Online Appendix
Synthetic Controls for Experimental Design

Alberto Abadie and Jinglong Zhao

February 4, 2026

OA.1. Other Versions of the Synthetic Control Design

Recall that in Figure 1, we discuss how to take into account the clustered nature of the data so 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. We provide a formulation of the synthetic control design in this setting.

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

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

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 (10) is given by:

$$\begin{aligned}
\min_{\substack{w_j, \forall j=1,2,..,J,\\ v_{ij}, \forall i,j=1,2,...,J}} \quad & \sum_{k=1}^{K} \left(\sum_{j \in \mathcal{I}_k} f_j \right) \left\{ \left\| \overline{\boldsymbol{X}}_k - \sum_{j \in \mathcal{I}_k} w_j \boldsymbol{X}_j \right\|^2 + \xi \sum_{j \in \mathcal{I}_k} w_j \left\| \boldsymbol{X}_j - \sum_{i,j \in \mathcal{I}_k} v_{ij} \boldsymbol{X}_i \right\|^2 \right\} \\
\text{s.t.} \quad & \sum_{j \in \mathcal{I}_k} w_j = 1, \quad \forall k = 1, \ldots, K, \\
& w_j \geq 0, \quad \forall j = 1, \ldots, J, \\
& \sum_{i=1}^{J} v_{ij} = 1, \quad \forall j \in \mathcal{J}_{\boldsymbol{w}} \\
& v_{ij} \geq 0, \quad \forall j \in \mathcal{J}_{\boldsymbol{w}}, \ i = 1, \ldots, J,
\end{aligned}$$

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

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

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

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

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 (7), or $\boldsymbol{X}_j$ for a treated unit in equation (10)) 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 (7), we obtain

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

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

$$\underline{m} \leq \|\boldsymbol{w}\|_0 \leq \overline{m}. \tag{OA.1}$$

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. See Abadie and L'Hour (2021) for details on penalized synthetic control estimators. In Section OA.2 below, we discuss how to apply the Abadie and L'Hour penalty to the other synthetic designs proposed in this article.

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 (7)). 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. In Section OA.2 below we provide details on how to apply bias correction techniques in a synthetic control design.

OA.2. Designs Based on Penalized and Bias-corrected Synthetic Control Methods

Consider the design problem in (10),

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

To apply the penalized synthetic control method of Abadie and L'Hour (2021) to this design, we replace the term (a) in (OA.2) with

$$\left\| \overline{\boldsymbol{X}} - \sum_{j=1}^{J} w_j \boldsymbol{X}_j \right\|^2 + \lambda_1 \sum_{j=1}^{J} w_j \|\overline{\boldsymbol{X}} - \boldsymbol{X}_j\|^2, \tag{OA.3}$$

and the terms (b) with

$$\left\| \boldsymbol{X}_j - \sum_{i=1}^{J} v_{ij} \boldsymbol{X}_i \right\|^2 + \lambda_2 \sum_{i=1}^{J} v_{ij} \|\boldsymbol{X}_j - \boldsymbol{X}_i\|^2. \tag{OA.4}$$

Here, λ_1 and λ_2 are positive constants that penalize discrepancies between the target values of the predictors ($\overline{\boldsymbol{X}}$ in (OA.3) and $\boldsymbol{X}_j$ in (OA.4)) 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.2). These terms can be adapted as in (OA.3) and (OA.4) 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.

OA.3. Approximate Validity when $\boldsymbol{\lambda}_t$ 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 (17) 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 (17) 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* (15)*, the* p*-values of equation* (17) *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*

$$\alpha - 2z_2 - \frac{1}{|\Pi|} \leq \Pr(\widehat{p} \leq \alpha) \leq \alpha + 2z_2, \tag{OA.5}$$

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

$$\Pr(\mathcal{C}) \geq 1 - 2J \exp\left(-\frac{z_1^2 \underline{\zeta}^2}{8\overline{\sigma}^2 \overline{\lambda}^4 F^2} T_{\mathcal{E}} \right) - \frac{z_1}{z_2} 4e \sqrt{2J(\min\{T - T_0, T_0 - T_{\mathcal{E}}\})^3}\, \kappa, \tag{OA.6}$$

where z_1, z_2 *are arbitrary positive choice parameters. In expression* (OA.6)*, the probability* $\Pr(\mathcal{C})$ *is over the distribution of* $\{\epsilon_{jt}\}_{j \in \{1,\ldots,J\}, t \in \{1,\ldots,T\}}$ *and* $\{\xi_{jt}\}_{j \in \{1,\ldots,J\}, t \in \{T_0+1,\ldots,T\}}$*. In expression* (OA.5)*, the probability* $\Pr(\widehat{p} \leq \alpha)$ *is over the distribution of* $\{\epsilon_{jt}\}_{j \in \{1,\ldots,J\}, t \in \{1,\ldots,T\}}$ *and* $\{\xi_{jt}\}_{j \in \{1,\ldots,J\}, t \in \{T_0+1,\ldots,T\}}$*, conditional on event* $\mathcal{C}$*.*

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 $T_{\mathcal{E}}$. Large $T_{\mathcal{E}}$ allows choices for z_1 and z_2 such that the bounds in (OA.5) are tight and the probability $\Pr(\mathcal{C})$ in (OA.6) is close to one.

We prove Theorem OA.1 in Section OA.7.4.

OA.4. 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)*

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

and (ii)

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

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,*

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

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

Theorem OA.2 *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 $t \geq T_0 + 1$,

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

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

We prove Theorem OA.2 in Section OA.7.5. Next, we provide the following result on inference.

Theorem OA.3 *Suppose that Assumptions 1, 2(ii) and 5(i) hold, and the noises* $\{\epsilon_{jt}\}_{t\in\mathcal{B}\cup\{T_0+1,\ldots,T\}}$ *and* $\{\xi_{jt}\}_{t\in\{T_0+1,\ldots,T\}}$ *have continuous distributions. Assume that* $\{\boldsymbol{\lambda}_t\}_{t\in\mathcal{B}\cup\{T_0+1,\ldots,T\}}$ *is a sequence of exchangeable random variables. Under the null hypothesis* (15), *for any* $\alpha \in [0, 1]$, *we have*

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

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

We prove Theorem OA.3 in Section OA.7.6.

OA.5. 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.6. Implementations of Different Optimization Formulations

To computationally solve (7), i.e., the *Unconstrained* design, we propose two methods. The first method is by enumeration, which takes advantage of the objective function of (7) 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 (7) 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 $k, l = 1, \ldots, 2J$,

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

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

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

Further define $\boldsymbol{e}_1 = (1, 1, ..., 1, 0, 0, ..., 0)'$ whose first J elements are 1 and last J elements 0; and $\boldsymbol{e}_2 = (0, 0, ..., 0, 1, 1, ..., 1)'$ whose first J elements are 0 and last J elements 1.

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

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

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

$$\begin{aligned} \min \quad & \tilde{\boldsymbol{W}}' P^0 \tilde{\boldsymbol{W}} + \boldsymbol{q}^{0'} \tilde{\boldsymbol{W}} \\ \text{s.t.} \quad & \boldsymbol{e}_1' \tilde{\boldsymbol{W}} = 1, \\ & \boldsymbol{e}_2' \tilde{\boldsymbol{W}} = 1, \\ & \tilde{\boldsymbol{W}}' P^1 \tilde{\boldsymbol{W}} = 0, \\ & \tilde{\boldsymbol{W}} \geq \boldsymbol{0}. \end{aligned} \tag{OA.7}$$

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.7) 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 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.

All the different versions of the synthetic control design are computationally implemented using either one of the above two methods. The *Unconstrained* design is implemented using the quadratic programming method. 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.7. Proofs

OA.7.1. Proof of Theorem 1

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

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

$$= \left(\sum_{j=1}^{J} w_j^* Y_{jt}^I - \sum_{j=1}^{J} f_j Y_{jt}^I\right) - \left(\sum_{j=1}^{J} v_j^* Y_{jt}^N - \sum_{j=1}^{J} f_j Y_{jt}^N\right). \quad \text{(OA.8)}$$

The first term in (OA.8) 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

$$\sum_{j=1}^{J} w_j^* Y_{jt}^I - \sum_{j=1}^{J} f_j Y_{jt}^I = \boldsymbol{\gamma}_t'\left(\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} f_j \boldsymbol{Z}_j\right) \\ + \boldsymbol{\eta}_t'\left(\sum_{j=1}^{J} w_j^* \boldsymbol{\mu}_j - \sum_{j=1}^{J} f_j \boldsymbol{\mu}_j\right) + \left(\sum_{j=1}^{J} w_j^* \xi_{jt} - \sum_{j=1}^{J} f_j \xi_{jt}\right) \quad \text{(OA.9)}$$

Similarly, using expression (12a), we obtain

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

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

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

$$+\boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\left(\sum_{j=1}^{J} w_j^*\boldsymbol{\epsilon}_j^{\mathcal{E}} - \sum_{j=1}^{J} f_j\boldsymbol{\epsilon}_j^{\mathcal{E}}\right). \tag{OA.10}$$

Equations (OA.9) and (OA.10) imply

$$\begin{aligned}\sum_{j=1}^{J} w_j^* Y_{jt}^I - \sum_{j=1}^{J} f_j Y_{jt}^I = {} & (\boldsymbol{\gamma}_t' - \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\theta}_{\mathcal{E}})\left(\sum_{j=1}^{J} w_j^*\boldsymbol{Z}_j - \sum_{j=1}^{J} f_j\boldsymbol{Z}_j\right)\\ & + \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\left(\sum_{j=1}^{J} w_j^*\boldsymbol{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} f_j\boldsymbol{Y}_j^{\mathcal{E}}\right)\\ & - \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\sum_{j=1}^{J} w_j^*\boldsymbol{\epsilon}_j^{\mathcal{E}}\\ & + \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\sum_{j=1}^{J} f_j\boldsymbol{\epsilon}_j^{\mathcal{E}}\\ & + \left(\sum_{j=1}^{J} w_j^*\xi_{jt} - \sum_{j=1}^{J} f_j\xi_{jt}\right).\end{aligned} \tag{OA.11}$$

If Assumption 3 holds, (OA.11) becomes

$$\begin{aligned}\sum_{j=1}^{J} w_j^* Y_{jt}^I - \sum_{j=1}^{J} f_j Y_{jt}^I = {} & -\boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\sum_{j=1}^{J} w_j^*\boldsymbol{\epsilon}_j^{\mathcal{E}}\\ & + \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\sum_{j=1}^{J} f_j\boldsymbol{\epsilon}_j^{\mathcal{E}}\\ & + \left(\sum_{j=1}^{J} w_j^*\xi_{jt} - \sum_{j=1}^{J} f_j\xi_{jt}\right).\end{aligned} \tag{OA.12}$$

Only the first term on the right-hand side of (OA.12) 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 (OA.8), we obtain

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

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

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

Similarly,

$$\left(\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_s\right)^2 \leq \left(\frac{\overline{\lambda}^2 F}{T_{\mathcal{E}}\underline{\zeta}}\right)^2. \tag{OA.13}$$

Let

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

Because $\overline{\epsilon}_{jt}^{\mathcal{E}}$ is a linear combination of independent sub-Gaussians with variance proxy $\overline{\sigma}^2$, it follows that $\overline{\epsilon}_{jt}^{\mathcal{E}}$ 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

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

An analogous argument yields

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

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 (OA.11). 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 $\overline{\gamma} + \overline{\theta}\frac{\overline{\lambda}\overline{\eta}F}{\underline{\zeta}}$. Cauchy–Schwarz inequality and Assumption 4 imply

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

and

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

Combining the last two displayed equations with (OA.11), we have

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

An analogous derivation produces

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

which finishes the proof of the theorem. ∎

OA.7.2. 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 $t \in \mathcal{B} \cup \{T_0 + 1, \dots, T\}$. For $t \in \{T_0 + 1, \dots, T\}$, $\widehat{u}_t$ 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 $t \in \{T_0{+}1, \dots, T\}$. The null hypothesis (15) and Assumption 2(ii) imply that $\{u_t\}_{t \in \mathcal{B} \cup \{T_0+1, \dots, T\}}$ is a sequence of exchangeable random variables. Additionally, Assumption 1 and the null hypothesis (15) imply

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

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

for $t \in \mathcal{B} \cup \{T_0+1, \ldots, T\}$, where the last equality is due to Assumption 3(i). So $\{\widehat{u}_t\}_{t \in \mathcal{B} \cup \{T_0+1,\ldots,T\}}$ is a sequence of exchangeable random variables.

