assessed Clean Energy," *Journal of Environmental Economics and Management*, 2014, *68*, 357–375.

Kleven, Henrik Jacobsen, Camille Landais, and Emmanuel Saez, "Taxation and International Migration of Superstars: Evidence from European Football Market," *American Economic Review*, 2013, *103* (5), 1892–1924.

Kreif, Noémi, Richard Grieve, Dominik Hangartner, Alex James Turner, Silviya Nikolova, and Matt Sutton, "Examination of the Synthetic Control Method for Evaluating Health Policies with Multiple Treated Units," *Health Economics*, 2015.

Li, Qi, "Economics Consequences of Civil Wars in the Post-World War II Period," *The Macrotheme Review*, 2012, *1* (1), 50–60.

Liu, Shimeng, "Spillovers from Universities: Evidence from the Land-Grant Program," *Journal of Urban Economics*, 2015, *87*, 25–41.

Mideksa, Torben K., "The Economic Impact of Natural Resources," *Journal of Environmental Economics and Management*, 2013, *65*, 277–289.

Montalvo, José G., "Voting after the Bombings: A Natural Experiment on the Effect of Terrorist Attacks on Democratic Elections," *Review of Economics and Statistics*, 2011, *93* (4), 1146–1154.

Newey, Whitney K. and Daniel McFadden, "Chapter 36 Large sample estimation and hypothesis testing," in "in," Vol. 4 of *Handbook of Econometrics*, Elsevier, 1994, pp. 2111 – 2245.

Pinotti, Paolo, "Organized Crime, Violence and the Quality of Politicians: Evidence from Southern Italy," 2012.

_ , "The Economic Costs of Organized Crime: Evidence from Southern Italy," April 2012. Temi di Discussione (Working Papers).

Powell, David, "Synthetic Control Estimation Beyond Case Studies: Does the Minimum Wage Reduce Employment?," 2016. RAND Corporation, WR-1142.

Ribeiro, Felipe, Guilherme Stein, and Thomas Kang, "The Cuban Experiment: Measuring the Role of the 1959 Revolution on Economic Performance using Synthetic Control," May 2013.

Sanso-Navarro, Marcos, "The effects on American Foreign Direct Investment in the United Kingdom from Not Adopting the Euro," *Journal of Common Markets Studies*, 2011, *49* (2), 463–483.

Saunders, Jessica, Russel Lundberg, Anthony A. Braga, Greg Ridgeway, and Jeremy Miles, "A Synthetic Control Approach to Evaluating Place-Based Crime Interventions," *Journal of Quantitative Criminology*, 2014.

Severnini, Edson R., "The Power of Hydroelectric Dams: Agglomeration Spillovers," March 2014. IZA Discussion Paper, No. 8082.

Sills, Erin O., Diego Herrera, A. Justin Kirkpatrick, Amintas Brandao, Rebecca Dickson, Simon Hall, Subhrendu Pattanayak, David Shoch, Mariana Vedoveto, Luisa Young, and Alexander Pfaff, "Estimating the Impact of a Local Policy Innovation: The Synthetic Control Method Applied to Tropica Desforestation," *PLOS One*, 2015.

