Synthetic Controls for Experimental Design

Alberto Abadie Jinglong Zhao
MIT Boston University

August 2021

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 induce large estimation biases under many or all possible treatment assignments. We propose a variety of synthetic control designs (Abadie et al., 2010, Abadie and Gardeazabal, 2003) as experimental designs to select treated units in non-randomized experiments with large aggregate units, as well as the untreated units to be used as a control group. Average potential outcomes are estimated as weighted averages of treated units for potential outcomes with treatment, and control units for potential outcomes without treatment. We analyze the properties of such estimators and propose inferential techniques.

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). 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 and implements an experimental evaluation where the new plan is deployed at a small scale, say, in one of the local markets (cities) in the US. 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 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

Alberto Abadie, Department of Economics, MIT, abadie@mit.edu. Jinglong Zhao, Questrom School of Business, Boston University, jinglong@bu.edu

arm of the experiment, which will result in biased experimental estimates.

A 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 window, 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 is different from the untreated units in the values of the features that affect the outcomes of interest.

As in the ride-sharing example where there is 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. 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 the use of the synthetic control method (Abadie et al., 2010, Abadie and Gardeazabal, 2003) as an experimental design to select treated units in non-randomized experiments, as well as the untreated units to be used as a comparison group. We use the name *synthetic control designs* to refer to the resulting experimental designs.[1,2]

In our framework, the choice of the treated unit (or treated units, if multiple treated units are desired) aims to accomplish two goals. First, it is often useful to select the treated units such that their features are representative of the features of an aggregate of interest, like an entire country market. The treatment effect for the treated units selected in this way may more accurately reflect the effect of the treatment on the entire aggregate of interest. Second, the treated units should not be idiosyncratic in the sense that their features cannot be closely approximated by the units in the control arm. Otherwise, the reliability of the estimate of the effect on the treated

[1]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 et al., 2010, Abadie and Gardeazabal, 2003, Doudchenko and Imbens, 2016), for which the identity of the treated unit(s) is taken as given.

[2]See, e.g., Abadie (2021), Amjad et al. (2018), Arkhangelsky et al. (2019), Doudchenko and Imbens (2016) for background material on synthetic controls and related methods.

unit may be questionable. We show how to achieve these two objectives, whenever they are possible to achieve, using synthetic control techniques.

While we are aware of the extensive use of synthetic control techniques for experimental design in business analytics units, especially in the tech industry,[3] the academic literature on this subject is at a nascent stage. There are, however, three publicly available studies that are connected to this article. To our knowledge, Doudchenko et al. (n.d.) is the first (and only) publicly available study on the topic of experimental design with synthetic controls, and it is closely related to the present article. The focus of Doudchenko et al. (n.d.) is on statistical power, which they calculate by simulation of the estimated effects of placebo interventions on 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 present article, we aim to take a different perspective on the problem of unit selection in experiments with synthetic controls; one that takes into account the extent to which different sets of treated and control units approximate an aggregate causal effect of interest. Agarwal et al. (2021) propose synthetic interventions, a framework related to synthetic controls, and apply it to estimate treatment effect heterogeneity in experimental setting with multiple treatments. Bottmer et al. (2021) is also related to the present article in the sense that they study synthetic control estimation in an experimental setting. Their article, however, considers only the case when the treatment is randomized, and is not concerned with issues of experimental design.

2. Synthetic Control Designs

We consider a setting with T time periods and J units, which may represent J local markets as in the example in the previous section. Let T_0 be the number of pre-experimental periods, with $1 \leq T_0 < T$, and let T_1 be the length of the experimental window, with $T_1 = T - T_0$. We consider the following problem. At the end of period T_0, an analyst plans out an experiment

[3]See, in particular, Jones and Barrows (2019), which also provides the basis for the ride-sharing example above.

to conduct during periods $T_0 + 1, T_0 + 2, \dots, T$. Using information available at T_0, the analyst aims to select the set of units that will be administered treatment (intervention) during the experimental periods.

To define causal parameters, we formally adopt a potential outcomes framework. For any $j \in \{1, \dots, J\}$ and any $t \in \{T_0 + 1, \dots, T\}$, let Y_{jt}^I be the potential outcome for unit j at time t when the unit is exposed to treatment starting at $T_0 + 1$. Similarly, for any $j \in \{1, \dots, J\}$ and any $t \in \{1, \dots, T\}$, let Y_{jt}^N be the potential outcome for unit j at time t under no treatment. We assume that the outcome variable of interest is scaled so that it does not depend on the size of the unit. In the ride-sharing example, Y_{jt}^I and Y_{jt}^N could measure net income divided by market size, under the active and the control treatment, respectively. Unit-level treatment effects are defined as

$$\alpha_{jt} = Y_{jt}^I - Y_{jt}^N,$$

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

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

for $t = T_0 + 1, \dots, T$. In this expression, f_j represents a set of known positive weights that are relevant to the definition of the average. In the ride-sharing example of the previous section, f_j could represent the size of local market j as a share of the national market. Without loss of generality, and because it is the case in many applications, we can assume that the weights f_j sum to one,

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

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

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

$$\sum_{j=1}^{J} w_j = 1,$$

$$\sum_{j=1}^{J} v_j = 1,$$

$$w_j \geq 0,\ \ v_j \geq 0,\ \text{ and } w_j v_j = 0, \tag{1}$$

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

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

$$\sum_{j=1}^{J} w_j Y_{jt}^{I} = \sum_{j=1}^{J} f_j Y_{jt}^{I}, \tag{2}$$

for $t = T_0 + 1, \ldots, T$. If equation (2) 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 (2). 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 will aim to approximate equation (2) using predictors observed at T_0 of the values of Y_{jt}^{I} for $t > T_0$. Notice also that it is not possible to use the solution $w_1 = f_1, \ldots, w_J = f_J$, because it would leave no units in the donor pool, making the set of units with $v_j > 0$ empty and violating the second condition in (1).

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

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

or, alternatively,

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

If equations (3) or (4) 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 (3)), or for the units selected for treatment (equation (4)). As in the previous case with treated outcomes, it is not feasible to directly choose $v_1, \ldots, v_J$ so that equations (3) or (4) are satisfied. Below we propose a variety of methods to approximate either (3) or (4) based on predictors observed at T_0 of Y_{jt}^N for $t > T_0$.

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 $t = 1, \ldots, T$. Y_{jt} is the outcome observed for unit $j = 1, \ldots, J$ at time $t = 1, \ldots, T$. We will 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 (2) and (3) hold. Then, $\tau_t(\boldsymbol{w}, \boldsymbol{v})$ is equal to the average treatment effect, τ_t. If equation (4) holds instead, then $\tau_t(\boldsymbol{w}, \boldsymbol{v})$ is equal to the average effect of the treatment on the treated,

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

As discussed above, we will choose w_j and v_j to match the pre-intervention values of the predictors of potential outcomes. Let $\boldsymbol{X}_j$ be a column vector of pre-intervention features of unit j. We see the features in $\boldsymbol{X}_j$ as predictors of post-intervention values of Y_{jt}^N and Y_{jt}^I, in a sense that will be made precise in Section 3. We will 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 $\boldsymbol{X}_j$.

We will consider a variety of selectors for the weights, $w_1, \ldots, w_J, v_1, \ldots, v_J$. For any real vector $\boldsymbol{x}$, let $\|\boldsymbol{x}\|$ be the Euclidean norm of $\boldsymbol{x}$, and let $\|\boldsymbol{x}\|_0$ be 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 selector of $\boldsymbol{w} = (w_1, \ldots, w_J)$ and $\boldsymbol{v} = (v_1, \ldots, v_J)$ is

$$
\begin{aligned}
\min_{\substack{w_1,\ldots,w_J,\\ v_1,\ldots,v_J}} \quad & \left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j \right\|^2 + \left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} v_j \boldsymbol{X}_j \right\|^2 \\
\text{s.t.} \quad & \sum_{j=1}^{J} w_j = 1, \\
& \sum_{j=1}^{J} v_j = 1, \\
& w_j, v_j \geq 0, \quad j = 1, \ldots, J, \\
& w_j v_j = 0, \quad j = 1, \ldots, J, \\
& \underline{m} \leq \|\boldsymbol{w}\|_0 \leq \overline{m}.
\end{aligned}
\tag{5}
$$

The two terms of the objective function in (5) measure the discrepancies between the population average of the covariates in $\boldsymbol{X}_j$ ($\boldsymbol{f}$-weighted) and the averages of the covariates for units assigned to the treatment ($\boldsymbol{w}$-weighted) and units assigned to the control group ($\boldsymbol{v}$-weighted), respectively. The first four constraints stipulate that the weights in $\boldsymbol{w}$, as well as the weights in $\boldsymbol{v}$, are non-negative and sum to one. They also stipulate that any unit selected for treatment cannot be utilized as a control unit — so, if $w_j > 0$, then $v_j = 0$. 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. The values $\underline{m} = 1$ and $\overline{m} = J - 1$ correspond to the unconstrained case. The last constraint in (5) 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 $\boldsymbol{\beta}'\boldsymbol{d} \leq b$. Where the j-th coordinate of $\boldsymbol{\beta}$ is equal to the cost of assigning unit j to treatment, $\boldsymbol{d}$ is a $(J \times 1)$-vector with ones for coordinates with $w_j > 0$, and zeros otherwise, and b is the experimenter's budget.

Let $\boldsymbol{w}^* = (w_1^*, \ldots, w_J^*)$, $\boldsymbol{v}^* = (v_1^*, \ldots, v_J^*)$ be a solution of the optimization problem in (5). 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 $\widehat{\tau}_t = \tau_t(\boldsymbol{w}^*, \boldsymbol{v}^*)$. That is,

$$\widehat{\tau}_t = \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt}. \tag{6}$$

This estimator is based on approximations to equations (2) and (3) that rely on $\boldsymbol{X}_j$, the observed predictors of the potential outcomes, Y_{jt}^N and Y_{jt}^I.

In what follows, we take the weight selector in (5) as a starting point for synthetic control designs, and modify it in several respects. First, notice that for every solution to (5) with $\underline{m} \leq \|\boldsymbol{v}\|_0 \leq \overline{m}$, there is always another solution that swaps the roles of the treated and the untreated in the experiment. Moreover, in some applications there may not be two disjoint sets of units that closely reproduce the values in $\overline{\boldsymbol{X}}$, or that produce the values in $\overline{\boldsymbol{X}}$ without heavily relying on interpolation between distant units. To address these two considerations, we modify the synthetic control design in (5) in the following manner. As for (5), the analyst selects a synthetic treated unit to match the average values of the characteristics in the population. However, unlike in the design in (5), the analyst chooses multiple synthetic controls, one for each unit that contributes to the synthetic treated unit. For any $(J \times 1)$-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, $\mathcal{J}_{\boldsymbol{w}} = \{j : w_j > 0\}$. Our second version of the synthetic experiment design is:

$$\begin{aligned}
\min_{\substack{w_1,\ldots,w_J,\\ v_1,\ldots,v_J}} \quad & \left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j \right\|^2 + \xi \sum_{j=1}^{J} w_j \left\| \boldsymbol{X}_j - \sum_{i=1}^{J} v_{ij} \boldsymbol{X}_i \right\|^2 \\
\text{s.t.} \quad & \sum_{j=1}^{J} w_j = 1, \\
& w_j \geq 0, \quad j = 1, \ldots, J, \\
& \sum_{i=1}^{J} v_{ij} = 1, \quad \forall j \in \mathcal{J}_{\boldsymbol{w}}, \\
& v_{ij} = 0, \quad \forall i \in \mathcal{J}_{\boldsymbol{w}},\ j = 1, \ldots, J \\
& v_{ij} \geq 0, \quad \forall j \in \mathcal{J}_{\boldsymbol{w}},\ i = 1, \ldots, J, \\
& v_{ij} = 0, \quad \forall j \notin \mathcal{J}_{\boldsymbol{w}},\ i = 1, \ldots, J, \\
& \underline{m} \leq \|\boldsymbol{w}\|_0 \leq \overline{m}.
\end{aligned} \tag{7}$$

The parameter $\xi > 0$ arbitrates potential 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 the treated units. A small value of ξ favors experimental designs with treated units that closely match $\overline{\boldsymbol{X}}$. A large value of ξ, instead, favors designs where the values of the predictors for the treated units are closely matched by their respective synthetic controls. While it is possible to use data-driven selectors of ξ, the rule of thumb $\xi = 1$ provides a natural choice, which equally weights the two terms in the objective function in (7).

