A Design-Based Perspective on Synthetic Control Methods*

Lea Bottmer† Guido Imbens‡ Jann Spiess§ Merrill Warnick¶

First version: November 2020
Current version: July 2021

Abstract

Since their introduction in Abadie and Gardeazabal (2003), Synthetic Control (SC) methods have quickly become one of the leading methods for estimating causal effects in observational studies with panel data. Formal discussions often motivate SC methods by the assumption that the potential outcomes were generated by a factor model. Here we study SC methods from a design-based perspective, assuming a model for the selection of the treated unit(s), *e.g.*, random selection as guaranteed in a randomized experiment. We show that SC methods offer benefits even in settings with randomized assignment, and that the design perspective offers new insights into SC methods for observational data. A first insight is that the standard SC estimator is not unbiased under random assignment. We propose a simple modification of the SC estimator that guarantees unbiasedness under random assignment and derive its exact, randomization-based, finite sample variance. We also propose an unbiased estimator for this variance. We show in settings with real data that under random assignment this Modified Unbiased Synthetic Control (MUSC) estimator can have a root mean-squared error that is substantially lower than that of the difference-in-means estimator. We show that such an improvement is weakly guaranteed if the treated period is similar to the other periods, for example, if the treated period was randomly selected. The improvement is most likely to be substantial if the number of pre-treatment periods is large relative to the number of control units.

Keywords: Randomization, Synthetic Controls, Panel Data, Causal Effects

*This research was generously supported by ONR grant N00014-17-1-2131.

†Department of Economics, Stanford University, lbottmer@stanford.edu.

‡Professor of Economics, Graduate School of Business, and Department of Economics, Stanford University, SIEPR, and NBER, imbens@stanford.edu.

§Assistant Professor of Operations, Information & Technology, Graduate School of Business, Stanford University, and SIEPR, jspiess@stanford.edu.

¶Department of Economics, Stanford University, mwarnick@stanford.edu.

1 Introduction

Synthetic Control (SC) methods for estimating causal effects have become commonplace in empirical work in the social sciences since their introduction by Alberto Abadie and coauthors (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015). Typically the properties of the original SC estimator, as well as those of the various modifications that have been proposed subsequently, are studied under model-based assumptions about the distribution of the potential outcomes in the absence of the intervention. A common approach is to assume that the potential outcomes follow a factor model plus noise.

In this article we take a different approach to studying the properties of SC estimators. Instead of taking a model-based approach by making assumptions about the distribution of the potential outcomes, we take a design-based approach by making assumptions about the assignment of the unit/time-period pairs to treatment. In particular, we consider the case with a single treated unit/time-period, with the treated unit selected at random from a set of units. We find that in this setting the original SC estimator is generally biased. We propose a simple modification of the SC estimator, labeled the Modified Unbiased Synthetic Control (MUSC) estimator, which is unbiased under random assignment of the treatment. We also propose a variance estimator that is unbiased for the variance of this estimator.

Studying the properties of SC-type estimators under design-based assumptions serves two distinct purposes. First, it suggests an important role for SC methods in the analysis of data from randomized experiments. Second, it leads to new insights into the properties of SC methods in observational studies.

Let us first consider the role of SC methods in randomized experiments. We show that SC methods are particularly valuable in experimental settings with relatively few units, in the presence of substantial correlation over time and across units. In such experiments, where our design assumptions hold by definition, SC-type methods can have substantially better root-mean-squared-error (RMSE) properties than the standard estimator based on the difference in means by treatment status, where we focus on the average treatment effect on the treated. However, it may be important to maintain the guaranteed unbiasedness that the standard estimator for randomized experiments, the difference in means, enjoys under randomization. The proposed MUSC estimator does so, and combines the typical improvement in terms of RMSE with unbiasedness under randomization.

To illustrate the benefits of the MUSC estimator in experimental settings, we simulate an experiment based on data on average log wages observed across 50 states over 40 years. We randomly select one state to be treated in the last period, and compare (*i*) the difference in means, (*ii*) the standard SC estimator, and (*iii*) our proposed MUSC estimator. There are three key findings, partially reported in Table 1 and expanded on in Section 4. First, the difference-in-means and the new MUSC estimator are unbiased by construction whereas the SC estimator is biased. Although in this example the bias of the SC estimator is modest, the bias need not be small in general. Second, the RMSE is substantially lower for the SC and MUSC estimators relative to the RMSE of the difference-in-means estimator. This shows that there can be considerable gains to using SC methods in randomized experiments, without a need to give up the unbiasedness guaranteed by the randomization. Third, the proposed variance estimators are accurate in this setting for all three estimators.

Table 1: Simulation Experiment Based on CPS Average Log Wage by State and Year

	DiM	SC	MUSC
Bias	0	-0.007	0
Root-mean-squared-error	0.105	0.051	0.048
Average standard error	0.105	0.051	0.048

DiM: Difference in Means estimator, SC: Synthetic Control estimator of Abadie et al. (2010), MUSC: Modified Unibased Synthetic Control Estimator.

The second contribution of the article concerns insights into SC methods in observational studies. These insights fall into four categories. First, we propose a new estimator (the MUSC estimaor) for those settings that come with additional robustness guarantees relative to the previously proposed SC estimators. Second, we develop new approaches to inference. Inference has been a challenge in the SC literature, especially in settings with a single treated unit and only a modest number of control units and time periods. In such cases, standard methods based on resampling units may not be applicable, and inferences based on asymptotic approximations which rely on large numbers of pre-treatment periods (Hahn and Shi, 2016; Chernozhukov et al., 2017) are not guaranteed to have good properties. We show that one set of methods for inference specifically designed for SC settings, the placebo method (*e.g.* Abadie et al., 2010; Doudchenko and Imbens, 2016; Ferman and Pinto, 2017), is also biased. With placebo methods, one estimates the variance by calculating the variance of the SC estimator over the distribution generated by applying the SC estimator to randomly selected control units, with or without the actual treated unit. Such methods can both under-estimate and over-estimate the true variance under our design assumptions. In contrast, our proposed variance estimator is unbiased in finite samples, even with a single treated unit, under the random assignment assumption, irrespective of any autocorrelation in the potential outcomes. Third, the design perspective highlights the importance of the choice of estimand. We define four different average treatment effects, either for the treated unit/period pairs, or averaged over units and or periods, and show how the choice of estimand relates to the assumptions and inferential methods. Fourth, we show that the criterion for choosing the weights has some optimality conditions under random assignment of the treated period.

We discuss extensions to settings with non-constant propensity scores as well as to the case where the estimand is the overall average treatment effect.

The article builds on the general SC literature started by Abadie and Gardeazabal (2003); Abadie et al. (2010, 2015). See Abadie (2019) for a recent survey. Recent work proposing new estimators in this general class include Doudchenko and Imbens (2016); Abadie and L'Hour (2017); Ferman and Pinto (2017); Arkhangelsky et al. (2019); Li (2020); Ben-Michael et al. (2020). The current paper also contributes to the literature on inference for SC estimators, which includes Abadie et al. (2010); Doudchenko and Imbens (2016); Ferman and Pinto (2017); Hahn and Shi (2016); Lei and Candès (2020); Chernozhukov et al. (2017). We also build on the general literature on randomization inference for causal effects, (*e.g.* Neyman, 1990; Imbens and Rubin, 2015; Abadie et al., 2020; Rambachan and Roth, 2020). In particular the discussion on the choice of estimands and its implications for randomization inference in Sekhon and Shem-Tov (2020) is relevant.

2 Set Up

We consider a setting with N units, for which we observe outcomes Y_{it} for T time periods, $i = 1, \ldots, N$, $t = 1, \ldots, T$. There is a binary treatment that varies by units and time periods, denoted by $W_{it} \in \{0, 1\}$, and a pair of potential outcomes $Y_{it}(0)$ and $Y_{it}(1)$ for all unit/period combinations (Rubin, 1974; Imbens and Rubin, 2015). We assume there are no dynamic effects for the time being, so the potential outcomes are indexed only by the contemporaneous treatment. In some of the settings we consider, the dynamic effects would simply change the interpretation of the estimand. There are no restrictions on the time path of the potential outcomes. The $N \times T$ matrices of treatments and potential outcomes are denoted by $\boldsymbol{W}$, $\boldsymbol{Y}(0)$ and $\boldsymbol{Y}(1)$ respectively. Given the treatment the realized/observed outcome matrix is $\boldsymbol{Y}$, with typical element

$$Y_{it} \equiv W_{it}Y_{it}(1) + (1 - W_{it})Y_{it}(0). \tag{2.1}$$

In contrast to most of the literature (with Athey and Imbens, 2018) an exception), we take the potential outcomes $\boldsymbol{Y}(0)$ and $\boldsymbol{Y}(1)$ as fixed in our analysis, and treat the assignment matrix $\boldsymbol{W}$ as stochastic. This in turn makes the realized outcomes $\boldsymbol{Y}$ stochastic.

For much of the discussion we focus on the case with a single treated unit and a single treated period. Many of the insights carry over to the case with a block of treated unit/time-period pairs, and we discuss explicitly the case with multiple treated units in Section 5.1. In the case we focus on with a single treated unit/time-period pair, the $N \times T$ matrix $\boldsymbol{W}$, with typical element $W_{it} \in \{0, 1\}$, satisfies $\sum_{i,t} W_{it} = 1$.

It is useful for our design-based analysis to separate out the assignment mechanism into the selection of the time period treated and the unit treated. For that purpose, exploiting the fact that there is only a single pair (i, t) with $W_{it} = 1$, we write

$$\boldsymbol{W} = \boldsymbol{U}\boldsymbol{V}^\top,$$

where $\boldsymbol{U}$ is an N-vector with typical element $U_i \in \{0, 1\}$ and $\sum_{i=1}^N U_i = 1$, and $\boldsymbol{V}$ is a T-vector with typical element $V_i \in \{0, 1\}$ and $\sum_{t=1}^T V_t = 1$. The V_t and U_i can be defined in terms of $\boldsymbol{W}$: $V_t = \sum_{i=1}^N W_{it}$ and $U_i = \sum_{t=1}^T W_{it}$.

In many cases the treated unit is exposed only in the last period, so $\boldsymbol{V}$ is non-stochastic, with last element equal to one and all other elements equal to zero. We examine this case separately, but additional insights are obtained by considering the more general case where both the treated unit and the treated period are stochastic.

2.1 Estimands

Next we define the estimands we consider in this article. Being precise about estimands will be important for the discussion of bias and variance, as well as clarify the role of various assumptions we introduce. The estimands are defined as averages over elements of the matrix $\boldsymbol{Y}(1) - \boldsymbol{Y}(0)$ with typical element $\tau_{it} \equiv Y_{it}(1) - Y_{it}(0)$. Which elements of this matrix of causal effects we average over may depend on $\boldsymbol{W}$, In that case the estimand may be stochastic. It is useful to write the estimands explicitly as functions of $\boldsymbol{U}$ and $\boldsymbol{V}$ to show that some of the estimands depend only on one component of $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{V}^\top$. For ease of exposition, the dependence of the estimands on the potential outcomes is suppressed in the notation.

The estimand that is the primary focus in this discussion is the causal effect for the single treated unit/time-period:

$$\tau \equiv \tau(\boldsymbol{U}, \boldsymbol{V}) \equiv \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t \Big(Y_{it}(1) - Y_{it}(0) \Big). \tag{2.2}$$

Because there is only a treated single unit/period, this is not an average but a single difference in potential outcomes. For the case with multiple treated units or periods this estimand could be generalized to

$$\tilde{\tau} \equiv \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t \Big(Y_{it}(1) - Y_{it}(0) \Big) \Big/ \left(\sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t \right). \tag{2.3}$$

We are interested in accurate estimation of τ as well as inference. In particular, we focus on two properties of estimators for τ: the (exact finite sample) bias and the variance. We also discuss estimation of the variance.