Recall that, 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 $|\pi| = T - T_0$. Recall that

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

and

$$S(\widehat{\boldsymbol{u}}_\pi) = \frac{1}{T - T_0} \sum_{t \in \pi} |\widehat{u}_t|.$$

Now define $k = |\Pi| - \lfloor \alpha |\Pi| \rfloor$. Define $S^{(k)}(\widehat{\boldsymbol{u}})$ to be the k-th smallest value in a small-to-large rearrangement of $\{S(\widehat{\boldsymbol{u}}_\pi)\}_{\pi \in \Pi}$. Because the noises $\{\epsilon_{jt}\}_{t \in \mathcal{B} \cup \{T_0+1,\ldots,T\}}$ and $\{\xi_{jt}\}_{t \in \{T_0+1,\ldots,T\}}$ have continuous distributions, $\{S(\widehat{\boldsymbol{u}}_\pi)\}_{\pi \in \Pi}$ are all unique with probability 1.

Using the above definitions, for any α,

$$\mathbb{1}\{\widehat{p} \leq \alpha\} = \mathbb{1}\{S(\widehat{\boldsymbol{u}}) > S^{(k)}(\widehat{\boldsymbol{u}})\}.$$

Note that for any $\pi \in \Pi$, we have $S^{(k)}(\widehat{\boldsymbol{u}}_\pi) = S^{(k)}(\widehat{\boldsymbol{u}})$. Then we have

$$\sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{\boldsymbol{u}}_\pi) > S^{(k)}(\widehat{\boldsymbol{u}}_\pi)\} = \sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{\boldsymbol{u}}_\pi) > S^{(k)}(\widehat{\boldsymbol{u}})\} = |\Pi| - k = \lfloor \alpha |\Pi| \rfloor.$$

Because $\{\widehat{u}_t\}_{t \in \mathcal{B} \cup \{T_0+1,\ldots,T\}}$ is a sequence of exchangeable random variables, $\mathbb{1}\{S(\widehat{\boldsymbol{u}}) > S^{(k)}(\widehat{\boldsymbol{u}})\}$

has the same distribution as $\mathbb{1}\{S(\widehat{\boldsymbol{u}}_\pi) > S^{(k)}(\widehat{\boldsymbol{u}}_\pi)\}$ for any $\pi \in \Pi$. So we have

$$\begin{aligned}\Pr(\widehat{p} \le \alpha) &= E[\mathbb{1}\{\widehat{p} \le \alpha\}] = E\Big[\mathbb{1}\{S(\widehat{\boldsymbol{u}}) > S^{(k)}(\widehat{\boldsymbol{u}})\}\Big] \\ &= E\Big[\frac{1}{|\Pi|}\sum_{\pi\in\Pi}\mathbb{1}\{S(\widehat{\boldsymbol{u}}_\pi) > S^{(k)}(\widehat{\boldsymbol{u}}_\pi)\}\Big] = \frac{\lfloor \alpha|\Pi| \rfloor}{|\Pi|}.\end{aligned}$$

Note that $\alpha|\Pi| - 1 \le \lfloor \alpha|\Pi| \rfloor \le \alpha|\Pi|$. This implies

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

∎

OA.7.3. Proof of Theorem 3

OA.7.3.1. A Technical Lemma

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

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

Lemma OA.4 *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 $q \in \mathbb{R}$,*

$$\left|\frac{1}{T_0 - T_\mathcal{E}}\sum_{t\in\mathcal{B}}\mathbb{1}\Big\{\Big|\sum_{j=1}^J w_j^* Y_{jt} - \sum_{j=1}^J v_j^* Y_{jt}\Big| \le q\Big\} - P_{\mathcal{E},q}\right| \le \epsilon_\mathcal{B}. \tag{OA.14}$$

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 $t \in \{T_0+1, ..., T\}$,

$$\left| \Pr\left(\Big| \sum_{j=1}^J w_j^* Y_{jt} - \sum_{j=1}^J v_j^* Y_{jt} - \tau_t \Big| \le q \right) - P_{\mathcal{E},q} \right| \le \epsilon_\mathcal{T}. \tag{OA.15}$$

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 $\big| \sum_{j=1}^J w_j^* \epsilon_{j*} - \sum_{j=1}^J v_j^* \epsilon_{j*} \big|$ has a continuous distribution. Then, for any $\alpha \in (0,1)$ and any $t \in \{T_0+1, ..., T\}$,

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

Note that Lemma OA.4 does not require Assumptions 1–3. But for Conditions (OA.14) and (OA.15) to hold, we will apply Assumptions 1–3. To prove Lemma OA.4, 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),

$$q_{1-\alpha} = \inf_{z \in \mathbb{R}} \left\{ \Pr\left(\Big| \sum_{j=1}^J w_j^* \epsilon_{j*} - \sum_{j=1}^J v_j^* \epsilon_{j*} \Big| \le z \right) \ge 1 - \alpha \right\} \tag{OA.16}$$

Intuitively, $\widehat{q}_{1-\alpha}$ as defined in (18) approximates $q_{1-\alpha}$ as defined in (OA.16).

Proof of Lemma OA.4. 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 (OA.14) to show that conditional on event $\mathcal{C}_{\mathcal{B}}$,

$$\begin{aligned}
\frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\Big\{ \Big| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \Big| \leq q_{1-\alpha-\epsilon_{\mathcal{B}}} - \frac{1}{k} \Big\} & \\
\leq & \Pr\left(\Big| \sum_{j=1}^{J} w_j^* \epsilon_{j*} - \sum_{j=1}^{J} v_j^* \epsilon_{j*} \Big| \leq q_{1-\alpha-\epsilon_{\mathcal{B}}} - \frac{1}{k} \right) + \epsilon_{\mathcal{B}} \\
< & 1 - \alpha - \epsilon_{\mathcal{B}} + \epsilon_{\mathcal{B}} \\
= & 1 - \alpha \\
\leq & \frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\Big\{ \Big| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \Big| \leq \widehat{q}_{1-\alpha} \Big\},
\end{aligned}$$

where the first inequality is due to Condition (OA.14); the second inequality is due to the infimum part of (OA.16) (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 (OA.16) is greater or equal to $1-\alpha$); the last inequality is due to the definition of $\widehat{q}_{1-\alpha}$ in (18).

The above inequality suggests that for any $k \in \mathbb{N}$, the event

$$\begin{aligned}
\mathcal{E}_k^{(\leq)} = \Bigg\{ & \frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\Big\{ \Big| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \Big| \leq q_{1-\alpha-\epsilon_{\mathcal{B}}} - \frac{1}{k} \Big\} \\
& \leq \frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\Big\{ \Big| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \Big| \leq \widehat{q}_{1-\alpha} \Big\} \Bigg\}
\end{aligned}$$

happens conditional on event $\mathcal{C}_{\mathcal{B}}$. Since the left hand side of the inequality inside event $\mathcal{E}_k^{(\leq)}$, which is $\frac{1}{T_0-T_{\mathcal{E}}} \sum_{t\in\mathcal{B}} \mathbb{1}\Big\{ \Big| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \Big| \leq q_{1-\alpha-\epsilon_{\mathcal{B}}} - \frac{1}{k} \Big\}$, 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

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

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 OA.4, 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,

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

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 (OA.15); the last inequality is due to the definition of $q_{1-\alpha-\epsilon_{\mathcal{B}}}$ in (OA.16).

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 (OA.15) to show that conditional on event $\mathcal{C}_{\mathcal{B}}$,

$$
\begin{aligned}
\frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t \in \mathcal{B}} \mathbb{1}\Big\{\Big| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \Big| \leq q_{1-\alpha+\epsilon_{\mathcal{B}}} \Big\} & \\
\geq & \Pr\left(\Big| \sum_{j=1}^{J} w_j^* \epsilon_{j*} - \sum_{j=1}^{J} v_j^* \epsilon_{j*} \Big| \leq q_{1-\alpha+\epsilon_{\mathcal{B}}} \right) - \epsilon_{\mathcal{B}} \\
\geq & 1 - \alpha + \epsilon_{\mathcal{B}} - \epsilon_{\mathcal{B}} \\
= & 1 - \alpha,
\end{aligned}
$$

where the first inequality is due to Condition (OA.14); the second inequality is due to the definition of $q_{1-\alpha+\epsilon_{\mathcal{B}}}$ in (OA.16);

Due to (18), 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 OA.4, 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,

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

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 (OA.15); the last inequality is due to the definition of $q_{1-\alpha+\epsilon_{\mathcal{B}}}$ in (OA.16). ∎

OA.7.3.2. Proof of Theorem 3

In this section, we use Lemma OA.4 to prove Theorem 3. Instead of proving exactly Theorem 3, we prove Theorem OA.5 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 OA.5 *Assume that Assumptions 1– 3 hold. Assume there exists a constant $\kappa < \infty$, such that for all $j = 1, \ldots, J$, $t = 1, \ldots, T$, ϵ_{jt} are continuously distributed with the probability density function upper bounded by κ. Assume that for $t = T_0+1, \ldots, T$, and $j = 1, \ldots, J$, ξ_{jt} has the same distribution as ϵ_{jt}. Then the confidence interval defined in* (19) *approximately achieves point-wise coverage, i.e., for any $\alpha \in (0,1)$ and any $t \in \{T_0+1, ..., T\}$,*

$$\left|\Pr\left(\tau_t \in \widehat{C}_{1-\alpha}(Y_{1t}, Y_{2t}, ..., Y_{Jt})\right) - (1-\alpha)\right|$$
$$\le \sqrt{\frac{1}{2(T_0 - T_{\mathcal{E}})}\log\left(\frac{2}{z_{\mathcal{B}}}\right)} + \kappa\sqrt{\frac{8eJ\overline{\sigma}^2\overline{\lambda}^2\overline{\eta}^2F^2}{\underline{\zeta}^2 T_{\mathcal{E}}}\log\left(\frac{2J}{z_{\mathcal{E}}}\right)} + 2\kappa\sqrt{\frac{8eJ\overline{\sigma}^2\overline{\lambda}^4F^2}{\underline{\zeta}^2 T_{\mathcal{E}}}\log\left(\frac{2J}{z_{\mathcal{E}}}\right)} + z_{\mathcal{B}} + 3z_{\mathcal{E}}.$$

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

Proof of Theorem OA.5. We outline the proof of Theorem OA.5 as follows. We first define four events. We then check Conditions (OA.14) under the first two events and (OA.15) under the last two events. Finally, we apply Lemma OA.4 and conclude the proof.

Step 1: We define the following four events. First, in the blank periods,

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

To analyze the event $\mathcal{E}_1$, condition on the weights $(\boldsymbol{w}^*, \boldsymbol{v}^*)$ obtained from the fitting periods and consider $\Pr(\mathcal{E}_1 \mid \boldsymbol{w}^*, \boldsymbol{v}^*)$.

For any $t \in \mathcal{B}$ in the blank periods, once we condition on $(\boldsymbol{w}^*, \boldsymbol{v}^*)$, the indicator

$$\mathbb{1}\Big\{\Big|\sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt}\Big| \leq q\Big\}$$

is a random variable whose only source of randomness comes from the idiosyncratic noises $\{\epsilon_{jt}\}_{j=1}^{J}$ in period t. Since for any $t \neq t'$, the noise vectors $\{\epsilon_{jt}\}_{j=1}^{J}$ and $\{\epsilon_{jt'}\}_{j=1}^{J}$ are independent, the corresponding indicators

$$\mathbb{1}\Big\{\Big|\sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt}\Big| \leq q\Big\} \quad \text{and} \quad \mathbb{1}\Big\{\Big|\sum_{j=1}^{J} w_j^* Y_{jt'} - \sum_{j=1}^{J} v_j^* Y_{jt'}\Big| \leq q\Big\}$$

are also independent. By Hoeffding's inequality for bounded random variables, the event $\mathcal{E}_1$ under the conditioning on $(\boldsymbol{w}^*, \boldsymbol{v}^*)$ occurs with probability

$$\Pr(\mathcal{E}_1 \mid \boldsymbol{w}^*, \boldsymbol{v}^*) \;\geq\; 1 - z_{\mathcal{B}}.$$

Using the law of total probability, because the inequality holds for any weights $(\boldsymbol{w}^*, \boldsymbol{v}^*)$, we conclude that, unconditionally, the event $\mathcal{E}_1$ happens with probability $\Pr(\mathcal{E}_1) \geq 1 - z_{\mathcal{B}}$.