Let $\{w_j^*, v_{ij}^*\}_{i,j=1,\ldots,J}$ be one optimal solution of the optimization problem in (7). As before, assign units with $w_j^* > 0$ to treatment in the experiment, and keep units with $w_j^* = 0$ untreated. Let

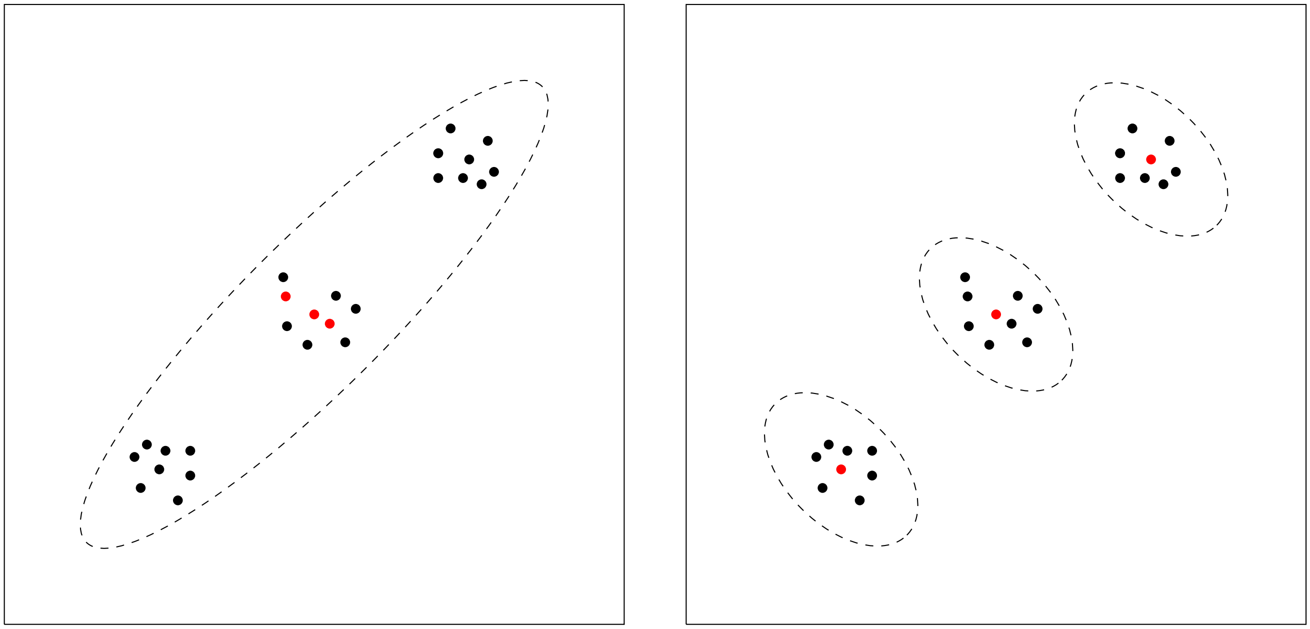
$$v_j^* = \sum_{i=1}^{J} w_i^* v_{ji}^*. \tag{8}$$

A synthetic control estimator of τ_t^T is

$$\begin{aligned} \widehat{\tau}_t^T &= \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \\ &= \sum_{j=1}^{J} w_j^* \left(Y_{jt} - \sum_{i=1}^{J} v_{ij}^* Y_{it} \right). \end{aligned} \tag{9}$$

This estimator is based on an approximation to equation (4) that relies on $\boldsymbol{X}_j$, the observed predictors of the potential outcomes Y_{jt}^N.

Our next adjustment to the synthetic control design is motivated by settings where the units may be naturally divided in 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, may influence the seasonality of the demand for ride-sharing services. In those cases, it is natural to treat each cluster as a distinct experimental design to ameliorate interpolation biases. Figure 2 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 of the units are represented by the coordinates of the points in the figure, which are the same in the two panels. Red dots represent units assigned to treatment. All other units are plotted as black dots. Panel (a)



(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.

visualizes the result of treating the entire sample a 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), we recognize the clustered nature of the data, and assign to treatment one unit per cluster. This provides a better approximation of the distribution of the predictor values for the entire sample, ameliorating concerns of interpolation biases.

Suppose we divide the set of J available units into K clusters. Let $\mathcal{I}_k$ be the set of indices for the units in cluster k. The cluster means are

$$\overline{\boldsymbol{X}}_k = \sum_{j\in\mathcal{I}_k} f_j \boldsymbol{X}_j \Big/ \sum_{j\in\mathcal{I}_k} f_j,$$

for $k = 1, \ldots, K$. For $j = 1, \ldots, J$, let $k(j)$ be the index of the cluster to which unit j belongs. A clustered version of the synthetic control design task is given by the following optimization problem:

$$
\begin{aligned}
\min_{\substack{K,\mathcal{I}_1,\ldots,\mathcal{I}_K,\\ w_1,\ldots,w_J,\\ v_1,\ldots,v_J}} \quad & \sum_{k=1}^{K}\left(\sum_{j\in\mathcal{I}_k} f_j\right)\left\{\left\|\overline{\boldsymbol{X}}_k-\sum_{j\in\mathcal{I}_k} w_j\boldsymbol{X}_j\right\|^2+\xi\sum_{j\in\mathcal{I}_k} w_j\left\|\boldsymbol{X}_j-\sum_{i,j\in\mathcal{I}_k} v_{ij}\boldsymbol{X}_i\right\|^2\right\}\\
\text{s.t.} \quad & \sum_{j\in\mathcal{I}_k} w_j=1, \quad k=1,\ldots,K,\\
& w_j\geq 0, \quad j=1,\ldots,J,\\
& \sum_{i=1}^{J} v_{ij}=1, \quad \forall j\in\mathcal{J}_{\boldsymbol{w}}\\
& v_{ij}\geq 0, \quad \forall j\in\mathcal{J}_{\boldsymbol{w}},\ i=1,\ldots,J,\\
& v_{ij}=0, \quad \forall j\notin\mathcal{J}_{\boldsymbol{w}},\ i=1,\ldots,J,\\
& v_{ij}=0, \quad \forall i\in\mathcal{J}_{\boldsymbol{w}},\ j=1,\ldots,J,\\
& v_{ij}=0, \quad k(i)\neq k(j),\\
& \underline{m}\leq\|\boldsymbol{w}\|_0\leq\overline{m}.
\end{aligned}
\tag{10}
$$

Although we state the clustered case in (10) as a one-step estimator, computational complexity can be substantially reduced by estimating the composition of the clusters in a first step (e.g., using K-means).

We conclude this section by introducing two additional modifications to the synthetic control design.

First, it is well known that synthetic control estimators may not be unique. Lack of uniqueness is typical in settings where the values of the predictors that a synthetic control is targeting (i.e., $\overline{\boldsymbol{X}}$ in equation (5), or $\boldsymbol{X}_j$ for a treated unit in equation (7)) fall inside the convex hull of the values of $\boldsymbol{X}_j$ for the units in the donor pool. To address this issue we adapt the penalized estimator of Abadie and L'Hour (2021) to the synthetic control designs proposed in this article. The penalized synthetic control estimator of Abadie and L'Hour (2021) is known to be unique as long as predictor values for the units in the donor pool are in general quadratic position (see Abadie and L'Hour, 2021, for details). Moreover, penalized synthetic controls favor solutions where the synthetic control is composed of units that have predictor values, $\boldsymbol{X}_j$, similar to the target values. Applying the penalized synthetic control of Abadie and L'Hour (2021) to the objective function of (5), we obtain

$$\begin{aligned}
\min_{\substack{w_1,\ldots,w_J,\\ v_1,\ldots,v_J}} \quad & \left\|\overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j\right\|^2 + \left\|\overline{\boldsymbol{X}} - \sum_{j=1}^{J} v_j \boldsymbol{X}_j\right\|^2 + \lambda_1 \sum_{j=1}^{J} w_j \|\overline{\boldsymbol{X}} - \boldsymbol{X}_j\|^2 + \lambda_2 \sum_{j=1}^{J} v_j \|\overline{\boldsymbol{X}} - \boldsymbol{X}_j\|^2 \\
\text{s.t.} \quad & \sum_{j=1}^{J} w_j = 1, \\
& \sum_{j=1}^{J} v_j = 1, \\
& w_j, v_j \geq 0, \quad j = 1, \ldots, J, \\
& w_j v_j = 0, \quad j = 1, \ldots, J, \\
& \underline{m} \leq \|\boldsymbol{w}\|_0 \leq \overline{m}.
\end{aligned} \tag{11}$$

Here, λ_1 and λ_2 are positive constants that penalize discrepancies between the target values of the predictor $\overline{\boldsymbol{X}}$ and the values of the predictors for the units that contribute to their synthetic counterparts.[4]

Finally, Abadie and L'Hour (2021), Arkhangelsky et al. (2019), and Ben-Michael et al. (2021) have proposed bias-correction techniques for synthetic control methods. Appendix A provides details on how to apply bias correction techniques in a synthetic control design.

3. Formal Results

We first introduce an extension of the linear factor model commonly employed in the synthetic control literature, which we will use to analyze the properties of estimators that are based on a synthetic control design.

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

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

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

[4] See Abadie and L'Hour (2021) for details on penalized synthetic control estimators. For simplicity, we have discussed design (5) only. Appendix A discusses how to apply the Abadie and L'Hour penalty to the other synthetic designs proposed in this article.

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; ϵ_{jt} and ξ_{jt} are unobserved random shocks.

Equation (12a) is the linear factor model for potential outcomes under no treatment that is commonly employed in the literature as a benchmark model to analyze the properties of synthetic control estimators (see, e.g., Abadie et al., 2010, Ferman, 2020). Equation (12b) extends the linear factor structure to potential outcomes under treatment. The reason for extending the linear factor model to treated outcomes is that, in contrast to the usual setting of synthetic control estimation with observational data, experimental synthetic control designs require the choice of a treatment group in addition to the choice of a comparison group. Agarwal et al. (2021) have also considered a linear factor model over multiple treatment arms to estimate individual responses under different interventions.

We will employ the covariates in $\boldsymbol{Z}_j$ as well as pre-experimental values of the outcome variables Y_{jt} to construct the vectors of predictors, $\boldsymbol{X}_j$. In particular, let $\mathcal{E} \subseteq \{1, \ldots, T_0\}$, $T_{\mathcal{E}} = |\mathcal{E}|$, and let $\boldsymbol{Y}_j^{\mathcal{E}}$ be the $(T_{\mathcal{E}} \times 1)$ vector of m pre-intervention outcomes for unit j and times indices in $\mathcal{E}$. We will 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-intervention 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) Let $\lambda_{t,f}$ and $\eta_{t,f}$ be the f-th coordinates of $\boldsymbol{\lambda}_t$ and $\boldsymbol{\eta}_t$, respectively. There exists $\bar{\lambda}$ and $\bar{\eta}$, such that $|\lambda_{t,f}| \leq \bar{\lambda}$, for $t = 1, \ldots, T$, $f = 1, \ldots, F$ and $|\eta_{t,f}| \leq \bar{\eta}$, for $t = T_0 + 1, \ldots, T$, $f = 1, \ldots, F$.

(ii) $T_{\mathcal{E}} \geq F$. *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,* $\underline{\zeta} = \zeta_{\mathcal{E}}/T_{\mathcal{E}} > 0$.

(iii) *For* $j = 1, \ldots, J$, $t = 1, \ldots, T$, ϵ_{jt} *are mean zero and i.i.d. sub-Gaussian random variables with variance proxy* $\bar{\sigma}^2$. *Similarly, for* $j = 1, \ldots, J$, $t = T_0 + 1, \ldots, T$, ξ_{jt} *are mean zero and i.i.d. sub-Gaussian random variables with variance proxy* $\bar{\sigma}^2$.

Assumptions 2 *(i)* and *(ii)* are regularity conditions similar to those in Abadie et al. (2010). The restrictions in Assumption 2 *(iii)* are similar to those invoked in Abadie et al. (2010), Doudchenko and Imbens (2016), Chernozhukov et al. (2021), and Arkhangelsky et al. (2019). Sub-Gaussianity is not strictly necessary, but it simplifies the form of our results. It can be relaxed by assuming bounded higher order moments (instead of bounding the entire moment generating function). Sub-Gaussianity is a relatively mild assumption. It holds for any Gaussian distribution with mean zero, as well any distribution with mean zero and bounded support. On the other hand, distributions with heavy tails, including the Cauchy distribution, are not sub-Gaussian.

The next assumption pertains to the quality of the synthetic control fit. For concreteness, we focus on the base design case, where $\boldsymbol{w}^* = (w_1^*, \ldots, w_J^*)$ and $\boldsymbol{v}^* = (v_1^*, \ldots, v_J^*)$ solve (5).

Assumption 3 *With probability one,*

$$\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j = \sum_{j=1}^{J} f_j \boldsymbol{Z}_j, \qquad \sum_{j=1}^{J} w_j^* Y_{jt} = \sum_{j=1}^{J} f_j Y_{jt}, \qquad \forall t \in \mathcal{E}, \tag{13a}$$

and

$$\sum_{j=1}^{J} v_j^* \boldsymbol{Z}_j = \sum_{j=1}^{J} f_j \boldsymbol{Z}_j, \qquad \sum_{j=1}^{J} v_j^* Y_{jt} = \sum_{j=1}^{J} f_j Y_{jt}, \qquad \forall t \in \mathcal{E}. \tag{13b}$$

Assumption 3 implies that the synthetic treated and control units defined by $\boldsymbol{w}^*$ and $\boldsymbol{v}^*$ provide a perfect fit for $\overline{\boldsymbol{X}}$. This assumption may only hold approximately in practice.

Theorem 1 *If Assumptions 1–3 hold, then for any positive integer* q,

$$|E[\widehat{\tau}_t - \tau_t]| \leq \frac{\bar{\lambda}(\bar{\eta} + \bar{\lambda})F}{\underline{\zeta}} J^{1/q} \sqrt{2}\bar{\sigma}(q\Gamma(q/2))^{1/q} \frac{1}{\sqrt{T_{\mathcal{E}}}}, \tag{14}$$

where the expectation is taken over the distributions of ϵ_{jt}, for $j = 1, \ldots, J$, $t = 1, \ldots, T$ and the distributions of ξ_{jt}, for $j = 1, \ldots, J$, $t = T_0 + 1, \ldots, T$.