There are three other estimands that are important to contrast with the primary estimand τ. First, the average effect for all N units in the treated period, the "vertical" effect, which only depends on $\boldsymbol{V}$ and not on $\boldsymbol{U}$:

$$\tau^{\mathrm{V}} \equiv \tau^{\mathrm{V}}(\boldsymbol{V}) \equiv \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} V_t \Big(Y_{it}(1) - Y_{it}(0) \Big),$$

Next, the average effect for the treated unit over all T periods, the "horizontal" effect. This only depends on $\boldsymbol{U}$ and not on $\boldsymbol{V}$:

$$\tau^{\mathrm{H}} \equiv \tau^{\mathrm{H}}(\boldsymbol{U}) \equiv \frac{1}{T} \sum_{i=1}^{N} \sum_{t=1}^{T} U_i \Big(Y_{it}(1) - Y_{it}(0) \Big).$$

Finally, the population average treatment effect, which depends on neither $\boldsymbol{U}$ nor $\boldsymbol{V}$:

$$\tau^{\mathrm{POP}} \equiv \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Big(Y_{it}(1) - Y_{it}(0) \Big).$$

An important conceptual advantage of focusing on τ rather than any of the other three estimands is that it frees us from conceptualizing $Y_{it}(1)$ for unit/time periods other than the unit/time pair that was actually treated. To make this explicit, we can write τ as

$$\tau = \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t Y_{it} - \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t Y_{it}(0),$$

which clarifies that the estimand does not depend on any unobserved $Y_{it}(1)$. This can be important because in some cases it is difficult to give meaning to $Y_{it}(1)$ for units other than the treated unit. For example, in the German reunification application in Abadie et al. (2015), it is difficult to conceptualize $Y_{it}(1)$ for countries other than West Germany: what does it mean for France to unify with East Germany? However, focusing on τ allows us to ignore all unobserved $Y_{it}(1)$ in this application; we only need to consider the value of West German GDP in the absence of the reunification.

2.2 Assumptions

Most of the literature on SC methods relies on *modeling* assumptions on the potential outcomes to derive properties of the proposed methods. For SC methods, the model often takes the form of an R-factor latent-factor model for the control outcome

$$Y_{it}(0) = \sum_{r=1}^{R} \gamma_i \beta_t + \varepsilon_{it},$$

in combination with independence assumptions on the noise components ε_{it} (Abadie et al., 2010; Athey et al., 2017; Amjad et al., 2018; Xu, 2017). Here we focus instead on *design* assumptions about the assignment process, that is, assumptions on the distribution of $\boldsymbol{W}$ (or, equivalently, the distributions of $\boldsymbol{U}$ and $\boldsymbol{V}$). In this approach, we place no restrictions on the potential outcomes, and in fact take those as fixed in the repeated sampling thought experiments. Such design-based, as opposed to model-based, approaches have been used in the general experimental design and program evaluation literature (*e.g.* Fisher, 1937; Neyman, 1990; Imbens and Rubin, 2015; Rosenbaum, 2002; Cunningham, 2018), as well more recently in regression settings (Abadie et al., 2020), and in some panel data analyses (Athey and Imbens, 2018). However, to the best of our knowledge, these methods have not yet been used to analyze the properties of SC estimators.

First, we consider random assignment of the units to the treatment.

Assumption 1. (Random Assignment of Units)

$$\mathrm{pr}(\boldsymbol{U} = \boldsymbol{u}) = \begin{cases} \frac{1}{N} & \text{if} \quad u_i \in \{0,1\}\ \forall i, \quad \sum_{i=1}^{N} u_i = 1, \\ 0 & \text{otherwise.} \end{cases}$$

We can also write this as $\boldsymbol{U} \perp\!\!\!\perp (\boldsymbol{Y}(0), \boldsymbol{Y}(1))$. Some version of this assumption underlies many approaches to inference in synthetic control estimation.

A second assumption we consider is less common. Here we assume that the treated period was randomly selected from the T periods under observation.

Assumption 2. (Random Assignment of Treated Period)

$$\mathrm{pr}(\boldsymbol{V} = \boldsymbol{v}) = \begin{cases} \frac{1}{T} & \text{if} \quad v_t \in \{0,1\}\ \forall t, \quad \sum_{t=1}^{T} v_t = 1, \\ 0 & \text{otherwise.} \end{cases}$$

Although this assumption is not plausible in many cases, as it is often the only last period(s) that are treated, it is useful to consider its implications. An implicit synthetic control assumption is that the within-period relation between outcomes for different units during the pre-treatment periods is similar to that same within-period relationship during the treated periods. Assumption 2 formalizes this.

We also consider the special case, which is common in empirical work, where the last element of $\boldsymbol{V}$ is equal to one and all other elements are equal to zero, so the treatment only occurs in the last period. In this case, the distribution of $\boldsymbol{V}$ is degenerate.

Assumption 3. (Last Period Assignment)

$$\text{pr}(\boldsymbol{V}=\boldsymbol{v})=\begin{cases}1 & \text{if } v_t=0\,\forall t=1,\ldots,T-1, v_T=1,\\ 0 & \text{otherwise.}\end{cases}$$

Though some properties we derive rely on both Assumption 1 and Assumption 2, other results are conditional on $\boldsymbol{V}$ and therefore apply to this common special case.

Most of our discussion concerns the finite-sample case. However, for some results it will be useful to consider large-T approximations. There are also other settings where it may be useful to consider large-N results, but we do not do so here. Large-N fixed-T settings require regularization of the SC weights because there more weights to be estimated than periods to estimate them on, and the properties of the estimators will depend directly on the specific regularization method used (*e.g.,* Abadie and L'Hour, 2017; Doudchenko and Imbens, 2016). For large-T results we consider a stationarity assumption. First define $\boldsymbol{Y}_{\cdot t}(0)$ to be the N vector with typical element $Y_{it}(0)$. Define the averages up to period t of the first and the centered second moment:

$$\hat{\mu}_t=\frac{1}{t}\sum_{s=1}^{t}\boldsymbol{Y}_{\cdot s}(0),\qquad \hat{\Sigma}_t=\frac{1}{t}\sum_{s=1}^{t}\left(\boldsymbol{Y}_{\cdot s}(0)-\hat{\mu}_t\right)\left(\boldsymbol{Y}_{\cdot s}(0)-\hat{\mu}_t\right)^{\top}.$$

Assumption 4. (Large-T Stationarity) *For some finite μ and Σ the sequence of populations indexed by T satisfies, as $T\to\infty$,*

$$\hat{\mu}_T\longrightarrow\mu,\qquad \hat{\Sigma}_T\longrightarrow\Sigma.$$

2.3 Synthetic-Control Type Estimators

In this section we introduce a class of SC type estimators. This class, which we refer to as Generalized Synthetic Control (GSC) estimators, includes the original SC estimator proposed by Abadie et al. (2010), as well as three modifications thereof. It also includes the simple Difference-in-Means (DiM) estimator. We define these estimators for all possible treatment assignment vectors $\boldsymbol{U}$ and $\boldsymbol{V}$, initially for the case where both $\boldsymbol{U}$ and $\boldsymbol{V}$ have a single element equal to one and all other elements equal to zero.

2.3.1 Estimators

We characterize the GSC estimators in terms of a set of weights $\boldsymbol{M}_{ijt}$, indexed by $i = 1, \ldots, N$, $j = 0, \ldots, N$, and $t = 1, \ldots, T$. Given a set of weights $\boldsymbol{M}$, treatment assignments $\boldsymbol{U}$, $\boldsymbol{V}$, and outcomes $\boldsymbol{Y}$, the GSC estimator has the form

$$\hat{\tau}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}) \equiv \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t \left\{ \boldsymbol{M}_{i0t} + \sum_{j=1}^{N} \boldsymbol{M}_{ijt} Y_{jt} \right\}. \tag{2.4}$$

If unit i is treated in period t, the estimator is a linear function of the outcomes for all units for that period. For all estimators we consider, the weights are a function of the baseline outcomes $\boldsymbol{Y}(0)$ during non-treated periods.

The DiM estimator, the average outcome for the treated unit minus the average outcome for all the control units in the same period,

$$\hat{\tau}^{\text{dim}}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}) = \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t Y_{it} - \frac{1}{N-1} \sum_{i=1}^{N} \sum_{t=1}^{T} V_t (1 - U_i) Y_{it}$$

corresponds to the case with weights $\boldsymbol{M}_{i0t} = 0$ and $\boldsymbol{M}_{ijt} = -1/(N-1)$ for $j \neq 0, i$, and $\boldsymbol{M}_{iit} = 1$ for all i.

The other estimators in the GSC class differ in the choice of the weights $\boldsymbol{M}$. There are generally two components to this choice. First, there is a set of possible weights, denoted by $\mathcal{M}$, over which we search for an optimal weight. These sets $\mathcal{M}$ differ between the estimators we consider, but in all cases the sets are non-stochastic and do not depend on either the actual assignment matrix $\boldsymbol{W}$ or the outcome data $\boldsymbol{Y}$ beyond the potential outcomes under the control treatment, $\boldsymbol{Y}(0)$. Second, there is an objective function that defines the chosen weight within the set of possible weights. This objective function is identical for all GSC estimators we consider in the current article.

We consider in this discussion only sets of weights that impose the following three restrictions. First, the weight for the treated unit is equal to one:

$$\boldsymbol{M}_{iit} = 1, \ \ \forall i \in \{1, \ldots, N\}, t \in \{1, \ldots, T\}. \tag{2.5}$$

Later when we discuss settings with multiple treated units we modify this to make this weight equal to one over the number of treated units. Second, the weight for unit j for the prediction of the causal effect for unit i is nonpositive:

$$\boldsymbol{M}_{ijt} \leq 0, \;\; \forall i \in \{1,\ldots,N\}, j \in \{1,\ldots,N\} \setminus \{i\}, t \in \{1,\ldots,T\}. \tag{2.6}$$

The third restriction requires that the weights for all units in the prediction for the causal effect for unit i in period t sum to zero. Because the weight for unit i in this prediction is restricted to be equal to one, this means that the weights for the control units sum to minus one:

$$\sum_{j=1}^{N} \boldsymbol{M}_{ijt} = 0 \; \forall i, t \tag{2.7}$$

Together, these three restrictions give the "adding up" restriction from standard SC: weights on control units sum to one. This new notation for synthetic control settings allows us flexibility in adding new restrictions that enforce unbiasedness under randomization and simplifies variance calculations.

Define $\mathcal{M}^0$ to be the set of possible weights that satisfy these three sets of restrictions:

$$\mathcal{M}^0 = \left\{ \boldsymbol{M} \left| \boldsymbol{M}_{iit} = 1, \forall i \geq 1, t; \boldsymbol{M}_{ijt} \leq 0, \forall i \geq 1, t; \sum_{j=1}^{N} \boldsymbol{M}_{ijt} = 0 \; \forall i \geq 1, t \right. \right\}. \tag{2.8}$$

We consider four estimators in this class, characterized by four sets of possible weights $\mathcal{M} \subset \mathcal{M}_0$, described in section 2.3.3.