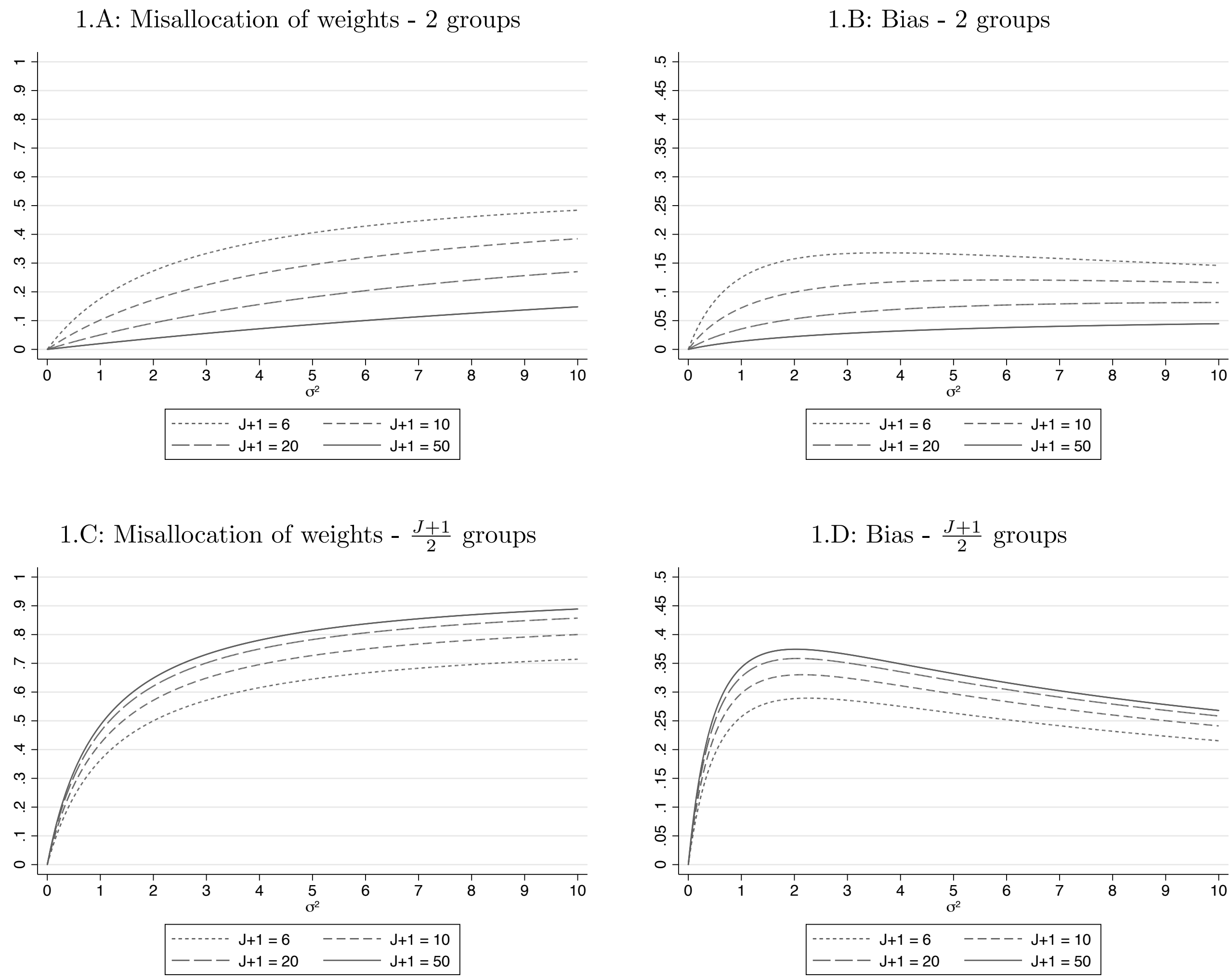
Smith, Brock, "The Resource Curse Exorcised: Evidence from a Panel of Countries," *Journal of Development Economics*, 2015, *116*, 57–73.

Wong, Laurence, "Three Essays in Causal Inference." PhD dissertation, Stanford University March 2015.

Xu, Yiqing, "Regional Policy Evaluation: Interative Fixed Effects and Synthetic Controls," *Political Analysis*, 2016. Forthcoming.

Yu, Jingwen and Chunchao Wang, "Political Risk and Economic Development: A Case Study of China," *Eknomska Istrazianja - Economic Research*, 2013, *26* (2), 35–50.

Figure 1: **Asymptotic Misallocation of Weights and Bias**



Notes: these figures present the asymptotic misallocation of weights and bias of the SC estimator as a function of the variance of the transitory shocks for different numbers of control units. Figures 1.A and 1.B report results when there are 2 groups of $\frac{J+1}{2}$ units each, while figures 1.C and 1.D report results when there are $\frac{J+1}{2}$ groups of 2 units each. The misallocation of weights is defined as the proportion of weight allocated to units that do not belong to the group of treated unit. The bias of the SC estimator is reported in terms of standard deviations of y_{jt} (which is equal to $\sqrt{1+\sigma_\epsilon^2}$) when the expected value of the common factor associated to the treated unit, conditional on this unit being treated, is equal to one standard deviation of the common factor.

