

CAUSAL INFERENCE WITH GROUPWISE MATCHING

Ratzanyel Rincón and Kyungchul Song

Vancouver School of Economics, University of British Columbia

March 24, 2026

ABSTRACT. This paper examines methods of causal inference based on groupwise matching when we observe multiple large groups of individuals over several periods. We formulate causal inference validity through a generalized matching condition, generalizing the parallel trend assumption in difference-in-differences designs. We show that difference-in-differences, synthetic control, and synthetic difference-in-differences designs are distinguished by the specific matching conditions that they invoke. Through regret analysis, we demonstrate that difference-in-differences and synthetic control with differencing are complementary; the former dominates the latter if and only if the latter's extrapolation error exceeds the former's matching error up to a term vanishing at the parametric rate. The analysis also reveals that synthetic control with differencing is equivalent to difference-in-differences when the parallel trend assumption holds for both the pre-treatment and post-treatment periods. We develop a statistical inference procedure based on synthetic control with differencing and present an empirical application demonstrating its usefulness.

KEY WORDS. Causal Inference; Difference-in-Differences; Synthetic Control Methods; Synthetic Difference-in-Differences; Parallel Trend Assumption; Generalized Matching Conditions

JEL CLASSIFICATION: C01, C18, C21

Date: March 24, 2026.

We thank Tom Chan, Hiro Kasahara, Yiqi Liu, Paul Schrimpf and participants in Econometrics Lunch Seminar at UBC for valuable comments. We also thank Chun Pang Chow for his excellent research assistance and outstanding work in building the R package implementing the SCD method proposed in this paper. The package is publicly available at <https://github.com/ratzanyelrincon/scd>. Song acknowledges that this research was supported by Social Sciences and Humanities Research Council of Canada. All errors are ours. Corresponding address: Kyungchul Song, Vancouver School of Economics, University of British Columbia, 6000 Iona Drive, Vancouver, BC, V6T 1L4, Canada. Email address: kysong@mail.ubc.ca.

1. Introduction

A recent stream of literature provides a systematic comparison between different causal inference designs, especially between the synthetic control (SC) and other designs such as difference-in-differences (DID) or matching (Doudchenko and Imbens (2017), Ferman and Pinto (2021), Kellogg *et al.* (2021), Arkhangelsky *et al.* (2021) and Chen (2023)). However, the comparison falls short of giving a full picture, because it assumes a data structure inspired by the SC methods. The data structure assumes cross-sectional units of similar or smaller magnitude than the time periods. Furthermore, it is not uncommon in this literature that the treatment occurs only for a single cross-sectional unit.

We take an opposite direction by studying the SC design and its variants from the DID perspective, assuming a data structure that involves multiple large groups of individuals observed over a short period of time. Recent advances in the literature of DID designs consider multiple untreated groups such as in settings with staggered adoption and heterogenous causal effects (see Callaway and Sant’Anna (2021), de Chaisemartin and D’Haultfœuille (2020), Goodman-Bacon (2021), Sun and Abraham (2021) and surveys by de Chaisemartin and D’Haultfœuille (2023) and Roth *et al.* (2023)). Thus, the SC approach naturally maps to this DID framework with multiple “donor groups”, by matching a counterfactual untreated group mean μ_0 to a weighted average of group means μ_j in the “donor pool”:

$$(1.1) \quad \mu_0 = \sum_j \mu_j w_j.$$

We call such causal inference methods *groupwise matching*.¹

As we show in this paper, the DID design can be thought of as arising from groupwise matching like SC. The main difference lies in the choice of the weights w_j . In the SC approach, the weights are chosen to minimize the pre-treatment matching errors, whereas in the DID approach, as this paper shows, the weights are chosen based on the fraction of the group sizes in the donor pool. The difference originates from two distinct thoughts on how we extrapolate the observed untreated outcomes to the counterfactual untreated outcomes for the treated units. The DID method matches the counterfactual mean untreated outcome to a pre-specified surrogate control group, whereas the SC method relies on the *stability of matching* as we move from the pre-treatment to the post-treatment periods.

In this paper, we formalize the complementarity of these two thoughts using a generalized version of the condition (1.1) that we call Generalized Matching Condition (GMC). More specifically, let $\mu_{j,t}(0)$ be a within-group-differenced, mean untreated potential outcome for

¹There are works that use groupwise matching in the SC approach (see Robbins *et al.* (2017), Xu (2017), and Sun *et al.* (2025)). See also Gunsilius (2023) who use quantiles instead of means in groupwise matching.

group j at time t . For a choice of weights w_j , the population-level matching error from matching to target group 0 is defined as follows:

$$(1.2) \quad e_t(w) = \mu_{0,t}(0) - \sum_j \mu_{j,t}(0)w_j,$$

where the sum is over the groups in the donor pool. Then the GMC simply says that $e_t(w) = 0$ for all post-treatment periods t .² Recent advances in the SC literature inspire various causal inference methods in this groupwise matching setting, including the classic synthetic control (SC), synthetic difference-in-differences (SDID), and synthetic control with differencing (SCD), and as we show later, the GMC captures their key identifying assumptions.³

Within this GMC framework, we focus on the SCD design which applies the SC weights after performing within-group differencing to eliminate time-invariant individual heterogeneity in potential outcomes. DID assigns weights based on the relative sizes of groups within the donor pool. In contrast, SCD chooses weights that best match the weighted average of donor group outcomes to the untreated outcomes for the treated group, yet this matching occurs only on the pre-treatment outcomes, not on the post-treatment outcomes. Consequently, SCD suffers from extrapolation error when the weights that achieve the best pre-treatment match fail to provide an adequate post-treatment match. On the other hand, DID's reliance on group-size-based weights makes it vulnerable to matching error if the surrogate control group is misspecified. Therefore, the relative performance of SCD versus DID depends fundamentally on SCD's extrapolation error against DID's matching error.⁴

We formalize this observation focusing on the setting with point-identified weights. Using the matching errors $e_t(w)$ in (1.2), we define the squared sum of matching errors:

$$\text{SSME}_d(w) = \frac{1}{|\mathcal{T}_d|} \sum_{t \in \mathcal{T}_d} e_t^2(w), \quad d = 0, 1,$$

where $\mathcal{T}_0$ denotes the set of pre-treatment periods and $\mathcal{T}_1$ that of post-treatment periods. From this, we construct two quantities that are used to evaluate the choice of the weight vector w :

$$\text{Matching Error in Regret: } \text{MER}_d(w) = \text{SSME}_d(w) - \inf_{\tilde{w} \in \Delta_{K-1}} \text{SSME}_d(\tilde{w}), \text{ and}$$

$$\text{Extrapolation Error: } \Delta\text{MER}(w) = \text{MER}_1(w) - \text{MER}_0(w).$$

²See [Shi et al. \(2022\)](#) for an investigation of primitive assumptions that yield this condition.

³The SDID design was proposed by [Arkhangelsky et al. \(2021\)](#) and the SCD was considered in their comparison studies in [Ferman and Pinto \(2021\)](#) and [Chen \(2023\)](#). Like other methods, they are distinguished by the way the weights and within-group differencing method are chosen in the GMC. Details follow below.

⁴Extrapolation in the SC literature usually refers to the use of a match lying outside the convex combination of the outcomes in the donor pool. On the other hand, extrapolation here refers to the use of the same weights obtained from the pre-treatment fit to produce a surrogate for the post-treatment counterfactual untreated mean outcome.

Thus, $\text{MER}_d(w)$ measures the matching error in regret form for the choice of weight w , whereas $\Delta\text{MER}(w)$ measures how well the matching error in regret is extrapolated from the pre-treatment periods to the post-treatment periods. Let w^{DID} be the population-level weights specified by the DID design and w^{SCD} those by the SCD design. Our main result shows that

$$\begin{aligned} \text{DID regret-dominates SCD, if } \Delta\text{MER}(w^{\text{SCD}}) &> \text{MER}_1(w^{\text{DID}}) + C\epsilon_n \text{ and} \\ \text{SCD regret-dominates DID, if } \Delta\text{MER}(w^{\text{SCD}}) &< \text{MER}_1(w^{\text{DID}}) - C\epsilon_n, \end{aligned}$$

where ϵ_n is a term that vanishes at the parametric rate (with respect to the size of the cross-sectional units) and C is a universal positive constant. Therefore, the domination of SCD over DID depends on the relative size of the matching error in regret to the extrapolation error.

One might wonder when the designs of DID and SCD are “equivalent”, in the sense that

$$w^{\text{DID}} = w^{\text{SCD}}.$$

We demonstrate that this equivalence holds when both pre-treatment and post-treatment parallel trend assumptions hold simultaneously. This latter condition is implicitly invoked in practice when researchers use pre-treatment parallel trend tests as supporting evidence for the post-treatment parallel trend assumption. Such usage assumes that satisfying the post-treatment parallel trend assumption necessarily implies satisfying the pre-treatment parallel trend assumption (Kahn-Lang and Lang (2020)).⁵ Under these conditions, our results show that DID and SCD employ identical weights and therefore rely on the same identifying assumption. Nevertheless, the finite-sample performance of estimates from these approaches may still differ.

Our complementarity result demonstrates that SCD emerges as a viable alternative to DID when the parallel trend assumption fails. Unlike approaches that robustify DID against the failure of the parallel trend assumption (see Manski and Pepper (2018) and Rambachan and Roth (2023)), SCD is inspired by the SC design and replaces the parallel trend assumption by the existence and stability of matching weights before and after the treatment.⁶ Just as the plausibility of the parallel trend assumption has to be examined in the specific context of application, so does the stable matching weight assumption of SCD.

While SCD has already been considered in the literature (Ferman and Pinto (2021) and Chen (2023)), the uniformly valid asymptotic inference for the SCD design for the groupwise

⁵See Bilinski and Hatfield (2019) also for issues with the usual pre-treatment tests and new proposals of tests addressing them.

⁶There have been variants of DID that do not require parallel trend assumption. For example, Freyaldenhoven *et al.* (2019) considered a linear panel framework where the violation of parallel trends is permitted and identification is achieved by removing possible confounding through the use of covariates. Kwon and Roth (2024) proposed an empirical Bayes approach.

matching setting has not been formally developed to the best of our knowledge. We fill this gap by applying the uniformly valid inference on the simplex-valued weights in [Canen and Song \(2025\)](#) and developing estimation and asymptotic inference methods for SCD. Our Monte Carlo simulations show how the complementarity between SCD and DID manifests in finite sample performance of the estimators.

To illustrate the usefulness of SCD as a causal inference method, we revisit the empirical setting analyzed in [Bohn *et al.* \(2014\)](#) and examine the impact of the 2007 Legal Arizona Workers Act (LAWA) on Arizona’s internal composition. We use CPS data between January 1998 and December 2009 and exploit its cross-sectional dimension to provide valid confidence intervals for the treatment effects estimated by SCD. Following the authors, we include 46 states in Arizona’s donor pool that did not implement any similar regulation during the period of analysis and focus on the population that is most likely to be affected by the policy change: non-citizen Hispanics. We find that Arizona’s share of this demographic group declined by 2 percentage points after LAWA’s enactment on average, consistent with the 1.5 percentage point reduction reported in [Bohn *et al.* \(2014\)](#). The average decrease is 1.2 percentage points larger when looking at Arizona’s proportion of low-educated non-citizen Hispanics among the population aged 15-45. These results are robust to alternative choices of the pre-treatment window and differencing parameters in the SCD design.

Related Literature The literature of SC designs and DID designs is vast and fast growing. We refer the readers to the survey papers by [Abadie \(2021\)](#) for the SC approaches, and [de Chaisemartin and D’Haultfœuille \(2023\)](#) and [Roth *et al.* \(2023\)](#) for the DID designs. Here, we will briefly focus only on some recent studies that attempt to synthesize and/or compare the SC and DID designs.

[Doudchenko and Imbens \(2017\)](#) presented a unifying framework that encompasses four major causal inference approaches (SC, DID, matching and regression). [Kellogg *et al.* \(2021\)](#) compared SC with matching methods in terms of extrapolation and interpolation bias and proposed a model average estimator of the two approaches. [Arkhangelsky *et al.* \(2021\)](#) synthesized SC and DID into what they called SDID (synthetic difference-in-differences). [Xu \(2017\)](#) assumed a linear factor structure for untreated potential outcomes and proposed extrapolating the estimated factor loadings and factors to accommodate time-varying confounders. In contrast, our focus is on formalizing the complementarity between DID and SC under short panels without relying on a factor structure.

Our findings contrast with recent work by [Ferman and Pinto \(2021\)](#) and [Chen \(2023\)](#), both of which showed that SCD dominates DID. [Ferman and Pinto \(2021\)](#) employ a linear factor model to demonstrate that SCD’s asymptotic mean-squared error dominates that of DID under

large- $|\mathcal{T}_0|$ asymptotics. Our analysis differs fundamentally: we impose no linear factor structure and examine environments with many individuals over short time periods. [Chen \(2023\)](#) compares SCD and DID through a regret analysis but assumes long time periods, rendering his results uninformative for our setting. Moreover, Chen’s risk definition assumes treatment timing is drawn randomly from an approximately uniform distribution. Our analysis takes the opposite extreme, assuming fully known treatment timing.

Closely related is [Sun *et al.* \(2025\)](#), who also consider short panels and propose identification and inference on the ATT accommodating both DID and SC settings. However, their parallel trend assumption is strong enough to identify the ATT under any design, whereas our comparison uses a weaker version sufficient for identification via DID but not necessarily SC. Also related is [Liu \(2025\)](#), who introduces a synthetic parallel trend assumption similar to our generalized matching condition. Independently, she found that DID methods can be viewed as matching with group-size-based weights. Her focus, however, is on the identified set of counterfactuals and inference thereon, without further restrictions on weights other than that they sum up to one or simplex-constrained. By contrast, our generalized matching condition is parametrized by weights and time-differencing methods, as our goal is to characterize the core identifying scheme underlying both DID and SC, two designs that differ precisely in these choices. The resulting complementarity and equivalence results between DID and SCD are, to our knowledge, new.

The paper is organized as follows. In [Section 2](#), we provide a basic set-up of causal inference with groupwise matching and the notion of GMC. In [Section 3](#), we show how GMC provides a unifying identification scheme that encompasses various causal inference designs. In [Section 4](#), we present a regret analysis that shows how the designs of DID and SCD are complementary to each other. In [Section 5](#), we provide estimation and inference of the SCD methods and results on asymptotic theory, followed by the Monte Carlo simulation results. In [Section 6](#), we present an empirical application. In [Section 7](#), the paper concludes. The mathematical proofs of the results in the paper are found in the Supplemental Note.

2. Causal Inference with Groupwise Matching

2.1. The Set-Up

We consider a setting with the set N of individuals i , divided into $K + 1$ disjoint groups. Let $\mathcal{G} := \{0, 1, \dots, K\}$ be a finite set of group indexes and denote $G_i = j \in \mathcal{G}$ if and only if the individual i belongs to group j . The individuals are observed over time $t \in \mathcal{T} := \{1, 2, \dots, T\}$. Each individual belongs either to the treatment group ($D_i = 1$) or the untreated group ($D_i = 0$). All the groups stay untreated until time $t = T^* > 1$, and at time T^* , those individuals

with $D_i = 1$ are treated. We partition $\mathcal{T}$ into $\mathcal{T}_0$ and $\mathcal{T}_1$, with

$$\mathcal{T}_0 = \{1, \dots, T^* - 1\}, \text{ and } \mathcal{T}_1 = \{T^*, \dots, T\}.$$

The set $\mathcal{T}_0$ collects the time periods before treatment occurs and $\mathcal{T}_1$ the time periods following treatment. Hereafter, we call $\mathcal{T}_0$ the pre-treatment periods and $\mathcal{T}_1$ the post-treatment periods.

The potential outcome of an individual i in time t when the individual is treated is denoted by $Y_{i,t}(1)$ and otherwise $Y_{i,t}(0)$. Define the average treatment effect on the treated in period t as

$$(2.1) \quad \theta_t^* = \mathbf{E}[Y_{i,t}(1) - Y_{i,t}(0) \mid D_i = 1].$$

The observed outcomes, $Y_{i,t}$, are defined as follows:

$$(2.2) \quad Y_{i,t} = D_i Y_{i,t}(1) + (1 - D_i) Y_{i,t}(0).$$

We introduce basic conditions maintained throughout the paper.

Assumption 2.1. (i) $P\{D_i = 1\} > 0$ and $P\{G_i = j, D_i = 0\} > 0$, for all $j = 1, \dots, K$ and $i \in N$.
(ii) For each $i \in N$ such that $D_i = 1$, we have $Y_{i,t}(1) = Y_{i,t}(0)$, whenever $t < T^*$.

Assumption 2.1(i) says that each group consists of a positive fraction of individuals in the population. Assumption 2.1(ii) supposes *no anticipation* of treatment. It says that each individual's potential outcome at time t before the treatment at time s is the same as that when the person is never treated. As we show in the following example, this setting accommodates the DID with discrete covariates and the staggered adoption in the DID literature (Callaway and Sant'Anna (2021), de Chaisemartin and D'Haultfœuille (2020), and Sun and Abraham (2021)).

Example 2.1 (Difference-in-Differences with Discrete Covariates). Consider the two-period DID setting with covariates. Suppose that $T = 2$ and each individual $i \in N$ is endowed with the discrete covariate $X_i \in \{x_1, \dots, x_K\}$, and belongs to either the treated group ($D_i = 1$) or the untreated group ($D_i = 0$). The potential outcomes are given as $Y_{i,t}(1)$ and $Y_{i,t}(0)$ for $t = 1, 2$. This setting is mapped to the above general setting, by setting $G_i = j$ if and only if $X_i = x_j$. The treatment time T^* is taken to be the second time, i.e., $T^* = 2$. ■

Example 2.2 (Difference-in-Differences with Staggered Adoption). In this example, each group may experience different treatment timing. First, we define $D_{j,t}$ as a binary variable equal to one if group j is treated at time t and zero otherwise. We decompose the group index set $\mathcal{G}$ as follows:

$$\mathcal{G} = \{0\} \cup \mathcal{G}_{\text{don}},$$

where $\mathcal{G}_{\text{don}} = \{1, 2, \dots, K\}$. Our focus is on the effect of the first-time treatment of the target group 0. Let T^* be the first time that group 0 is treated. The groups in $\mathcal{G}_{\text{don}}$ have not been treated until after $t = T$ and thus form a “donor pool” for the target group.

We assume that $T^* > 1$. For each individual $i \in N$ such that $D_{G_i,t} = 1$ for some $t \in \mathcal{T}$, let

$$T_i = \min\{t \in \mathcal{T} : D_{G_i,t} = 1\}$$

be the period in which individual i is first treated. Hence, $T_i = T^*$ for all i in the target group 0. We set $T_i = 0$ for an individual who is never treated. We let $Y_{i,t}^*(s)$ be the potential outcome of individual i at time t when group G_i is first treated at time $1 \leq s \leq T$. The quantity $Y_{i,t}^*(0)$ represents the potential outcome of the individual i when never treated.

Our parameter of interest is the average treatment effect on the target group 0:

$$\theta_t^* = \mathbf{E}[Y_{i,t}^*(T^*) - Y_{i,t}^*(0) \mid G_i = 0], \text{ for } t \in \mathcal{T}_1.$$

The observed outcomes are given as follows:

$$(2.3) \quad Y_{i,t} = D_{G_i,t} Y_{i,t}^*(T_i) + (1 - D_{G_i,t}) Y_{i,t}^*(0).$$

We consider the following assumptions.

- (i) $Y_{i,t}^*(s) = Y_{i,t}^*(0)$ for all $t < s$ with $t, s \in \mathcal{T}$ and $i \in N$.
- (ii) $D_{j,1} = 0$ for all $j \in \mathcal{G}$, if $T \geq 2$.
- (iii) $D_{j,t} \leq D_{j,t+1}$ for all $j \in \mathcal{G}$ and $t = 1, \dots, T-1$.
- (iv) (a) $T_i = T^*$ whenever $G_i = 0$, and (b) $T_i = 0$ whenever $G_i \in \mathcal{G}_{\text{don}}$.
- (v) For each $j \in \mathcal{G}$, $P\{G_i = j\} > 0$.

Condition (i) is the standard no-anticipation assumption. The next two conditions require that no group is treated at the initial period (ii) and that treatments arise in a staggered manner (iii). Furthermore, (iv) stipulates that T^* is the first time group 0 is treated and that the donor pool groups remain untreated through period T , while (v) ensures that each group has a positive membership probability. Note that under Condition (iii), $D_{G_i,t} = 0$ indicates that individual i has not been treated by time t , whereas $D_{G_i,t} = 1$ indicates treatment at or before time t .

Now, we show how this setting maps to the general setting above. We define

$$D_i = 1\{G_i = 0\}, \text{ for all } i \in N.$$

In this case, individuals with $D_i = 1$ represent people belonging to group 0, while those with $D_i = 0$ correspond to people who belong to groups that remain untreated until after T . Then, we set $Y_{i,1}(1) = Y_{i,1}(0) = Y_{i,1}^*(0)$ for all $i \in N$. And, for $1 < t \in \mathcal{T}$, we set

$$Y_{i,t}(1) = Y_{i,t}^*(T_i) \text{ and } Y_{i,t}(0) = Y_{i,t}^*(0), \text{ for all } i \in N.$$

Hence, for $t \in \mathcal{T}_1$, we can write

$$\theta_t^* = \mathbb{E}[Y_{i,t}(1) - Y_{i,t}(0) \mid D_i = 1].$$

Furthermore, the observed outcomes, $Y_{i,t}$, defined in (2.3) coincide with those defined in (2.2), for all $i \in N$ and $t \in \mathcal{T}$. It is not hard to see that Assumption 2.1 is also satisfied. ■

2.2. Generalized Matching

The causal effect of a treatment in an experimental setting is captured by the difference in outcomes between the treated and control groups. In a non-experimental setting, a control group is not available, which requires constructing a comparison group as a surrogate for the control group. This approach is valid only if the outcomes of the comparison group are “matched” to the counterfactual untreated outcomes of the treated group.

To express this idea, for each $j \in \mathcal{G}_{\text{don}}$, and $t \in \mathcal{T}$, define

$$m_{0,t}(0) = \mathbb{E}[Y_{i,t}(0) \mid D_i = 1] \text{ and } m_{j,t}(0) = \mathbb{E}[Y_{i,t}(0) \mid D_i = 0, G_i = j],$$

and their observed counterparts:

$$m_{0,t} = \mathbb{E}[Y_{i,t} \mid D_i = 1] \text{ and } m_{j,t} = \mathbb{E}[Y_{i,t} \mid D_i = 0, G_i = j].$$

Let $\Delta_{|\mathcal{T}_0|-1} \subset \mathbf{R}^{|\mathcal{T}_0|}$ be the simplex in $\mathbf{R}^{|\mathcal{T}_0|}$. Given $\lambda = (\lambda_s)_{s \in \mathcal{T}_0} \in \Delta_{|\mathcal{T}_0|-1}$, and $j \in \mathcal{G}$, we introduce a **within-group λ -differencing** of $m_{j,t}(0)$ and $m_{j,t}$ as follows:

$$\mu_{j,t}(0; \lambda) = m_{j,t}(0) - \sum_{s \in \mathcal{T}_0} m_{j,s}(0) \lambda_s \text{ and } \mu_{j,t}(\lambda) = m_{j,t} - \sum_{s \in \mathcal{T}_0} m_{j,s} \lambda_s.$$

One example is to subtract the most recent pre-treatment potential outcome so that

$$(2.4) \quad \mu_{j,t}(0; \lambda^{\text{DID}}) = m_{j,t}(0) - m_{j,T^*-1}(0),$$

where $\lambda_s^{\text{DID}} = 1\{s = T^* - 1\}$. This differencing is adopted in the DID designs of [Callaway and Sant’Anna \(2021\)](#), [Sun and Abraham \(2021\)](#), and [de Chaisemartin and D’Haultfœuille \(2020\)](#). Another example is the uniform differencing

$$\mu_{j,t}(0; \lambda^{\text{unif}}) = m_{j,t}(0) - \frac{1}{|\mathcal{T}_0|} \sum_{s \in \mathcal{T}_0} m_{j,s}(0),$$

where $\lambda_s^{\text{unif}} = 1/|\mathcal{T}_0|$ (see [Wooldridge \(2021\)](#) and [Lee and Wooldridge \(2024\)](#) for DID applications and [Arkhangelsky et al. \(2021\)](#) and [Chen \(2023\)](#) in the SDID and SCD designs).