2.3.2 The Objective Function

We start with the second component. For a given matrix of outcomes $\boldsymbol{Y}$, and a given set of possible weights $\mathcal{M}$, define the tensor $\boldsymbol{M}(\boldsymbol{Y}, \mathcal{M})$ with elements $\boldsymbol{M}_{ijt}$, for $i = 1,\ldots,N$, $j = 0,\ldots,N$, and $t = 1,\ldots,T$, through the minimization over the weights $\boldsymbol{M}$ within the set of weights $\mathcal{M}$:

$$\boldsymbol{M}(\boldsymbol{Y}, \mathcal{M}) \equiv \underset{\boldsymbol{M} \in \mathcal{M}}{\arg\min} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \sum_{s \neq t} \left(\boldsymbol{M}_{i0t} + \sum_{j=1}^{N} \boldsymbol{M}_{ijt} Y_{js} \right)^2 \right\}. \tag{2.9}$$

For the connection to the original setup for the SC estimator, see Abadie et al. (2010) and Doudchenko and Imbens (2016). This definition of the weights has some important features that drive some of the subsequent results. It is convenient to define the infeasible weights $\boldsymbol{M}^* \equiv \boldsymbol{M}(\boldsymbol{Y}(0), \mathcal{M})$. The reason that it is convenient to work with $\boldsymbol{M}^*$ is that $\boldsymbol{M}^*$ is non-stochastic, as a function of the fixed potential outcomes and the set $\mathcal{M}$. However, because we do not observe all elements of $\boldsymbol{Y}(0)$, we cannot calculate $\boldsymbol{M}^*_{ijt}$ for all i, j, and t. But, for the purpose of estimation we do not need to know every element of $\boldsymbol{M}^*$. In fact, we only need to know $\boldsymbol{M}^*_{ijt}$ for all j for the pair (i, t) such that $U_i = V_t = 1$, and these components are calculable given $\boldsymbol{W}$ and $\boldsymbol{Y}$.

Here we discuss how to motivate the objective function in (2.9). We may wish to choose a matrix $\boldsymbol{M}$ to minimize the expected variance, where the expectation is both over the random period that is treated and the random unit that is treated. The expected variance is

$$\mathbb{E}\mathbb{V}(\hat{\tau}; \mathcal{M}) = \mathbb{E}\left[(\hat{\tau}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}(0); \mathcal{M}) - \tau)^2\right] = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\boldsymbol{M}_{i0t} + \sum_{j=1}^{N}\boldsymbol{M}_{ijt}Y_{jt}(0)\right)^2.$$

Let $\boldsymbol{M}^*$ be the value of $\boldsymbol{M}$ that minimizes this infeasible objective function, $\boldsymbol{M}^* = \arg\min_{\mathcal{M}} \mathbb{E}\mathbb{V}(\hat{\tau}; \mathcal{M})$. Under time-randomization (Assumption 2), an unbiased estimator for this objective function is

$$\widehat{\mathbb{E}\mathbb{V}(\hat{\tau}; \mathcal{M})} = \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}(1 - V_t)\left(\boldsymbol{M}_{i0t} + \sum_{j=1}^{N}\boldsymbol{M}_{ijt}Y_{jt}\right)^2.$$

This is in fact the objective function in (2.9). Let $\widehat{\boldsymbol{M}}$ be the value that minimizes this empirical objective, $\widehat{\boldsymbol{M}} = \arg\min_{\mathcal{M}} \widehat{\mathbb{E}\mathbb{V}(\hat{\tau}; \mathcal{M})}$. $\widehat{\boldsymbol{M}}$ is not unbiased for $\boldsymbol{M}^\star$, but it is a natural approximation.

2.3.3 Feasible Weights

Next, we consider the sets of possible weights $\mathcal{M}$. For a given set $\mathcal{M}$ we can characterize the estimators as

$$\hat{\tau}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}(\boldsymbol{Y}, \mathcal{M})) = \sum_{i=1}^{N}\sum_{t=1}^{T}U_iV_t\left\{\boldsymbol{M}_{i0t}(\boldsymbol{Y}, \mathcal{M}) + \sum_{j=1}^{N}\boldsymbol{M}_{ijt}(\boldsymbol{Y}, \mathcal{M})Y_{jt}\right\}. \tag{2.10}$$

For this class of estimators we can view the weights as non-stochastic:

Lemma 1. *For all* $\boldsymbol{Y}(0)$, $\boldsymbol{Y}(1)$, $\boldsymbol{U}$, $\boldsymbol{V}$, *and* $\mathcal{M}$,

$$\hat{\tau}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}(\boldsymbol{Y}, \mathcal{M})) = \hat{\tau}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}(\boldsymbol{Y}(0), \mathcal{M})) \tag{2.11}$$

This representation is useful because the properties of $\hat{\tau}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}(\boldsymbol{Y}(0), \mathcal{M}))$ are easier to establish under assumptions on $\boldsymbol{V}$ and $\boldsymbol{U}$ than those of $\hat{\tau}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}(\boldsymbol{Y}, \mathcal{M}))$ for the general case.

The original SC estimator (Abadie and Gardeazabal, 2003; Abadie et al., 2010) corresponds to the estimator in (2.10) with the set $\mathcal{M}$ defined as

$$\mathcal{M}^{\text{SC}} = \left\{ \boldsymbol{M} \in \mathcal{M}^0 \middle| \boldsymbol{M}_{i0t} = 0 \, \forall i, t \right\}.$$

The modification introduced in Doudchenko and Imbens (2016) and Ferman and Pinto (2019), allows for an intercept by dropping the restriction $\boldsymbol{M}_{i0t} = 0$, leading to the Modified Synthetic Control (MSC) estimator:

$$\mathcal{M}^{\text{MSC}} = \mathcal{M}^0.$$

Allowing for this intercept in the weights has been proposed to make the SC estimator more robust. Arkhangelsky et al. (2019) show that the inclusion of the intercept can be interpreted as including a unit fixed effect in the regression function. In Section 3 we discuss how the inclusion of the intercept ties in with the time randomization assumption.

We also consider a second modification of the basic SC estimator, where we place an additional set of restrictions on the weights:

$$\sum_{i=1}^{N} \boldsymbol{M}_{ijt} = 0, \;\; \forall j = 1, \ldots, N, t = 1, \ldots, T. \tag{2.12}$$

We will later show that this restriction ensures unbiasedness of the corresponding estimator given randomization of the treated units. Formally the set of possible weights for the Unbiased

Synthetic Control (USC) estimator is

$$\mathcal{M}^{\mathrm{USC}} = \left\{ \boldsymbol{M} \in \mathcal{M}^0 \middle| \boldsymbol{M}_{i0t} = 0,\ \forall i, t, \sum_{i=1}^{N} \boldsymbol{M}_{ijt} = 0\, \forall j \geq 1, t \right\}.$$

Finally, we combine the two modifications of the SC estimator, the inclusion of the intercept and the additional restriction, leading to the Modified Unbiased Synthetic Control (MUSC) estimator:

$$\mathcal{M}^{\mathrm{MUSC}} = \left\{ \boldsymbol{M} \in \mathcal{M}^0 \middle| \sum_{i=1}^{N} \boldsymbol{M}_{ijt} = 0\, \forall j \geq 1, t \right\}. \tag{2.13}$$

Here, we drop the restriction that there is no intercept, $\boldsymbol{M}_{i0t} = 0 \forall i, t$. The optimization problem in (2.9) together with the restrictions in $\mathcal{M}^{\mathrm{MUSC}}$ then imply that intercepts do not contribute bias, $\sum_{i=1}^{N} \boldsymbol{M}_{i0t} = 0\, \forall t$.

These four sets of restrictions define four estimators: $\hat{\tau}^{\mathrm{SC}}$, $\hat{\tau}^{\mathrm{MSC}}$, $\hat{\tau}^{\mathrm{USC}}$, and $\hat{\tau}^{\mathrm{MUSC}}$, based on the sets $\mathcal{M}^{\mathrm{SC}}, \mathcal{M}^{\mathrm{MSC}}, \mathcal{M}^{\mathrm{USC}}$ and $\mathcal{M}^{\mathrm{MUSC}}$, respectively. Our focus is mainly on $\hat{\tau}^{\mathrm{SC}}$ and $\hat{\tau}^{\mathrm{MUSC}}$, with the comparison with $\mathcal{M}^{\mathrm{MSC}}$ and $\mathcal{M}^{\mathrm{USC}}$ serving to aid the interpretation of the restrictions that make up the difference between $\hat{\tau}^{\mathrm{SC}}$ and $\hat{\tau}^{\mathrm{MUSC}}$. Specifically, we show that relaxing the no-intercept restriction leads to unbiasedness under time randomization for large T, and imposing the second restriction (2.12) leads to unbiasedness under unit-randomization.

3 Properties

In this section, we investigate the properties of the various estimators given the unit and/or time randomization assumptions (Assumption 1 and Assumption 2). In some cases there are exact properties while in other cases these may depend on N and/or T being large.

3.1 The Bias of the SC Estimator

The first question we study is the bias of the four SC estimators, as estimators of the treatment effect for the treated unit, τ. We summarize the results in Table 2.

The basic SC estimator is not unbiased even if both the unit treated and the time period

Table 2: Bias Properties of SC Estimators for τ

Estimator ↓	Maintained Assumptions				
	Unit Rand	Time Rand	Time Rand Large T	Unit & Time Rand	Unit & Time Rand Large T
$\hat{\tau}^{\text{SC}}$	B	B	B	B	B
$\hat{\tau}^{\text{MSC}}$	B	B	U	B	U
$\hat{\tau}^{\text{USC}}$	U	B	B	U	U
$\hat{\tau}^{\text{MUSC}}$	U	B	U	U	U

treated were both randomly selected. Because this is perhaps surprising, it is useful to study the bias of the SC estimator in more detail. In order to guarantee unbiasedness given random selection of the treated unit, we require that $\sum_{i=1}^{N} \boldsymbol{M}_{ijt} = 0$ for all $j = 1, \ldots, N$, which holds for the USC and MUSC estimator. In order to achieve unbiasedness given random selection of the time period, we need an intercept in the weights estimated through minimization of (2.9), as in the MSC and MUSC estimators, as well as a large number of time periods.

Recall the general definition of the SC type estimators in (2.4). We can rewrite this as

$$\hat{\tau}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}) = \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t \left\{ \boldsymbol{M}_{i0t} + \sum_{j=1}^{N} \boldsymbol{M}_{ijt} Y_{jt} \right\}$$

$$= \sum_{t=1}^{T} \sum_{i=1}^{N} V_t U_i \left\{ Y_{it}(1) + \boldsymbol{M}_{i0t} + \sum_{j=1}^{N} (1 - U_j) \boldsymbol{M}_{ijt} Y_{jt}(0) \right\}.$$

We focus on the properties relative to our primary estimand, the treatment effect for the treated, $\tau(\boldsymbol{U}, \boldsymbol{V}) = \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t (Y_{it}(1) - Y_{it}(0))$. The estimation error is equal to

$$\hat{\tau}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}) - \tau(\boldsymbol{U}, \boldsymbol{V}) = \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t \left\{ Y_{it}(0) + \boldsymbol{M}_{i0t} + \sum_{j=1}^{N} (1 - U_j) \boldsymbol{M}_{ijt} Y_{jt}(0) \right\}.$$

We consider the bias of the SC estimators separately for the estimators without an intercept (the SC and USC estimators), and for the estimators with the intercept estimated through minimization of the objective function (2.9) (the MSC and MUSC estimators).

Lemma 2. *Suppose Assumption 1 (random assignment of units to treatment) holds. Then if*

one of the following two conditions holds:
(i) the intercept is zero, $\boldsymbol{M}_{i0t} = 0$ for all i, t,
or (ii) if the intercept is estimated through (2.9),
then the conditional (on $\boldsymbol{V}$) bias vanishes if $\mathcal{M}$ guarantees that

$$\sum_{i=1}^{N} \boldsymbol{M}_{ijt} = 0 \;\; \forall j = 1, \ldots, N, t = 1, \ldots, T.$$

This lemma shows that τ^{USC} and τ^{MUSC} are unbiased, because both estimators only search over weight sets that satisfy the adding-up condition (2.12). Intuitively, this weight restriction ensures that each unit on average is used the same amount in treatment and in control when forming the synthetic control estimator (and that the intercepts average out to zero).

Lemma 3. *For the SC estimator the conditional bias under Assumption 1 is*

$$\mathrm{Bias}^{\mathrm{SC}} = \mathbb{E}\left[\hat{\tau} - \tau \middle| \boldsymbol{V}\right] = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} V_t Y_{it}(0) \sum_{j=1}^{N} \boldsymbol{M}_{jit}^{\mathrm{SC}}.$$

The intuition for the bias of the SC estimator also holds for the simple matching estimator (*e.g.*, Abadie and Imbens, 2006), which is generally biased under randomization in finite samples. It also extends to other weighting estimators with the addition of the conditional bias term $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} V_t \boldsymbol{M}_{i0t}$ coming from the intercepts.