Table 1: **Misallocation of weights**

	$\sigma^2_\epsilon = 0.5$	$\sigma^2_\epsilon = 1$	$\sigma^2_\epsilon = 2$
	(1)	(2)	(3)
	Panel A: $K = 2$		
$T_0 = 20$	0.102	0.157	0.233
	[0.000]	[0.001]	[0.001]
$T_0 = 100$	0.059	0.098	0.156
	[0.000]	[0.001]	[0.001]
$T_0 = 1000$	0.034	0.061	0.105
	[0.000]	[0.001]	[0.001]
	Panel B: $K = 10$		
$T_0 = 20$	0.469	0.633	0.770
	[0.001]	[0.001]	[0.001]
$T_0 = 100$	0.350	0.508	0.664
	[0.001]	[0.001]	[0.001]
$T_0 = 1000$	0.310	0.466	0.623
	[0.001]	[0.001]	[0.001]

Notes: this table presents the proportion of misallocated weights in MC simulations of the SC estimator that uses all pre-treatment outcome lags as economic predictors for a given $(T_0, \sigma^2_\epsilon, K)$. In all simulations, we set $J + 1 = 20$. When $K = 2$, the proportion of misallocated weights is given by the sum of weights allocated to units 11 to 20. When $K = 10$, it is given by the sum of weights allocated to units 3 to 20.

Table 2: **Permutation test rejection rates - effects of asymptotic bias**

	$\sigma_\epsilon^2 = 0.5$	$\sigma_\epsilon^2 = 1$	$\sigma_\epsilon^2 = 2$
	(1)	(2)	(3)
	Panel A: $K = 2$		
$T_0 = 20$	0.050	0.050	0.050
	[0.000]	[0.000]	[0.000]
$T_0 = 100$	0.050	0.050	0.050
	[0.000]	[0.000]	[0.000]
$T_0 = 1000$	0.050	0.050	0.050
	[0.000]	[0.000]	[0.000]
	Panel B: $K = 10$		
$T_0 = 20$	0.088	0.087	0.079
	[0.001]	[0.001]	[0.001]
$T_0 = 100$	0.089	0.092	0.087
	[0.001]	[0.001]	[0.001]
$T_0 = 1000$	0.084	0.091	0.091
	[0.001]	[0.001]	[0.001]

Notes: this table presents the rejection rates in MC simulations of a permutation test with the SC estimator that uses all pre-treatment outcome lags as economic predictors for a given $(T_0, \sigma_\epsilon^2, K)$. In all simulations, we set $J + 1 = 20$. We set $\tilde{\lambda}_t^1 = \lambda_t^1 + 1$, so that the SC estimator is asymptotically biased if the SC weights do not reconstruct the factor loadings of the treated unit.

Table 3: **Permutation test rejection rates - effects of heteroskedasticity**

	no serial correlation	serial correlation=0.9
	(1)	(2)
	Panel A: $K = 2$	
$T_0 = 20$	-0.045	-0.047
	[0.000]	[0.000]
$T_0 = 100$	-0.009	-0.010
	[0.000]	[0.000]
$T_0 = 1000$	0.001	0.001
	[0.000]	[0.000]
	Panel B: $K = 10$	
$T_0 = 20$	-0.052	-0.055
	[0.000]	[0.000]
$T_0 = 100$	-0.013	-0.029
	[0.000]	[0.000]
$T_0 = 1000$	0.003	-0.010
	[0.000]	[0.000]

Notes: this table presents the difference in rejection rates when the variance of the treated unit is higher compared to when the variance of the treated unit is lower. In all simulations, we set $J+1 = 20$. Each unit is randomly selected to have $\sigma_\epsilon^2 = 0.5$ or $\sigma_\epsilon^2 = 2$. In column 1, the common factor is serially uncorrelated, while in column 2 it follows an AR(1) process with serial correlation equal to 0.9.

Table 4: **Permutation test rejection rates - modified test statistic**

	no serial correlation	serial correlation=0.9
	(1)	(2)
	Panel A: $K = 2$	
$T_0 = 100$	-0.004	0.001
	[0.000]	[0.000]
$T_0 = 1000$	0.000	0.001
	[0.000]	[0.000]
	Panel B: $K = 10$	
$T_0 = 100$	0.001	-0.010
	[0.000]	[0.000]
$T_0 = 1000$	0.002	-0.005
	[0.000]	[0.000]

Notes: this table replicates the results presented in Table 3 using a modified test statistic. We leave out the last 10 periods prior to T_0 from the minimization problem that estimates the SC weights, and calculate the test statistic using only these periods not used in the minimization problem to calculate the pre-intervention MSPE.