Second, in the blank periods,

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

Note that,

$$
\begin{aligned}
&\max_{t\in\mathcal{B}}\left|\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_\mathcal{E}'\boldsymbol{\lambda}_\mathcal{E})^{-1}\boldsymbol{\lambda}_\mathcal{E}'\sum_{j=1}^J(w_j^*-v_j^*)\boldsymbol{\epsilon}_j^\mathcal{E}\right| \\
&\qquad\le \max_{t\in\mathcal{B}}\sum_{j=1}^J|w_j^*-v_j^*|\sum_{s\in\mathcal{E}}|\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_\mathcal{E}'\boldsymbol{\lambda}_\mathcal{E})^{-1}\boldsymbol{\lambda}_s||\epsilon_{js}| \\
&\qquad\le \sum_{j=1}^J|w_j^*-v_j^*|\sum_{s\in\mathcal{E}}\frac{\overline{\lambda}^2F}{T_\mathcal{E}\underline{\zeta}}|\epsilon_{js}|,
\end{aligned}
$$

where the second inequality is due to (OA.13), and because $|w_j^*-v_j^*|\ge 0$ and $|\epsilon_{js}|\ge 0$. Therefore,

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

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

Third, in the experimental periods,

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

Note that,

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

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

where the second inequality is due to (OA.13), and because $|w_j^* - f_j| \geq 0$ and $|\epsilon_{js}| \geq 0$. Therefore,

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

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,

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

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

Step 2: Now we check Conditions (OA.14) and (OA.15). We first check Condition (OA.14). In the statement of Condition (OA.14), let $\mathcal{C}_\mathcal{B} = \mathcal{E}_1 \cap \mathcal{E}_2$. Note that

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

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

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

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

$$\begin{aligned}\Big| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \Big| &= \Big| \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} - \boldsymbol{\lambda}_t' (\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} (w_j^* - v_j^*) \boldsymbol{\epsilon}_j^{\mathcal{E}} \Big| \\ &\leq \Big| \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \Big| + \Big| \boldsymbol{\lambda}_t' (\boldsymbol{\lambda}_{\mathcal{E}}' \boldsymbol{\lambda}_{\mathcal{E}})^{-1} \boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} (w_j^* - v_j^*) \boldsymbol{\epsilon}_j^{\mathcal{E}} \Big| \\ &\leq \Big| \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \Big| + \sqrt{\frac{8 \overline{\sigma}^2 \overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log \left(\frac{2J}{z_{\mathcal{E}}} \right)}\end{aligned}$$

From the above inequality, for any $t \in \mathcal{B}$, due to Lemma OA.7-1 and Lemma OA.8, the probability density of $\big| \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \big|$ 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 $q \in \mathbb{R}$,

$$\begin{aligned}&\left| \Pr \left(\Big| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \Big| \leq q \right) - \Pr \left(\Big| \sum_{j=1}^{J} w_j^* \epsilon_{j*} - \sum_{j=1}^{J} v_j^* \epsilon_{j*} \Big| \leq q \right) \right| \\ =& \left| \Pr \left(\Big| \sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} \Big| \leq q \right) - \Pr \left(\Big| \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \Big| \leq q \right) \right| \\ \leq& \kappa \sqrt{\frac{8 e J \overline{\sigma}^2 \overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}} \log \left(\frac{2J}{z_{\mathcal{E}}} \right)}.\end{aligned}$$

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

$$\left|\frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t\in\mathcal{B}} \Pr\left(\Big|\sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt}\Big| \leq q\right) - \Pr\left(\Big|\sum_{j=1}^{J} w_j^* \epsilon_{j*} - \sum_{j=1}^{J} v_j^* \epsilon_{j*}\Big| \leq q\right)\right| \leq \kappa\sqrt{\frac{8eJ\overline{\sigma}^2\overline{\lambda}^4F^2}{\underline{\zeta}^2 T_{\mathcal{E}}}\log\left(\frac{2J}{z_{\mathcal{E}}}\right)}.$$

To conclude checking Condition (OA.14), 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 $q \in \mathbb{R}$,

$$\left|\frac{1}{T_0 - T_{\mathcal{E}}} \sum_{t\in\mathcal{B}} \mathbb{1}\left\{\Big|\sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt}\Big| \leq q\right\} - \Pr\left(\Big|\sum_{j=1}^{J} w_j^* \epsilon_{j*} - \sum_{j=1}^{J} v_j^* \epsilon_{j*}\Big| \leq q\right)\right| \leq \sqrt{\frac{1}{2(T_0 - T_{\mathcal{E}})}\log\left(\frac{2}{z_{\mathcal{B}}}\right)} + \kappa\sqrt{\frac{8eJ\overline{\sigma}^2\overline{\lambda}^4F^2}{\underline{\zeta}^2 T_{\mathcal{E}}}\log\left(\frac{2J}{z_{\mathcal{E}}}\right)}.$$

We then move on to check Condition (OA.15). In the statement of Condition (OA.15), 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

$$\begin{aligned}
&\sum_{j=1}^{J} w_j^* Y_{jt} - \sum_{j=1}^{J} v_j^* Y_{jt} - \tau_t \\
=&\sum_{j=1}^{J} (w_j^* - f_j) Y_{jt}^I - \sum_{j=1}^{J} (v_j^* - f_j) Y_{jt}^N \\
=&\sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} - \boldsymbol{\eta}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} (w_j^* - f_j)\boldsymbol{\epsilon}_j^{\mathcal{E}} + \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}' \sum_{j=1}^{J} (v_j^* - f_j)\boldsymbol{\epsilon}_j^{\mathcal{E}}.
\end{aligned}$$

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,

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

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 $q \in \mathbb{R}$,

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

Step 3: Now we apply Lemma OA.4. 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 OA.4,

$$\left|\Pr\left(\Big|\sum_{j=1}^J w_j^* Y_{jt} - \sum_{j=1}^J v_j^* Y_{jt} - \tau_t\Big| \leq \widehat{q}_{1-\alpha}\right) - (1-\alpha)\right|$$
$$\leq \sqrt{\frac{1}{2(T_0 - T_{\mathcal{E}})}\log\left(\frac{2}{z_{\mathcal{B}}}\right)} + \kappa\sqrt{\frac{8eJ\overline{\sigma}^2\overline{\lambda}^2\overline{\eta}^2 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}}\log\left(\frac{2J}{z_{\mathcal{E}}}\right)} + 2\kappa\sqrt{\frac{8eJ\overline{\sigma}^2\overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}}\log\left(\frac{2J}{z_{\mathcal{E}}}\right)} + z_{\mathcal{B}} + 3z_{\mathcal{E}}.$$

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

$$\left|\Pr\left(\tau_t \in \widehat{C}_{1-\alpha}(Y_{1t}, Y_{2t}, ..., Y_{Jt})\right) - (1-\alpha)\right|$$
$$\leq \sqrt{\frac{1}{2(T_0 - T_{\mathcal{E}})}\log\left(\frac{2}{z_{\mathcal{B}}}\right)} + \kappa\sqrt{\frac{8eJ\overline{\sigma}^2\overline{\lambda}^2\overline{\eta}^2 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}}\log\left(\frac{2J}{z_{\mathcal{E}}}\right)} + 2\kappa\sqrt{\frac{8eJ\overline{\sigma}^2\overline{\lambda}^4 F^2}{\underline{\zeta}^2 T_{\mathcal{E}}}\log\left(\frac{2J}{z_{\mathcal{E}}}\right)} + z_{\mathcal{B}} + 3z_{\mathcal{E}}.$$

∎

OA.7.4. Proof of Theorem OA.1

OA.7.4.1. 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

$$u_t = \sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \tag{OA.17}$$

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

$$u_t = \sum_{j=1}^{J} w_j^* \xi_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt} \tag{OA.18}$$

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

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

In addition, let $\boldsymbol{u} = (u_{T_0+1}, \ldots, u_T) = (\tau_{T_0+1}, \ldots, \tau_T)$. It is useful to observe that, under the null hypothesis in (15), 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

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

and

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

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 $\{S(\boldsymbol{u}_\pi)\}_{\pi \in \Pi}$ are well-separated from $S(\boldsymbol{u})$ under $\mathcal{C}_2$. In step three, we show that, conditional on $\mathcal{C}_1$ and $\mathcal{C}_2$, the ordering of $S(\boldsymbol{u}_\pi)$ and $S(\boldsymbol{u})$ will be the same as the ordering of $S(\widehat{\boldsymbol{u}}_\pi)$ and $S(\widehat{\boldsymbol{u}})$ for most $\pi \in \Pi$, which implies that $\widehat{F}(S(\widehat{\boldsymbol{u}}))$ and $\tilde{F}(S(\boldsymbol{u}))$ are also close to each other. In step four, we conclude the proof by linking $\widehat{F}(S(\widehat{\boldsymbol{u}}))$ to the estimated p-value, and $\tilde{F}(S(\boldsymbol{u}))$ to the nominal level α.

OA.7.4.2. 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.6 (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 $X = a_1X_1 + a_2X_2 + \ldots + a_nX_n$. 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.7 *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.7. 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.8 *Recall that u_t is defined as* (OA.17) *and* (OA.18)*, for the blank periods and the experimental periods, respectively. Under the null hypothesis* (15)*, the probability density of u_t can be bounded by*

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

Proof of Lemma OA.8. 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 $t \in \mathcal{B} \cup \{T_0+1,\dots,T\}$. Using Lemma OA.6 (Bobkov and Chistyakov, 2014, Corollary 2), let there be J variables ϵ_{jt} for any $j \in \{1,2,\dots,J\}$. For any $j \in \{1,2,\dots,J\}$, 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.7 Part 2; the second inequality is due to Lemma OA.6; 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,

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

where the first inequality is due to Lemma OA.7 Part 2; the second inequality is due to

Lemma OA.6; 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 $(\boldsymbol{w}^*,\boldsymbol{v}^*)$. ∎

OA.7.4.3. Proof of Theorem OA.1

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

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

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 (15), it follows that

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

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

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 (OA.13), 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 $k \in \{0, 1, 2, ..., T_p\}$, we define the following sets of permutations. First, define $\Pi_0 = \{\pi_0\}$, where π_0 is defined as the set of 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

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

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

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

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

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

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

where the first inequality is due to Lemma OA.7-1; the second inequality is due to Lemma OA.7-2;

the third inequality is due to Lemma OA.6; the last inequality is due to Lemma OA.8 and OA.7-1. 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

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

To conclude step two, define the event

$$\mathcal{C}_2=\left\{\sum_{k=1}^{T_p}\sum_{\pi\in\Pi_k}\mathbb{1}\left\{\left|\sum_{t\in\pi\setminus\pi_0}|u_t|-\sum_{t\in\pi_0\setminus\pi}|u_t|\right|\le 2kz_1\right\}<|\Pi|z_2\right\}.\tag{OA.19}$$

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

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

(Step three.) Conditional on event $\mathcal{C}_2$, fewer than $|\Pi|z_2$ of the absolute value terms in (OA.19) are such that $\left|\sum_{t\in\pi\setminus\pi_0}|u_t|-\sum_{t\in\pi_0\setminus\pi}|u_t|\right|\le 2kz_1$. For all the others, $\left|\sum_{t\in\pi\setminus\pi_0}|u_t|-\sum_{t\in\pi_0\setminus\pi}|u_t|\right|>2kz_1$.