Hence, there can be bias whenever $Y_{it}(0)\sum_{j=1}^{N} \boldsymbol{M}_{jit}^{\mathrm{SC}} \neq 0$ for some i and $V_t = 1$ While these weights could *happen* to sum to zero for a given SC estimator, this would require a high, non-generic level of symmetry. In general, the SC estimator will thus have bias under the randomization discussed in this article.

We can estimate the bias for the SC estimator as

$$\widehat{\mathrm{Bias}}^{\mathrm{SC}} = \frac{1}{N-1} \sum_{i=1}^{N} \sum_{t=1}^{T} V_t (1 - U_i) Y_{it} \sum_{j=1}^{N} \boldsymbol{M}_{jit}^{\mathrm{SC}}. \tag{3.1}$$

Because this estimator for the bias is unbiased, $\mathbb{E}[\widehat{\mathrm{Bias}}^{\mathrm{SC}} - \mathrm{Bias}^{\mathrm{SC}} | \boldsymbol{V}] = 0$, in principle an unbiased estimator can also be generated by subtracting this estimated bias from the standard SC estimator. However, the properties of this de-biased estimator are not very attractive in terms of RMSE.

To see the role the time randomization plays in the bias, consider the MSC estimator in a setting with large T, and random selection of the treated period. For ease of exposition, suppose unit N is the treated unit. We can view the MSC estimator as a regression estimator where we regress the outcomes $Y_{N1}, \ldots, Y_{NT}$ on the treatment indicator and the predictors $Y_{1t}, \ldots, Y_{N-1,t}$ and an intercept. It is well known that this leads to an estimator that is asymptotically unbiased in large samples (in this case meaning large T; Freedman, 2008; Lin, 2013; Imbens and Rubin, 2015).

3.2 The Exact Variance for GSC Estimators and an Unbiased Estimator for the Variance

Consider the GSC class estimator $\hat{\tau} = \hat{\tau}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M})$ as an estimator of τ.

Lemma 4. *Suppose Assumption 1 holds. Then*

$$\mathbb{V}(\boldsymbol{V}, \boldsymbol{M}) = \mathbb{E}\left[(\hat{\tau}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}) - \tau)^2 \middle| \boldsymbol{V} \right] = \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} V_t \left(\boldsymbol{M}_{i0t} + \sum_{j=1}^{N} \boldsymbol{M}_{ijt} Y_{jt}(0) \right)^2 .$$

For the unbiased estimators $\hat{\tau}^{\text{USC}}$ and $\hat{\tau}^{\text{MUSC}}$ this is also the variance. For the other estimators there is a bias, and so this is the expected squared error. Note that the synthetic-control objective in (2.9) that is used to choose the weights $\boldsymbol{M}$ is closely related to this expression for the variance.

The challenge is that the variance depends on components we do not observe. Suppose time period T is the treated period, and unit i is the treated unit. We do not observe the control outcome for unit i in period t, $Y_{iT}(0)$. Moreover, we do not know what the weights $M_{i'jT}$ are for the case that unit $i' \neq i$ is the treated unit. However, we can estimate this variance without bias under unit-randomization.

Proposition 1. *Suppose Assumption 1 holds. Then the estimator*

$$\hat{\mathbb{V}} = \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t \left\{ \frac{1}{N-3} \sum_{\substack{k=1 \\ k \neq i}}^{N} \left(\sum_{\substack{j=1 \\ j \neq i}}^{N} \boldsymbol{M}_{kjt}(Y_{kt} - Y_{jt}) \right)^2 - \frac{1}{(N-2)(N-3)} \sum_{\substack{k=1 \\ k \neq i}}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \boldsymbol{M}_{kjt}^2 (Y_{kt} - Y_{jt})^2 \right. \tag{3.2}$$

$$+\frac{2}{N-2}\sum_{\substack{k=1\\k\neq i}}^{N}\boldsymbol{M}_{k0t}\left(\sum_{\substack{j=1\\j\neq i}}^{N}\boldsymbol{M}_{kjt}(Y_{jt}-Y_{kt})\right)+\frac{1}{N}\sum_{k=1}^{N}\boldsymbol{M}_{k0t}^{2}\Bigg\},$$

is unbiased for $\mathbb{V}(\boldsymbol{V},\boldsymbol{M})$.

This result may be somewhat surprising. Note that in completely randomized experiments there is no unbiased estimator of the variance of the simple difference-in-means estimator for the average treatment effect (Imbens and Rubin, 2015). The intuition for why the current result is different is that here we focus on the average effect for the treated only. For that case there is an unbiased estimator for the variance of the difference-in-means estimator in the case of randomized experiments (*e.g.,* Sekhon and Shem-Tov (2020)).

The variance estimator in this proposition has three terms. The first takes the form of a leave-one-out estimator based on the control units excluding the treated unit. The remaining two terms correct for over-counting the diagonal elements in the inner square of the first term and additional terms for the intercept.

3.3 The Placebo Variance Estimator

To put the proposed variance estimator in (3.2) in perspective, we consider here an alternative approach for estimating the variance of SC estimator. Versions of this placebo variance estimator have been proposed previously both for testing zero effects (*e.g.* Abadie et al., 2010) and for constructing confidence intervals (*e.g.* Doudchenko and Imbens, 2016). Suppose unit i is the treated unit. We put this unit aside, and focus on the $N-1$ control units. For each of these $N-1$ control units (indexed by $j=1,\dots,N, j\neq i$) we recalculate the weights, leaving out the treated unit, and then estimate the treatment effect. For ease of exposition we focus on the case where the last period is the treated period, $V_T=1$.

We now define $(N+1)\times N$ weight matrices and sets of weight matrices $\boldsymbol{M}^{(i)}$ and $\mathcal{M}^{(i)}$. For $\boldsymbol{M}^{(i)}$ we have $\boldsymbol{M}_{ij}^{(i)}=\boldsymbol{M}_{ji}^{(i)}=0$ for all j. The weights are defined as

$$\boldsymbol{M}^{(i)}(\boldsymbol{Y},\mathcal{M}^{(i)})=\underset{\boldsymbol{M}^{(i)}\in\mathcal{M}^{(i)}}{\arg\min}\sum_{s=1}^{T-1}\sum_{\substack{j=1\\j\neq i}}^{N}\left(\boldsymbol{M}_{j0}^{(i)}+\sum_{\substack{k=1\\k\neq i}}^{N}\boldsymbol{M}_{jk}^{(i)}Y_{ks}\right)^{2}. \tag{3.3}$$

Given these weights the placebo estimator is

$$\hat{\tau}_j^{(i)} = \boldsymbol{M}_{j0}^{(i)} + \sum_{\substack{k=1 \\ k \neq i}}^{N} \boldsymbol{M}_{jk}^{(i)} Y_{kT}, \qquad \text{for } j \neq i.$$

Because unit j is a control unit, this is an estimator of zero, and the placebo variance estimator uses this it to estimate the variance of $\hat{\tau}$ as

$$\hat{\mathbb{V}}^{PCB} = \frac{1}{N-1} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} U_i \left(\hat{\tau}_j^{(i)}\right)^2 = \frac{1}{N-1} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} U_i \left(\boldsymbol{M}_{j0}^{(i)} + \sum_{\substack{k=1 \\ k \neq i}}^{N} \boldsymbol{M}_{jk}^{(i)} Y_{kT} \right)^2 .$$

The point we wish to make here is that this variance estimator is in general not unbiased. In fact, it can be upward as well as downward biased, depending on the potential outcomes $\boldsymbol{Y}(0)$. In order to demonstrate this we give two examples, one where the placebo variance estimator is biased downward, and one where it is biased upward.

Example 1. *Suppose $N = 4$ and $T = 3$, and suppose that $\boldsymbol{Y}(0)$ is*

$$\boldsymbol{Y}(0) = \begin{pmatrix} a & b & 0 \\ a & b & 1 \\ c & d & 0 \\ c & d & 1 \end{pmatrix},$$

for some arbitrary a, b, c, d. The solution for $\boldsymbol{M}$, within the set $\mathcal{M}^{musc}$ is

$$\boldsymbol{M} = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix},$$

so that the units are matched in pairs. If unit 1 is treated, the estimator is $Y_{13} - Y_{23}$, with error $Y_{13}(0) - Y_{23}(0) = -1$. Similar calculations for the other three units show that the squared error *is always equal to 1, and hence the true variance is 1.*

Now let us calculate the placebo variance. Here we exploit the fact that with three units the weights for all units are equal. This leads to

$$\boldsymbol{M}^{(1)} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ (a+b)-(c+d) & 0 & 1 & -1/2 & -1/2 \\ ((c+d)-(a+b))/2 & 0 & -1/2 & 1 & -1/2 \\ ((c+d)-(a+b))/2 & 0 & -1/2 & -1/2 & 1 \end{pmatrix},$$

Then the placebo variance is smaller in expectation than the true variance.

Example 2. *In the same setting of Assumption 3, the placebo variance can be biased upward, conditional on $\boldsymbol{V}$. Suppose $N = 4$,*

$$\boldsymbol{M} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}, \qquad \boldsymbol{Y}_{\cdot T} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix},$$

so the units are matched in pairs, and the matching is of perfect quality. Then the placebo variance is higher in expectation than the true variance.

4 An Illustration and Some Simulations

In this section we illustrate some of the methods proposed in this article.

4.1 The California Smoking Study

To illustrate some of the concepts developed in this article, we first turn to the data from the California smoking study (Abadie et al., 2010). In Figure 1 we compare the SC and MUSC estimates. We find that the pre-treatment fit as well as the point estimates are similar for the SC and MUSC estimators.

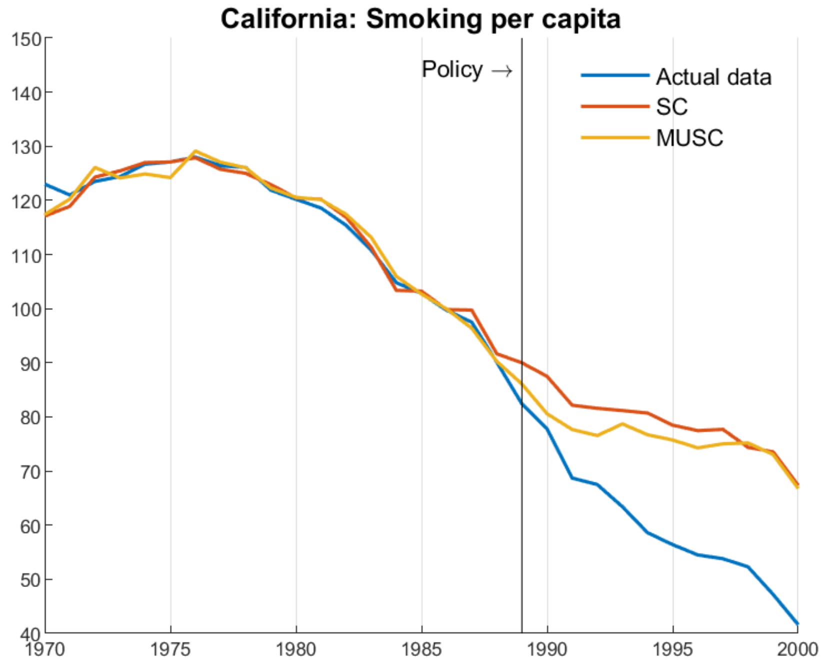


Figure 1: Pre- and Post-Treatment fit of SC and MUSC

4.2 A Simulation Study

We also perform a small simulation study to assess the properties of the MUSC estimator. Following Bertrand et al. (2004); Arkhangelsky et al. (2019), we use data from the Current Population Survey for 50 states and 40 years. The variables we analyze include state/year average log wages, hours, and the state/year unemployment rate. Note that there is no treatment, so the true treatment effects are all zero. For each of the variables, we estimate the treatment effects using the Difference-in-Means (DiM) estimator, the standard Synthetic Control (SC) estimator and the Modified Unbiased Synthetic Control (MUSC) estimator.