For each $w \in \Delta_{K-1}$, $\lambda \in \Delta_{|\mathcal{T}_0|-1}$ and $t \in \mathcal{T}$, we define a **between-group w -differencing** of the λ -differenced average potential outcomes as follows:

$$e_t(\lambda, w) = \mu_{0,t}(0; \lambda) - \sum_{j=1}^K \mu_{j,t}(0; \lambda) w_j.$$

We call the quantity $e_t(\lambda, w)$ the **matching error** from matching $\mu_{0,t}(0; \lambda)$ with a weighted average of group means $\mu_{j,t}(0; \lambda)$ in the donor pool.

Definition 2.1. Let $\lambda \in \Delta_{|\mathcal{T}_0|-1}$ be given.

(i) For $w \in \Delta_{K-1}$, we say that **Generalized Matching Condition (GMC) holds at (λ, w)** , if

$$e_t(\lambda, w) = 0, \text{ for all } t \in \mathcal{T}_1.$$

(ii) We say that **Stable Matching Condition (SMC) holds at λ** , if for some $w \in \Delta_{K-1}$,

$$e_t(\lambda, w) = 0, \text{ for all } t \in \mathcal{T}.$$

Suppose that GMC holds at (λ, w) . This means that we can *transfer* the w -weighted average of the expected untreated potential outcomes in the donor pool (after the within-group λ -differencing) to the corresponding counterfactual quantity in the target group. To see the role of GMC in identifying θ^* , we decompose the target parameter θ_t^* as follows:⁷

$$(2.5) \quad \theta_t^* = \theta_t(\lambda, w) - e_t(\lambda, w),$$

where

$$\theta_t(\lambda, w) = \mu_{0,t}(\lambda) - \sum_{j=1}^K \mu_{j,t}(\lambda) w_j.$$

Once we invoke GMC at (λ, w) , we obtain the following identification:

$$(2.6) \quad \theta^* = \theta(\lambda, w),$$

where

$$\theta(\lambda, w) = [\theta_{T^*}(\lambda, w), \dots, \theta_T(\lambda, w)]'.$$

As we will see later, many causal inference designs are distinguished by how λ and w are specified. Due to the use of groupwise matching, the estimand $\theta(\lambda, w)$ depends on the individual-level observations only through the group averages $m_{j,t}$. Hence, the causal inference framework accommodates both repeated cross-sections and panel data.

2.2.1. Generalized Matching Conditions under a Linear Factor Model. The literature often specifies the potential outcomes as a linear factor model to analyze a causal inference method (see [Abadie et al. \(2010\)](#), [Xu \(2017\)](#), [Ferman and Pinto \(2021\)](#), [Arkhangelsky et al. \(2021\)](#)). While our results do not rely on a linear factor model, it is interesting to study the implication of this model for GMC.

⁷The proof is simple and found in the Supplemental Note.

Consider the untreated potential outcomes specified as a factor model:

$$(2.7) \quad Y_{i,t}(0) = \Lambda_i' F_t + \varepsilon_{i,t} \text{ and } Y_{i,t}(1) = Y_{i,t}(0) + \tau_{i,t},$$

where $\Lambda_i \in \mathbf{R}^M$ denotes the factor loading of individual i , $F_t \in \mathbf{R}^M$, the factor at period t , $\varepsilon_{i,t}$, idiosyncratic components, and $\tau_{i,t}$ denotes the time-varying, heterogeneous treatment effects. We assume that $D_i = 1$ if and only if $G_i = 0$, so that there is a treated group $G_i = 0$ and all other groups are control groups. The number M represents the number of factors. As for the factor model, we make the following assumption.

Assumption 2.2. (i) The factor loadings, Λ_i , $i \in N$, are i.i.d.

(ii) The factors, F_t , $t \in \mathcal{T}$, are constants.

(iii) For each $i \in N$ and $t \in \mathcal{T}$, $\mathbf{E}[\varepsilon_{i,t} \mid G_i] = 0$.

The condition (ii) is motivated by the data structure of our setting where our observations span over only a short period. Hence, the distribution of the factors is not consistently estimable even if the factors are observed (see [Kuersteiner and Prucha \(2020\)](#)). The rest of the analysis carries over to the case of stochastic factors, once we replace probabilities and expectations by conditional probabilities and conditional expectations given the factors.

We introduce the λ -differenced versions of the factors:

$$F_t(\lambda) = F_t - \sum_{s \in \mathcal{T}_0} F_s \lambda_s,$$

and collect them into a matrix, $\mathbf{F}(\lambda) = [F_{T^*}(\lambda), \dots, F_T(\lambda)]$. Then, it follows that when $\mathbf{F}(\lambda)$ is full row rank, the GMC holds at some (λ, w) if and only if the GMC holds at $(\tilde{\lambda}, w)$ for all $\tilde{\lambda} \in \Delta_{|\mathcal{T}_0|-1}$. Hence, the choice of λ in the λ -differencing does not matter for identification, as long as the GMC holds at some λ . We formalize this into the following proposition.

Proposition 2.1. Suppose that Assumption 2.2 holds and let $\lambda \in \Delta_{|\mathcal{T}_0|-1}$ and $w \in \Delta_{K-1}$.

(i) GMC holds at (λ, w) .

(ii) GMC holds at $(\tilde{\lambda}, w)$ for all $\tilde{\lambda} \in \Delta_{|\mathcal{T}_0|-1}$.

(iii)

$$(2.8) \quad \mathbf{E}[\Lambda_i \mid G_i = 0] = \sum_{j=1}^K \mathbf{E}[\Lambda_i \mid G_i = j] w_j.$$

Then, (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i). If, furthermore, $\mathbf{F}(\lambda)$ is full row rank for some $\lambda \in \Delta_{|\mathcal{T}_0|-1}$, then the three statements are equivalent for all $w \in \Delta_{K-1}$.

The full row rank condition for $\mathbf{F}(\lambda)$ requires that $M \leq |\mathcal{T}_1|$, that is, the number of the factors is less than the number of the post-treatment periods. This condition is immediately satisfied in the case of a single-factor model. The full rank condition is not required for

identification of θ^* once (2.8) is satisfied for some (λ, w) . However, if the full rank condition holds, we have

$$\theta^* = \theta(\tilde{\lambda}, w) \text{ for all } \tilde{\lambda} \in \Delta_{|\mathcal{T}_0|-1},$$

i.e., θ^* is overidentified. For the identification, the researcher does not need to specify the within-group differencing, $\tilde{\lambda}$, that satisfies GMC.

3. Causal Inference Methods using Generalized Matching

In this section, we show how GMC is used as key identifying restrictions in various causal inference designs. We classify them into two categories, one using GMC with weights based on group sizes and the other using GMC with weights based on pre-treatment fit.

3.1. Matching with Weights Based on Group Sizes

3.1.1. Randomized Controlled Trials. First, note that the design of randomized control trials (RCT) can be viewed as a degenerate example of GMC, where we do not have the initial period of no treatment, i.e., $|\mathcal{T}_0| = 0$. The design assumes that the potential outcomes are independent of the treatment status and yields the following form of GMC at $(0, 1)$:

$$e_1(0, 1) = \mathbf{E}[Y_{i,1}(0) | D_i = 1] - \mathbf{E}[Y_{i,1}(0) | D_i = 0] = 0.$$

The parameter θ_t^* is identified as

$$\theta_1^* = \theta_1(0, 1) = \mathbf{E}[Y_{i,1} | D_i = 1] - \mathbf{E}[Y_{i,1} | D_i = 0].$$

3.1.2. Unconfoundedness Condition with Discrete Covariates. Consider the unconfoundedness condition on discrete random vector X_i with support $\{x_1, \dots, x_K\}$:

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i | X_i.$$

Our object of interest is the ATT, $\theta_1^* = \mathbf{E}[Y_{i,1}(1) - Y_{i,1}(0) | D_i = 1]$. We take $G_i = j$ if and only if $X_i = x_j$. The unconfoundedness condition, together with the overlap condition, $P\{D_i = 1 | G_i = j\} > 0$ for all j , yields the following:

$$\begin{aligned} 0 &= \mathbf{E}[Y_{i,1}(0) | D_i = 1] - \mathbf{E}[\mathbf{E}[Y_{i,1}(0) | D_i = 0, G_i] | D_i = 1] \\ &= \mathbf{E}[Y_{i,1}(0) | D_i = 1] - \sum_{j=1}^K \mathbf{E}[Y_{i,1}(0) | D_i = 0, G_i = j] w_j^C, \end{aligned}$$

where

$$(3.1) \quad w_j^C = P\{G_i = j | D_i = 1\}.$$

The identifying assumption is essentially GMC at $(0, w^C)$ with $w^C = (w_j^C)$.

3.1.3. Difference-in-Differences. First, consider the two-period setting $T = 2$ of the classic DID, where $Y_{i,t}(d)$ denotes the potential outcome at time $t = 1, 2$ at the treatment state $d \in \{0, 1\}$. We consider the following form of the RCT after within-group differencing:

$$Y_{i,t}(0; \lambda) \perp\!\!\!\perp D_i,$$

where $Y_{i,t}(0; \lambda) = Y_{i,t}(0) - \sum_{s \in \mathcal{T}_0} \lambda_s Y_{i,s}(0)$ and $\lambda = 1$ (since $|\mathcal{T}_0| = 1$). This yields the following parallel trend assumption:

$$\begin{aligned} 0 &= \mathbf{E}[Y_{i,t}(0; \lambda) \mid D_i = 1] - \mathbf{E}[Y_{i,t}(0; \lambda) \mid D_i = 0] \\ &= e_t(\lambda, 1) = e_t(1, 1). \end{aligned}$$

Thus, the parallel trend assumption is nothing but the GMC at $(1, 1)$.

3.1.4. Difference-in-Differences with Discrete Covariates. We consider the two-period setting as before, but consider the following form of the unconfoundedness condition instead:⁸

$$Y_{i,t}(0; \lambda) \perp\!\!\!\perp D_i \mid X_i,$$

with $X_i \in \{x_1, \dots, x_K\}$ being a discrete random vector. As before, if we let $G_i = j$ if and only if $X_i = x_j$, this condition, together with the overlap condition, yields GMC at (λ, w^C) as follows:

$$\begin{aligned} 0 &= \mathbf{E}[Y_{i,t}(0; \lambda) \mid D_i = 1] - \mathbf{E}[\mathbf{E}[Y_{i,t}(0; \lambda) \mid D_i = 0, G_i] \mid D_i = 1] \\ &= e_t(\lambda, w^C) = e_t(1, w^C), \end{aligned}$$

where $w^C = (w_j^C)$ with w_j^C defined in (3.1).

3.1.5. Difference-in-Differences with Staggered Adoption. We consider the staggered adoption setting of Example 2.2. We show that GMC characterizes the key identifying assumptions in the DID settings. Let us consider the parallel trend assumptions (PTA) used in the literature. Let $\Delta Y_{i,t}(0) = Y_{i,t}(0) - Y_{i,t-1}(0)$, for $t \in \{2, \dots, T\}$. Consider the two types of PTA as follows.

PTA-I: $\mathbf{E}[\Delta Y_{i,t}(0) \mid G_i = 0] = \mathbf{E}[\Delta Y_{i,t}(0) \mid G_i \in \mathcal{G}_{\text{don}}]$, for all $t \in \mathcal{T}_1$.

PTA-II: $\mathbf{E}[\Delta Y_{i,t}(0) \mid G_i = 0] = \mathbf{E}[\Delta Y_{i,t}(0) \mid G_i = j]$, for all $t \in \mathcal{T}_1$ and $j \in \mathcal{G}_{\text{don}}$.

PTA-I states that the average of the untreated potential outcomes of group 0 and those in its donor pool would have evolved in parallel in the absence of treatment (Callaway and

⁸The unconfoundedness condition for within-group differenced outcomes was studied by Heckman *et al.* (1997). They showed the efficacy of the differencing using Job Training Program Act (JTPA) data. See Smith and Todd (2005) for a similar observation using National Supported Work (NSW) data.

Sant’Anna (2021)). PTA-II is a stronger version of PTA-I, imposing parallel trends of untreated outcomes across all groups (similar to the exogeneity condition in de Chaisemartin and D’Haultfoeuille (2020) and Sun and Abraham (2021)).

The following result shows a close connection between the PTA and the GMC.⁹

Proposition 3.1. *Suppose that Assumption 2.1 holds, and let*

$$w^{\text{DID}} = [w_1^{\text{DID}}, \dots, w_K^{\text{DID}}]',$$

where

$$(3.2) \quad w_j^{\text{DID}} = P \{G_i = j \mid G_i \in \mathcal{G}_{\text{don}}\}.$$

Then, the following statements hold.

- (i) PTA-I holds if and only if GMC holds at $(\lambda^{\text{DID}}, w^{\text{DID}})$.
- (ii) PTA-II holds if and only if GMC holds at $(\lambda^{\text{DID}}, w)$ for all $w \in \Delta_{K-1}$.

This proposition shows that the PTA is represented as GMC. Thus, the target parameter θ^* is identified as $\theta(\lambda^{\text{DID}}, w^{\text{DID}})$ under either PTA-I or PTA-II. Instead of choosing the weight w based on the pre-treatment matching of the potential outcomes as in the SC design, the DID design simply chooses the weight w to be the group size-based one w^{DID} in (3.2). Under the stronger version PTA-II, the choice of the weight w is irrelevant, as GMC holds for all weights.

It is interesting to note that the identification scheme (2.6) is related to the proposal by Sun et al. (2025).¹⁰ The unconditional version of their model (without covariates) involves PTA-II, which is equivalent to GMC at $(\lambda^{\text{DID}}, w)$ for all $w \in \Delta_{K-1}$, and the SC assumption which is tantamount to SMC at 0, with w^{SC} identified as the weight w satisfying $e_t(0, w) = 0$ for all $t \in \mathcal{T}_0$. It is not hard to see that both PTA-II and the SC assumption imply GMC at $(\lambda^{\text{DID}}, w^{\text{SC}})$. Hence, we can identify θ^* as $\theta(\lambda^{\text{DID}}, w^{\text{SC}})$ when either PTA-II or the SC assumption holds. This is the essence of their doubly robust identification of the ATT in Sun et al. (2025). However, PTA-II is stronger than PTA-I and the latter is enough to identify the ATT in the DID design.

3.2. Matching with Weights Based on Pre-Treatment Fit

The literature of SC inspires various causal inference methods in the groupwise matching setting. These methods are distinct from the previous methods, as they rely on GMC with

⁹The connection between PTA and GMC can be viewed as an extension of the observation in Doudchenko and Imbens (2017) to the setting of multiple donor groups. See Liu (2025) for a related result.

¹⁰Sun et al. (2025) also considered conditioning on covariates, and for estimation, proposed using an estimated weight $\hat{w}^{\text{SC}}$ that is not restricted to the simplex Δ_{K-1} . For brevity, we do not consider conditioning on covariates throughout the paper and focus on the main conceptual difference between the two approaches of DID and SCD.

weights based on the pre-treatment fit of the outcomes. For the following examples, we focus on the setting of staggered adoption in Example 2.2 and Section 3.1.5.

3.2.1. Synthetic Control. The synthetic control method applied to the setting of groupwise matching identifies

$$\theta_t^* = \mathbb{E}[Y_{i,t} \mid G_i = 0] - \sum_{j=1}^K \mathbb{E}[Y_{i,t} \mid G_i = j] w_j^*(0),$$

where $w^*(0) = [w_1^*(0), \dots, w_K^*(0)]'$ is a minimizer of $Q(w)$ over $w = [w_1, \dots, w_K]' \in \Delta_{K-1}$, with

$$Q(w) = \sum_{t=1}^{T^*-1} \left(m_{0,t} - \sum_{j=1}^K m_{j,t} w_j \right)^2.$$

This identification scheme relies on GMC holding at $(0, w^*(0))$.

The choice of $w^*(0)$ is motivated as follows. First, consider the weight $w_j^*(0)$ that gives the population-level perfect pre-treatment matching: for all $t \in \mathcal{T}_0$,

$$(3.3) \quad \mathbb{E}[Y_{i,t} \mid G_i = 0] = \sum_{j=1}^K \mathbb{E}[Y_{i,t} \mid G_i = j] w_j^*(0).$$

Then, we assume that the same weight $w_j^*(0)$ yields the perfect post-treatment matching as well, i.e., (3.3) holds for $t \in \mathcal{T}_1$. In other words, the SC design relies on SMC at 0.

3.2.2. Synthetic Control with Differencing. The synthetic control with differencing (SCD) applies the SC design after applying a within-group λ -differencing of the potential outcomes (see Chen (2023) and references therein). Let λ be a researcher-chosen differencing method. For example, one may choose $\lambda = \lambda^{\text{DID}}$ or λ^{unif} . Define $Q : \Delta_{|\mathcal{T}_0|-1} \times \Delta_{K-1} \rightarrow \mathbf{R}$ as follows:

$$(3.4) \quad Q(\lambda, w) = \sum_{t=1}^{T^*-1} \left(\mu_{0,t}(\lambda) - \sum_{j=1}^K \mu_{j,t}(\lambda) w_j \right)^2,$$

and choose $w^*(\lambda)$ as a minimizer of $Q(\lambda, w)$:

$$(3.5) \quad w^*(\lambda) \in \arg \min_{w \in \Delta_{K-1}} Q(\lambda, w).$$

Then the SCD design invokes GMC at $(\lambda, w^*(\lambda))$ and identifies θ_t^* as follows:

$$\theta_t^* = \theta_t(\lambda, w^*(\lambda)) = \mu_{0,t}(\lambda) - \sum_{j=1}^K \mu_{j,t}(\lambda) w_j^*(\lambda).$$

The GMC at $(\lambda, w^*(\lambda))$ requires that the weight $w^*(\lambda)$ that achieves the optimal pre-treatment matching delivers the perfect post-treatment matching.

Again, the choice of w^* can be motivated in terms of SMC. We first consider $w_j^*(\lambda)$ such that

$$\mu_{0,t}(\lambda) = \sum_{j=1}^K \mu_{j,t}(\lambda) w_j^*(\lambda),$$

for $t \in \mathcal{T}_0$. Then, the SCD design assumes that this weight $w_j^*(\lambda)$ delivers the perfect post-treatment matching as well, i.e., SMC holds at λ .

3.2.3. Synthetic Difference-in-Differences. [Arkhangelsky et al. \(2021\)](#) developed the synthetic difference-in-differences (SDID) method, which integrates the synthetic control approach with the difference-in-differences design. While their original framework targeted a data structure different from our groupwise matching setting, the core idea of SDID can be adapted to this setting.

To facilitate the comparison, suppose that our target parameter is the same as before θ^* . Consider the time-differenced version of the population objective function in [Arkhangelsky et al. \(2021\)](#):

$$(3.6) \quad \tilde{Q}(\lambda, w) = \sum_{k=1}^K \left(\sum_{s \in \mathcal{T}_1} \left\{ \mu_{k,s}(\lambda) - \sum_{j=1}^K \mu_{j,s}(\lambda) w_j \right\} \right)^2.$$

We let λ^{unif} and w^{unif} be the uniform weights given by $\lambda_s^{\text{unif}} = 1/|\mathcal{T}_0|$ and $w_j^{\text{unif}} = 1/K$. Then, the identification strategy of the SDID can be formulated as follows:

$$\theta_t^* = \theta_t(\lambda^*(w^{\text{unif}}), w^*(\lambda^{\text{unif}})),$$

where

$$(3.7) \quad \lambda^*(w) = \arg \min_{\lambda \in \Delta_{|\mathcal{T}_0|-1}} \tilde{Q}(\lambda, w).$$

Thus, the identification strategy invokes GMC at $(\lambda^*(w^{\text{unif}}), w^*(\lambda^{\text{unif}}))$.¹¹

Note that unless $\lambda^{\text{unif}} = \lambda^*(w^{\text{unif}})$, the SDID design is not reduced to the SCD design in terms of GMC. More specifically, we cannot motivate the weight $w^*(\lambda^{\text{unif}})$ using SMC. This is because the within-group differencing used for the pre-treatment matching (λ^{unif}) is different from that used for the post-treatment matching ($\lambda^*(w^{\text{unif}})$). Since the within-group differencing changes after the treatment, we cannot say that SDID extrapolates the weight from the pre-treatment fit to the post-treatment periods like SCD. In other words, SCD and SDID are distinct designs.

¹¹Here the optimization problems defining $\lambda^*(w^{\text{unif}})$ and $w^*(\lambda^{\text{unif}})$ are equivalent to those proposed by [Arkhangelsky et al. \(2021\)](#) without regularization.

TABLE 1. Generalized Matching Conditions of Causal Inference Methods

Research Designs	Generalized Matching Conditions
<i>Using Weights Based on Group Sizes</i>	
RCT	GMC holds at $(0, 1)$
Unconfoundedness	GMC holds at $(0, w^C)$
DID with Two Periods	GMC holds at $(1, 1)$
DID with Two Periods and Discrete Covariates	GMC holds at $(1, w^C)$
DID with Staggered Adoption under PTA-I	GMC holds at $(\lambda^{\text{DID}}, w^{\text{DID}})$
DID with Staggered Adoption under PTA-II	GMC holds at $(\lambda^{\text{DID}}, w)$, for all $w \in \Delta_{K-1}$
<i>Using Weights Based on Pre-Treatment Fit</i>	
SC	SMC holds at 0 $\Rightarrow$ GMC holds at $(0, w^*(0))$
SCD	SMC holds at $\lambda \Rightarrow$ GMC holds at $(\lambda, w^*(\lambda))$
SDID	GMC holds at $(\lambda^*(w^{\text{unif}}), w^*(\lambda^{\text{unif}}))$

Notes: The table shows how GMC is used for various causal inference methods. Here, recall that $w_j^{\text{DID}} = P\{G_i = j \mid G_i \in \mathcal{G}_{\text{don}}\}$, $\lambda_s^{\text{DID}} = 1\{s = T^* - 1\}$, $w_j^{\text{unif}} = 1/K$ and $\lambda_s^{\text{unif}} = 1/(T^* - 1)$, and $w^*(\lambda)$ and $\lambda^*(w)$ are the solutions to the optimization problems in (3.5) and (3.7), respectively. The SCD method is based on the researcher-determined λ , for example, either $\lambda = \lambda^{\text{DID}}$ or $\lambda = \lambda^{\text{unif}}$. Note that while the DID method adopts the identified quantities w^{DID} and λ^{DID} directly, the SC, SDID and SCD methods need to invoke rank conditions to identify $w^*(\lambda)$ and $\lambda^*(w)$.

In summary, the major causal inference designs invoke different types of the GMC. Each type involves a choice of a differencing method (λ) and the groupwise matching weights (w). Table 1 summarizes the comparison of the designs in terms of GMC.

4. A Comparison Between DID and SCD

4.1. Extended Parallel Trend Assumption

In this section, we compare the two approaches of DID and SCD in the setting of staggered adoption. To facilitate the comparison, we introduce an extended form of PTA. First, for each $t \in \mathcal{T}$, we define

$$e_t^{\text{DID}}(\lambda) = \mathbf{E}[Y_{i,t}(0; \lambda) \mid G_i = 0] - \mathbf{E}[Y_{i,t}(0; \lambda) \mid G_i \in \mathcal{G}_{\text{don}}], \text{ and}$$

$$e_{j,t}^{\text{DID}}(\lambda) = \mathbf{E}[Y_{i,t}(0; \lambda) \mid G_i = 0] - \mathbf{E}[Y_{i,t}(0; \lambda) \mid G_i = j], \text{ for } j \in \mathcal{G}_{\text{don}}.$$

The quantities $e_t^{\text{DID}}(\lambda)$ and $e_{j,t}^{\text{DID}}(\lambda)$ represent matching errors from matching the λ -differenced average potential untreated outcome for the target group with that from the donor groups. Then, we consider the two types of PTA involving within-group differencing $\lambda \in \Delta_{|\mathcal{T}_0|-1}$.

