Synthetic Controls for Experimental Design

Alberto Abadie   Jinglong Zhao

MIT   Boston University

April 2025

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.

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 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.

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 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, as well as the untreated units to serve as a comparison group. We adopt the name *synthetic control designs* to refer to the resulting experimental

---

[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 ridesharing 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.

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 experimental 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 takes into account the extent to which different sets of treated and control units approximate an

---

[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), which also provides the basis for the ride-sharing example above.

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.

A related literature applies synthetic control methods and nearest-neighbor matching methods to select experimental sites in multi-site designs (Egami and Lee, 2024, Montiel Olea et al., 2024). In contrast, we examine settings where treatment occurs at the aggregate (i.e., site) level: each site receives only treatment or only control, precluding the estimation of site-level treatment effects.

Agarwal, Shah and Shen (2021) proposes synthetic interventions, a framework related to synthetic controls, and applies it to estimate treatment effect heterogeneity in an experimental setting with multiple treatments. Their work primarily focuses on the analysis of experimental data, but not on the design of experiments. Bottmer et al. (2021) is also related to the present article in the sense that it studies 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.

The research designs in this article are also related to ex-ante synthetic control designs for observational studies (see Abadie, 2021, Kasy and Lehner, 2023, and Chen, 2023, the latter for an online version of the problem) in that the synthetic weights are computed and can be pre-registered before post-intervention outcomes are realized. However, our methods differ significantly in one key aspect: they confront the challenge that experimenters face when selecting

---

[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).

specific units for exposure to the intervention of interest. 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-based aggregate counterfactuals in experimental settings where only a limited number of aggregate units can be treated.

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. Additionally, a stratified design with a small number of strata or treated units may result in the selection of a set of treated units that is not truly representative of the target population.

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

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 the ridesharing 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., $\boldsymbol{f} = (f_1, \ldots, f_J)$.

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

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

$$\sum_{j=1}^{J} v_j = 1, \tag{2}$$

$$w_j \geq 0, \ v_j \geq 0, \ \text{and} \ w_j v_j = 0, \ \forall j = 1, \ldots, J.$$

Units with $w_j > 0$ are units that will be assigned to the intervention of interest from $T_0 + 1$ to T, 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.

7

Instead, we propose a variety of methods to approximate either (4) or (5) based on predictors of $Y^N_{jT_0+1}, \ldots, Y^N_{jT}$.

For the treated units, we define $Y_{jt} = Y^N_{jt}$ if $t = 1, \ldots, T_0$, and $Y_{jt} = Y^I_{jt}$ if $t = T_0 + 1, \ldots, T$. For the untreated units, we define $Y_{jt} = Y^N_{jt}$, 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 effect of the treatment on the treated ($\boldsymbol{w}$-weighted),

$$\tau_t^T = \sum_{j=1}^{J} w_j \cdot (Y^I_{jt} - Y^N_{jt})$$

(Doudchenko et al., 2021).

We choose $\boldsymbol{w} = (w_1, \ldots, w_J)$ and $\boldsymbol{v} = (v_1, \ldots, v_J)$ to match the pre-intervention values of predictors of the potential outcomes Y^N_{jt} and Y^I_{jt} 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^N_{jt} and Y^I_{jt} 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 selector of $\boldsymbol{w} = (w_1, \ldots, w_J)$

and $\boldsymbol{v} = (v_1, \ldots, v_J)$ is

$$\min_{\substack{w_1,\ldots,w_J, \\ v_1,\ldots,v_J}} \left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j \right\|^2 + \left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} v_j \boldsymbol{X}_j \right\|^2$$

$$\text{s.t.} \quad \sum_{j=1}^{J} w_j = 1,$$

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

$$w_j, v_j \geq 0, \quad \forall j = 1, \ldots, J,$$

$$w_j v_j = 0, \quad \forall j = 1, \ldots, J,$$

$$\underline{m} \leq \|\boldsymbol{w}\|_0 \leq \overline{m}. \tag{6}$$

The first term of the objective function in (6) 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 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. We say that the design is *Unconstrained* if $\underline{m} = 1$ and $\overline{m} = J - 1$; otherwise, we say the design is *Constrained*. The last constraint in (6) 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 (6). 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.,

$$\widehat{\tau}_t = \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt}. \tag{7}$$

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 (6) 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 weight selector in (6) 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),

$$\min_{\substack{w_1,\ldots,w_J, \\ v_1,\ldots,v_J}} \left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j \right\|^2 + \beta \left\| \sum_{j=1}^{J} w_j \boldsymbol{X}_j - \sum_{j=1}^{J} v_j \boldsymbol{X}_j \right\|^2$$

$$\text{s.t.} \quad \sum_{j=1}^{J} w_j = 1,$$

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

$$w_j, v_j \geq 0, \quad \forall j = 1, \ldots, J,$$

$$w_j v_j = 0, \quad \forall j = 1, \ldots, J,$$

$$\underline{m} \leq \|\boldsymbol{w}\|_0 \leq \overline{m}. \tag{8}$$

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 $\overline{\boldsymbol{X}}$. A large value of β, on the other hand, favors designs with aggregate treated and aggregate control units that closely match each other. 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 (8). For this formulation of the synthetic control design, the treatment assignment

and estimation procedures follow the same steps as those used in the previous formulation. Large values for β produce estimators that target the $\boldsymbol{w}$-weighted average effect of the treatment on the treated, τ_t^T of Doudchenko et al. (2021). Small values of β prioritize estimation of the average treatment effect, τ_t.

In our third formulation of the synthetic control design, the experimenter selects a synthetic treated unit to match the average values of the characteristics in the population. However, unlike the design in (8), the experimenter chooses multiple synthetic controls, 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, $\mathcal{J}_{\boldsymbol{w}} = \{j : w_j > 0\}$. Our next version of the synthetic control design is:

$$\min_{\substack{w_j, \forall j=1,2,\ldots,J, \\ v_{ij}, \forall i,j=1,2,\ldots,J}} \left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j \right\|^2 + \xi \sum_{j=1}^{J} w_j \left\| \boldsymbol{X}_j - \sum_{i=1}^{J} v_{ij} \boldsymbol{X}_i \right\|^2$$

$$\text{s.t.} \quad \sum_{j=1}^{J} w_j = 1,$$

$$w_j \geq 0, \quad \forall j = 1, \ldots, J,$$

$$\sum_{i=1}^{J} v_{ij} = 1, \quad \forall j \in \mathcal{J}_{\boldsymbol{w}},$$

$$v_{ij} = 0, \quad \forall i \in \mathcal{J}_{\boldsymbol{w}}, \ j = 1, \ldots, J,$$

$$v_{ij} \geq 0, \quad \forall j \in \mathcal{J}_{\boldsymbol{w}}, \ i = 1, \ldots, J,$$

$$v_{ij} = 0, \quad \forall j \notin \mathcal{J}_{\boldsymbol{w}}, \ i = 1, \ldots, J,$$

$$\underline{m} \leq \|\boldsymbol{w}\|_0 \leq \overline{m}. \tag{9}$$

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, 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 (9). 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 $\overline{\boldsymbol{X}} -$
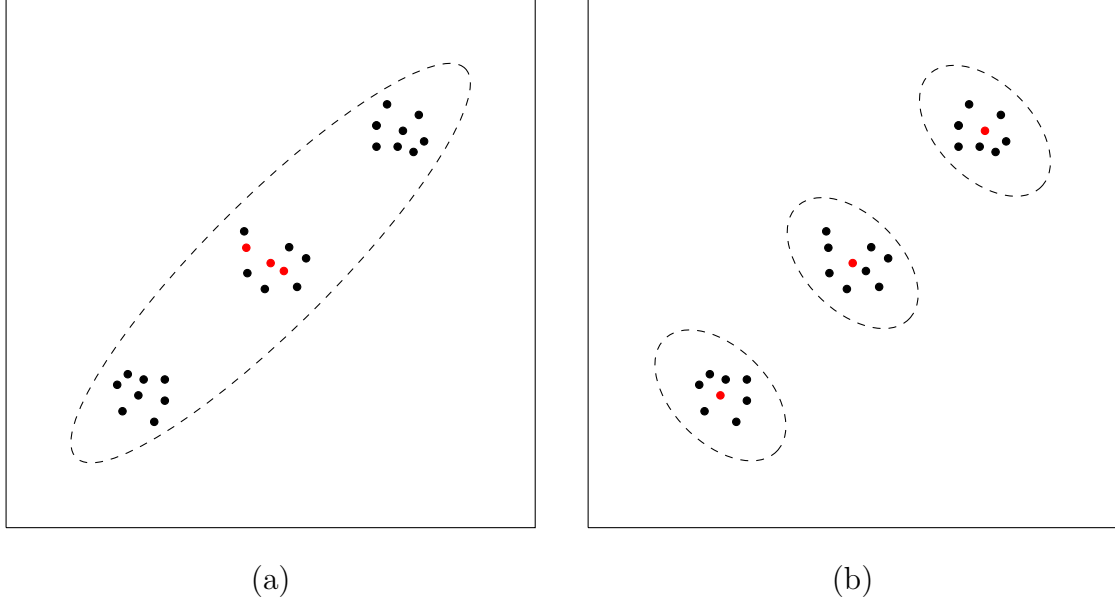
$$(a) \qquad\qquad\qquad (b)$$

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.

$\sum_{j=1}^{J} w_j^* \boldsymbol{X}_j \approx \boldsymbol{0}$ and $\boldsymbol{X}_j - \sum_{j=1}^{J} v_{ij}^* \boldsymbol{X}_j \approx \boldsymbol{0}$ 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

$$v_j^* = \sum_{i=1}^{J} w_i^* v_{ij}^*. \tag{10}$$

Then,

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

$$= \sum_{j=1}^{J} w_j^* \left(Y_{jt} - \sum_{i=1}^{J} v_{ij}^* Y_{it} \right).$$

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.

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 mean is

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

for each cluster $k = 1, \ldots, K$. For each index $i = 1, \ldots, J$, let $k(i)$ be the cluster to which unit i belongs, i.e., $i \in \mathcal{I}_{k(i)}$. A clustered version of the synthetic control design in (9) is given by:

$$\min_{\substack{w_j, \forall j=1,2,\ldots,J, \\ v_{ij}, \forall i,j=1,2,\ldots,J}} \sum_{k=1}^{K} \left(\sum_{j \in \mathcal{I}_k} f_j \right) \left\{ \left\| \overline{\boldsymbol{X}}_k - \sum_{j \in \mathcal{I}_k} w_j \boldsymbol{X}_j \right\|^2 + \xi \sum_{j \in \mathcal{I}_k} w_j \left\| \boldsymbol{X}_j - \sum_{i,j \in \mathcal{I}_k} v_{ij} \boldsymbol{X}_i \right\|^2 \right\}$$

$$\text{s.t.} \quad \sum_{j \in \mathcal{I}_k} w_j = 1, \quad \forall k = 1, \ldots, K,$$

$$w_j \geq 0, \quad \forall j = 1, \ldots, J,$$

$$\sum_{i=1}^{J} v_{ij} = 1, \quad \forall j \in \mathcal{J}_{\boldsymbol{w}}$$

$$v_{ij} \geq 0, \quad \forall j \in \mathcal{J}_{\boldsymbol{w}}, \; i = 1, \ldots, J,$$

$$v_{ij} = 0, \quad \forall j \notin \mathcal{J}_{\boldsymbol{w}}, \; i = 1, \ldots, J,$$

$$v_{ij} = 0, \quad \forall i \in \mathcal{J}_{\boldsymbol{w}}, \; j = 1, \ldots, J,$$

$$v_{ij} = 0, \quad \forall i, j, \text{ such that } k(i) \neq k(j),$$

13

$$\underline{m} \leq \|\boldsymbol{w}\|_0 \leq \overline{m}.$$

We conclude this section by discussing other possible extensions 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 (6), or $\boldsymbol{X}_j$ for a treated unit in equation (9)) fall inside the convex hull of the values of $\boldsymbol{X}_j$ for the units in the donor pool. To address the potential lack of uniqueness, 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 unique provided that 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 units are 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 (6), we obtain

$$\min_{\substack{w_1,\ldots,w_J, \\ v_1,\ldots,v_J}} \quad \left\|\overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j\right\|^2 + \left\|\overline{\boldsymbol{X}} - \sum_{j=1}^{J} v_j \boldsymbol{X}_j\right\|^2$$
$$+ \lambda_1 \sum_{j=1}^{J} w_j \left\|\overline{\boldsymbol{X}} - \boldsymbol{X}_j\right\|^2 + \lambda_2 \sum_{j=1}^{J} v_j \left\|\overline{\boldsymbol{X}} - \boldsymbol{X}_j\right\|^2$$

$$\text{s.t.} \quad \sum_{j=1}^{J} w_j = 1,$$
$$\sum_{j=1}^{J} v_j = 1,$$
$$w_j, v_j \geq 0, \quad \forall j = 1, \ldots, J,$$
$$w_j v_j = 0, \quad \forall j = 1, \ldots, J,$$
$$\underline{m} \leq \|\boldsymbol{w}\|_0 \leq \overline{m}. \tag{11}$$

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.[6]

---

[6]See Abadie and L'Hour (2021) for details on penalized synthetic control estimators. The synthetic control

Other types of penalization are possible. In particular, Doudchenko and Imbens (2016), Doudchenko et al. (2021), and others have proposed synthetic control estimators that use ridge or elastic net regularization on the synthetic control weights (e.g., on w_j and v_j in design (6)). The synthetic control designs proposed in this article can be modified to incorporate regularization on the weights.

Finally, Abadie and L'Hour (2021), Arkhangelsky et al. (2021), and Ben-Michael, Feller and Rothstein (2021) have proposed bias-correction techniques for synthetic control methods. Section OA.1 in the Online Appendix provides details on how to apply bias correction techniques in a synthetic control design.

3. Formal Results

We introduce 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 random shocks.

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.

---

design in (11) is a penalized version of (6). Section OA.1 in the Online Appendix discusses how to apply the Abadie and L'Hour penalty to the other synthetic designs proposed in this article.

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 $T_{\mathcal{E}} = |\mathcal{E}|$, 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

$$\boldsymbol{X}_j = \left(\begin{array}{c} \boldsymbol{Y}_j^{\mathcal{E}} \\ \boldsymbol{Z}_j \end{array} \right),$$

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 (6), and choose where $\boldsymbol{w}^* = (w_1^*, \ldots, w_J^*)$ 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)*

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

and (ii)

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

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 $|\eta_{tf}| \le \overline{\eta}$ for $t = T_0 + 1, \ldots, T$, $f = 1, \ldots, F$, $|\theta_{tr}| \le \overline{\theta}$ for $t = 1, \ldots, T$, $r = 1, \ldots, R$, and $|\gamma_{tr}| \le \overline{\gamma}$, 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 set-up of Section 2.

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

$$|E\left[\widehat{\tau}_t - \tau_t\right]| \leq \frac{\overline{\lambda}(\overline{\eta} + \overline{\lambda})F}{\underline{\zeta}}\sqrt{2\log{(2J)}}\frac{\overline{\sigma}}{\sqrt{T_{\mathcal{E}}}}. \tag{15}$$

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

$$|E\left[\widehat{\tau}_t - \tau_t\right]| \leq \left((\overline{\gamma} + \overline{\theta})R + \frac{\overline{\lambda}(\overline{\eta} + \overline{\lambda})F}{\underline{\zeta}}(1 + \overline{\theta}R)\right)d + \frac{\overline{\lambda}(\overline{\eta} + \overline{\lambda})F}{\underline{\zeta}}\sqrt{2\log{(2J)}}\frac{\overline{\sigma}}{\sqrt{T_{\mathcal{E}}}}. \tag{16}$$

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 (16) 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 (16) 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 (15) and (16) 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 (15) 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.[7]

---

[7]Shifting predictors from $\boldsymbol{Z}_j$ to $\boldsymbol{\mu}_j$ changes the bias bound (15) 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

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 $\mathcal{B} = \{1, \ldots, T_0\} \setminus \mathcal{E}$. We, 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 hypothesis:

> For $t = T_0 + 1, \ldots, T$, and $j = 1, \ldots, J$,
>
> $$Y_{jt}^I = \delta_t + \boldsymbol{\theta}_t' \boldsymbol{Z}_j + \boldsymbol{\lambda}_t' \boldsymbol{\mu}_j + \xi_{jt}, \tag{17}$$
>
> where ξ_{jt} has the same distribution as ϵ_{jt}.

Under the null hypothesis in (17), 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 (7) 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

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

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

---

components of $\boldsymbol{\theta}_t$ to $\boldsymbol{\lambda}_t$, which can change the value of ζ. Poincaré's separation theorem implies that ζ 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 (15) 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 (16), because a shift of components from $\boldsymbol{Z}_j$ to $\boldsymbol{\mu}_j$ decreases the value of R.

$\pi \in \Pi$, let $\pi(i)$ be the ith smallest value in π, and

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

In addition, let $\widehat{e} = (\widehat{u}_{T_0+1}, \ldots, \widehat{u}_T) = (\widehat{\tau}_{T_0+1}, \ldots, \widehat{\tau}_T)$. This is a vector of treatment effect estimates from the experimental periods. For any $(T-T_0)$-dimensional vector $e = (e_1, \ldots, e_{T-T_0})$, we adopt the test statistic,

$$S(e) = \frac{1}{T - T_0} \sum_{t=1}^{T-T_0} |e_t|. \tag{18}$$

Other choices of test statistics are possible, such as those based on an L_p-norm of e (Chernozhukov, Wüthrich and Zhu, 2021) and one-sided versions of the resulting test statistics (i.e., with the positive or the negative parts of e_t replacing $|e_t|$ in equation (18)).

The p-value of a permutation test on (18) is

$$\widehat{p} = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{e}_\pi) \geq S(\widehat{e})\} \tag{19}$$

Theorem 2 below shows that if λ_t are exchangeable random variables for $t \in \mathcal{B} \cup \{T_0+1, \ldots, T\}$, then a test of the null hypothesis in (17) based on the p-value in (19) is exact.

Theorem 2 *Suppose that Assumptions 1, 2(ii), and 3(i) hold. Assume that $\{\lambda_t\}_{t \in \mathcal{B} \cup \{T_0+1,\ldots,T\}}$ is a sequence of exchangeable random variables independent of $\{\epsilon_{jt}\}_{t \in \mathcal{B} \cup \{T_0+1,\ldots,T\}}$ and $\{\xi_{jt}\}_{t \in \{T_0+1,\ldots,T\}}$. Then under the null hypothesis (17), we have*

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

for any $\alpha \in [0,1]$, where $\Pr(\widehat{p} \leq \alpha)$ is taken over the distribution of $\{\xi_{jt}, \epsilon_{jt}, \lambda_t\}$.

Note that, under the assumptions of Theorem 2, the potential outcome series Y_{jt}^N is allowed to be non-stationary through the term $\delta_t + \theta_t' Z_j$ 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 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 (19) 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 (19).

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, Diamond 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.