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

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

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

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

Combining both cases, we know that conditional on $\mathcal{C}_1$ and when $\left|\sum_{t\in\pi\setminus\pi_0}|u_t| - \sum_{t\in\pi_0\setminus\pi}|u_t|\right| > 2kz_1$, the ordering of $S(\boldsymbol{u}_\pi)$ and $S(\boldsymbol{u})$ is the same as the ordering of $S(\widehat{\boldsymbol{u}}_\pi)$ and $S(\widehat{\boldsymbol{u}})$. As a result, for those π such that $\left|\sum_{t\in\pi\setminus\pi_0}|u_t| - \sum_{t\in\pi_0\setminus\pi}|u_t|\right| > 2kz_1$, we have $\mathbb{1}\{S(\widehat{\boldsymbol{u}}_\pi) \geq S(\widehat{\boldsymbol{u}})\} =$

$\mathbb{1}\left\{S(\boldsymbol{u}_\pi) \geq S(\boldsymbol{u})\right\}$. There are at most $|\Pi| z_2$ many π's that contribute to the following summation,

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

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

$$\begin{aligned}
&\left|\sum_{\pi \in \Pi} \left(\mathbb{1}\left\{S(\widehat{\boldsymbol{u}}_\pi) \geq S(\widehat{\boldsymbol{u}})\right\} - \mathbb{1}\left\{S(\boldsymbol{u}_\pi) \geq S(\boldsymbol{u})\right\}\right)\right| \\
=&\left|\sum_{k=0}^{T_p} \sum_{\pi \in \Pi_k} \left(\mathbb{1}\left\{S(\widehat{\boldsymbol{u}}_\pi) \geq S(\widehat{\boldsymbol{u}})\right\} - \mathbb{1}\left\{S(\boldsymbol{u}_\pi) \geq S(\boldsymbol{u})\right\}\right)\right| \\
<&|\Pi| z_2.
\end{aligned}$$

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

$$\begin{aligned}
&\left|\Pr\left(1 - \widehat{F}(S(\widehat{\boldsymbol{u}})) \leq \alpha\right) - \Pr\left(1 - \tilde{F}(S(\boldsymbol{u})) \leq \alpha\right)\right| \\
=&\left|\Pr\left(1 - \frac{1}{|\Pi|}\sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{\boldsymbol{u}}_\pi) < S(\widehat{\boldsymbol{u}})\} \leq \alpha\right) - \Pr\left(1 - \frac{1}{|\Pi|}\sum_{\pi \in \Pi} \mathbb{1}\{S(\boldsymbol{u}_\pi) < S(\boldsymbol{u})\} \leq \alpha\right)\right| \\
=&\left|\Pr\left(\sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{\boldsymbol{u}}_\pi) \geq S(\widehat{\boldsymbol{u}})\} \leq \alpha|\Pi|\right) - \Pr\left(\sum_{\pi \in \Pi} \mathbb{1}\{S(\boldsymbol{u}_\pi) \geq S(\boldsymbol{u})\} \leq \alpha|\Pi|\right)\right| \\
=&\left|E\left[\mathbb{1}\left\{\sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{\boldsymbol{u}}_\pi) \geq S(\widehat{\boldsymbol{u}})\} \leq \alpha|\Pi|\right\}\right] - E\left[\mathbb{1}\left\{\sum_{\pi \in \Pi} \mathbb{1}\{S(\boldsymbol{u}_\pi) \geq S(\boldsymbol{u})\} \leq \alpha|\Pi|\right\}\right]\right| \\
\leq& E\left|\mathbb{1}\left\{\sum_{\pi \in \Pi} \mathbb{1}\{S(\widehat{\boldsymbol{u}}_\pi) \geq S(\widehat{\boldsymbol{u}})\} \leq \alpha|\Pi|\right\} - \mathbb{1}\left\{\sum_{\pi \in \Pi} \mathbb{1}\{S(\boldsymbol{u}_\pi) \geq S(\boldsymbol{u})\} \leq \alpha|\Pi|\right\}\right| \\
\leq& \Pr\left(\left|\alpha|\Pi| - \sum_{\pi \in \Pi} \mathbb{1}\{S(\boldsymbol{u}_\pi) \geq S(\boldsymbol{u})\}\right| \leq \left|\sum_{\pi \in \Pi} \left(\mathbb{1}\left\{S(\widehat{\boldsymbol{u}}_\pi) \geq S(\widehat{\boldsymbol{u}})\right\} - \mathbb{1}\left\{S(\boldsymbol{u}_\pi) \geq S(\boldsymbol{u})\right\}\right)\right|\right),
\end{aligned}$$

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

$$\begin{aligned}
&\left| \Pr\left(1 - \widehat{F}(S(\widehat{\boldsymbol{u}})) \leq \alpha\right) - \Pr\left(1 - \tilde{F}(S(\boldsymbol{u})) \leq \alpha\right)\right| \\
&\qquad \leq \Pr\left(\left|\alpha|\Pi| - \sum_{\pi\in\Pi} \mathbb{1}\{S(\boldsymbol{u}_\pi) \geq S(\boldsymbol{u})\}\right| < |\Pi| z_2\right) \\
&\qquad \leq \frac{2|\Pi|z_2}{|\Pi|} \\
&\qquad = 2z_2, \qquad \text{(OA.20)}
\end{aligned}$$

where the last inequality is because $\sum_{\pi\in\Pi} \mathbb{1}\{S(\boldsymbol{u}_\pi) \geq S(\boldsymbol{u})\}$ 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 $\alpha \in (0, 1]$,

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

So conditional on events $\mathcal{C}_1$ and $\mathcal{C}_2$, (OA.20) implies

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

and

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

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

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

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

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

which finishes the proof. ∎

OA.7.5. Proof of Theorem OA.2

Proof of Theorem OA.2. 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

$$\begin{aligned}\sum_{j=1}^J w_j^* Y_{jt}^N - \sum_{j=1}^J v_j^* Y_{jt}^N = \boldsymbol{\theta}_t'\Big(\sum_{j=1}^J w_j^*\boldsymbol{Z}_j - \sum_{j=1}^J v_j^*\boldsymbol{Z}_j\Big)&\\ + \boldsymbol{\lambda}_t'\Big(\sum_{j=1}^J w_j^*\boldsymbol{\mu}_j - \sum_{j=1}^J v_j^*\boldsymbol{\mu}_j\Big) + \Big(\sum_{j=1}^J w_j^*\epsilon_{jt} - \sum_{j=1}^J v_j^*\epsilon_{jt}\Big).&\quad \text{(OA.21)}\end{aligned}$$

Similarly, using expression (12a), we obtain

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

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

$$\begin{aligned}\boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'&\left(\sum_{j=1}^{J} w_j^* \boldsymbol{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{Y}_j^{\mathcal{E}}\right)\\ &= \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\theta}_{\mathcal{E}}\left(\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{Z}_j\right)\\ &+ \boldsymbol{\lambda}_t'\left(\sum_{j=1}^{J} w_j^* \boldsymbol{\mu}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{\mu}_j\right)\\ &+ \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\left(\sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}}\right). \end{aligned} \tag{OA.22}$$

Equations (OA.21) and (OA.22) imply

$$\begin{aligned}\sum_{j=1}^{J} w_j^* Y_{jt}^N - \sum_{j=1}^{J} v_j^* Y_{jt}^N = {} & (\boldsymbol{\theta}_t' - \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\theta}_{\mathcal{E}})\Big(\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{Z}_j\Big)\\ &+ \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\Big(\sum_{j=1}^{J} w_j^* \boldsymbol{Y}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{Y}_j^{\mathcal{E}}\Big)\\ &- \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\Big(\sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}}\Big)\\ &+ \Big(\sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt}\Big). \end{aligned} \tag{OA.23}$$

If Assumption 3 holds, (OA.23) becomes

$$\begin{aligned}\sum_{j=1}^{J} w_j^* Y_{jt}^N - \sum_{j=1}^{J} v_j^* Y_{jt}^N = {} & - \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\Big(\sum_{j=1}^{J} w_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}} - \sum_{j=1}^{J} v_j^* \boldsymbol{\epsilon}_j^{\mathcal{E}}\Big)\\ &+ \Big(\sum_{j=1}^{J} w_j^* \epsilon_{jt} - \sum_{j=1}^{J} v_j^* \epsilon_{jt}\Big). \end{aligned} \tag{OA.24}$$

Only the first term on the right-hand side of (OA.24) has a non-zero mean (because the weights

w_j^* and v_j^*, depend on the error terms $\boldsymbol{\epsilon}_j^{\mathcal{E}}$). Therefore,

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

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

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

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

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

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.23). 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 $\overline{\theta}\Big(1 + \dfrac{\overline{\lambda}^2 F}{\underline{\zeta}}\Big)$. Cauchy–Schwarz inequality and Assumption 6 imply

$$\begin{aligned}
&\left|(\boldsymbol{\theta}_t' - \boldsymbol{\lambda}_t'(\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\lambda}_{\mathcal{E}})^{-1}\boldsymbol{\lambda}_{\mathcal{E}}'\boldsymbol{\theta}_{\mathcal{E}})\left(\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{Z}_j\right)\right| \\
&\qquad\qquad \leq \overline{\theta}\Big(1 + \frac{\overline{\lambda}^2 F}{\underline{\zeta}}\Big)\sqrt{R}\left\|\sum_{j=1}^{J} w_j^* \boldsymbol{Z}_j - \sum_{j=1}^{J} v_j^* \boldsymbol{Z}_j\right\|_2 \\
&\qquad\qquad \leq \overline{\theta}\Big(1 + \frac{\overline{\lambda}^2 F}{\underline{\zeta}}\Big)Rd,
\end{aligned}$$

and

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

Combining the last two displayed equations with (OA.23), we have

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

which finishes the proof of the theorem. ∎

OA.7.6. Proof of Theorem OA.3

Proof of Theorem OA.3. 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 (15) and the assumptions of Theorem OA.3 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 (15) imply

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

for $t \in \mathcal{B}\cup\{T_0+1,\ldots,T\}$, where the last equality is due to Assumption 5(i). So $\{\widehat{u}_t\}_{t\in\mathcal{B}\cup\{T_0+1,\ldots,T\}}$ is a sequence of exchangeable random variables. The result of the theorem follows now from the proof of Theorem 2. ∎

OA.8. Additional Results for the 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.

OA.9. Additional Simulation Results

OA.9.1. Results for a Single Simulation

In Section 5.1 the idiosyncratic shocks are i.i.d. Normal with variance $\sigma^2 = 1$. Figures OA.5 and OA.7 report results for $\sigma^2 = 5$ and $\sigma^2 = 10$, respectively. Figures OA.6 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 (15) increases. The deterioration in pre-treatment fit in Figures OA.5 and OA.7 provides a diagnosis of the accuracy of the respective estimates.

OA.9.2. Performance across Many Simulations

In this section, we present additional simulation results that compare the performance of the different versions of the synthetic control designs over 1000 simulations that independently generate the model primitives (i.e., the factor loadings, covariates, and error terms) of Assumption 1. The data generating process is the same as in Section 5.1.

We consider five versions of the synthetic control design:

1. *Unconstrained design*: This is the design in (7) without a cardinality constraint, so $\underline{m} = 1$ and $\overline{m} = J - 1 = 14$.
2. *Constrained design*: Same as the design in (7), but with $\underline{m} = 1$ and $\overline{m} = 1, \ldots, 7$.
3. *Weakly-targeted design*: This is the design in (9). We vary β from 0.01 to 100.
4. *Unit-level design*: This is the design in (10), which fits a different synthetic control to each unit assigned to treatment. We vary ξ from 0.01 to 100.

5. *Penalized design*: This is the design in (OA.1), 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 (11). 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 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.5 of the online appendix.

OA.9.2.1. Average Treatment Effects

Table OA.1 repeats Table 2 and includes additional results for the other synthetic control designs. The first panel of Table OA.1 reports average treatment effects, τ_t, over 1000 simulations. The second panel reports estimates of the average treatment effects, and then mean absolute error, root mean square error, and p-value, all averaged over 1000 simulations. The second last column reports the rejection rates. The last column reports the number of treated units averaged over

1000 simulations. Mean absolute error (MAE) and root mean square error are defined as

$$\text{MAE} = \frac{1}{T - T_0} \sum_{t=T_0+1}^{T} |\widehat{\tau}_t - \tau_t|, \qquad \text{RMSE} = \sqrt{\frac{1}{T - T_0} \sum_{t=T_0+1}^{T} (\widehat{\tau}_t - \tau_t)^2}, \qquad \text{(OA.25)}$$

and the p-value is defined as in (17). Because the treatment effect is not equal to zero in the simulation of Table OA.1, smaller p-values and larger rejection rates reflect better performance of the testing procedure for a particular design.

In Table OA.1, 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.

OA.9.2.2. 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 OA.9.2. We report the average treatment effects on the treated units in Table OA.2.

The first five columns in Table OA.2 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 differ across formulations of the synthetic control design. The next five columns report averages of $\widehat{\tau}_t$. They are the same as in Table OA.1, yet we use them as estimators for τ_t^T in Table OA.2. The next two columns of Table OA.2 report averages across simulations of the mean absolute error and the root mean square error, defined as in (OA.25) but with τ_t^T replacing τ_t. The last column reports the number of treated units averaged over 1000 simulations.