PTA(λ): $e_t^{\text{DID}}(\lambda) = 0$, for all $t \in \mathcal{T}_1$.

PTA-U(λ): $e_{j,t}^{\text{DID}}(\lambda) = 0$, for all $j \in \mathcal{G}_{\text{don}}$, and for all $t \in \mathcal{T}_1$.

The following proposition shows their connection with the PTA used in the literature.¹²

Proposition 4.1. (i) Suppose that $e_{T^*-1}^{\text{DID}}(\lambda) = 0$ holds for some $\lambda \in \Delta_{|\mathcal{T}_0|-1}$. Then, PTA-I holds if and only if PTA(λ) holds.

(ii) Suppose that for some $\lambda \in \Delta_{|\mathcal{T}_0|-1}$, $e_{j,T^*-1}^{\text{DID}}(\lambda) = 0$ holds for each $j \in \mathcal{G}_{\text{don}}$. Then, PTA-II holds if and only if PTA-U(λ) holds.

Certainly, we have $e_{j,T^*-1}^{\text{DID}}(\lambda^{\text{DID}}) = 0$ for all $j \in \mathcal{G}_{\text{don}}$. Hence, we have

$$\text{PTA-I} \Leftrightarrow \text{PTA}(\lambda^{\text{DID}}) \text{ and } \text{PTA-II} \Leftrightarrow \text{PTA-U}(\lambda^{\text{DID}}).$$

On the other hand, PTA(λ) and PTA-U(λ) allows other choices of λ . The comparison results below apply to such λ 's. From here on, we focus on PTA(λ).

4.2. Regret Analysis

In this section, we compare the research designs of DID and SCD in terms of regret in the staggered adoption setting in Example 2.2. Let $\mathcal{P}$ be the collection of the distributions of the variables under consideration. We fix a within-group differencing $\lambda \in \Delta_{|\mathcal{T}_0|-1}$ such that $e_{T^*-1}^{\text{DID}}(\lambda) = 0$. To facilitate the comparison, we introduce the squared sum of matching errors (SSME): for $w \in \Delta_{K-1}$ and $P \in \mathcal{P}$,

$$\text{SSME}_{d,P}(w) = \frac{1}{|\mathcal{T}_d|} \sum_{t \in \mathcal{T}_d} e_t^2(\lambda, w), \quad d = 0, 1.$$

We make explicit its dependence on $P \in \mathcal{P}$ through the matching errors $e_t(\lambda, w)$. Then, we define the matching error in regret (MER) and the extrapolation error of MER, respectively,

$$\text{MER}_d(\mathbf{w}) = \sup_{P \in \mathcal{P}} \left\{ \text{SSME}_{d,P}(w_P) - \inf_{w \in \Delta_{K-1}} \text{SSME}_{d,P}(w) \right\}, \text{ and}$$

$$\Delta \text{MER}(\mathbf{w}) = \text{MER}_1(\mathbf{w}) - \text{MER}_0(\mathbf{w}),$$

where $\mathbf{w} = (w_P)_{P \in \mathcal{P}}$. The quantity $\text{MER}_d(\mathbf{w})$ captures the matching error of the weights w_P , $P \in \mathcal{P}$, in the maximal regret form, whereas $\Delta \text{MER}(\mathbf{w})$ measures the stability of the MER as

¹²This result also suggests that when DID is used under PTA-I, the target parameter is overidentified using PTA(λ), for any λ satisfying $e_{T^*-1}^{\text{DID}}(\lambda) = 0$. Chen et al. (2025) proposed semiparametrically efficient estimation of ATT under overidentifying restrictions in PTA-U(λ) with covariates.

we move from the pre-treatment regime to the post-treatment regime. We tend to have small $\Delta\text{MER}(\mathbf{w})$ if the post-treatment matching errors are close to the pre-treatment matching errors. Thus, we call $\Delta\text{MER}(\mathbf{w})$ *the extrapolation error*, which essentially captures an error that arises from extrapolating the weight optimized for the pre-treatment data to the post-treatment outcomes.

We compare SCD and DID in terms of MER. We define the population version of the weights by DID and SCD: for each $P \in \mathcal{P}$,¹³

$$w_p^{\text{SCD}} = \arg \min_{w \in \Delta_{K-1}} \text{SSME}_{0,P}(w),$$

and $w_p^{\text{DID}} = [w_{1,p}^{\text{DID}}, \dots, w_{K,p}^{\text{DID}}]'$ with

$$(4.1) \quad w_{j,p}^{\text{DID}} := \frac{\sum_{i \in N} P\{G_i = j\}}{\sum_{i \in N} P\{G_i \in \mathcal{G}_{\text{don}}\}}.$$

We define $\mathbf{w}^{\text{DID}} = (w_p^{\text{DID}})_{p \in \mathcal{P}}$ and $\mathbf{w}^{\text{SCD}} = (w_p^{\text{SCD}})_{p \in \mathcal{P}}$. The SCD invokes the Stable Matching Condition (SMC) and DID the Parallel Trend Assumption (PTA). These assumptions can be formulated in terms of the matching errors:

$$(4.2) \quad \begin{aligned} \text{PTA}(\lambda) &\Leftrightarrow \text{SSME}_{1,P}(\mathbf{w}^{\text{DID}}) = 0 \text{ for all } P \in \mathcal{P}. \\ &\Rightarrow \text{MER}_1(\mathbf{w}^{\text{DID}}) = 0. \\ \text{SMC}(\lambda) &\Leftrightarrow \text{For some } \mathbf{w}, \text{MER}_0(\mathbf{w}) = 0 \text{ and } \text{MER}_1(\mathbf{w}) = 0. \\ &\Rightarrow \Delta\text{MER}(\mathbf{w}^{\text{SCD}}) = 0, \end{aligned}$$

where $\text{SMC}(\lambda)$ denotes that SMC holds at λ . Thus the DID design fails if $\text{MER}_1(\mathbf{w}^{\text{DID}}) \neq 0$ whereas the SCD design fails if $\Delta\text{MER}(\mathbf{w}^{\text{SCD}}) \neq 0$. We will now formalize this complementarity in terms of maximal regret.

For a concrete analysis, we define the sample analog estimator of $m_{j,t}$ and $\mu_{j,t}(\lambda)$ as follows:

$$(4.3) \quad \hat{m}_{j,t} = \frac{1}{n_j} \sum_{i \in N_j} Y_{i,t}, \text{ and } \hat{\mu}_{j,t}(\lambda) = \hat{m}_{j,t} - \sum_{s \in \mathcal{T}_0} \hat{m}_{j,s} \lambda_s,$$

where n_j denotes the number of individuals in the sample belonging to group j . Then, given within-group differencing λ and a choice of data-dependent weight $\hat{w}$, we can estimate θ_t^* as follows:

$$(4.4) \quad \hat{\theta}_t(\lambda, \hat{w}) = \hat{\mu}_{0,t}(\lambda) - \sum_{j=1}^K \hat{\mu}_{j,t}(\lambda) \hat{w}_j.$$

¹³In fact, when the minimizer $w^*(\lambda)$ in (3.5) is unique, the definition of w_p^{SCD} coincides with $w^*(\lambda)$. We assume uniqueness for this regret analysis.

Thus, the selection between the DID and SCD designs boils down to choosing the matching weight $\hat{w}$.

We define the weight $\hat{w}^{\text{DID}}$ for the DID design as follows: $\hat{w}^{\text{DID}} = [\hat{w}_1^{\text{DID}}, \dots, \hat{w}_K^{\text{DID}}]'$ with $\hat{w}_j^{\text{DID}}$ defined as the sample version of $w_{j,p}^{\text{DID}}$ in (4.1):

$$\hat{w}_j^{\text{DID}} := \frac{\sum_{i \in N} 1\{G_i = j\}}{\sum_{i \in N} 1\{G_i \in \mathcal{G}_{\text{don}}\}}, \text{ for } j \in \mathcal{G}_{\text{don}}.$$

The sample weight $\hat{w}_j^{\text{DID}}$ represents the sample fraction of individuals in group j relative to the total units in the donor pool. The DID design suggests estimating θ_t^* as $\hat{\theta}_t(\lambda, \hat{w}^{\text{DID}})$. When $\lambda = \lambda^{\text{DID}}$, this estimator can be viewed as a special case of an estimator proposed by [Callaway and Sant'Anna \(2021\)](#) without covariates. When $K = 1$ and $T^* = 2$ (i.e., the two periods and two groups setting), λ and $\hat{w}^{\text{DID}}$ are equal to one, so we obtain

$$\hat{\theta}_t(\lambda, \hat{w}^{\text{DID}}) = \Delta \bar{Y}_1 - \Delta \bar{Y}_0,$$

where $\Delta \bar{Y}_d$ denotes the first difference average outcomes for the group with treatment status $d = 0, 1$. Thus, we can view $\hat{\theta}_t(\lambda, \hat{w}^{\text{DID}})$ as an extension of the standard DID estimator of θ_t^* to the case with more than two periods and groups.

The SCD design uses the weight $\hat{w}^{\text{SCD}}$ defined as

$$\hat{w}^{\text{SCD}} \in \arg \min_{w \in \Delta_{K-1}} \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} \left(\hat{\mu}_{0,t}(\lambda) - \sum_{j=1}^K \hat{\mu}_{j,t}(\lambda) w_j \right)^2.$$

Hence, the weights $\hat{w}^{\text{SCD}}$ are chosen to minimize the sample version of the pre-treatment SSME. The SCD design suggests estimating θ_t^* by

$$(4.5) \quad \hat{\theta}_t(\lambda, \hat{w}^{\text{SCD}}) = \hat{\mu}_{0,t}(\lambda) - \sum_{j=1}^K \hat{\mu}_{j,t}(\lambda) \hat{w}_j^{\text{SCD}}.$$

Notice that the SCD estimator, $\hat{\theta}_t(\lambda, \hat{w}^{\text{SCD}})$, and the DID estimator, $\hat{\theta}_t(\lambda, \hat{w}^{\text{DID}})$, differ only by the choice of the estimated weights for the donor pool.

To build a decision-theoretic comparison between different research designs, we introduce the average squared error loss from estimating θ_t^* by $\hat{\theta}_t(\lambda, \hat{w})$:

$$\ell_1(\hat{w}) = \frac{1}{|\mathcal{T}_1|} \sum_{t \in \mathcal{T}_1} (\theta_t^* - \hat{\theta}_t(\lambda, \hat{w}))^2.$$

We define the maximal regret associated with the choice of $\hat{w}$:

$$\text{MaxRegret}(\hat{w}) = \sup_{P \in \mathcal{P}} \left\{ \mathbf{E}_P [\ell_1(\hat{w})] - \inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_P [\ell_1(\tilde{w}(Z))] \right\},$$

where $\mathcal{D}$ is the set of Δ_{K-1} -valued functions that are measurable with respect to Z and the random vector Z represents the vector of all the observed random variables.¹⁴ We compare the DID and SCD designs in terms of their maximal regrets.

We introduce assumptions used for the regret analysis. Let $Y_i^*(s) = (Y_{i,t}^*(s))_{t \in \mathcal{T}}$ and $Y_i^* = (Y_i^*(s))_{s \in \mathcal{T} \cup \{0\}}$.

Assumption 4.1. The random vectors, (Y_i^*, G_i) , are independent across $i \in N$, under each $P \in \mathcal{P}$.

This assumption requires that the variables be independent across the cross-sectional units. This condition allows for arbitrary dependence between the potential outcomes across different treatment timing or the time periods. The framework allows for both the settings of repeated cross-sections and panel data. It also allows for factor models for the potential outcomes; we can simply take the factors to be constants.

Assumption 4.2. There exist constants $\pi_0 > 0$, $\bar{m} \geq 1$, and $0 < c \leq 1$, such that for all $j \in \mathcal{G}$, $t \in \mathcal{T}$ and $s \in \mathcal{T} \cup \{0\}$, we have

$$(4.6) \quad \max_{1 \leq i \leq n} \sup_{P \in \mathcal{P}} \mathbf{E}_P[Y_{i,t}^4(s) | G_i = j] \leq \bar{m}^4 \text{ and } \inf_{P \in \mathcal{P}} \frac{1}{n} \sum_{i \in N} P\{G_i = j\} \geq \pi_0,$$

and

$$(4.7) \quad \inf_{P \in \mathcal{P}} \lambda_{\min}(\Gamma'_P \Gamma_P) \geq c|\mathcal{T}_0|,$$

where Γ_P is the $|\mathcal{T}_0| \times K$ matrix whose (t, j) -th entry is given by $\mu_{j,t}(\lambda)$.

The condition (4.6) in Assumption 4.2 requires the existence of uniform upper and lower bounds for the fourth moment of potential outcomes in each group and the probability of the group membership, respectively. The condition (4.7) says that the time-paths of the within-group differenced mean outcomes are not linearly dependent. This condition requires that $|\mathcal{T}_0| \geq K$ and ensures that w_p^{SCD} is identified.

The theorem below presents the regret-comparison result between the DID and SCD designs.

Theorem 4.1. Suppose that Assumptions 4.1-4.2 hold. For each $n \geq 1$, let

$$\epsilon_n := \frac{(K+1)\bar{m}^4}{c} \left\{ \frac{1}{\pi_0 \sqrt{n}} + \frac{1}{\pi_0^2 n} + \exp\left(-\frac{\pi_0 n}{8}\right) \right\},$$

where $\bar{m}$, π_0 and c are the constants in Assumption 4.2. Then, there exists a universal constant $C > 0$ such that the following statements hold for all $n \geq 1$.

¹⁴The expectation $\mathbf{E}_P[\ell_1(\hat{w})]$ is with respect to the distribution of both $\hat{w}$ and $\ell_1(\cdot)$.

(i) If $\Delta\text{MER}(\mathbf{w}^{\text{SCD}}) > \text{MER}_1(\mathbf{w}^{\text{DID}}) + C\epsilon_n$, then,

$$\text{MaxRegret}(\hat{\mathbf{w}}^{\text{SCD}}) > \text{MaxRegret}(\hat{\mathbf{w}}^{\text{DID}}).$$

(ii) If $\Delta\text{MER}(\mathbf{w}^{\text{SCD}}) < \text{MER}_1(\mathbf{w}^{\text{DID}}) - C\epsilon_n$, then,

$$\text{MaxRegret}(\hat{\mathbf{w}}^{\text{SCD}}) < \text{MaxRegret}(\hat{\mathbf{w}}^{\text{DID}}).$$

The result shows that the DID design regret-dominates the SCD design if and only if the extrapolation error of the SCD design dominates the post-treatment MER of the DID design up to a term that vanishes at the parametric rate $\sqrt{n}$.

The DID design specifies the matching weights to be the group size-based weights, and hence does not need to invoke extrapolation of weights from the pre-treatment fit. On the other hand, the SCD design obtains the weights that exhibit a best pre-treatment fit, and extrapolates the weights to the post-treatment periods. The comparison shows when the DID design or the SCD design is appropriate or not. The DID design is not appropriate in a setting where it is doubtful that the relevance of each group in matching is proportional to the size of the group, whereas the SCD design is not appropriate if the relevance of the groups in matching is not stable before and after the treatment.

4.3. Equivalence of DID and SCD

One might wonder when the DID and SCD designs are equivalent in terms of GMC. The analysis in (4.2) gives an answer. Let us introduce pre-treatment PTA with within-group differencing λ as follows:

Pre-treatment PTA(λ): $e_t^{\text{DID}}(\lambda) = 0$, for all $t \in \mathcal{T}_0$.

Then, following the same arguments in the proof of Proposition 4.1, we can show that the pre-treatment PTA(λ) is equivalent to the following:

Pre-treatment PTA-I: $\mathbb{E}[\Delta Y_{i,t}(0) \mid G_i = 0] = \mathbb{E}[\Delta Y_{i,t}(0) \mid G_i \in \mathcal{G}_{\text{don}}]$, for all $t \in \mathcal{T}_0 \setminus \{1\}$.

Thus, Pre-treatment PTA-I refers to the parallel trend assumption that holds for the pre-treatment periods. Now, suppose that both the PTA(λ) and the pre-treatment PTA(λ) hold. This implies that

$$\text{MER}_1(\mathbf{w}^{\text{DID}}) = \text{MER}_0(\mathbf{w}^{\text{DID}}) = 0.$$

On the other hand, by the definition of $\mathbf{w}^{\text{SCD}}$, we have

$$\text{MER}_0(\mathbf{w}^{\text{SCD}}) = 0.$$

Since there is a unique $\mathbf{w}$ such that $\text{MER}_0(\mathbf{w}) = 0$ by (4.7) in Assumption 4.2, we must have $\mathbf{w}^{\text{SCD}} = \mathbf{w}^{\text{DID}}$. We formalize this into the following proposition.

Proposition 4.2. *Suppose that Assumptions 4.1-4.2 hold, and the $PTA(\lambda)$ and the pre-treatment $PTA(\lambda)$ hold. Then,*

$$\mathbf{w}^{\text{SCD}} = \mathbf{w}^{\text{DID}}.$$

Hence, the DID and SCD designs are equivalent in terms of GMC.

The result shows that when we use the same differencing method for both DID and SCD designs, and the PTA holds at all periods, the two designs are equivalent in terms of GMC. For the identification of the ATT, we do not require the pre-treatment PTA. However, it is a common practice to perform a pre-trend PTA test to assess the plausibility of the post-treatment PTA. This procedure is valid only if PTA implies the pre-treatment PTA. Then, the proposition says that under this implication, if the DID identifies the ATT through the post-treatment PTA, this means that the weights used by the DID are exactly the same as the weights chosen by the SCD. Hence, both designs are equivalent in terms of GMC.

Now, when the post-treatment PTA fails, the equivalence between the DID and the SCD breaks down, and the SCD can be an alternative to the DID design. The GMC for the SCD emerges as an alternative identifying assumption replacing PTA.

5. Inference for Synthetic Control with Differencing

We saw that the SCD can serve as an alternative to DID when the parallel trend assumption fails. To the best of our knowledge, the estimation and inference methods for SCD in our data structure have not been developed. Thus, we present the methods here together with their asymptotic properties. The proofs of the results are found in the Supplemental Note.

5.1. The Sampling Process and Estimation

5.1.1. The Sampling Process. In this section, we explain the sampling process that links the population objects to the sample. As for the population objects, we first assume that the random vectors, (Y_i^*, G_i) , are i.i.d. across $i \in N$, under each $P \in \mathcal{P}$. Let $P_{j,t}$ be the conditional distribution of $Y_{i,t}$ given $G_i = j$, and let $p_j = P\{G_i = j\}$.

For each $t \in \mathcal{T}$, we first draw $G_{i,t} \in \mathcal{G}$, i.i.d. across $i \in N_t$, with probability $P\{G_{i,t} = j\}$ equal to p_j for each $j \in \mathcal{G}$. Then, we draw $Y_{i,t}$, $i \in N_t$, i.i.d. from the conditional distribution $P_{j,t}$. By the sampling process, for each $t \in \mathcal{T}$, and $i \in N_t$, we have

$$m_{j,t} = \mathbb{E}[Y_{i,t} \mid G_{i,t} = j].$$

We let $N = \bigcup_{t=1}^T N_t$ and $n = |N|$. We also define $N_{j,t} = \{i \in N : G_{i,t} = j\}$ and $n_{j,t} = |N_{j,t}|$.

This sampling process accommodates the empirical setting where the size of cross-sectional units varies over time. It also accommodates both balanced or unbalanced panel settings and

repeated cross-sections. In the balanced panel setting, we have $N_t = N$ for all $t \in \mathcal{T}$, and assume that the random vectors

$$[(Y_{i,1}, G_{i,1}), \dots, (Y_{i,T}, G_{i,T})]$$

are i.i.d. across $i \in N$, whereas in the repeated cross-sections setting, we assume that $(Y_{i,t}, G_{i,t})$ are i.i.d. across $i \in N_t$ and independent across $t \in \mathcal{T}$.

For estimation and inference, we fix within-group differencing $\lambda \in \Delta_{|\mathcal{T}_0|-1}$ and assume that we are under SMC at λ so that we have

$$(5.1) \quad \mu_{0,t}(\lambda) = \sum_{j=1}^K \mu_{j,t}(\lambda) w_j,$$

for all $t \in \mathcal{T}$, for some $w \in \Delta_{K-1}$. Thus, SCD recovers this weight using (3.5).¹⁵

5.1.2. Estimation. First, we consider a setting where θ^* is identified. Since θ^* is equal to $\theta(\lambda, w^*(\lambda))$, the identification of θ^* boils down to that of $w^*(\lambda)$. We simply write

$$\hat{\mu}_{j,t} = \hat{\mu}_{j,t}(\lambda),$$

where $\hat{\mu}_{j,t}(\lambda)$ is defined in (4.3). We propose the following estimator of the weight $w^*(\lambda)$:

$$\hat{w}(\lambda) = \arg \min_{w \in \Delta_{K-1}} \hat{Q}(\lambda, w),$$

where, with $\hat{\mu}_t = [\hat{\mu}_{1,t}, \dots, \hat{\mu}_{K,t}]'$,

$$\hat{Q}(\lambda, w) = \sum_{t=1}^{T^*-1} (\hat{\mu}_{0,t} - \hat{\mu}_t' w)^2.$$

Lastly, we consider the following estimator for the target parameter θ_t^* :

$$(5.2) \quad \hat{\theta}_t(\hat{w}(\lambda)) = \hat{\mu}_{0,t} - \hat{\mu}_t' \hat{w}(\lambda).$$

Then, we can show that our estimators for the weight $w^*(\lambda)$ defined in (3.5) and the target parameter θ_t^* are $\sqrt{n}$ -consistent.

Theorem 5.1. *Suppose that Assumptions 2.1, 4.1 and 4.2 hold. Then, for all $t \in \mathcal{T}_1$, as $n \rightarrow \infty$,*

$$\hat{w}(\lambda) = w^*(\lambda) + O_p(n^{-1/2}) \text{ and } \hat{\theta}_t(\hat{w}(\lambda)) = \theta_t(\lambda, w^*(\lambda)) + O_p(n^{-1/2}).$$

¹⁵Note that SCD extrapolates the weight obtained from the pre-treatment matching to the post-treatment periods. It appears strange that the weight that did not give a perfect pre-treatment matching now achieves a perfect post-treatment matching. Hence, we assume that the weight gives a perfect matching on both pre- and post-treatment periods, i.e., SMC at λ .

5.2. Inference

We consider statistical inference on vector θ^* without assuming its point-identification, adapting the proposal of [Canen and Song \(2025\)](#) to our setting.