The first simulation study we conduct compares the performance of the three estimators based on the RMSE. The study is designed as follows. For each treated period $T \geq 21$, we use the $T-1$ pre-treatment periods to estimate the weights for the different estimators. Then, we pretend that each state has been selected for treatment once, calculate the corresponding estimated treatment effects and compare those to the true ones. Lastly, we average over all states to summarize the performance in terms of RMSE for a single treated period. In Table 3, we report the results. We find that the RMSE is substantially lower for the SC and MUSC estimator compared to the DiM estimator for all variables. Moreover, this holds for all years. Note that the SC and MUSC estimator have similar RMSE.

Table 3: Simulation Experiment Based on CPS Data by State and Year – RMSE

Treated Period	**Log Wages**			**Hours**			**Unemployment Rate**		
	DiM	SC	MUSC	DiM	SC	MUSC	DiM	SC	MUSC
$T = 21$	0.1157	0.0543	0.0550	1.4838	0.9797	0.9340	0.0123	0.0111	0.0102
$T = 22$	0.1112	0.0496	0.0502	1.3764	0.8204	0.9170	0.0107	0.0086	0.0084
$T = 23$	0.1089	0.0435	0.0484	1.2821	0.9091	0.9009	0.0114	0.0111	0.0116
$T = 24$	0.1143	0.0421	0.0450	1.2400	0.9569	0.8829	0.0145	0.0146	0.0142
$T = 25$	0.1136	0.0465	0.0500	1.3125	0.9074	0.8896	0.0124	0.0110	0.0115
$T = 26$	0.1111	0.0487	0.0471	1.1424	0.8509	0.8347	0.0132	0.0136	0.0146
$T = 27$	0.1105	0.0481	0.0493	1.1926	0.8612	0.8141	0.0155	0.0145	0.0149
$T = 28$	0.1014	0.0551	0.0674	1.0627	0.7611	0.7660	0.0136	0.0105	0.0106
$T = 29$	0.0964	0.0473	0.0594	1.1342	0.8552	0.8686	0.0137	0.0124	0.0118
$T = 30$	0.0980	0.0505	0.0516	1.0741	0.7765	0.7755	0.0137	0.0141	0.0147
$T = 31$	0.1084	0.0410	0.0393	1.2340	0.9700	0.9537	0.0209	0.0180	0.0175
$T = 32$	0.1049	0.0460	0.0468	1.0571	0.8496	0.8209	0.0214	0.0182	0.0167
$T = 33$	0.1046	0.0530	0.0536	1.2947	1.0018	0.9696	0.0215	0.0161	0.0151
$T = 34$	0.1060	0.0625	0.0662	1.1640	0.9174	0.9007	0.0182	0.0131	0.0129
$T = 35$	0.1053	0.0560	0.0537	1.2233	1.0824	0.9914	0.0201	0.0137	0.0137
$T = 36$	0.0985	0.0594	0.0535	1.1267	0.8083	0.7396	0.0150	0.0124	0.0119
$T = 37$	0.1017	0.0605	0.0579	1.3066	1.2080	1.1235	0.0156	0.0128	0.0131
$T = 38$	0.0929	0.0580	0.0615	0.9917	0.7742	0.7929	0.0126	0.0116	0.0117
$T = 39$	0.0853	0.0459	0.0554	0.8437	0.7979	0.9475	0.0112	0.0120	0.0122
$T = 40$	0.1051	0.0517	0.0479	1.4048	1.2714	1.2382	0.0126	0.0112	0.0106
Average	0.1047	0.0510	0.0530	1.1404	0.8743	0.9031	0.0150	0.0130	0.0129

DiM: Difference in Means estimator, SC: Synthetic Control estimator of Abadie et al. (2010), MUSC: Modified Unibased Synthetic Control Estimator.

The second simulation study demonstrates the properties of our proposed unbiased variance estimator and the placebo variance estimator. Here we focus on the average log wages and fix $T = 40$ as the treated period. Moreover, we decrease the sample to $N = 20$ units in total. For ease of readability, we choose to report the standard error based on our variance estimator instead of the variances themselves. Table 4 reports the true variance along with the estimates based on our variance estimator and using the placebo approach. We find that our estimator is indeed unbiased. The placebo approach is clearly biased - the direction of the bias depends on which estimator is used. For the DiM and the SC estimator, the placebo estimator is upward biased; for the MUSC estimator it is downward biased.

4.3 Inference

Under the null hypothesis that the treatment year is as good as random, we can calculate p-values for the test $H_0 : \tau = 0$ against $H_a : \tau \neq 0$. These p-values are calculated in a similar spirit to placebo variance estimators. However, instead of using only random treatment years to compare the same unit across time, we leverage randomization across units as well and compare treatment and control estimates during pre-treatment periods.

For $s \geq 1$, calculate $\tau^{\text{MUSC}}(N)$ for the treated unit in period $T - s$. Then, calculate τ^{MUSC} for each control unit in the same period, and obtain $\tau^{\text{MUSC}}(i)$ for $i = 1, \ldots, N - 1$. Then under the null hypotheses of time and unit randomization, $Pr(|\tau^{\text{MUSC}}(N)| > |\tau^{\text{MUSC}}(i)|)$ gives a proper p-value. Note that this algorithm does not take any computation other than computing the MUSC estimator since the objective function necessarily finds weights simultaneously for all units, both treated and and control.

In Table 5, we show p-values calculated in these way for $s = 1, \ldots, 5$. All but the latest calculated p-value are not significant at the 10% level. However, the significance of the latest time period, $T - 1$, may give some indication of treatment anticipation.

If p-values are calculated this way for multiple values of s, meaning that p-values are calculated for multiple pre-treatment periods, the p-values are not independent, so they cannot be viewed as separate tests. However, calculating all the p-values is still useful and gives some idea of whether the treatment effect is nonzero.

5 Generalizations and Extensions

In this section we look at generalizations of the set up considered so far with a single treated unit and single treated period where the estimand was the average effect for the treated. First, we consider the case with multiple treated units. Second, we consider the case where the estimand is the average effect for all units in the treated period. Both of these generalizations create conceptual complications. Third, we consider the case of non-constant propensity scores.

Table 4: Simulation Experiment Based on CPS Data (Log Wages) by State and Year for $N = 20$ units – Standard Error; Treated Period $T = 40$

Unit	$\sqrt{\hat{\mathrm{V}}}_{\mathrm{DiM},true}$	$\sqrt{\hat{\mathrm{V}}}_{\mathrm{DiM}}$	$\sqrt{\hat{\mathrm{V}}}_{\mathrm{DiM,PCB}}$	$\sqrt{\hat{\mathrm{V}}}_{\mathrm{SC},true}$	$\sqrt{\hat{\mathrm{V}}}_{\mathrm{SC}}$	$\sqrt{\hat{\mathrm{V}}}_{\mathrm{SC,PCB}}$	$\sqrt{\hat{\mathrm{V}}}_{\mathrm{MUSC},true}$	$\sqrt{\hat{\mathrm{V}}}_{\mathrm{MUSC}}$	$\sqrt{\hat{\mathrm{V}}}_{\mathrm{MUSC,PCB}}$
$i = 1$	0.1157	0.1163	0.1165	0.0518	0.0521	0.0533	0.0490	0.0484	0.0459
$i = 2$	0.1157	0.1176	0.1178	0.0518	0.0497	0.0529	0.0490	0.0476	0.0498
$i = 3$	0.1157	0.1129	0.1131	0.0518	0.0515	0.0520	0.0490	0.0499	0.0486
$i = 4$	0.1157	0.1184	0.1186	0.0518	0.0526	0.0542	0.0490	0.0481	0.0515
$i = 5$	0.1157	0.1173	0.1174	0.0518	0.0530	0.0558	0.0490	0.0524	0.0503
$i = 6$	0.1157	0.1147	0.1149	0.0518	0.0526	0.0514	0.0490	0.0510	0.0487
$i = 7$	0.1157	0.1158	0.1160	0.0518	0.0412	0.0440	0.0490	0.0466	0.0479
$i = 8$	0.1157	0.1166	0.1167	0.0518	0.0517	0.0523	0.0490	0.0458	0.0504
$i = 9$	0.1157	0.1182	0.1183	0.0518	0.0540	0.0531	0.0490	0.0530	0.0514
$i = 10$	0.1157	0.1176	0.1178	0.0518	0.0523	0.0519	0.0490	0.0468	0.0500
$i = 11$	0.1157	0.1188	0.1190	0.0518	0.0528	0.0510	0.0490	0.0436	0.0466
$i = 12$	0.1157	0.1186	0.1188	0.0518	0.0527	0.0525	0.0490	0.0449	0.0491
$i = 13$	0.1157	0.1118	0.1119	0.0518	0.0538	0.0592	0.0490	0.0553	0.0493
$i = 14$	0.1157	0.1186	0.1188	0.0518	0.0513	0.0537	0.0490	0.0491	0.0507
$i = 15$	0.1157	0.1157	0.1158	0.0518	0.0523	0.0512	0.0490	0.0449	0.0474
$i = 16$	0.1157	0.1188	0.1190	0.0518	0.0509	0.0492	0.0490	0.0442	0.0426
$i = 17$	0.1157	0.1171	0.1172	0.0518	0.0543	0.0531	0.0490	0.0492	0.0504
$i = 18$	0.1157	0.1115	0.1117	0.0518	0.0538	0.0519	0.0490	0.0488	0.0484
$i = 19$	0.1157	0.1109	0.1110	0.0518	0.0504	0.0501	0.0490	0.0456	0.0474
$i = 20$	0.1157	0.1054	0.1055	0.0518	0.0510	0.0475	0.0490	0.0614	0.0510
Average[1]	0.1157	0.1157	0.1158	0.0518	0.0518	0.0521	0.0490	0.0490	0.0489

[1] To illustrate unbiasedness, we calculate this average by taking the square root of the mean over all unit variances.

Table 5: p-values for California Smoking Data

	$pr((\lvert\hat{\tau}(N)\rvert > \hat{\tau}(i))$
T-1	0.0789
T-2	0.3947
T-3	0.1579
T-4	0.1053
T-5	0.4737

5.1 Multiple Treated Units

In this section we look at the case with multiple treated units. We fix the number of treated units at N_{T}. The estimand is, as before, the average effect for the N_{T} treated units:

$$\tau = \tau(\boldsymbol{U}, \boldsymbol{V}) \equiv \frac{1}{N_{\mathrm{T}}} \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t \Big(Y_{it}(1) - Y_{it}(0) \Big),$$

We modify Assumption 1 to

Assumption 5. (Random Assignment of Units)

$$\mathrm{pr}(\boldsymbol{U} = \boldsymbol{u}) = \begin{cases} \left(\frac{N!}{N_{\mathrm{T}}! N_{\mathrm{C}}!} \right)^{-1} & \text{if} \quad u_i \in \{0,1\}\ \forall i, \quad \sum_{i=1}^{N} u_i = N_{\mathrm{T}}, \\ 0 & \text{otherwise.} \end{cases}$$

In this case it is convenient to work with the sets of units assigned to the treatment, rather than the individual units. There are $K = N!/(N_{\mathrm{T}}! N_{\mathrm{C}}!) = \binom{N}{N_T}$ such sets. Of these, $(N-1)!/((N_{\mathrm{T}}-1)! N_{\mathrm{C}}!) = \binom{N-1}{N_T-1}$ include a given unit, such as unit 1, since there are $\binom{N-1}{N_T-1}$ combinations of the remaining units if that unit is treated. This represents a fraction N_{T}/N of the total number of sets of N_{T} treated units. Hence the fraction of sets that does not include unit 1 is N_{C}/N.