The results in Table OA.2 are similar to those for τ_t in Table OA.1. This is because in this simulation units are i.i.d. and carry equal weights, making the average treatment effect on the treated, τ_t^T, almost identical to τ_t. Section OA.9.4 reports results for a design with unequal weights, which breaks this close correspondence between τ_t^T and τ_t.

OA.9.2.3. Test size

In this section, we generate the model primitives under the null hypothesis (15). 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 OA.3, 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 OA.3 are close to zero. The same is true for the averages of $\widehat{\tau}_t$ for all designs. Under the null hypothesis (15), the p-value should approximately follow a uniform distribution between zero and one. The results in Table OA.3 show 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.

OA.9.3. Performance across Many Simulations 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,

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

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

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 the factor loadings and the values of the covariates. To control the magnitude of the exponential terms in the nonlinear design, we draw $\boldsymbol{\theta}_t$, $\boldsymbol{\gamma}_t$, $\boldsymbol{\lambda}_t$, and $\boldsymbol{\eta}_t$ as vectors of i.i.d. Uniform$(0, 3)$ random variables, instead of Uniform$(0, 10)$. Similarly, we draw $\boldsymbol{Z}_j$ and $\boldsymbol{\mu}_j$ as vectors of i.i.d. Uniform$(-0.5, 0.5)$ random variables, instead of Uniform$(0, 1)$.

OA.9.3.1. Average Treatment Effects

Table OA.4 reports simulation results for average treatment effects under the nonlinear model.

In comparison to the results in Table OA.1, we see that Table OA.4 presents similar results. 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.

OA.9.3.2. Average Treatment Effects on the Treated

Table OA.5 reports simulation results for the average treatment effects on the treated units under the nonlinear model.

In comparison to the results in Table OA.2, we now see that the *Weakly targeted*, *Unit-level*, and *Penalized* designs can easily improve the performance of the *Unconstrained* design in many cases. The *Weakly targeted* design easily outperforms the *Unconstrained* design for larger values of β, as it puts more emphasis on the average treatment effects on the treated units. The *Unit-level* design can ameliorate interpolation biases induced by the aggregation of $\boldsymbol{X}_j$, by fitting each treated unit with a unit-specific synthetic control. Although the *Unit-level* design is outperformed by the *Weakly targeted* design, it selects fewer treated units. Finally, the *Penalized* design outperforms the *Unconstrained* design in some cases, while always selecting fewer treated units than the *Unconstrained* design.

OA.9.3.3. Comparison to Randomized Treatment Assignments

In this section, we follow the simulation setup as in Section OA.9.3. We consider randomized treatment assignment with $\overline{m}$ treated units. In comparison to the results in Table 3, we see that Table OA.6 presents similar results. 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 Tables 1 and 3, 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.

OA.9.4. Performance across Many Simulations with Unequal Weights

We now examine the behavior of estimators based on synthetic control designs when the weights f_j in expression (1) are equal. All parameter values are the same as in the simulation setup of Section 5.1, except for the weights f_j. The weights f_j are chosen to be proportional to $\frac{1}{1}, \frac{1}{2}, ..., , \frac{1}{15}$ where the sum of weights $\sum_{j=1}^{J} f_j = 1$ is equal to 1.

OA.9.4.1. Average Treatment Effects

In this section, we estimate the average treatment effects by conducting simulations following the simulation setup as in Section OA.9.4. We report the average treatment effects under unequal weights in Table OA.7.

In comparison to the results in Table OA.1, we see that Table OA.7 presents similar results. 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.

OA.9.4.2. 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 OA.9.4. We report the average treatment effects on the treated units under unequal weights in Table OA.8.

Table OA.8 reports the results for τ_t^T. In comparison to the results in Table OA.2, we now see that the *Weakly targeted* design improves the performance of the *Unconstrained* design for larger values of β, as it puts more emphasis on the average treatment effects on the treated units. The *Unit-level* and the *Penalized* designs, however, have worse performance than the *Unconstrained* design.

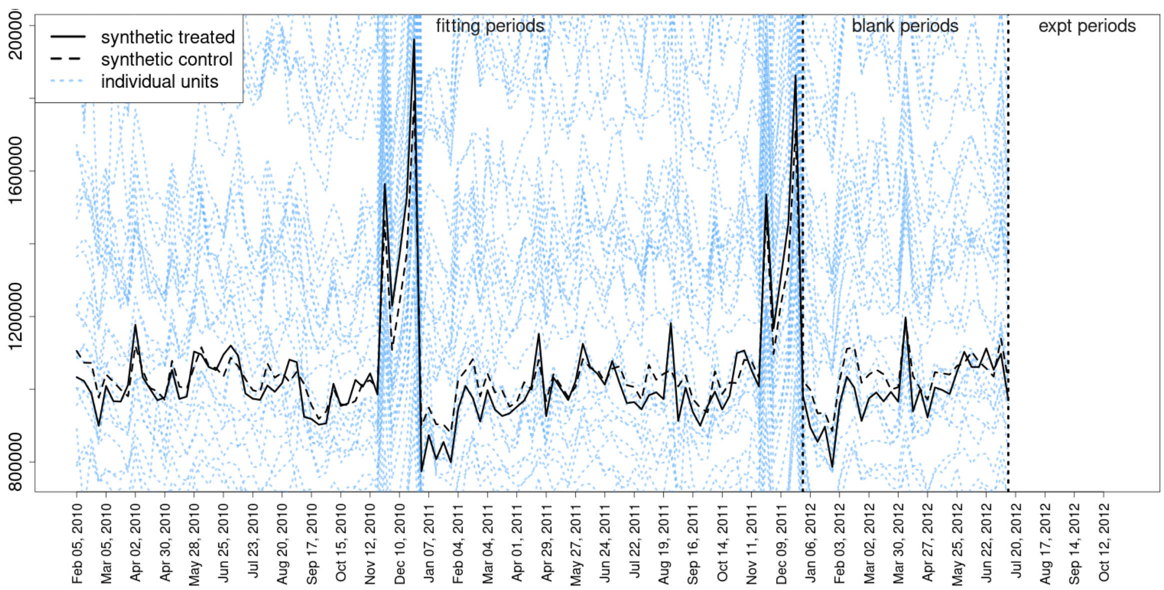

Figure OA.1: Synthetic Treated Unit and Synthetic Control Unit, $\overline{m} = 1$

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.

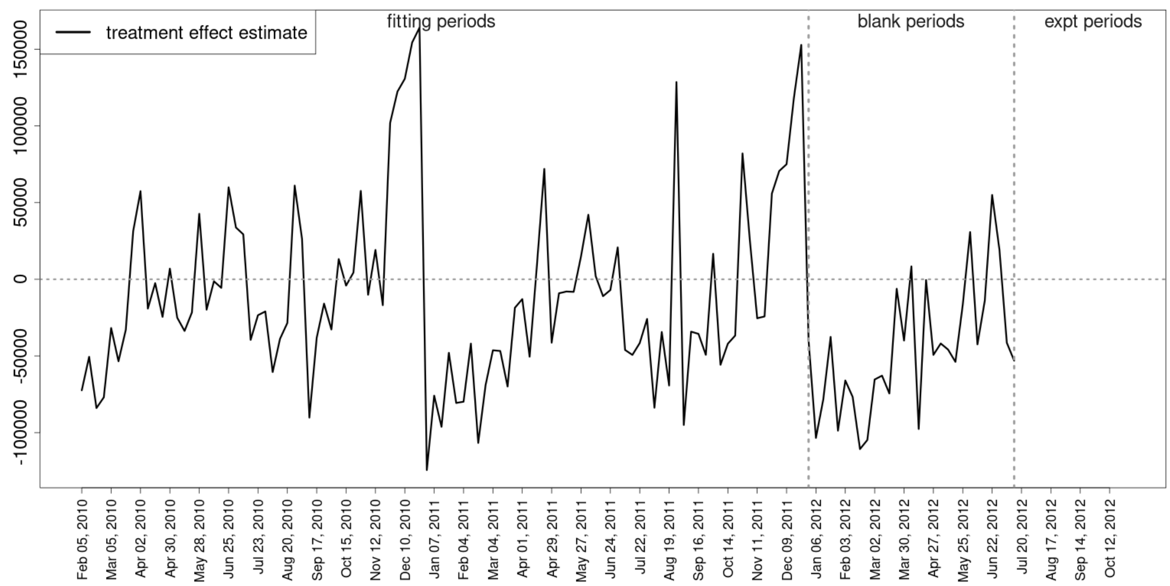

Figure OA.2: Placebo Treatment Effects, $\overline{m} = 1$

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

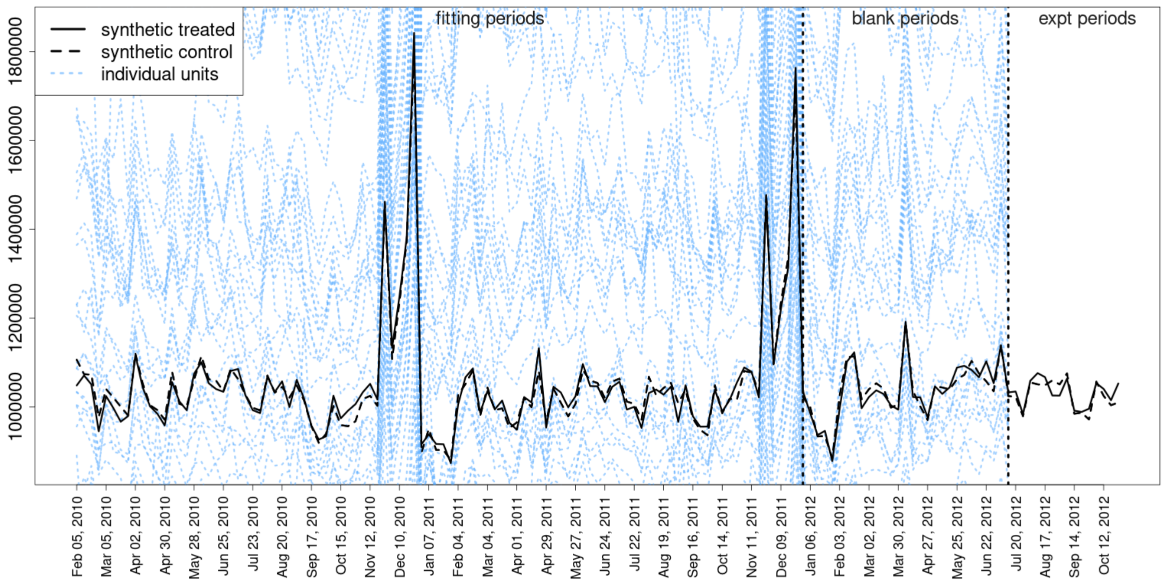

Figure OA.3: Synthetic Treatment Unit and Synthetic Control Unit, when $\overline{m} = 3$.

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.

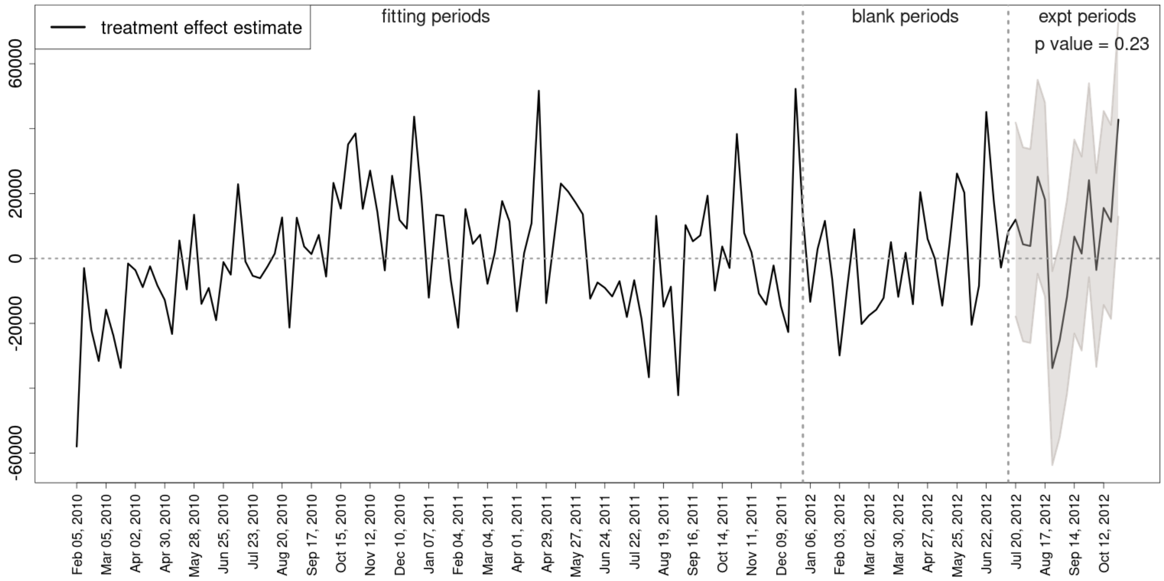

Figure OA.4: Treatment Effect Estimate, when $\overline{m} = 3$.