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

$$\widehat{q}_{1-\alpha} = \inf_{z \in \mathbb{R}} \left\{ \frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\left\{ \left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| \leq z \right\} \geq 1 - \alpha \right\} \tag{20}$$

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

$$\widehat{C}_{1-\alpha}(Y_{1t}, Y_{2t}, ..., Y_{Jt}) = \left[\sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} - \widehat{q}_{1-\alpha}, \ \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} + \widehat{q}_{1-\alpha} \right]. \tag{21}$$

We next show that the confidence interval defined in (21) approximately achieves correct point-wise 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 (21) approximately achieves point-wise coverage, i.e., for any $\alpha \in (0,1)$ and any $t \in \{T_0 + 1, ..., T\}$, as $(T_0 - T_{\mathcal{E}}), T_{\mathcal{E}} \to +\infty$,*

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

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, in the presence of a good pre-intervention fit, 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 (17) for the placebo intervention.
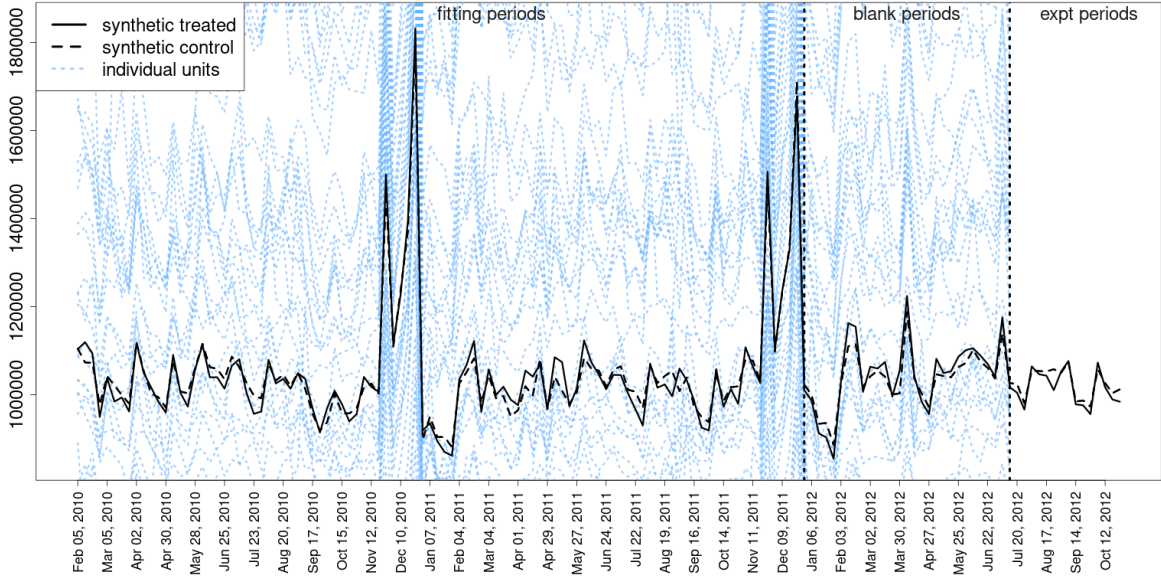
Figure 2: Synthetic Treatment Unit and Synthetic Control Unit, $\overline{m} = 2$

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.

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. For the purpose of estimating the synthetic treated and synthetic control weights, we normalize each of the 100 pre-experimental outcomes to have a unit variance.

We compute synthetic treated and control units that apply the *Constrained* formulation in (6) 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. 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
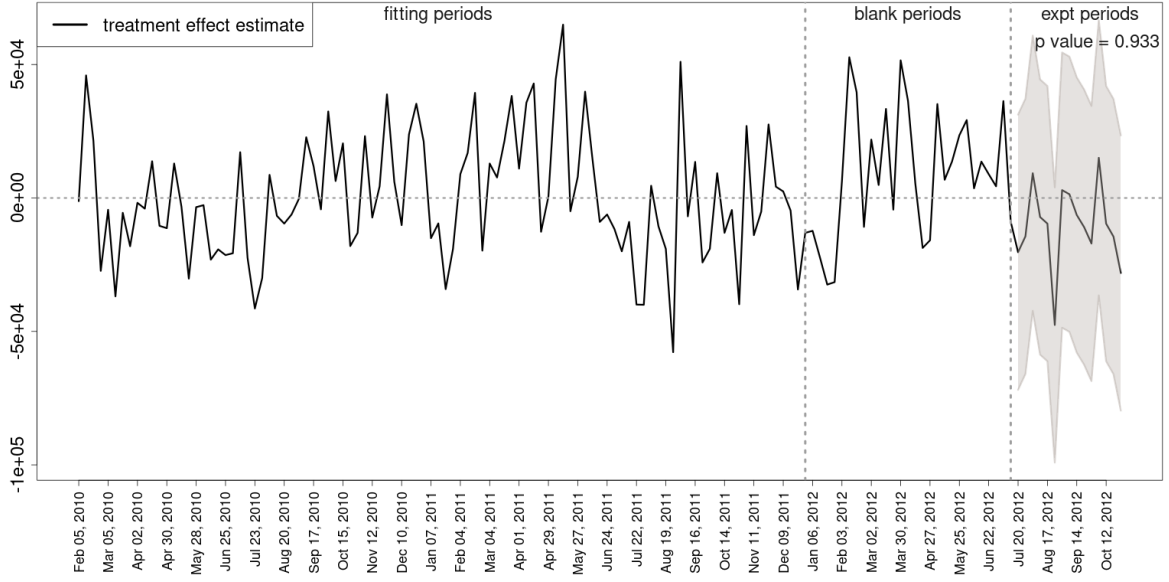
23

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.

(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 (19), calculated over the residuals of Figure 3, is equal to 0.933, which results in a failure to reject the null hypothesis (17). Confidence intervals based on equation (21) 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 5.2.4 for a precise definition of the estimators and RMSE performance metric). For each of the

24

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.

three 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. We consider a setting with $J = 15$ units, $R = 7$ observable covariates, and $F = 11$ unobservable covariates. We simulate data for a total of $T = 30$ periods, comprising $T_0 = 25$ pre-experimental periods and $T - T_0 = 5$ 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. We set the weights f_j in expression (1) to be $f_j = 1/J$, for all $j = 1, ..., J$.

For our baseline simulation design, 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 rearrangements 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 $\sigma^2 = 1$. We present additional simulation results of alternative values of the noise parameter σ^2 in Section OA.7 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 (6), 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 (17).

5.2. Performance Across Many Simulations

This section compares the performance of the different varieties 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. The data generating process is the same as in Section 5.1.

We consider five varieties of the synthetic control design:

1. *Unconstrained design*: This is the design in (6) without a cardinality constraint, so $\underline{m} = 1$ and $\overline{m} = J - 1 = 14$.

2. *Constrained design*: Same as the design in (6), but with $\underline{m} = 1$ and $\overline{m} = 1, \ldots, 7$.

3. *Weakly-targeted design*: This is the design in (8). We vary β from 0.01 to 100.

4. *Unit-level design*: This is the design in (9), which fits a different synthetic control to each
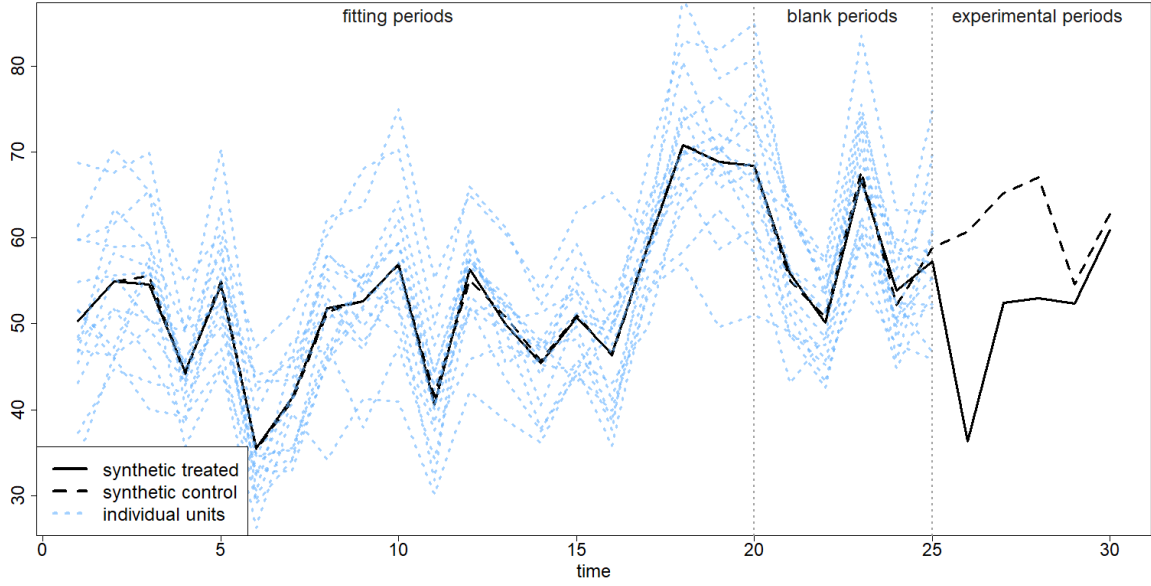
26

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).

unit assigned to treatment. We vary ξ from 0.01 to 100.

5. *Penalized design*: This is the design in (11), with $\lambda = \lambda_1 = \lambda_2$. We vary λ from 0.01 to 100.

The *Constrained* design imposes sparsity in the synthetic treatment weights through a hard cardinality constraint specified by the integer $\overline{m}$. The *Weakly-targeted* design targets the average treatment effect for small values of β and a weighted average effect for the treated for large values of β. For the *Unit-level* design, large values of ξ generate sparsity in the synthetic treated weights. A sufficiently large value of ξ produces a *Unit-level* design where the only single treated unit can be closely fitted by a convex combination of the other units. 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 the *Unit-level* design, synthetic control weights are aggregated as in (10). For the *Unconstrained* and *Penalized* designs, the synthetic treated and synthetic control weights can always be swapped without changing the objective values for their respective designs. For the *Constrained* design, the weights can be swapped when $\|\boldsymbol{v}^*\|_0 \leq \overline{m}$. When it is possible to swap synthetic
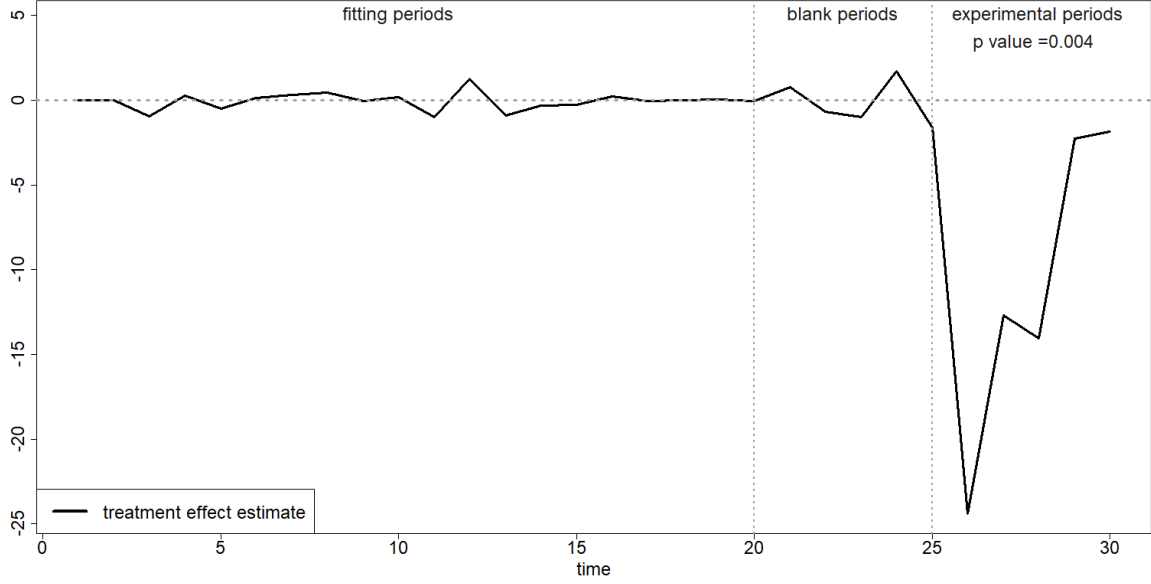
27

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.

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 Section OA.4 of the Online Appendix.

5.2.1. Average Treatment Effects

The first panel of Table 2 reports average treatment effects, τ_t, over 1000 simulations. The second panel reports estimates of the average treatment effects, mean absolute error, root mean square error, and p-value, all averaged over 1000 simulations, as well as rejection rates. Mean absolute error (MAE) and root mean square error are defined as

$$\text{MAE} = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} |\widehat{\tau}_t - \tau_t|, \qquad \text{RMSE} = \sqrt{\frac{1}{T - T_0} \sum_{t=T_0+1}^{T} (\widehat{\tau}_t - \tau_t)^2}, \qquad (22)$$

and the p-value is defined as in (19). 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.

28

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	$RMSE$	$\widehat{p}$	$\widehat{p} < 0.05$
		$t = 26$	$t = 27$	$t = 28$	$t = 29$	$t = 30$				
Unconstrained		-13.57	-10.97	-8.37	-5.06	-2.52	0.83	0.97	0.014	0.946
Constrained	$\overline{m} = 1$	-13.61	-10.97	-8.39	-4.86	-2.41	2.93	3.45	0.057	0.668
	$\overline{m} = 2$	-13.58	-10.90	-8.43	-5.01	-2.40	1.69	2.00	0.028	0.854
	$\overline{m} = 3$	-13.56	-11.00	-8.38	-5.05	-2.52	1.26	1.49	0.019	0.916
	$\overline{m} = 4$	-13.59	-11.06	-8.40	-4.99	-2.50	1.06	1.25	0.016	0.935
	$\overline{m} = 5$	-13.57	-11.01	-8.37	-5.02	-2.48	0.93	1.09	0.015	0.933
	$\overline{m} = 6$	-13.51	-10.95	-8.29	-5.01	-2.47	0.87	1.02	0.015	0.942
	$\overline{m} = 7$	-13.57	-10.96	-8.37	-5.06	-2.52	0.83	0.97	0.014	0.946
Weakly-targeted	$\beta = 0.01$	-13.58	-10.95	-8.38	-4.99	-2.53	1.18	1.38	0.018	0.920
	$\beta = 0.1$	-13.57	-11.00	-8.34	-4.98	-2.52	0.93	1.08	0.014	0.949
	$\beta = 1$	-13.56	-10.98	-8.32	-4.93	-2.44	0.86	1.01	0.013	0.951
	$\beta = 10$	-13.57	-10.98	-8.38	-5.01	-2.51	0.94	1.10	0.013	0.955
	$\beta = 100$	-13.60	-10.98	-8.39	-5.07	-2.52	1.01	1.18	0.013	0.951
Unit-level	$\xi = 0.01$	-13.60	-10.95	-8.39	-5.04	-2.53	0.95	1.13	0.014	0.938
	$\xi = 0.1$	-13.58	-10.97	-8.35	-4.97	-2.47	0.91	1.07	0.015	0.942
	$\xi = 1$	-13.57	-10.99	-8.39	-4.99	-2.49	1.34	1.58	0.020	0.899
	$\xi = 10$	-13.60	-10.93	-8.45	-5.05	-2.52	2.16	2.57	0.030	0.829
	$\xi = 100$	-13.61	-10.86	-8.48	-5.02	-2.54	2.76	3.27	0.040	0.770
Penalized	$\lambda = 0.01$	-13.59	-10.98	-8.35	-5.05	-2.48	0.88	1.02	0.014	0.950
	$\lambda = 0.1$	-13.64	-11.03	-8.43	-5.03	-2.50	1.21	1.43	0.019	0.904
	$\lambda = 1$	-13.67	-10.96	-8.41	-4.87	-2.45	2.08	2.46	0.037	0.791
	$\lambda = 10$	-13.68	-11.04	-8.37	-4.79	-2.45	3.72	4.40	0.091	0.542
	$\lambda = 100$	-13.64	-10.94	-8.42	-4.86	-2.50	4.17	4.93	0.111	0.490

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

In Table 2, the *Unconstrained* design has a strong relative performance. The performance of the *Constrained* design improves for larger $\overline{m}$, and is virtually identical to the performance of the *Unconstrained* design when $\overline{m} = 7$. The performance of *Weakly-targeted* and *Unit-level* designs is best when β and ξ take intermediate values. The *Penalized* design yields results similar to those of the *Unconstrained* design for small values of the penalization parameter λ.

5.2.2. Performance with Nonlinearities

We now examine the behavior of estimators based on synthetic control designs under deviations from the linear model in (12a) and (12b). We consider a nonlinear data generating process,

$$Y_{jt}^N = \delta_t + \exp\left(\boldsymbol{\theta}_t' \boldsymbol{Z}_j\right) + \exp\left(\boldsymbol{\lambda}_t' \boldsymbol{\mu}_j\right) + \epsilon_{jt}, \tag{23a}$$

$$Y_{jt}^I = \upsilon_t + \exp\left(\boldsymbol{\gamma}_t' \boldsymbol{Z}_j\right) + \exp\left(\boldsymbol{\eta}_t' \boldsymbol{\mu}_j\right) + \xi_{jt}. \tag{23b}$$

The motivation to study a nonlinear model is that nonlinearities may induce interpolation biases, affecting the relative performance of the different designs. All parameter values are the same as in the simulation setup of section 5.1, except for the values of $\boldsymbol{\theta}_t$, $\boldsymbol{\gamma}_t$, $\boldsymbol{\lambda}_t$, and $\boldsymbol{\eta}_t$, which are chosen to be random vectors of i.i.d. Uniform $(0,3)$ random variables, instead of Uniform $(0,10)$, to control the magnitude of the exponential components in the nonlinear design.

Table 3 reports the results for τ_t. In comparison to the results in Table 2, we now see that the *Unit-level* and *Penalized* designs can easily match and in some cases improve the performance of the *Unconstrained* design. By fitting each treated unit with a unit-specific synthetic control, the *Unit-level* design can ameliorate interpolation biases induced by the aggregation of $\boldsymbol{X}_j$. The *Penalized* design selects synthetic treated and control units close to $\overline{\boldsymbol{X}}$ in the space of the predictors, which can reduce interpolation biases at the potential cost of lower precision for large values of λ (in which case, the *Penalized* design employs a small number of units in the synthetic treated and synthetic control).

5.2.3. Test size

In this section, we generate the model primitives under the null hypothesis (17). That is, we employ a data generating process such that the values of the common factors and the distributions of the idiosyncratic error variables are unaffected by the intervention.

We report the simulation results in Table 4, which organizes information in the same way as in Table 2. Because the data are generated from the same distribution under treatment and under no treatment, the average treatment effects in Table 4 are close to zero. The same is true for the averages of $\widehat{\tau}_t$ for all designs. Under the null hypothesis (17), the p-value should approximately follow a uniform distribution between zero and one. The results in Table 4 show

Table 3: Average Treatment Effects, Nonlinear Model (Averages over 1000 Simulations)

		τ_t								
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$				
		-13.18	-10.72	-7.96	-5.47	-2.43				
		$\widehat{\tau}_t$					MAE	RMSE	$\widehat{p}$	$\widehat{p} < 0.05$
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$				
Unconstrained		-13.44	-10.92	-8.18	-5.85	-2.78	1.99	2.54	0.059	0.741
Constrained	$\overline{m}=1$	-15.70	-13.18	-10.50	-7.76	-4.78	3.51	4.27	0.061	0.717
	$\overline{m}=2$	-14.27	-11.86	-8.90	-6.44	-3.34	2.64	3.29	0.061	0.725
	$\overline{m}=3$	-13.69	-11.38	-8.38	-5.95	-2.97	2.23	2.83	0.058	0.745
	$\overline{m}=4$	-13.58	-11.09	-8.23	-5.89	-2.75	2.10	2.67	0.058	0.754
	$\overline{m}=5$	-13.37	-10.97	-8.14	-5.79	-2.88	2.05	2.61	0.060	0.747
	$\overline{m}=6$	-13.54	-11.03	-8.31	-5.86	-2.86	2.00	2.56	0.060	0.738
	$\overline{m}=7$	-13.49	-10.94	-8.17	-5.86	-2.78	1.98	2.53	0.058	0.743
Weakly-targeted	$\beta=0.01$	-11.66	-9.02	-6.37	-3.87	-1.00	2.59	3.24	0.116	0.604
	$\beta=0.1$	-12.08	-9.60	-6.87	-4.31	-1.47	2.15	2.74	0.083	0.680
	$\beta=1$	-12.51	-10.13	-7.35	-4.81	-1.91	1.97	2.51	0.057	0.761
	$\beta=10$	-13.03	-10.51	-7.81	-5.25	-2.32	2.19	2.76	0.031	0.854
	$\beta=100$	-13.28	-10.72	-8.00	-5.43	-2.59	2.45	3.11	0.024	0.886
Unit-level	$\xi=0.01$	-11.76	-9.15	-6.51	-3.91	-1.15	2.57	3.22	0.118	0.593
	$\xi=0.1$	-13.11	-10.59	-7.82	-5.15	-2.29	2.06	2.64	0.060	0.754
	$\xi=1$	-13.74	-11.12	-8.42	-5.75	-2.84	2.37	3.02	0.029	0.850
	$\xi=10$	-13.74	-11.20	-8.55	-5.89	-3.09	3.02	3.77	0.028	0.866
	$\xi=100$	-13.79	-11.16	-8.54	-5.90	-3.08	3.20	4.00	0.029	0.863
Penalized	$\lambda=0.01$	-13.40	-10.93	-8.32	-5.82	-2.82	1.97	2.53	0.055	0.759
	$\lambda=0.1$	-13.33	-10.79	-8.13	-5.56	-2.65	2.07	2.65	0.045	0.779
	$\lambda=1$	-13.32	-10.84	-8.15	-5.39	-2.60	3.08	3.84	0.056	0.738
	$\lambda=10$	-13.39	-10.82	-7.95	-5.34	-2.58	3.85	4.80	0.103	0.595
	$\lambda=100$	-13.35	-10.82	-8.00	-5.29	-2.57	4.10	5.11	0.117	0.562

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

good behavior of our testing procedure under the null hypothesis: average p-values and rejection rates are close to 0.5 and 0.05, respectively.