Let $\tilde{\boldsymbol{U}}$ be the vector of length K of indicators that denotes which set of N_{T} units is treated. Let e_k denote the K-component vector with all elements equal to zero, other than the k-th component which is equal to one. By construction, $\sum_{k=1}^{K} \tilde{\boldsymbol{U}}_k = 1$, and $\tilde{\boldsymbol{U}}_k \in \{0,1\}$. Assumption 5 implies that the probability that $\tilde{\boldsymbol{U}}_k = 1$ is equal to $N_{\mathrm{T}}! N_{\mathrm{C}}!/N! = 1/K$. Let $u_i(\tilde{\boldsymbol{U}}) \in \{0,1\}$ be an indicator for unit i being treated given the assignment vector $\tilde{\boldsymbol{U}}$. In this notation we can rewrite τ as

$$\tau = \tau(\tilde{\boldsymbol{U}}, \boldsymbol{V}) = \frac{1}{N_{\mathrm{T}}} \sum_{k=1}^{K} \sum_{t=1}^{T} \tilde{\boldsymbol{U}}_k V_t \sum_{i=1}^{N} u_i(\tilde{\boldsymbol{U}}_k) \Big(Y_{it}(1) - Y_{it}(0) \Big).$$

Instead of the tensors $\boldsymbol{M}$ with dimension $N \times (N+1) \times T$, we now have tensors with dimension $K \times (N+1) \times T$, with one row for each of the $K = N!/(N_{\mathrm{T}}!N_{\mathrm{C}}!)$ possible sets of treated units. The estimators we consider are of the form

$$\hat{\tau}(\tilde{\boldsymbol{U}}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}) \equiv \sum_{k=1}^{K} \sum_{t=1}^{T} \tilde{\boldsymbol{U}}_k V_t \left\{ \boldsymbol{M}_{k0t} + \sum_{j=1}^{N} \boldsymbol{M}_{kjt} Y_{jt} \right\}. \tag{5.1}$$

This formulation suggest the restriction that $\boldsymbol{M}_{kjt} = 1/N_T$, for all j such that $u_j(e_k) = 1$ and $\boldsymbol{M}_{kjt} \leq 0$, whenever $u_j(e_k) = 0$. The set of such $\boldsymbol{M}$ we consider is for the generalized modified unbiased synthetic control (MUSC) estimator is

$$\mathcal{M}^{\mathrm{MUSC}} = \left\{ \boldsymbol{M} \middle| \sum_{j=1}^{N} \boldsymbol{M}_{kjt} = 0 \, \forall k, t, \sum_{k=1}^{K} \boldsymbol{M}_{kjt} = 0 \, \forall j \geq 1, t \right\}.$$

The objective function for choosing $\boldsymbol{M}$ is now

$$\boldsymbol{M}(\boldsymbol{Y}, \mathcal{M}^{\mathrm{MUSC}}) = \underset{\boldsymbol{M} \in \mathcal{M}^{\mathrm{MUSC}}}{\arg\min} \sum_{k=1}^{K} \sum_{t=1}^{T} \left\{ \sum_{s \neq t} \left(\boldsymbol{M}_{k0t} + \sum_{j=1}^{N} \boldsymbol{M}_{kjt} Y_{js} \right)^2 \right\}.$$

Lemma 5. *Suppose that Assumption 5 holds. Then*
(i) the estimator $\hat{\tau}(\tilde{\boldsymbol{U}}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}(\boldsymbol{Y}, \mathcal{M}^{\mathrm{MUSC}}))$ is unbiased conditional on $\boldsymbol{V}$:

$$\mathbb{E}\left[\hat{\tau}(\tilde{\boldsymbol{U}}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}(\boldsymbol{Y}, \mathcal{M}^{\mathrm{MUSC}})) - \tau(\boldsymbol{U}, \boldsymbol{V}) \middle| \boldsymbol{V} \right] = 0,$$

(ii) the variance of $\hat{\tau}(\tilde{\boldsymbol{U}}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}(\boldsymbol{Y}, \mathcal{M}^{\mathrm{MUSC}})$ is

$$\mathbb{V}\left(\hat{\tau}(\tilde{\boldsymbol{U}}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}(\boldsymbol{Y}, \mathcal{M}^{\mathrm{MUSC}})) \middle| \boldsymbol{V} \right) = \frac{1}{K} \sum_{t=1}^{T} V_t \sum_{k=1}^{K} \left(\boldsymbol{M}_{k0t} + \sum_{j=1}^{N} \boldsymbol{M}_{kjt} Y_{ts}(0) \right)^2,$$

and (iii), the variance can be estimated without bias (conditional on $\boldsymbol{V}$) by a generalization of the variance estimator in Proposition 1,

$$\hat{\mathbb{V}} = \sum_{k=1}^{K}\sum_{t=1}^{T} \tilde{\boldsymbol{U}}_k V_t \Bigg(\sum_{\substack{k'=1; \\ u_i(\tilde{\boldsymbol{U}}_k)+u_i(\tilde{\boldsymbol{U}}_{k'})\leq 1 \forall i}}^{K} \Bigg\{ \frac{1}{\binom{N_C-2}{N_T}} \left(\sum_{j=1}^{N} (1-u_j(\tilde{\boldsymbol{U}}_k))\boldsymbol{M}_{k'jt}(Y_{jt}-\overline{Y}_{k't}) \right)^2$$

$$- \frac{N_T}{(N_C-1)\binom{N_C-2}{N_T}} \sum_{j=1}^{N} (1-u_j(\tilde{\boldsymbol{U}}_k))\boldsymbol{M}^2_{k'jt}(Y_{jt}-\overline{Y}_{k't})^2$$

$$+ \frac{2}{\binom{N_C-1}{N_T}} \boldsymbol{M}_{k'0t} \sum_{j=1}^{N} (1-u_j(\tilde{\boldsymbol{U}}_k))\boldsymbol{M}_{k'jt}(Y_j-\overline{Y}_{k'}) \Bigg\} + \frac{1}{K}\sum_{k'=1}^{K} \boldsymbol{M}^2_{k'0t} \Bigg)$$

for $\overline{Y}_{k't} = \frac{1}{N_T}\sum_{j=1}^{N} u_j(\tilde{\boldsymbol{U}}_{k'})Y_{jt}$.

So far the estimators we consider have imposed the restriction that all the treated units receive equal weight,

$$\boldsymbol{M}_{kjt} = \frac{1}{N_T}$$

In the case with a single treated unit that restriction was natural, but here we could relax this to requiring only that the sum of the weights for the treated units is restricted to unit:

$$\sum_{j=1}^{N} u_j(e_k)\boldsymbol{M}_{kjt} = 1.$$

This allows us to choose the weights for the treated units to reduce the variance. Yet changing the weights on treated units also affects the expectation under unit randomization. Specifically, we can understand the resulting estimator as estimating the weighted estimand

$$\sum_{k=1}^{K}\sum_{t=1}^{T} \tilde{\boldsymbol{U}}_k V_t \sum_{i=1}^{N} u_i(\tilde{\boldsymbol{U}}_k)\boldsymbol{M}_{kit}\Big(Y_{it}(1)-Y_{it}(0)\Big).$$

While the estimator is unbiased relative to this estimand, and we can estimate its variance in the same way as in Lemma 5, we cannot generally estimate its error relative to the equally-weighted average treatment effect on the treated τ.

To see the issue of unequally weighted treatment units, let us focus on the simplest case with two treated units and a single control unit. In that case there are $K = 3$ possible sets of two treated units. If we impose the restrictions that the weights for the treated units sum to one and the weights for the control units sum to minus one, there is only one free parameter in each row of the weight matrix. Consider the first row of the weight matrix, with the third unit the control unit. The estimator in that case is

$$\hat{\tau} = \boldsymbol{M}_{11T} Y_1(1) + \boldsymbol{M}_{12T} Y_2(1) - Y_3(0).$$

The error is

$$\hat{\tau} - \frac{1}{2}(Y_1(1) + Y_2(1)) + \frac{1}{2}(Y_1(0) + Y_2(0))$$

$$= (\boldsymbol{M}_{11T} - 1/2) Y_1(1) + (\boldsymbol{M}_{12T} - 1/2) Y_2(1) + \frac{1}{2}\left(Y_1(0) + Y_2(0)\right) - Y_3(0).$$

Hence the expected squared error over the three assignments is

$$\frac{1}{3}\Bigg\{\left((\boldsymbol{M}_{11T} - 1/2) Y_1(1) + (\boldsymbol{M}_{12T} - 1/2) Y_2(1) + \frac{1}{2}\left(Y_1(0) + Y_2(0)\right) - Y_3(0)\right)^2$$

$$+ \left((\boldsymbol{M}_{21T} - 1/2) Y_1(1) + (\boldsymbol{M}_{23T} - 1/2) Y_3(1) + \frac{1}{2}\left(Y_1(0) + Y_3(0)\right) - Y_2(0)\right)^2$$

$$+ \left((\boldsymbol{M}_{32T} - 1/2) Y_2(1) + (\boldsymbol{M}_{33T} - 1/2) Y_3(1) + \frac{1}{2}\left(Y_2(0) + Y_3(0)\right) - Y_1(0)\right)^2\Bigg\}.$$

The complication is that there is no unbiased estimator for this error because it involves cross-products of $Y_i(0)$ and $Y_i(1)$ which cannot be estimated. If we impose the restriction that the weights for all the treated units are equal, the dependence of the error on the $Y_i(1)$ vanishes, and the error can in general be estimated without bias.

5.2 The Average Effect for All Units

Here we look at the case where the estimand changes from the average effect for the treated unit(s) to the average effect over all units in the treated periods. For ease of exposition we continue to focus on the case with a single treated period and a single treated unit. The extension to the case with multiple treated units is conceptually clear based on the discussion in the previous subsection. Formally, the estimand is

$$\tau^{\mathrm{V}} = \tau^{\mathrm{V}}(\boldsymbol{V}) \equiv \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} V_t \Big(Y_{it}(1) - Y_{it}(0) \Big),$$

We can separate this into two components, the effect for the treated unit,

$$\tau^{\mathrm{T}} \equiv \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t \Big(Y_{it}(1) - Y_{it}(0) \Big),$$

(which is the same as τ before), and the average effect for the control units:

$$\tau^{\mathrm{C}} \equiv \frac{1}{N-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (1 - U_i) V_t \Big(Y_{it}(1) - Y_{it}(0) \Big),$$

with

$$\tau^{\mathrm{V}} = \frac{1}{N} \tau^{\mathrm{T}} + \frac{N-1}{N} \tau^{\mathrm{C}}.$$

Consider, as before, an estimator of the form

$$\hat{\tau}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}, \boldsymbol{M}) = \sum_{i=1}^{N} \sum_{t=1}^{T} U_i V_t \left\{ \boldsymbol{M}_{i0t} + \sum_{j=1}^{N} \boldsymbol{M}_{ijt} Y_{jt} \right\}$$

$$= \sum_{t=1}^{T} \sum_{i=1}^{N} V_t U_i \left\{ \boldsymbol{M}_{i0t} + \boldsymbol{M}_{iit} Y_{it}(1) + \sum_{j=1}^{N} (1 - U_j) \boldsymbol{M}_{ijt} Y_{jt}(0) \right\}.$$

the restrictions $\boldsymbol{M}_{iit} = 1 \, \forall i, t$, $\sum_{i=1}^{N} \boldsymbol{M}_{ijt} = 0 \, \forall j, t$ (including the intercept) still imply unbiasedness conditional on $\boldsymbol{V}$, and the MUSC remains unbiased for τ^{V}. Yet the variance (and more generally the conditional expected loss of such a weighted estimator) is now

$$\begin{aligned}
&\mathbb{E}\left[\left(\hat{\tau}(\boldsymbol{U},\boldsymbol{V},\boldsymbol{Y},\boldsymbol{M})-\tau^V\right)^2\middle|\boldsymbol{V}\right]\\
&=\frac{1}{N}\sum_{i=1}^N\sum_{t=1}^T V_t\Bigg(\boldsymbol{M}_{i0t}+\left(\boldsymbol{M}_{iit}-\frac{1}{N}\right)Y_{it}(1)-\sum_{j=1}^N(1-U_j)\frac{1}{N}Y_{jt}(1)\\
&\qquad\qquad+\frac{1}{N}Y_{it}(0)+\sum_{j=1}^N(1-U_j)\left(\boldsymbol{M}_{ijt}+\frac{1}{N}\right)Y_{jt}(0)\Bigg)^2
\end{aligned} \tag{5.2}$$

which depends on treated and untreated potential outcomes. This creates two related challenges: First, since the expression depends on treated outcomes, there is no immediate sample analogue available that corresponds to minimizing expected error, even under time randomization. Second, the variance cannot generally be estimated without bias, since it depends not only on the variation of the $Y_{it}(0)$ (which can be estimated), but also on the variation of the $Y_{it}(1)$ and their covariance with the $Y_{it}(0)$ (neither of which is identified from the data).