The bias bound in Theorem 1 depends on the ratio between the scale of ϵ_{jt}, represented in (14) by $\sqrt{2}\bar{\sigma}(q\Gamma(q/2))^{1/q}$, and the number of fitting periods $T_{\mathcal{E}}$.[5] Intuitively, the bias of the synthetic control estimator is small when a good fit in pre-intervention outcomes (Assumption 3) is obtained by implicitly fitting the values of the latent variables, μ_j. Overfitting happens when pre-intervention 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 enough 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 created by increased variability in ϵ_{jt} over larger donor pools. Finally, the number of unobserved factors, F, enters the bound linearly, which highlights the importance of including the observed predictors, $\boldsymbol{Z}_j$ —other than pre-intervention outcomes— in the vector of fitting variables, $\boldsymbol{X}_j$. Under the factor model in equations (12a) and (12b), observed predictors not included in Z_j are shifted to μ_j, increasing F and the magnitude of the bias bound.

We next turn our attention to inference. We will utilize a set of "blank periods," $\mathcal{B} \subseteq \{1, \ldots, T_0\} \setminus \mathcal{E}$, which comprise pre-intervention periods whose outcomes Y_{jt} have not been used to calculate $\boldsymbol{w}^*$ or $\boldsymbol{v}^*$. Because pre-intervention periods that are not in $\mathcal{E}$ or $\mathcal{B}$ could always be discarded from the data, we can consider $\mathcal{B} = \{1, \ldots, T_0\} \setminus \mathcal{E}$ only, without loss of generality. We will, therefore, 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 null:

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

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

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

Under the null hypothesis in (15), the distribution of Y^I_{jt} is the same as the distribution of

[5] Recall that if X is sub-Gaussian with variance proxy $\bar{\sigma}^2$, then $(E[|X|^q])^{1/q} \leq \sqrt{2}\bar{\sigma}(q\Gamma(q/2))^{1/q}$. See, e.g., Lemma 1.4 in Rigollet and Hütter (2019).

Y_{jt}^N, for $t = T_0 + 1, \ldots, T$, and $j = 1, \ldots, J$.

For $t \in \mathcal{B}$, let

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

Such $\widehat{u}_t$ for $t \in \mathcal{B}$ are "placebo" treatment effects estimated for the blank periods. Under the null hypothesis in (15), these placebo treatment effects should come from the same distribution as the estimates of the treatment effects $\widehat{\tau}_t$ for all $t = T_0 + 1, \ldots, T$, as defined in (6). To make this explicit, let $t_1, \ldots, t_{T_{\mathcal{B}}}$ be the set of indices of the time periods in $\mathcal{B}$, and define

$$\begin{aligned}\widehat{\boldsymbol{r}} &= (\widehat{r}_1, \ldots, \widehat{r}_{T-T_{\mathcal{E}}}) \\ &= (\widehat{\tau}_{T_0+1}, \ldots, \widehat{\tau}_T, \widehat{u}_{t_1}, \ldots, \widehat{u}_{t_{T_{\mathcal{B}}}}).\end{aligned}$$

Recall that $T_1 = T - T_0$. The first T_1 coordinates of $\widehat{\boldsymbol{r}}$ are the post-intervention estimates of the treatment effects. The last $T_{\mathcal{B}}$ coordinates of $\widehat{\boldsymbol{r}}$ are placebo treatment effects estimated for the blank periods.

Let Π be the set of all T_1-combinations of $\{1, \ldots, T - T_{\mathcal{E}}\}$. That is, for each $\pi \in \Pi$, π is a subset of indices from $\{1, \ldots, T - T_{\mathcal{E}}\}$, such that $|\pi| = T_1$. Then, Π is a set of such subsets with cardinality $|\Pi| = (T - T_{\mathcal{E}})!/(T_1!(T_0 - T_{\mathcal{E}})!)$. For each $\pi \in \Pi$, let $\pi(i)$ be the i^{th} smallest value in π. We define the $(T_1 \times 1)$-vector

$$\widehat{\boldsymbol{e}}_\pi = (\widehat{r}_{\pi(1)}, \widehat{r}_{\pi(2)}, ..., \widehat{r}_{\pi(T_1)}).$$

In addition, let $\widehat{\boldsymbol{e}} = (\widehat{r}_1, \ldots, \widehat{r}_{T_1}) = (\widehat{\tau}_{T_0+1}, \ldots, \widehat{\tau}_T)$. This is a vector of treatment effect estimates from the post-intervention periods For any $(T_1 \times 1)$-vector $\boldsymbol{e} = (e_1, \ldots, e_{T_1})$, define the test statistic,

$$S(\boldsymbol{e}_\pi) = \frac{1}{T_1} \sum_{t=1}^{T_1} |e_t| \,. \tag{16}$$

Other choices of test statistics are possible, such as those based on an L_p-norm of $\boldsymbol{e}$ and one-sided versions of the resulting test statistics (e.g., with e_t or $-e_t$ replacing $|e_t|$ in equation (16)).

Define the p-value:

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

The next theorem shows that the p-value in (17) is approximately valid.

Theorem 2 *Suppose that Assumptions 1–3 hold, and that for* $j = 1, \ldots, J$*,* $t = 1, \ldots, T$*,* ϵ_{jt} *are continuously distributed with (a version of) the probability density function bounded by a constant* $\kappa < \infty$*. Then, the* p*-values of equation* (17) *are approximately valid. In particular, for any* $\alpha \in (0, 1]$*, we have*

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

with probability equal or greater than $1 - c_1(T_{\mathcal{E}})$*, where*

$$c_1(T_{\mathcal{E}}) = 2J(T - T_{\mathcal{E}}) \exp\left(-\frac{\underline{\zeta}^2}{8\bar{\sigma}^2\bar{\lambda}^4 F^2}(T_{\mathcal{E}})^{1/2}\right) + 2JT_1\kappa\,|\Pi|^2\,(T_{\mathcal{E}})^{-1/4}.$$

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

The inferential techniques proposed in this article are related to but distinct from the permutation methods in Abadie et al. (2010), Chernozhukov et al. (2019, 2021), Firpo and Possebom (2018), Lei and Candès (2020), and others. Inferential methods that reassign treatment across units (e.g., Abadie et al., 2010) are unfit for the designs of Section 2, which explicitly select treated and control units to satisfy an optimality criterion. Instead, like in Chernozhukov et al. (2021), our methods are based on rearrangements of estimated treatment effects across time periods, where some of the time periods correspond to the post-intervention window and some correspond to pre-intervention blank periods that are left-out for the calculation of the synthetic treatment and synthetic control weights. Relative to Chernozhukov et al. (2021), the generative models of equations (12a) and (12b), which allow for unobserved factors, and the finite sample nature of the result in Theorem 2 require both a novel testing procedure that takes advantage of the availability of blank periods and proof techniques that are, to our knowledge, new to the literature.

4. Simulation Study

4.1. Base Results

We report in this section the results of a simulation study that illustrates the behavior of the estimators proposed in this article. We consider a setting with $J = 15$ units, $r = 7$ observable covariates and $F = 11$ unobservable covariates. We consider $T = 30$ periods in total, with $T_0 = 25$ pre-intervention periods, and $T_1 = 5$ experimental or post-intervention periods. We compute weights during the first $T_{\mathcal{E}} = 20$ periods, and leave periods $t = 21, \ldots, 25$ as blank periods. Periods $t = 26, \ldots, 30$ are the experimental periods.

We use the factor model in Assumption 1 to generate potential outcomes. For $t = 1, \ldots, T$, we generate the series δ_t and υ_t as small-to-large re-arrangements of T 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 $(0, \sigma^2)$ random variables, with $\sigma^2 = 1$. In Appendix D we present additional simulation results of alternative values of the noise parameter σ^2.

Using the data generating process described above, we draw a single sample and conduct the synthetic control design in (5), with parameters $\underline{m} = 1$ and $\overline{m} = 14$ (no constraint on the number of treated units). We report the results in Figures 2 and 3. In Figure 2, the solid line represents the synthetic treated unit ($\sum_{j=1}^{J} w_j^* Y_{jt}$, for $t = 1, \ldots, T$). The dashed line represents the synthetic control unit ($\sum_{j=1}^{J} v_j^* Y_{jt}$, for $t = 1, \ldots, T$). The two lines closely track each other in the pre-experimental periods. They diverge in 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 3 reports the difference between the synthetic treated and the synthetic control outcomes. The inferential procedure of Section 3 produces p-value equal to 0.0198, for the null hypothesis of no treatment effect in (15).

4.2. A Comparison of Different Synthetic Control Designs

In this section we compare the performance of different synthetic control designs under the same data generating process as in Section 4.1. We consider four formulations of the synthetic control

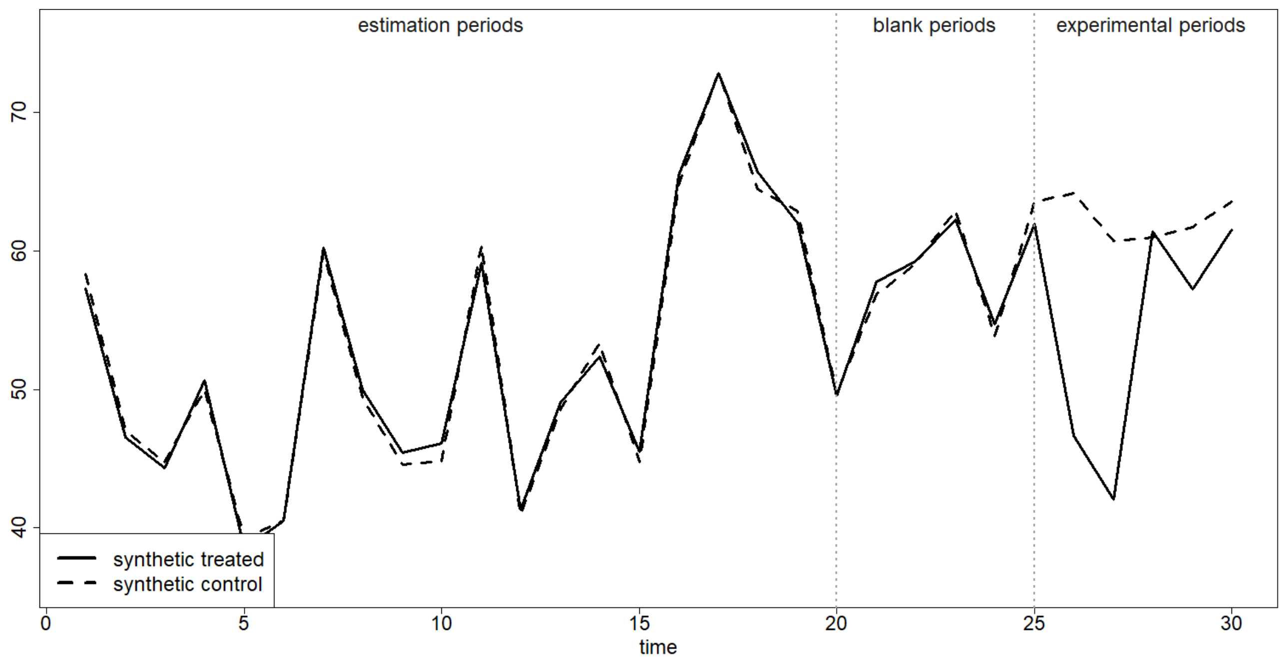


Figure 2: Synthetic Treatment Unit and Synthetic Control Unit, when $\sigma^2 = 1$.

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

design:

1. *Unconstrained synthetic control*: This is the design in (5) without a cardinality constraint, so $\underline{m} = 1$ and $\overline{m} = J - 1 = 14$.
2. *Constrained synthetic control*: Same as the design in (5), but with $\underline{m} = 1$ and $\overline{m} = 1, \ldots, 7$.
3. *Penalized synthetic control*: This is the design in (11), with $\lambda = \lambda_1 = \lambda_2$. We will vary λ from 10^{-4} to 10^3.
4. *Unit-level synthetic control*: This is the design in (7), which fits a different synthetic control to each of the units assigned to treatment. We will vary ξ from 10^{-4} to 10^3.