5.2.4. 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 set-up as for Table 2. We consider randomized treatment

Table 4: Average Treatment Effects Under the Null Hypothesis (17) (Averages over 1000 Simulations)

		τ_t								
		$t = 26$	$t = 27$	$t = 28$	$t = 29$	$t = 30$				
		-0.01	0.00	0.00	0.01	-0.01				
		$\widehat{\tau}_t$					MAE	$RMSE$	$\widehat{p}$	$\widehat{p} < 0.05$
		$t = 26$	$t = 27$	$t = 28$	$t = 29$	$t = 30$				
Unconstrained		-0.01	0.00	-0.03	-0.04	-0.07	0.97	1.13	0.495	0.060
Constrained	$\overline{m} = 1$	0.20	0.14	0.02	-0.08	0.05	3.00	3.55	0.495	0.056
	$\overline{m} = 2$	-0.02	-0.01	-0.02	-0.09	-0.03	1.80	2.13	0.497	0.038
	$\overline{m} = 3$	-0.09	-0.07	-0.02	-0.05	-0.02	1.37	1.62	0.505	0.048
	$\overline{m} = 4$	-0.02	-0.02	0.00	-0.01	-0.01	1.19	1.41	0.494	0.054
	$\overline{m} = 5$	0.01	-0.02	0.03	0.00	-0.05	1.07	1.25	0.496	0.057
	$\overline{m} = 6$	0.07	0.06	0.10	-0.01	-0.03	0.99	1.17	0.484	0.054
	$\overline{m} = 7$	-0.01	0.00	-0.02	-0.04	-0.07	0.96	1.13	0.495	0.059
Weakly-targeted	$\beta = 0.01$	0.01	0.03	-0.06	0.00	-0.03	1.27	1.50	0.503	0.042
	$\beta = 0.1$	0.02	-0.01	0.01	0.04	-0.03	1.03	1.21	0.498	0.055
	$\beta = 1$	0.00	0.01	0.03	0.10	0.04	0.95	1.11	0.501	0.044
	$\beta = 10$	-0.07	-0.01	-0.03	0.04	0.00	0.94	1.09	0.485	0.061
	$\beta = 100$	-0.09	-0.08	-0.06	-0.05	-0.04	0.95	1.11	0.493	0.051
Unit-level	$\xi = 0.01$	0.00	0.03	-0.04	-0.04	-0.02	1.05	1.25	0.511	0.053
	$\xi = 0.1$	0.00	0.00	0.03	0.02	0.02	1.05	1.24	0.500	0.049
	$\xi = 1$	0.01	0.02	-0.06	-0.05	-0.03	1.38	1.63	0.499	0.046
	$\xi = 10$	0.18	0.00	-0.02	-0.15	-0.02	1.97	2.33	0.496	0.038
	$\xi = 100$	0.19	-0.03	-0.02	-0.18	-0.03	2.34	2.77	0.502	0.053
Penalized	$\lambda = 0.01$	0.00	0.00	0.02	-0.04	-0.01	1.01	1.18	0.494	0.051
	$\lambda = 0.1$	-0.07	-0.05	-0.07	-0.11	-0.10	1.32	1.56	0.505	0.041
	$\lambda = 1$	0.02	0.07	-0.07	-0.07	0.01	2.17	2.57	0.495	0.045
	$\lambda = 10$	0.16	0.03	-0.11	-0.08	-0.08	3.79	4.48	0.514	0.045
	$\lambda = 100$	0.22	0.15	-0.14	-0.14	-0.08	4.22	5.00	0.515	0.041

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

assignment with $\overline{m}$ treated units. D_j is a treatment indicator that equals one if unit j is randomized into the treated group and zero otherwise. We study the performance of the following 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

Table 5: 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.

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*: Divide the sample in $\overline{m}$ strata, such that each stratum has at least two units. In each stratum, one unit is assigned to treatment at random. The composition of the strata is chosen to minimize the maximal within-strata 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 cluster k. Let J_k be the number of units in stratum k.

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

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 as 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.

Results are reported in Table 5. Across all values of $\overline{m}$, the synthetic control design outperforms randomized assignment, including variants that incorporate pre-stratification, post-stratification, or regression adjustment. Taken together with the findings in Table 1, these results underscore the potential of synthetic controls as a more effective design strategy in experiments involving aggregate units and a limited number of treated 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 County." *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.

Agarwal, Anish, Devavrat Shah, and Dennis Shen. 2021. "Synthetic Interventions." *arXiv e-prints*, 2006.07691.

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.

Bottmer, Lea, Guido Imbens, Jann Spiess, and Merrill Warnick. 2021. "A Design-Based Perspective on Synthetic Control Methods." *arXiv e-prints*, 2101.09398.

Chen, Jiafeng. 2023. "Synthetic Control as Online Linear Regression." *Econometrica*, 91(2): 465–491.

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.

Egami, Naoki, and Diana Da In Lee. 2024. "Designing Multi-Site Studies for External Validity: Site Selection via Synthetic Purposive Sampling." `https://naokiegami.com/paper/sps.pdf`.

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, Max, and Lukas Lehner. 2023. "Employing the Unemployed of Marienthal: Evaluation of a Guaranteed Job Program." `https://maxkasy.github.io/home/files/papers/Jobguarantee_marienthal.pdf`.

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.

Montiel Olea, José Luis, Brenda Prallon, Chen Qiu, Jörg Stoye, and Yiwei Sun. 2024. "Externally Valid Selection of Experimental Sites via the k-Median Problem." *arXiv e-prints*, 2101.09398.

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.

Appendix

A. Proofs

A.1. Proof of Theorem 1

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

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

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

The first term in (A.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

$$
\sum_{j=1}^{J} w_j^* Y_{jt}^I - \sum_{j=1}^{J} f_j Y_{jt}^I = \boldsymbol{\gamma}_t' \left(\sum_{j=1}^{J} w_j^* \mathbf{Z}_j - \sum_{j=1}^{J} f_j \mathbf{Z}_j \right)
$$

$$
+ \boldsymbol{\eta}_t' \left(\sum_{j=1}^{J} w_j^* \boldsymbol{\mu}_j - \sum_{j=1}^{J} f_j \boldsymbol{\mu}_j \right) + \left(\sum_{j=1}^{J} w_j^* \xi_{jt} - \sum_{j=1}^{J} f_j \xi_{jt} \right) \tag{A.2}
$$

Similarly, using expression (12a), we obtain

$$
\sum_{j=1}^{J} w_j^* \mathbf{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} f_j \mathbf{Y}_j^{\mathcal{E}} = \boldsymbol{\theta}_{\mathcal{E}} \left(\sum_{j=1}^{J} w_j^* \mathbf{Z}_j - \sum_{j=1}^{J} f_j \mathbf{Z}_j \right)
$$

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

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

$$
\boldsymbol{\eta}_t' (\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \left(\sum_{j=1}^{J} w_j^* \mathbf{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} f_j \mathbf{Y}_j^{\mathcal{E}} \right)
$$

$$
= \boldsymbol{\eta}_t' (\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\theta}_{\mathcal{E}} \left(\sum_{j=1}^{J} w_j^* \mathbf{Z}_j - \sum_{j=1}^{J} f_j \mathbf{Z}_j \right)
$$

$$+ \boldsymbol{\eta}_t' \left(\sum_{j=1}^{J} w_j^* \boldsymbol{\mu}_j - \sum_{j=1}^{J} f_j \boldsymbol{\mu}_j \right)$$

$$+ \boldsymbol{\eta}_t' (\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \left(\sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} - \sum_{j=1}^{J} f_j \boldsymbol{\epsilon}_j^{\mathcal{E}} \right). \tag{A.3}$$

Equations (A.2) and (A.3) imply

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

$$+ \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \left(\sum_{j=1}^{J} w_j^* \boldsymbol{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} f_j \boldsymbol{Y}_j^{\mathcal{E}} \right)$$

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

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

$$+ \left(\sum_{j=1}^{J} w_j^* \xi_{jt} - \sum_{j=1}^{J} f_j \xi_{jt} \right). \tag{A.4}$$

If Assumption 3 holds, (A.4) becomes

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

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

$$+ \left(\sum_{j=1}^{J} w_j^* \xi_{jt} - \sum_{j=1}^{J} f_j \xi_{jt} \right). \tag{A.5}$$

Only the first term on the right-hand side of (A.5) has a non-zero mean (because the weights, w_j^*, depend on the error terms $\boldsymbol{\epsilon}_j^{\mathcal{E}}$). Therefore,

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

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

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

39

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

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

Similarly,

$$\left(\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_s\right)^2 \leq \left(\frac{\overline{\lambda}^2 F}{T_{\mathcal{E}}\underline{\zeta}}\right)^2. \tag{A.6}$$

Let

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

Because $\bar{\epsilon}_{jt}^{\mathcal{E}}$ is a linear combination of independent sub-Gaussians with variance proxy $\overline{\sigma}^2$, it follows that $\bar{\epsilon}_{jt}^{\mathcal{E}}$ is sub-Gaussian with variance proxy $(\overline{\eta}\,\overline{\lambda}F/\underline{\zeta})^2\overline{\sigma}^2/T_{\mathcal{E}}$. Let $\mathcal{S} = \{\boldsymbol{w} \in \mathbb{R}^J : \sum_{j=1}^J w_j = 1\}$. Theorem 1.16 from Rigollet and Hütter (2019) implies

$$\left|E\left[\sum_{j=1}^J w_j^* Y_{jt}^I - \sum_{j=1}^J f_j Y_{jt}^I\right]\right|$$
$$= \left|E\left[\sum_{j=1}^J w_j^* \bar{\epsilon}_{jt}^{\mathcal{E}}\right]\right| \leq E\left[\max_{\boldsymbol{w} \in \mathcal{S}} \left|\sum_{j=1}^J w_j \bar{\epsilon}_{jt}^{\mathcal{E}}\right|\right] \leq \frac{\overline{\eta}\,\overline{\lambda}F}{\underline{\zeta}}\sqrt{2\log(2J)}\frac{\overline{\sigma}}{\sqrt{T_{\mathcal{E}}}}.$$

An analogous argument yields

$$\left|E\left[\sum_{j=1}^J v_j^* Y_{jt}^N - \sum_{j=1}^J f_j Y_{jt}^N\right]\right| \leq \frac{\overline{\lambda}^2 F}{\underline{\zeta}}\sqrt{2\log(2J)}\frac{\overline{\sigma}}{\sqrt{T_{\mathcal{E}}}},$$

which completes the proof of the theorem.

Suppose now Assumption 4 holds (but Assumption 3 does not). To obtain a bound on the bias, we bound the first two terms in (A.4). Recall that

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

Therefore, the absolute value of each element in vector $(\boldsymbol{\gamma}'_t - \boldsymbol{\eta}'_t(\boldsymbol{\lambda}'_\mathcal{E}\boldsymbol{\lambda}_\mathcal{E})^{-1}\boldsymbol{\lambda}'_\mathcal{E}\boldsymbol{\theta}_\mathcal{E})$ is bounded by $\overline{\gamma} + \overline{\theta}\dfrac{\overline{\lambda}\overline{\eta}F}{\underline{\zeta}}$. Cauchy–Schwarz inequality and Assumption 4 imply

$$
\left| (\boldsymbol{\gamma}'_t - \boldsymbol{\eta}'_t(\boldsymbol{\lambda}'_\mathcal{E}\boldsymbol{\lambda}_\mathcal{E})^{-1}\boldsymbol{\lambda}'_\mathcal{E}\boldsymbol{\theta}_\mathcal{E}) \left(\sum_{j=1}^{J} w^*_j \boldsymbol{Z}_j - \sum_{j=1}^{J} f_j \boldsymbol{Z}_j \right) \right|
$$
$$
\leq \left(\overline{\gamma} + \overline{\theta}\dfrac{\overline{\lambda}\overline{\eta}F}{\underline{\zeta}} \right) \sqrt{R} \left\| \sum_{j=1}^{J} w^*_j \boldsymbol{Z}_j - \sum_{j=1}^{J} f_j \boldsymbol{Z}_j \right\|_2
$$
$$
\leq \left(\overline{\gamma} + \overline{\theta}\dfrac{\overline{\lambda}\overline{\eta}F}{\underline{\zeta}} \right) Rd,
$$

and

$$
\left| \boldsymbol{\eta}'_t(\boldsymbol{\lambda}'_\mathcal{E}\boldsymbol{\lambda}_\mathcal{E})^{-1}\boldsymbol{\lambda}'_\mathcal{E} \left(\sum_{j=1}^{J} w^*_j \boldsymbol{Y}^\mathcal{E}_j - \sum_{j=1}^{J} f_j \boldsymbol{Y}^\mathcal{E}_j \right) \right| \leq \dfrac{\overline{\lambda}\overline{\eta}F}{\underline{\zeta}}d.
$$

Combining the last two displayed equations with (A.4), we have

$$
\left| E\left[\sum_{j=1}^{J} w^*_j Y^I_{jt} - \sum_{j=1}^{J} f_j Y^I_{jt} \right] \right| \leq \left(\overline{\gamma}R + \dfrac{\overline{\lambda}\overline{\eta}F}{\underline{\zeta}}(1 + \overline{\theta}R) \right)d + \dfrac{\overline{\lambda}\overline{\eta}F}{\underline{\zeta}}\sqrt{2\log(2J)}\dfrac{\overline{\sigma}}{\sqrt{T_\mathcal{E}}}.
$$

An analogous derivation produces

$$
\left| E\left[\sum_{j=1}^{J} v^*_j Y^N_{jt} - \sum_{j=1}^{J} f_j Y^N_{jt} \right] \right| \leq \left(\overline{\theta}R + \dfrac{\overline{\lambda}^2 F}{\underline{\zeta}}(1 + \overline{\theta}R) \right)d + \dfrac{\overline{\lambda}^2 F}{\underline{\zeta}}\sqrt{2\log(2J)}\dfrac{\overline{\sigma}}{\sqrt{T_\mathcal{E}}},
$$

which finishes the proof of the theorem. ∎

A.2. Proof of Theorem 2

Proof of Theorem 2. Recall that

$$
\widehat{u}_t = \sum_{j=1}^{J} w^*_j Y_{jt} - \sum_{j=1}^{J} v^*_j Y_{jt},
$$

for $t \in \mathcal{B} \cup \{T_0 + 1, \ldots, T\}$. For $t \in \{T_0 + 1, \ldots, T\}$, $\widehat{u}_t$ 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. Let

$$
u_t = \sum_{j=1}^{J} w^*_j \epsilon_{jt} - \sum_{j=1}^{J} v^*_j \epsilon_{jt}
$$

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

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

for $t \in \{T_0 + 1, \ldots, T\}$. The null hypothesis (17) and the assumptions of Theorem 2 imply that $\{u_t\}_{t \in \mathcal{B} \cup \{T_0+1,\ldots,T\}}$ is a sequence of exchangeable random variables. Additionally, Assumption 1 and the null hypothesis (17) imply

$$\widehat{u}_t = \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} = \boldsymbol{\theta}_t' \sum_{j=1}^{J} (w_j^* - v_j^*) \boldsymbol{Z}_j + \boldsymbol{\lambda}_t' \sum_{j=1}^{J} (w_j^* - v_j^*) \boldsymbol{\mu}_j + u_t,$$

for $t \in \mathcal{B} \cup \{T_0 + 1, \ldots, T\}$. The result of the theorem then follows from Theorem D.1 in Chernozhukov, Wüthrich and Zhu (2021). ∎

A.3. Proof of Theorem 3

A.3.1. A Technical Lemma

We first define the following quantity and present a technical lemma. Let $\boldsymbol{\epsilon}_* = (\epsilon_{1*}, \epsilon_{2*}, \ldots, \epsilon_{J*})$ be an i.i.d. copy of $(\epsilon_{1t}, \epsilon_{2t}, \ldots, \epsilon_{Jt})$ the idiosyncratic noises. Using the definition of $\boldsymbol{\epsilon}_*$ and conditioning on the weights $(\boldsymbol{w}^*, \boldsymbol{v}^*)$, we define, for any $q \in \mathbb{R}$,

$$P_{\mathcal{E},q} = \Pr\left(\left| \sum_{j=1}^{J} w_j^* \epsilon_{j*} - \sum_{j=1}^{J} v_j^* \epsilon_{j*} \right| \le q \right).$$

Lemma A.1 *Assume there exist parameters $\epsilon_\mathcal{B}$ and $\epsilon_\mathcal{T}$, as well as events $\mathcal{C}_\mathcal{B}$ and $\mathcal{C}_\mathcal{T}$, such that the following two conditions hold:*

1. *There exists a high probability event $\mathcal{C}_\mathcal{B}$ such that conditional on this event, for any weights $(\boldsymbol{w}^*, \boldsymbol{v}^*)$ and any $q \in \mathbb{R}$,*

$$\left| \frac{1}{T_0 - T_\mathcal{E}} \sum_{t \in \mathcal{B}} \mathbb{1}\left\{ \left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| \le q \right\} - P_{\mathcal{E},q} \right| \le \epsilon_\mathcal{B}. \tag{A.7}$$

2. *Recall that $\tau_t = \sum_{j=1}^{J} f_j(Y_{jt}^I - Y_{jt}^N)$. There exists a high probability event $\mathcal{C}_\mathcal{T}$ such that conditional on this event, for any weights $(\boldsymbol{w}^*, \boldsymbol{v}^*)$, any $q \in \mathbb{R}$, and any $t \in \{T_0 + 1, \ldots, T\}$,*

$$\left| \Pr\left(\left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} - \tau_t \right| \le q \right) - P_{\mathcal{E},q} \right| \le \epsilon_\mathcal{T}. \tag{A.8}$$

Assume that the joint event $\mathcal{C}_\mathcal{B} \cap \mathcal{C}_\mathcal{T}$ happens with probability at least $1 - \delta_\mathcal{B}(\epsilon_\mathcal{B}) - \delta_\mathcal{T}(\epsilon_\mathcal{T})$, where we use $\delta_\mathcal{B}(\epsilon_\mathcal{B})$ and $\delta_\mathcal{T}(\epsilon_\mathcal{T})$ to stand for two quantities that each depends on $\epsilon_\mathcal{B}$ and $\epsilon_\mathcal{T}$, respectively. In addition, assume that $\left| \sum_{j=1}^J w_j^* \epsilon_{j*} - \sum_{j=1}^J v_j^* \epsilon_{j*} \right|$ has a continuous distribution. Then, for any $\alpha \in (0,1)$ and any $t \in \{T_0 + 1, ..., T\}$,

$$\left| \Pr\left(\left| \sum_{j=1}^J w_j^* Y_{jt} - \sum_{j=1}^J v_j^* Y_{jt} - \tau_t \right| \le \widehat{q}_{1-\alpha} \right) - (1-\alpha) \right| \le \epsilon_\mathcal{B} + \epsilon_\mathcal{T} + \delta_\mathcal{B}(\epsilon_\mathcal{B}) + \delta_\mathcal{T}(\epsilon_\mathcal{T}).$$

Note that Lemma A.1 does not require Assumptions 1–3. But for Conditions (A.7) and (A.8) to hold, we will apply Assumptions 1–3. To prove Lemma A.1, we borrow the proof techniques from Oliveira et al. (2022). We first define the following quantile on the probability distribution (instead of the empirical distribution),

$$q_{1-\alpha} = \inf_{z \in \mathbb{R}} \left\{ \Pr\left(\left| \sum_{j=1}^J w_j^* \epsilon_{j*} - \sum_{j=1}^J v_j^* \epsilon_{j*} \right| \le z \right) \ge 1 - \alpha \right\} \tag{A.9}$$

Intuitively, $\widehat{q}_{1-\alpha}$ as defined in (20) approximates $q_{1-\alpha}$ as defined in (A.9).