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.

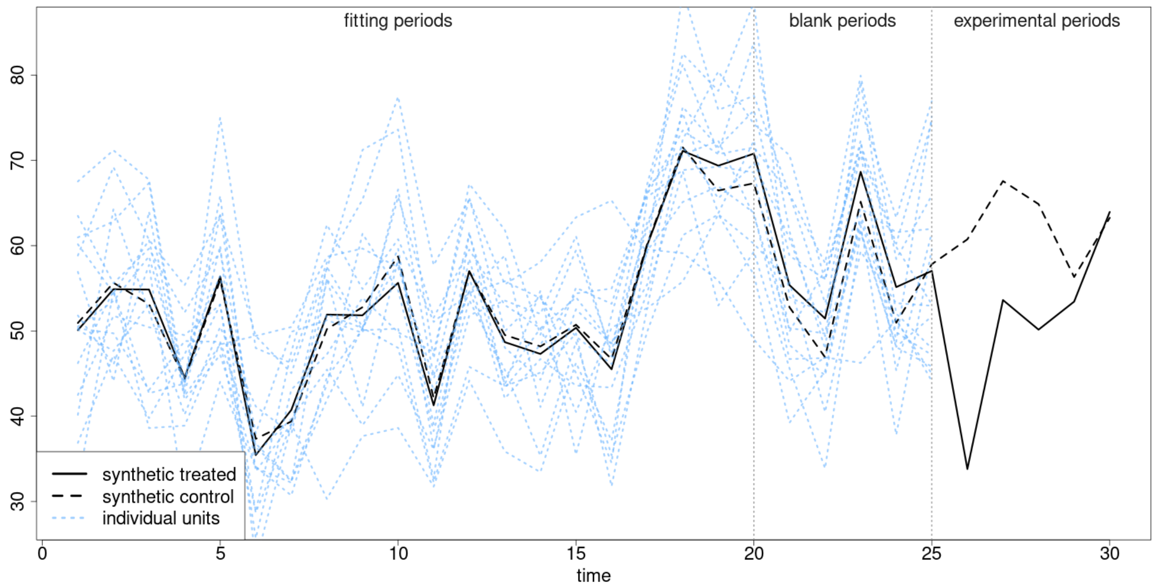

Figure OA.5: Synthetic Treatment Unit and Synthetic Control Unit, when $\sigma^2 = 5$.

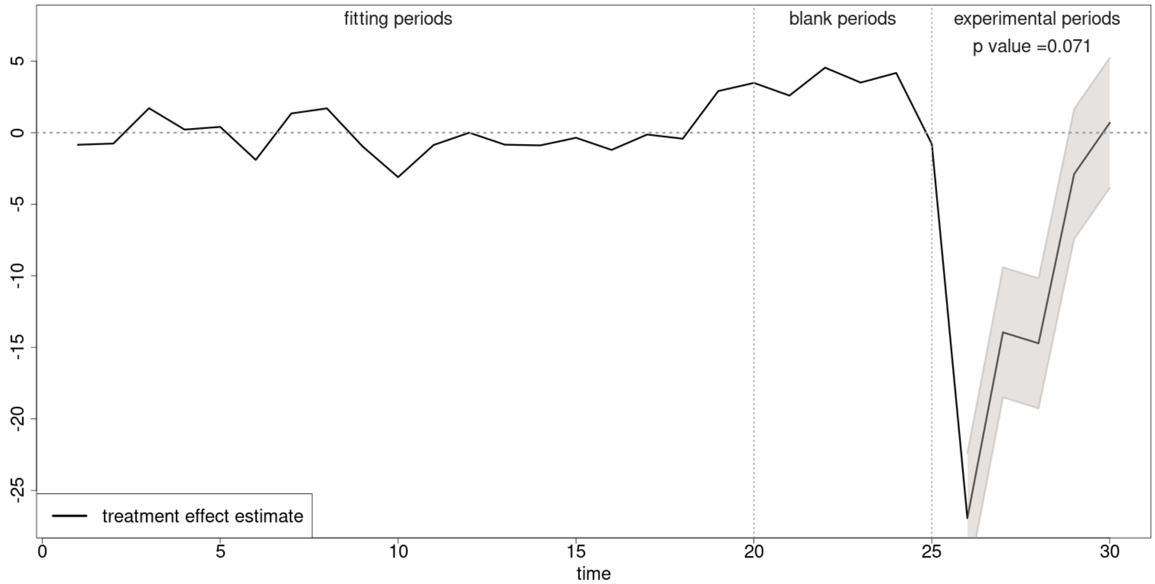

Figure OA.6: Treatment Effect Estimate, when $\sigma^2 = 5$.

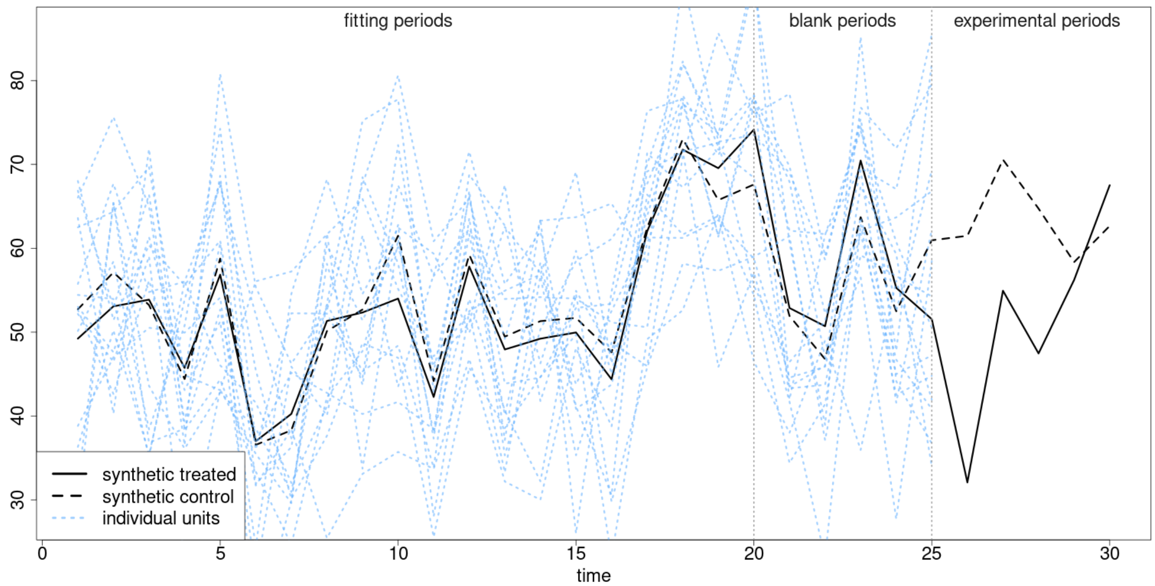

Figure OA.7: Synthetic Treatment Unit and Synthetic Control Unit, when $\sigma^2 = 10$.

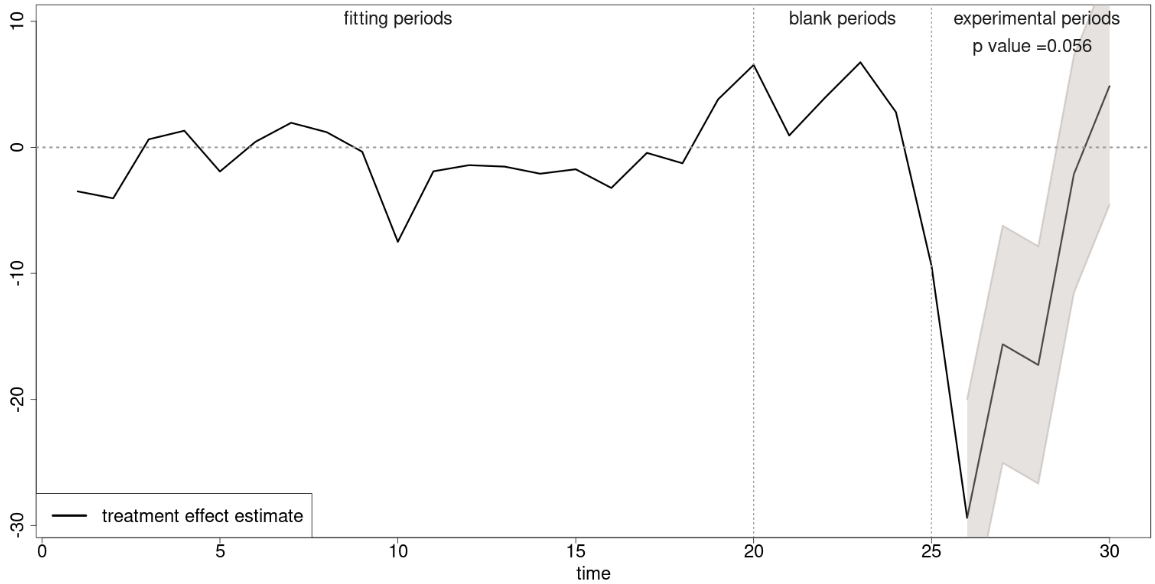

Figure OA.8: Treatment Effect Estimate, when $\sigma^2 = 10$.

Table OA.1: Additional Results for 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*	$\widehat{p}$	$\widehat{p}<0.05$	$\Vert\boldsymbol{w}\Vert_0$
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$					
Unconstrained		-13.57	-10.97	-8.38	-5.07	-2.53	0.83	0.97	0.014	0.944	6.76
Constrained	$\overline{m}=1$	-13.61	-10.97	-8.39	-4.86	-2.41	2.93	3.45	0.057	0.668	1
	$\overline{m}=2$	-13.58	-10.90	-8.43	-5.01	-2.40	1.69	2.00	0.028	0.854	2
	$\overline{m}=3$	-13.56	-11.00	-8.38	-5.05	-2.52	1.26	1.49	0.019	0.916	3
	$\overline{m}=4$	-13.59	-11.06	-8.40	-4.99	-2.50	1.06	1.25	0.016	0.935	4
	$\overline{m}=5$	-13.57	-11.01	-8.37	-5.02	-2.48	0.93	1.09	0.015	0.933	5.00
	$\overline{m}=6$	-13.51	-10.95	-8.29	-5.01	-2.47	0.87	1.02	0.015	0.942	5.97
	$\overline{m}=7$	-13.57	-10.96	-8.37	-5.06	-2.52	0.83	0.97	0.014	0.946	6.76
Weakly-targeted	$\beta=0.01$	-13.57	-10.95	-8.38	-4.99	-2.53	1.17	1.38	0.018	0.920	11.49
	$\beta=0.1$	-13.56	-10.99	-8.34	-4.97	-2.52	0.93	1.08	0.014	0.951	9.75
	$\beta=1$	-13.55	-10.98	-8.32	-4.95	-2.44	0.87	1.01	0.013	0.954	8.49
	$\beta=10$	-13.56	-10.96	-8.37	-4.99	-2.48	0.94	1.10	0.014	0.953	8.15
	$\beta=100$	-13.61	-10.99	-8.42	-5.06	-2.53	1.00	1.18	0.013	0.953	7.91
Unit-level	$\xi=0.01$	-13.60	-10.95	-8.39	-5.04	-2.53	0.95	1.13	0.014	0.938	10.16
	$\xi=0.1$	-13.58	-10.97	-8.35	-4.97	-2.47	0.91	1.07	0.015	0.942	7.30
	$\xi=1$	-13.57	-10.99	-8.38	-4.99	-2.49	1.34	1.58	0.020	0.900	4.50
	$\xi=10$	-13.60	-10.93	-8.45	-5.06	-2.52	2.16	2.57	0.030	0.829	2.11
	$\xi=100$	-13.61	-10.86	-8.48	-5.02	-2.54	2.76	3.27	0.040	0.770	1.15
Penalized	$\lambda=0.01$	-13.58	-10.97	-8.34	-5.05	-2.47	0.88	1.03	0.014	0.950	6.70
	$\lambda=0.1$	-13.64	-11.03	-8.43	-5.03	-2.50	1.21	1.42	0.019	0.904	5.43
	$\lambda=1$	-13.67	-10.96	-8.41	-4.87	-2.45	2.08	2.46	0.037	0.791	2.95
	$\lambda=10$	-13.68	-11.04	-8.37	-4.79	-2.45	3.72	4.40	0.091	0.542	1.11
	$\lambda=100$	-13.64	-10.94	-8.42	-4.86	-2.50	4.17	4.93	0.111	0.490	1

Note: In this table, all designs use $\underline{m}=1$ and $\overline{m}=14$.