4.2.1. Estimated Average Treatment Effects

The first panel of Table 1 reports true average treatment effects, τ_t. The first five columns in the second panel report estimated average treatment effects, $\widehat{\tau}_t$, for periods $(T_0 + 1) = 26$ through $T = 30$, under different synthetic control designs. The last column in the second panel reports

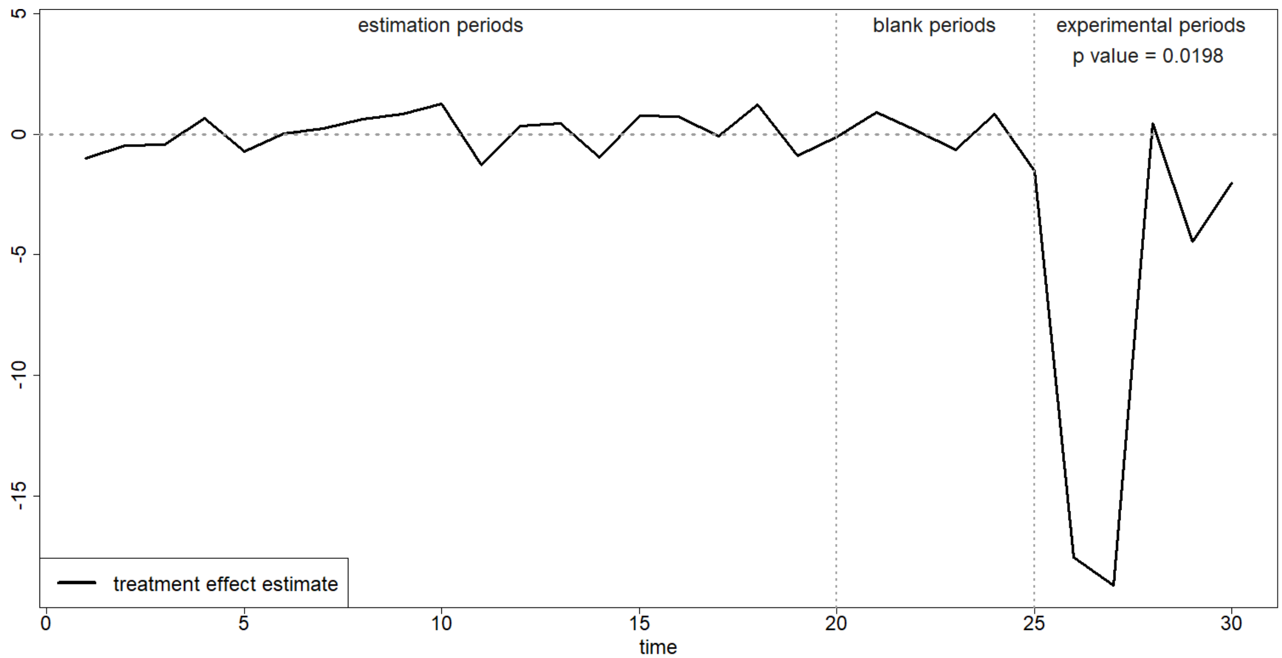


Figure 3: Treatment Effect Estimate, when $\sigma^2 = 1$.

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.

mean absolute error (MAE), defined as

$$MAE = \frac{1}{T_1} \sum_{t=T_0+1}^{T} |\widehat{\tau}_t - \tau_t|.$$

Because Table 1 reports outcomes for a single simulation only, the results in this table may not be reflective of general patterns across many simulations, even for the particular data generating process employed to produce the data. However, we use this table to illustrate patterns in the estimates that are induced directly by the features of their respective estimators. In particular, the results of Table 1 suggest that exceedingly large values of λ or ξ damper the performance of synthetic control designs. Large values of λ encourage synthetic control and synthetic treated units that put their entire weights in only one of the sample units (see Section 4.2.2 below), worsening performance relative to designs with larger number of units in the treated and control groups. Large values of ξ have the same effect for the synthetic treated unit. On the other hand, however, exceedingly small values of ξ produce designs that ignore the quality of the fit of the unit-level synthetic controls for each of the treated units, resulting in a negative performance of the estimator.

Table 1: Average Treatment Effects Estimates

		τ_t					
		$t = 26$	$t = 27$	$t = 28$	$t = 29$	$t = 30$	
		-15.55	-17.76	2.52	-4.92	-3.27	
		$\widehat{\tau}_t$					*MAE*
		$t = 26$	$t = 27$	$t = 28$	$t = 29$	$t = 30$	
Unconstrained		-17.54	-18.70	0.46	-4.47	-2.02	1.34
Constrained	$\overline{m} = 1$	-19.14	-18.47	2.14	0.76	0.88	2.90
	$\overline{m} = 2$	-19.16	-18.31	3.73	-0.22	1.58	2.98
	$\overline{m} = 3$	-13.88	-17.72	3.13	-6.56	-4.23	0.98
	$\overline{m} = 4$	-14.31	-18.51	2.46	-5.94	-5.05	0.97
	$\overline{m} = 5$	-13.43	-18.16	2.54	-5.77	-4.52	0.93
	$\overline{m} = 6$	-18.11	-19.50	0.38	-4.54	-2.01	1.62
	$\overline{m} = 7$	-17.54	-18.70	0.46	-4.47	-2.02	1.34
Penalized	$\lambda = 10^{-4}$	-17.53	-18.69	0.46	-4.46	-2.01	1.34
	$\lambda = 10^{-3}$	-17.48	-18.64	0.50	-4.43	-1.98	1.32
	$\lambda = 10^{-2}$	-16.83	-17.94	1.11	-3.95	-1.44	1.13
	$\lambda = 10^{-1}$	-17.40	-19.28	1.74	-2.53	-0.37	1.89
	$\lambda = 1$	-20.30	-18.36	2.78	0.67	-0.47	2.80
	$\lambda = 10$	-19.50	-20.15	0.31	1.80	-2.07	3.29
	$\lambda = 10^{2}$	-18.06	-21.37	-1.00	5.62	-2.86	4.12
	$\lambda = 10^{3}$	-18.06	-21.37	-1.00	5.62	-2.86	4.12
Unit-level	$\xi = 10^{-4}$	-10.36	-12.96	7.66	-4.18	3.64	4.55
	$\xi = 10^{-3}$	-13.34	-16.47	4.02	-2.95	0.25	2.10
	$\xi = 10^{-2}$	-14.23	-18.20	3.25	-3.50	0.93	1.62
	$\xi = 10^{-1}$	-15.62	-17.15	4.43	-4.28	1.08	1.51
	$\xi = 1$	-16.65	-18.15	3.91	-2.09	-0.09	1.78
	$\xi = 10$	-17.65	-19.71	2.04	0.81	3.01	3.31
	$\xi = 10^{2}$	-17.65	-19.71	2.04	0.81	3.01	3.31
	$\xi = 10^{3}$	-17.65	-19.71	2.04	0.81	3.01	3.31

Note: Unless otherwise noted, all designs use $\underline{m} = 1$ and $\overline{m} = 14$.

4.2.2. Synthetic Treated and Synthetic Control Weights

Tables 2 and 3 report the synthetic treated and synthetic control weights ($\boldsymbol{w}^*$ and $\boldsymbol{v}^*$, respectively) for the same designs as in Table 1. For the *Unit-level* design, synthetic control weights are aggregated as in (8). *Unconstrained*, *Constrained* and *Penalized* synthetic treated and synthetic control weights can always be switched without changing the value of the objective functions for their respective designs. We choose between switches to minimize the number of treated units.

The *Constrained* design imposes sparsity in the synthetic treatment weights through a hard cardinality constraint: the density of synthetic treatment weights for this design cannot be larger than $\overline{m}$. For $\overline{m} = 7$, the *Constrained* and the *Unconstrained* weights coincide. For large values of λ, the matching discrepancies between $\overline{\boldsymbol{X}}$ and $\boldsymbol{X}_j$ dominate the objective function of the *Penalized* design. As a result, for large values of λ, the *Penalized* design behaves like a one-to-one matching design, assigning all the weight to one treated and one control unit. For small values of λ, the *Penalized* design weights are close to the *Unconstrained* design weights. For the *Unit-level* design, the synthetic treated weights become sparse for large ξ. Large values of ξ encourage assignment to treatment of units that can be closely fitted by a synthetic control, even if the resulting treated units are not able to closely fit the population average $\overline{\boldsymbol{X}}$. As ξ becomes large, the *Unit-level* design assigns to treatment units with values of $\boldsymbol{X}_j$ that can be closely fitted by a synthetic control.

5. 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 large classes of problems, but becomes inadequate when treating more than a few units is unfeasible, which 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

Table 2: Synthetic Treated Weights

		w_1^*	w_2^*	w_3^*	w_4^*	w_5^*	w_6^*	w_7^*	w_8^*	w_9^*	w_{10}^*	w_{11}^*	w_{12}^*	w_{13}^*	w_{14}^*	w_{15}^*
Unconstrained		0	0.16	0	0.09	0.08	0	0	0.15	0	0	0	0.11	0.19	0.22	0
Constrained	$\overline{m} = 1$	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	$\overline{m} = 2$	0	0	0	0	0	0	0	0	0	0	0	0	0.30	0.70	0
	$\overline{m} = 3$	0	0	0	0	0	0	0	0	0.34	0	0.28	0.38	0	0	0
	$\overline{m} = 4$	0	0	0	0	0.10	0	0	0	0.24	0	0.35	0.31	0	0	0
	$\overline{m} = 5$	0	0	0	0	0.11	0	0	0	0.28	0	0.27	0.22	0	0	0.11
	$\overline{m} = 6$	0	0.13	0	0.12	0.15	0	0	0.13	0	0	0	0	0.23	0.24	0
	$\overline{m} = 7$	0	0.16	0	0.09	0.08	0	0	0.15	0	0	0	0.11	0.19	0.22	0
Penalized	$\lambda = 10^{-4}$	0	0.16	0	0.09	0.08	0	0	0.15	0	0	0	0.11	0.19	0.22	0
	$\lambda = 10^{-3}$	0	0.16	0	0.08	0.08	0	0	0.14	0	0	0	0.12	0.19	0.23	0
	$\lambda = 10^{-2}$	0	0.20	0	0.03	0	0	0	0.12	0	0	0	0.19	0.18	0.28	0
	$\lambda = 10^{-1}$	0	0.18	0	0	0	0	0	0	0	0.08	0	0	0.27	0.47	0
	$\lambda = 1$	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	$\lambda = 10$	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	$\lambda = 10^2$	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	$\lambda = 10^3$	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
Unit-level	$\xi = 10^{-4}$	0.09	0.09	0.03	0.04	0.06	0.04	0.06	0	0.12	0	0.10	0.08	0.11	0.09	0.11
	$\xi = 10^{-3}$	0.10	0.11	0.02	0	0.03	0.03	0.04	0	0.14	0	0.09	0.13	0.10	0.08	0.14
	$\xi = 10^{-2}$	0.08	0.11	0	0	0.07	0.03	0	0	0.14	0	0.11	0.14	0.11	0.08	0.13
	$\xi = 10^{-1}$	0.08	0.16	0	0	0	0	0	0.14	0	0	0	0.14	0.14	0.25	0.11
	$\xi = 1$	0	0	0	0	0	0	0	0.16	0	0	0.18	0.14	0	0.51	0
	$\xi = 10$	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	$\xi = 10^2$	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	$\xi = 10^3$	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

Note: Unless otherwise noted, all designs use $\underline{m} = 1$ and $\overline{m} = 14$.

Table 3: Synthetic Control Weights

		v_1^*	v_2^*	v_3^*	v_4^*	v_5^*	v_6^*	v_7^*	v_8^*	v_9^*	v_{10}^*	v_{11}^*	v_{12}^*	v_{13}^*	v_{14}^*	v_{15}^*
Unconstrained		0.17	0	0.09	0	0	0	0.08	0	0.19	0.09	0.15	0	0	0	0.23
Constrained	$\overline{m} = 1$	0.10	0.09	0	0.03	0.07	0.06	0.06	0	0.14	0	0.11	0.11	0.12	0	0.11
	$\overline{m} = 2$	0.13	0.07	0.06	0.02	0	0	0.08	0	0.19	0.04	0.14	0.13	0	0	0.14
	$\overline{m} = 3$	0.10	0.11	0.08	0.07	0	0	0.08	0.08	0	0.03	0	0	0.13	0.21	0.11
	$\overline{m} = 4$	0.10	0.11	0.08	0.07	0	0	0.08	0.08	0	0.03	0	0	0.13	0.21	0.11
	$\overline{m} = 5$	0.09	0.12	0.06	0.11	0	0	0.12	0.08	0	0.06	0	0	0.15	0.20	0
	$\overline{m} = 6$	0.13	0	0.06	0	0	0	0.08	0	0.22	0.05	0.17	0.10	0	0	0.18
	$\overline{m} = 7$	0.17	0	0.09	0	0	0	0.08	0	0.19	0.09	0.15	0	0	0	0.23
Penalized	$\lambda = 10^{-4}$	0.17	0	0.09	0	0	0	0.08	0	0.19	0.09	0.15	0	0	0	0.23
	$\lambda = 10^{-3}$	0.17	0	0.09	0	0	0	0.08	0	0.19	0.09	0.16	0	0	0	0.23
	$\lambda = 10^{-2}$	0.19	0	0.06	0	0	0	0.10	0	0.18	0.10	0.18	0	0	0	0.19
	$\lambda = 10^{-1}$	0.13	0	0	0	0	0	0.14	0	0.19	0	0.28	0.26	0	0	0
	$\lambda = 1$	0.21	0.33	0	0	0	0	0	0	0	0	0	0.29	0.17	0	0
	$\lambda = 10$	0.52	0.08	0	0	0	0	0	0	0	0	0	0.22	0.18	0	0
	$\lambda = 10^2$	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	$\lambda = 10^3$	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unit-level	$\xi = 10^{-4}$	0	0	0	0	0	0	0	0.40	0	0.60	0	0	0	0	0
	$\xi = 10^{-3}$	0	0	0	0.23	0	0	0	0.55	0	0.22	0	0	0	0	0
	$\xi = 10^{-2}$	0	0	0.14	0.26	0	0	0.13	0.35	0	0.12	0	0	0	0	0
	$\xi = 10^{-1}$	0	0	0.10	0.17	0.04	0.10	0.12	0	0.17	0.12	0.17	0	0	0	0
	$\xi = 1$	0.12	0.09	0.04	0.10	0.01	0.10	0.11	0	0.15	0.13	0	0	0.08	0	0.08
	$\xi = 10$	0.22	0	0	0.08	0	0.05	0.18	0	0.21	0.14	0	0.02	0	0	0.09
	$\xi = 10^2$	0.22	0	0	0.08	0	0.05	0.18	0	0.21	0.14	0	0.02	0	0	0.09
	$\xi = 10^3$	0.22	0	0	0.08	0	0.05	0.18	0	0.21	0.14	0	0.02	0	0	0.09

Note: Unless otherwise noted, all designs use $\underline{m} = 1$ and $\overline{m} = 14$.

population, or when the units in the control arm fail to approximate the outcomes that treated units would be experienced without treatment.