Proof of Lemma A.1. This proof proceeds in two parts.

Part 1: Consider the event

$$\mathcal{E}_1 = \left\{ \widehat{q}_{1-\alpha} \ge q_{1-\alpha-\epsilon_\mathcal{B}} \right\}.$$

We aim to show that event $\mathcal{E}_1$ occurs given event $\mathcal{C}_\mathcal{B}$. For any positive integer $k \in \mathbb{N}$, we can use Condition (A.7) to show that conditional on event $\mathcal{C}_\mathcal{B}$,

$$\frac{1}{T_0 - T_\mathcal{E}} \sum_{t \in \mathcal{B}} \mathbb{1}\left\{ \left| \sum_{j=1}^J w_j^* Y_{jt} - \sum_{j=1}^J v_j^* Y_{jt} \right| \le q_{1-\alpha-\epsilon_\mathcal{B}} - \frac{1}{k} \right\}$$

$$\le \Pr\left(\left| \sum_{j=1}^J w_j^* \epsilon_{j*} - \sum_{j=1}^J v_j^* \epsilon_{j*} \right| \le q_{1-\alpha-\epsilon_\mathcal{B}} - \frac{1}{k} \right) + \epsilon_\mathcal{B}$$

$$< 1 - \alpha - \epsilon_\mathcal{B} + \epsilon_\mathcal{B}$$

$$= 1 - \alpha$$

$$\le \frac{1}{T_0 - T_\mathcal{E}} \sum_{t \in \mathcal{B}} \mathbb{1}\left\{ \left| \sum_{j=1}^J w_j^* Y_{jt} - \sum_{j=1}^J v_j^* Y_{jt} \right| \le \widehat{q}_{1-\alpha} \right\},$$

43

where the first inequality is due to Condition (A.7); the second inequality is due to the infimum part of (A.9) (because $q_{1-\alpha-\epsilon_{\mathcal{B}}} - \frac{1}{k} < q_{1-\alpha-\epsilon_{\mathcal{B}}}$ which is the infimum value such that the probability in (A.9) is greater or equal to $1 - \alpha$); the last inequality is due to the definition of $\widehat{q}_{1-\alpha}$ in (20). The above inequality suggests that for any $k \in \mathbb{N}$, the event

$$\mathcal{E}_k^{(\leq)} = \left\{ \frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\left\{ \left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| \leq q_{1-\alpha-\epsilon_{\mathcal{B}}} - \frac{1}{k} \right\} \right.$$

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

happens conditional on event $\mathcal{C}_{\mathcal{B}}$. Since the left hand side of the inequality inside event $\mathcal{E}_k^{(\leq)}$, which is $\frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\left\{ \left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| \leq q_{1-\alpha-\epsilon_{\mathcal{B}}} - \frac{1}{k} \right\}$, is increasing in k, so the probability $\mathcal{E}_k^{(\leq)}$ decreases in k. Given that the lower bound of $\Pr(\mathcal{E}_k^{(\leq)})$ exists, the limit of $\lim_{k \to +\infty} \Pr(\mathcal{E}_k^{(\leq)})$ exists, i.e.,

$$1 - \delta_{\mathcal{B}} \leq \lim_{k \to +\infty} \Pr(\mathcal{E}_k^{(\leq)}) = \Pr(\mathcal{E}_\infty^{(\leq)}),$$

where we use $\Pr(\mathcal{E}_\infty^{(\leq)})$ to stand for the limiting event

$$\mathcal{E}_\infty^{(\leq)} = \left\{ \frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\left\{ \left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| \leq q_{1-\alpha-\epsilon_{\mathcal{B}}} \right\} \right.$$

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

This means that, event $\mathcal{E}_1 = \{\widehat{q}_{1-\alpha} \geq q_{1-\alpha-\epsilon_{\mathcal{B}}}\}$ happens conditional on event $\mathcal{C}_{\mathcal{B}}$. Due to the assumption of Lemma A.1, event $\mathcal{C}_{\mathcal{T}} \cap \mathcal{E}_1$ happens with probability at least $1 - \delta_{\mathcal{B}}(\epsilon_{\mathcal{B}}) - \delta_{\mathcal{T}}(\epsilon_{\mathcal{T}})$. Next we have, for any $t \in \{T_0 + 1, ..., T\}$ in the experimental periods,

$$\Pr\left(\left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} - \tau_t \right| \leq \widehat{q}_{1-\alpha} \right)$$

$$\geq \Pr\left(\left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} - \tau_t \right| \leq \widehat{q}_{1-\alpha} \bigcap (\mathcal{C}_{\mathcal{T}} \cap \mathcal{E}_1) \right) - \delta_{\mathcal{B}}(\epsilon_{\mathcal{B}}) - \delta_{\mathcal{T}}(\epsilon_{\mathcal{T}})$$

$$\geq \Pr\left(\left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} - \tau_t \right| \leq q_{1-\alpha-\epsilon_{\mathcal{B}}} \right) - \delta_{\mathcal{B}}(\epsilon_{\mathcal{B}}) - \delta_{\mathcal{T}}(\epsilon_{\mathcal{T}})$$

44

$$\geq \Pr\left(\left|\sum_{j=1}^{J} w_j^* \epsilon_{j*} - \sum_{j=1}^{J} v_j^* \epsilon_{j*}\right| \leq q_{1-\alpha-\epsilon_{\mathcal{B}}}\right) - \epsilon_{\mathcal{T}} - \delta_{\mathcal{B}}(\epsilon_{\mathcal{B}}) - \delta_{\mathcal{T}}(\epsilon_{\mathcal{T}})$$

$$\geq 1 - \alpha - \epsilon_{\mathcal{B}} - \epsilon_{\mathcal{T}} - \delta_{\mathcal{B}}(\epsilon_{\mathcal{B}}) - \delta_{\mathcal{T}}(\epsilon_{\mathcal{T}}).$$

where the second inequality is because the probability decreases if we decrease from $\widehat{q}_{1-\alpha}$ to $q_{1-\alpha-\epsilon_{\mathcal{B}}}$; the third inequality is due to Condition (A.8); the last inequality is due to the definition of $q_{1-\alpha-\epsilon_{\mathcal{B}}}$ in (A.9).

Part 2: Consider the event

$$\mathcal{E}_2 = \left\{\widehat{q}_{1-\alpha} \leq q_{1-\alpha+\epsilon_{\mathcal{B}}}\right\}.$$

We wish to show that event $\mathcal{E}_1$ happens conditional on event $\mathcal{C}_{\mathcal{B}}$. We use Condition (A.8) to show that conditional on event $\mathcal{C}_{\mathcal{B}}$,

$$\frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\left\{\left|\sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt}\right| \leq q_{1-\alpha+\epsilon_{\mathcal{B}}}\right\}$$

$$\geq \Pr\left(\left|\sum_{j=1}^{J} w_j^* \epsilon_{j*} - \sum_{j=1}^{J} v_j^* \epsilon_{j*}\right| \leq q_{1-\alpha+\epsilon_{\mathcal{B}}}\right) - \epsilon_{\mathcal{B}}$$

$$\geq 1 - \alpha + \epsilon_{\mathcal{B}} - \epsilon_{\mathcal{B}}$$

$$= 1 - \alpha,$$

where the first inequality is due to Condition (A.7); the second inequality is due to the definition of $q_{1-\alpha+\epsilon_{\mathcal{B}}}$ in (A.9);

Due to (20), since $\widehat{q}_{1-\alpha}$ is the smallest value satisfying this condition, we have that event $\mathcal{E}_2 = \{\widehat{q}_{1-\alpha} \leq q_{1-\alpha+\epsilon_{\mathcal{B}}}\}$ happens conditional on event $\mathcal{C}_{\mathcal{B}}$. Due to the assumption of Lemma A.1, event $\mathcal{C}_{\mathcal{T}} \cap \mathcal{E}_2$ happens with probability at least $1 - \delta_{\mathcal{B}}(\epsilon_{\mathcal{B}}) - \delta_{\mathcal{T}}(\epsilon_{\mathcal{T}})$.

Then, for any $t \in \{T_0 + 1, ..., T\}$ in the experimental periods,

$$\Pr\left(\left|\sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} - \tau_t\right| \leq \widehat{q}_{1-\alpha}\right)$$

$$\leq \Pr\left(\left|\sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} - \tau_t\right| \leq \widehat{q}_{1-\alpha}\bigcap(\mathcal{C}_{\mathcal{T}} \cap \mathcal{E}_2)\right) + \delta_{\mathcal{B}}(\epsilon_{\mathcal{B}}) + \delta_{\mathcal{T}}(\epsilon_{\mathcal{T}})$$

$$\leq \Pr\left(\left|\sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} - \tau_t\right| \leq q_{1-\alpha+\epsilon_{\mathcal{B}}}\right) + \delta_{\mathcal{B}}(\epsilon_{\mathcal{B}}) + \delta_{\mathcal{T}}(\epsilon_{\mathcal{T}})$$

$$\leq \Pr\left(\left|\sum_{j=1}^{J} w_j^* \epsilon_{j*} - \sum_{j=1}^{J} v_j^* \epsilon_{j*}\right| \leq q_{1-\alpha+\epsilon_{\mathcal{B}}}\right) + \epsilon_{\mathcal{T}} + \delta_{\mathcal{B}}(\epsilon_{\mathcal{B}}) + \delta_{\mathcal{T}}(\epsilon_{\mathcal{T}})$$

$$\leq 1 - \alpha + \epsilon_{\mathcal{B}} + \epsilon_{\mathcal{T}} + \delta_{\mathcal{B}}(\epsilon_{\mathcal{B}}) + \delta_{\mathcal{T}}(\epsilon_{\mathcal{T}}).$$

where the second inequality is because the probability increases if we increase from $\widehat{q}_{1-\alpha}$ to $q_{1-\alpha+\epsilon_{\mathcal{B}}}$; the third inequality is due to Condition (A.8); the last inequality is due to the definition of $q_{1-\alpha+\epsilon_{\mathcal{B}}}$ in (A.9). ∎

A.3.2. Proof of Theorem 3

In this section, we use Lemma A.1 to prove Theorem 3. Instead of proving exactly Theorem 3, we prove Theorem A.2 below with all the constants provided. Then, setting $z_{\mathcal{B}} = (T_0 - T_{\mathcal{E}})^{-\frac{1}{2}}$ and $z_{\mathcal{E}} = T_{\mathcal{E}}^{-\frac{1}{2}}$ we prove Theorem 3.

Theorem A.2 *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 (21) approximately achieves point-wise coverage, i.e., for any $\alpha \in (0, 1)$ and any $t \in \{T_0 + 1, ..., T\}$,*

$$\left|\Pr\left(\tau_t \in \widehat{C}_{1-\alpha}(Y_{1t}, Y_{2t}, ..., Y_{Jt})\right) - (1 - \alpha)\right|$$

$$\leq \sqrt{\frac{1}{2(T_0 - T_{\mathcal{E}})} \log\left(\frac{2}{z_{\mathcal{B}}}\right)} + \kappa\sqrt{\frac{8eJ\bar{\sigma}^2\bar{\lambda}^2\bar{\eta}^2 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}}\right)} + 2\kappa\sqrt{\frac{8eJ\bar{\sigma}^2\bar{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}}\right)} + z_{\mathcal{B}} + 3z_{\mathcal{E}}.$$

where $z_{\mathcal{B}}$ and $z_{\mathcal{E}}$ are arbitrary positive constants.

Proof of Theorem A.2. We outline the proof of Theorem A.2 as follows. We first define four events. We then check Conditions (A.7) under the first two events and (A.8) under the last two events. Finally, we apply Lemma A.1 and conclude the proof.

Step 1: We define the following four events. First, in the blank periods and conditioning on the weights $(\boldsymbol{w}^*, \boldsymbol{v}^*)$ that we obtain from the fitting periods,

$$\mathcal{E}_1 = \left\{ \left| \frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\left\{ \left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| \le q \right\} \right. \right.$$
$$\left. \left. - \frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \Pr\left(\left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| \le q \right) \right| \le \sqrt{\frac{1}{2(T_0 - T_{\mathcal{E}})} \log\left(\frac{2}{z_{\mathcal{B}}}\right)} \right\}.$$

Due to Hoeffding inequality for bounded random variables (conditioning on the weights $(\boldsymbol{w}^*, \boldsymbol{v}^*)$, these indicators are independent), we have that event $\mathcal{E}_1$ happens with probability $1 - z_{\mathcal{B}}$.

Second, in the blank periods,

$$\mathcal{E}_2 = \left\{ \forall t \in \mathcal{B}, \left| \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J}(w_j^* - v_j^*)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right| \le \sqrt{\frac{8\overline{\sigma}^2\overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}}\right)} \right\}$$

$$= \left\{ \max_{t \in \mathcal{B}} \left| \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J}(w_j^* - v_j^*)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right| \le \sqrt{\frac{8\overline{\sigma}^2\overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}}\right)} \right\}.$$

Note that,

$$\max_{t \in \mathcal{B}} \left| \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J}(w_j^* - v_j^*)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right|$$

$$\le \max_{t \in \mathcal{B}} \sum_{j=1}^{J} |w_j^* - v_j^*| \sum_{s \in \mathcal{E}} |\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_s||\epsilon_{js}|$$

$$\le \sum_{j=1}^{J} |w_j^* - v_j^*| \sum_{s \in \mathcal{E}} \frac{\overline{\lambda}^2 F}{T_{\mathcal{E}}\underline{\zeta}} |\epsilon_{js}|,$$

where the second inequality is due to (A.6), and because $|w_j^* - v_j^*| \ge 0$ and $|\epsilon_{js}| \ge 0$. Therefore,

$$\Pr(\mathcal{E}_2) \ge 1 - \Pr\left(\sum_{j=1}^{J} \frac{|w_j^* - v_j^*|}{2} \sum_{s \in \mathcal{E}} \frac{\overline{\lambda}^2 F}{T_{\mathcal{E}}\underline{\zeta}} |\epsilon_{js}| > \sqrt{\frac{2\overline{\sigma}^2\overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}}\right)} \right)$$

$$\ge 1 - \sum_{j=1}^{J} \Pr\left(\sum_{s \in \mathcal{E}} \frac{\overline{\lambda}^2 F}{T_{\mathcal{E}}\underline{\zeta}} |\epsilon_{js}| > \sqrt{\frac{2\overline{\sigma}^2\overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}}\right)} \right)$$

$$\ge 1 - z_{\mathcal{E}},$$

where the second inequality follows from union bound, and the third inequality is the Chernoff bound for sub-Gaussian random variables.

47

Third, in the experimental periods,

$$
\mathcal{E}_3 = \left\{ \forall t \in \{T_0+1,\ldots,T\}, \left| \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J}(w_j^* - f_j)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right| \le \sqrt{\frac{8\overline{\sigma}^2\overline{\lambda}^2\overline{\eta}^2 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}}\right)} \right\}
$$

$$
= \left\{ \max_{t\in\{T_0+1,\ldots,T\}} \left| \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J}(w_j^* - f_j)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right| \le \sqrt{\frac{8\overline{\sigma}^2\overline{\lambda}^2\overline{\eta}^2 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}}\right)} \right\}.
$$

Note that,

$$
\max_{t\in\{T_0+1,\ldots,T\}} \left| \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J}(w_j^* - f_j)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right|
$$

$$
\le \max_{t\in\{T_0+1,\ldots,T\}} \sum_{j=1}^{J}|w_j^* - f_j| \sum_{s\in\mathcal{E}} |\boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_s||\epsilon_{js}|
$$

$$
\le \sum_{j=1}^{J}|w_j^* - f_j| \sum_{s\in\mathcal{E}} \frac{\overline{\lambda}\overline{\eta}F}{T_{\mathcal{E}}\underline{\zeta}} |\epsilon_{js}|,
$$

where the second inequality is due to (A.6), and because $|w_j^* - f_j| \ge 0$ and $|\epsilon_{js}| \ge 0$. Therefore,

$$
\Pr(\mathcal{E}_3) \ge 1 - \Pr\left(\sum_{j=1}^{J} \frac{|w_j^* - f_j|}{2} \sum_{s\in\mathcal{E}} \frac{\overline{\lambda}\overline{\eta}F}{T_{\mathcal{E}}\underline{\zeta}} |\epsilon_{js}| > \sqrt{\frac{2\overline{\sigma}^2\overline{\lambda}^2\overline{\eta}^2 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}}\right)} \right)
$$

$$
\ge 1 - \sum_{j=1}^{J} \Pr\left(\sum_{s\in\mathcal{E}} \frac{\overline{\lambda}\overline{\eta}F}{T_{\mathcal{E}}\underline{\zeta}} |\epsilon_{js}| > \sqrt{\frac{2\overline{\sigma}^2\overline{\lambda}^2\overline{\eta}^2 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}}\right)} \right)
$$

$$
\ge 1 - z_{\mathcal{E}},
$$

where the second inequality follows from union bound, and the third inequality is the Chernoff bound for sub-Gaussian random variables.

Fourth, in the experimental periods,

$$
\mathcal{E}_4 = \left\{ \forall t \in \{T_0+1,\ldots,T\}, \left| \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J}(v_j^* - f_j)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right| \le \sqrt{\frac{8\overline{\sigma}^2\overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}}\right)} \right\}.
$$

Similar to the event $\mathcal{E}_3$, we can show that $\Pr(\mathcal{E}_4) \ge 1 - z_{\mathcal{E}}$.

Step 2: Now we check Conditions (A.7) and (A.8). We first check Condition (A.7). In the statement of Condition (A.7), let $\mathcal{C}_{\mathcal{B}} = \mathcal{E}_1 \cap \mathcal{E}_2$. Note that

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

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

Assumption 3 implies

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

For $t \in \mathcal{B}$, we have

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

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

Conditional on event $\mathcal{E}_2$, we have for any $t \in \mathcal{B}$ in the blank periods,

$$\left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| = \left| \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} - \boldsymbol{\lambda}_t' (\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} (w_j^* - v_j^*) \boldsymbol{\epsilon}_j^{\mathcal{E}} \right|$$

$$\leq \left| \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \right| + \left| \boldsymbol{\lambda}_t' (\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} (w_j^* - v_j^*) \boldsymbol{\epsilon}_j^{\mathcal{E}} \right|$$

$$\leq \left| \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \right| + \sqrt{\frac{8 \bar{\sigma}^2 \overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log \left(\frac{2J}{z_{\mathcal{E}}} \right)}.$$

From the above inequality, for any $t \in \mathcal{B}$, due to Lemma OA.3-1 and Lemma OA.4, the probability density of $\left| \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \right|$ is upper bounded by $\kappa \sqrt{eJ}$, where $e \approx 2.718$ is the base of the natural logarithm. This implies that, conditional on event $\mathcal{E}_2$, for any $t \in \mathcal{B}$ and any $q \in \mathbb{R}$,

$$\left| \Pr \left(\left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| \leq q \right) - \Pr \left(\left| \sum_{j=1}^{J} w_j^* \epsilon_{j*} - \sum_{j=1}^{J} v_j^* \epsilon_{j*} \right| \leq q \right) \right|$$

$$= \left| \Pr \left(\left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| \leq q \right) - \Pr \left(\left| \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \right| \leq q \right) \right|$$

$$\leq \kappa \sqrt{\frac{8eJ \bar{\sigma}^2 \overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log \left(\frac{2J}{z_{\mathcal{E}}} \right)}.$$

This means that, conditional on event $\mathcal{E}_2$, for any $t \in \mathcal{B}$ and any $q \in \mathbb{R}$,

$$\left| \frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \Pr\left(\left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| \le q \right) - \Pr\left(\left| \sum_{j=1}^{J} w_j^* \epsilon_{j*} - \sum_{j=1}^{J} v_j^* \epsilon_{j*} \right| \le q \right) \right|$$

$$\le \kappa \sqrt{\frac{8e J \bar{\sigma}^2 \overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}} \right)}.$$

To conclude checking Condition (A.7), we see that conditional on event $\mathcal{C}_{\mathcal{B}} = \mathcal{E}_1 \cap \mathcal{E}_2$, for any weights $(\boldsymbol{w}^*, \boldsymbol{v}^*)$ and $q \in \mathbb{R}$,

$$\left| \frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\left\{ \left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \right| \le q \right\} - \Pr\left(\left| \sum_{j=1}^{J} w_j^* \epsilon_{j*} - \sum_{j=1}^{J} v_j^* \epsilon_{j*} \right| \le q \right) \right|$$

$$\le \sqrt{\frac{1}{2(T_0 - T_{\mathcal{E}})} \log\left(\frac{2}{z_{\mathcal{B}}} \right)} + \kappa \sqrt{\frac{8e J \bar{\sigma}^2 \overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}} \right)}.$$

We then move on to check Condition (A.8). In the statement of Condition (A.8), let $\mathcal{C}_{\mathcal{T}} = \mathcal{E}_3 \cap \mathcal{E}_4$. For any $t \in \{T_0 + 1, ..., T\}$ in the experimental periods, we have

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

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

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

where the third equality is using Assumption 3 and using the assumption that ξ_{jt} has the same distribution as ϵ_{jt} for $t = T_0 + 1, \ldots, T$, and $j = 1, \ldots, J$.