First, we construct a confidence set for $w^*(\lambda)$. Define

$$\hat{H} = \frac{1}{T^* - 1} \sum_{t=1}^{T^*-1} \hat{H}_t \text{ and } \hat{\mathbf{h}} = \frac{1}{T^* - 1} \sum_{t=1}^{T^*-1} \hat{\mathbf{h}}_t,$$

where $\hat{H}_t = \hat{\mu}_t \hat{\mu}_t'$ and $\hat{\mathbf{h}}_t = \hat{\mu}_{0,t} \hat{\mu}_t'$. Let

$$\hat{\varphi}(w) = \hat{H}w - \hat{\mathbf{h}} \text{ and } \varphi_p(w) = Hw - \mathbf{h}.$$

Let $B = [\mathbf{1}/\sqrt{K}, B_2]$ be the $K \times K$ orthogonal matrix B such that $B'B = I_K$ and $B_2'B_2 = I_{K-1}$, where $\mathbf{1}$ denotes the K -dimensional vector of ones.¹⁶ Note that B_2 is a $K \times (K-1)$ matrix. First, note that for each $i \in N$ and $t \in \mathcal{T}_0$,

$$\sqrt{n}(\hat{\mu}_{j,t} - \mu_{j,t}) = \frac{1}{\sqrt{n}} \sum_{i \in N} \psi_{ij,t} + o_p(1), \text{ as } n \rightarrow \infty,$$

where $\psi_{ij,t} = \psi_{ij,t}^* - \sum_{s \in \mathcal{T}_0} \lambda_s \psi_{ij,s}^*$, with¹⁷

$$\psi_{ij,t}^* = \frac{n}{n_t} \frac{1\{G_{i,t} = j\}}{p_j} (Y_{i,t} - m_{j,t}).$$

For notational brevity, we define

$$\mathbf{z}_{ij} = \frac{1}{T^* - 1} \sum_{t=1}^{T^*-1} \mu_t \psi_{ij,t} \text{ and } \mathbf{z}_i(w) = \sum_{j=1}^K w_j \mathbf{z}_{ij} - \mathbf{z}_{i0}.$$

Using this, we find that

$$\sqrt{n}B_2'(\hat{\varphi}(w) - \varphi_p(w)) \rightarrow_d N(0, V_p(w)),$$

where $V_p(w) = B_2' \text{Var}_p(\mathbf{z}_i(w)) B_2$. Let us estimate $V_p(w)$ by $\hat{V}(w)$ as follows: with $\hat{\mathbf{z}}_{ij} = \frac{1}{T^*-1} \sum_{t=1}^{T^*-1} \hat{\mu}_t \hat{\psi}_{ij,t}$ and $\hat{\mathbf{z}}_i(w) = \sum_{j=1}^K w_j \hat{\mathbf{z}}_{ij} - \hat{\mathbf{z}}_{i0}$,

$$(5.3) \quad \hat{V}(w) = B_2' \left(\frac{1}{n} \sum_{i \in N} \hat{\mathbf{z}}_i(w) \hat{\mathbf{z}}_i'(w) \right) B_2,$$

¹⁶The matrix B_2 can be computed as follows. First, we obtain a spectral decomposition: $I_K - \mathbf{1}\mathbf{1}'/K = UDU'$, where $\mathbf{1}$ denotes the K -dimensional vector of ones. From this, we set B_2 to be the $K \times (K-1)$ matrix after removing the eigenvector from U that corresponds to the zero diagonal element of D .

¹⁷In the case of a balanced panel setting with $N = N_t$ and $G_{i,t} = G_i$ for all $t \in \mathcal{T}$ and $i \in N$, we have $\psi_{ij,t} = 1\{G_{i,t} = j\}(y_{i,t} - \mu_{j,t})/p_j$, with $y_{i,t} = Y_{i,t} - \sum_{s \in \mathcal{T}_0} \lambda_s Y_{i,s}$.

where $\hat{\psi}_{ij,t}$ is the same as $\psi_{ij,t}$ except that p_j and $m_{j,t}$ are replaced by $\hat{p}_{j,t} = n_{j,t}/n_t$ and $\hat{m}_{j,t} = (1/n_{j,t}) \sum_{i \in N_{j,t}} Y_{i,t}$.

When the sample is repeated cross-sections, the observations are independent across time. In this case, we can obtain sharper inference by modifying $\hat{V}(w)$ as follows:

$$(5.4) \quad \hat{V}_{RC}(w) = \frac{1}{T^* - 1} \sum_{t=1}^{T^*-1} \left\{ \frac{1}{n} \sum_{i \in N} \left(\sum_{j=1}^K w_j \hat{\psi}_{ij,t}^* - \hat{\psi}_{i0,t}^* \right)^2 \right. \\ \left. \times B_2' \left(\frac{\hat{\mu}_t \hat{\mu}_t'}{T^* - 1} - \lambda_t (\bar{\mu} \hat{\mu}_t' + \hat{\mu}_t \bar{\mu}') + (T^* - 1) \lambda_t^2 \bar{\mu} \bar{\mu}' \right) B_2 \right\},$$

where

$$\hat{\psi}_{ij,t}^* = \frac{n}{n_t} \frac{1\{G_{i,t} = j\}}{\hat{p}_{j,t}} (Y_{i,t} - \hat{m}_{j,t}) \text{ and } \bar{\mu} = \frac{1}{T^* - 1} \sum_{t=1}^{T^*-1} \hat{\mu}_t.$$

For each $w \in \Delta_{K-1}$, define¹⁸

$$(5.5) \quad \hat{r}(w) = \arg \min_r (\hat{H}w - \hat{h} - r)' B_2 \hat{V}^{-1}(w) B_2' (\hat{H}w - \hat{h} - r),$$

where the minimization over r is under the constraint that $w'r = 0$ and $r \geq 0$. We define

$$\hat{d}(w) = |\{j = 1, \dots, K : \hat{\gamma}_j(w) = 0 \text{ and } w_j = 0\}|,$$

where

$$\hat{\gamma}(w) = B_2 \hat{V}^{-1}(w) B_2' (\hat{H}w - \hat{h} - \hat{r}(w)).$$

We set $\hat{c}_{1-\kappa, \text{bf}}(w)$ to be the $(1 - \kappa)$ -th quantile of the χ^2 distribution with degrees of freedom equal to

$$(5.6) \quad \hat{k}(w) := \max\{K - 1 - \hat{d}(w), 1\}.$$

Then, the confidence set for $w^*(\lambda)$ is given by

$$\tilde{C}_{1-\kappa} = \{w \in \Delta_{K-1} : T_{\text{bf}}(w) \leq \hat{c}_{1-\kappa, \text{bf}}(w)\},$$

where

$$T_{\text{bf}}(w) = n(\hat{H}w - \hat{h} - \hat{r}(w))' B_2 \hat{V}^{-1}(w) B_2' (\hat{H}w - \hat{h} - \hat{r}(w)).$$

¹⁸Due to the constraint, we have $\hat{r}(w) = 0$ if all entries of w are positive. Hence, we perform the numerical optimization only if some of the entries of w are zeros.

Algorithm 1 Computing Confidence Intervals for θ_t^* : Bonferroni Method

Input: Consistent estimator $\hat{w}(\lambda)$ of $w^*(\lambda)$, $\hat{\sigma}_t(\hat{w}(\lambda))$ and $\hat{V}(\hat{w}(\lambda))$.

- 1: Draw $w_1, \dots, w_R \in \Delta_{K-1}$ i.i.d. from a distribution that has a full support on Δ_{K-1} .
- 2: Compute $T_{\text{bf}}(w_r)$ and $\hat{c}_{1-\kappa, \text{bf}}(w_r)$ with $\hat{V}(w)$ replaced by $\hat{V}(\hat{w}(\lambda))$ for each $r = 1, \dots, R$.
- 3: Let

$$c_{U,R} = \max_{1 \leq r \leq R: T_{\text{bf}}(w_r) \leq \hat{c}_{1-\kappa, \text{bf}}(w_r)} \left\{ \hat{\theta}_t(w_r) + \frac{z_{1-\beta(\alpha, \kappa)} \hat{\sigma}_t(\hat{w}(\lambda))}{\sqrt{n}} \right\} \text{ and}$$

$$c_{L,R} = \min_{1 \leq r \leq R: T_{\text{bf}}(w_r) \leq \hat{c}_{1-\kappa, \text{bf}}(w_r)} \left\{ \hat{\theta}_t(w_r) - \frac{z_{1-\beta(\alpha, \kappa)} \hat{\sigma}_t(\hat{w}(\lambda))}{\sqrt{n}} \right\},$$

where $\beta(\alpha, \kappa) = (\alpha - \kappa)/2$.

Output: Confidence interval for θ_t^* :

$$C_{1-\alpha, R} = [c_{L,R}, c_{U,R}].$$

Notes: The algorithm assumes that the weight $w^*(\lambda)$ is point-identified. When it is partially identified, we can modify the above algorithm by replacing $\hat{V}(\hat{w}(\lambda))$ and $\hat{\sigma}_t(\hat{w}(\lambda))$ in Steps 2 and 3 by $\hat{V}(w_r)$ and $\hat{\sigma}_t(w_r)$.

Let $\tilde{C}_{1-\kappa}$ be the confidence set for $w^*(\lambda)$. We construct a confidence interval for θ_t^* . Note that

$$\sqrt{n}(\hat{\theta}_t(w^*(\lambda)) - \theta_t^*) = \frac{1}{\sqrt{n}} \sum_{i \in N} \psi_{it, \theta}(w^*(\lambda)) + o_p(1),$$

where

$$\psi_{it, \theta}(w^*(\lambda)) = \psi_{i0, t} - \sum_{j=1}^K \psi_{ij, t} w_j^*(\lambda).$$

Define

$$\hat{\sigma}_t^2(w) = \frac{1}{n} \sum_{i \in N} \hat{\psi}_{it, \theta}^2(w),$$

where $\hat{\psi}_{it, \theta}(w) = \hat{\psi}_{i0, t} - \sum_{j=1}^K \hat{\psi}_{ij, t} w_j$. Then, the confidence interval for θ_t^* is given as follows: with $\kappa \in (0, \alpha)$, (say, $\kappa = 0.005$)

$$C_{1-\alpha}^{\text{bf}} = \left\{ \tau \in \mathbf{R} : \inf_{w \in \tilde{C}_{1-\kappa}} \left| \frac{\sqrt{n}(\hat{\theta}_t(w) - \tau)}{\hat{\sigma}_t(w)} \right| \leq z_{1-\beta(\alpha, \kappa)} \right\},$$

where $\beta(\alpha, \kappa) = (\alpha - \kappa)/2$.

The computation of a confidence interval for θ_t^* involves inverting a test for the weight vector. For the case of point-identified $w^*(\lambda)$, we present an algorithm that computes the convex hull of $C_{1-\alpha}$ directly without constructing $\tilde{C}_{1-\kappa}$ first. See Algorithm 1. Computational

experiments in Section 5.4 below demonstrate that the algorithm computes the confidence set efficiently in practical data dimensions ($n_t = 14,000 \sim 130,000$, $T = 84$ and $K = 46$).

5.3. Asymptotic Validity

Let us introduce assumptions we use for the uniform asymptotic validity of the confidence set for θ_t^* , without requiring the point-identification of $w^*(\lambda)$:

Assumption 5.1. There exist constants $C, c > 0$ such that for all $n \geq 1$,

$$\sup_{P \in \mathcal{P}} \sup_{w \in \Delta_{K-1}} \|V_p(w)\| < C \text{ and } \inf_{P \in \mathcal{P}} \inf_{w \in \Delta_{K-1}} \lambda_{\min}(V_p(w)) > c.$$

Assumption 5.1 requires that the asymptotic variance $V_p(w)$ is well behaved uniformly over $P \in \mathcal{P}$ and $w \in \Delta_{K-1}$: it should be both bounded and non-singular.

Under these conditions, we obtain the following validity result.

Theorem 5.2. Suppose that Assumptions 2.1, 4.1, 5.1, (4.6) in Assumption 4.2, and SMC holds at λ . Then, for all $t \in \mathcal{T}_1$, as $n \rightarrow \infty$, we have

$$\liminf_{n \rightarrow \infty} \inf_{P \in \mathcal{P}} P \{ \theta_t^* \in C_{1-\alpha} \} \geq 1 - \alpha.$$

The proofs are found in the Supplemental Note.

5.4. Monte Carlo Simulations

In this subsection we study the finite sample properties of our estimator of the target parameter. Our focus is on comparing SCD and DID and examining their complementarities. We consider a short panel setting of length T where individual data is available and analyze the simple case of one treated group and K untreated groups. Group 0 becomes treated at time T^* , and the remaining groups $\{1, \dots, K\}$ form the donor pool $\mathcal{G}_{\text{don}}$. We set $T^* = T \in \{50, 100\}$ and $K \in \{10, 30\}$.¹⁹ We compare the performance of our SCD estimator with the standard DID estimator in terms of the mean absolute deviation (MAD), the empirical coverage probability (ECP), and the average length of the 95% confidence intervals. We consider a sample size of $n \in \{1500, 3000\}$ and set the number of Monte Carlo simulations to 1,000.

When comparing SCD and DID approaches, we implement the DID estimator of Callaway and Sant'Anna (2021) without covariates. In the study, we consider three different scenarios: one (Scenario A) in which PTA holds and there are parallel pre-trends, a second one (Scenario B) in which PTA is violated but SMC holds throughout all time periods, and a last one (Scenario C) where PTA holds, SMC holds, but the weights for donors cannot be recovered from pre-treatment data. Thus, in Scenario A, both DID and SCD produce consistent and

¹⁹Here the length of the post-treatment window is equal to one.

asymptotically normal estimators of the treatment effect. However, in Scenario B, while SCD works, DID is not consistent. The opposite occurs in Scenario C.

We now describe the data generating process used in the simulations. First, for the baseline setting, we define the probability of an individual belonging to group $j \in \mathcal{G} = \{0, 1, \dots, K\}$ as simply $1/(K + 1)$, so that G_i is drawn i.i.d. from the uniform distribution over $\mathcal{G}$ with probability $1/(K + 1)$. As for the generation of potential outcomes, we adopt a factor model:

$$(5.7) \quad Y_{i,t}(0) = \Lambda_i' F_t + \varepsilon_{i,t},$$

where, conditional on G_i , $\Lambda_i \sim N(m_{G_i}, I_3)$.²⁰ Here m_j denotes the population mean of factor loadings in group j , whose components are drawn independently from a normal distribution with mean zero and variance 2.5^2 . In addition, time factors and idiosyncratic shocks are generated as $F_t \sim N(0.02\sqrt{t} \cdot \mathbf{1}_3, 0.5^2 \cdot I_3)$ and $\varepsilon_{i,t} \sim N(0, 1)$. Lastly, we set treated potential outcomes for individuals in group 0 as

$$(5.8) \quad Y_{i,t}(T^*) = Y_{i,t}(0) + \tau_{i,t}(T^*),$$

where $\tau_{i,t}(T^*) = \eta_{i,t}^2$, and each $\eta_{i,t}$ is drawn from a normal distribution with mean zero and variance 0.1. This setup implies that $\theta_{T^*}^* = 0.1$. In other words, the average treatment effect for individuals in group 0 is equal to the variance of the random variable $\eta_{i,t}$.

We select different combinations of the differencing parameter ($\lambda \in \Delta_{|\mathcal{T}_0|-1}$) and the population mean of individual factor loadings in the treated group ($m_0 \in \mathbf{R}^3$) across scenarios. In Scenario A, we choose

$$\lambda = \lambda^{\text{DID}} \text{ and } m_0 = \sum_{j=1}^K m_j w_j^{\text{DID}},$$

where $w^{\text{DID}} = [1/K, \dots, 1/K]' \in \mathbf{R}^K$ by the simulation design. By choosing these values, we guarantee that PTA is satisfied, parallel pre-trends are present, and GMC holds in both the pre- and post-treatment periods at $(\lambda^{\text{DID}}, w^{\text{DID}})$. In Scenario B, we let

$$\lambda = \lambda^{\text{unif}} \text{ and } m_0 = \sum_{j=1}^K m_j w_j^{\text{SCD}},$$

where $w^{\text{SCD}} = [0, \dots, 0, 0.1, 0.9]' \in \mathbf{R}^K$ is a vector whose first $K - 2$ entries are zero. In this case, PTA is violated since $w^{\text{SCD}} \neq w^{\text{DID}}$, but GMC still holds at $(\lambda^{\text{unif}}, w^{\text{SCD}})$. Lastly, we consider a Scenario C where PTA and GMC hold at $(\lambda^{\text{DID}}, w^{\text{DID}})$ for the post-treatment period, but the SCD approach is unable to recover w^{DID} using pre-treatment data. More precisely, we

²⁰We use $\mathbf{1}_m$ and I_m to denote the m -dimensional vector of ones and the $m \times m$ identity matrix, respectively.

TABLE 2. Results for Monte Carlo Simulations.

Parameters			MAD		Coverage Probability		CI Length	
K	T	n	SCD	CSDID	SCD	CSDID	SCD	CSDID
<i>Scenario A: PTA and SMC hold</i>								
10	50	1500	0.079	0.145	0.999	0.998	1.475	1.166
10	50	3000	0.059	0.105	1.000	0.999	0.967	0.828
10	100	1500	0.079	0.148	1.000	0.998	1.397	1.228
10	100	3000	0.054	0.100	1.000	1.000	0.907	0.875
30	50	1500	0.128	0.229	1.000	0.997	2.456	1.770
30	50	3000	0.094	0.163	1.000	1.000	1.824	1.279
30	100	1500	0.130	0.228	1.000	0.999	2.280	1.864
30	100	3000	0.090	0.157	1.000	0.999	1.715	1.349
<i>Scenario B: PTA fails but SMC holds</i>								
10	50	1500	0.092	1.914	0.999	-	1.002	-
10	50	3000	0.069	1.909	0.996	-	0.683	-
10	100	1500	0.092	1.861	0.997	-	0.965	-
10	100	3000	0.062	1.857	0.998	-	0.652	-
30	50	1500	0.140	2.083	0.999	-	1.590	-
30	50	3000	0.100	2.073	0.997	-	1.072	-
30	100	1500	0.137	2.006	0.992	-	1.378	-
30	100	3000	0.098	1.982	0.993	-	0.927	-
<i>Scenario C: PTA holds but SMC fails</i>								
10	50	1500	1.163	0.145	-	0.999	-	1.167
10	50	3000	1.159	0.105	-	0.999	-	0.828
10	100	1500	1.105	0.148	-	0.998	-	1.229
10	100	3000	1.130	0.100	-	1.000	-	0.875
30	50	1500	1.317	0.229	-	0.998	-	1.771
30	50	3000	1.340	0.163	-	1.000	-	1.278
30	100	1500	1.209	0.228	-	0.999	-	1.866
30	100	3000	1.198	0.157	-	0.999	-	1.350

Notes: This table presents the simulation results for three scenarios: Scenario A assumes both Parallel Trends Assumption (PTA) and Stable Market Condition (SMC) hold, Scenario B assumes PTA fails but SMC holds, and Scenario C assumes PTA holds but SMC fails. The table reports Mean Absolute Deviation (MAD), Coverage Probability, and Confidence Interval (CI) Length for our proposed method of Synthetic Control with Differencing (SCD) and the Difference-in-Differences estimator proposed by [Callaway and Sant’Anna \(2021\)](#) (CSDID). Inference for SCD is conducted using Algorithm 1. Inference results are only reported for cases where the target parameter is identified. The number of MC simulations is 1,000.

allow for time-varying factor loadings for individuals in the treated group as follows

$$\Lambda_{i,t} = \tilde{\Lambda}_i 1\{t \leq T^* - 2\} + \Lambda_i 1\{t \geq T^* - 1\},$$

where Λ_i is defined as in Scenario A, but, conditional on G_i , $\tilde{\Lambda}_i \sim N(\tilde{m}_{G_i}, I_3)$, and

$$\tilde{m}_0 = \sum_{j=1}^K m_j w_j^{\text{OUT}},$$

with $w^{\text{OUT}} = [0, \dots, 0, -0.3, 0.4, 0.9]' \in \mathbf{R}^K$. In this case, we allow individual factor loadings to be drawn from different distributions between pre- and post-treatment periods so that weights for control groups cannot be estimated by SCD using pre-treatment data.

The simulation results from this baseline setting are reported in Table 2. Inference results for SCD are based on the Bonferroni approach detailed in Algorithm 1. In all scenarios, when the number of donor groups K increases (keeping fixed T and n), the accuracy of the estimators in terms of MAD deteriorates for both SCD and DID. In particular, in Scenario A, we have $w^{\text{SCD}} = w^{\text{DID}}$, so both designs generate a consistent estimator of θ_{T*}^* and the confidence intervals are asymptotically valid as $n \rightarrow \infty$. As the number of groups increases, the estimation error of the weights accumulates. This explains the performance deterioration as K increases from 10 to 30. When T increases, the performance of SCD in terms of MAD remains similar, while confidence intervals become narrower.²¹ This pattern arises primarily because our setting is a panel design. In a repeated cross-section setting, the observations are independent across time and the accuracy would have increased as T increased. The empirical coverage probability of DID and SCD shows conservativeness in this scenario, whereas CS-DID produces confidence bands that are, on average, 24% narrower than those obtained with SCD.

In Scenario B, our simulation design is chosen so that PTA fails but SMC holds at λ^{unif} . As expected, in this case we find that SCD outperforms DID in terms of MAD, and it also exhibits some conservativeness as in Scenario A. In Scenario C, we consider an opposite setting, that is, PTA holds but the stability of the weights fails. In this case, only DID provides a consistent estimate of θ_{T*}^* . As a result, DID displays a smaller MAD than SCD. The performance of SCD in Scenario C still appears better than that of DID in Scenario B. However, we believe this difference is largely driven by the simulation design.

In Tables D.1–D.3 of Appendix D, we conduct three robustness checks for our simulation results. First, we depart from the equal-size group setting and instead consider a donor pool in which the last group is relatively larger than the others. Specifically, we set the group membership probabilities to $p = [0.8/K, \dots, 0.8/K, 0.2]' \in \mathbf{R}^{K+1}$ for $K = 10$, and $p = [0.925/K, \dots, 0.925/K, 0.075]' \in \mathbf{R}^{K+1}$ for $K = 30$. Secondly, we allow each component

²¹Since the simulation includes only one post-treatment period, an increase in T reflects an increase in the number of pre-treatment periods.

of time factors F_t to have a different population mean by drawing it from a multivariate normal distribution with mean $\xi\sqrt{t}$ and covariance matrix 0.5^2I_3 , where $\xi = [0.01, 0.02, 0.04]'$. Lastly, we use a multivariate t distribution with 5 degrees of freedom, mean m_{G_i} , and scale matrix I_3 , as the conditional distribution of individual factor loadings Λ_i . Across all these additional exercises, the main findings reported in Table 2 continue to hold.

Overall, our simulation results highlight the complementarity between SCD and DID. Each method performs well when its identifying assumptions hold and deteriorates when they fail. In particular, SCD provides a reliable alternative when PTA is violated and performs as well as DID when both sets of identifying assumptions are satisfied.

5.5. Computation Time

Our SCD method relies on a Bonferroni approach to construct a confidence set for the weight w^* , so a natural concern is whether the computational cost of this procedure is prohibitive in practice. In this section, we demonstrate that Algorithm 1, which uses simulated draws from the simplex, is computationally feasible for data dimensions commonly found in practice.

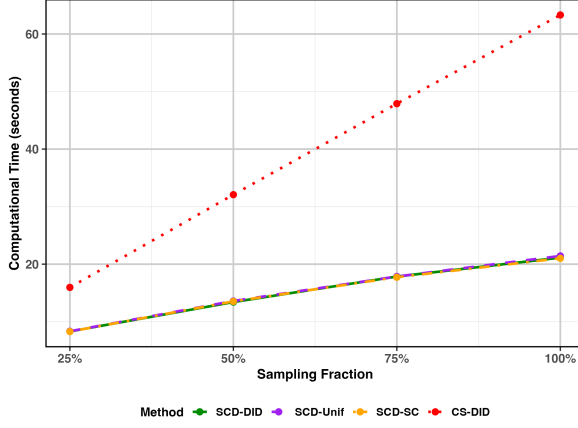
We first report the average computation times of constructing confidence intervals using Algorithm 1 on the monthly CPS data used in our empirical application in Section 6. We restrict the time periods to January 2003-December 2009 and consider Arizona as treated after July 2007, so that $T = 84$ and $T^* = 55$. In this application the number of donors is $K = 46$ states. We also consider two outcome variables: the indicator of the individual being non-U.S. citizen Hispanic (binary outcome) and the individual's log of weekly earnings (continuous outcome). We restrict our attention to individuals with strictly positive weekly earnings when analyzing the continuous outcome variable. On average, the total number of cross-sectional units per month is 128,932 for the case with a binary outcome, and 13,977 when considering a continuous outcome. To assess scalability, four sample fractions are considered: 25%, 50%, 75%, and 100%. For a given fraction, we randomly draw that share of the full data five times and implement our SCD method on each draw. The reported computation time is the average time across these five repetitions.²²

The first set of results is shown in panels (a) and (b) of Figure 1. In panel (a), we document that SCD takes on average 20 seconds to construct the confidence intervals of treatment effects with a discrete outcome and the full sample. For comparison, we also report computation times for the `did` R package version 2.3.0 of Callaway and Sant'Anna (2021) (CS-DID).²³ In this case, their package performs similar to SCD for small samples, but their computation time

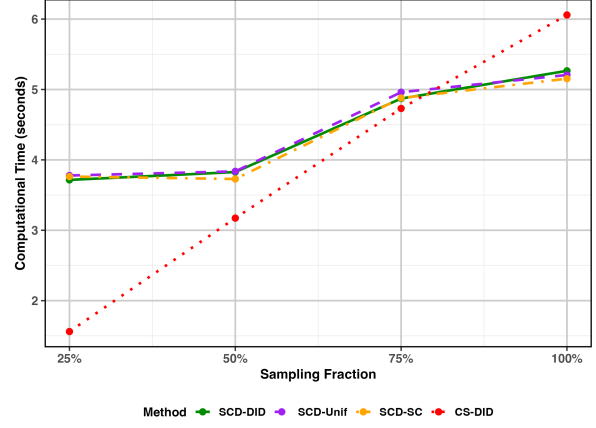
²²All computations are performed on an Apple M4 Max with 64GB of RAM.