A Supplemental Appendix: Revisiting the Synthetic Control Estimator

A.1 Example: SC Estimator vs DID Estimator

We provide an example in which the asymptotic bias of the SC estimator can higher than the asymptotic bias of the DID estimator. Assume we have 1 treated and 4 control units in a model with 2 common factors. For simplicity, assume that there is no additive fixed effects and that $E[\lambda_t] = 0$. We have that the factor loadings are given by:

$$\mu_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \mu_2 = \begin{pmatrix} 0.5 \\ 1 \end{pmatrix}, \mu_3 = \begin{pmatrix} 1.5 \\ 1 \end{pmatrix}, \mu_4 = \begin{pmatrix} 0.5 \\ 0 \end{pmatrix}, \mu_5 = \begin{pmatrix} 1.5 \\ 1 \end{pmatrix} \tag{27}$$

Note that the linear combination $0.5\mu_2 + w_1^3\mu_3 + w_1^5\mu_5 = \mu_1$ with $w_1^3 + w_1^5 = 0.5$ satisfy assumption 1. Note also that DID equal weights would set the first factor loading to 1, which is equal to μ_1^1, but the second factor loading would be equal to $0.75 \neq \mu_1^2$. We want to show that the SC weights would improve the construction of the second factor loading but it will distort the combination for the first factor loading. If we set $\sigma_\epsilon^2 = E[(\lambda_t^1)^2] = E[(\lambda_t^2)^2] = 1$, then the factor loadings of the SC unit would be given by $(1.038, 0.8458)$. Therefore, there is small loss in the construction of the first factor loading and a gain in the construction of the second factor loading. Therefore, if selection into treatment is correlated with the common shock λ_t^1, then the SC estimator would be more asymptotically biased than the DID estimator.

A.2 Definition: Asymptotically Unbiased

We now show that the expected value of the asymptotic distribution will be the same as the limit of the expected value of the SC estimator. Let γ be the expected value of the asymptotic distribution of $\hat{\alpha}_{1t} - \alpha_{1t}$. Therefore, we have that:

$$\begin{aligned} E[\hat{\alpha}_{1t} - \alpha_{1t}] &= \gamma + E\left[\sum_{j\neq 1}(\bar{w}_1^j - \hat{w}_1^j)\epsilon_{jt}\right] + E\left[\lambda_t \sum_{j\neq 1}(\bar{w}_1^j - \hat{w}_1^j)\mu_j\right] \\ &= \gamma + \sum_{j\neq 1} E\left[(\bar{w}_1^j - \hat{w}_1^j)\epsilon_{jt}\right] + \sum_{j\neq 1} E\left[\lambda_t(\bar{w}_1^j - \hat{w}_1^j)\right]\mu_j \end{aligned}$$

Given that $\hat{w}_1^j$ is a consistent estimator for $\bar{w}_1^j$, if we have that ϵ_{it} has finite variance, then:

$$\left|E\left[(\bar{w}_1^j - \hat{w}_1^j)\epsilon_{jt}\right]\right| \leq E\left[\left|(\bar{w}_1^j - \hat{w}_1^j)\epsilon_{jt}\right|\right] \leq \sqrt{E\left[(\bar{w}_1^j - \hat{w}_1^j)^2\right]E\left[(\epsilon_{jt})^2\right]} \to 0$$

Similarly, if λ_t^f has finite variance for all $f = 1, ..., F$, then $E\left[\lambda_t(\bar{w}_1^j - \hat{w}_1^j)\right]\mu_j \to 0$.

A.3 Minimum Distance Problem

Using the notation of Abadie et al. (2010), the SC weights will solve the following optimization problem:

$$\|X_1 - X_0 W\|_V$$

where $\sum\limits_{j=2}^{J} w_1^j = 1$ and $w_1^j > 0$ for all $j = 2, ..., J$, and

$$X_1 - X_0 W = \begin{bmatrix} Z_1 - \sum\limits_{j \neq 1} w_1^j Z_j \\ \sum\limits_{s=1}^{T_0} k_s^1 Y_{1s} - \sum\limits_{j \neq 1} w_1^j \sum\limits_{s=1}^{T_0} k_s^1 Y_{1s} \\ \vdots \\ \sum\limits_{s=1}^{T_0} k_s^K Y_{1s} - \sum\limits_{j \neq 1} w_1^j \sum\limits_{s=1}^{T_0} k_s^K Y_{1s} \end{bmatrix}$$