We briefly discuss two assumptions on the (non-stochastic) correlation of treatment and control outcomes, and what they imply for estimation. First, if treatment effects are constant within time period (and treatment and control potential outcomes thus perfectly correlated, $Y_{it}(1)-Y_{it}(0)=\tau$), then (5.2) becomes

$$\frac{1}{N}\sum_{i=1}^N\sum_{t=1}^T V_t\left(\boldsymbol{M}_{i0t}+\sum_{j=1}^N\boldsymbol{M}_{ijt}Y_{jt}(0)\right)^2$$

as before, suggesting the MUSC estimator. If, on the other hand, treated outcomes are uncorrelated to control outcomes, and we focus on unbiased estimators (so in particular $\boldsymbol{M}_{iit}=1$), then (5.2) becomes

$$\frac{1}{N}\sum_{i=1}^N\sum_{t=1}^T V_t\left(\boldsymbol{M}_{i0t}+\frac{1}{N}\sum_{j=1}^N Y_{jt}(0)+\sum_{j=1}^N(1-U_j)\boldsymbol{M}_{ijt}Y_{jt}(0)\right)^2+\text{const.},$$

suggesting an alternative MUSC-type estimator that minimizes the sample analogue in non-treated time periods over weights $\mathcal{M}^{\text{MUSC}}$, which could effectively shrink the MUSC weights on control units towards the DiM weights $-\frac{1}{N-1}$. (Indeed, one feasible set of weights would yield the estimator $\hat{\tau}=\frac{1}{N}\hat{\tau}^{\text{MUSC}}+\frac{N-1}{N}\hat{\tau}^{\text{DiM}}$ of $\tau^V=\frac{1}{N}\tau^T+\frac{N-1}{N}\tau^C$ corresponding to control weights

$\frac{1}{N}\boldsymbol{M}_{ijt}^{\text{MUSC}} - \frac{1}{N}$. This solution would likely be suboptimal because it enforces $\boldsymbol{M}_{ijt} \leq -\frac{1}{N}$ for control weights.)

5.3 Non-Constant Propensity Scores

Throughout this article, we have assumed that treatment is assigned with equal probability across units, time periods, or unit–time pairs. Yet the theory we develop generalizes to non-constant propensity scores. For example, assume that treatment is assigned randomly to unit i with probability p_i (or, similarly, to a time period or a unit–time pair), where $\sum_{i=1}^N p_i = 1$. A natural analogue of the MUSC estimator is then

$$\boldsymbol{M}_p^{\text{MUSC}}(\boldsymbol{Y}, \mathcal{M}_p^{\text{MUSC}}) \equiv \underset{\boldsymbol{M} \in \mathcal{M}_p^{\text{MUSC}}}{\arg\min} \sum_{i=1}^N p_i \sum_{t=1}^T \left\{ \sum_{s \neq t} \left(\boldsymbol{M}_{i0t} + \sum_{j=1}^N \boldsymbol{M}_{ijt} Y_{js} \right)^2 \right\} \tag{5.3}$$

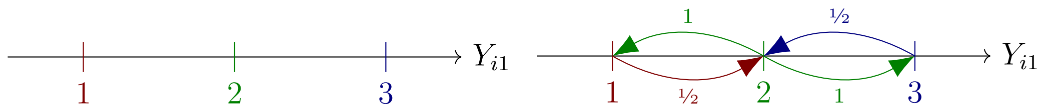
where unbiasedness is guaranteed by the constraints

$$\mathcal{M}_p^{\text{MUSC}} = \left\{ \boldsymbol{M} \middle| \sum_{j=1}^N \boldsymbol{M}_{ijt} = 0 \; \forall i, t, \sum_{i=1}^N p_i \boldsymbol{M}_{ijt} = 0 \; \forall j, t \right\}.$$

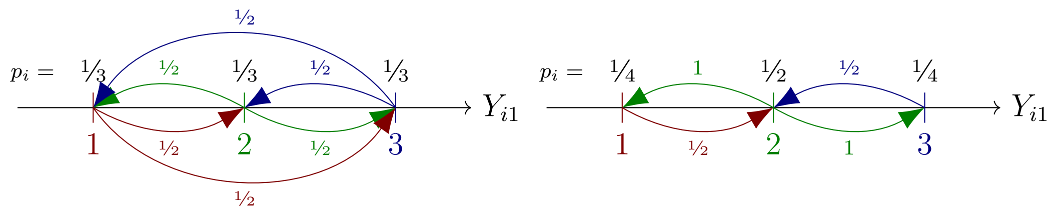
Such an estimator could be used when treatment is assigned randomly. When the analyst has a choice over the treatment assignment, and $t = T$, the optimization in (5.3) could also include the choice of propensity score.

We now illustrate how varying propensity scores can affect the synthetic control estimator in a simple three-unit, two-period example where treatment is assigned at time $t = 2$ with probabilities p_i and we have one pre-treatment outcome Y_{i1} available to construct synthetic-control weights $\boldsymbol{M}_{ij2}$. For simplicity, we do not include an intercept, and assume that $Y_{i1} < Y_{i2} < Y_{i3}$ are equally spaced (Figure 2a). In this example, we consider the standard SC estimator, the USC estimator with equal propensities, and the USC estimator with non-constant propensities.

The weights of the standard SC estimator are represented in Figure 2b. When the central unit ($i = 2$) is treated, equal weights are placed on the outcomes of units $1, 3$. However, when the peripheral units 1 or 3 are treated, full weight is placed on the central unit 2, leading to potential bias when treatment is assigned with equal probability since unit 2 is used more often as a control than units 1 and 3. The USC estimator corrects this bias by enforcing balance



(a) Pre-treatment outcomes **(b)** Weights of the SC estimator



(c) USC, constant propensity **(d)** USC, varying propensity

Figure 2: Pre-treatment outcome in three-unit, two-period example with varying treatment propensities. An outgoing arrow represents the weight assigned to that unit when the target of the arrow is treated, and arrows are colored by the unit the respective weight is put on.

between the probability of being treated and being used as a control (Figure 2c), enforcing in this simple example that all untreated units are used as control with weight ½ in all cases. However, when the central unit is treated with higher probability of ½ (Figure 2), then the weight matrix that only uses the closest units as control in each case is the optimal unbiased solution. In this specific example, this solution also coincides with the standard SC solution, but this is not generally the case.

This example suggests ways in which considering varying propensity scores can be helpful when analyzing SC-type estimators. First, when treatment is randomized according to a known probability distribution, then those probabilities affect the optimal USC and MUSC weights. Second, when we choose propensity scores in the design of an experiment and plan to use an SC-type estimator, then we can optimize the choice of propensities based on past outcomes to be better suited to their relationship, assigning more central units higher probabilities. Finally, even in the observational case, varying propensities could be used when some units can be considered to be more likely to receive treatment or to be more appropriate as controls, allowing to replace binary inclusion criteria by treatment propensities.

6 Conclusion

In this article we study Synthetic Control methods from a design perspective. We show that when a randomized experiment is conducted, the standard SC estimator is biased. However, a minor modification of the SC estimator is unbiased under randomization, and in cases with few treated units can have RMSE properties superior to those of the standard Difference-in-Means estimator. We show that the design perspective also has implications for observational studies. We propose a variance estimator that is validated by randomization.

References

Abadie, A. (2019). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*.

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica*, 88(1):265–296.

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

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

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

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1):235–267.

Abadie, A. and L'Hour, J. (2017). A penalized synthetic control estimator for disaggregated data. *Work. Pap., Mass. Inst. Technol., Cambridge, MA*.

Amjad, M., Shah, D., and Shen, D. (2018). Robust synthetic control. *The Journal of Machine Learning Research*, 19(1):802–852.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2019). Synthetic difference in differences. Technical report, National Bureau of Economic Research.

Athey, S., Bayati, M., Doudchenko, N., Imbens, G., and Khosravi, K. (2017). Matrix completion methods for causal panel data models. *arXiv preprint arXiv:1710.10251*.

Athey, S. and Imbens, G. W. (2018). Design-based analysis in difference-in-differences settings with staggered adoption. Technical report, National Bureau of Economic Research.

Ben-Michael, E., Feller, A., and Rothstein, J. (2020). The augmented synthetic control method. *arXiv preprint arXiv:1811.04170*.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275.

Chernozhukov, V., Wuthrich, K., and Zhu, Y. (2017). An exact and robust conformal inference method for counterfactual and synthetic controls. *arXiv preprint arXiv:1712.09089*.

Cunningham, S. (2018). *Causal inference: The mixtape*. Yale University Press.

Doudchenko, N. and Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. Technical report, National Bureau of Economic Research.

Ferman, B. and Pinto, C. (2017). Placebo tests for synthetic controls.

Ferman, B. and Pinto, C. (2019). Synthetic controls with imperfect pre-treatment fit. *arXiv preprint arXiv:1911.08521*.

Fisher, R. A. (1937). *The design of experiments*. Oliver And Boyd; Edinburgh; London.

Freedman, D. A. (2008). On regression adjustments in experiments with several treatments. *The annals of applied statistics*, 2(1):176–196.

Hahn, J. and Shi, R. (2016). Synthetic control and inference. *Available at UCLA*.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Lei, L. and Candès, E. J. (2020). Conformal inference of counterfactuals and individual treatment effects. *arXiv preprint arXiv:2006.06138*.

Li, K. T. (2020). Statistical inference for average treatment effects estimated by synthetic control methods. *Journal of the American Statistical Association*, 115(532):2068–2083.

Lin, W. (2013). Agnostic notes on regression adjustments for experimental data: Reexamining freedman's critique. *The Annals of Applied Statistics*, 7(1):295–318.

Neyman, J. (1923/1990). On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, 5(4):465–472.

Rambachan, A. and Roth, J. (2020). Design-Based Uncertainty for Quasi-Experiments. *arXiv preprint arXiv:2008.00602*.

Rosenbaum, P. R. (2002). *Observational Studies*. Springer.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Sekhon, J. S. and Shem-Tov, Y. (2020). Inference on a new class of sample average treatment effects. *Journal of the American Statistical Association*, pages 1–7.