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

		τ_t^T					$\widehat{\tau}_t$					MAE^T	$RMSE^T$	$\|\boldsymbol{w}\|_0$
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$	$t=26$	$t=27$	$t=28$	$t=29$	$t=30$			
Unconstrained		-13.58	-10.97	-8.34	-5.00	-2.46	-13.57	-10.97	-8.38	-5.07	-2.53	1.01	1.18	6.76
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	1
	$\overline{m}=2$	-13.53	-10.89	-8.42	-4.93	-2.40	-13.58	-10.90	-8.43	-5.01	-2.40	1.80	2.13	2
	$\overline{m}=3$	-13.39	-10.92	-8.35	-4.97	-2.54	-13.56	-11.00	-8.38	-5.05	-2.52	1.39	1.64	3
	$\overline{m}=4$	-13.54	-11.05	-8.42	-4.95	-2.52	-13.59	-11.06	-8.40	-4.99	-2.50	1.22	1.44	4
	$\overline{m}=5$	-13.57	-11.03	-8.42	-5.02	-2.45	-13.57	-11.01	-8.37	-5.02	-2.48	1.11	1.30	5.00
	$\overline{m}=6$	-13.61	-11.06	-8.36	-4.99	-2.46	-13.51	-10.95	-8.29	-5.01	-2.47	1.03	1.22	5.97
	$\overline{m}=7$	-13.58	-10.97	-8.34	-5.00	-2.46	-13.57	-10.96	-8.37	-5.06	-2.52	1.01	1.18	6.76
Weakly-targeted	$\beta=0.01$	-13.59	-10.99	-8.31	-5.00	-2.51	-13.57	-10.95	-8.38	-4.99	-2.53	1.30	1.54	11.49
	$\beta=0.1$	-13.59	-10.98	-8.37	-5.01	-2.51	-13.56	-10.99	-8.34	-4.97	-2.52	1.07	1.26	9.75
	$\beta=1$	-13.54	-10.98	-8.36	-5.01	-2.48	-13.55	-10.98	-8.32	-4.95	-2.44	0.99	1.16	8.49
	$\beta=10$	-13.50	-10.96	-8.33	-5.03	-2.51	-13.56	-10.96	-8.37	-4.99	-2.48	1.00	1.16	8.15
	$\beta=100$	-13.49	-10.90	-8.31	-5.02	-2.51	-13.61	-10.99	-8.42	-5.06	-2.53	1.01	1.17	7.91
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	10.16
	$\xi=0.1$	-13.56	-10.99	-8.38	-4.95	-2.50	-13.58	-10.97	-8.35	-4.97	-2.47	1.08	1.28	7.30
	$\xi=1$	-13.57	-11.02	-8.31	-4.91	-2.48	-13.57	-10.99	-8.38	-4.99	-2.49	1.40	1.66	4.50
	$\xi=10$	-13.78	-10.98	-8.37	-4.87	-2.52	-13.60	-10.93	-8.45	-5.06	-2.52	1.96	2.33	2.11
	$\xi=100$	-13.81	-10.90	-8.39	-4.80	-2.55	-13.61	-10.86	-8.48	-5.02	-2.54	2.35	2.78	1.15
Penalized	$\lambda=0.01$	-13.59	-10.99	-8.35	-5.00	-2.47	-13.58	-10.97	-8.34	-5.05	-2.47	1.05	1.23	6.70
	$\lambda=0.1$	-13.57	-10.98	-8.36	-4.91	-2.42	-13.64	-11.03	-8.43	-5.03	-2.50	1.36	1.61	5.43
	$\lambda=1$	-13.69	-11.01	-8.33	-4.80	-2.48	-13.67	-10.96	-8.41	-4.87	-2.45	2.19	2.59	2.95
	$\lambda=10$	-13.88	-11.06	-8.23	-4.73	-2.40	-13.68	-11.04	-8.37	-4.79	-2.45	3.81	4.50	1.11
	$\lambda=100$	-13.90	-11.10	-8.27	-4.75	-2.46	-13.64	-10.94	-8.42	-4.86	-2.50	4.24	5.00	1

Table OA.3: Average Treatment Effects Under the Null Hypothesis (15) (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*	$\widehat{p}$	$\widehat{p} < 0.05$	$\lVert \boldsymbol{w} \rVert_0$
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$					
Unconstrained		-0.01	0.00	-0.03	-0.05	-0.07	0.96	1.13	0.495	0.061	6.76
Constrained	$\overline{m}=1$	0.20	0.14	0.02	-0.08	0.05	3.00	3.55	0.495	0.056	1
	$\overline{m}=2$	-0.02	-0.01	-0.02	-0.09	-0.03	1.80	2.13	0.497	0.038	2
	$\overline{m}=3$	-0.09	-0.07	-0.02	-0.05	-0.02	1.37	1.62	0.505	0.048	3
	$\overline{m}=4$	-0.02	-0.02	0.00	-0.01	-0.01	1.19	1.41	0.494	0.054	4
	$\overline{m}=5$	0.01	-0.02	0.03	0.00	-0.05	1.07	1.25	0.496	0.057	5.00
	$\overline{m}=6$	0.07	0.06	0.10	-0.01	-0.03	0.99	1.17	0.484	0.054	5.97
	$\overline{m}=7$	-0.01	0.00	-0.02	-0.04	-0.07	0.96	1.13	0.495	0.059	6.76
Weakly-targeted	$\beta=0.01$	0.01	0.04	-0.06	0.00	-0.02	1.27	1.50	0.504	0.042	11.49
	$\beta=0.1$	0.03	-0.01	0.01	0.04	-0.02	1.03	1.21	0.499	0.055	9.75
	$\beta=1$	0.00	0.00	0.03	0.08	0.04	0.95	1.11	0.499	0.047	8.49
	$\beta=10$	-0.03	0.01	-0.03	0.04	0.03	0.95	1.10	0.486	0.053	8.15
	$\beta=100$	-0.08	-0.08	-0.09	-0.04	-0.02	0.96	1.11	0.490	0.050	7.91
Unit-level	$\xi=0.01$	0.00	0.03	-0.04	-0.04	-0.02	1.05	1.25	0.511	0.053	10.16
	$\xi=0.1$	0.00	0.00	0.03	0.02	0.02	1.05	1.24	0.500	0.049	7.30
	$\xi=1$	0.01	0.02	-0.06	-0.05	-0.03	1.38	1.63	0.498	0.046	4.50
	$\xi=10$	0.18	0.00	-0.02	-0.15	-0.02	1.97	2.33	0.496	0.038	2.11
	$\xi=100$	0.19	-0.03	-0.02	-0.18	-0.03	2.34	2.77	0.502	0.053	1.15
Penalized	$\lambda=0.01$	0.01	0.01	0.02	-0.03	-0.01	1.01	1.18	0.493	0.052	6.70
	$\lambda=0.1$	-0.07	-0.05	-0.07	-0.11	-0.10	1.32	1.56	0.505	0.041	5.43
	$\lambda=1$	0.02	0.07	-0.07	-0.07	0.01	2.17	2.57	0.495	0.045	2.95
	$\lambda=10$	0.16	0.03	-0.11	-0.08	-0.08	3.79	4.48	0.514	0.045	1.11
	$\lambda=100$	0.22	0.15	-0.14	-0.14	-0.08	4.22	5.00	0.515	0.041	1

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

Table OA.4: 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*	$\widehat{p}$	$\widehat{p} < 0.05$	$\lVert \boldsymbol{w} \rVert_0$
		$t = 26$	$t = 27$	$t = 28$	$t = 29$	$t = 30$					
Unconstrained		-13.45	-10.90	-8.18	-5.88	-2.75	1.99	2.54	0.059	0.741	6.49
Constrained	$\overline{m} = 1$	-15.70	-13.18	-10.50	-7.76	-4.78	3.51	4.27	0.061	0.717	1
	$\overline{m} = 2$	-14.27	-11.86	-8.90	-6.44	-3.34	2.64	3.29	0.061	0.725	2
	$\overline{m} = 3$	-13.69	-11.38	-8.38	-5.95	-2.97	2.23	2.83	0.058	0.745	3
	$\overline{m} = 4$	-13.58	-11.09	-8.23	-5.89	-2.75	2.10	2.67	0.058	0.754	4
	$\overline{m} = 5$	-13.37	-10.97	-8.14	-5.79	-2.88	2.05	2.61	0.060	0.747	5.00
	$\overline{m} = 6$	-13.54	-11.03	-8.31	-5.86	-2.86	2.00	2.56	0.060	0.738	5.93
	$\overline{m} = 7$	-13.49	-10.94	-8.16	-5.85	-2.77	1.98	2.53	0.058	0.743	6.56
Weakly-targeted	$\beta = 0.01$	-11.67	-9.02	-6.37	-3.87	-1.00	2.59	3.24	0.116	0.603	11.14
	$\beta = 0.1$	-12.08	-9.60	-6.86	-4.31	-1.47	2.15	2.73	0.083	0.678	9.56
	$\beta = 1$	-12.50	-10.12	-7.37	-4.81	-1.93	1.97	2.52	0.056	0.762	8.30
	$\beta = 10$	-13.00	-10.54	-7.79	-5.24	-2.32	2.19	2.77	0.031	0.851	7.93
	$\beta = 100$	-13.28	-10.76	-8.00	-5.44	-2.55	2.45	3.11	0.024	0.884	7.47
Unit-level	$\xi = 0.01$	-11.76	-9.15	-6.51	-3.91	-1.15	2.57	3.22	0.118	0.593	10.06
	$\xi = 0.1$	-13.11	-10.59	-7.82	-5.15	-2.29	2.06	2.64	0.060	0.754	7.17
	$\xi = 1$	-13.74	-11.12	-8.42	-5.75	-2.84	2.37	3.02	0.029	0.850	4.08
	$\xi = 10$	-13.74	-11.20	-8.55	-5.89	-3.09	3.02	3.77	0.028	0.866	1.74
	$\xi = 100$	-13.79	-11.16	-8.54	-5.90	-3.08	3.20	4.00	0.029	0.863	1.08
Penalized	$\lambda = 0.01$	-13.41	-10.93	-8.32	-5.82	-2.82	1.97	2.53	0.055	0.759	6.44
	$\lambda = 0.1$	-13.32	-10.79	-8.12	-5.56	-2.64	2.07	2.64	0.045	0.777	5.69
	$\lambda = 1$	-13.32	-10.84	-8.15	-5.39	-2.60	3.08	3.84	0.056	0.738	3.04
	$\lambda = 10$	-13.39	-10.82	-7.95	-5.34	-2.58	3.85	4.80	0.103	0.595	1.09
	$\lambda = 100$	-13.35	-10.82	-8.00	-5.29	-2.57	4.10	5.11	0.117	0.562	1

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

Table OA.5: Average Treatment Effects on the Treated, Nonlinear Model (Averages over 1000 Simulations)