Conditional on event $\mathcal{E}_3 \cap \mathcal{E}_4$, we have for any $t \in \{T_0 + 1, ..., T\}$ in the experimental periods,

$$\left| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} - \tau_t \right|$$

$$\le \left| \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \right| + \sqrt{\frac{8 \bar{\sigma}^2 \overline{\lambda}^2 \overline{\eta}^2 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}} \right)} + \sqrt{\frac{8 \bar{\sigma}^2 \overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log\left(\frac{2J}{z_{\mathcal{E}}} \right)}.$$

50

Following the same argument, we see that conditional on event $\mathcal{C}_\mathcal{T} = \mathcal{E}_3 \cap \mathcal{E}_4$, for any $t \in \{T_0 + 1, ..., T\}$ and any $q \in \mathbb{R}$,

$$\left| \Pr\left(\left| \sum_{j=1}^J w_j^* Y_{jt} - \sum_{j=1}^J v_j^* Y_{jt} - \tau_t \right| \le q \right) - \Pr\left(\left| \sum_{j=1}^J w_j^* \epsilon_{j*} - \sum_{j=1}^J v_j^* \epsilon_{j*} \right| \le q \right) \right|$$

$$\le \kappa \sqrt{\frac{8eJ\overline{\sigma}^2 \overline{\lambda}^2 \overline{\eta}^2 F^2}{\underline{\varsigma}^2 T_\mathcal{E}} \log\left(\frac{2J}{z_\mathcal{E}}\right)} + \kappa \sqrt{\frac{8eJ\overline{\sigma}^2 \overline{\lambda}^4 F^2}{\underline{\varsigma}^2 T_\mathcal{E}} \log\left(\frac{2J}{z_\mathcal{E}}\right)}.$$

Step 3: Now we apply Lemma A.1. Note that, the joint event $\mathcal{C}_\mathcal{B} \cap \mathcal{C}_\mathcal{T} = \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \cap \mathcal{E}_4$ happens with probability at least $1 - z_\mathcal{B} - 3z_\mathcal{E}$. Due to Lemma A.1,

$$\left| \Pr\left(\left| \sum_{j=1}^J w_j^* Y_{jt} - \sum_{j=1}^J v_j^* Y_{jt} - \tau_t \right| \le \widehat{q}_{1-\alpha} \right) - (1-\alpha) \right|$$

$$\le \sqrt{\frac{1}{2(T_0 - T_\mathcal{E})} \log\left(\frac{2}{z_\mathcal{B}}\right)} + \kappa \sqrt{\frac{8eJ\overline{\sigma}^2 \overline{\lambda}^2 \overline{\eta}^2 F^2}{\underline{\varsigma}^2 T_\mathcal{E}} \log\left(\frac{2J}{z_\mathcal{E}}\right)} + 2\kappa \sqrt{\frac{8eJ\overline{\sigma}^2 \overline{\lambda}^4 F^2}{\underline{\varsigma}^2 T_\mathcal{E}} \log\left(\frac{2J}{z_\mathcal{E}}\right)} + z_\mathcal{B} + 3z_\mathcal{E}.$$

Because $\left| \sum_{j=1}^J w_j^* Y_{jt} - \sum_{j=1}^J v_j^* Y_{jt} - \tau_t \right| \le \widehat{q}_{1-\alpha}$ is equivalent to $\tau_t \in \widehat{C}_{1-\alpha}(Y_{1t}, Y_{2t}, ..., Y_{Jt})$, this implies

$$\left| \Pr\left(\tau_t \in \widehat{C}_{1-\alpha}(Y_{1t}, Y_{2t}, ..., Y_{Jt}) \right) - (1-\alpha) \right|$$

$$\le \sqrt{\frac{1}{2(T_0 - T_\mathcal{E})} \log\left(\frac{2}{z_\mathcal{B}}\right)} + \kappa \sqrt{\frac{8eJ\overline{\sigma}^2 \overline{\lambda}^2 \overline{\eta}^2 F^2}{\underline{\varsigma}^2 T_\mathcal{E}} \log\left(\frac{2J}{z_\mathcal{E}}\right)} + 2\kappa \sqrt{\frac{8eJ\overline{\sigma}^2 \overline{\lambda}^4 F^2}{\underline{\varsigma}^2 T_\mathcal{E}} \log\left(\frac{2J}{z_\mathcal{E}}\right)} + z_\mathcal{B} + 3z_\mathcal{E}.$$

∎

<div align="center">
Online Appendix
Synthetic Controls for Experimental Design

Alberto Abadie and Jinglong Zhao
April 24, 2025
</div>

OA.1. Designs Based on Penalized and Bias-corrected Synthetic Control Methods

Consider the design problem in (9),

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

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

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

and the terms (b) with

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

Here, λ_1 and λ_2 are positive constants that penalize discrepancies between the target values of the predictors ($\overline{\boldsymbol{X}}$ in (OA.2) and $\boldsymbol{X}_j$ in (OA.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 (OA.1). These terms can be adapted as in (OA.2) and (OA.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

$$\widehat{\tau}_t^{BC} = \sum_{j=1}^{J} w_j^*(Y_{jt} - \widehat{\mu}_{0t}(\boldsymbol{X}_j)) - \sum_{j=1}^{J} v_j^*(Y_{jt} - \widehat{\mu}_{0t}(\boldsymbol{X}_j)),$$

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 unit's characteristics, $\boldsymbol{X}_j$. To avoid over-fitting biases, $\widehat{\mu}_{0t}(\boldsymbol{X}_j)$ can be cross-fitted for the untreated.

<div align="center">OA.1</div>

OA.2. Approximate Validity when λ_t are not Exchangeable

Recall that in Theorem 2 we have shown that when $\boldsymbol{\lambda}_t$ are exchangeable for $t \in \mathcal{B} \cup \{T_0+1, \ldots, T\}$ the p-value in (19) is exact. In this section, we discuss the case when $\boldsymbol{\lambda}_t$ are not necessarily exchangeable. We show below in Theorem OA.1 that the p-value in (19) is approximately valid when $T_{\mathcal{E}}$ is large.

Theorem OA.1 *Assume that Assumptions 1 – 3 hold. Assume there exists a constant $\kappa < \infty$, such that for $j = 1, \ldots, J$, $t = 1, \ldots, T$, ϵ_{jt} are continuously distributed with (a version of) the probability density function upper bounded by κ. Then, under the null hypothesis (17), the p-values of equation (19) are approximately valid. In particular, there is an event $\mathcal{C}$, such that conditional on $\mathcal{C}$, for any $\alpha \in (0, 1]$, we have*

$$\alpha - 2z_2 - \frac{1}{|\Pi|} \leq \Pr(\widehat{p} \leq \alpha) \leq \alpha + 2z_2,$$

and the event $\mathcal{C}$ happens with probability at least

$$\Pr(\mathcal{C}) \geq 1 - 2J \exp\left(-\frac{z_1^2 \zeta^2}{8\bar{\sigma}^2 \bar{\lambda}^4 F^2} T_{\mathcal{E}}\right) - \frac{z_1}{z_2} 4e\sqrt{2J(\min\{T - T_0, T_0 - T_{\mathcal{E}}\})^3}\, \kappa,$$

where z_1, z_2 are arbitrary positive constants.

A limitation of the result in Theorem OA.1 is that there are values of the parameters of the data generating for which the result of the theorem provides a tight bound on test size only for large values of $T_{\mathcal{E}}$. We prove Theorem OA.1 next.

OA.2.1. Definitions

First, define $T_p = \min\{T - T_0, T_0 - T_{\mathcal{E}}\}$. Next, recall that

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

for $t \in \mathcal{B} \cup \{T_0 + 1, \ldots, T\}$. For $t \in \{T_0 + 1, \ldots, T\}$, $\widehat{u}_t$ 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.

Let

$$u_t = \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \tag{OA.4}$$

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

$$u_t = \sum_{j=1}^{J} w_j^* \xi_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \tag{OA.5}$$

for $t \in \{T_0 + 1, \ldots, T\}$. For each $\pi \in \Pi$, similar to our definition of $\widehat{e}_\pi$, define the $(T - T_0)$-dimensional vector

$$\boldsymbol{e}_\pi = \left(u_{\pi(1)}, u_{\pi(2)}, \ldots, u_{\pi(T-T_0)} \right).$$

In addition, let $\boldsymbol{e} = (u_1, \ldots, u_{T-T_0}) = (\tau_{T_0+1}, \ldots, \tau_T)$. It is useful to observe that, under the null hypothesis in (17), the random variables u_t for $t \in \mathcal{B} \cup \{T_0 + 1, \ldots, T\}$ are independent and identically distributed.

Next, define the following two functions. Let

$$\widehat{F}(x) = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{1}\left\{ S(\widehat{\boldsymbol{e}}_\pi) < x \right\},$$

and

$$\tilde{F}(x) = \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{1}\left\{ S(\boldsymbol{e}_\pi) < x \right\}.$$

The proof of Theorem OA.1 proceeds in four steps. In step one, we define a high probability event, $\mathcal{C}_1$, such that u_t and $\widehat{u}_t$ are close to each other under $\mathcal{C}_1$. In step two, we define a high probability event, $\mathcal{C}_2$, such that many components of $\{S(\boldsymbol{e}_\pi)\}_{\pi \in \Pi}$ are well-separated from $S(\boldsymbol{e})$ under $\mathcal{C}_2$. In step three, we show that, conditional on $\mathcal{C}_1$ and $\mathcal{C}_2$, the ordering of $S(\boldsymbol{e}_\pi)$ and $S(\boldsymbol{e})$ will be the same as the ordering of $S(\widehat{\boldsymbol{e}}_\pi)$ and $S(\widehat{\boldsymbol{e}})$ for most $\pi \in \Pi$, which implies that $\widehat{F}(S(\widehat{\boldsymbol{e}}))$ and $\tilde{F}(S(\boldsymbol{e}))$ are also close to each other. In step four, we conclude the proof by linking $\widehat{F}(S(\widehat{\boldsymbol{e}}))$ to the estimated p-value, and $\tilde{F}(S(\boldsymbol{e}))$ to the nominal level α.

OA.2.2. Lemmas for the Proof of Theorem OA.1

For each continuously distributed random variable X with a density f_X, define Λ_X to be the smallest upper bound on the probability density f_X.

Lemma OA.2 (Corollary 2, Bobkov and Chistyakov (2014)) *Let $X_1, X_2, \ldots, X_n$ be independent and continuously distributed random variables with densities $f_{X_1}, f_{X_2}, \ldots, f_{X_n}$. For any $k \in \{1, 2, \ldots, n\}$, let Λ_{X_k} be the smallest upper bound on the probability density f_{X_k}. For any $a_1, a_2, \ldots, a_n$, let $X = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n$. Suppose for any $k \in \{1, 2, \ldots, n\}$, $\Lambda_{X_k} \leq \kappa$; and if $\sum_{k=1}^{n} a_k^2 = 1$,*

$$\Lambda_X \leq \sqrt{e}\kappa.$$

Lemma OA.3 *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 $\Lambda_{aX} \leq \Lambda_X/|a|$.*

Proof of Lemma OA.3. 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 $v \geq 0$,

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

∎

Lemma OA.4 *Recall that u_t is defined as (OA.4) and (OA.5), for the blank periods and the experimental periods, respectively. Under the null hypothesis (17), the probability density of u_t can be bounded by*

$$\Lambda_{u_t} \leq \frac{1}{2}\sqrt{eJ}\kappa.$$

OA.4

Proof of Lemma OA.4. This proof consists of two steps. In Step 1, we prove a version of the lemma after conditioning on $(\boldsymbol{w}^*, \boldsymbol{v}^*)$. In Step 2, we apply the law of total probability to obtain a bound on the unconditional density of u_t.

Step 1. We condition on $(\boldsymbol{w}^*, \boldsymbol{v}^*)$ and write $u_t|(\boldsymbol{w}^*, \boldsymbol{v}^*)$ to indicate that we are conditional on $(\boldsymbol{w}^*, \boldsymbol{v}^*)$.

Fix any $t \in \mathcal{B} \cup \{T_0 + 1, \ldots, T\}$. Using Lemma OA.2 (Bobkov and Chistyakov, 2014, Corollary 2), let there be J variables ϵ_{jt} for any $j \in \{1, 2, \ldots, J\}$. For any $j \in \{1, 2, \ldots, J\}$, define $a_j = \frac{w_j^* - v_j^*}{\sqrt{\sum_{j=1}^J (w_j^* - v_j^*)^2}}$ such that $\sum_{j=1}^J a_j^2 = 1$. Using a_j, we can write u_t as $u_t = \sqrt{\sum_{j=1}^J (w_j^* - v_j^*)^2} \cdot \sum_{j=1}^J a_j \epsilon_{jt}$. When J is even,

$$
\begin{aligned}
\Lambda_{u_t|(\boldsymbol{w}^*, \boldsymbol{v}^*)} &\leq \frac{1}{\sqrt{\sum_{j=1}^J (w_j^* - v_j^*)^2}} \cdot \Lambda_{\sum_{j=1}^J a_j \epsilon_{jt}} \\
&\leq \frac{1}{\sqrt{\sum_{j=1}^J (w_j^* - v_j^*)^2}} \cdot \sqrt{e\kappa} \\
&\leq \frac{1}{\sqrt{\sum_{j=1}^J (\frac{2}{J})^2}} \cdot \sqrt{e\kappa} \\
&= \frac{\sqrt{J}}{2} \sqrt{e\kappa},
\end{aligned}
$$

where the first inequality is due to Lemma OA.3 Part 2; the second inequality is due to Lemma OA.2; the third inequality is due to convexity and Jensen's inequality, and the worst case is taken when $w_j^* = 2/J$ for one half of total units and $v_j^* = 2/J$ for the other half. When J is odd,

$$
\begin{aligned}
\Lambda_{u_t|(\boldsymbol{w}^*, \boldsymbol{v}^*)} &\leq \frac{1}{\sqrt{\sum_{j=1}^J (w_j^* - v_j^*)^2}} \cdot \Lambda_{\sum_{j=1}^J a_j \epsilon_{jt}} \\
&\leq \frac{1}{\sqrt{\sum_{j=1}^J (w_j^* - v_j^*)^2}} \cdot \sqrt{e\kappa} \\
&\leq \frac{1}{\sqrt{\frac{J+1}{2}(\frac{2}{J+1})^2 + \frac{J-1}{2}(\frac{2}{J-1})^2}} \cdot \sqrt{e\kappa} \\
&= \sqrt{\frac{J^2 - 1}{J}} \cdot \frac{\sqrt{e\kappa}}{2},
\end{aligned}
$$

$$\leq \frac{\sqrt{J}}{2}\sqrt{e}\kappa,$$

where the first inequality is due to Lemma OA.3 Part 2; the second inequality is due to Lemma OA.2; the third inequality is due to convexity and Jensen's inequality, and the worst case is taken when $w_j^* = 2/(J+1)$ for $(J+1)/2$ of total units and $v_j^* = 2/(J-1)$ for the other $(J-1)/2$ of total units.

Step 2. Using the law of total probability, we show

$$
\begin{aligned}
f_{u_t}(u) &= \int_{(\boldsymbol{w}^*,\boldsymbol{v}^*)} f\big(u|(\boldsymbol{w}^*,\boldsymbol{v}^*)\big)\,\mathrm{d}P(\boldsymbol{w}^*,\boldsymbol{v}^*)\\
&\leq \int_{(\boldsymbol{w}^*,\boldsymbol{v}^*)} \frac{\sqrt{J}}{2}\sqrt{e}\kappa\,\mathrm{d}P(\boldsymbol{w}^*,\boldsymbol{v}^*)\\
&= \frac{\sqrt{J}}{2}\sqrt{e}\kappa,
\end{aligned}
$$

where we use $P(\boldsymbol{w}^*,\boldsymbol{v}^*)$ to stand for the joint distribution of $(\boldsymbol{w}^*,\boldsymbol{v}^*)$. $\blacksquare$

OA.2.3.   Proof of Theorem OA.1

Proof of Theorem OA.1. (Step one.) Note 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

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

Under the null hypothesis (17), it follows that

$$
\begin{aligned}
\widehat{u}_t &= \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt}\\
&= -\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\sum_{j=1}^{J}(w_j^* - v_j^*)\boldsymbol{\epsilon}_j^{\mathcal{E}} + u_t,
\end{aligned}
$$

OA.6

for $t \in \mathcal{B} \cup \{T_0 + 1, \ldots, T\}$. We next define an event

$$\mathcal{C}_1 = \left\{ \forall t \in \mathcal{B} \cup \{T_0 + 1, \ldots, T\}, \left| \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J}(w_j^* - v_j^*)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right| \leq z_1 \right\}$$

$$= \left\{ \max_{t \in \mathcal{B} \cup \{T_0+1,\ldots,T\}} \left| \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J}(w_j^* - v_j^*)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right| \leq z_1 \right\}.$$

Note that,

$$\max_{t \in \mathcal{B} \cup \{T_0+1,\ldots,T\}} \left| \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J}(w_j^* - v_j^*)\boldsymbol{\epsilon}_j^{\mathcal{E}} \right|$$

$$\leq \max_{t \in \mathcal{B} \cup \{T_0+1,\ldots,T\}} \sum_{j=1}^{J} |w_j^* - v_j^*| \sum_{s \in \mathcal{E}} |\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_s||\epsilon_{js}|$$

$$\leq \sum_{j=1}^{J} |w_j^* - v_j^*| \sum_{s \in \mathcal{E}} \frac{\overline{\lambda}^2 F}{T_{\mathcal{E}}\underline{\zeta}} |\epsilon_{js}|,$$

where the second inequality is due to (A.6), and because $|w_j^* - v_j^*| \geq 0$ and $|\epsilon_{js}| \geq 0$. Therefore,

$$\Pr(\mathcal{C}_1) \geq 1 - \Pr\left(\sum_{j=1}^{J} \frac{|w_j^* - v_j^*|}{2} \sum_{s \in \mathcal{E}} \frac{\overline{\lambda}^2 F}{T_{\mathcal{E}}\underline{\zeta}} |\epsilon_{js}| > \frac{z_1}{2} \right)$$

$$\geq 1 - \sum_{j=1}^{J} \Pr\left(\sum_{s \in \mathcal{E}} \frac{\overline{\lambda}^2 F}{T_{\mathcal{E}}\underline{\zeta}} |\epsilon_{js}| > \frac{z_1}{2} \right)$$

$$\geq 1 - 2J \exp\left(-\frac{z_1^2 \underline{\zeta}^2}{8\overline{\sigma}^2 \overline{\lambda}^4 F^2} T_{\mathcal{E}} \right),$$

where the second inequality follows from union bound, and the third inequality is the Chernoff bound for sub-Gaussian random variables.