We prove the properties of the M-estimator for the weights for the special case in which we use all the pre-treatment periods as predictors. In this case, V becomes the identity matrix, and the optimization problem for this particular case is:

$$\arg\min_{w \in W} \frac{\sum_{t_0=1}^{T} \left[\left(y_{1t} - \sum\limits_{j \neq 1} w_1^j y_{jt} \right)' \left(y_{1t} - \sum\limits_{j \neq 1} w_1^j y_{jt} \right) \right]}{T_0}$$

subject to $\sum\limits_{j=2}^{J} w_1^j = 1$ and $w_1^j > 0$ for all $j = 2, ..., J$.. Define the vector Jx1 $\widehat{\mathbf{w}} \equiv \{\hat{w}_1^j\}_{j \neq 1}$ as the solution of this minimization problem.[21] Using the population model for y_{it}, we can write this optimization problem as:

$$\arg\min_{w \in W} \frac{\sum_{t_0=1}^{T} \left[\left(\epsilon_{1t} - \sum\limits_{j \neq 1} w_1^j \epsilon_{jt} \right) + \lambda_t \left(\mu_{1t} - \sum\limits_{j \neq 1} w_1^j \mu_{jt} \right) \right]^2}{T_0}$$

In order to show the uniform convergence of the objective function, we need to impose assumptions about the stochastic processes of $\{\epsilon_{jt}\}_{t=1}^{T_0}$ and $\{\lambda_t\}_{t=1}^{T_0}$.

Assumption 1: $(\epsilon_{jt}, \lambda_t)'$ is weakly stationary and second moment ergodic.

Lemma 1 *Define* $g(y_{1t}, y_{jt}, w) \equiv \left[\left(\epsilon_{1t} - \sum\limits_{j \neq 1} w_1^j \epsilon_{jt} \right) + \lambda_t \left(\mu_{1t} - \sum\limits_{j \neq 1} w_1^j \mu_{jt} \right) \right]^2$. *Under assumption 1,*

$$\sup_{w \in W} \left\| \frac{1}{T_0} \sum_{t_0=1}^{T} g(y_{1t}, y_{jt}, w) - \mathbb{E}[g(y_{1t}, y_{jt}, w)] \right\| \to_p 0 \tag{28}$$

Proof. Note that $g(y_{1t}, y_{jt}, w)$ is continuous a each set of $\left\{ \widehat{w}_1^j \right\}_{j=2}^{J}$. In addition,

$$\begin{aligned} \| g(y_{1t}, y_{jt}, w) \| &\leq \left\| y_{1t} - \sum_{j=2}^{J} w_1^j y_{jt} \right\|' \left\| y_{1t} - \sum_{j=2}^{J} w_1^j y_{jt} \right\| \\ &\leq C \end{aligned}$$

By lemma 2.4 of Newey and McFadden (1994), we have uniform convergence:

$$\sup_{w \in W} \left\| \frac{1}{T_0} \sum_{t_0=1}^{T} g(y_{1t}, y_{jt}, w) - \mathbb{E}[g(y_{1t}, y_{jt}, w)] \right\| \to_p 0$$

[21] As ???? show, if the number of control units is greater than the number of pre-treatment periods, then the solution to this minimization problem might not be unique. However, since we consider the asymptotics with $T_0 \to \infty$, then we guarantee that, for large enough T_0, the solution will be unique.

■

Now, we need to show that there is one only set of $\mathbf{w}_0 \equiv \left\{w_1^j\right\}_{j=2}^{J}$ that maximizes $\mathbb{E}\left[g\left(y_{1t}, y_{jt}, w\right)\right]$.

$$\arg\min_{w\in W} \mathbb{E}\left[\left[\left(\epsilon_{1t} - \sum_{j\neq 1} w_1^j \epsilon_{jt}\right) + \lambda_t \left(\mu_{1t} - \sum_{j\neq 1} w_1^j \mu_{jt}\right)\right]^2\right]$$

In order to simplify the problem, we impose assumptions about the second moments of $\{\epsilon_{jt}\}_{t=1}^{T_0}$ and $\{\lambda_t\}_{t=1}^{T_0}$.

