Synthetic Controls for Experimental Design

Alberto Abadie Jinglong Zhao

MIT Boston University

February 2026

Abstract

This article studies experimental design in settings where the experimental units are large aggregate entities (e.g., markets), and only one or a small number of units can be exposed to the treatment. In such settings, randomization of the treatment may result in treated and control groups with substantially different baseline characteristics, inducing biases. We propose a variety of experimental non-randomized synthetic control designs (Abadie, Diamond and Hainmueller, 2010, Abadie and Gardeazabal, 2003) that select the units to be treated, as well as the untreated units to be used as a control group. Average potential outcomes with treatment are estimated as weighted averages of observed outcomes for treated units, and average potential outcomes without treatment as weighted averages of observed outcomes for control units. We analyze the properties of estimators based on synthetic control designs and propose new inferential techniques. We show that in experimental settings with aggregate units, synthetic control designs can substantially reduce estimation biases in comparison to randomization of the treatment.

Alberto Abadie, Department of Economics, MIT, abadie@mit.edu. Jinglong Zhao, Questrom School of Business, Boston University, jinglong@bu.edu. The authors are grateful to Victor Chernozhukov, Guido Imbens, Rahul Mazumder, Jaume Vives-i-Bastida, Yinchu Zhu and participants at Amazon.com, Brown, Harvard, National University of Singapore, Princeton, Stanford, the Online Causal Inference Seminar, USC, and the 2025 Joint Statistical Meetings for helpful comments and discussions. The replication code can be found at GitHub. Abadie gratefully acknowledges funding from NSF (SES-1756692) and ONR (N000142412687).

1. Introduction

Consider the problem of a ride-sharing company choosing between two compensation plans for drivers (Doudchenko et al., n.d.; Jones and Barrows, 2019 for related examples). The company can either keep the current compensation plan or adopt a new one with higher incentives. In order to estimate the effect of a change in compensation plans on profits, the company's data science unit designs an experimental evaluation where the new plan is deployed at a small scale, say, in one of the local markets (cities) in the country. In this setting, a randomized control trial — or A/B test, where drivers in a local market are randomized into the new plan (active treatment arm) or the status quo (control treatment arm) — is problematic. On the one hand, such an experiment raises equity concerns, as drivers in the same local market but in different treatment arms obtain different compensations for the same jobs. On the other hand, if drivers in the active treatment arm respond to higher incentives by working longer hours, they will effectively steal business from drivers in the control arm of the experiment, resulting in biased experimental estimates.[1]

One possible approach to this problem is to assign an entire local market to treatment, and use the rest of the local markets, which remain under the current compensation plan during the experimental periods, as potential comparison units. In this setting, using randomization to assign the active treatment allows ex-ante (i.e., pre-randomization) unbiased estimation of the effect of the active treatment. However, ex-post (i.e., post-randomization) biases can be large if, at baseline, the treated unit differs from the untreated units in the values of the features that affect the outcomes of interest. We document the magnitude and practical relevance of these biases in Sections 4 and 5.

[1] A randomized evaluation across many markets is a potential solution to the problem of experimental interference between drivers. In practice, however, large-scale market-level randomized evaluations are often unfeasible. In the context of the ride-sharing company example, large-scale market-level randomized evaluations *(i)* could be prohibitively expensive, *(ii)* could still raise substantial equity concerns, *(iii)* could negatively affect morale for the large number of drivers in the treated cities if the program is rolled back after experimentation, and *(iv)* in some cases, the number of cities where the company operates could be too small for effective randomization.

As in the ride-sharing example with only one treated local market, large biases may arise more generally in randomized studies when either the treatment arm or the control arm contains a small number of units, so randomized treatment assignment may not produce treated and control groups that are similar in their features (see, e.g., Bruhn and McKenzie, 2009). In those cases, the fact that estimation biases would have averaged out over alternative treatment assignments is of little comfort to a researcher who, in practice, is limited to one assignment only.

To address these challenges, we propose using the synthetic control method (Abadie, Diamond and Hainmueller, 2010, Abadie and Gardeazabal, 2003) as an experimental design to select treated units in non-randomized experiments, and the untreated units to serve as a comparison group. We adopt the name *synthetic control designs* for the resulting experimental designs.[2,3]

In our framework, the choice of the treated unit (or treated units, if multiple treated units are desired) aims to accomplish two goals. First, the treated units should be representative of an aggregate of interest, such as a national market, so that the estimated effect reflects the aggregate impact of the treatment. Second, the treated units should not be idiosyncratic in the sense that the untreated units cannot closely approximate their features. Otherwise, the reliability of the estimate of the effect on the treated unit may be questionable. We show how to achieve these two objectives, whenever they are possible to achieve, using synthetic control methods.

While we are aware of the extensive use of synthetic control methods for experimental design in data science units, especially in the technology industry,[4] the academic literature on this subject is at a nascent stage. There are, however, a few publicly available studies that are connected to this article. Aside from the present article, to our knowledge, Doudchenko et al. (n.d.) and Doudchenko et al. (2021) are the only other publicly available studies on the topic of experi-

[2]While we leave the "experimental" qualifier implicit in "synthetic control design", it should be noted that the synthetic control designs proposed in this article differ from observational synthetic control designs (e.g., Abadie, Diamond and Hainmueller, 2010, Abadie and Gardeazabal, 2003, Doudchenko and Imbens, 2016), for which the identity of the treated unit(s) is taken as given.

[3]See, e.g., Abadie (2021), Amjad, Shah and Shen (2018), Arkhangelsky et al. (2021), Doudchenko and Imbens (2016) for background material on synthetic controls and related methods.

[4]See, in particular, Jones and Barrows (2019) for applications of synthetic control methods in the ride-sharing context, and Ma (2017) for applications in the internet retail context.

mental design with synthetic controls. The focus of Doudchenko et al. (n.d.) is on statistical power, which they calculate by simulating the estimated effects of placebo interventions using historical (pre-experimental) data. That is, the selection of treated units is based on a measure of statistical power implied by the distribution of the placebo estimates for each unit. As a result, estimates based on the procedure in Doudchenko et al. (n.d.) target the effect of the treatment for the unit or units that are most closely tracked in the placebo distribution. In the same spirit, the target parameter in Doudchenko et al. (2021) is the treatment effect for a weighted average of treated units that can be closely matched in their pre-treatment outcomes by a weighted average of untreated units. In the present article, we aim to take a different perspective on the problem of unit selection in experiments with synthetic controls; one that can take into account the extent to which different sets of treated and control units approximate an aggregate causal effect of interest chosen by the analyst, such as the average treatment effect for the relevant population.[5] The inferential methods in the present article also differ from those in the related literature. In particular, Doudchenko et al. (2021) proposes a permutation procedure for inference that requires that potential outcomes without the treatment are independent and identically distributed (i.i.d.) in time. In contrast, the inferential procedure proposed in the present article allows for time series dependence and non-stationarity in outcomes, which are pervasive features of time-series data. Another important difference between the present article and Doudchenko et al. (n.d.) and Doudchenko et al. (2021) is that Doudchenko et al. (n.d.) and Doudchenko et al. (2021) make use of pre-treatment outcomes only to select treated and control units, while our method allows the use of other observed features of the units.

In a wider context, our methods are rooted in the broader framework of experimental non-randomized designs (see, e.g., Armstrong and Kolesár, 2018, Kasy, 2016, Thorlund et al., 2020). Yet, they diverge by addressing a distinct challenge: estimating synthetic control counterfactuals in experimental settings where only a limited number of aggregate units can be treated.

[5]Consistent with the majority of literature on synthetic controls, our focus is primarily on average treatment effects. For an analysis of distributional effects using synthetic controls, see Gunsilius (2023).

An alternative approach to control post-randomization bias involves stratifying units based on covariate values prior to randomization of treatment within each stratum. Stratification can significantly reduce post-randomization biases if units have similar covariate values within strata. However, traditional stratification methods do not adapt to the setting considered in this article, which features a limited number of large aggregate entities as units of analysis and a single unit or a handful chosen for treatment. Because every stratum in stratified designs must have at least one unit randomized into treatment, the number of strata cannot exceed the desired number of treated units in the experiment. In the case of only one treated unit, we would be limited to a single stratum. This may lead to significant variation in units' characteristics within strata, reducing the appeal of stratification procedures.

The rest of the article is organized as follows: Section 2 presents and discusses the synthetic control designs proposed in this article. Section 3 details the formal properties of estimators based on synthetic control designs and proposes inferential methods. In Section 4, we report the findings from an empirical validation of synthetic control designs using sales data from a sample of Walmart stores. Section 5 discusses the results of simulation studies. Finally, Section 6 provides concluding remarks. The appendix contains proofs and supplemental materials.

2. Synthetic Control Designs

2.1. Setup and Notation

We consider a setting with T time periods and J units, which may represent J local markets as in the ride-sharing example in the previous section. Let T_0 be the number of pre-experimental periods, with $1 \leq T_0 < T$. At the end of period T_0, a researcher designs an experiment to conduct during periods $T_0+1, T_0+2, \ldots, T$. Using the information available at T_0, the experimenter aims to select the set of units that will receive the treatment (intervention) during the experimental periods.

To define causal parameters, we formally adopt a potential outcomes framework. For any

$j \in \{1, \ldots, J\}$ and any $t \in \{T_0 + 1, \ldots, T\}$, Y_{jt}^I is the potential outcome for unit j at time t when unit j is exposed to treatment starting at $T_0 + 1$. Similarly, for any $j \in \{1, \ldots, J\}$ and any $t \in \{1, \ldots, T\}$, Y_{jt}^N is the potential outcome for unit j at time t under no treatment in all periods. In the ride-sharing example, Y_{jt}^I and Y_{jt}^N could measure net revenue divided by market size under the active and the control treatment, respectively. Unit-level treatment effects are defined as

$$Y_{jt}^I - Y_{jt}^N,$$

for $j = 1, \ldots, J$ and $t = T_0 + 1, \ldots, T$. They represent the effect of switching unit j to the active treatment at time $T_0 + 1$ on the outcome of unit j at time $t > T_0$. We aim to estimate the average treatment effect

$$\tau_t = \sum_{j=1}^{J} f_j \cdot (Y_{jt}^I - Y_{jt}^N), \tag{1}$$

for $t = T_0 + 1, \ldots, T$. In this expression, $f_1, \ldots, f_J$ are known positive weights that define the average of interest. In the ride-sharing example from the previous section, f_j may represent the size of local market j as a share of the national market. Without loss of generality, and because it is often the case in applications, we can assume that the weights f_j sum to one,

$$\sum_{j=1}^{J} f_j = 1.$$

When units are equally weighted, we set $f_j = 1/J$ for $j = 1, \ldots, J$. We use the notation $\boldsymbol{f}$ for a vector that collects the values of f_j for all the units, i.e., $\boldsymbol{f} = (f_1, \ldots, f_J)$.