²³The package is available on the website: <https://cran.r-project.org/web/packages/did/index.html>

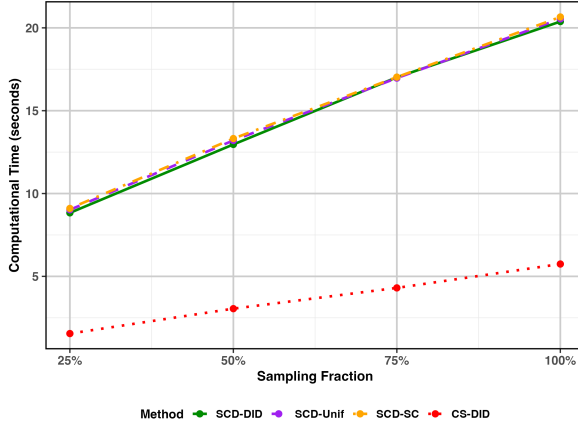
FIGURE 1. Computation Time.



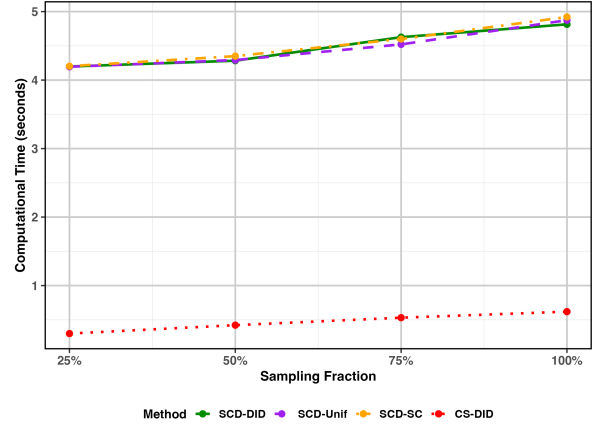
(a) Discrete outcome, Repeated Cross-Section



(b) Continuous outcome, Repeated Cross-Section



(c) Discrete outcome, Panel

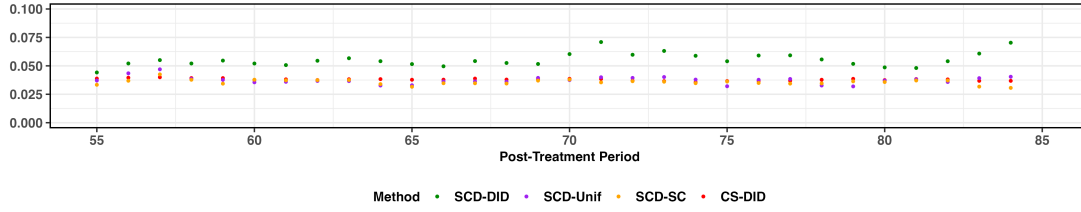


(d) Continuous outcome, Panel

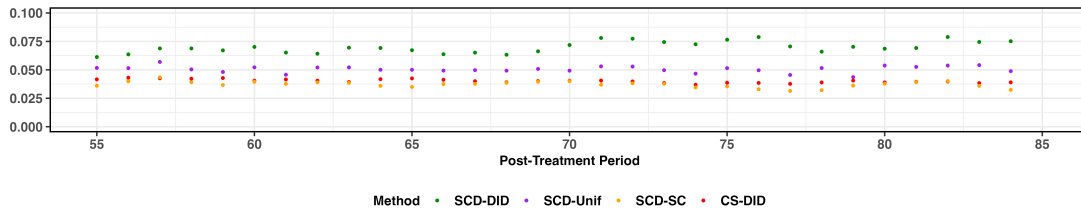
Notes: The figure reports the computation time required to implement the SCD estimator across different combinations of sample size, outcome variables, and data structures. As a benchmark, we also report the computation time of the `did` R package version 2.3.0 of [Callaway and Sant'Anna \(2021\)](#) (CS-DID). We consider two outcomes: (i) a discrete outcome equal to an indicator for being a non-U.S.-citizen Hispanic, and (ii) a continuous outcome given by the log of weekly earnings. Results are reported for both repeated cross-section data from the CPS between January 2003 and December 2009 and for a fictitious panel constructed from the same CPS data. For each sample fraction, we randomly draw that share five times and implement SCD on each draw. The reported computation time is the average across these five repetitions. SCD-DID refers to SCD with $\lambda = \lambda^{\text{DID}}$, SCD-Unif uses $\lambda = \lambda^{\text{unif}}$, and SCD-SC corresponds to SCD without differencing.

increases faster than our SCD method as the sample becomes larger. This difference likely reflects the greater generality of the [Callaway and Sant'Anna \(2021\)](#) package, which accommodates multiple covariates and various estimation options. A similar pattern is observed in panel (b), where we consider a continuous outcome variable. CS-DID starts outperforming

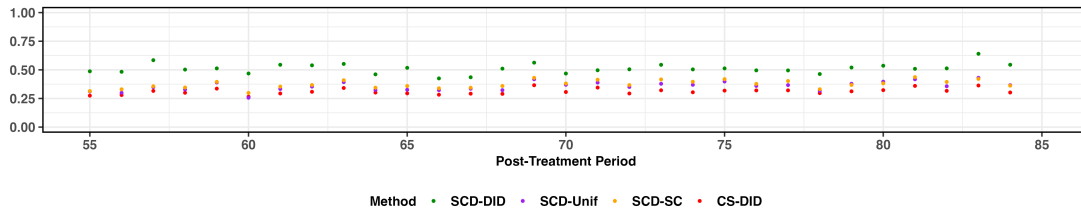
FIGURE 2. Length of Post-Treatment Confidence Intervals.



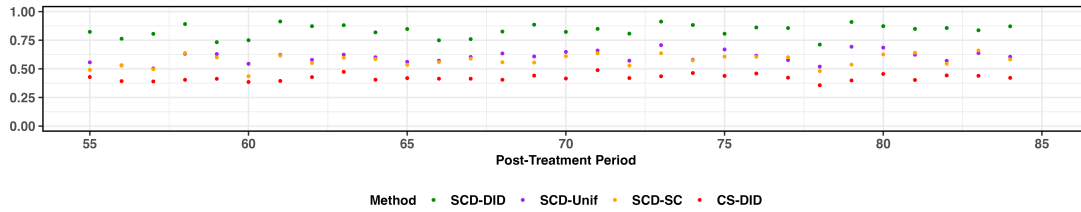
(a) Discrete outcome, Repeated Cross-Section



(b) Discrete outcome, Panel



(c) Continuous outcome, Repeated Cross-Section



(d) Continuous outcome, Panel

Notes: The figures report the length of the confidence intervals only for the confidence intervals in the post-treatment periods. Repeated cross-sectional data comes from the monthly CPS between January 2003 and December 2009. The panel data is created by independently drawing the same number of individuals per state-month from the CPS. Treatment occurs in July 2007 ($T^* = 55$) in Arizona, after LAWA was passed. The number of donors is $K = 46$ states. Panels (a) and (c) use as main outcome the indicator of the individual being non-U.S. citizen Hispanic, and panels (b) and (d) use the individual's log of weekly earnings. SCD inference is based on Algorithm 1. CS-DID refers to the did R package of [Callaway and Sant'Anna \(2021\)](#). SCD-DID refers to the method of SCD using $\lambda = \lambda^{\text{DID}}$, and SCD-Unif using $\lambda = \lambda^{\text{unif}}$.

our procedure by a couple of seconds in small samples but its computation time grows faster, taking almost one second more than SCD once we consider the full sample.

We next compare SCD and CS-DID under a panel data structure. To this end, we construct a balanced panel from the CPS used before by randomly drawing, in each month, the same number of individuals within a given state. For each state, the cross-sectional dimension of the panel is defined as the minimum number of observations available in that state across all months in the CPS. Sampled individuals are then assigned a panel identifier by interacting the state code with a within-state row number, producing individual ids that are consistent across time periods. The resulting panel contains $n = 112,744$ observations every period for the discrete case, and $n = 8,969$ for the continuous case. The results, shown in panels (c) and (d) of Figure 1, indicate that although SCD remains computationally feasible, CS-DID outperforms our approach in all cases. This reversal relative to the repeated cross-section results arises because our inference procedure exploits the time-independence across samples that is present in repeated cross-section data, an advantage that disappears in panel settings. We also find that computation times are not affected by the choice of differencing parameter and that the computational cost of SCD does not increase exponentially with the number of cross-sectional units.

Finally, we compare SCD and CS-DID in terms of the average length of their confidence intervals for the post-treatment periods across the different data structures and outcome variables considered above. The results are reported in Figure 2. Consistent with the findings in Table 2, SCD produces, on average, longer confidence intervals than CS-DID. Interestingly, the gap between the two methods becomes less pronounced when the uniform differencing parameter is used or when no differencing is applied. These patterns hold in both repeated cross-section and panel settings.

6. Empirical Application

To illustrate our method, we revisit the empirical setting analyzed by [Bohn *et al.* \(2014\)](#) and study the effects of the 2007 Legal Arizona Workers Act (LAWA) on Arizona’s internal composition. LAWA was passed in July 2007 and prohibited businesses from knowingly hiring unauthorized workers after December 31, 2007. In addition, this new law required all Arizona employers to verify the identity and work eligibility of new hires using an online system (called E-Verify) that cross-checks employee information against federal earnings and immigration databases. Employers who did not comply with the new rules faced sanctions like suspensions or permanent revocation of their business licenses. As one of the strictest state-level immigration laws at the time, it raised the costs of unauthorized employment for both employers and undocumented immigrants.

TABLE 3. Summary Statistics.

	Arizona			Donor pool		
	2006	2009	Diff.	2006	2009	Diff.
Age	35.168	35.674	0.506	36.369	36.814	0.445
Female	0.503	0.503	0.000	0.511	0.510	-0.001
<i>Educational attainment</i>						
Less than high school	0.413	0.375	-0.038	0.362	0.349	-0.013
High school graduate	0.227	0.211	-0.016	0.240	0.237	-0.003
Some college	0.201	0.223	0.022	0.205	0.209	0.004
College or more	0.159	0.190	0.031	0.193	0.205	0.012
Employment	0.462	0.462	0.000	0.485	0.470	-0.015
Non-citizen Hispanic	0.095	0.063	-0.032	0.043	0.042	-0.001
Observations	1,944	1,627		127,040	124,880	

Notes: Cells for age display the mean and cells for other variables show proportions. Arizona's donor pool consists of 46 states without a similar regulation during the period analyzed. Columns 2 and 5 report January 2006 CPS statistics; Columns 3 and 6 report January 2009 CPS statistics; Columns 4 and 7 report the change between 2009 and 2006. Survey weights are used.

In this context, the group membership variable (G_i) is defined as the state in which individual i lives, the treated group is Arizona and the post-treatment period begins once LAWA is passed in July 2007. We use CPS microdata from January 1998 to December 2009 and follow the authors in considering 46 states (K) in Arizona's donor pool that did not implement any similar regulation during the period of analysis.²⁴ Nevertheless, unlike [Bohn et al. \(2014\)](#), we do not aggregate the monthly CPS data to the annual level, which allows us to point identify the weights for Arizona's donor pool using SCD. Our dataset contains 114 months and 30 months in the pre and post-treatment periods, respectively, with a total of 144 time periods, i.e., $T^* = 115$ and $T = 144$. We focus on the population that is most likely to be affected by the policy change, so our primary outcome of interest, $Y_{i,t}$, is defined as an indicator variable equal to one if individual i is Hispanic but not a U.S. citizen at time t and zero otherwise. We apply SCD with the DID differencing parameter (λ^{DID}) to estimate the average treatment effect on the treated, which, in this case, captures the causal impact of LAWA on the share of non-citizen Hispanic individuals in Arizona.

²⁴The excluded states are Mississippi, Rhode Island, South Carolina, and Utah. CPS data is provided in the replication package of [Bohn et al. \(2014\)](#).

TABLE 4. Arizona’s Donors with Positive SCD Weights.

State	Weights
Connecticut	0.204
Florida	0.029
Georgia	0.066
Idaho	0.047
Illinois	0.037
Kansas	0.086
New Jersey	0.378
New York	0.017
Washington	0.138

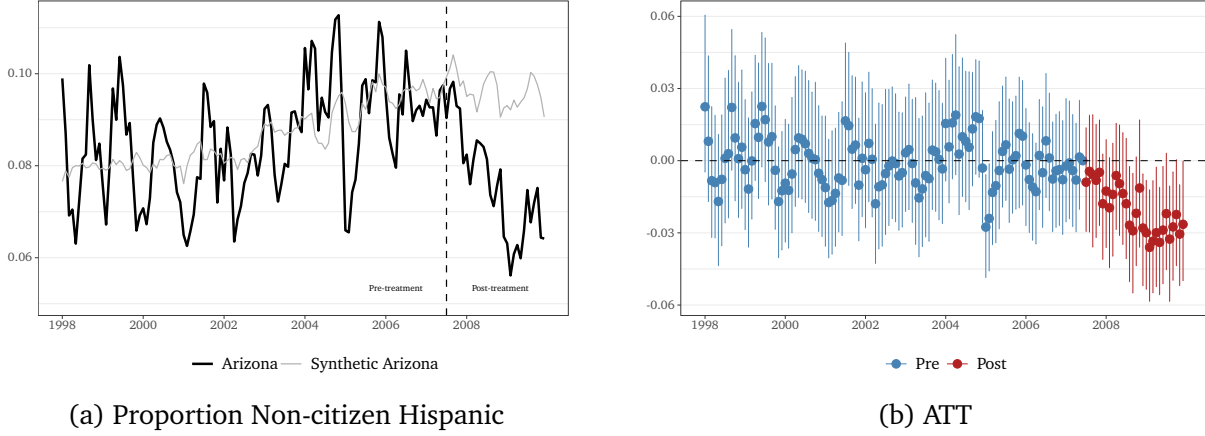
Notes: Weights are obtained by applying SCD to Arizona and its donor pool, using as main outcome the proportion of non-citizen Hispanic and setting $\lambda = \lambda^{\text{DID}}$. Arizona’s donor pool consists of 46 states without any similar regulation during the period analyzed. Data come from the monthly CPS between January 1998 and December 2009. The sum of weights may differ from one due to rounding. Survey weights are used.

Table 3 presents descriptive statistics for Arizona and its donor pool one and a half years before and after LAWA’s enactment. We observe small changes over time in both Arizona and its donor pool in terms of age, gender composition, and the employment-to-population ratio. In contrast, changes in Arizona’s educational attainment distribution are more pronounced than in the donor pool between 2006 and 2009. In particular, the share of low-educated individuals (those with a high school diploma or less) declined by 5.4 percentage points in Arizona, compared to a 1.6 percentage-point reduction among donor states. Likewise, the variable of interest, the proportion of non-citizen Hispanic, fell by 3.2 percentage points (a 34% drop) in Arizona, whereas the donor pool experienced only a marginal 0.1 percentage-point (a 2% fall) decrease over the same period. These patterns are in line with the hypothesis that LAWA reshaped Arizona’s demographic composition by tightening immigrants’ access to employment opportunities. In the next subsection we provide an estimate of LAWA’s causal effect on the internal composition of Arizona using SCD.

6.1. Results

Table 4 reports the subset of states in Arizona’s donor pool with positive weights from the SCD estimation using as main outcome the proportion of non-citizen Hispanic. The largest weight is assigned to New Jersey, followed by Connecticut and Washington. Smaller positive weights are assigned to Kansas, Georgia, Idaho, Illinois, Florida, and New York. Interestingly, the fact that all of Arizona’s neighboring states receive a zero weight by SCD in the construction of synthetic Arizona suggests the presence of potential spillover effects following LAWA’s enactment. In addition, none of the three states with positive SC weights found by [Bohn et al.](#)

FIGURE 3. Estimated Effects on Arizona's Share of Non-citizen Hispanic.



Notes: Panel (a) in this figure shows the evolution of Arizona's share of non-citizen Hispanic individuals compared to its synthetic version and panel (b) displays the corresponding ATT after LAWA's enactment in July 2007. We apply SCD with λ^{DID} , where Arizona's donor pool consists of 46 states without any similar regulation during the period analyzed. The states in the donor pool with positive SCD weights are Connecticut (0.204), Florida (0.029), Georgia (0.066), Idaho (0.047), Illinois (0.037), Kansas (0.086), New Jersey (0.378), New York (0.017), and Washington (0.138). Data come from the monthly CPS between January 1998 and December 2009. The blue and red lines correspond to 95% CIs constructed using Algorithm 1 for repeated cross-sectional data. Survey weights are used.

(2014) (California, Maryland, and North Carolina) are shown in Table 4. Two main factors contribute to this discrepancy. First, our identification strategies are different. We invoke GMC with λ^{DID} , so we need trends in averaged untreated potential outcomes to match between Arizona and its donor pool, which is less restrictive than the traditional SC approach that matches averaged untreated potential outcomes between Arizona and its donors directly. Secondly, the authors combine the CPS data at the annual level before applying SC, whereas we exploit the frequency of the CPS to obtain point-identification of SCD weights.²⁵

Figure 3 shows our main results for the share of non-citizen Hispanic after applying SCD. Panel (a) mirrors the standard plot used in the SC literature, displaying two time-series lines: one for Arizona (black) and another for its synthetic control (grey).²⁶ Overall, both lines follow a similar trend during the pre-treatment period, with Arizona's series exhibiting higher volatility than its synthetic counterpart. On the other hand, following the passage of LAWA,

²⁵In their main SC analysis, the authors also incorporate covariates such as state unemployment rates and industrial composition of the workforce, yet their results remain virtually unchanged when these covariates are excluded.

²⁶In SCD, the synthetic counterfactual outcomes for the treated unit are computed as follows:

$$\mathbb{E}[Y_{i,t}(0) | G_i = 0] = \hat{m}_{0,t} - \hat{\theta}_t(\hat{w}), \text{ for all } t \in \mathcal{T}.$$

TABLE 5. SCD Weights for Arizona’s Donors Across Robustness Exercises.

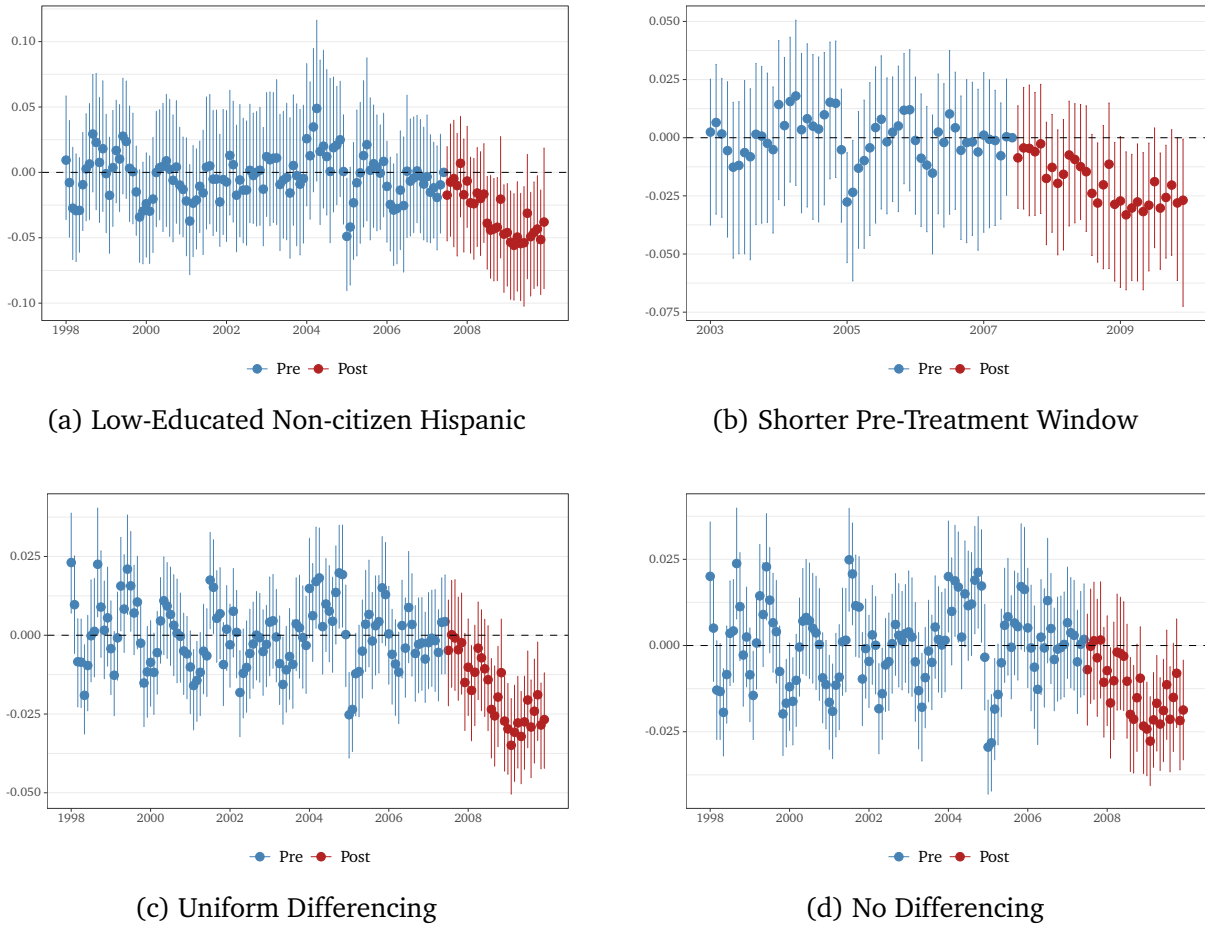
State	Non-citizen Hispanic Low-Educated	Shorter Pre-Treatment	Uniform Differencing	No Differencing
Alabama	0.195	0	0	0
Arkansas	0	0.010	0	0
California	0	0	0	0.621
Colorado	0.070	0.247	0	0
Connecticut	0	0	0.134	0
District of Columbia	0	0	0.060	0
Florida	0	0	0	0.213
Georgia	0.168	0	0.155	0
Idaho	0	0	0.103	0
Illinois	0.013	0	0.010	0
Kansas	0	0.096	0	0
Louisiana	0	0.201	0.037	0
Missouri	0	0	0.023	0
New Jersey	0.245	0.382	0.265	0.166
New York	0	0	0.050	0
North Carolina	0.120	0.055	0	0
South Dakota	0	0	0.025	0
Washington	0.188	0	0.139	0
Wisconsin	0	0.009	0	0

Notes: Cells contain the SCD weights for each robustness exercise described at the top of each column. The second column uses as main outcome the number of non-citizen Hispanics with a high-school diploma or less as a proportion of the prime-working age (15-45) state population. The third column applies SCD on the proportion of non-citizen Hispanics with a shorter pre-treatment window that starts in January 2003. The fourth column shows the states’ weights after applying SCD with a uniform differencing parameter (λ^{unif}), and the last column does not apply any differencing to the data. Arizona’s donor pool consists of 46 states without any similar regulation during the period analyzed. Data come from the monthly CPS between January 1998 and December 2009. The sum of weights may differ from one due to rounding. Survey weights are used.

we observe a big drop in Arizona’s proportion of non-citizen Hispanic relative to its synthetic control, going from 9.1% to 6.4% between June 2006 and December 2009.