Assumption 2: ϵ_{jt} is uncorrelated with λ_t for $t = 1, ..., T_0$. In addition, $Var\left[\epsilon_{jt}\right] = \sigma^2$ and $\mathbb{E}\left[\lambda_t' \lambda_t\right] = \Omega$.

Under assumption 2, the population objective function simplifies to:

$$\mathbb{E}\left[g\left(y_{1t}, y_{jt}, w\right)\right] = \sigma^2 \left(1 + \sum_{j\neq 1}\left(w_1^j\right)^2\right) + \left(\mu_{1t} - \sum_{j\neq 1} w_1^j \mu_{jt}\right)' \Omega \left(\mu_{1t} - \sum_{j\neq 1} w_1^j \mu_{jt}\right)$$

Note that the first element of this expression is a constant, and it does not matter for the optimization problem. Except for the constant, we can represent this objective function using matrices. Define $\mathbf{w}$ as a vector $(J \times 1)$ of the weights, $\{w_1^j\}_{j\neq 1}$, μ_1 is a vector $(K \times 1)$ with the factor loadings for the treated units and μ_0 is a matrix $(K \times J)$ that contains the factor loadings for the all the control units, we can write population optimization problem as:

$$\arg\min_{w\in W} \mathbf{w}'\mathbf{w} + \left(\mu_1 - \mu_0 \mathbf{w}\right)' \Omega \left(\mu_1 - \mu_0 \mathbf{w}\right)$$

subject to $W = \{\mathbf{w} : \mathbf{w}'\iota = 1, \mathbf{w} \geq 0\}$, with ι being a vector $(J \times 1)$ of 1's. This is a minimization of a quadratic function in a convex space, and has a unique solution $\mathbf{w}_0$.

Using the results above, we could use the theory about M-estimator to show consistent of $\widehat{\mathbf{w}} \equiv \left\{\widehat{w}_1^j\right\}_{j=2}^{J}$.

Theorem 2 *Under assumptions 1 and 2,* $\widehat{\mathbf{w}} \rightarrow_p \mathbf{w}_0$

Proof. *Using the results of previous lemma and the fact that* $\mathbf{w}_0$ *is the unique maximum of* $Q_0(w) \equiv \mathbb{E}\left[g\left(y_{1t}, y_{jt}, \boldsymbol{w}_0\right)\right]$ *and* W *is compact, we can use Theorem* 2.1 *of Newey and McFadden (1994) to show that* $\widehat{\mathbf{w}} \rightarrow_p \mathbf{w}_0$. ■

A.4 Relation with Powell (2016) and Wong (2015)

In this section of the Appendix, we show how the proofs in Wong (2015) and Powell (2016) differ from our approach.

In the third chapter of his thesis, Wong (2015) shows in Section 3.9 that the SC estimator of the weights is given by:

$$\widehat{\mathbf{w}} - \mathbf{w} = \left((Y'Y)^{-1} - (Y'Y)^{-1}\mathbf{j}(\mathbf{j}'(Y'Y)^{-1}\mathbf{j})^{-1}\mathbf{j}'(Y'Y)^{-1}\right) Y'(\zeta - Y'\mathbf{w}) \tag{29}$$

where ζ is a $(T_0 \times 1)$ vector with the pre-intervention outcomes for the treated group (with elements y_{1t}), while Y is a $(T_0 \times J)$ matrix with the pre-intervention outcomes for the control units (with rows $\mathbf{y}_t'$). Also, let $\mathbf{j}$ be a $(J \times 1)$ vector of ones.

Let $E[y_{1t}] = y_{1t}^*$ and $E[\mathbf{y}_t] = \mathbf{y}_t^*$, so that $y_{1t} = y_{1t}^* + \epsilon_{1t}$ and $\mathbf{y}_t = \mathbf{y}_t^* + \epsilon_t$. The SC assumption in his model states that there exists weights $\mathbf{w}$ such that $y_{1t}^* = \mathbf{y}_t^{*\prime}\mathbf{w}$. Assuming $(y_{1t}, \mathbf{y}_t')$ stationary and ergodic, they show that $\frac{1}{T_0}Y'Y \rightarrow E[\mathbf{y}_t\mathbf{y}_t']$ and $\frac{1}{T_0}Y'(\zeta - Y\mathbf{w}) \rightarrow E[\mathbf{y}_t(y_{1t} - \mathbf{y}_t'\mathbf{w})]$. Wong (2015) argues that $E[\mathbf{y}_t(y_{1t} - \mathbf{y}_t'\mathbf{w})] = 0$. However, we have that:

$$\begin{aligned} E[\mathbf{y}_t(y_{1t} - \mathbf{y}_t'\mathbf{w})] &= E[\mathbf{y}_t y_{1t}] - E[\mathbf{y}_t\mathbf{y}_t'\mathbf{w}] = E[(\mathbf{y}_t^* + \epsilon_t)(y_{1t}^* + \epsilon_{1t})] - E[(\mathbf{y}_t^* + \epsilon_t)(\mathbf{y}_t^* + \epsilon_t)'\mathbf{w}] \\ &= \mathbf{y}_t^* y_{1t}^* - \mathbf{y}_t^*\mathbf{y}_t^{*\prime}\mathbf{w} - E[\epsilon_t\epsilon_t']\mathbf{w} = -E[\epsilon_t\epsilon_t']\mathbf{w} \end{aligned} \tag{30}$$

Therefore, this term will only be equal to zero if $var(\epsilon_t) = 0$, which is exactly the condition we find so that the SC weights would be consistent.

In another article, Powell (2016) proposes a generalization of the SC method where the treatment can be multivalued and more than one unit may be treated. He jointly estimates the treatment effect and the SC weights, and argues that the estimator for the treatment effect is consistent. In Theorem 3.1 of his paper, he argues that the following objective function has a unique minimum at $b = \alpha_0$ (although there might be multiple choices of weights):

$$\Gamma(b, \{w_i^j\}) = E\left[||Y_{it} - D_{it}'b - \sum_{j \neq i} \left(w_i^j (Y_{jt} - D_{jt}'b) \right) || \right] \tag{31}$$

where D_{it} is a $(K \times 1)$ vector of treatment variables and α_0 is the $(K \times 1)$ vector of treatment effects.

We show that this generally will not be the case. For simplicity, we assume that μ_i is fixed and that $\mu_i - \sum_{j \neq i} {w_i^j}^* \mu_j = 0$ for some $\{{w_i^j}^*\}_{j \neq i}$. Therefore:

$$\begin{aligned} \Gamma(b, \{w_i^j\}) &= E\left[\left(\epsilon_i - \sum_{j \neq i} w_i^j \epsilon_j \right)^2 \right] \\ &+ \left(\mu_i - \sum_{j \neq i} w_i^j \mu_j \right)' E[\lambda_t' \lambda_t] \left(\mu_i - \sum_{j \neq i} w_i^j \mu_j \right) \\ &+ (\alpha_0 - b)' \left(D_{it} - \sum_{j \neq i} w_i^j D_{jt} \right) \left(D_{it} - \sum_{j \neq i} w_i^j D_{jt} \right)' (\alpha_0 - b) \\ &+ \left(\mu_i - \sum_{j \neq i} w_i^j \mu_j \right)' cov\left(\lambda_t', \left(D_{it} - \sum_{j \neq i} w_i^j D_{jt} \right)' \right) (\alpha_0 - b) \end{aligned} \tag{32}$$

Note that we can set the second, third, and the forth terms of this objective function equal to zero by setting $w_i^j = {w_i^j}^*$ and $b = \alpha_0$. However, there is a first order gain in moving in the direction of weights that minimize the first term. Therefore, there is a set of parameters $\tilde{w}_i^j \neq {w_i^j}^*$ and $b = \alpha_0$ that attains a lower value than ${w_i^j}^*$ and $b = \alpha_0$ (unless ${w_i^j}^*$ minimizes the first term). Since $b = \alpha_0$ minimizes the objective function conditional on setting $w_i^j = {w_i^j}^*$, then it cannot be that the optimal weights will be given by ${w_i^j}^*$. Let $\tilde{\tilde{w}}_i^j$ be the weights that minimize the objective function. Therefore, $\mu_i - \sum_{j \neq i} \tilde{\tilde{w}}_i^j \mu_j \neq 0$.

Now we consider whether $\tilde{\tilde{w}}_i^j$ and $b = \alpha_0$ can be the solution to the problem. Note that the third term can be set to zero by choosing $b = \alpha_0$. However, if treatment assignment is correlated with λ_t, then we could make the forth term lower than zero. Since the first order effect of moving away from $b = \alpha_0$ on the third term is equal to zero, while we can have a first order gain in the forth term, then α_0 would not be the solution to this minimization problem. Note that $b = \alpha_0$ minimizes this problem if treatment assignment is uncorrelated with the common factors. Again, this is consistent with the results we find that the SC is asymptotically unbiased in this case.