2.2. The Experimenter's Problem

At time T_0, in order to estimate the treatment effect τ_t for $t = T_0 + 1, \ldots, T$, the experimenter chooses $\boldsymbol{w} = (w_1, \ldots, w_J)$ and $\boldsymbol{v} = (v_1, \ldots, v_J)$, such that

$$\begin{aligned} \sum_{j=1}^{J} w_j &= 1, \\ \sum_{j=1}^{J} v_j &= 1, \\ w_j \geq 0,\ v_j \geq 0,\ &\text{and } w_j v_j = 0,\ \forall j = 1, \ldots, J. \end{aligned} \tag{2}$$

Units with $w_j > 0$ are units that will be assigned to the intervention of interest from $T_0 + 1$ to T, and will be used to estimate average outcomes under the intervention. Units with $w_j = 0$ constitute an untreated reservoir of potential control units (a "donor pool"). Among units with $w_j = 0$, those with $v_j > 0$ will be used to estimate average outcomes under no intervention.

The first goal of the experimenter is to choose $w_1, \ldots, w_J$ such that

$$\sum_{j=1}^{J} w_j Y_{jt}^I = \sum_{j=1}^{J} f_j Y_{jt}^I, \tag{3}$$

for $t = T_0 + 1, \ldots, T$. If equation (3) holds, a weighted average of outcomes for the units selected for treatment reproduces the average outcome with treatment for the entire population of J units. In practice, however, the choice of $w_1, \ldots, w_J$ cannot directly rely on matching the population average of Y_{jt}^I, as in equation (3). The quantities Y_{jt}^I are unobserved before time $T_0 + 1$, and will remain unobserved in the experimental periods for the units that are not exposed to the treatment. Instead, we aim to approximate equation (3) using predictors observed at T_0 of the values of $Y_{jT_0+1}^I, \ldots, Y_{jT}^I$. Note also that it is not possible to use the weights $w_1 = f_1, \ldots, w_J = f_J$, because it would leave no units in the donor pool, making the set of units with $v_j > 0$ empty and violating equation (2).

The second goal of the experimenter is to choose $v_1, \ldots, v_J$ such that

$$\sum_{j=1}^{J} v_j Y_{jt}^N = \sum_{j=1}^{J} f_j Y_{jt}^N, \tag{4}$$

or, alternatively,

$$\sum_{j=1}^{J} v_j Y_{jt}^N = \sum_{j=1}^{J} w_j Y_{jt}^N. \tag{5}$$

If equations (4) or (5) hold, a weighted average of outcomes for the units in the donor pool reproduces the average outcome without treatment for the entire population of J units (equation (4)), or for the units selected for treatment (equation (5)). Like in the previous case with treated outcomes, it is not feasible to directly choose $v_1, \ldots, v_J$ so that equation (4) or (5) is satisfied. Instead, we propose a variety of methods to approximate either (4) or (5) based on predictors of $Y_{jT_0+1}^N, \ldots, Y_{jT}^N$.

For the treated units, we define $Y_{jt} = Y_{jt}^N$ if $t = 1, \ldots, T_0$, and $Y_{jt} = Y_{jt}^I$ if $t = T_0 + 1, \ldots, T$. For the untreated units, we define $Y_{jt} = Y_{jt}^N$, for all $t = 1, \ldots, T$. That is, Y_{jt} is the outcome observed for unit $j = 1, \ldots, J$ at time $t = 1, \ldots, T$. We say that

$$\sum_{j=1}^{J} w_j Y_{jt} \quad \text{and} \quad \sum_{j=1}^{J} v_j Y_{jt}$$

are the synthetic treated and synthetic control outcomes, respectively. The difference between these two quantities is

$$\tau_t(\boldsymbol{w}, \boldsymbol{v}) = \sum_{j=1}^{J} w_j Y_{jt} - \sum_{j=1}^{J} v_j Y_{jt},$$

for $t = T_0 + 1, \ldots, T$. Suppose that equations (3) and (4) hold. Then, $\tau_t(\boldsymbol{w}, \boldsymbol{v})$ is equal to the average treatment effect, τ_t. If equation (5) holds instead, then $\tau_t(\boldsymbol{w}, \boldsymbol{v})$ is equal to the average

treatment effect on the treated ($\boldsymbol{w}$-weighted),

$$\tau_t^T = \sum_{j=1}^{J} w_j \cdot (Y_{jt}^I - Y_{jt}^N) \tag{6}$$

(Doudchenko et al., 2021).

2.3. Synthetic Control Designs

The experimenter chooses $\boldsymbol{w} = (w_1, \ldots, w_J)$ and $\boldsymbol{v} = (v_1, \ldots, v_J)$ to match the pre-intervention values of predictors of the potential outcomes Y_{jt}^N and Y_{jt}^I for $t > T_0$.

Let $\boldsymbol{X}_j$ be a column vector of pre-intervention features of unit j. We view the features in $\boldsymbol{X}_j$ as predictors of the values of Y_{jt}^N and Y_{jt}^I in the experimental periods, in a sense that will be made precise in Section 3. We use the notation

$$\overline{\boldsymbol{X}} = \sum_{j=1}^{J} f_j \boldsymbol{X}_j.$$

That is, $\overline{\boldsymbol{X}}$ is the vector of population values for the predictors in $\boldsymbol{X}_j$. For any real vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|$ is the Euclidean norm of $\boldsymbol{x}$, and $\|\boldsymbol{x}\|_0$ is the number of non-zero coordinates of $\boldsymbol{x}$. Let $\underline{m}$ and $\overline{m}$ be positive integers such that $1 \leq \underline{m} \leq \overline{m} \leq J - 1$.

A simple way to choose weights $\boldsymbol{w} = (w_1, \ldots, w_J)$ and $\boldsymbol{v} = (v_1, \ldots, v_J)$ is to solve the optimization problem

$$\begin{aligned} \min_{\substack{w_1,\ldots,w_J,\\ v_1,\ldots,v_J}} \quad & \left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j \right\|^2 + \left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} v_j \boldsymbol{X}_j \right\|^2 \\ \text{s.t.} \quad & \sum_{j=1}^{J} w_j = 1, \\ & \sum_{j=1}^{J} v_j = 1, \end{aligned}$$

$$
\begin{aligned}
& w_j, v_j \geq 0, \quad \forall j = 1, \ldots, J, \\
& w_j v_j = 0, \quad \forall j = 1, \ldots, J, \\
& \underline{m} \leq \|\boldsymbol{w}\|_0 \leq \overline{m}.
\end{aligned} \tag{7}
$$

The first term of the objective function in (7) measures the discrepancies between the population average of the features in $\boldsymbol{X}_j$ ($\boldsymbol{f}$-weighted) and the averages of the features for units assigned to the treatment group ($\boldsymbol{w}$-weighted). The second term is analogous but with the second average taken over the units assigned to no intervention ($\boldsymbol{v}$-weighted). The first four constraints require that the weights in $\boldsymbol{w}$, as well as the weights in $\boldsymbol{v}$, are non-negative and sum to one. They also require that any unit selected for treatment cannot serve as a control unit, i.e., if $w_j > 0$, then $v_j = 0$, and vice versa. The last constraint allows a minimum and maximum number of units assigned to treatment. This restriction is of practical importance in a variety of contexts, especially when experimentation is costly and the experimenter is restricted in the number of units that may receive the treatment. We say that the design is *Unconstrained* if $\underline{m} = 1$ and $\overline{m} = J - 1$; otherwise, we say the design is *Constrained*. The cardinality constraint in (7) is not the only conceivable restriction to the size or cost of the experiment. An explicit upper bound on the cost of an experiment would be given by $\boldsymbol{c}'\boldsymbol{d} \leq B$, where the j-th coordinate of $\boldsymbol{c}$ is equal to the cost of assigning unit j to treatment, $\boldsymbol{d}$ is a J-dimensional vector with ones at coordinates where $w_j > 0$, and zeros otherwise, and B is the experimenter's budget.

Let $\boldsymbol{w}^* = (w_1^*, \ldots, w_J^*)$ and $\boldsymbol{v}^* = (v_1^*, \ldots, v_J^*)$ be a solution to the optimization problem in (7). In practice, we do not require optimality of $(\boldsymbol{w}^*, \boldsymbol{v}^*)$, as long as $(\boldsymbol{w}^*, \boldsymbol{v}^*)$ is feasible and satisfies $\overline{\boldsymbol{X}} - \sum_{j=1}^J w_j^* \boldsymbol{X}_j \approx \boldsymbol{0}$ and $\overline{\boldsymbol{X}} - \sum_{j=1}^J v_j^* \boldsymbol{X}_j \approx \boldsymbol{0}$, where $\boldsymbol{0}$ is a vector of zeros of the same dimension as $\boldsymbol{X}_j$. Suppose that units with $w_j^* > 0$ are assigned to treatment in the experiment, and units with $w_j^* = 0$ are kept untreated. A synthetic control estimator of τ_t is $\widehat{\tau}_t = \tau_t(\boldsymbol{w}^*, \boldsymbol{v}^*)$,

i.e.,

$$\widehat{\tau}_t = \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt}. \tag{8}$$

This estimator is based on approximations to equations (3) and (4) that rely on $\boldsymbol{X}_j$, the observed predictors of the potential outcomes, Y_{jt}^N, and Y_{jt}^I. Note that for every solution to (7) with $\underline{m} \leq \|\boldsymbol{v}\|_0 \leq \overline{m}$, there exists another solution that swaps the roles of the treated and the untreated in the experiment without altering the value of the objective function.

In what follows, we take the formulation in (7) as a starting point for synthetic control designs and modify it in several ways. A second formulation of the synthetic control design is based on equations (3) and (5), which we refer to as the *Weakly targeted* design,

$$\begin{aligned}
\min_{\substack{w_1,\ldots,w_J,\\ v_1,\ldots,v_J}} \quad & \left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j \right\|^2 + \beta \left\| \sum_{j=1}^{J} w_j \boldsymbol{X}_j - \sum_{j=1}^{J} v_j \boldsymbol{X}_j \right\|^2 \\
\text{s.t.} \quad & \sum_{j=1}^{J} w_j = 1, \\
& \sum_{j=1}^{J} v_j = 1, \\
& w_j, v_j \geq 0, \quad \forall j = 1, \ldots, J, \\
& w_j v_j = 0, \quad \forall j = 1, \ldots, J, \\
& \underline{m} \leq \|\boldsymbol{w}\|_0 \leq \overline{m}.
\end{aligned} \tag{9}$$

The parameter $\beta > 0$ reflects the trade-off between selecting treated units to fit the aggregate value of the predictors $\overline{\boldsymbol{X}}$, and selecting control units to fit the aggregate value of the treated units. A small value of β favors designs with treated units that closely match $\overline{\boldsymbol{X}}$, and prioritizes estimation of τ_t, the average treatment effect. A large value of β, on the other hand, favors designs with aggregate treated and aggregate control units that closely match each other, and prioritizes estimation of τ_t^T, the $\boldsymbol{w}$-weighted average treatment effect on the treated. While it

is possible to use data-driven selectors of β, the rule of thumb $\beta = 1$ provides a natural choice that equally weights the two terms in the objective function in (9). For this formulation of the synthetic control design, the treatment assignment and estimation procedures follow the same steps as those used in the formulation in (7).