(Step two.) Define $\tilde{z}_1 = 2z_1 > 0$, and $T_p = \min\{T - T_0, T_0 - T_{\mathcal{E}}\}$. For each $k \in \{0, 1, 2, \ldots, T_p\}$, we define the following sets of permutations. First, define $\Pi_0 = \{\pi_0\}$, where π_0 is defined as the set of post-intervention indices $\pi_0 = \{T_0 + 1, \ldots, T\}$. Then, for any $k \in \{1, 2, \ldots, T_p\}$, define

$$\Pi_k = \left\{ \pi \in \Pi \,\middle|\, |\pi \setminus \pi_0| = k \right\}$$

to be the set of $(T - T_0)$-combinations with exactly k many indices from the blank periods. Using the above definitions, we can decompose Π into

$$\Pi = \bigcup_{k=0}^{T_p} \Pi_k.$$

OA.7

Then, for any $k \in \{1, 2, ..., T_p\}$ and $\pi \in \Pi_k$, we focus on the following indicator

$$\mathbb{1}\left\{\left|\sum_{t \in \pi \setminus \pi_0} |u_t| - \sum_{t \in \pi_0 \setminus \pi} |u_t|\right| \le 2kz_1\right\}.$$

The above indicator involves $2k$ instances of $|u_t|$'s. Intuitively, it is obtained by canceling out common terms in $S(e_\pi)$ and $S(e)$.

Below we focus on the properties of the sum of such indicators. First, focus on the probability density of $\left|\sum_{t \in \pi \setminus \pi_0} |u_t| - \sum_{t \in \pi_0 \setminus \pi} |u_t|\right|$. We have

$$\begin{aligned}
\Lambda_{\left|\sum_{t \in \pi \setminus \pi_0} |u_t| - \sum_{t \in \pi_0 \setminus \pi} |u_t|\right|} &\le 2\Lambda_{\sum_{t \in \pi \setminus \pi_0} |u_t| - \sum_{t \in \pi_0 \setminus \pi} |u_t|} \\
&\le 2\sqrt{2k}\Lambda_{\sum_{t \in \pi \setminus \pi_0} \frac{1}{\sqrt{2k}} |u_t| - \sum_{t \in \pi_0 \setminus \pi} \frac{1}{\sqrt{2k}} |u_t|} \\
&\le 2\sqrt{2k}\sqrt{e}\Lambda_{|u_t|} \\
&\le 2\sqrt{2k}\sqrt{e}\sqrt{eJ}\kappa \\
&= 2\sqrt{2Jk}e\kappa,
\end{aligned}$$

where the first inequality is due to Lemma OA.3-1; the second inequality is due to Lemma OA.3-2; the third inequality is due to Lemma OA.2; the last inequality is due to Lemma OA.4 and OA.3-1. We obtain

$$\Pr\left(\left|\sum_{t \in \pi \setminus \pi_0} |u_t| - \sum_{t \in \pi_0 \setminus \pi} |u_t|\right| \le 2kz_1\right) \le 4e\sqrt{2Jk^3}z_1\kappa.$$

Next, due to Markov inequality, for any constant $z_2 > 0$, we have

$$\begin{aligned}
&\Pr\left(\sum_{k=1}^{T_p} \sum_{\pi \in \Pi_k} \mathbb{1}\left\{\left|\sum_{t \in \pi \setminus \pi_0} |u_t| - \sum_{t \in \pi_0 \setminus \pi} |u_t|\right| \le 2kz_1\right\} \ge |\Pi|z_2\right) \\
&\qquad \le \frac{1}{|\Pi|z_2} \sum_{k=1}^{T_p} \sum_{\pi \in \Pi_k} E\left[\mathbb{1}\left\{\left|\sum_{t \in \pi \setminus \pi_0} |u_t| - \sum_{t \in \pi_0 \setminus \pi} |u_t|\right| \le 2kz_1\right\}\right] \\
&\qquad \le \frac{\sum_{k=1}^{T_p} |\Pi_k| 4e\sqrt{2Jk^3}z_1\kappa}{|\Pi|z_2}.
\end{aligned}$$

To conclude step two, define the event

$$\mathcal{C}_2 = \left\{\sum_{k=1}^{T_p} \sum_{\pi \in \Pi_k} \mathbb{1}\left\{\left|\sum_{t \in \pi \setminus \pi_0} |u_t| - \sum_{t \in \pi_0 \setminus \pi} |u_t|\right| \le 2kz_1\right\} < |\Pi|z_2\right\}. \qquad \text{(OA.6)}$$

OA.8

The probability that event $\mathcal{C}_2$ happens is at least

$$\Pr(\mathcal{C}_2) \geq 1 - \frac{\sum_{k=1}^{T_p} |\Pi_k| \sqrt{k^3}}{|\Pi|} \frac{4e\sqrt{2J}z_1\kappa}{z_2}.$$

(Step three.) Conditional on event $\mathcal{C}_2$, fewer than $|\Pi|z_2$ of the absolute value terms in (OA.6) are such that $\left|\sum_{t\in\pi\backslash\pi_0} |u_t| - \sum_{t\in\pi_0\backslash\pi} |u_t|\right| \leq 2kz_1$. For all the others, $\left|\sum_{t\in\pi\backslash\pi_0} |u_t| - \sum_{t\in\pi_0\backslash\pi} |u_t|\right| > 2kz_1$.

Conditional on event $\mathcal{C}_1$, we know that $|\widehat{u}_t - u_t| \leq z_1$ for any $t \in \mathcal{B} \cup \{T_0+1,\ldots,T\}$. So we have that $\sum_{t\in\pi\backslash\pi_0} |u_t| - \sum_{t\in\pi_0\backslash\pi} |u_t| > 2kz_1$ implies

$$\begin{aligned}
S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}) &= \frac{1}{T-T_0}\sum_{t\in\pi} |\widehat{u}_t| - \frac{1}{T-T_0}\sum_{t\in\pi_0} |\widehat{u}_t| \\
&= \frac{1}{T-T_0}\left(\sum_{t\in\pi\backslash\pi_0} |\widehat{u}_t| - \sum_{t\in\pi_0\backslash\pi} |\widehat{u}_t|\right) \\
&\geq \frac{1}{T-T_0}\left(\sum_{t\in\pi\backslash\pi_0} (|u_t|-z_1) - \sum_{t\in\pi_0\backslash\pi} (|u_t|+z_1)\right) \\
&> \frac{1}{T-T_0}(2kz_1 - 2kz_1) \\
&= 0,
\end{aligned}$$

where the first equality is due to definition $S(\boldsymbol{e}_\pi) = \frac{1}{T-T_0}\sum_{t\in\pi} |u_t|$. Similarly, $\mathcal{C}_1$ and $\sum_{t\in\pi\backslash\pi_0} |u_t| - \sum_{t\in\pi_0\backslash\pi} |u_t| < -2kz_1$ imply

$$\begin{aligned}
S(\widehat{\boldsymbol{e}}_\pi) - S(\widehat{\boldsymbol{e}}) &= \frac{1}{T-T_0}\left(\sum_{t\in\pi\backslash\pi_0} |\widehat{u}_t| - \sum_{t\in\pi_0\backslash\pi} |\widehat{u}_t|\right) \\
&\leq \frac{1}{T-T_0}\left(\sum_{t\in\pi\backslash\pi_0} (|u_t|+z_1) - \sum_{t\in\pi_0\backslash\pi} (|u_t|-z_1)\right) \\
&< \frac{1}{T-T_0}(-2kz_1 + 2kz_1) \\
&= 0.
\end{aligned}$$

Combining both cases, we know that conditional on $\mathcal{C}_1$ and when $\left|\sum_{t\in\pi\backslash\pi_0} |u_t| - \sum_{t\in\pi_0\backslash\pi} |u_t|\right| > 2kz_1$, the ordering of $S(\boldsymbol{e}_\pi)$ and $S(\boldsymbol{e})$ is the same as the ordering of $S(\widehat{\boldsymbol{e}}_\pi)$ and $S(\widehat{\boldsymbol{e}})$. As a

result, for those π such that $\left| \sum_{t \in \pi \backslash \pi_0} |u_t| - \sum_{t \in \pi_0 \backslash \pi} |u_t| \right| > 2kz_1$, we have $\mathbb{1}\{S(\widehat{e}_\pi) \geq S(\widehat{e})\} = \mathbb{1}\{S(e_\pi) \geq S(e)\}$. There are at most $|\Pi|z_2$ many π's that contribute to the following summation,

$$\left| \sum_{k=1}^{T_p} \sum_{\pi \in \Pi_k} \left(\mathbb{1}\{S(\widehat{e}_\pi) \geq S(\widehat{e})\} - \mathbb{1}\{S(e_\pi) \geq S(e)\} \right) \right| < |\Pi|z_2.$$

Note that $S(\widehat{e}_{\pi_0}) = S(\widehat{e})$ and $S(e_{\pi_0}) = S(e)$, so $\mathbb{1}\{S(\widehat{e}_\pi) > S(\widehat{e})\} = \mathbb{1}\{S(e_{\pi_0}) > S(e)\}$ is always true. Combining π_0 we have

$$\left| \sum_{\pi \in \Pi} \left(\mathbb{1}\{S(\widehat{e}_\pi) \geq S(\widehat{e})\} - \mathbb{1}\{S(e_\pi) \geq S(e)\} \right) \right|$$

$$= \left| \sum_{k=0}^{T_p} \sum_{\pi \in \Pi_k} \left(\mathbb{1}\{S(\widehat{e}_\pi) \geq S(\widehat{e})\} - \mathbb{1}\{S(e_\pi) \geq S(e)\} \right) \right|$$

$$< |\Pi|z_2.$$

We conclude step three using the following block of inequalities. For any $\alpha \in (0, 1]$,

$$\left| \Pr\left(1 - \widehat{F}(S(\widehat{e})) \leq \alpha\right) - \Pr\left(1 - \tilde{F}(S(e)) \leq \alpha\right) \right|$$

$$= \left| \Pr\left(1 - \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{e}_\pi) < S(\widehat{e})\} \leq \alpha\right) - \Pr\left(1 - \frac{1}{|\Pi|} \sum_{\pi \in \Pi} \mathbb{1}\{S(e_\pi) < S(e)\} \leq \alpha\right) \right|$$

$$= \left| \Pr\left(\sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{e}_\pi) \geq S(\widehat{e})\} \leq \alpha|\Pi|\right) - \Pr\left(\sum_{\pi \in \Pi} \mathbb{1}\{S(e_\pi) \geq S(e)\} \leq \alpha|\Pi|\right) \right|$$

$$= \left| E\left[\mathbb{1}\left\{\sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{e}_\pi) \geq S(\widehat{e})\} \leq \alpha|\Pi|\right\}\right] - E\left[\mathbb{1}\left\{\sum_{\pi \in \Pi} \mathbb{1}\{S(e_\pi) \geq S(e)\} \leq \alpha|\Pi|\right\}\right] \right|$$

$$\leq E\left| \mathbb{1}\left\{\sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{e}_\pi) \geq S(\widehat{e})\} \leq \alpha|\Pi|\right\} - \mathbb{1}\left\{\sum_{\pi \in \Pi} \mathbb{1}\{S(e_\pi) \geq S(e)\} \leq \alpha|\Pi|\right\} \right|$$

$$\leq \Pr\left(\left| \alpha|\Pi| - \sum_{\pi \in \Pi} \mathbb{1}\{S(e_\pi) \geq S(e)\} \right| \leq \left| \sum_{\pi \in \Pi} \left(\mathbb{1}\{S(\widehat{e}_\pi) \geq S(\widehat{e})\} - \mathbb{1}\{S(e_\pi) \geq S(e)\} \right) \right| \right),$$

where the second inequality is due to the following: $|\mathbb{1}\{a \leq c\} - \mathbb{1}\{b \leq c\}| \leq \mathbb{1}\{|c - b| \leq |a - b|\}$.

Conditional on events $\mathcal{C}_1$ and $\mathcal{C}_2$, we obtain

$$\left| \Pr\left(1 - \widehat{F}(S(\widehat{e})) \leq \alpha\right) - \Pr\left(1 - \tilde{F}(S(e)) \leq \alpha\right) \right|$$

$$\leq \Pr\left(\left| \alpha|\Pi| - \sum_{\pi \in \Pi} \mathbb{1}\{S(e_\pi) \geq S(e)\} \right| < |\Pi|z_2 \right)$$

OA.10

$$\leq \frac{2|\Pi|z_2}{|\Pi|}$$
$$= 2z_2, \tag{OA.7}$$

where the last inequality is because $\sum_{\pi \in \Pi} \mathbb{1}\{S(\boldsymbol{e}_\pi) \geq S(\boldsymbol{e})\}$ is a discrete uniform distribution over $\{1, 2, ..., |\Pi|\}$, and that there are at most $2|\Pi|z_2$ many integers centered around $\alpha|\Pi|$.

(Step four.) Note that, for any $\alpha \in (0, 1]$,

$$\alpha - \frac{1}{|\Pi|} \leq \Pr\left(1 - \tilde{F}(S(\boldsymbol{e})) \leq \alpha\right) \leq \alpha.$$

So conditional on events $\mathcal{C}_1$ and $\mathcal{C}_2$, (OA.7) implies

$$\Pr\left(1 - \widehat{F}(S(\widehat{\boldsymbol{e}})) \leq \alpha\right) \leq \Pr\left(1 - \tilde{F}(S(\boldsymbol{e})) \leq \alpha\right) + 2z_2 \leq \alpha + 2z_2$$

and

$$\Pr\left(1 - \widehat{F}(S(\widehat{\boldsymbol{e}})) \leq \alpha\right) \geq \Pr\left(1 - \tilde{F}(S(\boldsymbol{e})) \leq \alpha\right) - 2z_2 \geq \alpha - 2z_2 - \frac{1}{|\Pi|}.$$

Combining both parts, conditional on $\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2$, we have

$$\alpha - 2z_2 - \frac{1}{|\Pi|} \leq \Pr(\widehat{p} \leq \alpha) = \Pr\left(1 - \widehat{F}(S(\widehat{\boldsymbol{e}})) \leq \alpha\right) \leq \alpha + 2z_2,$$

and $\mathcal{C}$ happens with probability at least

$$\Pr(\mathcal{C}_1 \cap \mathcal{C}_2) \geq (1 - \Pr(\mathcal{C}_1)) + (1 - \Pr(\mathcal{C}_2)) - 1$$
$$\geq 1 - 2J \exp\left(-\frac{z_1^2 \underline{\zeta}^2}{8\overline{\sigma}^2 \overline{\lambda}^4 F^2} T_{\mathcal{E}}\right) - \frac{\sum_{k=1}^{T_p} |\Pi_k|\sqrt{k^3}}{|\Pi|} \cdot \frac{z_1}{z_2} \cdot 4e\sqrt{2J}\kappa,$$

which finishes the proof. ∎

OA.3. Estimating the Average Effect of Treatment on the Treated Units

In Section 3, we have shown formal results of the bias bounds in estimating the average treatment effect. In this section, we present similar results for estimating the average effect of treatment on the treated units. Similar to Assumption 3, we begin with the assumption of perfect fit.

Assumption 5 *With probability one, (i)*

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

and (ii)

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

In practice, Assumption 3 may only hold approximately. The next assumption accommodates settings with imperfect fit.

Assumption 6 *There exists a positive constant $d > 0$, such that with probability one,*

$$\left\| \sum_{j=1}^{J} w_j^* \mathbf{Z}_j - \sum_{j=1}^{J} v_j^* \mathbf{Z}_j \right\|_2^2 \leq Rd^2, \qquad \left\| \sum_{j=1}^{J} w_j^* \mathbf{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \mathbf{Y}_j^{\mathcal{E}} \right\|_2^2 \leq T_{\mathcal{E}} d^2.$$

Using the above assumptions, we are able to provide the following bias bounds.

Theorem OA.5 *If Assumptions 1, 2, and 5 hold, then for any $t \geq T_0 + 1$,*

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

If Assumptions 1, 2, and 6 hold, then for any $t \geq T_0 + 1$,

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

We provide the following result on inference.

Theorem OA.6 *Suppose that Assumptions 1 and 5(i) hold. Assume that $\{\boldsymbol{\lambda}_t\}_{t \in \mathcal{B} \cup \{T_0+1,\dots,T\}}$ is a sequence of exchangeable random variables independent of $\{\epsilon_{jt}\}_{t \in \mathcal{B} \cup \{T_0+1,\dots,T\}}$ and $\{\xi_{jt}\}_{t \in \{T_0+1,\dots,T\}}$. Assume also that for each $j = 1, \dots, J$, $\{\epsilon_{jt}\}_{t \in \mathcal{B} \cup \{T_0+1,\dots,T\}}$ and $\{\xi_{jt}\}_{t \in \{T_0+1,\dots,T\}}$ are two sequences of exchangeable random variables, respectively. Then under the null hypothesis (17), we have*

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

for any $\alpha \in [0,1]$, where $\Pr(\widehat{p} \leq \alpha)$ is taken over the distributions of $\{\xi_{jt}, \epsilon_{jt}, \boldsymbol{\lambda}_t\}$.

If Assumption 1 holds but 5(i) is violated, the same result holds if $\{(\boldsymbol{\theta}_t, \boldsymbol{\lambda}_t)\}_{t \in \mathcal{B} \cup \{T_0+1,...,T\}}$ is a sequence of exchangeable random variables independent of $\{\epsilon_{jt}\}_{t \in \mathcal{B} \cup \{T_0+1,...,T\}}$ and $\{\xi_{jt}\}_{t \in \{T_0+1,...,T\}}$, and if for each $j = 1, \ldots, J$, $\{\epsilon_{jt}\}_{t \in \mathcal{B} \cup \{T_0+1,...,T\}}$ and $\{\xi_{jt}\}_{t \in \{T_0+1,...,T\}}$ are two sequences of exchangeable random variables, respectively. Here, $\Pr(\widehat{p} \leq \alpha)$ is taken over the distributions of $\{\xi_{jt}, \epsilon_{jt}, \boldsymbol{\lambda}_t, \boldsymbol{\theta}_t\}$.

OA.3.1. Simulation Results for Average Treatment Effects on the Treated

In this section, we estimate the average treatment effects on the treated units by conducting simulations following the simulation setup as in Section 5.2. Recall that we compare the performance of different synthetic control designs over 1000 simulations that independently generate the model primitives (i.e., the factor loadings, covariates, and error terms) of Assumption 1. The data generating process for each one of the 1000 simulations is the same as in Section 5.1. The five varieties of the synthetic control design are described in Section 5.2.

We report the average treatment effects on the treated units in Table OA.1 as well as the average treatment effects on the treated units under the nonlinear model in Table OA.2.

OA.3.1.1. *Average Treatment Effects on the Treated*

The first five columns in Table OA.1 in the Online Appendix report averages of τ_t^T, the average effect of treatment on the treated units. These quantities depend on the weights for the treated units, which are different across different formulations of the synthetic control design. The next five columns report averages of $\widehat{\tau}_t$. They are the same as in Table 2, yet we use them as estimators for τ_t^T in Table OA.1. The last two columns of Table OA.1 report averages across simulations of the mean absolute error and the root mean square error, defined as in (22) but with τ_t^T replacing τ_t.

The results in Table OA.1 are qualitatively similar to those for τ_t in Table 2, with one notable exception. As expected, for intermediate and large values of β, the *Weakly-targeted* design outperforms the other designs when the goal is to estimate τ_t^T. This is because in the *Weakly-targeted* design the synthetic control weights are targeted to τ_t^T (and more so as β becomes large).

OA.3.1.2. Performance with Nonlinearities

We now study the behavior of the estimators based on synthetic control designs under deviations from the linear model in (12a) and (12b). We consider a nonlinear data generating process as defined in (23a) and (23b).

Table OA.2 reports the results for τ_t^T. In comparison to the results in Tables OA.1, we now see that the *Unit-level* and *Penalized* designs can easily match and in some cases improve the performance of the *Unconstrained* design, especially for the estimation of τ_t^T. By fitting each treated unit with a unit-specific synthetic control, the *Unit-level* design can ameliorate interpolation biases induced by the aggregation of $\boldsymbol{X}_j$. Like in Table OA.1, the *Weakly targeted* design easily outperforms the unconstrained estimator for large values of β.