In this article we have applied synthetic control techniques, widely used in observational studies, to the design of experiments when 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 aimed 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. In addition, we report simulation results that demonstrate the applicability and computational feasibility of the methods proposed in this article.

Corporate research units and academic investigators are often confronted with settings where interventions at the level of micro-units (i.e., customers, workers, or families) are unfeasible, impractical or ineffective (see, e.g., Duflo et al., 2007, Jones and Barrows, 2019). There is, in consequence, a wide range of potential applications of experimental design methods for large aggregate entities, like the ones proposed in this article.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature 59*(2).

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

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the Basque County. *American Economic Review 93*(1), 113–132.

Abadie, A. and J. L'Hour (2021). A penalized synthetic control estimator for disaggregated data. `https://economics.mit.edu/files/18642`.

Agarwal, A., D. Shah, and D. Shen (2021). Synthetic interventions. *arXiv e-print*, arXiv:2006.07691.

Amjad, M., D. Shah, and D. Shen (2018). Robust synthetic control. *Journal of Machine Learning Research 19*(22), 1–51.

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

Ben-Michael, E., A. Feller, and J. Rothstein (2021). The augmented synthetic control method. *Journal of the American Statistical Association*. Forthcoming.

Bottmer, L., G. Imbens, J. Spiess, and M. Warnick (2021). A design-based perspective on synthetic control methods. *arXiv e-prints*, arXiv:2101.09398.

Chernozhukov, V., K. Wüthrich, and Y. Zhu (2019). Distributional conformal prediction. *arXiv preprint arXiv:1909.07889*.

Chernozhukov, V., K. Wüthrich, and Y. Zhu (2021). An exact and robust conformal inference method for counterfactual and synthetic controls. *Journal of the American Statistical Association*, 1–16.

Doudchenko, N., S. Gilinson, David; Taylor, and N. Wernerfelt (n.d.). Designing experiments with synthetic controls. `https://mackinstitute.wharton.edu/wp-content/uploads/2020/03/Wernerfelt-Nils-Doudchenko-Nick-Gilinson-David-and-Taylor-Sean_Designing-Experiments-with-Synthetic-Controls.pdf`.

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

Duflo, E., R. Glennerster, and M. Kremer (2007). Using randomization in development economics research: A toolkit. Volume 4 of *Handbook of Development Economics*, Chapter 61, pp. 3895–3962. Elsevier.

Ferman, B. (2020). On the properties of the synthetic control estimator with many periods and many controls. *arXiv e-prints*, arXiv:1906.06665.

Firpo, S. and V. Possebom (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *Journal of Causal Inference 6*(2), 20160026.

Jones, N. and S. Barrows (2019). Synthetic control and alternatives to A/B testing at Uber. Presented at PyData Amsterdam 2019. `https://youtu.be/j5DoJV5S2Ao`.

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

Rigollet, P. and J.-C. Hütter (2019). High dimensional statistics. `http://www-math.mit.edu/~rigollet/PDFs/RigNotes17.pdf`.

A. Designs based on penalized and bias-corrected synthetic control methods

Consider the design problem in (7),

$$\underbrace{\left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j \right\|^2}_{\text{(a)}} + \xi \sum_{j=1}^{J} w_j \underbrace{\left\| \boldsymbol{X}_j - \sum_{i=1}^{J} v_{ij} \boldsymbol{X}_i \right\|^2}_{\text{(b)}}. \tag{A.1}$$

To apply the penalized synthetic control method of Abadie and L'Hour (2021) to this design, we replace the term (a) in (A.1) with

$$\left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j \right\|^2 + \lambda_1 \sum_{j=1}^{J} w_j \|\overline{\boldsymbol{X}} - \boldsymbol{X}_j\|^2, \tag{A.2}$$

and the terms (b) with

$$\left\| \boldsymbol{X}_j - \sum_{i=1}^{J} v_{ij} \boldsymbol{X}_i \right\|^2 + \lambda_2 \sum_{i=1}^{J} v_{ij} \|\boldsymbol{X}_j - \boldsymbol{X}_i\|^2. \tag{A.3}$$

Here, λ_1 and λ_2 are positive constants that penalize discrepancies between the target values of the predictors ($\overline{\boldsymbol{X}}$ in (A.2) and $\boldsymbol{X}_j$ in (A.3)) and the values of the predictors for the units that contribute to their synthetic counterparts.

All designs of Section 2 depend on terms akin to (a) and (b) in (A.1). These terms can be adapted as in (A.2) and (A.3) to implement the penalized synthetic control design of Abadie and L'Hour (2021).

For all the designs in Section 2, the bias-corrected estimator of Abadie and L'Hour (2021) is

$$\widehat{\tau}_t^{BC} = \sum_{j=1}^{J} w_j^* (Y_{jt} - \widehat{\mu}_0(\boldsymbol{X}_j)) - \sum_{j=1}^{J} v_j^* (Y_{jt} - \widehat{\mu}_0(\boldsymbol{X}_j)),$$

where $t \geq T_0+1$ and the terms $\widehat{\mu}_{0t}(\boldsymbol{X}_j)$ are the fitted values of a regression of untreated outcomes, Y_{jt}^N, on units characteristics, $\boldsymbol{X}_j$. To avoid over-fitting biases, $\widehat{\mu}_{0t}(\boldsymbol{X}_j)$ can be cross-fitted for the untreated.

B. Proofs

Proof of Theorem 1. For any period $t = T_0 + 1, \ldots, T$ we decompose $\widehat{\tau}_t$ as follows,

$$\widehat{\tau}_t - \tau_t = \left(\sum_{j=1}^{J} w_j^* Y_{jt}^I - \sum_{j=1}^{J} v_j^* Y_{jt}^N \right) - \left(\sum_{j=1}^{J} f_j Y_{jt}^I - \sum_{j=1}^{J} f_j Y_{jt}^N \right)$$

$$= \left(\sum_{j=1}^{J} w_j^* Y_{jt}^I - \sum_{j=1}^{J} f_j Y_{jt}^I \right) - \left(\sum_{j=1}^{J} v_j^* Y_{jt}^N - \sum_{j=1}^{J} f_j Y_{jt}^N \right). \tag{B.1}$$

The first term in (B.1) measures the difference between the synthetic treatment outcome and the aggregated treatment outcomes. The second term measures the difference between the synthetic control outcome and the aggregate control outcomes. We bound these two terms separately. From (12b), we obtain

$$\begin{aligned}
&\sum_{j=1}^{J} w_j^* Y_{jt}^I - \sum_{j=1}^{J} f_j Y_{jt}^I \\
&\qquad = \boldsymbol{\gamma}_t' \left(\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} f_j \boldsymbol{Z}_j \right) + \boldsymbol{\eta}_t' \left(\sum_{j=1}^{J} w_j^* \boldsymbol{\mu}_j - \sum_{j=1}^{J} f_j \boldsymbol{\mu}_j \right) + \left(\sum_{j=1}^{J} w_j^* \xi_{jt} - \sum_{j=1}^{J} f_j \xi_{jt} \right)
\end{aligned} \tag{B.2}$$

Similarly, using expression (12a), we obtain

$$\begin{aligned}
&\sum_{j=1}^{J} w_j^* \boldsymbol{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} f_j \boldsymbol{Y}_j^{\mathcal{E}} \\
&\qquad = \boldsymbol{\theta}_{\mathcal{E}} \left(\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} f_j \boldsymbol{Z}_j \right) + \boldsymbol{\lambda}_{\mathcal{E}} \left(\sum_{j=1}^{J} w_j^* \boldsymbol{\mu}_j - \sum_{j=1}^{J} f_j \boldsymbol{\mu}_j \right) + \left(\sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} - \sum_{j=1}^{J} f_j \boldsymbol{\epsilon}_j^{\mathcal{E}} \right),
\end{aligned}$$

where $\boldsymbol{\theta}_{\mathcal{E}}$ is the $(T_{\mathcal{E}} \times r)$ matrix with rows equal to the $\boldsymbol{\theta}_t$'s indexed by $\mathcal{E}$, and $\boldsymbol{\epsilon}_j^{\mathcal{E}}$ is defined analogously. Pre-multiplying by $\boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'$ yields

$$\begin{aligned}
\boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \left(\sum_{j=1}^{J} w_j^* \boldsymbol{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} f_j \boldsymbol{Y}_j^{\mathcal{E}} \right) &= \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\theta}_{\mathcal{E}} \left(\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} f_j \boldsymbol{Z}_j \right) \\
&\quad + \boldsymbol{\eta}_t' \left(\sum_{j=1}^{J} w_j^* \boldsymbol{\mu}_j - \sum_{j=1}^{J} f_j \boldsymbol{\mu}_j \right) \\
&\quad + \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \left(\sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} - \sum_{j=1}^{J} f_j \boldsymbol{\epsilon}_j^{\mathcal{E}} \right).
\end{aligned} \tag{B.3}$$

Subtract (B.3) from (B.2) and apply Assumption 3 to obtain

$$\sum_{j=1}^{J} w_j^* Y_{jt}^I - \sum_{j=1}^{J} f_j Y_{jt}^I = -\boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}}$$

$$
\begin{aligned}
&+ \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} f_j \boldsymbol{\epsilon}_j^{\mathcal{E}} \\
&+ \left(\sum_{j=1}^{J} w_j^* \xi_{jt} - \sum_{j=1}^{J} f_j \xi_{jt} \right). \qquad \text{(B.4)}
\end{aligned}
$$

Only the first term on the right-hand side of (B.4) has a non-zero mean (because the weights, w_j^*, depend on the error terms $\boldsymbol{\epsilon}_j^{\mathcal{E}}$). Therefore,

$$
\left| E\left[\sum_{j=1}^{J} w_j^* Y_{jt}^I - \sum_{j=1}^{J} f_j Y_{jt}^I \right] \right| = \left| E\left[\boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} \right] \right|. \qquad \text{(B.5)}
$$

Using the same line of reasoning for the second term on the right-hand side of (B.1), we obtain

$$
\left| E\left[\sum_{j=1}^{J} v_j^* Y_{jt}^N - \sum_{j=1}^{J} f_j Y_{jt}^N \right] \right| = \left| E\left[\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} v_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} \right] \right|. \qquad \text{(B.6)}
$$

For any $t \geq T_0 + 1$ and $s \in T_{\mathcal{E}}$, under Assumption 2 (ii), we apply Cauchy-Schwarz inequality and the eigenvalue bound on the Rayleigh quotient to obtain

$$
\begin{aligned}
\left(\boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_s\right)^2 &\leq \left(\boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\eta}_t\right)\left(\boldsymbol{\lambda}_s'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_s\right) \\
&\leq \left(\frac{\bar{\eta}^2 F}{T_{\mathcal{E}}\underline{\zeta}}\right)\left(\frac{\bar{\lambda}^2 F}{T_{\mathcal{E}}\underline{\zeta}}\right).
\end{aligned}
$$

Similarly,

$$
\left(\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_s\right)^2 \leq \left(\frac{\bar{\lambda}^2 F}{T_{\mathcal{E}}\underline{\zeta}}\right)^2.
$$

Let

$$
\begin{aligned}
\bar{\epsilon}_{jt}^{\mathcal{E}} &= \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\epsilon}_j^{\mathcal{E}} \\
&= \sum_{s\in\mathcal{E}} \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_s \epsilon_{js}.
\end{aligned}
$$

Because $\bar{\epsilon}_{jt}^{\mathcal{E}}$ is a linear combination of independent sub-Gaussians with variance proxy $\bar{\sigma}^2$, it follows that $\bar{\epsilon}_{jt}^{\mathcal{E}}$ is sub-Gaussian with variance proxy $(\bar{\eta}\bar{\lambda}F/\underline{\zeta})^2\bar{\sigma}^2/T_{\mathcal{E}}$. Let $q \geq 1$ be any positive integer. We obtain

$$
\left| E\left[\sum_{j=1}^{J} w_j^* Y_{jt}^I - \sum_{j=1}^{J} f_j Y_{jt}^I \right] \right| \leq E\left[\sum_{j=1}^{J} w_j^* |\bar{\epsilon}_{jt}^{\mathcal{E}}| \right]
$$

$$
\begin{aligned}
&\leq E\left[\left(\sum_{j=1}^{J} |\bar{\epsilon}_{jt}^{\mathcal{E}}|^q\right)^{1/q}\right] \\
&\leq \left(E\left[\sum_{j=1}^{J} |\bar{\epsilon}_{jt}^{\mathcal{E}}|^q\right]\right)^{1/q} \\
&= \left(\sum_{j=1}^{J} E\left[|\bar{\epsilon}_{jt}^{\mathcal{E}}|^q\right]\right)^{1/q} \\
&\leq \left(J(2(\bar{\eta}\bar{\lambda}F/\underline{\zeta})^2\bar{\sigma}^2/T_{\mathcal{E}})^{q/2} q\Gamma(q/2)\right)^{1/q} \\
&= \frac{\bar{\lambda}\bar{\eta}F}{\underline{\zeta}} J^{1/q}\sqrt{2}\bar{\sigma}(q\Gamma(q/2))^{1/q}\frac{1}{\sqrt{T_{\mathcal{E}}}}. \qquad \text{(B.7)}
\end{aligned}
$$

The first inequality is due to triangular inequality when we exchange the absolute value and expectation; the second inequality is due to Holder's inequality; the third is due to Jensen's inequality because for any $q \geq 1$, $x^{1/q}$ is concave in x; the first equality is due to linearity of expectations; and the last inequality is bounding the absolute moments of sub-Gaussian random variables (see, e.g., Rigollet and Hütter, 2019, Lemma 1.4). An analogous argument yields,

$$
\left| E\left[\sum_{j=1}^{J} v_j^* Y_{jt}^N - \sum_{j=1}^{J} f_j Y_{jt}^N\right]\right| \leq \frac{\bar{\lambda}^2 F}{\underline{\zeta}} J^{1/q}\sqrt{2}\bar{\sigma}(q\Gamma(q/2))^{1/q}\frac{1}{\sqrt{T_{\mathcal{E}}}}. \qquad \text{(B.8)}
$$

Equations (B.7) and (B.8) directly yield the result of the theorem.