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

Appendix

Proof of Proposition 1. We first consider the case without intercept. As a preliminary calculation, note that for $k, j, j' \in \{1, \ldots, n\}$

$$\sum_{i=1}^{N} \sum_{k,j,j' \neq i} \frac{1}{N - |\{k, j, j'\}|} a_{kjj'} = \sum_{k,j,j'} a_{kjj'}, \tag{A.1}$$

since every term kjj' term appears $N - |\{k, j, j'\}|$ times in the sum on the left. Let now

$$a_{kjj'} = M_{kj}(Y_j(0) - Y_k(0)) \cdot M_{kj'}(Y_{j'}(0) - Y_k(0)),$$

where for simplicity we fix the period t, drop all time indices to set $M_{ij} = \boldsymbol{M}_{ijt}$, and write $\hat{\mathbb{V}}_i$ for the variance estimator when $U_i = 1$. Then $a_{kjj'} = 0$ for $k \in \{j, j'\}$ and thus

$$\frac{1}{N} \sum_{i=1}^{N} \underbrace{\left(\frac{1}{N-3} \sum_{k \neq i} \left(\sum_{j \neq i} M_{kj}(Y_j(0) - Y_k(0)) \right)^2 - \frac{1}{(N-3)(N-2)} \sum_{k,j \neq i} M_{kj}^2 (Y_j(0) - Y_k(0))^2 \right)}_{=\hat{\mathbb{V}}_i}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \Big(\sum_{k,j,j' \neq i} \frac{1}{N-3} a_{kjj'} - \sum_{\substack{k,j,j' \neq i \\ j = j'}} \underbrace{\frac{1}{(N-3)(N-2)}}_{= \frac{1}{N-3} - \frac{1}{N-2}} a_{kjj'} \Big)$$

$$= \frac{1}{N} \Big(\sum_{i=1}^{N} \sum_{\substack{k,j,j' \neq i \\ |\{k,j,j'\}| = 3}} \frac{1}{N-3} a_{kjj'} + \sum_{\substack{k,j,j' \neq i \\ |\{k,j,j'\}| = 2}} \frac{1}{N-2} \underbrace{a_{kjj'}}_{=0 \text{ for } j \neq j'} + \sum_{\substack{k,j,j' \neq i \\ |\{k,j,j'\}| = 1}} \frac{1}{N-1} \underbrace{a_{kjj'}}_{=0} \Big)$$

$$= \frac{1}{N} \sum_{i=1}^{N} \sum_{k,j,j' \neq i} \frac{1}{N - |\{k, j, j'\}|} a_{kjj'} \overset{\text{(A.1)}}{=} \frac{1}{N} \sum_{k,j,j'} a_{kjj'}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} M_{ij}(Y_j(0) - Y_i(0)) \right)^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\sum_{j=1}^{N} M_{ij} Y_j(0) \right)^2 = \mathbb{V}.$$

Here, we have used that $\sum_{j=1}^{N} M_{ij} = 0$.

With an intercept we note that

$$\mathbb{V} = \frac{1}{N}\sum_{i=1}^N \left(M_{i0} + \sum_{j=1}^N M_{ij} Y_j(0) \right)^2 = \frac{1}{N}\sum_{i=1}^N \left(M_{i0} + \sum_{j=1}^N M_{ij} (Y_j(0) - Y_i(0)) \right)^2$$
$$= \frac{1}{N}\sum_{i=1}^N M_{i0}^2 + \frac{2}{N}\sum_{i=1}^N M_{i0} \left(\sum_{j=1}^N M_{ij} (Y_j(0) - Y_i(0)) \right) + \frac{1}{N} \left(\sum_{j=1}^N M_{ij} (Y_j(0) - Y_i(0)) \right)^2,$$

where

$$\frac{2}{N-2}\sum_{k \neq i} M_{k0} \left(\sum_{j \neq i} M_{kj} (Y_j(0) - Y_k(0)) \right)$$

is unbiased for the middle term, using that $M_{kj}(Y_j(0) - Y_k(0)) = 0$ for $k = j$. It follows that

$$\hat{\mathbb{V}}_i = \frac{1}{N{-}3}\sum_{k \neq i} \left(\sum_{j \neq i} M_{kj} (Y_j(0) - Y_k(0)) \right)^2 - \frac{1}{(N{-}3)(N{-}2)} \sum_{k,j \neq i} M_{kj}^2 (Y_j(0) - Y_k(0))^2$$
$$+ \frac{2}{N-2}\sum_{k \neq i} M_{k0} \left(\sum_{j \neq i} M_{kj} (Y_j(0) - Y_k(0)) \right) + \frac{1}{N}\sum_k M_{k0}^2$$

is an unbiased estimator of the conditional variance $\mathbb{V}$. □

Proof of the variance expression in Lemma 5. This proof generalized the proof of Proposition 1 above. Specifically, for $[N] = \{1, \ldots, N\}$,

$$\sum_{\substack{k \subseteq [N];\\ |k| = N_T}} \sum_{\substack{i \subseteq [N] \setminus k;\\ |i| = N_T}} \sum_{j,j' \in [N] \setminus k \cup \{0\}} \frac{1}{\binom{|[N] \setminus (k \cup [\{j,j'\}])|}{N_T}} a_{i,j,j'} = \sum_{\substack{k \subseteq \{1,\ldots,N\};\\ |k| = N_T}} \sum_{j,j' \in [N] \cup \{0\}} a_{k,j,j'} \tag{A.2}$$

for a conformal tensor a.

For fixed t as above consider weights M_{kj} indexed by $k \subseteq [N]$ with $|k| = N_T$ and $j \in [N] \cup 0$, for which (a) $\sum_{j=1}^N M_{kj} = 0$ and (b) $M_{kj} = 1$ for $j \in k$, and potential outcomes $Y_j(0)$ with $j \in [N]$. Write $\overline{Y}_k(0) = \frac{1}{N_T}\sum_{j \in k} Y_j(0)$. (This approach generalizes to cases where treated units are themselves weighted, in which case we would replace $\overline{Y}_k(0)$ by the corresponding weighted average.) Let

$$b_{k,j} = \begin{cases} 0, & j \in k, \\ M_{kj}(Y_j(0) - \overline{Y}_k(0)), & j \in [N] \setminus k, \\ M_{k0}, & j = 0, \end{cases} \qquad a_{k,j,j'} = b_{k,j} b_{k,j'}. \tag{A.3}$$

Then, for $K = \binom{N}{N_T}$, and using that (c) $a_{k,j,j'} = 0$ whenever j or j' are in $[N] \setminus k$ and (d) $a_{k,j,j'} = a_{k,j',j}$,

$$\begin{aligned}
\mathbb{V} &= \frac{1}{K} \sum_{\substack{k \subseteq [N]; \\ |k| = N_T}} \left(M_{k0} + \sum_{j=1}^{N} M_{kj} Y_j(0) \right)^2 \overset{\text{(a)}}{=} \frac{1}{K} \sum_{\substack{k \subseteq [N]; \\ |k| = N_T}} \left(M_{k0} + \sum_{j=1}^{N} M_{kj} (Y_j(0) - \overline{Y}_k(0)) \right)^2 \\
&\overset{\text{(b)}}{=} \frac{1}{K} \sum_{\substack{k \subseteq [N]; \\ |k| = N_T}} \left(M_{k0} + \sum_{j \in [N] \setminus k} M_{kj} (Y_j(0) - \overline{Y}_k(0)) \right)^2 \overset{\text{(A.3.1)}}{=} \frac{1}{K} \sum_{\substack{k \subseteq [N]; \\ |k| = N_T}} \left(\sum_{j=0}^{N} b_{kj} \right)^2 \\
&\overset{\text{(A.3.2)}}{=} \frac{1}{K} \sum_{\substack{k \subseteq [N]; \\ |k| = N_T}} \sum_{j,j' \in [N] \cup \{0\}} a_{k,j,j'} \overset{\text{(A.2)}}{=} \frac{1}{K} \sum_{\substack{k \subseteq [N]; \\ |k| = N_T}} \sum_{\substack{i \subseteq [N] \setminus k; \\ |i| = N_T}} \sum_{j,j' \in [N] \setminus k \cup \{0\}} \frac{1}{\binom{|[N] \setminus (k \cup [\{j,j'\})|}{N_T}} a_{i,j,j'} \\
&\overset{\text{(c)}}{=} \frac{1}{K} \sum_{\substack{k \subseteq [N]; \\ |k| = N_T}} \sum_{\substack{i \subseteq [N] \setminus k; \\ |i| = N_T}} \left(\frac{a_{i,0,0}}{\binom{N_C}{N_T}} + \sum_{j \in [N] \setminus k} \frac{a_{i,j,0} + a_{i,0,j} + a_{i,j,j}}{\binom{N_C - 1}{N_T}} + \sum_{\substack{j,j' \in [N] \setminus k \\ j \neq j'}} \frac{a_{i,j,j'}}{\binom{N_C - 2}{N_T}} \right) \\
&\overset{\text{(d)}}{=} \frac{1}{K} \sum_{\substack{k \subseteq [N]; \\ |k| = N_T}} \left(a_{k,0,0} + \sum_{\substack{i \subseteq [N] \setminus k; \\ |i| = N_T}} \left(\sum_{j \in [N] \setminus k} \left(\frac{2 a_{i,j,0}}{\binom{N_C - 1}{N_T}} - \underbrace{\frac{N_T}{(N_C - 1)\binom{N_C - 2}{N_T}}}_{= \frac{1}{\binom{N_C - 2}{N_T}} - \frac{1}{\binom{N_C - 1}{N_T}}} a_{i,j,j} \right) + \sum_{j,j' \in [N] \setminus k} \frac{a_{i,j,j'}}{\binom{N_C - 2}{N_T}} \right) \right) \\
&\overset{\text{(A.3.2)}}{=} \frac{1}{K} \sum_{\substack{k \subseteq [N]; \\ |k| = N_T}} \left(b_{k,0}^2 + \sum_{\substack{i \subseteq [N] \setminus k; \\ |i| = N_T}} \left(\frac{2 b_{i,0} \sum_{j \in [N] \setminus k} b_{i,j}}{\binom{N_C - 1}{N_T}} - \frac{N_T \sum_{j \in [N] \setminus k} b_{i,j}^2}{(N_C - 1)\binom{N_C - 2}{N_T}} + \frac{\left(\sum_{j \in [N] \setminus k} b_{i,j} \right)^2}{\binom{N_C - 2}{N_T}} \right) \right) \\
&\overset{\text{(A.3.1)}}{=} \frac{1}{K} \left(\frac{1}{K} \sum_{\substack{i \subseteq [N]; \\ |i| = N_T}} M_{i0}^2 + \sum_{\substack{i \subseteq [N] \setminus k; \\ |i| = N_T}} \left\{ \frac{2}{\binom{N_C - 1}{N_T}} M_{i0} \sum_{j \in [N] \setminus k} M_{ij} (Y_j - \overline{Y}_i) \right. \right. \\
&\qquad \left. \left. + \frac{1}{\binom{N_C - 2}{N_T}} \left(\sum_{j \in [N] \setminus k} M_{ij} (Y_j - \overline{Y}_i) \right)^2 - \frac{N_T}{(N_C - 1)\binom{N_C - 2}{N_T}} \sum_{j \in [N] \setminus k} M_{ij}^2 (Y_j - \overline{Y}_i)^2 \right\} \right) \\
&= \frac{1}{K} \sum_{\substack{k \subseteq [N]; \\ |k| = N_T}} \hat{\mathbb{V}}_k,
\end{aligned}$$

so the proposed estimator is unbiased for the variance.

An alternative estimator that emphasizes the leave-fold-out nature of this construction is

$$\sum_{\substack{i\subseteq[N]\setminus k;\\ |i|=N_T}} \Bigg\{ \frac{1}{\binom{N_C}{N_T}} M_{i0}^2 + \frac{2}{\binom{N_C-1}{N_T}} M_{i0} \sum_{j\in[N]\setminus k} M_{ij}(Y_j - \overline{Y}_i)$$
$$+ \frac{1}{\binom{N_C-2}{N_T}} \left(\sum_{j\in[N]\setminus k} M_{ij}(Y_j - \overline{Y}_i) \right)^2 - \frac{N_T}{(N_C-1)\binom{N_C-2}{N_T}} \sum_{j\in[N]\setminus k} M_{ij}^2 (Y_j - \overline{Y}_i)^2 \Bigg\}.$$

It has the additional advantage that it does not use the weights on the treated observations when constructing the variance for a specific draw. Here, the first average of squared intercepts could be replaced by the overall average of squared intercepts, as in the main variance estimator above. □