		τ_t^T					$\widehat{\tau}_t$					MAE^T	$RMSE^T$	$\|\boldsymbol{w}\|_0$
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$	$t=26$	$t=27$	$t=28$	$t=29$	$t=30$			
Unconstrained		-13.30	-10.65	-7.99	-5.45	-2.41	-13.45	-10.90	-8.18	-5.88	-2.75	2.31	3.01	6.49
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	1
	$\overline{m}=2$	-13.19	-10.88	-7.93	-5.42	-2.47	-14.27	-11.86	-8.90	-6.44	-3.34	2.86	3.61	2
	$\overline{m}=3$	-13.17	-10.82	-8.07	-5.54	-2.44	-13.69	-11.38	-8.38	-5.95	-2.97	2.52	3.25	3
	$\overline{m}=4$	-13.29	-10.82	-8.03	-5.59	-2.31	-13.58	-11.09	-8.23	-5.89	-2.75	2.44	3.16	4
	$\overline{m}=5$	-13.23	-10.77	-7.88	-5.58	-2.30	-13.37	-10.97	-8.14	-5.79	-2.88	2.39	3.10	5.00
	$\overline{m}=6$	-13.36	-10.71	-7.99	-5.50	-2.36	-13.54	-11.03	-8.31	-5.86	-2.86	2.32	3.02	5.93
	$\overline{m}=7$	-13.31	-10.68	-7.98	-5.47	-2.40	-13.49	-10.94	-8.16	-5.85	-2.77	2.30	3.01	6.56
Weakly-targeted	$\beta=0.01$	-13.14	-10.67	-7.97	-5.47	-2.45	-11.67	-9.02	-6.37	-3.87	-1.00	2.68	3.36	11.14
	$\beta=0.1$	-13.20	-10.66	-7.91	-5.47	-2.45	-12.08	-9.60	-6.86	-4.31	-1.47	2.32	2.94	9.56
	$\beta=1$	-13.21	-10.75	-7.90	-5.45	-2.45	-12.50	-10.12	-7.37	-4.81	-1.93	1.90	2.43	8.30
	$\beta=10$	-13.30	-10.69	-7.95	-5.46	-2.49	-13.00	-10.54	-7.79	-5.24	-2.32	1.33	1.66	7.93
	$\beta=100$	-13.29	-10.75	-8.07	-5.46	-2.53	-13.28	-10.76	-8.00	-5.44	-2.55	1.14	1.40	7.47
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	10.06
	$\xi=0.1$	-13.35	-10.81	-8.07	-5.53	-2.55	-13.11	-10.59	-7.82	-5.15	-2.29	1.98	2.51	7.17
	$\xi=1$	-13.40	-10.77	-8.05	-5.47	-2.56	-13.74	-11.12	-8.42	-5.75	-2.84	1.47	1.81	4.08
	$\xi=10$	-13.35	-10.66	-7.99	-5.38	-2.71	-13.74	-11.20	-8.55	-5.89	-3.09	1.49	1.80	1.74
	$\xi=100$	-13.39	-10.63	-7.95	-5.37	-2.69	-13.79	-11.16	-8.54	-5.90	-3.08	1.56	1.89	1.08
Penalized	$\lambda=0.01$	-13.33	-10.70	-8.07	-5.50	-2.42	-13.41	-10.93	-8.32	-5.82	-2.82	2.25	2.93	6.44
	$\lambda=0.1$	-13.17	-10.72	-8.09	-5.59	-2.55	-13.32	-10.79	-8.12	-5.56	-2.64	1.87	2.38	5.69
	$\lambda=1$	-13.40	-10.80	-7.98	-5.35	-2.66	-13.32	-10.84	-8.15	-5.39	-2.60	1.97	2.44	3.04
	$\lambda=10$	-13.41	-10.89	-8.00	-5.32	-2.64	-13.39	-10.82	-7.95	-5.34	-2.58	2.79	3.46	1.09
	$\lambda=100$	-13.42	-10.87	-8.02	-5.26	-2.63	-13.35	-10.82	-8.00	-5.29	-2.57	3.06	3.80	1

Table OA.6: RMSE for Different Experimental Designs and Estimators, Nonlinear Model (Averages over 1000 Simulations)

	SC	RND	STR	REG	1-NN	5-NN
$\overline{m} = 1$	4.27	12.71	12.71	14.26	8.02	7.42
$\overline{m} = 2$	3.29	9.79	9.68	10.11	6.51	6.20
$\overline{m} = 3$	2.83	8.48	9.08	8.93	5.65	5.54
$\overline{m} = 4$	2.67	8.01	8.95	8.46	5.24	5.22
$\overline{m} = 5$	2.61	7.88	8.07	7.87	4.95	5.00
$\overline{m} = 6$	2.56	7.71	8.92	7.42	4.69	4.81
$\overline{m} = 7$	2.53	7.43	8.05	7.27	4.54	4.68

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.

Table OA.7: Average Treatment Effects, Unequal Weights (Averages over 1000 Simulations)

		τ_t									
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$					
		-13.64	-10.96	-8.34	-4.99	-2.52					
		$\widehat{\tau}_t$					*MAE*	*RMSE*	$\widehat{p}$	$\widehat{p} < 0.05$	$\lVert \boldsymbol{w} \rVert_0$
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$					
Unconstrained		-13.61	-10.95	-8.35	-5.00	-2.48	1.16	1.36	0.019	0.913	6.36
Constrained	$\overline{m}=1$	-13.80	-11.11	-8.34	-5.05	-2.42	2.94	3.45	0.062	0.646	1
	$\overline{m}=2$	-13.58	-11.01	-8.30	-4.98	-2.48	1.70	2.01	0.031	0.836	2
	$\overline{m}=3$	-13.63	-10.96	-8.29	-5.00	-2.46	1.38	1.62	0.022	0.891	3
	$\overline{m}=4$	-13.65	-10.91	-8.32	-4.97	-2.49	1.20	1.43	0.020	0.906	4
	$\overline{m}=5$	-13.61	-10.93	-8.32	-4.97	-2.49	1.17	1.38	0.019	0.918	4.99
	$\overline{m}=6$	-13.59	-10.96	-8.32	-4.98	-2.46	1.12	1.33	0.019	0.914	5.91
	$\overline{m}=7$	-13.58	-10.96	-8.32	-5.01	-2.46	1.13	1.34	0.018	0.915	6.47
Weakly-targeted	$\beta=0.01$	-13.65	-11.09	-8.44	-5.09	-2.58	1.44	1.70	0.023	0.883	10.65
	$\beta=0.1$	-13.64	-11.05	-8.40	-5.08	-2.57	1.25	1.47	0.019	0.915	9.06
	$\beta=1$	-13.62	-11.06	-8.33	-5.06	-2.50	1.21	1.44	0.016	0.934	7.89
	$\beta=10$	-13.53	-10.97	-8.27	-4.93	-2.42	1.37	1.62	0.014	0.947	7.61
	$\beta=100$	-13.52	-10.96	-8.32	-4.96	-2.48	1.48	1.75	0.014	0.952	7.65
Unit-level	$\xi=0.01$	-13.52	-10.97	-8.27	-4.97	-2.44	1.42	1.67	0.021	0.887	9.33
	$\xi=0.1$	-13.60	-11.04	-8.31	-5.01	-2.48	1.33	1.57	0.020	0.910	6.84
	$\xi=1$	-13.61	-10.99	-8.38	-5.04	-2.49	1.56	1.85	0.022	0.897	4.32
	$\xi=10$	-13.62	-10.98	-8.46	-5.08	-2.54	2.42	2.85	0.031	0.825	2.12
	$\xi=100$	-13.62	-10.85	-8.49	-5.01	-2.54	3.01	3.54	0.040	0.768	1.16
Penalized	$\lambda=0.01$	-13.67	-10.97	-8.36	-5.01	-2.53	1.18	1.39	0.019	0.917	6.28
	$\lambda=0.1$	-13.68	-11.02	-8.40	-5.07	-2.40	1.42	1.67	0.022	0.897	5.13
	$\lambda=1$	-13.68	-10.92	-8.33	-4.96	-2.34	2.25	2.66	0.042	0.757	2.72
	$\lambda=10$	-13.75	-10.85	-8.21	-4.85	-2.25	3.76	4.44	0.095	0.528	1.08
	$\lambda=100$	-13.79	-10.94	-8.20	-4.84	-2.26	4.18	4.93	0.114	0.470	1

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

Table OA.8: Average Treatment Effects on the Treated, Unequal Weights (Averages over 1000 Simulations)

		τ_t^T					$\widehat{\tau}_t$					MAE^T	$RMSE^T$	$\|\boldsymbol{w}\|_0$
		$t=26$	$t=27$	$t=28$	$t=29$	$t=30$	$t=26$	$t=27$	$t=28$	$t=29$	$t=30$			
Unconstrained		-13.58	-10.96	-8.27	-4.96	-2.52	-13.61	-10.95	-8.35	-5.00	-2.48	1.34	1.58	6.36
Constrained	$\overline{m}=1$	-13.66	-11.03	-8.41	-4.83	-2.35	-13.80	-11.11	-8.34	-5.05	-2.42	3.14	3.70	1
	$\overline{m}=2$	-13.60	-10.94	-8.25	-4.99	-2.56	-13.58	-11.01	-8.30	-4.98	-2.48	1.93	2.27	2
	$\overline{m}=3$	-13.61	-10.95	-8.31	-4.92	-2.56	-13.63	-10.96	-8.29	-5.00	-2.46	1.59	1.88	3
	$\overline{m}=4$	-13.66	-10.95	-8.33	-4.97	-2.53	-13.65	-10.91	-8.32	-4.97	-2.49	1.40	1.65	4
	$\overline{m}=5$	-13.65	-10.97	-8.24	-5.02	-2.55	-13.61	-10.93	-8.32	-4.97	-2.49	1.36	1.60	4.99
	$\overline{m}=6$	-13.62	-10.99	-8.32	-5.00	-2.50	-13.59	-10.96	-8.32	-4.98	-2.46	1.31	1.55	5.91
	$\overline{m}=7$	-13.60	-10.97	-8.31	-4.97	-2.53	-13.58	-10.96	-8.32	-5.01	-2.46	1.31	1.55	6.47
Weakly-targeted	$\beta=0.01$	-13.62	-10.94	-8.31	-4.96	-2.52	-13.65	-11.09	-8.44	-5.09	-2.58	1.53	1.81	10.65
	$\beta=0.1$	-13.62	-10.94	-8.31	-4.98	-2.52	-13.64	-11.05	-8.40	-5.08	-2.57	1.35	1.59	9.06
	$\beta=1$	-13.58	-10.95	-8.29	-4.99	-2.51	-13.62	-11.06	-8.33	-5.06	-2.50	1.18	1.38	7.89
	$\beta=10$	-13.58	-10.95	-8.29	-4.99	-2.54	-13.53	-10.97	-8.27	-4.93	-2.42	1.05	1.23	7.61
	$\beta=100$	-13.55	-10.92	-8.29	-4.98	-2.55	-13.52	-10.96	-8.32	-4.96	-2.48	1.00	1.17	7.65
Unit-level	$\xi=0.01$	-13.65	-10.96	-8.36	-4.98	-2.54	-13.52	-10.97	-8.27	-4.97	-2.44	1.50	1.77	9.33
	$\xi=0.1$	-13.62	-10.93	-8.38	-4.94	-2.52	-13.60	-11.04	-8.31	-5.01	-2.48	1.44	1.70	6.84
	$\xi=1$	-13.59	-11.07	-8.36	-4.88	-2.48	-13.61	-10.99	-8.38	-5.04	-2.49	1.51	1.78	4.32
	$\xi=10$	-13.78	-10.99	-8.39	-4.87	-2.53	-13.62	-10.98	-8.46	-5.08	-2.54	1.97	2.34	2.12
	$\xi=100$	-13.82	-10.90	-8.40	-4.79	-2.56	-13.62	-10.85	-8.49	-5.01	-2.54	2.35	2.79	1.16
Penalized	$\lambda=0.01$	-13.58	-10.99	-8.28	-4.96	-2.51	-13.67	-10.97	-8.36	-5.01	-2.53	1.37	1.61	6.28
	$\lambda=0.1$	-13.58	-10.99	-8.31	-4.87	-2.51	-13.68	-11.02	-8.40	-5.07	-2.40	1.54	1.82	5.13
	$\lambda=1$	-13.64	-11.04	-8.30	-4.79	-2.44	-13.68	-10.92	-8.33	-4.96	-2.34	2.39	2.83	2.72
	$\lambda=10$	-13.61	-11.00	-8.31	-4.81	-2.35	-13.75	-10.85	-8.21	-4.85	-2.25	3.93	4.63	1.08
	$\lambda=100$	-13.60	-11.01	-8.29	-4.77	-2.37	-13.79	-10.94	-8.20	-4.84	-2.26	4.35	5.13	1