Lemma B.1 *Let X be a continuously distributed random variable with a density f_X. Let Λ_X be the smallest upper bound on the probability density f_X.*

1. *The random variable $|X|$ has a density $f_{|X|}$ bounded by $\Lambda_{|X|} \leq 2\Lambda_X$;*
2. *For any constant $a \neq 0$, the random variable aX has a density, f_{aX}, bounded by $\Lambda_{aX} \leq \Lambda_X/a$;*
3. *Let X and Y be two independent continuous random variables with densities bounded by Λ_X and Λ_Y, respectively. Then, $Z = X + Y$ has a density, f_Z, bounded by*

$$
\Lambda_Z \leq \min\{\Lambda_X, \Lambda_Y\}.
$$

4. *Let $A_1, \ldots, A_n$ be n random variables (potentially correlated, not necessarily continuous). Let $X_1, \ldots, X_n$ be n independent continuous random variables with densities bounded by $\Lambda_{X_1}, \ldots, \Lambda_{X_n}$, respectively. Let $X_1, \ldots, X_n$ be independent of $A_1, \ldots, A_n$. Then, $Z = \sum_{i=1}^{n} |A_i + X_i|$ has a density, f_Z, bounded by*

$$\Lambda_Z \leq 2 \min_{i=1,\ldots,n} \{\Lambda_{X_i}\}.$$

Proof of Lemma B.1. To prove 1, note that for any $v \geq 0$,

$$f_{|X|}(v) = f_X(v) + f_X(-v) \leq 2\Lambda_X.$$

To prove 2, note that for any $v \geq 0$,

$$f_{aX}(v) = \frac{1}{|a|} f_X(v/a) \leq \frac{1}{|a|} \Lambda_X.$$

To prove 3, note that for any $v \geq 0$,

$$\begin{aligned} f_Z(v) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(v-x) \mathrm{d}x \\ &\leq \int_{-\infty}^{+\infty} f_X(x) \Lambda_Y \mathrm{d}x \\ &= \Lambda_Y. \end{aligned}$$

Similarly, we have $f_Z(v) \leq \Lambda_X$ for any $v \in \mathbb{R}$. So $\Lambda(Z) \leq \min\{\Lambda(X), \Lambda(Y)\}$.

To prove 4, we use the following notations. Let $\Omega(\cdot, \ldots, \cdot)$ be the σ-field that defines the joint distribution of $A_1, \ldots, A_n$; let $\mathbb{P}(\cdot, \ldots, \cdot)$ be the probability measure of $A_1, \ldots, A_n$. Since $X_1, \ldots, X_n$ are independent continuous random variables, we may use the convolution formula to obtain, for any $v \geq 0$,

$$\begin{aligned} & f_Z(v) \\ =& \int_{\Omega(\cdot,\ldots,\cdot)} f_Z(v | A_1 = a_1, \ldots, A_n = a_n) \mathrm{d}\mathbb{P}(a_1, \ldots, a_n) \\ =& \int_{\Omega(\cdot,\ldots,\cdot)} \left(\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{|a_1+X_1|}(x_1) \ldots f_{|a_{n-1}+X_{n-1}|}(x_{n-1}) f_{|a_n+X_n|}(v - \sum_{i=1}^{n-1} x_i) \mathrm{d}x_1 \ldots \mathrm{d}x_{n-1} \right) \mathrm{d}\mathbb{P}(a_1, \ldots, a_n) \\ \leq& \int_{\Omega(\cdot,\ldots,\cdot)} \left(\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f_{|a_1+X_1|}(x_1) \ldots f_{|a_{n-1}+X_{n-1}|}(x_{n-1}) \ (2\Lambda_{X_n}) \ \mathrm{d}x_1 \ldots \mathrm{d}x_{n-1} \right) \mathrm{d}\mathbb{P}(a_1, \ldots, a_n) \end{aligned}$$

$= 2\Lambda_{X_n},$

where the inequality is due to Part 2 of Lemma B.1. Similarly, we have $f_Z(v) \leq 2\Lambda_{X_i}$ for any $i = 1, \ldots, n-1$ and $v \in \mathbb{R}$. So $\Lambda_Z \leq 2\min_{i=1,\ldots,n}\{\Lambda_{X_i}\}$.

Proof of Theorem 2. Define $z = (T_{\mathcal{E}})^{-1/4}$. Let

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

for $t \in \mathcal{B} \cup \{T_0+1, \ldots, T\}$. For $t = T_0+1, \ldots, T$, $\widehat{u}_t$ are the post-intervention estimates of the treatment effects; and for $t \in \mathcal{B}$, $\widehat{u}_t$ are the placebo treatment effects estimated for the blank periods. Using this notation, $\widehat{\boldsymbol{r}} = (\widehat{u}_{T_0+1}, \ldots, \widehat{u}_T, \widehat{u}_{t_1}, \ldots, \widehat{u}_{t_{T_{\mathcal{B}}}})$. Let

$$u_t = \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \tag{B.9}$$

for $t \in \mathcal{B}$, and

$$u_t = \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \xi_{jt} \tag{B.10}$$

for $t = T_0+1, \ldots, T$. Under the null hypothesis in (15), the random variables u_t for $t \in \mathcal{B} \cup \{T_0+1, \ldots, T\}$ are i.i.d. We will compare this sequence of random variables to $\widehat{\boldsymbol{r}}$. Notice that

$$\sum_{j=1}^{J} w_j^* \boldsymbol{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{Y}_j^{\mathcal{E}} = \boldsymbol{\theta}_{\mathcal{E}} \left(\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{Z}_j \right) + \boldsymbol{\lambda}_{\mathcal{E}} \left(\sum_{j=1}^{J} w_j^* \boldsymbol{\mu}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{\mu}_j \right) + \left(\sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} \right).$$

Assumption 3 implies

$$\left(\sum_{j=1}^{J} w_j^* \boldsymbol{\mu}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{\mu}_j \right) = -(\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \left(\sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} \right).$$

It follows that

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

$$
\begin{aligned}
&= -\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_\mathcal{E}'\boldsymbol{\lambda}_\mathcal{E})^{-1}\boldsymbol{\lambda}_\mathcal{E}'\left(\sum_{j=1}^J w_j^*\boldsymbol{\epsilon}_j^\mathcal{E} - \sum_{j=1}^J v_j^*\boldsymbol{\epsilon}_j^\mathcal{E}\right) + u_t \\
&= -\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_\mathcal{E}'\boldsymbol{\lambda}_\mathcal{E})^{-1}\boldsymbol{\lambda}_\mathcal{E}'\sum_{j=1}^J (w_j^* - v_j^*)\boldsymbol{\epsilon}_j^\mathcal{E} + u_t, \qquad \text{(B.11)}
\end{aligned}
$$

for $t \in \mathcal{B}\cup\{T_0+1,\ldots,T\}$. We next find a high probability bound for $\left|\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_\mathcal{E}'\boldsymbol{\lambda}_\mathcal{E})^{-1}\boldsymbol{\lambda}_\mathcal{E}'\sum_{j=1}^J(w_j^* - v_j^*)\boldsymbol{\epsilon}_j^\mathcal{E}\right|$. Similar to the proof of Theorem 1, for $t \in \mathcal{B} \cup \{T_0+1,\ldots,T\}$, let

$$
\breve{\epsilon}_{jt}^\mathcal{E} = \sum_{s\in\mathcal{E}} \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_\mathcal{E}'\boldsymbol{\lambda}_\mathcal{E})^{-1}\boldsymbol{\lambda}_s\epsilon_{js}.
$$

Because $\breve{\epsilon}_{jt}^\mathcal{E}$ is a linear combination of independent sub-Gaussians with variance proxy $\bar{\sigma}^2$, $\breve{\epsilon}_{jt}^\mathcal{E}$ is sub-Gaussian with variance proxy $(\bar{\lambda}^2 F/\underline{\zeta})^2\bar{\sigma}^2/T_\mathcal{E}$. Notice that

$$
\begin{aligned}
\left|\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_\mathcal{E}'\boldsymbol{\lambda}_\mathcal{E})^{-1}\boldsymbol{\lambda}_\mathcal{E}'\sum_{j=1}^J(w_j^* - v_j^*)\boldsymbol{\epsilon}_j^\mathcal{E}\right| &\le \sum_{j=1}^J w_j^*\left|\breve{\epsilon}_{jt}^\mathcal{E}\right| + \sum_{j=1}^J v_j^*\left|\breve{\epsilon}_{jt}^\mathcal{E}\right| \\
&\le 2\max_{\breve{\boldsymbol{w}}\in\mathcal{S}_{J-1}}\sum_{j=1}^J \breve{w}_j\left|\breve{\epsilon}_{jt}^\mathcal{E}\right|,
\end{aligned}
$$

where $\mathcal{S}_{J-1}$ stands for a simplex with $(J-1)$ degrees of freedom, and $\breve{\boldsymbol{w}}$ is any vector that lives in this simplex. For any $z > 0$ and any $\breve{\boldsymbol{w}} \in \mathcal{S}_{J-1}$,

$$
\begin{aligned}
\Pr\left(2\sum_{j=1}^J \breve{w}_j\left|\breve{\epsilon}_{jt}^\mathcal{E}\right| \ge z\right) &\le \sum_{j=1}^J \Pr\left(\left|\breve{\epsilon}_{jt}^\mathcal{E}\right| \ge \frac{z}{2}\right) \\
&= J\Pr\left(\left|\breve{\epsilon}_{jt}^\mathcal{E}\right| \ge \frac{z}{2}\right) \\
&\le 2J\exp\left(-\frac{z^2\underline{\zeta}^2}{8\bar{\sigma}^2\bar{\lambda}^4F^2}T_\mathcal{E}\right), \qquad \text{(B.12)}
\end{aligned}
$$

where the first inequality follows $\Pr(\sum_{j=1}^J \breve{w}_j x_j \ge c) \le \sum_{j=1}^J \Pr(x_j \ge c)$, where $\sum_{j=1}^J \breve{w}_j = 1$, the equality holds because ϵ_{jt} are i.i.d. random variables, and the the last inequality is the Chernoff bound for sub-Gaussian variables. Using what we have defined for $z = (T_\mathcal{E})^{-1/4}$, the bound in (B.12) becomes

$$
c_2(T_\mathcal{E}) = 2J\exp\left(-\frac{\underline{\zeta}^2}{8\bar{\sigma}^2\bar{\lambda}^4F^2}\sqrt{T_\mathcal{E}}\right).
$$

As a result, for any $t \in \mathcal{B} \cup \{T_0+1,\dots,T\}$, we have $|\widehat{u}_t - u_t| \le (T_{\mathcal{E}})^{-1/4}$ with probability at least $1 - c_2(T_{\mathcal{E}})$.

To conclude the proof of Theorem 2, we define the following two clean events. First, define the event

$$\begin{aligned}
\mathcal{C}_1 &= \left\{ \forall t \in \mathcal{B} \cup \{T_0+1,\dots,T\}, \left| \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} (w_j^* - v_j^*)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right| \le (T_{\mathcal{E}})^{-1/4} \right\} \\
&= \bigcap_{t \in B \cup \{T_0+1,\dots,T\}} \left\{ \left| \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} (w_j^* - v_j^*)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right| \le (T_{\mathcal{E}})^{-1/4} \right\} \\
&= \left(\bigcup_{t \in B \cup \{T_0+1,\dots,T\}} \left\{ \left| \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} (w_j^* - v_j^*)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right| > (T_{\mathcal{E}})^{-1/4} \right\} \right)^{\mathsf{c}}.
\end{aligned}$$

Due to union bound, the event $\mathcal{C}_1$ happens with probability at least

$$\begin{aligned}
\Pr(\mathcal{C}_1) &\ge 1 - \sum_{t \in B \cup \{T_0+1,\dots,T\}} \Pr\left(\left| \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} (w_j^* - v_j^*)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right| > (T_{\mathcal{E}})^{-1/4} \right) \\
&\ge 1 - (T - T_{\mathcal{E}}) c_2(T_{\mathcal{E}}).
\end{aligned}$$

Second, define the event

$$\begin{aligned}
\mathcal{C}_2 &= \left\{ \forall \pi, \pi' \in \Pi, |S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}_{\pi'})| > 2(T_{\mathcal{E}})^{-1/4} \right\} \\
&= \left(\bigcup_{\pi,\pi' \in \Pi} \left\{ |S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}_{\pi'})| \le 2(T_{\mathcal{E}})^{-1/4} \right\} \right)^{\mathsf{c}}.
\end{aligned}$$

Due to union bound, the event $\mathcal{C}_2$ happens with probability at least

$$\begin{aligned}
\Pr(\mathcal{C}_2) &\ge 1 - \sum_{\pi,\pi' \in \Pi} \Pr\left(|S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}_{\pi'})| \le 2(T_{\mathcal{E}})^{-1/4} \right) \\
&\ge 1 - |\Pi|^2 \, 2J(T - T_0)\kappa (T_{\mathcal{E}})^{-1/4},
\end{aligned}$$

where the first inequality is due to union bound. The second inequality is because for any two distinct $\pi, \pi' \in \Pi$, $|S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}_{\pi'})|$ has bounded probability density of at most $2J(T - T_0)\kappa$ so integrating over an interval of length $2(T_{\mathcal{E}})^{-1/4}$ would lead to at most $4J(T - T_0)\kappa(T_{\mathcal{E}})^{-1/4}$ probability; and there are fewer than $|\Pi|^2/2$ distinct pairs of combinations.