Panel (b) in Figure 3 shows LAWA’s causal effects (estimated by $\hat{\theta}_t(\hat{w})$) on Arizona’s internal composition of non-citizen Hispanic along with a 95% confidence band obtained via Bonferroni’s method described in Algorithm 1 for repeated cross-sectional data. Similar to the usual pre-trend test used in empirical DID studies (Callaway and Sant’Anna, 2021; Sun and Abraham, 2021; Roth *et al.*, 2023; Borusyak *et al.*, 2024), the non-statistically significant estimates of the treatment effect during the pre-treatment period provide evidence in favour of the stability assumption of matching weights in SCD. Additionally, the SCD estimates for the post-treatment period indicate a negative effect of LAWA on Arizona’s share of non-citizen

FIGURE 4. Robustness Checks for ATT.



Notes: Panel (a) in this figure shows the ATT on Arizona's share of non-citizen Hispanic with a high-school diploma or less with respect to the prime-working age (15-45) state population. Panel (b) presents the ATT on Arizona's proportion of non-citizen Hispanic after applying SCD with a shorter pre-treatment window that starts in January 2003. Panel (c) displays the ATT after applying SCD with a uniform differencing parameter (λ^{unif}), and panel (d) uses no differencing at all. The blue and red lines correspond to 95% CIs constructed using Algorithm 1 for repeated cross-sectional data. Arizona's donor pool consists of 46 donor states. Donors with positive SCD weights for each of the robustness exercises are shown in Table 5. Data come from the monthly CPS. Survey weights are used.

Hispanics. In line with the 1.5 percentage point reduction reported by [Bohn et al. \(2014\)](#), we find that the proportion of this demographic group declined by 2 percentage points on average after July 2007 (equivalent to 125,000 fewer individuals with respect to Arizona's CPS population in June 2007).

6.2. Robustness Checks

To complement our main findings, we conduct four robustness exercises. First, since illegal immigrants tend to be low-educated, we refine our main outcome variable and use Arizona’s share of non-citizen Hispanic with high school or less among the population aged 15-45. Second, we shorten the pre-treatment window by starting the sample in January 2003 instead of January 1998. This relaxes the trend-matching requirement in SCD and provides a robustness check against potential overfitting of early-period dynamics. Third, we test how our results change when SCD matches on trends relative to the pre-treatment average rather than the last pre-treatment period by applying our SCD method with a uniform differencing parameter (λ^{unif}). Lastly, we report results using SCD without applying any differencing to the outcome, which is equivalent to the SC approach.

Table 5 reports the subset of Arizona’s donors with positive SCD weights for each robustness exercise. In general, donors contributing to synthetic Arizona differ across specifications. For instance, in column three where we adopt a shorter pre-treatment window, only Kansas and New Jersey also appear in Table 4. When we apply SC to our monthly CPS data, we obtain three donors with positive weights (California, Florida, and New Jersey), two of which coincide with those reported in Table 4. Interestingly, New Jersey is the only state with positive SCD weights across all exercises, highlighting its relevance as a comparison group for Arizona.

We present the ATT results for each robustness exercise in Figure 4. Panel (a) reveals that Arizona’s share of low-educated non-citizen Hispanics is reduced by 4.6 percentage points after one and a half years of LAWA’s enactment, suggesting that the policy’s impact was concentrated among less-educated immigrants. Next, in panel (b), we find that reducing the number of pre-treatment periods does not affect our baseline results and estimate an average post-treatment decline of 1.9 percentage points in Arizona’s share of non-citizen Hispanic. Finally, panels (c) and (d) show that our main ATT estimates in subsection 6.1 remain robust to alternative differencing choices in SCD. Specifically, we estimate an average post-LAWA decline of 1.8 percentage points when using uniform differencing and 1.3 percentage points when no differencing is applied.

Overall, these results show that Arizona experienced a significant change in its internal composition after LAWA, suggesting that the policy discouraged undocumented workers from residing in the state.

7. Conclusion

This paper considers the general framework of causal inference with groupwise matching. We show that many widely used designs, including RCTs, difference-in-differences (DID),

synthetic control (SC), synthetic control with differencing (SCD), and synthetic difference-in-differences (SDID), can be understood through the lenses of a Generalized Matching Condition (GMC) and the choice of within-group differencing parameters and matching weights. This perspective provides a clear distinction of how these methods extrapolate untreated outcomes. In particular, we show that DID relies on group-size based weights justified by the parallel trend assumption (PTA), whereas SCD relies on weights chosen by pre-treatment fit and justified by the stability of matching weights across time.

We examine the complementarity between DID and SCD through a regret analysis, showing that DID regret-dominates SCD when SCD's extrapolation error exceeds DID's matching error up to a term that vanishes at the parametric rate. In the opposite case, SCD regret-dominates DID, so neither method uniformly dominates the other. We further establish that, under conditions requiring both pre- and post-treatment parallel trends, DID and SCD are equivalent in terms of GMC. In this case, the weights estimated by SCD using pre-treatment data coincide with the group-size weights implicit in DID. When PTA fails, however, SCD provides an alternative to DID where identification no longer relies on parallel trends, but instead on a Stable Matching Condition (SMC). We develop a uniformly valid inference procedure for SCD based on a Bonferroni approach. Finally, our theoretical insights are illustrated via Monte Carlo experiments and revisiting the empirical application of [Bohn et al. \(2014\)](#), who study the impact of the 2007 Legal Arizona Workers Act (LAWA) on Arizona's internal composition.

While the GMC is formulated in a setting where the target parameter is the average causal effect of a treatment, it is conceivable that the condition extends to the setting where the target parameter is the distribution of the causal effect. This opens up the question for future research of how the causal inference designs such as changes in changes of [Athey and Imbens \(2006\)](#) and the distributional synthetic control of [Gunsilius \(2023\)](#) compare.

References

- ABADIE, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, **59**, 391–425.
- , DIAMOND, A. and HAINMUELLER, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, **105**, 493–505.
- ARKHANGELSKY, D., ATHEY, S., HIRSHBERG, D. A., IMBENS, G. W. and WAGER, S. (2021). Synthetic difference-in-differences. *American Economic Review*, **111**, 4088–4118.
- ATHEY, S. and IMBENS, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, **74**, 431–497.

- BILINSKI, A. and HATFIELD, L. A. (2019). Nothing to see here? non-inferiority approaches to parallel trends and other model assumptions. *arXiv:1805.03273v5 [stat.ME]*.
- BOHN, S., LOFSTROM, M. and RAPHAEL, S. (2014). Did the 2007 legal arizona workers act reduce the state's unauthorized immigrant population? *The Review of Economics and Statistics*, **96**, 258–269.
- BORUSYAK, K., JARAVEL, X. and SPIESS, J. (2024). Revisiting event-study designs: Robust and efficient estimation. *The Review of Economic Studies*, **91**, 3253–3285.
- CALLAWAY, B. and SANT'ANNA, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, **225**, 200–230.
- CANEN, N. and SONG, K. (2025). Simple inference on a simplex-valued weight. *arXiv:2501.15692v1 [econ.EM]*.
- CHEN, J. (2023). Synthetic control as online linear regression. *Econometrica*, **91**, 465–491.
- CHEN, X., SANT'ANNA, P. H. and XIE, H. (2025). Efficient difference-in-differences and event study estimators. *arXiv:2506.17729v1 [econ.EM]*.
- CHUNG, F. and LU, L. (2002). Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics* 2002 6:2, **6**, 125–145.
- CLARKE, F. H. (1990). *Optimization and Nonsmooth Analysis, Classics in Applied Mathematics*, vol. 5. New York.
- DE CHAISEMARTIN, C. and D'HAULTFŒUILLE, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, **110**, 2964–96.
- and — (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: a survey. *The Econometrics Journal*, **26**, C1–C30.
- DOUDCHENKO, N. and IMBENS, G. W. (2017). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. *arXiv*, pp. 1–36.
- FERMAN, B. and PINTO, C. (2021). Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, **12**, 1197–1221.
- FREYALDENHOVEN, S., HANSEN, C. and SHAPIRO, J. M. (2019). Pre-event trends in the panel event-study design. *American Economic Review*, **109**, 3307–3338.
- GOODMAN-BACON, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, **225**, 254–277.
- GUNSILIUS, F. F. (2023). Distributional synthetic controls. *Econometrica*, **91**, 1105–1117.
- HECKMAN, J. J., ICHIMURA, H. and TODD, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, **64**, 605–654.
- KAHN-LANG, A. and LANG, K. (2020). The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications. *Journal of Business & Economic Statistics*, **38**, 613–620.

- KELLOGG, M., MOGSTAD, M., POULIOT, G. A. and TORGOVITSKY, A. (2021). Combining matching and synthetic control to tradeoff biases from extrapolation and interpolation. *Journal of the American Statistical Association*, **116**, 1804–1816.
- KUERSTEINER, G. M. and PRUCHA, I. R. (2020). Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. *Econometrica*, **88**, 2109–2146.
- KWON, S. and ROTH, J. (2024). (empirical) bayes approaches to parallel trends. *American Economic Association Papers and Proceedings*, **114**, 606–609.
- LEE, S. J. and WOOLDRIDGE, J. M. (2024). A simple transformation approach to difference-in-difference estimation for panel data. *Working Paper*.
- LIU, Y. (2025). Synthetic parallel trends. *arXiv:2511.05870v1 [econ.EM]*.
- MANSKI, C. F. and PEPPER, J. V. (2018). How do right-to-carry laws affect crime rates? coping with ambiguity using bounded-variation assumptions. *Review of Economics and Statistics*, **100**, 232–244.
- RAMBACHAN, A. and ROTH, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, **90**, 2555–2591.
- ROBBINS, M. W., SAUNDERS, J. and KILMER, B. (2017). A framework for synthetic control methods with high-dimensional, micro-level data: Evaluating a neighborhood-specific crime intervention. *Journal of the American Statistical Association*, **112**, 109–126.
- ROTH, J., SANT’ANNA, P. H., BILINSKI, A. and POE, J. (2023). What’s trending in difference-in-differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, **235**.
- SHI, C., SRIDHAR, D., MISRA, V. and BLEI, D. M. (2022). On the assumptions of synthetic control methods. *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS)*.
- SMITH, J. A. and TODD, P. E. (2005). Does matching overcome lalonde’s critique of nonexperimental estimators? *Journal of Econometrics*, **125**, 305–353.
- SUN, L. and ABRAHAM, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, **225**, 175–199.
- SUN, Y., XIE, H. and ZHANG, Y. (2025). Difference-in-differences meets synthetic control: Doubly robust identification and estimation. *arXiv:2503.11375v1 [econ.EM]*.
- WOOLDRIDGE, J. M. (2021). Two-way fixed effects, the two-way mundlak regression, and difference-in-differences estimators. *SSRN 3906345*.
- XU, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, **25**, 57–76.

SUPPLEMENTAL NOTE TO “CAUSAL INFERENCE WITH GROUPWISE MATCHING”

Ratzanyel Rincón and Kyungchul Song
Vancouver School of Economics, University of British Columbia

Appendix A. Proofs of the Results in Sections 2 and 3

Lemma A.1. Suppose that Assumption 2.1 holds. Then, for $t \in \mathcal{T}$,

$$(A.1) \quad \theta_t^* = \theta_t(\lambda, w) - e_t(\lambda, w).$$

Proof: By (2.2) and Assumption 2.1(ii), for all $t \in \mathcal{T}_0$, $m_{0,t} = m_{0,t}(0)$. Furthermore, for all $j \in \mathcal{G}_{\text{don}}$, $m_{j,t} = m_{j,t}(0)$ for all $t \in \mathcal{T}$, by (2.2) and the definition of $m_{j,t}(0)$. Hence,

$$\begin{aligned} \theta_t(\lambda, w) - e_t(\lambda, w) &= m_{0,t} - m_{0,t}(0) - \sum_{j=1}^K (m_{j,t} - m_{j,t}(0))w_j \\ &= m_{0,t} - m_{0,t}(0). \end{aligned}$$

Now by (2.2), if we let $m_{0,t}(1) = \mathbf{E}[Y_{i,t}(1) \mid D_i = 1]$, we can write the last difference as $m_{0,t}(1) - m_{0,t}(0)$. This delivers the desired result. ■

Proof of Proposition 2.1: First, note that

$$(A.2) \quad e_t(\lambda, w) = \left(\mathbf{E}[\Lambda'_i \mid G_i = 0] - \sum_{j=1}^K \mathbf{E}[\Lambda'_i \mid G_i = j]w_j \right) F_t(\lambda).$$

Thus, (iii) implies (ii), which implies (i). Now, suppose that the full row rank condition holds. Define $\mathbf{e}(\lambda, w) = [e_{T^*}(\lambda, w), \dots, e_T(\lambda, w)]'$. Then, we have

$$\mathbf{E}[\Lambda'_i \mid G_i = 0] - \sum_{j=1}^K \mathbf{E}[\Lambda'_i \mid G_i = j]w_j = \mathbf{e}(\lambda, w)' \mathbf{F}(\lambda)' (\mathbf{F}(\lambda) \mathbf{F}(\lambda)')^{-1}.$$

Now, suppose that (i) holds. Then, this implies (iii), completing the proof. ■

Lemma A.2. (i) Suppose that $e_{T^*-1}^{\text{DID}}(\lambda) = 0$ holds for some $\lambda \in \Delta_{|\mathcal{T}_0|-1}$. Then, PTA-I holds if and only if PTA(λ) holds.

(ii) Suppose that for some $\lambda \in \Delta_{|\mathcal{T}_0|-1}$, $e_{j,T^*-1}^{\text{DID}}(\lambda) = 0$ holds for each $j \in \mathcal{G}_{\text{don}}$. Then, PTA-II holds if and only if PTA-U(λ) holds.

Proof: (i) We assume that $e_{T^*-1}^{\text{DID}}(\lambda) = 0$ for some $\lambda \in \Delta_{|\mathcal{T}_0|-1}$. Now, suppose that PTA(λ) holds. Then, $e_t^{\text{DID}}(\lambda) = 0$ for all $t \in \mathcal{T}_1 \cup \{T^* - 1\}$. Hence, for all $t \in \mathcal{T}_1$,

$$\begin{aligned} \mathbb{E}[Y_{i,t}(0; \lambda) | G_i = 0] &= \mathbb{E}[Y_{i,t}(0; \lambda) | G_i \in \mathcal{G}_{\text{don}}] \text{ and} \\ \mathbb{E}[Y_{i,t-1}(0; \lambda) | G_i = 0] &= \mathbb{E}[Y_{i,t-1}(0; \lambda) | G_i \in \mathcal{G}_{\text{don}}]. \end{aligned}$$

Subtracting the second equation from the first one, we obtain

$$\mathbb{E}[\Delta Y_{i,t}(0) | G_i = 0] = \mathbb{E}[\Delta Y_{i,t}(0) | G_i \in \mathcal{G}_{\text{don}}],$$

for all $t \in \mathcal{T}_1$. Hence, PTA-I holds.

Conversely, suppose that PTA-I holds. Then,

$$\begin{aligned} \mathbb{E}[Y_{i,t}(0; \lambda) | G_i = 0] &= \mathbb{E}[Y_{i,t}(0) - Y_{i,T^*-1}(0) | G_i = 0] + \mathbb{E}\left[Y_{i,T^*-1}(0) - \sum_{s \in \mathcal{T}_0} Y_{i,s}(0)\lambda_s | G_i = 0\right] \\ &= \mathbb{E}[Y_{i,t}(0) - Y_{i,T^*-1}(0) | G_i = 0] + \mathbb{E}\left[Y_{i,T^*-1}(0) - \sum_{s \in \mathcal{T}_0} Y_{i,s}(0)\lambda_s | G_i \in \mathcal{G}_{\text{don}}\right], \end{aligned}$$

because $e_{T^*-1}^{\text{DID}}(\lambda) = 0$. The last sum of two conditional expectations is written as

$$\begin{aligned} &\sum_{\ell=T^*}^t \mathbb{E}[\Delta Y_{i,\ell}(0) | G_i = 0] + \mathbb{E}\left[Y_{i,T^*-1}(0) - \sum_{s \in \mathcal{T}_0} Y_{i,s}(0)\lambda_s | G_i \in \mathcal{G}_{\text{don}}\right] \\ &= \sum_{\ell=T^*}^t \mathbb{E}[\Delta Y_{i,\ell}(0) | G_i \in \mathcal{G}_{\text{don}}] + \mathbb{E}\left[Y_{i,T^*-1}(0) - \sum_{s \in \mathcal{T}_0} Y_{i,s}(0)\lambda_s | G_i \in \mathcal{G}_{\text{don}}\right] \\ &= \mathbb{E}\left[Y_{i,t}(0) - \sum_{s \in \mathcal{T}_0} Y_{i,s}(0)\lambda_s | G_i \in \mathcal{G}_{\text{don}}\right] = \mathbb{E}[Y_{i,t}(0; \lambda) | G_i \in \mathcal{G}_{\text{don}}]. \end{aligned}$$

The first equality follows by PTA-I. Hence, PTA(λ) holds.

(ii) We assume that $e_{j,T^*-1}^{\text{DID}}(\lambda) = 0$ for each $j \in \mathcal{G}_{\text{don}}$, for some $\lambda \in \Delta_{|\mathcal{T}_0|-1}$. First, assume that PTA-U(λ) holds. We choose $j \in \mathcal{G}_{\text{don}}$, and replace the event $G_i \in \mathcal{G}_{\text{don}}$ by the event $G_i = j$ in the proof of (i), to obtain that

$$\mathbb{E}[\Delta Y_{i,t}(0) | G_i = 0] = \mathbb{E}[\Delta Y_{i,t}(0) | G_i = j],$$

for all $t \in \mathcal{T}_1$. Since the choice of j was arbitrary, we obtain PTA-II. Conversely, suppose that PTA-II holds. Again, choose $j \in \mathcal{G}_{\text{don}}$, and replace the event $G_i \in \mathcal{G}_{\text{don}}$ by the event $G_i = j$ in the proof of (i) to obtain

$$\mathbb{E}[Y_{i,t}(0; \lambda) | G_i = 0] = \mathbb{E}[Y_{i,t}(0; \lambda) | G_i = j]$$

for each $j \in \mathcal{G}_{\text{don}}$. ■

Lemma A.3. Suppose that Assumption 2.1 holds, and let

$$w^{\text{DID}} = [w_1^{\text{DID}}, \dots, w_K^{\text{DID}}]',$$

where w_j^{DID} is as defined in (3.2). Then, the following statements hold.

- (i) PTA(λ) holds if and only if GMC holds at $(\lambda, w^{\text{DID}})$.
- (ii) PTA-U(λ) holds if and only if GMC holds at (λ, w) for all $w \in \Delta_{K-1}$.

Proof: (i) Notice that $e_t(\lambda, w)$ and $e_t^{\text{DID}}(\lambda)$ have the following relationship:

$$\begin{aligned} e_t^{\text{DID}}(\lambda) &= \mathbf{E}[Y_{i,t}(0; \lambda) \mid G_i = 0] - \mathbf{E}[Y_{i,t}(0; \lambda) \mid G_i \in \mathcal{G}_{\text{don}}] \\ &= \mathbf{E}[Y_{i,t}(0; \lambda) \mid G_i = 0] - \sum_{j=1}^K \mathbf{E}[Y_{i,t}(0; \lambda) \mid G_i = j] w_j^{\text{DID}} = e_t(\lambda, w^{\text{DID}}), \quad \text{for all } t \in \mathcal{T}, \end{aligned}$$

where $w_j^{\text{DID}} = P\{G_i = j \mid G_i \in \mathcal{G}_{\text{don}}\}$. Hence,

$$\text{PTA}(\lambda) \text{ holds.} \Leftrightarrow \text{GMC}(\lambda, w^{\text{DID}}) \text{ holds.}$$

(ii) Note that

(A.3)

$$\begin{aligned} \sum_{j=1}^K e_{j,t}^{\text{DID}}(\lambda) w_j &= \sum_{j=1}^K (\mathbf{E}[Y_{i,t}(0; \lambda) \mid G_i = 0] - \mathbf{E}[Y_{i,t}(0; \lambda) \mid G_i = j]) w_j \\ &= \mathbf{E}[Y_{i,t}(0; \lambda) \mid G_i = 0] - \sum_{j=1}^K \mathbf{E}[Y_{i,t}(0; \lambda) \mid G_i = j] w_j = e_t(\lambda, w), \quad \text{for all } t \in \mathcal{T}. \end{aligned}$$

Hence,

$$\begin{aligned} \text{PTA-U}(\lambda) \text{ holds.} &\Rightarrow e_t(\lambda, w) = 0, \quad \text{for all } t \in \mathcal{T}_1 \text{ and all } w \in \Delta_{K-1}, \\ &\text{i.e., GMC}(\lambda, w) \text{ holds for all } w \in \Delta_{K-1}. \end{aligned}$$

Conversely, suppose that $\text{GMC}(\lambda, w)$ holds for all $w \in \Delta_{K-1}$. Then, for any $\ell = 1, \dots, K$, we have $\tilde{w}^\ell \in \Delta_{K-1}$, where $\tilde{w}^\ell$ denotes the K -dimensional vector of zeros except for the ℓ -th entry which is equal to one. Then,

$$\text{GMC holds at } (\lambda, \tilde{w}^\ell) \Rightarrow e_{\ell,t}^{\text{DID}}(\lambda) = \sum_{j=1}^K e_{j,t}^{\text{DID}}(\lambda) \tilde{w}_j^\ell = e_t(\lambda, \tilde{w}^\ell) = 0,$$

for all $t \in \mathcal{T}_1$, where the last equality follows from (A.3). We can repeat this for all $\ell = 1, \dots, K$, to obtain that PTA-U(λ) holds. ■

Proof of Proposition 3.1: Note that $e_{T^*-1}^{\text{DID}}(\lambda^{\text{DID}}) = 0$. Hence, the desired result follows by Lemmas A.2 and A.3. ■

Appendix B. Proofs of the Results in Section 4

Proof of Proposition 4.1: The proposition is the same as Lemma A.2. ■

We turn to the proof of Theorem 4.1. For the results below, we assume that the assumptions of the theorem hold. Recall the definition $\text{MER}_{d,p}(w) = \eta_{d,p}(w) - \inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_p[\eta_{d,p}(\tilde{w}(Z))]$,

where $\eta_{d,p}(w) = \text{SSME}_{d,p}(w)$, $d = 0, 1$. For $d = 0, 1$, we define

$$\hat{\eta}_d(w) = \frac{1}{|\mathcal{T}_d|} \sum_{t \in \mathcal{T}_d} \hat{e}_t^2(\lambda, w).$$

Despite the notation, we cannot actually recover $\hat{\eta}_1(w)$ from data, because we do not observe the untreated potential outcomes for the treated group. That is, we do not observe $Y_{i,t}(0)$ for $t \in \mathcal{T}_1$. However, we can construct $\hat{\eta}_0(w)$ from data.

Also, let

$$\hat{e}_{j,t} = \hat{\mu}_{j,t}(\lambda) - \mu_{j,t}(\lambda) \text{ and } \hat{e}_{j,t}(0) = \hat{\mu}_{j,t}(0; \lambda) - \mu_{j,t}(0; \lambda),$$

where $\hat{\mu}_{j,t}(0; \lambda)$ is defined to be the same as $\hat{\mu}_{j,t}(\lambda)$ except that $Y_{i,t}$ is replaced by $Y_{i,t}(0)$.

Lemma B.1. *For each $P \in \mathcal{P}$,*

$$\begin{aligned} \inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_P[\hat{\eta}_0(\tilde{w}(Z))] &= \inf_{\tilde{w} \in \mathcal{D}_0} \mathbf{E}_P[\hat{\eta}_0(\tilde{w}(Z_0))] = \mathbf{E}_P \left[\inf_{w \in \Delta_{K-1}} \hat{\eta}_0(w) \right], \text{ and} \\ \inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_P[\eta_{0,p}(\tilde{w}(Z))] &= \inf_{w \in \Delta_{K-1}} \eta_{0,p}(w). \end{aligned}$$

Proof: First, note that

$$\begin{aligned} \text{(B.1)} \quad \inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_P[\hat{\eta}_0(\tilde{w}(Z))] &\leq \inf_{\tilde{w} \in \mathcal{D}_0} \mathbf{E}_P[\hat{\eta}_0(\tilde{w}(Z_0))] \\ &= \mathbf{E}_P \left[\inf_{w \in \Delta_{K-1}} \hat{\eta}_0(w) \right] \leq \inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_P[\hat{\eta}_0(\tilde{w}(Z))], \end{aligned}$$

where $\mathcal{D}_0$ denotes the Δ_{K-1} -valued maps that are measurable with respect to the σ -field generated by the pre-treatment data Z_0 .²⁷ The equality above follows because $\hat{\eta}_0(\cdot)$ is measurable with respect to the σ -field generated by Z_0 .