In our third formulation of the synthetic control design, we select treated units to fit the aggregate value of the predictors $\overline{\boldsymbol{X}}$. But unlike the design in (9), we choose multiple synthetic control units, one for each unit that contributes to the synthetic treated unit. For any J-dimensional vector of non-negative coordinates, $\boldsymbol{w} = (w_1, \ldots, w_J)$, let $\mathcal{J}_{\boldsymbol{w}}$ be the set of the indices with non-zero coordinates, $\mathcal{J}_{\boldsymbol{w}} = \{j : w_j > 0\}$. We refer to this formulation as the *Unit-level* design. Like the *Weakly targeted* design, this formulation estimates the average treatment effect on the treated, as defined in equation (6), but typically relies on a small set of treated units even in the absence of a sparsity constraint on $\boldsymbol{w}$ (i.e., when $\overline{m} = J - 1$).

$$
\begin{aligned}
\min_{\substack{w_j, \forall j=1,2,..,J, \\ v_{ij}, \forall i,j=1,2,...,J}} \quad & \left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j \right\|^2 + \xi \sum_{j=1}^{J} w_j \left\| \boldsymbol{X}_j - \sum_{i=1}^{J} v_{ij} \boldsymbol{X}_i \right\|^2 \\
\text{s.t.} \quad & \sum_{j=1}^{J} w_j = 1, \\
& w_j \geq 0, \quad \forall j = 1, \ldots, J, \\
& \sum_{i=1}^{J} v_{ij} = 1, \quad \forall j \in \mathcal{J}_{\boldsymbol{w}}, \\
& v_{ij} = 0, \quad \forall i \in \mathcal{J}_{\boldsymbol{w}},\ j = 1, \ldots, J, \\
& v_{ij} \geq 0, \quad \forall j \in \mathcal{J}_{\boldsymbol{w}},\ i = 1, \ldots, J, \\
& v_{ij} = 0, \quad \forall j \notin \mathcal{J}_{\boldsymbol{w}},\ i = 1, \ldots, J, \\
& \underline{m} \leq \|\boldsymbol{w}\|_0 \leq \overline{m}.
\end{aligned} \tag{10}
$$

The parameter $\xi > 0$ reflects the trade-offs between selecting treated units to fit the aggregate

value of the predictors $\overline{\boldsymbol{X}}$, and selecting control units to fit the values of the predictors for each treated unit. A small value of ξ favors experimental designs with treated units that closely match $\overline{\boldsymbol{X}}$. A large value of ξ, on the other hand, favors designs where the values of the predictors for the treated units are closely matched by those of their synthetic controls.

Let $\{w_j^*, v_{ij}^*\}_{i,j=1,\ldots,J}$ be a solution of the optimization problem in (10). As before, we do not strictly require optimality of $\{w_j^*, v_{ij}^*\}_{i,j=1,\ldots,J}$, provided $\{w_j^*, v_{ij}^*\}_{i,j=1,\ldots,J}$ is feasible and $\overline{\boldsymbol{X}} - \sum_{j=1}^J w_j^* \boldsymbol{X}_j \approx \mathbf{0}$ and $\boldsymbol{X}_j - \sum_{j=1}^J v_{ij}^* \boldsymbol{X}_j \approx \mathbf{0}$ for all j such that $w_j^* > 0$. Assign units with $w_j^* > 0$ to treatment in the experiment, and keep units with $w_j^* = 0$ untreated. Let

$$v_j^* = \sum_{i=1}^J w_i^* v_{ij}^*. \tag{11}$$

Then,

$$\begin{aligned}\widehat{\tau}_t &= \sum_{j=1}^J w_j^* Y_{jt} - \sum_{j=1}^J v_j^* Y_{jt} \\ &= \sum_{j=1}^J w_j^* \left(Y_{jt} - \sum_{i=1}^J v_{ij}^* Y_{it} \right).\end{aligned}$$

Our next adjustment to the synthetic control design is motivated by settings where experimental units may be naturally divided into clusters with similar values in the predictors, $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_J$. For example, weather patterns, which may be highly dependent across cities in the same region (e.g., Northeast, Midwest, etc., in the US), may influence the seasonality of the demand for ride-sharing services. In those cases, it is natural to treat each cluster (each region, in our example) as a distinct experimental design to ameliorate interpolation biases. Figure 1 illustrates this point. Panels (a) and (b) depict identical samples in the space of the predictors. In this simple example, we have two predictors only, and their values for each unit are represented by the coordinates of the dots in the figure. Red dots represent units assigned to treatment. All other units are plotted as black dots. Panel (a) visualizes the result of treating the entire sample

as one cluster. Three units are assigned to treatment. They closely reproduce the value of $\overline{\boldsymbol{X}}$, but they all fall in the same central cluster, far away from observations in other clusters. In panel (b), assignment to treatment takes into account the clustered nature of the data, and one unit is treated per cluster. This provides a better approximation of the distribution of the predictor values for the entire sample, ameliorating concerns of interpolation biases.

Section 5 and the online appendix present extensive simulation evidence on the relative performance of different formulations of the synthetic control design. In the simulations, the baseline formulation in (7) generally has the most accurate estimates of the average treatment effect, as it fits both the synthetic treated and control units to reproduce the population-level predictor vector $\overline{\boldsymbol{X}}$. By prioritizing the match of aggregate predictor values for the treated, the *Weakly targeted* formulation in (9) achieves high relative accuracy for the average treatment effect on the treated when β is large. With large ξ, the *Unit-level* formulation (10) also yields accurate estimates, but remains less accurate than the *Weakly targeted* design. Its main advantage is parsimony: even without a hard cardinality constraint, it selects fewer treated units by emphasizing unit-level fits. Finally, the online appendix introduces a penalized version of the synthetic control design, which further limits the number of units in both the synthetic treated and control groups, even in the absence of a hard cardinality constraint.

3. Formal Results

This section introduces an extension of the linear factor model commonly employed in the synthetic control literature and use it to analyze the properties of estimators based on synthetic control designs.

Assumption 1 *Potential outcomes follow a linear factor model,*

$$Y_{jt}^N = \delta_t + \boldsymbol{\theta}_t' \boldsymbol{Z}_j + \boldsymbol{\lambda}_t' \boldsymbol{\mu}_j + \epsilon_{jt}, \tag{12a}$$

$$Y_{jt}^I = \upsilon_t + \boldsymbol{\gamma}_t' \boldsymbol{Z}_j + \boldsymbol{\eta}_t' \boldsymbol{\mu}_j + \xi_{jt}, \tag{12b}$$

where $\boldsymbol{Z}_j$ is a $(R \times 1)$ vector of observed covariates, $\boldsymbol{\theta}_t$ and $\boldsymbol{\gamma}_t$ are $(R \times 1)$ vectors of unknown parameters, $\boldsymbol{\mu}_j$ is a $(F \times 1)$ vector of unobserved covariates, $\boldsymbol{\lambda}_t$ and $\boldsymbol{\eta}_t$ are $(F \times 1)$ vectors of unknown parameters, and ϵ_{jt} and ξ_{jt} are unobserved idiosyncratic noise terms whose distributions do not depend on the other components of the linear factor model.

Equation (12a) is the linear factor model for potential outcomes under no treatment, a benchmark commonly used in the literature to analyze the properties of synthetic control estimators (see, e.g., Abadie, Diamond and Hainmueller, 2010, Ferman, 2021). Equation (12b) extends the linear factor structure to potential outcomes under treatment. The reason for this extension is that, in contrast to synthetic control estimation with observational data, synthetic control designs require the choice of a treatment group in addition to the choice of a comparison group.

We employ the covariates in $\boldsymbol{Z}_j$ as well as pre-experimental values of the outcome variable Y_{jt} to construct the vectors of predictors, $\boldsymbol{X}_j$. In particular, let $\mathcal{E} \subseteq \{1, \ldots, T_0\}$, let $T_{\mathcal{E}} = |\mathcal{E}|$ (i.e., the cardinality of $\mathcal{E}$), and let $\boldsymbol{Y}_j^{\mathcal{E}}$ be the $(T_{\mathcal{E}} \times 1)$ vector of $T_{\mathcal{E}}$ pre-experimental outcomes for unit j and time indices in $\mathcal{E}$. We define

$$\boldsymbol{X}_j = \begin{pmatrix} \boldsymbol{Y}_j^{\mathcal{E}} \\ \boldsymbol{Z}_j \end{pmatrix},$$

for $j = 1, \ldots, J$. That is, the vector of predictors $\boldsymbol{X}_j$ collects the covariates in $\boldsymbol{Z}_j$ and the pre-experimental outcome values Y_{jt} for the *fitting periods* in $\mathcal{E}$. In practice, the values in $\boldsymbol{X}_j$ are often scaled to make them independent of units of measurement or to reflect the relative importance of each of the predictors (see, e.g., Abadie, 2021).

The next assumption gathers regularity conditions on model primitives.

Assumption 2

(i) $F \leq T_{\mathcal{E}}$. *Moreover, let* $\boldsymbol{\lambda}_{\mathcal{E}}$ *be the* $(T_{\mathcal{E}} \times F)$ *matrix with rows equal to the* $\boldsymbol{\lambda}_t$*'s indexed by* $\mathcal{E}$. *Let* $\zeta_{\mathcal{E}}$ *be the smallest eigenvalue of* $\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}}$. *Then,* $\underline{\zeta} = \zeta_{\mathcal{E}}/T_{\mathcal{E}} > 0$.

(ii) For each $j = 1, \ldots, J$, $\epsilon_{j1}, \ldots, \epsilon_{jT}$ *is a sequence of i.i.d. sub-Gaussian random variables with mean zero and variance proxy* $\overline{\sigma}^2$. *For any* $j = 1, \ldots, J$, $\xi_{jT_0+1}, \ldots, \xi_{jT}$ *is a sequence of i.i.d. sub-Gaussian random variables with mean zero, variance proxy* $\overline{\sigma}^2$, *and independent of* $\epsilon_{j1}, \ldots, \epsilon_{jT}$.