To see why $|S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}_{\pi'})|$ has bounded density $2J(T-T_0)\kappa$, first note that, under Lemma B.1, it suffices to show that for any $\pi \in \Pi$, $S(\widehat{\boldsymbol{e}}_\pi)$ has bounded density $J(T-T_0)\kappa$.

Note that, under Lemma B.1, u_t has bounded probability density of at most

$$\Lambda_{u_t} \le \min_{j=1,\ldots,J} \left\{ \min \left\{ \frac{1}{w_j^*}, \frac{1}{v_j^*} \right\} \right\} \kappa \le \frac{J}{2}\kappa, \tag{B.13}$$

because u_t is a weighted average of independent random noises, with the weights being w_j^* and v_j^* for all $j = 1, \ldots, J$ (we refer to $1/0 = +\infty$). The worst case is when $w_j^* = 2/J$ for one half of total units and $v_j^* = 2/J$ for the other half.

Next, conditional on any $(\boldsymbol{w}, \boldsymbol{v})$, we focus on $S(\widehat{\boldsymbol{e}}_\pi)$. Recall that $\widehat{\boldsymbol{r}} = (\widehat{u}_{T_0+1}, \ldots, \widehat{u}_T, \widehat{u}_{t_1}, \ldots, \widehat{u}_{t_{T_\mathcal{B}}})$. For any $\pi \in \Pi$ and any $t = 1, \ldots, T-T_0$, denote $\widehat{r}_{\pi(t)} = \widehat{u}_{\mathfrak{r}(t)}$.

$$\begin{aligned} S(\widehat{\boldsymbol{e}}_\pi) &= \frac{1}{T-T_0} \sum_{t=1}^{T-T_0} \left| \widehat{r}_{\pi(t)} \right| \\ &= \frac{1}{T-T_0} \sum_{t=1}^{T-T_0} \Bigg| \underbrace{-\boldsymbol{\lambda}'_{\mathfrak{r}(t)} (\boldsymbol{\lambda}'_\mathcal{E} \boldsymbol{\lambda}_\mathcal{E})^{-1} \boldsymbol{\lambda}'_\mathcal{E} \sum_{j=1}^J (w_j - v_j) \boldsymbol{\epsilon}_j^\mathcal{E}}_{A} + \underbrace{u_{\mathfrak{r}(t)}}_{X} \Bigg|. \end{aligned}$$

Conditional on any $(\boldsymbol{w}, \boldsymbol{v})$, we know $-\boldsymbol{\lambda}'_{\mathfrak{r}(t)} (\boldsymbol{\lambda}'_\mathcal{E} \boldsymbol{\lambda}_\mathcal{E})^{-1} \boldsymbol{\lambda}'_\mathcal{E} \sum_{j=1}^J (w_j - v_j) \boldsymbol{\epsilon}_j^\mathcal{E}$ and $u_{\mathfrak{r}(t)}$ are independent, and $u_{\mathfrak{r}(t)}$ are independent across time. Due to Lemma B.1, thinking of the first term as A term and $u_{\mathfrak{r}(t)}$ as X term, we know $\Lambda_{S(\widehat{e})} \le 2(T-T_0) \min_{t \in \{1,\ldots,T-T_0\}} \{\Lambda_{u_{\mathfrak{r}(t)}}\}$. Combining with (B.13), we know $S(\widehat{\boldsymbol{e}})$ has bounded density of at most $J(T-T_0)\kappa$.

Now that we have lower bounded the probability of each event $\mathcal{C}_1$ and $\mathcal{C}_2$ happening, we know that the probability of both events $\mathcal{C}_1$ and $\mathcal{C}_2$ happening is at least $1-(T-T_\mathcal{E})c_2(T_\mathcal{E}) - |\Pi|^2 2J(T-T_0)\kappa(T_\mathcal{E})^{-1/4}$. More specifically, if we pick

$$c_1(T_\mathcal{E}) = 2J(T-T_\mathcal{E}) \exp\left(-\frac{\zeta^2}{8\bar{\sigma}^2 \bar{\lambda}^4 F^2} (T_\mathcal{E})^{1/2} \right) + |\Pi|^2 2J(T-T_0)\kappa(T_\mathcal{E})^{-1/4},$$

then with probability at least $1 - c_1(T_\mathcal{E})$ the two clean events hold.

Conditional on both $\mathcal{C}_1$ and $\mathcal{C}_2$, we proceed with the following analysis. Define

$$\begin{aligned} \boldsymbol{r} &= (r_1, \ldots, r_{T-T_\mathcal{E}}) \\ &= (u_{T_0+1}, \ldots, u_T, u_{t_1}, \ldots, u_{T_\mathcal{B}}). \end{aligned}$$

Based on $\boldsymbol{r}$, for any $\pi \in \Pi$, define the following $(T_1 \times 1)$ vector,

$$\boldsymbol{u}_\pi = (r_{\pi(1)}, \ldots, r_{\pi(T_1)}).$$

Now we focus on the rank of $S(\widehat{\boldsymbol{e}}_\pi)$ and $S(\boldsymbol{u}_\pi)$. First there are no ties with probability one, i.e., for any two distinct combinations $\pi, \pi' \in \Pi$, $S(\widehat{\boldsymbol{e}}_\pi) \neq S(\widehat{\boldsymbol{e}}_{\pi'})$ and $S(\boldsymbol{u}_\pi) \neq S(\boldsymbol{u}_{\pi'})$ with probability one. This is because $S(\widehat{\boldsymbol{e}}_\pi)$ and $S(\boldsymbol{u}_\pi)$ are continuous random variables.

Next, we argue that with probability $1 - c_1(T_\mathcal{E})$, for any two distinct combinations $\pi, \pi' \in \Pi$, $S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}_{\pi'})$ and $S(\boldsymbol{u}_\pi) - S(\boldsymbol{u}_{\pi'})$ have the same sign. Without loss of generality suppose that $S(\widehat{\boldsymbol{e}}_\pi) > S(\widehat{\boldsymbol{e}}_{\pi'})$. Conditional on $\mathcal{C}_2$, this implies

$$S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}_{\pi'}) > 2(T_\mathcal{E})^{-1/4}.$$

Further conditional on $\mathcal{C}_1$, $|\widehat{u}_t - u_t| \leq (T_\mathcal{E})^{-1/4}$ for any $t \in \mathcal{B} \cup \{T_0 + 1, \ldots, T\}$, so that for any π, we have

$$|S(\widehat{\boldsymbol{e}}_\pi) - S(\boldsymbol{u}_\pi)| \leq \frac{1}{T - T_0} \sum_{t=1}^{T-T_0} \left| \widehat{u}_{\pi(t)} - u_{\pi(t)} \right| \leq (T_\mathcal{E})^{-1/4},$$

where the first inequality is due to triangular inequality. Combining both events we have that, conditional on both $\mathcal{C}_1$ and $\mathcal{C}_2$ (which happens with probability at least $1 - c_1(T_\mathcal{E})$),

$$[S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}_{\pi'})][S(\boldsymbol{u}_\pi) - S(\boldsymbol{u}_{\pi'})] \geq [S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}_{\pi'})][S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}_{\pi'}) - 2(T_\mathcal{E})^{-1/4}] > 0 \tag{B.14}$$

happens for all $\pi, \pi' \in \Pi$.

To conclude the proof, note that there are $|\Pi|$ many distinct combinations, which we denote as $\Pi = \{\pi_1, ..., \pi_{|\Pi|}\}$. Define $\mathfrak{P}$ to be any permutation, $\mathfrak{P} : \{1, \ldots, |\Pi|\} \to \{1, \ldots, |\Pi|\}$. Again conditional on both $\mathcal{C}_1$ and $\mathcal{C}_2$ (which happens with probability at least $1 - c_1(T_\mathcal{E})$),

$$\begin{aligned} \Pr\left(S(\widehat{\boldsymbol{e}}_{\pi_{\mathfrak{P}(1)}}) > \ldots > S(\widehat{\boldsymbol{e}}_{\pi_{\mathfrak{P}(|\Pi|)}})\right) &= \Pr\left(S(\boldsymbol{u}_{\pi_{\mathfrak{P}(1)}}) > \ldots > S(\boldsymbol{u}_{\pi_{\mathfrak{P}(|\Pi|)}})\right) \\ &= \frac{1}{|\Pi|!}, \end{aligned} \tag{B.15}$$

where the first equality is due to (B.14) that $S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}_{\pi'})$ and $S(\boldsymbol{u}_\pi) - S(\boldsymbol{u}_{\pi'})$ have the same sign for any two distinct combinations $\pi, \pi' \in \Pi$; and the second equality is because $\boldsymbol{u}$ is a sequence of i.i.d. random variables so all the rankings happen with equal probability.

Let $S^{(1)}(\widehat{\boldsymbol{e}}) > S^{(2)}(\widehat{\boldsymbol{e}}) > \ldots > S^{(|\Pi|)}(\widehat{\boldsymbol{e}})$ denote the decreasing rearrangement of $\{S(\widehat{\boldsymbol{e}}_\pi)\}_{\pi\in\Pi}$, which exists with probability one. For any $\alpha \in (0,1]$, let $k(\alpha) = \lfloor |\Pi|\alpha \rfloor$. We have that, conditional on both $\mathcal{C}_1$ and $\mathcal{C}_2$ (which happens with probability at least $1 - c_1(T_{\mathcal{E}})$),

$$
\begin{aligned}
\Pr(\widehat{p} \leq \alpha) &= \Pr\left(\sum_{\pi\in\Pi} \mathbb{1}\{S(\widehat{\boldsymbol{e}}_\pi) \geq S(\widehat{\boldsymbol{e}})\} \leq \alpha|\Pi|\right) \\
&= \Pr\left(\sum_{\pi\in\Pi} \mathbb{1}\{S(\widehat{\boldsymbol{e}}_\pi) \geq S(\widehat{\boldsymbol{e}})\} \leq k(\alpha)\right) \\
&= \Pr\left(S(\widehat{\boldsymbol{e}}) \geq S^{(k(\alpha))}(\widehat{\boldsymbol{e}})\right) \\
&= \frac{k(\alpha)}{|\Pi|}.
\end{aligned}
$$

The first equality holds because of the definition of p-value, the second and third equalities are implied by the definition of $k(\alpha)$, and the last equality holds because, by equation (B.15), all rankings happen with the same probability. To finish the proof, note that

$$
\alpha - \frac{1}{|\Pi|} \leq \frac{k(\alpha)}{|\Pi|} = \frac{\lfloor |\Pi|\alpha \rfloor}{|\Pi|} \leq \alpha.
$$

C. Implementation of the Optimization Method

To computationally solve (5), we propose two methods. The first method is by enumeration, which takes advantage of the objective function of (5) being separated between $\boldsymbol{w}$ and $\boldsymbol{v}$. If we knew which units were to receive treatment and which units were to receive control, then we could decompose (5) into two classical synthetic control problems and solve both of them efficiently. We brute force enumerate all the possible combinations of the treatment units and control units. Because the roles of treatment and control units can be switched (see Section 2), we only enumerate combinations such that the cardinality of the treatment group is smaller or equal to the cardinality of the control group. In cases when the cardinality constraint $\bar{m}$ is small, this brute force enumeration is very efficient.

The second method solves an unconstrained optimization problem of (5), by converting the optimization problem into the canonical form of a Quadratic Constraint Quadratic Program (QCQP), which we will detail below. The decision variables are w_j and $v_j, \forall\ j = 1, \ldots, J$. For simplicity, we write it in a vector form $\tilde{\boldsymbol{W}} = (w_1, w_2, ..., w_J, v_1, v_2, ..., v_J)$.