The second statement follows because

$$\inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_P[\eta_{0,p}(\tilde{w}(Z))] \leq \inf_{w \in \Delta_{K-1}} \eta_{0,p}(w) = \mathbf{E}_P \left[\inf_{w \in \Delta_{K-1}} \eta_{0,p}(w) \right] \leq \inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_P[\eta_{0,p}(\tilde{w}(Z))].$$

The first inequality follows because $\mathcal{D}$ includes constant maps taking values in Δ_{K-1} , and the equality follows because $\eta_{0,p}(\cdot)$ is nonstochastic. ■

Lemma B.2. *For $c > 0$ in Assumption 4.2,*

$$\text{SSME}_{0,p}(w) - \inf_{\tilde{w} \in \Delta_{K-1}} \text{SSME}_{0,p}(\tilde{w}) \geq c \|w - w_p^{\text{SCD}}\|^2.$$

²⁷Note that $\hat{\eta}_0(w)$ is continuous in w everywhere. Hence, $\inf_{w \in \Delta_{K-1}} \hat{\eta}_0(w) = \inf_{w \in \Delta_{K-1} \cap \mathbf{Q}^K} \hat{\eta}_0(w)$, where $\mathbf{Q}$ is the set of rational numbers. Therefore, $\inf_{w \in \Delta_{K-1}} \hat{\eta}_0(w)$ is a random variable.

Proof: Let $h(\lambda) = [\mu_{0,1}(\lambda), \dots, \mu_{0,T^*-1}(\lambda)]'$. Note that

$$\begin{aligned} & \text{SSME}_{0,p}(w) - \inf_{\tilde{w} \in \Delta_{K-1}} \text{SSME}_{0,p}(\tilde{w}) \\ &= \frac{1}{|\mathcal{T}_0|} (\Gamma_p(w - w_p^{\text{SCD}}))' (\Gamma_p(w - w_p^{\text{SCD}})) + \frac{2}{|\mathcal{T}_0|} (\Gamma_p w_p^{\text{SCD}} - h(\lambda))' \Gamma_p(w - w_p^{\text{SCD}}) \\ &\geq \frac{1}{|\mathcal{T}_0|} (\Gamma_p(w - w_p^{\text{SCD}}))' (\Gamma_p(w - w_p^{\text{SCD}})) \geq \frac{\lambda_{\min}(\Gamma_p' \Gamma_p)}{|\mathcal{T}_0|} \|w - w_p^{\text{SCD}}\|^2, \end{aligned}$$

where the inequality follows because $(\Gamma_p w_p^{\text{SCD}} - h(\lambda))' \Gamma_p(w - w_p^{\text{SCD}}) \geq 0$ by the optimality of w_p^{SCD} (e.g., Propositions 2.1.5 and 2.3.2 of [Clarke \(1990\)](#)). ■

Define

$$\text{Regret}_{1,p}(\hat{w}) = \mathbf{E}_p[\ell_1(\hat{w})] - \inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_p[\ell_1(\tilde{w}(Z))],$$

and let

$$R_{n,p}(w) = -\frac{2}{|\mathcal{T}_1|} \sum_{t \in \mathcal{T}_1} e_t(\lambda, w) (\theta_t(\lambda, w) - \hat{\theta}_t(\lambda, w)) + \frac{1}{|\mathcal{T}_1|} \sum_{t \in \mathcal{T}_1} (\theta_t(\lambda, w) - \hat{\theta}_t(\lambda, w))^2.$$

Lemma B.3. For any estimator $\hat{w} \in \Delta_{K-1}$ and for each $P \in \mathcal{P}$,

$$|\text{Regret}_{1,p}(\hat{w}) - \mathbf{E}_p[\text{MER}_1(\hat{w})]| \leq 2\mathbf{E}_p \left[\sup_{w \in \Delta_{K-1}} |R_{n,p}(w)| \right].$$

Proof: Since $e_t(\lambda, w) = \theta_t(\lambda, w) - \theta_t^*$ by Lemma A.1, we write

$$\begin{aligned} \mathbf{E}_p \left[(\theta_t^* - \hat{\theta}_t(\lambda, \hat{w}))^2 \right] &= \mathbf{E}_p \left[(\theta_t^* - \theta_t(\lambda, \hat{w}) + \theta_t(\lambda, \hat{w}) - \hat{\theta}_t(\lambda, \hat{w}))^2 \right] \\ &= \mathbf{E}_p [e_t^2(\lambda, \hat{w})] - 2\mathbf{E}_p [e_t(\lambda, \hat{w})(\theta_t(\lambda, \hat{w}) - \hat{\theta}_t(\lambda, \hat{w}))] \\ &\quad + \mathbf{E}_p [(\theta_t(\lambda, \hat{w}) - \hat{\theta}_t(\lambda, \hat{w}))^2]. \end{aligned}$$

Hence,

$$\begin{aligned} \text{Regret}_{1,p}(\hat{w}) &= \mathbf{E}_p [\eta_{1,p}(\hat{w}) + R_{n,p}(\hat{w})] - \inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_p [\eta_{1,p}(\tilde{w}(Z)) + R_{n,p}(\tilde{w}(Z))] \\ &= \mathbf{E}_p [\text{MER}_1(\hat{w})] + \mathbf{E}_p [R_{n,p}(\hat{w})] \\ &\quad - \left\{ \inf_{\tilde{w} \in \mathcal{D}} (\mathbf{E}_p [\eta_{1,p}(\tilde{w}(Z))] + \mathbf{E}_p [R_{n,p}(\tilde{w}(Z))]) - \inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_p [\eta_{1,p}(\tilde{w}(Z))] \right\}. \end{aligned}$$

As for the last term,

$$\begin{aligned} & \left| \inf_{\tilde{w} \in \mathcal{D}} (\mathbf{E}_p [\eta_{1,p}(\tilde{w}(Z))] + \mathbf{E}_p [R_{n,p}(\tilde{w}(Z))]) - \inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_p [\eta_{1,p}(\tilde{w}(Z))] \right| \\ &\leq \inf_{\tilde{w} \in \mathcal{D}} \left(\mathbf{E}_p [\eta_{1,p}(\tilde{w}(Z))] + \mathbf{E}_p \left[\sup_{w \in \Delta_{K-1}} |R_{n,p}(w)| \right] \right) - \inf_{\tilde{w} \in \mathcal{D}} \mathbf{E}_p [\eta_{1,p}(\tilde{w}(Z))] = \mathbf{E}_p \left[\sup_{w \in \Delta_{K-1}} |R_{n,p}(w)| \right]. \end{aligned}$$

Thus, we obtain the desired bound. ■

We define

$$\hat{D}_{d,1} = \frac{1}{|\mathcal{T}_d|} \sum_{t \in \mathcal{T}_d} \sum_{j=0}^K |\hat{\varepsilon}_{j,t}| \text{ and } \hat{D}_{d,2}^2 = \frac{1}{|\mathcal{T}_d|} \sum_{t \in \mathcal{T}_d} \sum_{j=0}^K \hat{\varepsilon}_{j,t}^2.$$

We define similarly $\hat{D}_{d,1}(0)$ and $\hat{D}_{d,2}^2(0)$ with $\hat{\mu}_{j,t}(\lambda)$ and $\mu_{j,t}(\lambda)$ replaced by $\hat{\mu}_{j,t}(0; \lambda)$ and $\mu_{j,t}(0; \lambda)$.

Lemma B.4. *For each $P \in \mathcal{P}$ and $w \in \Delta_{K-1}$, the following statements hold.*

$$(i) \quad |R_{n,p}(w)| \leq 8\bar{m}\hat{D}_{1,1} + 2\hat{D}_{1,2}^2.$$

$$(ii) \quad |\hat{\eta}_0(w) - \eta_{0,p}(w)| \leq 8\bar{m}\hat{D}_{0,1}(0) + 2\hat{D}_{0,2}^2(0).$$

Proof: (i) First, by (4.6), we have $|e_t(\lambda, w)| \leq 4\bar{m}$, for all $w \in \Delta_{K-1}$. Hence,

$$|R_{n,p}(\hat{w})| \leq \frac{8\bar{m}}{|\mathcal{T}_1|} \sum_{t \in \mathcal{T}_1} |\theta_t(\lambda, \hat{w}) - \hat{\theta}_t(\lambda, \hat{w})| + \frac{1}{|\mathcal{T}_1|} \sum_{t \in \mathcal{T}_1} (\theta_t(\lambda, \hat{w}) - \hat{\theta}_t(\lambda, \hat{w}))^2.$$

Note that

$$|\theta_t(\lambda, \hat{w}) - \hat{\theta}_t(\lambda, \hat{w})| \leq \sum_{j=0}^K |\hat{\varepsilon}_{j,t}|.$$

Also, note that

$$\begin{aligned} (\theta_t(\lambda, \hat{w}) - \hat{\theta}_t(\lambda, \hat{w}))^2 &\leq 2\hat{\varepsilon}_{0,t}^2 + 2\left(\sum_{j=1}^K \hat{\varepsilon}_{j,t} \hat{w}_j\right)^2 \\ &\leq 2\hat{\varepsilon}_{0,t}^2 + 2\sum_{j=1}^K \hat{\varepsilon}_{j,t}^2 \hat{w}_j \leq 2\sum_{j=0}^K \hat{\varepsilon}_{j,t}^2. \end{aligned}$$

The second inequality follows from Jensen's inequality. Combining these, we obtain the desired result.

(ii) Note that

$$\begin{aligned} |\hat{\eta}_0(w) - \eta_{0,p}(w)| &\leq \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} |\hat{e}_t^2(\lambda, w) - e_t^2(\lambda, w)| \\ &\leq \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} (\hat{e}_t(\lambda, w) - e_t(\lambda, w))^2 \\ &\quad + \frac{2}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} |e_t(\lambda, w)| |\hat{e}_t(\lambda, w) - e_t(\lambda, w)|. \end{aligned}$$

Now, observe that

$$|\hat{e}_t(\lambda, w) - e_t(\lambda, w)| \leq \sum_{j=0}^K |\hat{\varepsilon}_{j,t}(0)|,$$

and

$$(\hat{e}_t(\lambda, w) - e_t(\lambda, w))^2 \leq 2\hat{\varepsilon}_{0,t}^2(0) + 2 \sum_{j=1}^K \hat{\varepsilon}_{j,t}^2(0) \leq 2 \sum_{j=0}^K \hat{\varepsilon}_{j,t}^2(0).$$

Therefore, $|\hat{\eta}_0(w) - \eta_{0,p}(w)| \leq 8\bar{m}\hat{D}_{0,1}(0) + 2\hat{D}_{0,2}^2(0)$. ■

Lemma B.5. For each $P \in \mathcal{P}$,

$$\begin{aligned} & \hat{\eta}_0(w_p^{\text{SCD}}) - \eta_{0,p}(w_p^{\text{SCD}}) - (\hat{\eta}_0(\hat{w}^{\text{SCD}}) - \eta_{0,p}(\hat{w}^{\text{SCD}})) \\ & \leq 4(\hat{D}_{0,2}^2 + 4\bar{m}\hat{D}_{0,1}) \sum_{j=1}^K |\hat{w}_j^{\text{SCD}} - w_{j,p}^{\text{SCD}}|. \end{aligned}$$

Proof: We first write $\hat{\eta}_0(w_p^{\text{SCD}}) - \eta_{0,p}(w_p^{\text{SCD}}) - (\hat{\eta}_0(\hat{w}^{\text{SCD}}) - \eta_{0,p}(\hat{w}^{\text{SCD}}))$ as

$$\begin{aligned} & \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} (\hat{e}_t(\lambda, w_p^{\text{SCD}}) + \hat{e}_t(\lambda, \hat{w}^{\text{SCD}})) (\hat{e}_t(\lambda, w_p^{\text{SCD}}) - \hat{e}_t(\lambda, \hat{w}^{\text{SCD}})) \\ & - \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} (e_t(\lambda, w_p^{\text{SCD}}) + e_t(\lambda, \hat{w}^{\text{SCD}})) (e_t(\lambda, w_p^{\text{SCD}}) - e_t(\lambda, \hat{w}^{\text{SCD}})) \\ & = \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} (\hat{e}_t(\lambda, w_p^{\text{SCD}}) + \hat{e}_t(\lambda, \hat{w}^{\text{SCD}}) - (e_t(\lambda, w_p^{\text{SCD}}) + e_t(\lambda, \hat{w}^{\text{SCD}}))) (\hat{e}_t(\lambda, w_p^{\text{SCD}}) - \hat{e}_t(\lambda, \hat{w}^{\text{SCD}})) \\ & + \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} (e_t(\lambda, w_p^{\text{SCD}}) + e_t(\lambda, \hat{w}^{\text{SCD}})) (\hat{e}_t(\lambda, w_p^{\text{SCD}}) - \hat{e}_t(\lambda, \hat{w}^{\text{SCD}}) - (e_t(\lambda, w_p^{\text{SCD}}) - e_t(\lambda, \hat{w}^{\text{SCD}}))). \end{aligned}$$

By rearranging terms, we can write the sum of the last sums as

$$\begin{aligned} & \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} \left\{ \left(2\hat{\varepsilon}_{0,t} - \sum_{j=1}^K \hat{\varepsilon}_{j,t}(\hat{w}_j^{\text{SCD}} + w_{j,p}^{\text{SCD}}) \right) \times \sum_{j=1}^K \hat{\mu}_{j,t}(\lambda)(\hat{w}_j^{\text{SCD}} - w_{j,p}^{\text{SCD}}) \right. \\ & \quad \left. + \left(2\mu_{0,t}(\lambda) - \sum_{j=1}^K \mu_{j,t}(\lambda)(\hat{w}_j^{\text{SCD}} + w_{j,p}^{\text{SCD}}) \right) \times \sum_{j=1}^K \hat{\varepsilon}_{j,t}(\hat{w}_j^{\text{SCD}} - w_{j,p}^{\text{SCD}}) \right\} \\ & = \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} \left\{ \left(2\hat{\varepsilon}_{0,t} - \sum_{j=1}^K \hat{\varepsilon}_{j,t}(\hat{w}_j^{\text{SCD}} + w_{j,p}^{\text{SCD}}) \right) \times \sum_{j=1}^K \hat{\varepsilon}_{j,t}(\hat{w}_j^{\text{SCD}} - w_{j,p}^{\text{SCD}}) \right. \\ & \quad + \left(2\hat{\varepsilon}_{0,t} - \sum_{j=1}^K \hat{\varepsilon}_{j,t}(\hat{w}_j^{\text{SCD}} + w_{j,p}^{\text{SCD}}) \right) \times \sum_{j=1}^K \mu_{j,t}(\lambda)(\hat{w}_j^{\text{SCD}} - w_{j,p}^{\text{SCD}}) \\ & \quad \left. + \left(2\mu_{0,t}(\lambda) - \sum_{j=1}^K \mu_{j,t}(\lambda)(\hat{w}_j^{\text{SCD}} + w_{j,p}^{\text{SCD}}) \right) \times \sum_{j=1}^K \hat{\varepsilon}_{j,t}(\hat{w}_j^{\text{SCD}} - w_{j,p}^{\text{SCD}}) \right\} \\ & \leq \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} \left\{ 4 \max_{0 \leq j \leq K} |\hat{\varepsilon}_{j,t}|^2 + 16\bar{m} \max_{1 \leq j \leq K} |\hat{\varepsilon}_{j,t}| \right\} \times \sum_{j=1}^K |\hat{w}_j^{\text{SCD}} - w_{j,p}^{\text{SCD}}|. \end{aligned}$$

Since

$$\frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} \max_{1 \leq j \leq K} |\hat{\varepsilon}_{j,t}| \leq \hat{D}_{0,1} \text{ and } \frac{1}{|\mathcal{T}_0|} \sum_{t \in \mathcal{T}_0} \max_{1 \leq j \leq K} |\hat{\varepsilon}_{j,t}|^2 \leq \hat{D}_{0,2}^2,$$

we obtain the desired result. ■

Lemma B.6. *For each $P \in \mathcal{P}$,*

$$\sum_{j=1}^K |\hat{w}_j^{\text{SCD}} - w_{j,P}^{\text{SCD}}| \leq \frac{4K(\hat{D}_{0,2}^2 + 4\bar{m}\hat{D}_{0,1})}{c}.$$

Proof: Note that

$$\begin{aligned} & \hat{\eta}_0(w_p^{\text{SCD}}) - \eta_{0,P}(w_p^{\text{SCD}}) - (\hat{\eta}_0(\hat{w}^{\text{SCD}}) - \eta_{0,P}(\hat{w}^{\text{SCD}})) \\ & \geq \eta_{0,P}(\hat{w}^{\text{SCD}}) - \eta_{0,P}(w_p^{\text{SCD}}) \geq c \|\hat{w}^{\text{SCD}} - w_p^{\text{SCD}}\|^2 \geq \frac{c}{K} \left(\sum_{j=1}^K |\hat{w}_j^{\text{SCD}} - w_{j,P}^{\text{SCD}}| \right)^2, \end{aligned}$$

by Lemma B.2. From Lemma B.5, we obtain the desired result. ■

Lemma B.7. *For each $P \in \mathcal{P}$,*

$$|\eta_{1,P}(\hat{w}^{\text{SCD}}) - \eta_{1,P}(w_p^{\text{SCD}})| \leq \frac{64\bar{m}^2 K}{c} (\hat{D}_{0,2}^2 + 4\bar{m}\hat{D}_{0,1}).$$

Proof: First, note that $\eta_{1,P}(\hat{w}^{\text{SCD}}) - \eta_{1,P}(w_p^{\text{SCD}})$ is equal to

$$\begin{aligned} & \frac{1}{|\mathcal{T}_1|} \sum_{t \in \mathcal{T}_1} (e_t^2(\lambda, \hat{w}^{\text{SCD}}) - e_t^2(\lambda, w_p^{\text{SCD}})) \\ & = \frac{1}{|\mathcal{T}_1|} \sum_{t \in \mathcal{T}_1} \left(2\mu_{0,t}(0; \lambda) - \sum_{j=1}^K \mu_{j,t}(0; \lambda) (\hat{w}_j^{\text{SCD}} + w_{j,P}^{\text{SCD}}) \right) \left(\sum_{j=1}^K \mu_{j,t}(0; \lambda) (w_{j,P}^{\text{SCD}} - \hat{w}_j^{\text{SCD}}) \right). \end{aligned}$$

Hence, by Assumption 4.2,

$$|\eta_{1,P}(\hat{w}^{\text{SCD}}) - \eta_{1,P}(w_p^{\text{SCD}})| \leq 16\bar{m}^2 \cdot \sum_{j=1}^K |w_{j,P}^{\text{SCD}} - \hat{w}_j^{\text{SCD}}|.$$

The desired result follows by Lemma B.6. ■

Lemma B.8. *There exists a universal constant $C > 0$ such that for each $P \in \mathcal{P}$*

$$\begin{aligned} & |\text{Regret}_{1,P}(\hat{w}^{\text{SCD}}) - \text{MER}_{1,P}(w_p^{\text{SCD}})| \\ & \leq C\bar{m}\mathbf{E}_P[\hat{D}_{1,1}] + C\mathbf{E}_P[\hat{D}_{1,2}^2] + \frac{C\bar{m}^2 K}{c} (\mathbf{E}_P[\hat{D}_{0,2}^2] + C\bar{m}\mathbf{E}_P[\hat{D}_{0,1}]) \\ & \quad + C\bar{m}\mathbf{E}_P[\hat{D}_{0,1}(0)] + C\mathbf{E}_P[\hat{D}_{0,2}^2(0)]. \end{aligned}$$

Proof: Note that

$$(B.2) \quad \left| \text{Regret}_{1,p}(\hat{w}^{\text{SCD}}) - \text{MER}_{1,p}(w_p^{\text{SCD}}) \right| \leq \left| \text{Regret}_{1,p}(\hat{w}^{\text{SCD}}) - \mathbb{E}_p [\text{MER}_{1,p}(\hat{w}^{\text{SCD}})] \right| + \left| \mathbb{E}_p [\text{MER}_{1,p}(\hat{w}^{\text{SCD}})] - \text{MER}_{1,p}(w_p^{\text{SCD}}) \right|.$$

As for the leading term on the right-hand side, by Lemmas B.3 and B.4(i),

$$\begin{aligned} \left| \text{Regret}_{1,p}(\hat{w}^{\text{SCD}}) - \mathbb{E}_p [\text{MER}_{1,p}(\hat{w}^{\text{SCD}})] \right| &\leq 2\mathbb{E}_p \left[\sup_{w \in \Delta_{K-1}} |R_{n,p}(w)| \right] \\ &\leq 2 \left(8\bar{m}\mathbb{E}_p [\hat{D}_{1,1}] + 2\mathbb{E}_p [\hat{D}_{1,2}^2] \right). \end{aligned}$$

It remains to deal with the last term in (B.2). First, we focus on $\text{MER}_{1,p}(\hat{w}^{\text{SCD}})$. Define

$$(B.3) \quad \text{Regret}_{0,p}^+(\hat{w}^{\text{SCD}}) = \mathbb{E}_p [\eta_{0,p}(\hat{w}^{\text{SCD}})] - \inf_{\tilde{w} \in \mathcal{D}} \mathbb{E}_p [\hat{\eta}_0(\tilde{w}(Z))].$$

We write

$$(B.4) \quad \begin{aligned} \text{MER}_{1,p}(\hat{w}^{\text{SCD}}) &= \text{MER}_{1,p}(\hat{w}^{\text{SCD}}) - \text{MER}_{0,p}(\hat{w}^{\text{SCD}}) + \text{MER}_{0,p}(\hat{w}^{\text{SCD}}) \\ &\quad - \text{Regret}_{0,p}^+(\hat{w}^{\text{SCD}}) + \text{Regret}_{0,p}^+(\hat{w}^{\text{SCD}}). \end{aligned}$$

We look into $\text{Regret}_{0,p}^+(\hat{w}^{\text{SCD}})$. Note that

$$(B.5) \quad \begin{aligned} \text{Regret}_{0,p}^+(\hat{w}^{\text{SCD}}) &\leq \mathbb{E}_p [\hat{\eta}_0(\hat{w}^{\text{SCD}})] - \inf_{\tilde{w} \in \mathcal{D}} \mathbb{E}_p [\hat{\eta}_0(\tilde{w}(Z))] + \mathbb{E}_p \left[\sup_{w \in \Delta_{K-1}} |\hat{\eta}_0(w) - \eta_{0,p}(w)| \right] \\ &= \mathbb{E}_p \left[\inf_{w \in \Delta_{K-1}} \hat{\eta}_0(w) \right] - \inf_{\tilde{w} \in \mathcal{D}} \mathbb{E}_p [\hat{\eta}_0(\tilde{w}(Z))] + \mathbb{E}_p \left[\sup_{w \in \Delta_{K-1}} |\hat{\eta}_0(w) - \eta_{0,p}(w)| \right] \\ &= \mathbb{E}_p \left[\sup_{w \in \Delta_{K-1}} |\hat{\eta}_0(w) - \eta_{0,p}(w)| \right], \end{aligned}$$

where the first equality uses the definition of $\hat{w}^{\text{SCD}}$ and the second equality follows by Lemma B.1. Since

$$\mathbb{E}_p [\text{MER}_{0,p}(\hat{w}^{\text{SCD}})] = \mathbb{E}_p [\eta_{0,p}(\hat{w}^{\text{SCD}})] - \inf_{\tilde{w} \in \mathcal{D}} \mathbb{E}_p [\eta_{0,p}(\tilde{w}(Z))],$$

by the definition of $\text{Regret}_{0,p}^+(\hat{w}^{\text{SCD}})$ in (B.3), we also have

$$(B.6) \quad \begin{aligned} \mathbb{E}_p [\text{MER}_{0,p}(\hat{w}^{\text{SCD}})] - \text{Regret}_{0,p}^+(\hat{w}^{\text{SCD}}) &= \inf_{\tilde{w} \in \mathcal{D}} \mathbb{E}_p [\hat{\eta}_0(\tilde{w}(Z))] - \inf_{\tilde{w} \in \mathcal{D}} \mathbb{E}_p [\eta_{0,p}(\tilde{w}(Z))] \\ &= \mathbb{E}_p \left[\inf_{w \in \Delta_{K-1}} \hat{\eta}_0(w) \right] - \inf_{w \in \Delta_{K-1}} \eta_{0,p}(w) \\ &\leq \mathbb{E}_p \left[\sup_{w \in \Delta_{K-1}} |\hat{\eta}_0(w) - \eta_{0,p}(w)| \right], \end{aligned}$$

where the second equality follows from Lemma B.1. Therefore, by (B.5) and (B.6),

$$|\mathbf{E}_P [\text{MER}_{0,P}(\hat{w}^{\text{SCD}})]| \leq 2\mathbf{E}_P \left[\sup_{w \in \Delta_{K-1}} |\hat{\eta}_0(w) - \eta_{0,P}(w)| \right].$$