Assumption 2*(i)* is similar to conditions in Abadie, Diamond and Hainmueller (2010). Assumption 2*(ii)* is similar to conditions in Abadie, Diamond and Hainmueller (2010), Doudchenko and Imbens (2016), Chernozhukov, Wüthrich and Zhu (2021), and Arkhangelsky et al. (2021). Sub-Gaussianity is not strictly necessary, but it simplifies the form of our results. It can be relaxed by assuming bounded finite-order moments (instead of bounding the entire moment generating function). At the same time, sub-Gaussianity is a relatively mild assumption. It holds for any Gaussian distribution, as well as any distribution with a bounded support. Distributions with heavy tails, such as the Cauchy distribution, are not sub-Gaussian. Notably, Assumption 2*(ii)* allows for dependence of ϵ_{jt} and ξ_{jt} across units.

Unless otherwise noted, all probability statements are over the joint distribution of ϵ_{jt} and ξ_{jt} and conditional on the values of the other components on the right-hand sides of equations (12a) and (12b). The next assumption pertains to the quality of the synthetic control fit. For concreteness, we focus on the base design in (7), and choose $\boldsymbol{w}^* = (w_1^*, \ldots, w_J^*)$ and $\boldsymbol{v}^* = (v_1^*, \ldots, v_J^*)$ so that the synthetic treated and synthetic control units reproduce the average values of $\boldsymbol{X}_j$.

Assumption 3 *With probability one, (i)*

$$\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j = \sum_{j=1}^{J} v_j^* \boldsymbol{Z}_j = \sum_{j=1}^{J} f_j \boldsymbol{Z}_j,$$

and (ii)

$$\sum_{j=1}^{J} w_j^* \boldsymbol{Y}_j^{\mathcal{E}} = \sum_{j=1}^{J} v_j^* \boldsymbol{Y}_j^{\mathcal{E}} = \sum_{j=1}^{J} f_j \boldsymbol{Y}_j^{\mathcal{E}}.$$

Assumption 3 implies that the synthetic treated and control units defined by $\boldsymbol{w}^*$ and $\boldsymbol{v}^*$ provide a perfect fit for $\overline{\boldsymbol{X}}$. Assumption 3 is a strong restriction, which may only hold approximately in practice. The next assumption relaxes the perfect fit condition in Assumption 3.

Assumption 4 *There exists a positive constant $d > 0$, such that with probability one, (i)*

$$\Big\| \sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} f_j \boldsymbol{Z}_j \Big\|_2^2 \leq Rd^2, \qquad \Big\| \sum_{j=1}^{J} v_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} f_j \boldsymbol{Z}_j \Big\|_2^2 \leq Rd^2,$$

and (ii)

$$\Big\| \sum_{j=1}^{J} w_j^* \boldsymbol{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} f_j \boldsymbol{Y}_j^{\mathcal{E}} \Big\|_2^2 \leq T_{\mathcal{E}} d^2, \qquad \Big\| \sum_{j=1}^{J} v_j^* \boldsymbol{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} f_j \boldsymbol{Y}_j^{\mathcal{E}} \Big\|_2^2 \leq T_{\mathcal{E}} d^2.$$

Let $\lambda_{t,f}$ be the f-th coordinate of $\boldsymbol{\lambda}_t$, and

$$\overline{\lambda} = \max_{\substack{t=1,\ldots,T \\ f=1,\ldots,F}} |\lambda_{tf}|.$$

We define η_{tf}, θ_{tr}, γ_{tr}, $\overline{\eta}$, $\overline{\theta}$ and $\overline{\gamma}$ analogously, so $|\eta_{tf}| \leq \overline{\eta}$ for $t = T_0 + 1, \ldots, T$, $f = 1, \ldots, F$, $|\theta_{tr}| \leq \overline{\theta}$ for $t = 1, \ldots, T$, $r = 1, \ldots, R$, and $|\gamma_{tr}| \leq \overline{\gamma}$, for $t = T_0 + 1, \ldots, T$, $r = 1, \ldots, R$. Next theorem extends results on the bias of synthetic control estimators (see, e.g., Abadie, Diamond and Hainmueller, 2010, Vives-i-Bastida, 2022) to the experimental setup of Section 2.

Theorem 1 *If Assumptions 1 – 3 hold, then for any $t \geq T_0 + 1$,*

$$|E\left[\widehat{\tau}_t - \tau_t\right]| \leq \frac{\overline{\lambda}(\overline{\eta} + \overline{\lambda})F}{\underline{\zeta}} \sqrt{2 \log(2J)} \frac{\overline{\sigma}}{\sqrt{T_{\mathcal{E}}}}. \tag{13}$$

If Assumptions 1, 2, and 4 hold, then for any $t \geq T_0 + 1$,

$$|E\left[\widehat{\tau}_t - \tau_t\right]| \leq \left((\overline{\gamma} + \overline{\theta})R + \frac{\overline{\lambda}(\overline{\eta} + \overline{\lambda})F}{\underline{\zeta}}(1 + \overline{\theta}R)\right)d + \frac{\overline{\lambda}(\overline{\eta} + \overline{\lambda})F}{\underline{\zeta}}\sqrt{2\log(2J)}\frac{\overline{\sigma}}{\sqrt{T_{\mathcal{E}}}}. \quad (14)$$

Expectations are over the distribution of $\{\epsilon_{jt}\}_{j\in\{1,\ldots,J\},t\in\{1,\ldots,T\}}$ *and* $\{\xi_{jt}\}_{j\in\{1,\ldots,J\},t\in\{T_0+1,\ldots,T\}}$.

Note that, while the factor model in equations (12a) and (12b) leave the sign and scale of $\boldsymbol{\lambda}_t$ and $\boldsymbol{\eta}_t$ free (e.g., multiplying $\boldsymbol{\lambda}_t$ and dividing $\boldsymbol{\mu}_t$ by the same non-zero constant does not change the value of $\boldsymbol{\lambda}_t'\boldsymbol{\mu}_j$), the value of the bound in Theorem 1 is invariant to changes in the sign or the scale of $\boldsymbol{\lambda}_t$ and $\boldsymbol{\eta}_t$. Moreover, the bound in (14) does not depend on the scale of $\boldsymbol{Z}_j$, because changing the scale of $\boldsymbol{Z}_j$ leaves the product $\overline{\theta}d$ unchanged. The scale of Y_{jt} does affect the bound in (14) because the treatment effect τ_t is measured in the same units as Y_{jt}. The results in Theorem 1 do not depend on the specific formulation of the synthetic control design (e.g., *Constrained* vs. *Unconstrained*).

The bias bounds (13) and (14) depend on the ratio between the scale of ϵ_{jt}, represented by $\overline{\sigma}$, and the number of fitting periods $T_{\mathcal{E}}$. Intuitively, the bias of the synthetic control estimator is small when a good fit in pre-experimental outcomes (Assumption 3) is obtained by implicitly fitting the values of the latent variables, μ_j. Overfitting happens when pre-experimental outcomes are instead fitted out of the variability in the individual transitory shocks, ϵ_{jt}. A small number of fitting periods $T_{\mathcal{E}}$ combined with large variability in ϵ_{jt} increases the risk of overfitting and, as a result, increases the bias bound. Similarly, for any fixed value of $T_{\mathcal{E}}$, the bias bound increases with J, reflecting the increased risk of over-fitting caused by increased variability in ϵ_{jt} over larger donor pools. Finally, the number of unobserved factors F enters the bound (13) linearly, which highlights the importance of including the observed predictors $\boldsymbol{Z}_j$ — other than pre-experimental outcomes — in the vector of fitting variables $\boldsymbol{X}_j$. Under the factor model in equations (12a) and (12b), observed predictors not included in $\boldsymbol{Z}_j$ are shifted to $\boldsymbol{\mu}_j$, increasing F and the magnitude

of the bound.[6]

We next turn our attention to inference. We utilize a set of *blank periods*, $\mathcal{B} \subseteq \{1, \ldots, T_0\} \setminus \mathcal{E}$, which comprise pre-experimental periods whose outcomes Y_{jt} have not been used to calculate $\boldsymbol{w}^*$ or $\boldsymbol{v}^*$. Because pre-experimental periods that are not in $\mathcal{E}$ or $\mathcal{B}$ are not used in our procedure, without loss of generality, we consider $\mathcal{B} = \{1, \ldots, T_0\} \setminus \mathcal{E}$. Therefore, we assume that the number of elements of $\mathcal{B}$ is $T_{\mathcal{B}} = |\mathcal{B}| = T_0 - T_{\mathcal{E}}$. We aim to test the following null hypothesis:

For $t = T_0 + 1, \ldots, T$, and $j = 1, \ldots, J$,

$$Y_{jt}^I = \delta_t + \boldsymbol{\theta}_t' \boldsymbol{Z}_j + \boldsymbol{\lambda}_t' \boldsymbol{\mu}_j + \xi_{jt}, \tag{15}$$

where ξ_{jt} has the same distribution as ϵ_{jt}.

Under the null hypothesis in (15), the distribution of Y_{jt}^I is the same as the distribution of Y_{jt}^N, for $t = T_0 + 1, \ldots, T$, and $j = 1, \ldots, J$. But the realized values of Y_{jt}^I and Y_{jt}^N may differ.

Recall from (8) that, for $t \in \{T_0 + 1, \ldots, T\}$, a synthetic control estimator is defined as

$$\widehat{\tau}_t = \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt}.$$

Let $\widehat{u}_t = \widehat{\tau}_t, \forall t \in \{T_0 + 1, \ldots, T\}$ be the synthetic control estimator in the experimental periods. Similarly, for each $t \in \mathcal{B}$ in the blank periods, let

$$\widehat{u}_t = \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt}.$$

[6]Shifting predictors from $\boldsymbol{Z}_j$ to $\boldsymbol{\mu}_j$ changes the bias bound (13) in a more complex manner than what might be inferred from a cursory look at the bias formula. First, moving predictors from $\boldsymbol{Z}_j$ to $\boldsymbol{\mu}_j$ also means shifting components of $\boldsymbol{\theta}_t$ to $\boldsymbol{\lambda}_t$, which can change the value of $\underline{\zeta}$. Poincaré's separation theorem implies that $\underline{\zeta}$ cannot increase as a result of this shift. Moreover, moving predictors from $\boldsymbol{Z}_j$ to $\boldsymbol{\mu}_j$ cannot decrease the values of $\overline{\lambda}$ and $\overline{\eta}$. Overall, the value of the bias bound in (13) cannot decrease and will typically increase by moving predictors from $\boldsymbol{Z}_j$ to $\boldsymbol{\mu}_j$. This is not necessarily true for the bound in (14), because a shift of components from $\boldsymbol{Z}_j$ to $\boldsymbol{\mu}_j$ decreases the value of R.

Such $\widehat{u}_t$ for $t \in \mathcal{B}$ are placebo treatment effects estimated for the blank periods. We study the properties of a test based on combinations from the set $\{\widehat{u}_t\}_{t\in\mathcal{B}\cup\{T_0+1,\ldots,T\}}$.