Define $P^0 = \{P^0_{k,l}\}_{k,l=1,\dots,2J} \in \mathbb{R}^{2J\times 2J}$, such that P^0 has only two diagonal blocks, while the two off-diagonal blocks are zero. Define $\forall k, l = 1, \dots, 2J$,

$$P^0_{k,l} = \begin{cases} \sum_{i=1}^{M} X_{i,k}X_{i,l}, & k,l = 1,\dots,J; \\ \sum_{i=1}^{M} X_{i,(k-J)}X_{i,(l-J)}, & k,l = J+1,\dots,2J; \\ 0, & \text{otherwise}. \end{cases}$$

Define $\boldsymbol{q}^0 \in \mathbb{R}^{2J}$, such that $\forall k = 1, \dots, 2J$

$$q^0_k = \begin{cases} -2\sum_{i=1}^{M} X_{i,k} \cdot (\sum_{j=1}^{J} f_j X_{i,j}), & k = 1,\dots,J; \\ -2\sum_{i=1}^{M} X_{i,k-J} \cdot (\sum_{j=1}^{J} f_j X_{i,j}), & k \in= J+1,\dots,2J. \end{cases}$$

Further define $\boldsymbol{e}_1 = (1,1,...,1,0,0,...,0)'$ whose first J elements are 1 and last J elements 0; and $\boldsymbol{e}_2 = (0,0,...,0,1,1,...,1)'$ whose first J elements are 0 and last J elements 1.
Finally, define $P^1 = \{P^1_{k,l}\}_{k,l=1,\dots,2J} \in \mathbb{R}^{2J\times 2J}$ such that P^1 only has non-zero values in the two off-diagonal blocks, i.e., $\forall\ k, l = 1, \dots, 2J$,

$$P^1_{k,l} = \begin{cases} 1, & k = l + J; \\ 1, & k = l - J; \\ 0, & \text{otherwise}. \end{cases}$$

Using the above notations we re-write the (non-convex) QCQP as follows,

$$\begin{aligned} \min \quad & \tilde{\boldsymbol{W}}'P^0\tilde{\boldsymbol{W}} + \boldsymbol{q}^{0'}\tilde{\boldsymbol{W}} \\ \text{s.t.} \quad & \boldsymbol{e}_1'\tilde{\boldsymbol{W}} = 1, \\ & \boldsymbol{e}_2'\tilde{\boldsymbol{W}} = 1, \\ & \tilde{\boldsymbol{W}}'P^1\tilde{\boldsymbol{W}} = 0, \\ & \tilde{\boldsymbol{W}} \geq \boldsymbol{0}. \end{aligned} \tag{C.1}$$

Remark 1 *The above problem* (C.1) *can be solved using Gurobi 9.0 solver.* ∎

D. Additional Simulation Results

In this section we present additional simulation results that complement the results in Section 4.1. In Section 4.1 we have prescribed the noises to be i.i.d. Normal $(0, 1)$ variables. In this section we change the noises to be i.i.d. Normal $(0, \sigma^2)$ variables, where we set $\sigma^2 \in \{5, 10\}$ to different scales in this experiment. The simulation results of the synthetic treated unit and the synthetic control unit are in Figures 4 and 5.

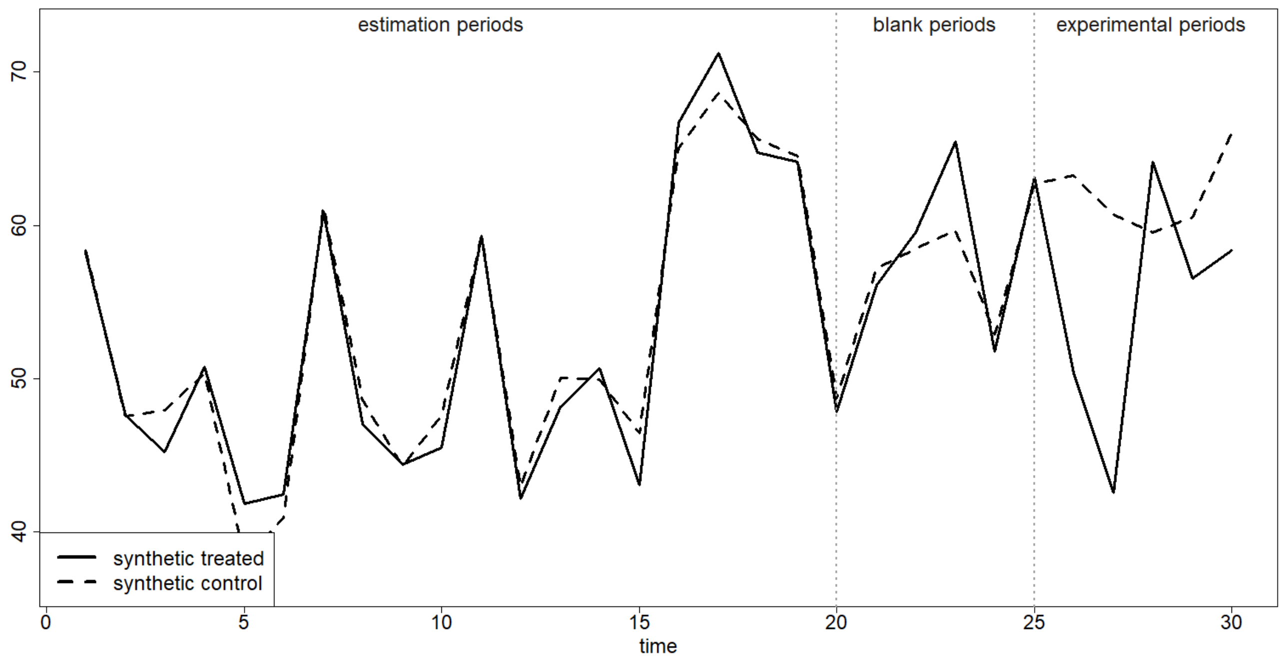


Figure 4: Synthetic Treatment Unit and Synthetic Control Unit, when $\sigma^2 = 5$.

We also present the difference between the synthetic treatment unit and the synthetic control unit when we change the noises to be i.i.d. Normal $(0, 5)$ and $(0, 10)$ variables. The simulation results are in Figures 6 and 7. If we further compare with Figure 2, we notice that when σ^2 increases from 1 to 10, the p-value also increases, indicating less power when the noise is stronger.

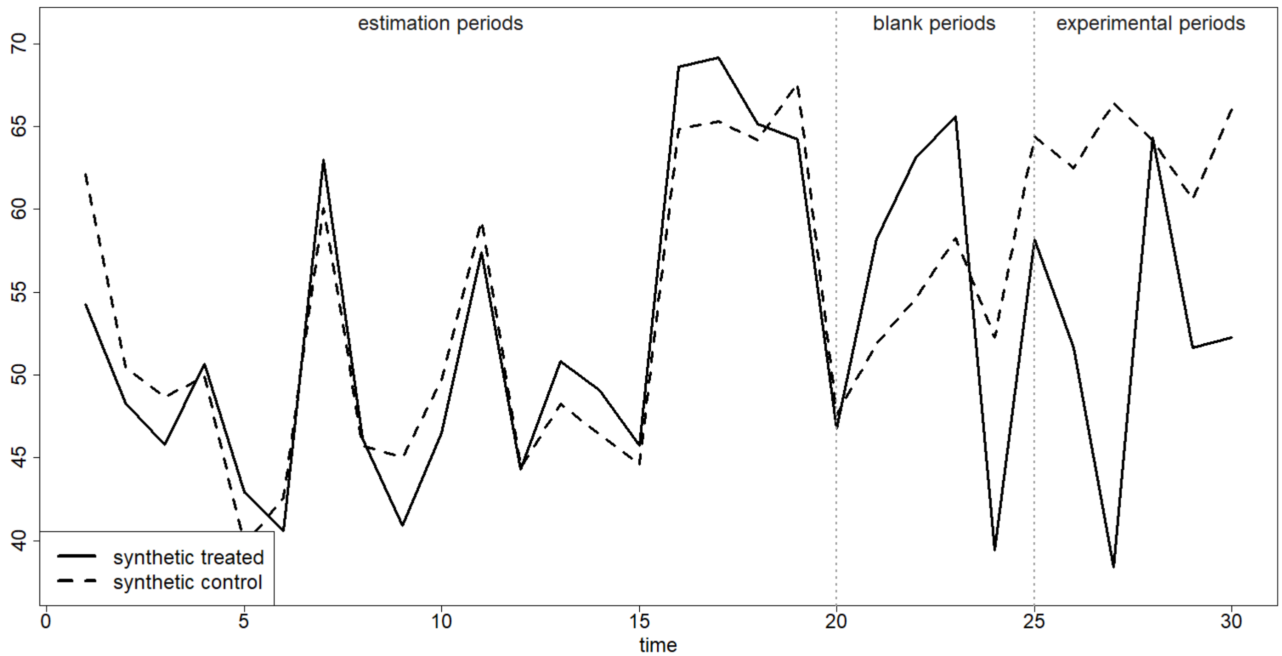


Figure 5: Synthetic Treatment Unit and Synthetic Control Unit, when $\sigma^2 = 10$.

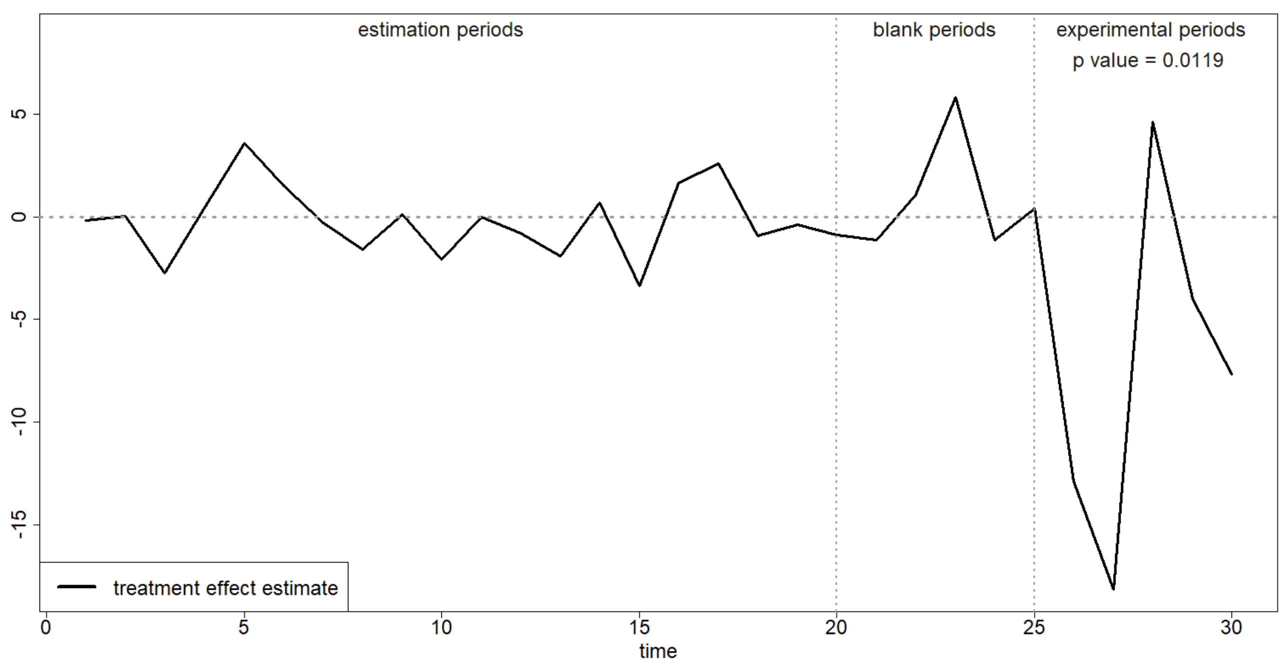


Figure 6: Treatment Effect Estimate, when $\sigma^2 = 5$.

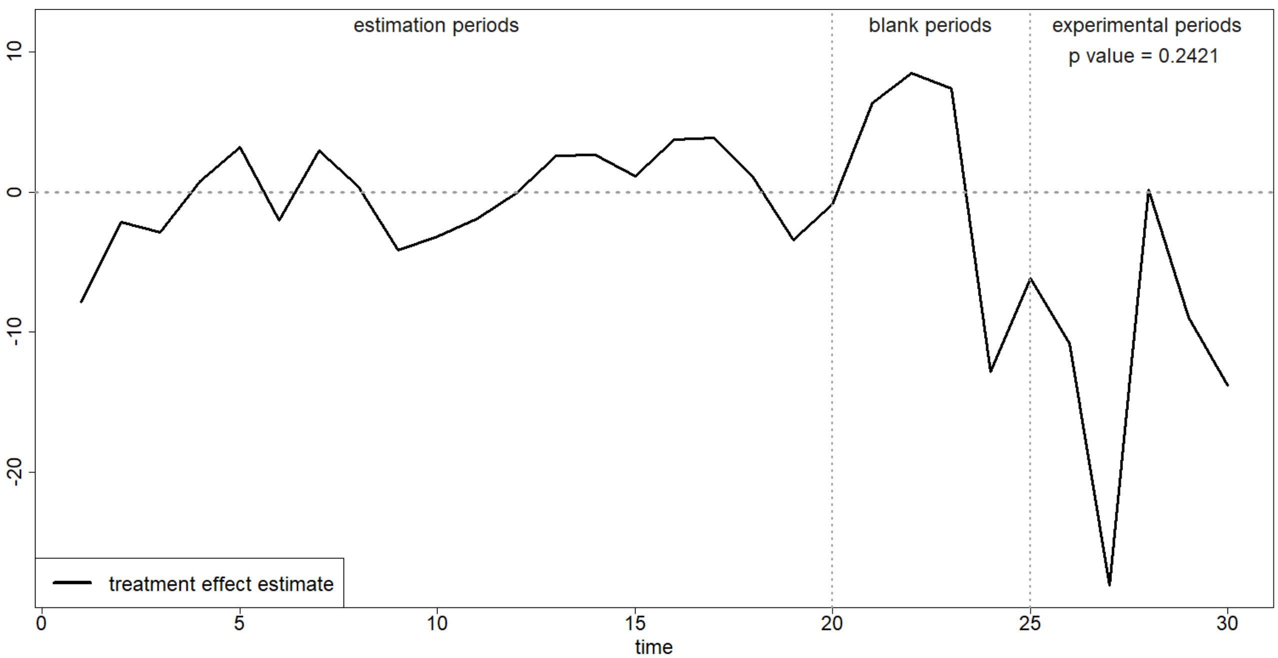


Figure 7: Treatment Effect Estimate, when $\sigma^2 = 10$.