Table OA.1: Average Treatment Effects on the Treated (Averages over 1000 Simulations)

		τ_t^T					$\widehat{\tau}_t$					MAE^T	$RMSE^T$
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$	$t=26$	$t=27$	$t=28$	$t=29$	$t=30$		
Unconstrained		-13.58	-10.98	-8.34	-5.00	-2.46	-13.57	-10.97	-8.37	-5.06	-2.52	1.01	1.18
Constrained	$\overline{m}=1$	-13.80	-11.09	-8.37	-4.82	-2.50	-13.61	-10.97	-8.39	-4.86	-2.41	3.02	3.57
	$\overline{m}=2$	-13.53	-10.89	-8.42	-4.93	-2.40	-13.58	-10.90	-8.43	-5.01	-2.40	1.80	2.13
	$\overline{m}=3$	-13.39	-10.92	-8.35	-4.97	-2.54	-13.56	-11.00	-8.38	-5.05	-2.52	1.39	1.64
	$\overline{m}=4$	-13.54	-11.05	-8.42	-4.95	-2.52	-13.59	-11.06	-8.40	-4.99	-2.50	1.22	1.44
	$\overline{m}=5$	-13.57	-11.03	-8.42	-5.02	-2.45	-13.57	-11.01	-8.37	-5.02	-2.48	1.11	1.30
	$\overline{m}=6$	-13.61	-11.06	-8.36	-4.99	-2.46	-13.51	-10.95	-8.29	-5.01	-2.47	1.03	1.22
	$\overline{m}=7$	-13.58	-10.97	-8.34	-5.00	-2.46	-13.57	-10.96	-8.37	-5.06	-2.52	1.01	1.18
Weakly-targeted	$\beta=0.01$	-13.59	-10.99	-8.31	-5.00	-2.51	-13.58	-10.95	-8.38	-4.99	-2.53	1.31	1.55
	$\beta=0.1$	-13.59	-10.98	-8.37	-5.01	-2.51	-13.57	-11.00	-8.34	-4.98	-2.52	1.07	1.26
	$\beta=1$	-13.55	-10.99	-8.35	-5.01	-2.48	-13.56	-10.98	-8.32	-4.93	-2.44	0.99	1.16
	$\beta=10$	-13.45	-10.96	-8.35	-5.06	-2.50	-13.57	-10.98	-8.38	-5.01	-2.51	0.99	1.15
	$\beta=100$	-13.49	-10.90	-8.33	-5.01	-2.48	-13.60	-10.98	-8.39	-5.07	-2.52	1.00	1.16
Unit-level	$\xi=0.01$	-13.60	-10.98	-8.35	-4.98	-2.51	-13.60	-10.95	-8.39	-5.04	-2.53	1.09	1.29
	$\xi=0.1$	-13.56	-10.99	-8.38	-4.95	-2.50	-13.58	-10.97	-8.35	-4.97	-2.47	1.08	1.28
	$\xi=1$	-13.57	-11.02	-8.31	-4.91	-2.48	-13.57	-10.99	-8.39	-4.99	-2.49	1.40	1.66
	$\xi=10$	-13.78	-10.98	-8.37	-4.87	-2.52	-13.60	-10.93	-8.45	-5.05	-2.52	1.96	2.33
	$\xi=100$	-13.81	-10.90	-8.39	-4.80	-2.55	-13.61	-10.86	-8.48	-5.02	-2.54	2.35	2.78
Penalized	$\lambda=0.01$	-13.59	-10.99	-8.35	-5.00	-2.47	-13.59	-10.98	-8.35	-5.05	-2.48	1.05	1.23
	$\lambda=0.1$	-13.57	-10.98	-8.36	-4.91	-2.42	-13.64	-11.03	-8.43	-5.03	-2.50	1.36	1.61
	$\lambda=1$	-13.69	-11.01	-8.33	-4.80	-2.48	-13.67	-10.96	-8.41	-4.87	-2.45	2.19	2.59
	$\lambda=10$	-13.88	-11.06	-8.23	-4.73	-2.40	-13.68	-11.04	-8.37	-4.79	-2.45	3.81	4.50
	$\lambda=100$	-13.90	-11.10	-8.27	-4.75	-2.46	-13.64	-10.94	-8.42	-4.86	-2.50	4.24	5.00

Table OA.2: Average Treatment Effects on the Treated, Nonlinear Model (Averages over 1000 Simulations)

		τ_t^T					$\widehat{\tau}_t$					MAE^T	$RMSE^T$
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$	$t=26$	$t=27$	$t=28$	$t=29$	$t=30$		
Unconstrained		-13.30	-10.65	-7.98	-5.47	-2.41	-13.44	-10.92	-8.18	-5.85	-2.78	2.31	3.01
Constrained	$\overline{m}=1$	-13.41	-10.81	-8.01	-5.25	-2.61	-15.70	-13.18	-10.50	-7.76	-4.78	3.53	4.28
	$\overline{m}=2$	-13.19	-10.88	-7.93	-5.42	-2.47	-14.27	-11.86	-8.90	-6.44	-3.34	2.86	3.61
	$\overline{m}=3$	-13.17	-10.82	-8.07	-5.54	-2.44	-13.69	-11.38	-8.38	-5.95	-2.97	2.52	3.25
	$\overline{m}=4$	-13.29	-10.82	-8.03	-5.59	-2.31	-13.58	-11.09	-8.23	-5.89	-2.75	2.44	3.16
	$\overline{m}=5$	-13.23	-10.77	-7.88	-5.58	-2.30	-13.37	-10.97	-8.14	-5.79	-2.88	2.39	3.10
	$\overline{m}=6$	-13.36	-10.71	-7.99	-5.50	-2.36	-13.54	-11.03	-8.31	-5.86	-2.86	2.32	3.02
	$\overline{m}=7$	-13.31	-10.68	-7.97	-5.46	-2.39	-13.49	-10.94	-8.17	-5.86	-2.78	2.30	3.01
Weakly-targeted	$\beta=0.01$	-13.14	-10.67	-7.97	-5.47	-2.45	-11.66	-9.02	-6.37	-3.87	-1.00	2.68	3.36
	$\beta=0.1$	-13.20	-10.66	-7.92	-5.47	-2.44	-12.08	-9.60	-6.87	-4.31	-1.47	2.31	2.94
	$\beta=1$	-13.20	-10.71	-7.91	-5.47	-2.46	-12.51	-10.13	-7.35	-4.81	-1.91	1.91	2.43
	$\beta=10$	-13.27	-10.67	-7.96	-5.45	-2.49	-13.03	-10.51	-7.81	-5.25	-2.32	1.33	1.66
	$\beta=100$	-13.29	-10.77	-8.04	-5.47	-2.53	-13.28	-10.72	-8.00	-5.43	-2.59	1.14	1.40
Unit-level	$\xi=0.01$	-13.20	-10.76	-7.98	-5.44	-2.48	-11.76	-9.15	-6.51	-3.91	-1.15	2.68	3.35
	$\xi=0.1$	-13.35	-10.81	-8.07	-5.53	-2.55	-13.11	-10.59	-7.82	-5.15	-2.29	1.98	2.51
	$\xi=1$	-13.40	-10.77	-8.05	-5.47	-2.56	-13.74	-11.12	-8.42	-5.75	-2.84	1.47	1.81
	$\xi=10$	-13.35	-10.66	-7.99	-5.38	-2.71	-13.74	-11.20	-8.55	-5.89	-3.09	1.49	1.80
	$\xi=100$	-13.39	-10.63	-7.95	-5.37	-2.69	-13.79	-11.16	-8.54	-5.90	-3.08	1.56	1.89
Penalized	$\lambda=0.01$	-13.32	-10.69	-8.07	-5.50	-2.41	-13.40	-10.93	-8.32	-5.82	-2.82	2.25	2.93
	$\lambda=0.1$	-13.17	-10.72	-8.09	-5.59	-2.55	-13.33	-10.79	-8.13	-5.56	-2.65	1.87	2.39
	$\lambda=1$	-13.40	-10.80	-7.98	-5.35	-2.66	-13.32	-10.84	-8.15	-5.39	-2.60	1.97	2.44
	$\lambda=10$	-13.41	-10.89	-8.00	-5.32	-2.64	-13.39	-10.82	-7.95	-5.34	-2.58	2.79	3.46
	$\lambda=100$	-13.42	-10.87	-8.02	-5.26	-2.63	-13.35	-10.82	-8.00	-5.29	-2.57	3.06	3.80

OA.3.2. Proofs of Theorem OA.5 and Theorem OA.6

Proof of Theorem OA.5. For any period $t = T_0 + 1, \ldots, T$ we decompose $(\widehat{\tau}_t^T - \tau_t^T)$ as follows,

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

From (12a), we obtain

$$\sum_{j=1}^{J} w_j^* Y_{jt}^N - \sum_{j=1}^{J} v_j^* Y_{jt}^N = \boldsymbol{\theta}_t' \Big(\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{Z}_j \Big)$$

$$+ \boldsymbol{\lambda}_t' \Big(\sum_{j=1}^{J} w_j^* \boldsymbol{\mu}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{\mu}_j \Big) + \Big(\sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \Big). \quad \text{(OA.8)}$$

Similarly, using expression (12a), we obtain

$$\sum_{j=1}^{J} w_j^* \boldsymbol{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{Y}_j^{\mathcal{E}} = \boldsymbol{\theta}_{\mathcal{E}} \Big(\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{Z}_j \Big)$$

$$+ \boldsymbol{\lambda}_{\mathcal{E}} \Big(\sum_{j=1}^{J} w_j^* \boldsymbol{\mu}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{\mu}_j \Big) + \Big(\sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} \Big),$$

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 $\boldsymbol{\lambda}_t' (\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}'$ yields

$$\boldsymbol{\lambda}_t' (\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \Big(\sum_{j=1}^{J} w_j^* \boldsymbol{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{Y}_j^{\mathcal{E}} \Big)$$

$$= \boldsymbol{\lambda}_t' (\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\theta}_{\mathcal{E}} \Big(\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{Z}_j \Big)$$

$$+ \boldsymbol{\lambda}_t' \Big(\sum_{j=1}^{J} w_j^* \boldsymbol{\mu}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{\mu}_j \Big)$$

$$+ \boldsymbol{\lambda}_t' (\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \Big(\sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} \Big). \quad \text{(OA.9)}$$

Equations (OA.8) and (OA.9) imply

$$\sum_{j=1}^{J} w_j^* Y_{jt}^N - \sum_{j=1}^{J} v_j^* Y_{jt}^N = (\boldsymbol{\theta}_t' - \boldsymbol{\lambda}_t' (\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\theta}_{\mathcal{E}}) \Big(\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{Z}_j \Big)$$

$$+ \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\Big(\sum_{j=1}^{J} w_j^* \boldsymbol{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{Y}_j^{\mathcal{E}}\Big)$$

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

$$+ \Big(\sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt}\Big). \tag{OA.10}$$

If Assumption 3 holds, (OA.10) becomes

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

$$+ \Big(\sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt}\Big). \tag{OA.11}$$

Only the first term on the right-hand side of (OA.11) has a non-zero mean (because the weights w_j^* and v_j^*, depend on the error terms $\boldsymbol{\epsilon}_j^{\mathcal{E}}$). Therefore,

$$\Big|E\Big[\sum_{j=1}^{J} w_j^* Y_{jt}^N - \sum_{j=1}^{J} v_j^* Y_{jt}^N\Big]\Big| = \Big|E\Big[\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\Big(\sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}}\Big)\Big]\Big|$$

$$\leq \Big|E\Big[\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}}\Big]\Big| + \Big|E\Big[\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\sum_{j=1}^{J} v_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}}\Big]\Big|.$$

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

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

Let

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

Because $\overline{\epsilon}_{jt}^{\mathcal{E}}$ is a linear combination of independent sub-Gaussians with variance proxy $\overline{\sigma}^2$, we know $\overline{\epsilon}_{jt}^{\mathcal{E}}$ is sub-Gaussian with variance proxy $(\overline{\lambda}^2 F/\underline{\zeta})^2\overline{\sigma}^2/T_{\mathcal{E}}$. Let $\mathcal{S} = \{\boldsymbol{w} \in \mathbb{R}^J : \sum_{j=1}^{J} w_j = 1\}$ be the unit simplex. Theorem 1.16 from Rigollet and Hütter (2019) implies

$$\Big|E\Big[\sum_{j=1}^{J} w_j^* Y_{jt}^N - \sum_{j=1}^{J} v_j^* Y_{jt}^N\Big]\Big| \leq \Big|E\Big[\sum_{j=1}^{J} w_j^* \overline{\epsilon}_{jt}^{\mathcal{E}}\Big]\Big| + \Big|E\Big[\sum_{j=1}^{J} v_j^* \overline{\epsilon}_{jt}^{\mathcal{E}}\Big]\Big|$$

OA.18

$$\leq E\left[\max_{\boldsymbol{w}\in\mathcal{S}}\left|\sum_{j=1}^{J}w_j\bar{\epsilon}_{jt}^{\mathcal{E}}\right|\right]+E\left[\max_{\boldsymbol{v}\in\mathcal{S}}\left|\sum_{j=1}^{J}v_j\bar{\epsilon}_{jt}^{\mathcal{E}}\right|\right]$$

$$\leq\frac{\overline{\lambda}^2 F}{\underline{\zeta}}2\sqrt{2\log(2J)}\frac{\overline{\sigma}}{\sqrt{T_{\mathcal{E}}}},$$

which finishes the proof of the theorem.

Suppose now Assumption 6 holds (but Assumption 5 does not). To obtain a bound on the bias we need to bound the first two terms in (OA.10). Recall that

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

Therefore, the absolute value of each element in vector $(\boldsymbol{\theta}_t'-\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\theta}_{\mathcal{E}})$ is bounded by $\overline{\theta}\left(1+\dfrac{\overline{\lambda}^2 F}{\underline{\zeta}}\right)$. Cauchy–Schwarz inequality and Assumption 6 imply

$$\left|(\boldsymbol{\theta}_t'-\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\theta}_{\mathcal{E}})\left(\sum_{j=1}^{J}w_j^*\boldsymbol{Z}_j-\sum_{j=1}^{J}v_j^*\boldsymbol{Z}_j\right)\right|$$

$$\leq\overline{\theta}\left(1+\frac{\overline{\lambda}^2 F}{\underline{\zeta}}\right)\sqrt{R}\left\|\sum_{j=1}^{J}w_j^*\boldsymbol{Z}_j-\sum_{j=1}^{J}v_j^*\boldsymbol{Z}_j\right\|_2$$

$$\leq\overline{\theta}\left(1+\frac{\overline{\lambda}^2 F}{\underline{\zeta}}\right)Rd,$$

and

$$\left|\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\left(\sum_{j=1}^{J}w_j^*\boldsymbol{Y}_j^{\mathcal{E}}-\sum_{j=1}^{J}v_j^*\boldsymbol{Y}_j^{\mathcal{E}}\right)\right|\leq\frac{\overline{\lambda}^2 F}{\underline{\zeta}}d.$$

Combining the last two displayed equations with (OA.10), we have

$$\left|E\left[\sum_{j=1}^{J}w_j^*Y_{jt}^I-\sum_{j=1}^{J}f_jY_{jt}^I\right]\right|\leq\left(\overline{\theta}R+\frac{\overline{\lambda}^2 F}{\underline{\zeta}}(1+\overline{\theta}R)\right)d+\frac{\overline{\lambda}^2 F}{\underline{\zeta}}2\sqrt{2\log(2J)}\frac{\overline{\sigma}}{\sqrt{T_{\mathcal{E}}}},$$

which finishes the proof of the theorem. ∎

Proof of Theorem OA.6. Recall that

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

OA.19

for $t \in \mathcal{B} \cup \{T_0 + 1, \ldots, T\}$. For $t \in \{T_0 + 1, \ldots, T\}$, $\widehat{u}_t$ 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. Let

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

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

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

for $t \in \{T_0+1, \ldots, T\}$. The null hypothesis (17) and the assumptions of Theorem OA.6 imply that $\{u_t\}_{t \in \mathcal{B} \cup \{T_0+1, \ldots, T\}}$ is a sequence of exchangeable random variables. Additionally, Assumption 1 and the null hypothesis (17) imply

$$\widehat{u}_t = \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} = \boldsymbol{\theta}_t' \sum_{j=1}^{J} (w_j^* - v_j^*) \boldsymbol{Z}_j + \boldsymbol{\lambda}_t' \sum_{j=1}^{J} (w_j^* - v_j^*) \boldsymbol{\mu}_j + u_t,$$

for $t \in \mathcal{B} \cup \{T_0 + 1, \ldots, T\}$. The result of the theorem follows now from Theorem D.1 in Chernozhukov, Wüthrich and Zhu (2021). ∎

OA.4. Swapping Treated and Control Weights

Recall that 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 or not to swap using the following rule. For the *Unconstrained* design, we choose the treated group to be the one with the smallest index among the units with positive weights. We use the same procedure based on the lowest index for *Constrained* with $\overline{m} = 7$ (highest value) and *Penalized* with $\lambda = 0.01$ (lowest value). Then, starting from $\overline{m} = 7$ and for smaller values of $\overline{m}$, we assign to the treated group the set of weights that is most similar to the weights obtained for $\|\boldsymbol{w}^*\|_0 \leq \overline{m} + 1$ (in terms of what units obtain positive weights). In those cases where the two sets of swappable weights for $\|\boldsymbol{w}^*\|_0 \leq \overline{m}$ are equally similar to the synthetic treated weights for $\|\boldsymbol{w}^*\|_0 \leq \overline{m} + 1$, we select the set of weights with the smallest index. We follow the analogous procedure for $\lambda > 0.01$, starting from smaller values of λ.

OA.5.   Implementations of the Optimization Formulations

To computationally solve (6), i.e., the *Unconstrained* design, we propose two methods. The first method is by enumeration, which takes advantage of the objective function of (6) 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 (6) 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 two groups of treated and control units can be swapped, we only enumerate combinations such that the cardinality of the treated group is smaller than or equal to the cardinality of the control group. When the cardinality of the treated group is equal to the cardinality of the control group, we prioritize the treated group to be the one with the smallest index among the units with positive weights.

The second method solves a constrained optimization problem, by converting it into the canonical form of a Quadratic Constraint Quadratic Program (QCQP), which we 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 $\tilde{\boldsymbol{W}} = (w_1, w_2, ..., w_J, v_1, v_2, ..., v_J)$.

Let M be the dimension of the predictors $\boldsymbol{X}_j$. Let X be an $M \times J$ matrix, each column of which is $\boldsymbol{X}_j$, which stands for the predictors of unit j.

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

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

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

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

OA.21

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 $P^1 = \{P^1_{k,l}\}_{k,l=1,...,2J} \in \mathbb{R}^{2J \times 2J}$ such that P^1 only has non-zero values in the two off-diagonal blocks, i.e., for any $k, l = 1, \ldots, 2J$,

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

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

$$\begin{aligned} \min \quad & \tilde{\boldsymbol{W}}' P^0 \tilde{\boldsymbol{W}} + \boldsymbol{q}^{0'} \tilde{\boldsymbol{W}} & \text{(OA.12)} \\ \text{s.t.} \quad & \boldsymbol{e}_1' \tilde{\boldsymbol{W}} = 1, \\ & \boldsymbol{e}_2' \tilde{\boldsymbol{W}} = 1, \\ & \tilde{\boldsymbol{W}}' P^1 \tilde{\boldsymbol{W}} = 0, \\ & \tilde{\boldsymbol{W}} \geq \boldsymbol{0}. \end{aligned}$$

The first computational method (enumeration) solves two synthetic control problems in each iteration. The synthetic control problems can be efficiently solved. We implement the synthetic control problem using the "lsei" function from "limSolve" package in R 4.0.2. For the second computational method (quadratic programming), the problem (OA.12) is implemented using Gurobi 9.0.2 in R 4.0.2. Since the QCQP is non-convex, the computation leads to some numerical errors up to 0.001 in finding the treated and control weights. So we round the treated and control weights to the nearest 2-digits in the implementation of the QCQP. Moreover, for all the weights that are less than or equal to 0.01, we trim the weights to zero. This is because smaller weights suffer from greater impacts of numerical errors, and that numerical errors could cause zero weights to be non-zero, thus having a non-negligible impact on the swapping rule.

To conclude, we compare the treated and control weights calculated from both methods. Both methods yield the same treated and control weights up to some negligible rounding error, while the first method takes longer computational time.