We define Π as the set of all $(T-T_0)$-combinations of $\mathcal{B}\cup\{T_0+1,\ldots,T\}$. That is, for each $\pi \in \Pi$, π is a subset of indices from the blank periods and the experimental periods $\mathcal{B}\cup\{T_0+1,\ldots,T\}$, such that $|\pi| = T - T_0$. The cardinality of Π is $|\Pi| = (T - T_{\mathcal{E}})!/((T - T_0)!(T_0 - T_{\mathcal{E}})!)$. For each $\pi \in \Pi$, let $\pi(i)$ be the i^{th} smallest value in π, and

$$\widehat{\boldsymbol{u}}_\pi = (\widehat{u}_{\pi(1)}, \widehat{u}_{\pi(2)}, ..., \widehat{u}_{\pi(T-T_0)}).$$

In addition, let $\widehat{\boldsymbol{u}} = (\widehat{u}_{T_0+1}, \ldots, \widehat{u}_T) = (\widehat{\tau}_{T_0+1}, \ldots, \widehat{\tau}_T)$. This is a vector of treatment effect estimates from the experimental periods. We adopt the following test statistic,

$$S(\widehat{\boldsymbol{u}}) = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} |\widehat{u}_t|. \tag{16}$$

Other choices of test statistics are possible, such as those based on an L_p-norm of $\widehat{\boldsymbol{u}}$ (Chernozhukov, Wüthrich and Zhu, 2021) and one-sided versions of the resulting test statistics (i.e., with the positive or the negative parts of $\widehat{u}_t$ replacing $|\widehat{u}_t|$ in equation (16)).

For $\widehat{\boldsymbol{u}}_\pi \in \Pi$, define $S(\widehat{\boldsymbol{u}}_\pi) = \frac{1}{T-T_0} \sum_{t\in\pi} |\widehat{u}_t|$. The p-value of a permutation test on (16) is

$$\widehat{p} = \frac{1}{|\Pi|} \sum_{\pi\in\Pi} \mathbb{1}\{S(\widehat{\boldsymbol{u}}_\pi) \geq S(\widehat{\boldsymbol{u}})\} \tag{17}$$

Theorem 2 below shows that if $\boldsymbol{\lambda}_t$ are exchangeable random variables for $t \in \mathcal{B} \cup \{T_0+1,\ldots,T\}$, then a test of the null hypothesis in (15) based on the p-value in (17) is exact.

Theorem 2 *Suppose that Assumptions 1, 2(ii), and 3(i) hold, and the noises* $\{\epsilon_{jt}\}_{t\in\mathcal{B}\cup\{T_0+1,\ldots,T\}}$ *and* $\{\xi_{jt}\}_{t\in\{T_0+1,\ldots,T\}}$ *have continuous distributions. Assume that* $\{\boldsymbol{\lambda}_t\}_{t\in\mathcal{B}\cup\{T_0+1,\ldots,T\}}$ *is a sequence*

of exchangeable random variables. Under the null hypothesis (15)*, for any* $\alpha \in [0,1]$*, we have*

$$\alpha - \frac{1}{|\Pi|} \leq \Pr(\widehat{p} \leq \alpha) \leq \alpha,$$

where $\Pr(\widehat{p} \leq \alpha)$ *is over the distribution of* $\{\epsilon_{jt}\}_{j\in\{1,\ldots,J\},t\in\{1,\ldots,T\}}$, $\{\xi_{jt}\}_{j\in\{1,\ldots,J\},t\in\{T_0+1,\ldots,T\}}$ *and* $\{\boldsymbol{\lambda}_t\}_{t\in\{1,\ldots,T\}}$.

Under the assumptions of Theorem 2, the potential outcome series Y_{jt}^N is allowed to be non-stationary through the term $\delta_t + \boldsymbol{\theta}_t'\boldsymbol{Z}_j$ in equation (12a). This is in contrast to a related result in Doudchenko et al. (2021), which requires that the potential outcomes Y_{jt}^N are i.i.d. over time.

The assumptions in Theorem 2 build upon those in Theorem 1. Although these assumptions are simple and sufficient for the result of the theorem, they can be substantially relaxed. Under exchangeability of $\boldsymbol{\lambda}_t$, if the i.i.d. condition in Assumption 2*(ii)* is violated, the result for Theorem 2 holds if for each $j = 1,\ldots,J$, $\{\epsilon_{jt}\}_{t\in\mathcal{B}\cup\{T_0+1,\ldots,T\}}$ and $\{\xi_{jt}\}_{t\in\{T_0+1,\ldots,T\}}$ are sequences of exchangeable random variables. Second, if Assumption 3*(i)* is violated, the result for Theorem 2 holds if $\{(\boldsymbol{\theta}_t,\boldsymbol{\lambda}_t)\}_{t\in\mathcal{B}\cup\{T_0+1,\ldots,T\}}$ is a sequence of exchangeable random variables independent of $\{\epsilon_{jt}\}_{t\in\mathcal{B}\cup\{T_0+1,\ldots,T\}}$ and $\{\xi_{jt}\}_{t\in\{T_0+1,\ldots,T\}}$. In the above two cases under exchangeability of $\boldsymbol{\lambda}_t$, we still have exact p-value. Finally, exchangeability of $\boldsymbol{\lambda}_t$ is a strong restriction. Theorem OA.1 in the online appendix relaxes this restriction by showing that for fixed $\boldsymbol{\lambda}_t$ (i.e., without resorting to exchangeability of $\boldsymbol{\lambda}_t$), the p-value in (17) is still approximately valid for large $T_{\mathcal{E}}$.

In some settings, the number of possible combinations, $|\Pi|$, could be very large, making exact calculation of $\widehat{p}$ computationally expensive. In those instances, random samples from Π can be used to approximate the p-value in equation (17).

The inferential technique proposed in this article is related to, but distinct from, the permutation methods in Abadie, Diamond and Hainmueller (2010), Chernozhukov, Wüthrich and Zhu (2019), Chernozhukov, Wüthrich and Zhu (2021), Firpo and Possebom (2018), Lei and Candès (2021), and others. Inferential methods that reassign treatment across units (e.g., Abadie, Dia-

mond and Hainmueller, 2010) are not appropriate for the designs of Section 2, which explicitly select treated and control units to satisfy an optimality criterion.

Similar to Chernozhukov, Wüthrich and Zhu (2021), our method is based on rearrangements of estimated treatment effects across time periods. But unlike Chernozhukov, Wüthrich and Zhu (2021), which proposes permutations over all periods, including the pre-intervention periods, our inferential method permutes only over the blank periods and post-intervention periods, which are not used to estimate the weights in the synthetic control design. Relative to Chernozhukov, Wüthrich and Zhu (2021), the generative models of equations (12a) and (12b), which allow for unobserved factors, and the finite sample nature of the results require a novel testing procedure that, similar to split conformal prediction methods (Lei et al., 2018, Vovk, Gammerman and Shafer, 2005), takes advantage of the availability of blank periods. Lei and Sudijono (2025) introduces a refinement to conformal inference in synthetic control settings, particularly effective when the number of rearrangements (in our case, $|\Pi|$) is small.

Confidence intervals for τ_t can be constructed using split conformal inference methods. For any $\alpha \in (0,1)$, let

$$\widehat{q}_{1-\alpha} = \inf_{z \in \mathbb{R}} \left\{ \frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\left\{ \left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| \leq z \right\} \geq 1 - \alpha \right\} \tag{18}$$

be the empirical $(1-\alpha)$-quantile on the absolute values of placebo treatment effects in the blank periods, and

$$\widehat{C}_{1-\alpha}(Y_{1t}, Y_{2t}, ..., Y_{Jt}) = \left[\sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} - \widehat{q}_{1-\alpha},\ \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} + \widehat{q}_{1-\alpha} \right]. \tag{19}$$

We next show that the confidence interval defined in (19) approximately achieves correct pointwise coverage in large samples if treatment does not change the distribution of the idiosyncratic noises.

Theorem 3 *Assume that Assumptions 1– 3 hold. Assume there exists a constant $\kappa < \infty$, such that for all $j = 1, \ldots, J$, $t = 1, \ldots, T$, ϵ_{jt} are continuously distributed with the probability density function upper bounded by κ. Assume that for $t = T_0 + 1, \ldots, T$, and $j = 1, \ldots, J$, ξ_{jt} has the same distribution as ϵ_{jt}. Then the confidence interval defined in* (19) *approximately achieves point-wise coverage, i.e., for any $\alpha \in (0, 1)$ and any $t \in \{T_0 + 1, ..., T\}$, as $(T_0 - T_{\mathcal{E}}), T_{\mathcal{E}} \to +\infty$,*

$$\left| \Pr\Big(\tau_t \in \widehat{C}_{1-\alpha}(Y_{1t}, Y_{2t}, ..., Y_{Jt})\Big) - (1-\alpha) \right| \\ = O\Big(\big(\log(T_0 - T_{\mathcal{E}})/(T_0 - T_{\mathcal{E}})\big)^{1/2} + \big(\log T_{\mathcal{E}}/T_{\mathcal{E}}\big)^{1/2}\Big) \longrightarrow 0.$$

where the probability $\Pr\big(\tau_t \in \widehat{C}_{1-\alpha}(Y_{1t}, Y_{2t}, ..., Y_{Jt})\big)$ *is over the distribution of* $\{\epsilon_{jt}\}_{j\in\{1,...,J\},t\in\{1,...,T\}}$ *and* $\{\xi_{jt}\}_{j\in\{1,...,J\},t\in\{T_0+1,...,T\}}$.

4. Empirical Illustration Using Walmart Data

In this section, we illustrate the applicability of the methods in this article using store-level data from Walmart (Prakash, 2023). The dataset is a balanced panel of weekly sales for $J = 45$ Walmart stores and $T = 143$ weeks, spanning the period from the week of February 5, 2010, to the week of October 26, 2012. We estimate the effect of a placebo intervention and show that the methods of Section 3 produce point estimates that are close to zero and a test result that does not reject the null hypothesis in (15) for the placebo intervention.

We consider the design of a fictitious experiment across stores taking place on July 20, 2012 (week 129 in the data). Out of the $T_0 = 128$ pre-experimental weeks, we take the first $T_{\mathcal{E}} = 100$ weeks as the fitting period, and the last $(T_0 - T_{\mathcal{E}}) = 28$ weeks as the blank period. The number of weeks in the experimental period is $T - T_0 = 15$. The outcomes $\{Y_{jt}\}_{j=1,...,J,t=1,...,T}$ are weekly sales (units of revenue are undisclosed in the data). We use uniform weights $f_j = 1/J$ for $j = 1, ..., J$, to average sales across all stores. To estimate the synthetic treated and synthetic

control weights, we normalize each of the 100 pre-experimental outcomes to have unit variance.