A.5 Demeaned Estimator

In this section, we formalize the alternative SC estimator that we propose in section ?? of the paper. In this new method, before finding the weights, we calculate the pre-treatment average of all units and demean the data. The "within-model" for treatment and control units are, respectively:

$$y_{it}^{C} - \overline{y}_i = \left(\lambda_t - \overline{\lambda}\right)' \mu_i + \left(\epsilon_{it} - \overline{\epsilon}_i\right)$$

$$y_{it}^{T} - \overline{y}_i = \alpha_{it} + \left(y_{it}^{C} - \overline{y}_i\right)$$

where $\overline{y}_i = \frac{1}{T_0}\sum\limits_{t=1}^{T_0} y_{it}$, $\overline{\lambda} = \frac{1}{T_0}\sum\limits_{t=1}^{T_0} \lambda_t$ and $\overline{\epsilon}_i = \frac{1}{T_0}\sum\limits_{t=1}^{T_0} \epsilon_{it}$.

Note that we can write this model as,

$$\widetilde{y}_{it}^{C} = \widetilde{\lambda}_t' \widetilde{\mu}_i + \widetilde{\varepsilon}_{it}$$

$$\widetilde{y}_{it}^{T} = \alpha_{it} + \widetilde{y}_{it}^{C}$$

where $\widetilde{\lambda}_t$ does not include any time-invariant common factor, and $\widetilde{\mu}_i$ does not involve factor loadings associated with a constant common factor. This model is the same as before, but using the demeaned variables. In this case,

$$\widehat{\alpha}_{1t} \to \alpha_{1t} + \left(\widetilde{\epsilon}_{1t} \sum_{j\neq 1} \overline{w}_j^1 \widetilde{\epsilon}_{jt}\right) + \widetilde{\lambda}_t' \left(\widetilde{\mu}_1 \sum_{j\neq 1} \overline{w}_j^1 \widetilde{\mu}_j\right)$$

Under the assumptions of the Difference-in-Difference Model,

$$\mathbb{E}\left[\widetilde{\lambda}_t\right] = 0$$

and

$$\mathbb{E}\left[\widetilde{\epsilon}_{1t} - \sum_{j\neq 1} \overline{w}_j^1 \widetilde{\epsilon}_{jt}\right] = 0$$

In this case, the SC estimator is asymptotically unbiased.

Notation

Variable	Dimension	Description
y_{it}	(1×1)	Outcome for unit i at time t
y_{it}^C	(1×1)	Potential outcome for unit i at time t if not treated
y_{it}^T	(1×1)	Potential outcome for unit i at time t if treated
$\mathbf{Y}_1^P$	$(T_0 \times 1)$	Vector of pre-treatment outcome for the treated
$\mathbf{Y}_0^P$	$(T_0 \times J)$	Matrix of pre-treatment outcome for the controls
$\mathbf{y}_t$	$(J \times 1)$	Vector of outcomes for the controls at time t
Z_i	$(R \times 1)$	Vector of covariates
X_1	$(K \times 1)$	Vector of economic predictors for the treated
X_0	$(K \times J)$	Matrix of economic predictors for the controls
λ_t	$(1 \times F)$	Vector of common factors
Ω	$(F \times F)$	$E[\lambda_t' \lambda_t]$
μ_i	$(F \times 1)$	Vector of factor loadings
μ_0	$(F \times J)$	Matrix of factor loadings for the controls
α_{it}	(1×1)	Treatment effect for unit i at time t
$\mathbf{w}$ or $\{w_1^j\}_{j \neq 1}$	$(J \times 1)$	Vector of weights
$\widehat{\mathbf{w}}$ or $\{\hat{w}_1^j\}_{j \neq 1}$	$(J \times 1)$	M-estimator of weights
$\bar{\mathbf{w}}$ or $\{\bar{w}_1^j\}_{j \neq 1}$	$(J \times 1)$	Probability limit of the M-estimator of weights
ϵ_{it}	(1×1)	Idiosyncratic error for unit i at time t
ϵ_t	$(J \times 1)$	Idiosyncratic error for the control units at time t