The other designs are computationally implemented using either one of the above two methods. The *Constrained* design is implemented using the enumeration method. In cases when the cardinality constraint $\overline{m}$ is small, this brute force enumeration is very efficient. The *Weakly-targeted* design is implemented using the quadratic programming method. In the QCQP formulation, the objective function has both a different quadratic term P^0 and a different linear term $\boldsymbol{q}^0$. The *Unit-Level* design is implemented using the enumeration method. The *Penalized* design is implemented using the quadratic programming method. In the QCQP formulation, the objective function has the same quadratic term P^0 and a different linear term $\boldsymbol{q}^0$.

OA.6. Additional Illustrations Using Walmart Data

In this section, we present results for $\overline{m} = 1$ and $\overline{m} = 3$. Using only one treated unit $(\overline{m} = 1)$ fails to produce a good fit between the treated and synthetic control unit in the fitting periods. For the case of $\overline{m} = 1$, Figures OA.1 and OA.2 reveal a substantial gap with a clear seasonal trend between the two synthetic units. Figures OA.3 and OA.4 report results for $\overline{m} = 3$. Increasing $\overline{m}$ from $\overline{m} = 2$ to $\overline{m} = 3$ results in a minor improvement in fit, and leaves estimation results substantively unchanged.
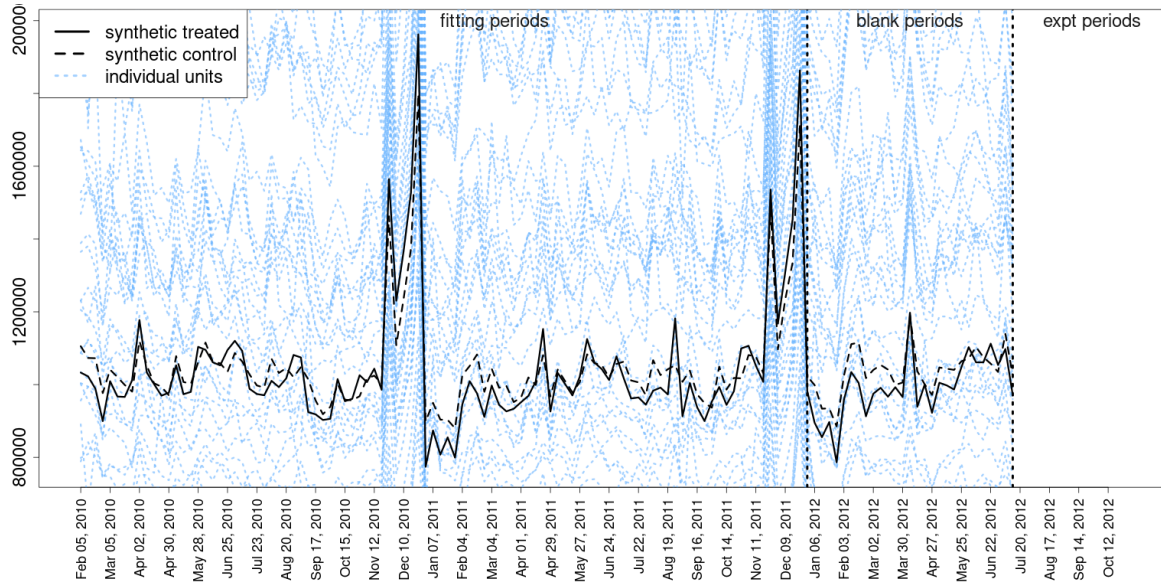
Figure OA.1: Synthetic Treated Unit and Synthetic Control Unit, $\overline{m} = 1$

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.
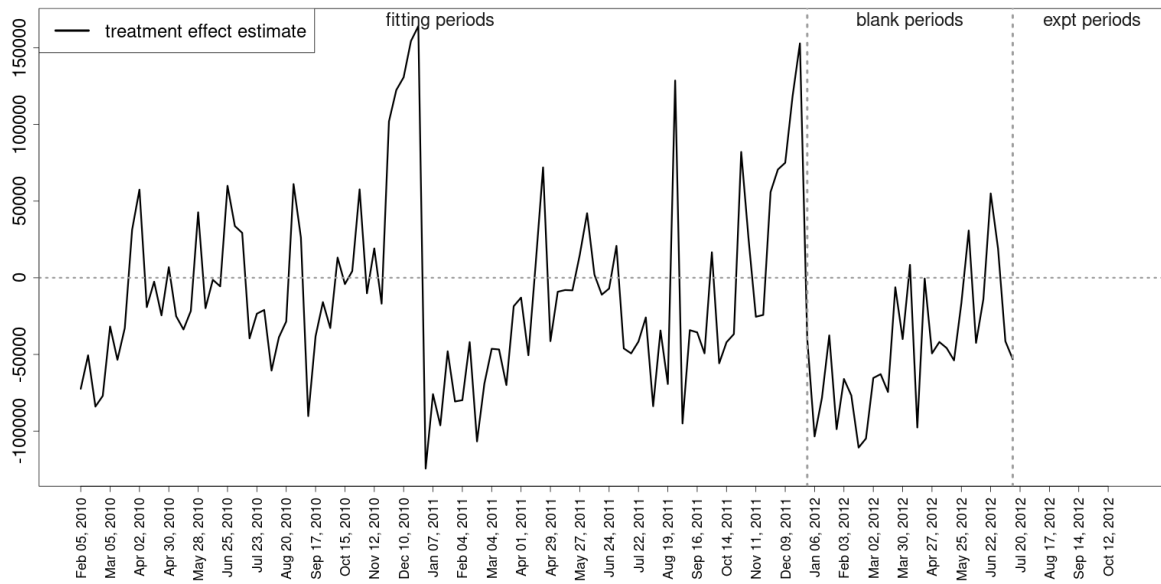
Figure OA.2: Placebo Treatment Effects, $\overline{m} = 1$

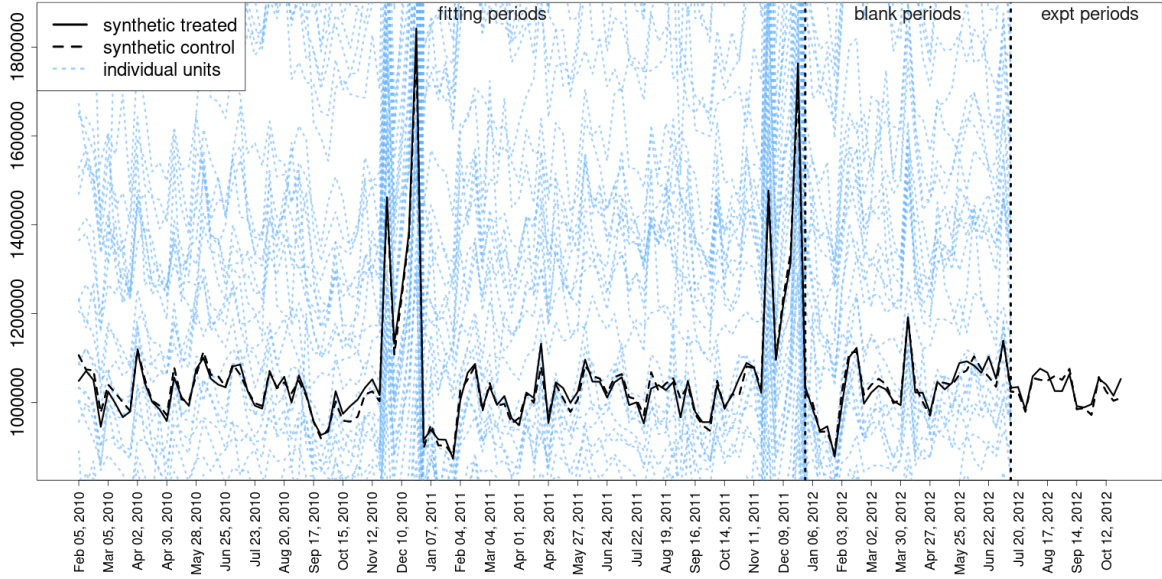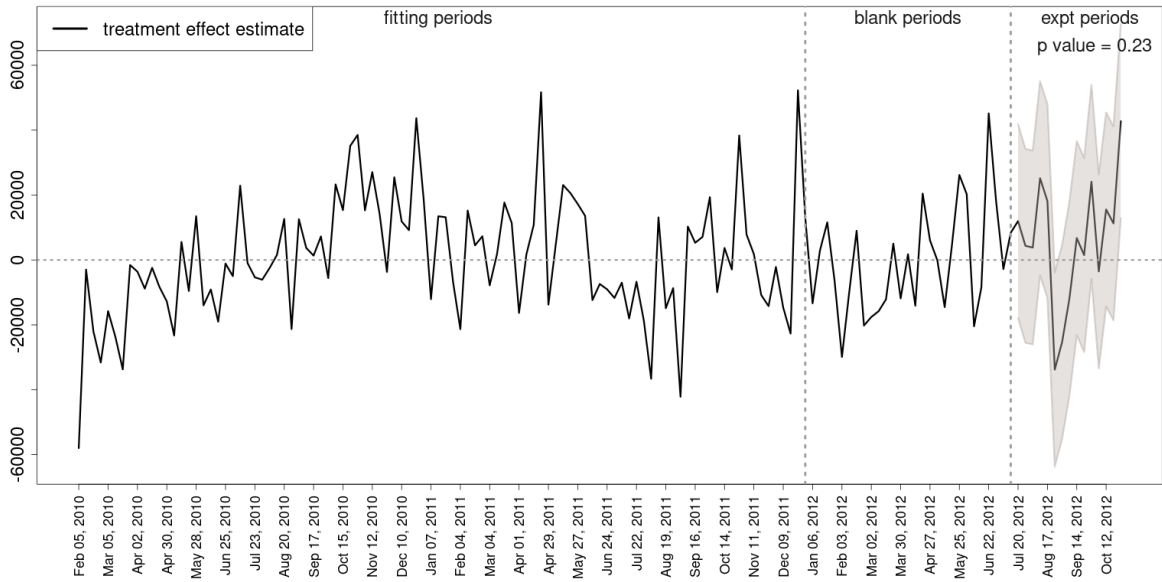*Note:* This figure reports the difference between the synthetic treated and synthetic control outcomes of Figure OA.1.

OA.25

Figure OA.3: Synthetic Treatment Unit and Synthetic Control Unit, when $\overline{m} = 3$.

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 OA.4: Treatment Effect Estimate, when $\overline{m} = 3$.

Note: This figure reports the difference between the synthetic treated and synthetic control outcomes of Figure OA.3. For the experimental periods, this is the treatment effect estimate.

OA.26

OA.7. Additional Simulation Results

In Section 5.1 the idiosyncratic shocks are i.i.d. Normal with variance $\sigma^2 = 1$. Figures OA.5 and OA.6 report results for $\sigma^2 = 5$ and $\sigma^2 = 10$, respectively. Figures OA.7 and OA.8 report differences between the outcomes for the synthetic treated and the synthetic control units for the same values for σ^2. As the value of σ^2 increases, the quality of the post-treatment estimation and inference deteriorates, and the p-value for the null hypotheses of in (17) increases. The deterioration in pre-treatment fit in Figures OA.5 and OA.6 provides a diagnosis of the accuracy of the respective estimates.
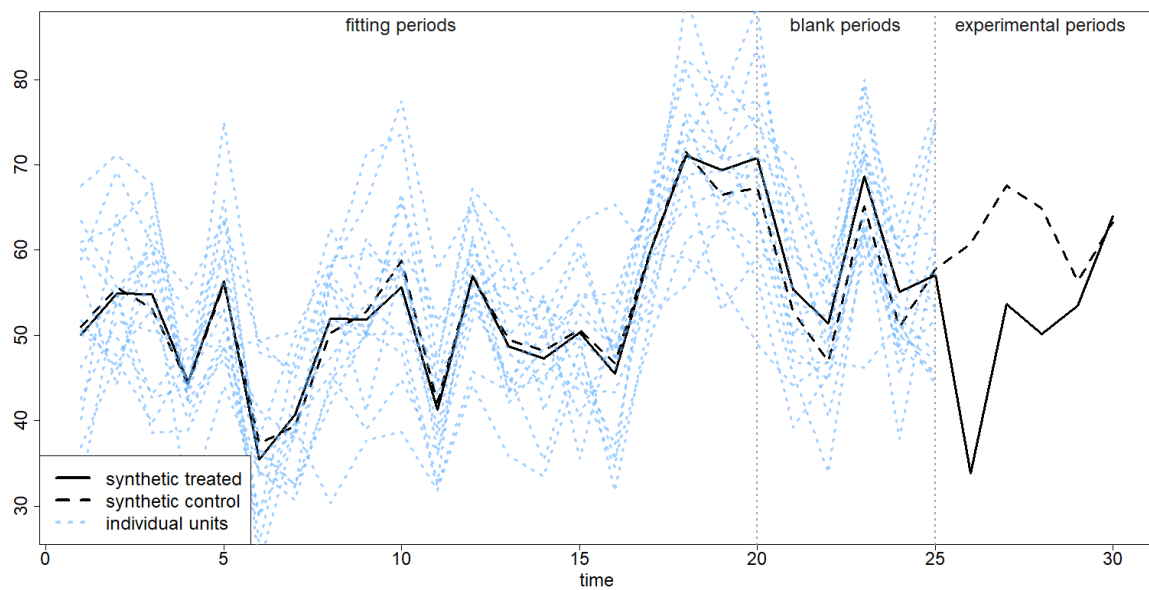
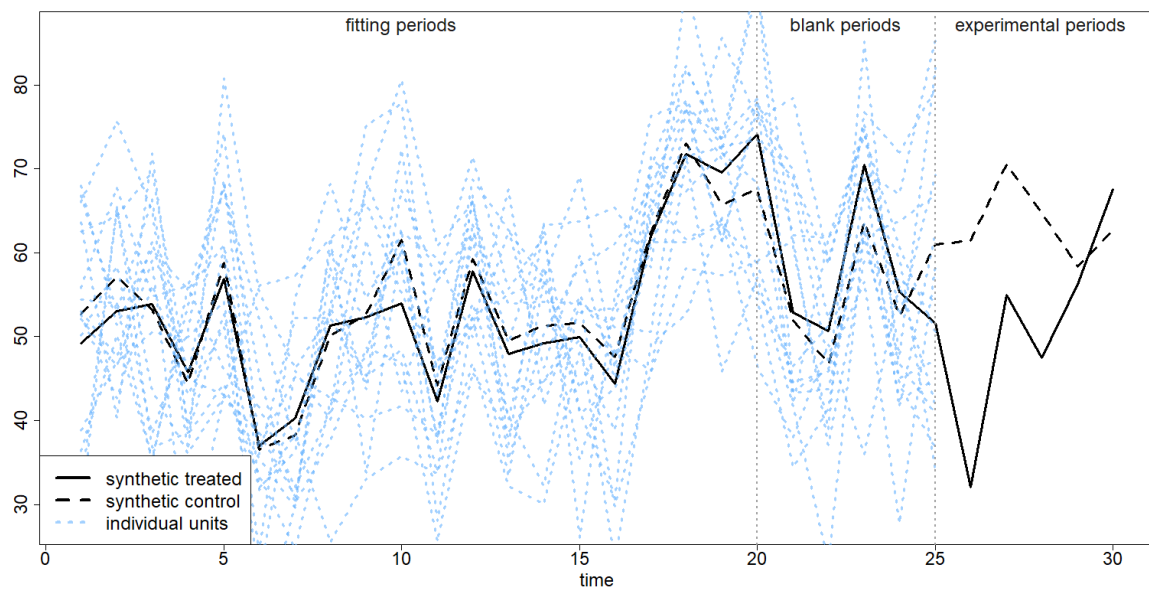Figure OA.5: Synthetic Treatment Unit and Synthetic Control Unit, when $\sigma^2 = 5$.

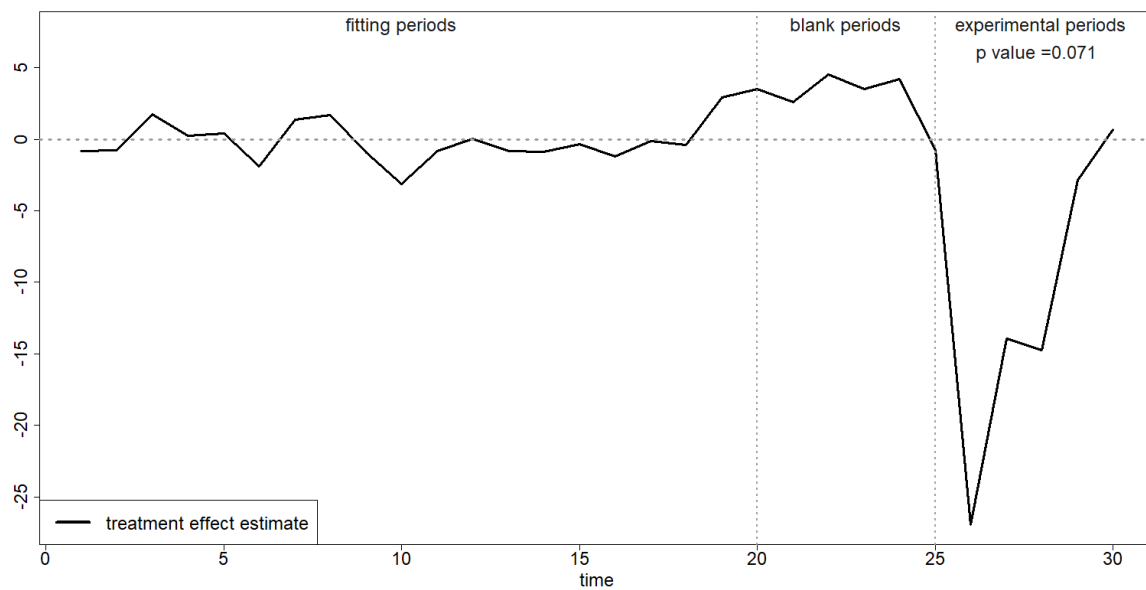Figure OA.6: Synthetic Treatment Unit and Synthetic Control Unit, when $\sigma^2 = 10$.

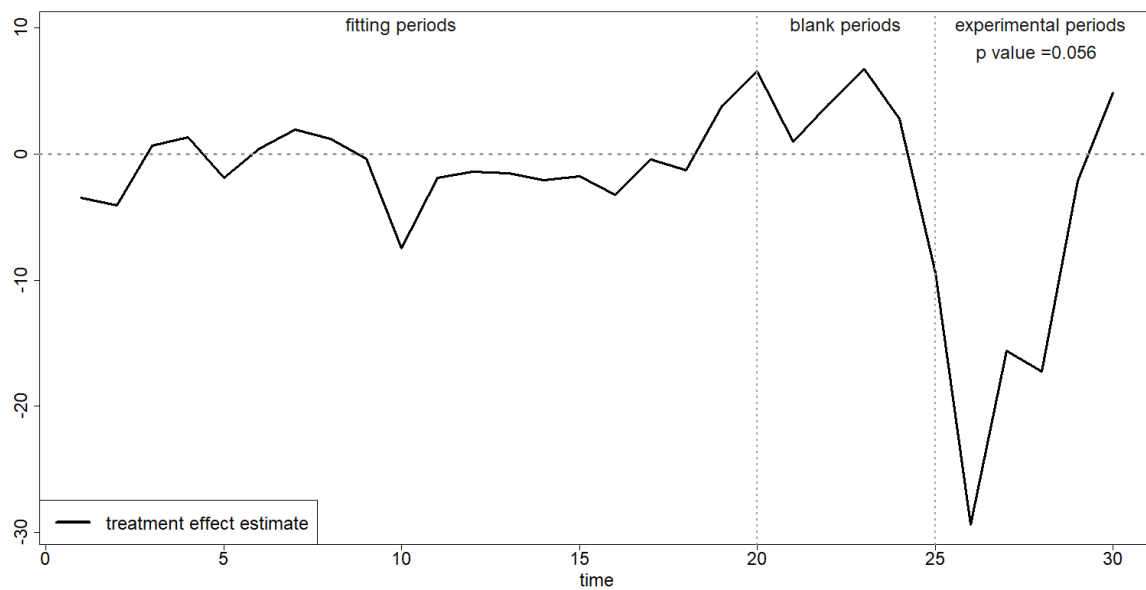Figure OA.7: Treatment Effect Estimate, when $\sigma^2 = 5$.

Figure OA.8: Treatment Effect Estimate, when $\sigma^2 = 10$.

OA.29