We compute synthetic treated and control units that apply the *Constrained* formulation in (7) with $\overline{m} = 2$. We adopt $\overline{m} = 2$ because using only one store for the synthetic treated fails to produce a good fit between the resulting synthetic treated and synthetic control units during the fitting period (as measured by root-mean square error or mean absolute error). Increasing to $\overline{m} = 3$ brings only marginal improvements in fit. Figures 2 and 3 report results for $\overline{m} = 2$. Results for $\overline{m} = 1$ and $\overline{m} = 3$ appear in the online appendix.

Figure 2 reports the time series of weekly sales for the synthetic treated unit (black solid line), the synthetic control unit (black dashed line), and for each individual store in the dataset (blue dashed lines). Weekly sales for the synthetic treated and the synthetic control units closely follow each other during the fitting period. The gap between the two synthetic units remains small after the fitting period, indicating good out-of-sample predictive power in the absence of intervention.

Figure 3 reports the difference in weekly sales between the synthetic treated and the synthetic control units. The p-value of equation (17), calculated over the residuals of Figure 3, is equal to 0.933, which results in a failure to reject the null hypothesis (15). Confidence intervals based on equation (19) cover zero for all t in the experimental period.

Table 1 compares the performance of the synthetic control design to those of straight randomization followed by difference-in-means, randomization after stratification on pre-intervention outcomes followed by difference-in-means, and 1- and 5-nearest neighbor adjustment after randomization. In particular, Table 1 reports out-of-sample root mean square error (RMSE) over the post-intervention period, normalized by the post-intervention outcome mean (see Section 5.2.2 for a precise definition of the estimators and RMSE performance metric). For each of the four randomization-based estimators, the reported RMSE is the average over 1000 randomized treatment assignments. The synthetic control design dominates all other alternatives, even when it uses only the outcomes in the fitting periods to construct the synthetic treated and synthetic

control units, whereas stratification and nearest-neighbor adjustment utilize all pre-intervention outcomes.

5. Simulation Study

This section presents simulation results that showcase the behavior of estimators based on synthetic control designs. The simulation design employs $J = 15$ units, $R = 7$ observed covariates, and $F = 11$ unobserved covariates. Each simulated sample covers $T = 30$ periods, with $T_0 = 25$ periods before the intervention and five after the intervention. Synthetic control weights are estimated using the first $T_{\mathcal{E}} = 20$ periods, leaving $t = 21, \ldots, 25$ as blank periods. The weights f_j in expression (1) are set to $f_j = 1/J$ for all $j = 1, \ldots, J$.

For our baseline simulation design, we use the linear factor model in Assumption 1 to generate potential outcomes. For $t = 1, \ldots, T$, we generate the series δ_t as small-to-large re-arrangements of T i.i.d. Uniform $(0, 20)$ random variables. For $t = T_0 + 1, \ldots, T$, we generate the series υ_t as small-to-large re-arrangements of $T - T_0$ i.i.d. Uniform $(0, 20)$ random variables. For $j = 1, \ldots, J$, we set both $\boldsymbol{Z}_j$ and $\boldsymbol{\mu}_j$ to be random vectors of i.i.d. Uniform $(0, 1)$ random variables. For $t = 1, \ldots, T$, we set $\boldsymbol{\theta}_t$, $\boldsymbol{\gamma}_t$, $\boldsymbol{\lambda}_t$, and $\boldsymbol{\eta}_t$ to be random vectors of i.i.d. Uniform $(0, 10)$ random variables. Finally, for $j = 1, \ldots, J$, and any $t = 1, \ldots, T$, we set ϵ_{jt} and ξ_{jt} to be i.i.d. Normal $\mathcal{N}(0, \sigma^2)$ random variables, with $\sigma^2 = 1$. We present additional simulation results of alternative values of the noise parameter σ^2 in the online appendix.

The simulations in this section cover the *Unconstrained* and the *Constrained* designs. For the *Constrained* design, we consider $\overline{m} = 1, \ldots, 7$. For the *Unconstrained* design, the synthetic treated and synthetic control weights can always be swapped without changing the value of the objective function in (7). For the *Constrained* design, the weights can be swapped when $\|\boldsymbol{v}^*\|_0 \leq \overline{m}$. When it is possible to swap synthetic treated and synthetic control weights, we choose the treated units so that the number of units with positive weights in $\boldsymbol{w}^*$ is smaller than the number of units with positive weights in $\boldsymbol{v}^*$. When $\|\boldsymbol{w}^*\|_0 = \|\boldsymbol{v}^*\|_0$, we determine whether

to swap using a specific rule described in the online appendix.

5.1. Results for a Single Simulation

Using the data generating process described above, we draw a single sample and conduct the synthetic control design in (7), with parameters $\underline{m} = 1$ and $\overline{m} = 14$, i.e., no constraint on the number of treated units. We report the results in Figures 4 and 5. In Figure 4, each blue dashed line represents an outcome trajectory Y_{jt}, for $t = 1, \ldots, T$ and $j = 1, \ldots J$. The solid black line represents the trajectory of the synthetic treated unit $\sum_{j=1}^{J} w_j^* Y_{jt}$, for $t = 1, \ldots, T$. The black dashed line represents the trajectory of the synthetic control unit $\sum_{j=1}^{J} v_j^* Y_{jt}$, for $t = 1, \ldots, T$. The synthetic treated and synthetic control units closely track each other in the pre-experimental periods. They diverge during the experimental periods, when a treatment effect emerges as a result of the differences in the parameters of the data-generating processes for Y_{jt}^N and Y_{jt}^I. Figure 5 reports the difference between the synthetic treated and the synthetic control outcomes. The inferential procedure of Section 3 produces p-value equal to 0.004 for the null hypothesis of no treatment effect in (15).

5.2. Performance across Many Simulations

This section compares the performance of the different versions of the synthetic control designs over 1000 simulations that independently generate the model primitives (i.e., the factor loadings, covariates, and error terms) of Assumption 1.

5.2.1. Average Treatment Effects

The first panel of Table 2 reports true values of τ_t, averaged over 1000 simulations. The first seven columns of the second panel report $\widehat{\tau}_t$ for each of the five experimental periods, mean absolute error, and root mean square error, averaged over 1000 simulations. Mean absolute error

(MAE) and root mean square error (RMSE) are defined as

$$\text{MAE} = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} |\widehat{\tau}_t - \tau_t|, \qquad \text{RMSE} = \sqrt{\frac{1}{T - T_0} \sum_{t=T_0+1}^{T} (\widehat{\tau}_t - \tau_t)^2},$$

The last three columns of Table 2 report average p-values, rejection rates, and average number of treated units, over 1000 simulations.

Because the treatment effect is not equal to zero in the simulation of Table 2, smaller p-values and larger rejection rates reflect better performance of the testing procedure for a particular design.

Table 2 shows how $\overline{m}$ affects the performance of the synthetic control design. As expected, the *Unconstrained* design delivers the best performance across all measures, at the cost of using a larger number of treated units than the *Constrained* design. Relative to the single-treated-unit case, performance metrics improve substantially when allowing $\overline{m} = 2$ or $\overline{m} = 3$, with smaller gains beyond that point.

5.2.2. Comparison to Randomized Treatment Assignment

Randomized treatment assignment produces ex-ante (pre-randomization) unbiased estimation of the average treatment effect. As we show below, however, ex-post (post-randomization) biases can be large, especially when only a small number of units are treated.

In this section, we adopt the same setup as in the previous section to compare the performance of the synthetic control design with that of randomized designs, with or without pre- or post-stratification, in settings with a small number of treated units.

Let D_j denote a treatment indicator equal to one if unit j is randomized into the treated group and zero otherwise. Table 3 evaluates the performance of the following experimental design and estimation strategies:

1. *SC*: *Constrained* formulation of the synthetic control design. The results reproduce those

of Table 2.

2. *RND*: Randomized assignment of $\overline{m}$ units to treatment followed by the difference in means estimator,

$$\frac{1}{\overline{m}} \sum_{j=1}^{J} D_j Y_{jt} - \frac{1}{J - \overline{m}} \sum_{j=1}^{J} (1 - D_j) Y_{jt}.$$

3. *STR*: Stratification in $\overline{m}$ blocks, with at least two units per block. In each block, one unit is assigned to treatment at random. Blocks are chosen to minimize the maximal within-block discrepancy in the covariates, $\boldsymbol{Z}_j$, and pre-experimental outcomes (all normalized to have unit variance). Let B_{jk} be a binary variable that equals one if and only if unit j belongs to block k. The estimator of τ_t is

$$\sum_{k=1}^{\overline{m}} \frac{J_k}{J} \left(\sum_{j=1}^{J} B_{jk} D_j Y_{jt} - \frac{1}{J_k - 1} \sum_{j=1}^{k} B_{jk} (1 - D_j) Y_{jt} \right),$$

   where J_k represents the number of units within the k-th block.

4. *REG*: Randomized assignment of $\overline{m}$ units to treatment followed by regression adjustment on the covariates, $\boldsymbol{Z}_j$. Ordinary least-squares adjustment on all pre-treatment outcomes is unfeasible because the number of pre-treatment outcomes exceeds the number of units in the sample.

5. *1-NN* and *5-NN*: Randomized assignment of $\overline{m}$ units to treatment followed by 1-nearest neighbor and 5-nearest neighbor matching, respectively, on all pre-experimental outcomes and covariates. In both cases, predictors are rescaled to have unit variance.

Across all values of $\overline{m}$, the synthetic control design outperforms randomized assignment, including variants that incorporate pre-stratification, post-stratification, or regression adjustment. The online appendix reports similar results from simulations based on a nonlinear data-generating

process. Combined with the results in Table 1, these findings demonstrate the power of synthetic controls as an effective design strategy for experiments with aggregate units.

6. Conclusions

Experimental design methods have largely been concerned with settings where a large number of experimental units are randomly assigned to a treatment arm, and a similarly large number of experimental units are assigned to a control arm. This focus on large samples and randomization has proven to be enormously useful in various classes of problems, but becomes inadequate when treating more than a few units is unfeasible, as is often the case in experimental studies with large aggregate units (e.g., markets). In that case, randomized designs may produce estimators that are substantially biased (post-randomization) relative to the average treatment effect or to the average treatment effect on the treated. Large biases can be expected when the unit or units assigned to treatment fail to approximate average outcomes under treatment for the entire population or when the units in the control arm fail to approximate the outcomes that treated units would experience without treatment.

In this article, we have proposed synthetic control techniques, widely used in observational studies, to design experiments when the treatment can only be applied to a small number of experimental units. The synthetic control design optimizes jointly over the identities of the units assigned to the treatment and the control arms and over the weights that determine the relative contribution of those units to reproduce the counterfactuals of interest. We propose various designs to estimate average treatment effects, analyze the properties of such designs and the resulting estimators, and devise inferential methods to test a null hypothesis of no treatment effects and construct confidence intervals. In addition, we report results from an application to retail sales data and simulation results that demonstrate the applicability and computational feasibility of the methods proposed in this article. We show that synthetic control design can substantially outperform randomized designs in experimental settings with a small number of

treated units.

Corporate researchers, policymakers, and academic investigators are often confronted with settings where interventions at the micro-unit level (e.g., customers, workers, or families) are unfeasible, impractical, or ineffective (see, e.g., Duflo, Glennerster and Kremer, 2007, Jones and Barrows, 2019). Consequently, there is broad scope for experimental design methods targeting large aggregate entities (such as regional markets, school districts, or states), a setting where synthetic control designs offer a powerful tool for data-driven evaluation of treatment effects.

References

Abadie, Alberto. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature*, 59(2): 391–425.

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

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

Abadie, Alberto, and Jérémy L'Hour. 2021. "A Penalized Synthetic Control Estimator for Disaggregated Data." *Journal of the American Statistical Association*, 116(536): 1817–1834.

Amjad, Muhammad, Devavrat Shah, and Dennis Shen. 2018. "Robust Synthetic Control." *Journal of Machine Learning Research*, 19(22): 1–51.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2021. "Synthetic Difference-in-Differences." *American Economic Review*, 111(12): 4088–4118.

Armstrong, Timothy B., and Michal Kolesár. 2018. "Finite-Sample Optimal Estimation and Inference on Average Treatment Effects Under Unconfoundedness." *arXiv e-prints*, 1712.04594v2.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2021. "The Augmented Synthetic Control Method." *Journal of the American Statistical Association*, 116(536): 1789–1803.

Bobkov, SG, and GP Chistyakov. 2014. "Bounds on the Maximum of the Density for Sums of Independent Random Variables." *Journal of Mathematical Sciences*, 199(2): 100–106.

Bruhn, Miriam, and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics*, 1(4): 200–232.

Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu. 2019. "Distributional Conformal Prediction." *arXiv e-prints*, 1909.07889.

Chernozhukov, Victor, Kaspar Wüthrich, and Yinchu Zhu. 2021. "An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls." *Journal of the American Statistical Association*, 116(536): 1849–1864.

Doudchenko, Nick, Khashayar Khosravi, Jean Pouget-Abadie, Sebastien Lahaie, Miles Lubin, Vahab Mirrokni, Jann Spiess, et al. 2021. "Synthetic Design: An Optimization Approach to Experimental Design with Synthetic Controls." *Advances in Neural Information Processing Systems*, 34.

Doudchenko, Nikolay, and Guido W Imbens. 2016. "Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis." National Bureau of Economic Research.

Doudchenko, Nikolay, David Gilinson, Sean Taylor, and Nils Wernerfelt. n.d. "Designing Experiments with Synthetic Controls." https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2020/03/Wernerfelt-Nils-Doudchenko-Nick-Gilinson-David-and-Taylor-Sean_Designing-Experiments-with-Synthetic-Controls.pdf.

Duflo, Esther, Rachel Glennerster, and Michael Kremer. 2007. "Using Randomization in Development Economics Research: A Toolkit." *Handbook of development economics*, 4: 3895–3962.

Ferman, Bruno. 2021. "On the Properties of the Synthetic Control Estimator with Many Periods and Many Controls." *Journal of the American Statistical Association*, 116(536): 1764–1772.

Firpo, Sergio, and Vitor Possebom. 2018. "Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets." *Journal of Causal Inference*, 6(2): 20160026.

Gunsilius, F. F. 2023. "Distributional Synthetic Controls." *Econometrica*, 91(3): 1105–1117.

Jones, Nick, and Sam Barrows. 2019. "Synthetic Control and Alternatives to A/B Testing at Uber." *Presented at PyData Amsterdam 2019*, https://youtu.be/j5DoJV5S2Ao.

Kasy, Maximilian. 2016. "Why Experimenters Might Not Always Want to Randomize, and What They Could Do Instead." *Political Analysis*, 24(3): 324–338.

Lei, Jing, Max G'Sell, Alessandro Rinaldo, Ryan J Tibshirani, and Larry Wasserman. 2018. "Distribution-Free Predictive Inference for Regression." *Journal of the American Statistical Association*, 113(523): 1094–1111.

Lei, Lihua, and Emmanuel J Candès. 2021. "Conformal Inference of Counterfactuals and Individual Treatment Effects." *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(5): 911–938.

Lei, Lihua, and Timothy Sudijono. 2025. "Inference for Synthetic Controls via Refined Placebo Tests." *arXiv e-prints*, 2401.07152.

Ma, Emma. 2017. "Using Geographic Splitting & Optimization Techniques to Measure Marketing Performance." https://www.aboutwayfair.com/tech-innovation/using-geographic-splitting-optimization-techniques-to-measure-marketing-performance.

Oliveira, Roberto I, Paulo Orenstein, Thiago Ramos, and João Vitor Romano. 2022. "Split conformal prediction for dependent data." *arXiv e-prints*, 2203.15885.

Prakash, Sourav. 2023. "Walmart Condensed Sales Data." https://www.kaggle.com/ds/3471234.

Rigollet, Philippe, and Jan-Christian Hütter. 2019. "High Dimensional Statistics." http://www-math.mit.edu/~rigollet/PDFs/RigNotes17.pdf.

Thorlund, Kristian, Louis Dron, Jay JH Park, and Edward J Mills. 2020. "Synthetic and external controls in clinical trials–a primer for researchers." *Clinical epidemiology*, 457–467.

Vives-i-Bastida, Jaume. 2022. "Predictor Selection for Synthetic Controls." *arXiv e-prints*, 2203.11576.

Vovk, Vladimir, Alex Gammerman, and Glenn Shafer. 2005. *Algorithmic Learning in a Random World.* Springer Science & Business Media.

Table 1: Out-of-Sample Normalized Root Mean Square Error

	SC	RND	STR	1-NN	5-NN
$\overline{m} = 1$	0.052	0.452	0.452	0.096	0.082
$\overline{m} = 2$	0.018	0.312	0.299	0.070	0.063
$\overline{m} = 3$	0.019	0.254	0.173	0.059	0.053
$\overline{m} = 4$	0.027	0.223	0.181	0.052	0.048
$\overline{m} = 5$	0.012	0.202	0.164	0.047	0.043

Note: Root mean square error divided by the average outcome in the experimental periods. $\overline{m}$ stands for the maximum number of treated units. SC: *Constrained* formulation of the synthetic control design. RND: Randomized treatment assignment followed by the difference-in-means estimator. STR: Stratified randomization, followed by difference in means in each stratum. 1-NN: Randomized treatment assignment followed by 1-nearest neighbor matching, using all pre-experimental outcomes. 5-NN: Randomized treatment assignment followed by 5-nearest neighbor matching, using all pre-experimental outcomes.

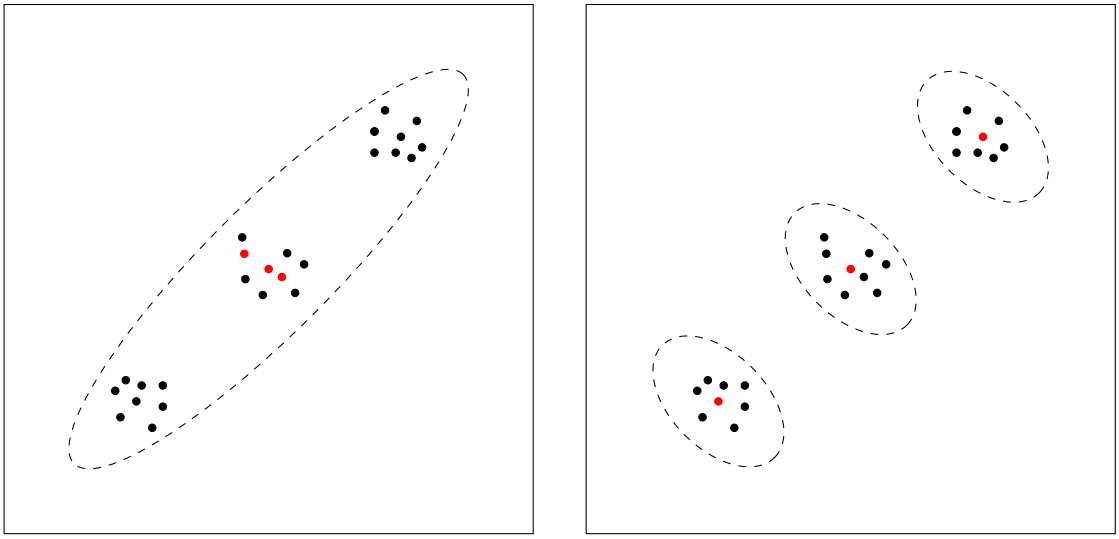
Table 2: Average Treatment Effects (Averages over 1000 Simulations)

		τ_t									
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$					
		-13.58	-10.99	-8.35	-5.00	-2.50					
		$\widehat{\tau}_t$					*MAE*	*RMSE*	$\widehat{p}$	$\widehat{p}<0.05$	$\|\boldsymbol{w}\|_0$
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$					
Unconstrained		-13.57	-10.97	-8.38	-5.07	-2.53	0.83	0.97	0.014	0.944	6.76
Constrained	$\overline{m}=1$	-13.61	-10.97	-8.39	-4.86	-2.41	2.93	3.45	0.057	0.668	1
	$\overline{m}=2$	-13.58	-10.90	-8.43	-5.01	-2.40	1.69	2.00	0.028	0.854	2
	$\overline{m}=3$	-13.56	-11.00	-8.38	-5.05	-2.52	1.26	1.49	0.019	0.916	3
	$\overline{m}=4$	-13.59	-11.06	-8.40	-4.99	-2.50	1.06	1.25	0.016	0.935	4
	$\overline{m}=5$	-13.57	-11.01	-8.37	-5.02	-2.48	0.93	1.09	0.015	0.933	5.00
	$\overline{m}=6$	-13.51	-10.95	-8.29	-5.01	-2.47	0.87	1.02	0.015	0.942	5.97
	$\overline{m}=7$	-13.57	-10.96	-8.37	-5.06	-2.52	0.83	0.97	0.014	0.946	6.76

Table 3: RMSE for Different Experimental Designs and Estimators (Averages over 1000 Simulations)

	SC	RND	STR	REG	1-NN	5-NN
$\overline{m} = 1$	3.45	6.35	6.35	8.14	5.28	4.40
$\overline{m} = 2$	2.00	4.70	3.53	6.00	3.69	3.20
$\overline{m} = 3$	1.49	3.91	2.75	5.11	3.02	2.66
$\overline{m} = 4$	1.25	3.49	2.44	4.49	2.67	2.40
$\overline{m} = 5$	1.09	3.22	2.07	4.12	2.38	2.28
$\overline{m} = 6$	1.02	3.04	1.95	3.87	2.24	2.23
$\overline{m} = 7$	0.97	3.01	1.85	3.90	2.18	2.32

Note: SC: *Constrained* formulation of the synthetic control design. RND: Randomized treatment assignment followed by the difference-in-means estimator. STR: Stratified randomization, followed by difference in means in each stratum. REG: Randomized treatment assignment followed by regression adjustment. 1-NN: Randomized treatment assignment followed by 1-nearest neighbor matching. 5-NN: Randomized treatment assignment followed by 5-nearest neighbor matching. SC uses outcomes in the fitting periods and covariates as predictors. STR, 1-NN, and 5-NN use all pre-intervention outcomes and covariates. REG adjusts for the covariates only.



(a) (b)

Figure 1: Clustering in a synthetic control design

Note: Panels (a) and (b) plot the values of the predictors in $\boldsymbol{X}_j$, which is bivariate in this simple example. Units assigned to treatment are drawn in red. In panel (a), we treat the entire sample as a single cluster. In panel (b), we divide the sample into three clusters and assign one unit in each cluster to the treatment.

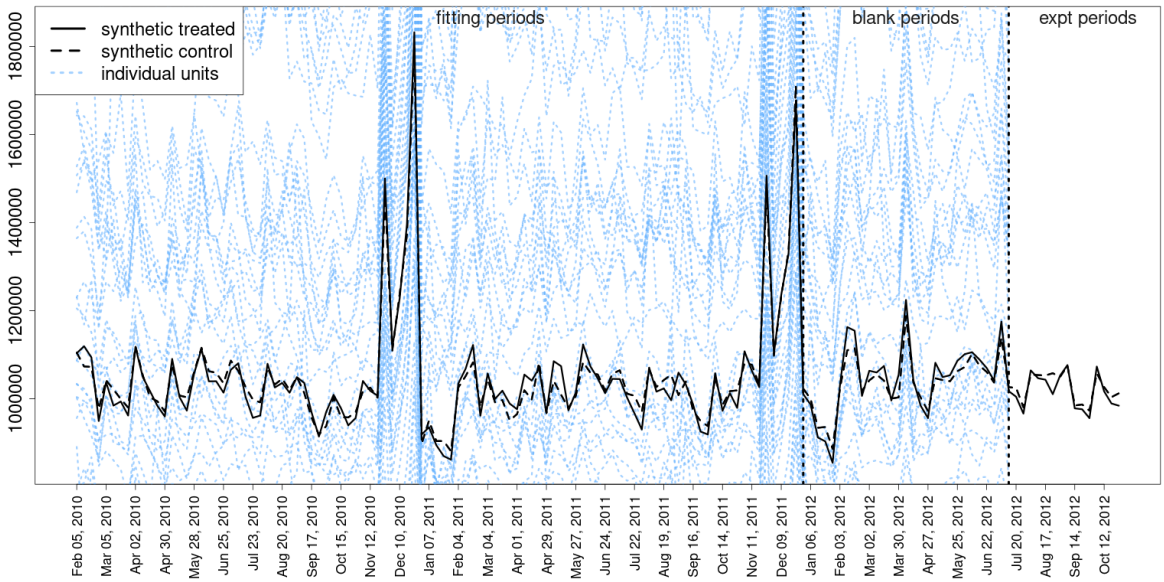


Figure 2: Synthetic Treatment Unit and Synthetic Control Unit, $\overline{m} = 2$

Note: The black solid line represents the synthetic treated outcome. The black dashed line represents the synthetic control outcome. The blue dashed lines are individual stores' sales.

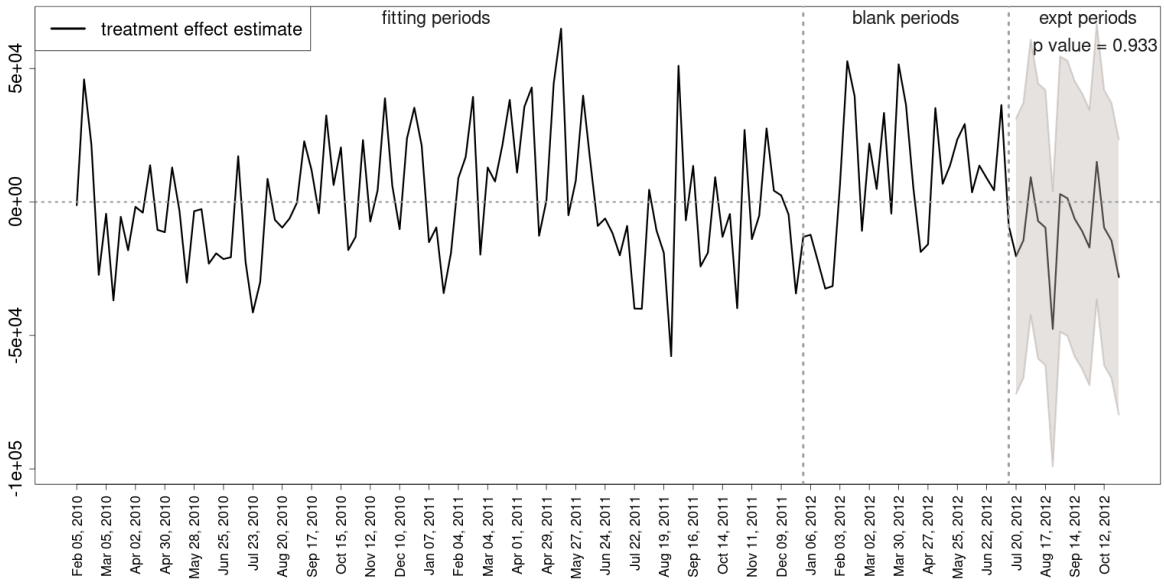


Figure 3: Treatment Effect Estimate, when $\overline{m} = 2$.

Note: This figure reports the difference between the synthetic treated and synthetic control outcomes of Figure 2. For the experimental periods, this is the treatment effect estimate. The shaded region indicates the 95% confidence interval for each of the experimental periods.

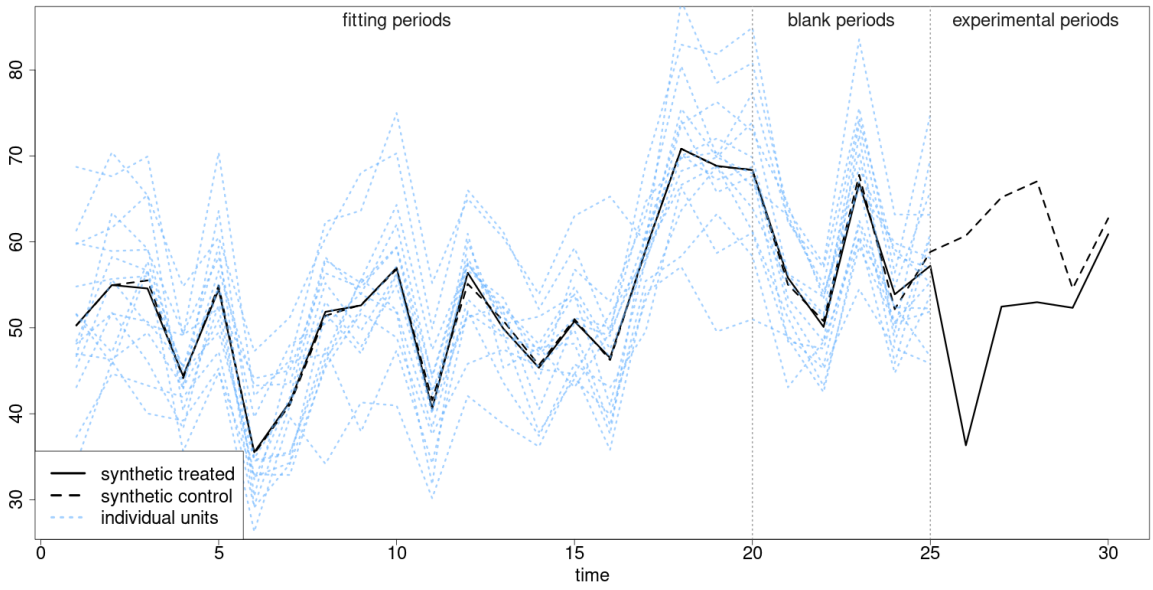


Figure 4: Synthetic Treatment Unit and Synthetic Control Unit, $\sigma^2 = 1$

Note: The black solid line represents the synthetic treated outcome ($\boldsymbol{w}^*$-weighted). The black dashed line represents the synthetic control outcome ($\boldsymbol{v}^*$-weighted).

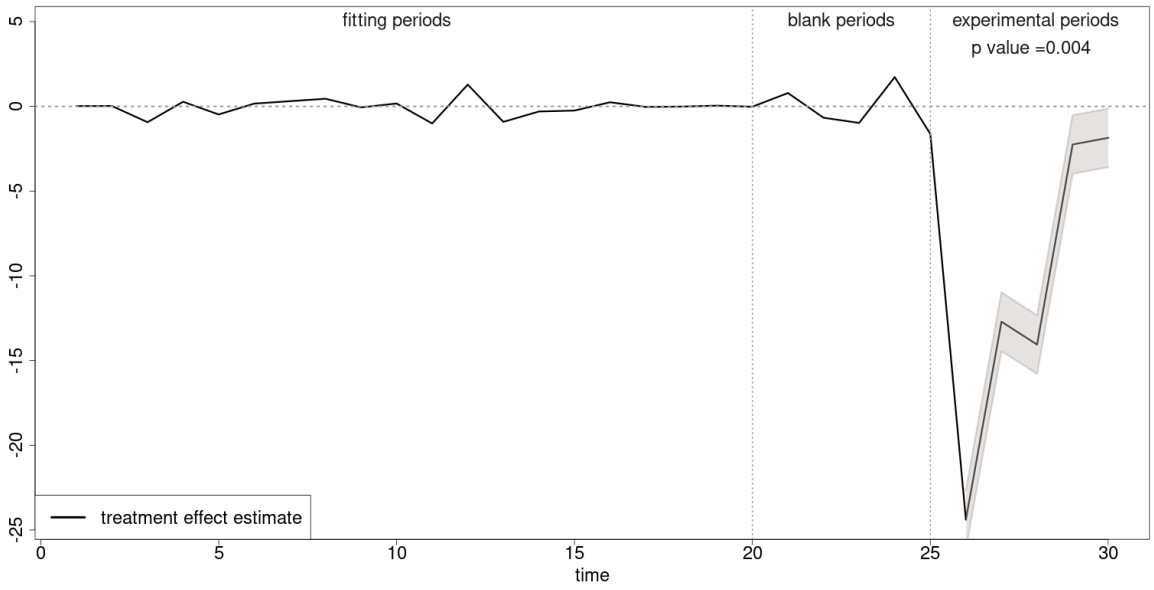


Figure 5: Treatment Effect Estimate, when $\sigma^2 = 1$

Note: This figure reports the difference between the synthetic treated and synthetic control outcomes of Figure 4. For the experimental periods, this is the treatment effect estimate.