Thus, as for the last term (B.2), noting that $\text{MER}_{0,P}(w_P^{\text{SCD}}) = 0$,

$$\begin{aligned} \text{(B.7)} \quad & |\mathbf{E}_P [\text{MER}_{1,P}(\hat{w}^{\text{SCD}})] - \text{MER}_{1,P}(w_P^{\text{SCD}})| \\ & \leq |\mathbf{E}_P [\text{MER}_{1,P}(\hat{w}^{\text{SCD}}) - \text{MER}_{0,P}(\hat{w}^{\text{SCD}})] - (\text{MER}_{1,P}(w_P^{\text{SCD}}) - \text{MER}_{0,P}(w_P^{\text{SCD}}))| \\ & \quad + 2\mathbf{E}_P \left[\sup_{w \in \Delta_{K-1}} |\hat{\eta}_0(w) - \eta_{0,P}(w)| \right]. \end{aligned}$$

Now note that

$$\begin{aligned} & \text{MER}_{1,P}(\hat{w}^{\text{SCD}}) - \text{MER}_{0,P}(\hat{w}^{\text{SCD}}) - (\text{MER}_{1,P}(w_P^{\text{SCD}}) - \text{MER}_{0,P}(w_P^{\text{SCD}})) \\ & = \eta_{1,P}(\hat{w}^{\text{SCD}}) - \eta_{0,P}(\hat{w}^{\text{SCD}}) - (\eta_{1,P}(w_P^{\text{SCD}}) - \eta_{0,P}(w_P^{\text{SCD}})). \end{aligned}$$

Hence,

$$\begin{aligned} & |\mathbf{E}_P [\text{MER}_{1,P}(\hat{w}^{\text{SCD}}) - \text{MER}_{0,P}(\hat{w}^{\text{SCD}})] - (\text{MER}_{1,P}(w_P^{\text{SCD}}) - \text{MER}_{0,P}(w_P^{\text{SCD}}))| \\ & \leq |\mathbf{E}_P [\eta_{1,P}(\hat{w}^{\text{SCD}}) - \eta_{1,P}(w_P^{\text{SCD}})] - \mathbf{E}_P [\eta_{0,P}(\hat{w}^{\text{SCD}}) - \eta_{0,P}(w_P^{\text{SCD}})]|. \end{aligned}$$

Note that

$$\begin{aligned} 0 & \leq \eta_{0,P}(\hat{w}^{\text{SCD}}) - \eta_{0,P}(w_P^{\text{SCD}}) \\ & = \eta_{0,P}(\hat{w}^{\text{SCD}}) - \hat{\eta}_0(\hat{w}^{\text{SCD}}) + \hat{\eta}_0(\hat{w}^{\text{SCD}}) - \eta_{0,P}(w_P^{\text{SCD}}) \\ & \leq \sup_{w \in \Delta_{K-1}} |\eta_{0,P}(w) - \hat{\eta}_0(w)| + \hat{\eta}_0(w_P^{\text{SCD}}) - \eta_{0,P}(w_P^{\text{SCD}}) \leq 2 \sup_{w \in \Delta_{K-1}} |\eta_{0,P}(w) - \hat{\eta}_0(w)|. \end{aligned}$$

Combining this with Lemma B.7, we find that

$$\begin{aligned} & |\mathbf{E}_P [\text{MER}_{1,P}(\hat{w}^{\text{SCD}}) - \text{MER}_{0,P}(\hat{w}^{\text{SCD}})] - (\text{MER}_{1,P}(w_P^{\text{SCD}}) - \text{MER}_{0,P}(w_P^{\text{SCD}}))| \\ & \leq \frac{64\bar{m}^2 K}{c} (\mathbf{E}_P [\hat{D}_{0,2}^2] + 4\bar{m}\mathbf{E}_P [\hat{D}_{0,1}]) + 2\mathbf{E}_P \left[\sup_{w \in \Delta_{K-1}} |\eta_{0,P}(w) - \hat{\eta}_0(w)| \right] \\ & \leq \frac{64\bar{m}^2 K}{c} (\mathbf{E}_P [\hat{D}_{0,2}^2] + 4\bar{m}\mathbf{E}_P [\hat{D}_{0,1}]) + 16\bar{m}\mathbf{E}_P [\hat{D}_{0,1}(0)] + 4\mathbf{E}_P [\hat{D}_{0,2}^2(0)]. \end{aligned}$$

The last inequality follows by Lemma B.4. In light of (B.7), this yields the desired result. ■

Lemma B.9. *There exists a universal constant $C > 0$ such that for $d = 0, 1$ and each $P \in \mathcal{P}$,*

$$\begin{aligned} \max \{ \mathbf{E}_P [\hat{D}_{d,1}], \mathbf{E}_P [\hat{D}_{d,1}(0)] \} & \leq C\bar{m} \sum_{j=0}^K \inf_{v>0} \left\{ \frac{1}{v\sqrt{n}} + \exp \left(-\frac{n(\tilde{p}_j - v)^2}{2\tilde{p}_j} \right) \right\}, \text{ and} \\ \max \{ \mathbf{E}_P [\hat{D}_{d,2}^2], \mathbf{E}_P [\hat{D}_{d,2}^2(0)] \} & \leq C\bar{m}^2 \sum_{j=0}^K \inf_{v>0} \left\{ \frac{1}{v^2 n} + \exp \left(-\frac{n(\tilde{p}_j - v)^2}{2\tilde{p}_j} \right) \right\}, \end{aligned}$$

where $\tilde{p}_j = \frac{1}{n} \sum_{i \in N} P\{G_i = j\}$.

Proof: We focus on $\hat{D}_{d,1}$. Note that

$$\mathbf{E}_P \left[\left| \hat{m}_{j,t} - m_{j,t} \right| \right] = A_{n,1} + A_{n,2},$$

where, with $\hat{p}_j = \frac{1}{n} \sum_{i \in N} 1\{G_i = j\}$, we define

$$A_{n,1} = \mathbf{E}_P \left[\left| \hat{m}_{j,t} - m_{j,t} \right| 1\{\hat{p}_j \geq \nu\} \right], \text{ and}$$

$$A_{n,2} = \mathbf{E}_P \left[\left| \hat{m}_{j,t} - m_{j,t} \right| 1\{\hat{p}_j < \nu\} \right].$$

We define $\varepsilon_{ij,t} = Y_{i,t} - \mathbf{E}_P[Y_{i,t} \mid G_i = j]$, and, using $\hat{m}_{j,t} - m_{j,t} = \frac{1}{n} \sum_{i \in N} (Y_{i,t} - \mu_{j,t}) 1\{G_i = j\} / \hat{p}_j$, bound

$$A_{n,1} \leq \frac{1}{\nu} \mathbf{E}_P \left[\left| \frac{1}{n} \sum_{i \in N} \varepsilon_{ij,t} 1\{G_i = j\} \right| \right] \leq \frac{\bar{m}}{\nu \sqrt{n}},$$

where the last inequality follows because

$$\begin{aligned} \mathbf{E}_P \left[\left(\frac{1}{n} \sum_{i \in N} \varepsilon_{ij,t} 1\{G_i = j\} \right)^2 \right] &\leq \frac{1}{n^2} \sum_{i \in N} \mathbf{E}_P \left[\varepsilon_{ij,t}^2 1\{G_i = j\} \right] \\ &\leq \frac{1}{n^2} \sum_{i \in N} \mathbf{E}_P \left[Y_{i,t}^2 1\{G_i = j\} \right] \leq \frac{\bar{m}^2}{n}. \end{aligned}$$

(Note that $\varepsilon_{ij,t} 1\{G_i = j\}$ have mean zero and are independent across i 's by Assumption 4.1.)

Let us turn to $A_{n,2}$. Note that

$$\begin{aligned} A_{n,2} &\leq \mathbf{E}_P \left[\frac{\sum_{i \in N} |\varepsilon_{ij,t}| 1\{G_i = j\}}{\sum_{i \in N} 1\{G_i = j\}} 1\{\hat{p}_j < \nu\} \right] \\ &\leq \mathbf{E}_P \left[\frac{\sum_{i \in N} \mathbf{E}_P \left[|\varepsilon_{ij,t}| \mid G_1, \dots, G_n \right] 1\{G_i = j\}}{\sum_{i \in N} 1\{G_i = j\}} 1\{\hat{p}_j < \nu\} \right] \\ &\leq \mathbf{E}_P \left[\frac{\sum_{i \in N} \mathbf{E}_P \left[|\varepsilon_{ij,t}| \mid G_i \right] 1\{G_i = j\}}{\sum_{i \in N} 1\{G_i = j\}} 1\{\hat{p}_j < \nu\} \right] \leq \bar{m} P\{\hat{p}_j < \nu\}, \end{aligned}$$

because $\mathbf{E}_P[\varepsilon_{ij,t}^2 \mid G_i] \leq \bar{m}^2$. Now, note that by Chernoff's bound (see, e.g., Lemma 2.1 of Chung and Lu (2002)),

$$\begin{aligned} P\{\hat{p}_j < \nu\} &= P \left\{ \sum_{i \in N} 1\{G_i = j\} - n\tilde{p}_j < n\nu - n\tilde{p}_j \right\} \\ &\leq \exp \left(-\frac{n(\tilde{p}_j - \nu)^2}{2\tilde{p}_j} \right). \end{aligned}$$

Therefore, we have

$$\mathbf{E}_P \left[\left| \hat{m}_{j,t} - m_{j,t} \right| \right] \leq \bar{m} \inf_{v>0} \left(\frac{1}{v\sqrt{n}} + \exp \left(-\frac{n(\tilde{p}_j - v)^2}{2\tilde{p}_j} \right) \right).$$

Hence,

$$\mathbf{E}_P \left[\left| \hat{\mu}_{j,t}(\lambda) - \mu_{j,t}(\lambda) \right| \right] \leq 2\bar{m} \inf_{v>0} \left(\frac{1}{v\sqrt{n}} + \exp \left(-\frac{n(\tilde{p}_j - v)^2}{2\tilde{p}_j} \right) \right).$$

Using the same arguments, we obtain the same bound for $\mathbf{E}_P \left[\left| \hat{\mu}_{j,t}(0; \lambda) - \mu_{j,t}(0; \lambda) \right| \right]$. This gives the first bound of the lemma.

As for the second bound, we consider

$$\mathbf{E}_P \left[\left(\hat{m}_{j,t} - m_{j,t} \right)^2 \right] = B_{n,t}^a + B_{n,t}^b,$$

where

$$\begin{aligned} B_{n,t}^a &= \mathbf{E}_P \left[\left(\hat{m}_{j,t} - m_{j,t} \right)^2 \mathbf{1}\{\hat{p}_j \geq v\} \right], \text{ and} \\ B_{n,t}^b &= \mathbf{E}_P \left[\left(\hat{m}_{j,t} - m_{j,t} \right)^2 \mathbf{1}\{\hat{p}_j < v\} \right]. \end{aligned}$$

Similarly as before, note that

$$B_{n,t}^a \leq \frac{1}{v^2 n^2} \sum_{i \in N} \mathbf{E}_P \left[\varepsilon_{ij,t}^2 \mathbf{1}\{G_i = j\} \right] \leq \frac{\bar{m}^2}{v^2 n},$$

and

$$B_{n,t}^b \leq 4\bar{m}^2 P\{\hat{p}_j < v\} \leq 4\bar{m}^2 \exp \left(-\frac{n(\tilde{p}_j - v)^2}{2\tilde{p}_j} \right).$$

Hence,

$$\begin{aligned} \mathbf{E}_P \left[\left(\hat{\mu}_{j,t}(\lambda) - \mu_{j,t}(\lambda) \right)^2 \right] &\leq 2(B_{n,t}^a + B_{n,t}^b) + 2 \sum_{s=1}^{T^*-1} (B_{n,s}^a + B_{n,s}^b) \lambda_s \\ &\leq 2(B_{n,t}^a + B_{n,t}^b) + 2 \max_{s=1, \dots, T^*-1} (B_{n,s}^a + B_{n,s}^b) \\ &\leq 16\bar{m}^2 \left\{ \frac{1}{v^2 n} + \exp \left(-\frac{n(\tilde{p}_j - v)^2}{2\tilde{p}_j} \right) \right\}, \end{aligned}$$

where the first inequality follows by Jensen's inequality. Using the same arguments, we obtain the same bound for $\mathbf{E}_P \left[\left(\hat{\mu}_{j,t}(0; \lambda) - \mu_{j,t}(0; \lambda) \right)^2 \right]$. This gives the second bound of the lemma. $\blacksquare$

Proposition B.1. *There exists a universal constant $C > 0$ such that*

$$\sup_{P \in \mathcal{P}} \left| \text{Regret}_{1,P}(\hat{w}^{\text{SCD}}) - \Delta \text{MER}_P(w_P^{\text{SCD}}) \right| \leq \frac{C(K+1)\bar{m}^4}{c} \left\{ \frac{1}{\pi_0 \sqrt{n}} + \frac{1}{\pi_0^2 n} + \exp \left(-\frac{\pi_0 n}{8} \right) \right\}.$$

Proof: Since we can take $\nu = \tilde{p}_j/2$ in the bound in Lemma B.9, we have

$$\begin{aligned} \inf_{\nu>0} \left\{ \frac{1}{\nu\sqrt{n}} + \exp\left(-\frac{n(\tilde{p}_j - \nu)^2}{2\tilde{p}_j}\right) \right\} &\leq \frac{2}{\tilde{p}_j\sqrt{n}} + \exp\left(-\frac{n\tilde{p}_j}{8}\right) \\ &\leq \frac{2}{\pi_0\sqrt{n}} + \exp\left(-\frac{\pi_0 n}{8}\right) =: A_n, \end{aligned}$$

where the last inequality uses $\tilde{p}_j \geq \pi_0$. Similarly, we have

$$\begin{aligned} \inf_{\nu>0} \left\{ \frac{1}{\nu^2 n} + \exp\left(-\frac{n(\tilde{p}_j - \nu)^2}{2\tilde{p}_j}\right) \right\} &\leq \frac{4}{\tilde{p}_j^2 n} + \exp\left(-\frac{n\tilde{p}_j}{8}\right) \\ &\leq \frac{4}{\pi_0^2 n} + \exp\left(-\frac{\pi_0 n}{8}\right) =: B_n. \end{aligned}$$

By Lemmas B.8 and B.9 and the bounds above,

$$\begin{aligned} \sup_{p \in \mathcal{P}} |\text{Regret}_{1,p}(\hat{w}^{\text{SCD}}) - \text{MER}_{1,p}(w_p^{\text{SCD}})| &\leq C \left(\bar{m} + \frac{\bar{m}^3 K}{c} \right) \bar{m} \cdot A_n \\ &\quad + C \left(1 + \frac{\bar{m}^2 K}{c} \right) \bar{m}^2 \cdot B_n, \end{aligned}$$

with some universal constant $C > 0$. The desired result follows because $\text{MER}_{0,p}(w_p^{\text{SCD}}) = 0$ and $\bar{m} \geq 1$ and $c \leq 1$. ■

Lemma B.10. Suppose that Assumption 4.1 holds. Then,

$$\sup_{p \in \mathcal{P}} \max_{1 \leq j \leq K} \mathbf{E}_p \left[\left| \hat{w}_j^{\text{DID}} - w_{j,p}^{\text{DID}} \right| \right] \leq \frac{2}{\pi_0\sqrt{n}} + \exp\left(-\frac{\pi_0 n}{8}\right).$$

Proof: Let $u_{ij} = 1\{G_i = j\} - P\{G_i = j \mid G_i \in \mathcal{G}_{\text{don}}\}$. We write

$$\hat{w}_j^{\text{DID}} - w_{j,p}^{\text{DID}} = \frac{\sum_{i \in N} u_{ij} 1\{G_i \in \mathcal{G}_{\text{don}}\}}{\sum_{i \in N} 1\{G_i \in \mathcal{G}_{\text{don}}\}}.$$

Hence, with $\hat{p} := \frac{1}{n} \sum_{i \in N} 1\{G_i \in \mathcal{G}_{\text{don}}\}$, and arbitrarily chosen $\nu > 0$, we have

$$\mathbf{E}_p \left[\left| \hat{w}_j^{\text{DID}} - w_{j,p}^{\text{DID}} \right| \right] \leq C_{n,1} + C_{n,2},$$

where

$$\begin{aligned} C_{n,1} &= \mathbf{E}_p \left[\left| \frac{\sum_{i \in N} u_{ij} 1\{G_i \in \mathcal{G}_{\text{don}}\}}{\sum_{i \in N} 1\{G_i \in \mathcal{G}_{\text{don}}\}} \right| 1\{\hat{p} \geq \nu\} \right] \text{ and} \\ C_{n,2} &= \mathbf{E}_p \left[\left| \frac{\sum_{i \in N} u_{ij} 1\{G_i \in \mathcal{G}_{\text{don}}\}}{\sum_{i \in N} 1\{G_i \in \mathcal{G}_{\text{don}}\}} \right| 1\{\hat{p} < \nu\} \right]. \end{aligned}$$

Using the same arguments as before, we find that

$$C_{n,1} \leq \frac{1}{\nu\sqrt{n}},$$

because $|u_{ij}| \leq 1$, and that

$$C_{n,2} \leq \exp\left(-\frac{n(p-\nu)^2}{2p}\right),$$

where $p := \frac{1}{n} \sum_{i \in N} P\{G_i \in \mathcal{G}_{\text{don}}\}$. By taking $\nu = p/2$ and noting $p \geq \pi_0$, we find that

$$\sup_{P \in \mathcal{P}} \mathbf{E}_P \left[\left| \hat{w}_j^{\text{DID}} - w_{j,P}^{\text{DID}} \right| \right] \leq \frac{2}{\pi_0 \sqrt{n}} + \exp\left(-\frac{\pi_0 n}{8}\right).$$

■

Proposition B.2. *There exists a universal constant $C > 0$ such that*

$$\sup_{P \in \mathcal{P}} \left| \text{Regret}_{1,p}(\hat{w}^{\text{DID}}) - \text{MER}_{1,p}(w_p^{\text{DID}}) \right| \leq C \bar{m}^2 (K+1) \left\{ \frac{1}{\pi_0 \sqrt{n}} + \frac{1}{\pi_0^2 n} + \exp\left(-\frac{\pi_0 n}{8}\right) \right\}.$$

Proof: By Lemmas B.3, B.4(i), and B.9 with $\nu = \pi_0/2$ in the bound, we have

$$\begin{aligned} \sup_{P \in \mathcal{P}} \left| \text{Regret}_{1,p}(\hat{w}^{\text{DID}}) - \mathbf{E}_P [\text{MER}_{1,p}(\hat{w}^{\text{DID}})] \right| &\leq 2 \sup_{P \in \mathcal{P}} \mathbf{E}_P \left[\sup_{w \in \Delta_{K-1}} |R_{n,p}(w)| \right] \\ &\leq C' \bar{m}^2 (K+1) \left\{ \frac{1}{\pi_0 \sqrt{n}} + \frac{1}{\pi_0^2 n} + \exp\left(-\frac{\pi_0 n}{8}\right) \right\}, \end{aligned}$$

for some universal constant C' . We bound $\mathbf{E}_P [\text{MER}_{1,p}(\hat{w}^{\text{DID}})] - \text{MER}_{1,p}(w_p^{\text{DID}})$ as follows:

$$\begin{aligned} \text{(B.8)} \quad &\mathbf{E}_P [\eta_{1,p}(\hat{w}^{\text{DID}}) - \eta_{1,p}(w_p^{\text{DID}})] \\ &\leq \frac{1}{|\mathcal{T}_1|} \sum_{t \in \mathcal{T}_1} \mathbf{E}_P \left[|e_t(\lambda, \hat{w}^{\text{DID}}) + e_t(\lambda, w_p^{\text{DID}})| |e_t(\lambda, \hat{w}^{\text{DID}}) - e_t(\lambda, w_p^{\text{DID}})| \right]. \end{aligned}$$

Note that

$$\begin{aligned} |e_t(\lambda, \hat{w}^{\text{DID}}) - e_t(\lambda, w_p^{\text{DID}})| &\leq \sum_{j=1}^K |\mu_{j,t}(0; \lambda)| |\hat{w}_j^{\text{DID}} - w_{j,p}^{\text{DID}}| \\ &\leq 2\bar{m} \sum_{j=1}^K |\hat{w}_j^{\text{DID}} - w_{j,p}^{\text{DID}}|. \end{aligned}$$

On the other hand,

$$|e_t(\lambda, \hat{w}^{\text{DID}}) + e_t(\lambda, w_p^{\text{DID}})| \leq 2|\mu_{0,t}(0; \lambda)| + \sum_{j=1}^K |\mu_{j,t}(0; \lambda)| (\hat{w}_j^{\text{DID}} + w_{j,p}^{\text{DID}}) \leq 8\bar{m}.$$

Hence, the last term in (B.8) is bounded by

$$16\bar{m}^2 \sup_{P \in \mathcal{P}} \sum_{j=1}^K \mathbf{E}_P \left[\left| \hat{w}_j^{\text{DID}} - w_{j,P}^{\text{DID}} \right| \right] \leq 16\bar{m}^2 K \left\{ \frac{2}{\pi_0 \sqrt{n}} + \exp\left(-\frac{\pi_0 n}{8}\right) \right\},$$

by Lemma B.10. Thus, since $\bar{m} \geq 1$, we have the desired result. ■

Proof of Theorem 4.1: The desired result follows from Propositions B.1 and B.2. ■

Appendix C. Proofs of the Results in Section 5

Proof of Theorem 5.1: The first result follows from Lemmas B.6 and B.9. We can see that the second result follows from the first result using the standard arguments. Details are omitted. ■

Theorem C.1. Suppose that Assumptions 4.1 and 5.1, and (4.6) in Assumption 4.2 hold. Then, for any $\alpha \in (0, 1)$, as $n \rightarrow \infty$, we have

$$\liminf_{n \rightarrow \infty} \inf_{P \in \mathcal{P}} P \left\{ w^*(\lambda) \in \tilde{C}_{1-\alpha} \right\} \geq 1 - \alpha.$$

Proof: For any vector $x = (x_k)_{k=1}^K \in \mathbf{R}^K$, we write $J_0[x] = \{1 \leq k \leq K : x_k = 0\}$. We define

$$\begin{aligned} \Lambda(w) &= \{B_2' \lambda \in \mathbf{R}^{K-1} : w' \lambda = 0, \lambda \geq 0\} \text{ and} \\ \Lambda^\circ(w, \hat{V}(w)) &= \{x \in \mathbf{R}^{K-1} : [B_2 \hat{V}^{-1}(w)x]_{J_0[w]} \leq 0\}. \end{aligned}$$

It is not hard to see that $\Lambda(w)$ is a polyhedral cone and $\Lambda^\circ(w, \hat{V}(w))$ its polar cone along $\|\cdot\|_{\hat{V}(w)}$, where $\|x\|_{\hat{V}(w)}^2 = x' \hat{V}^{-1}(w)x$. We let $Y_n(w) = \sqrt{n} B_2' \hat{\varphi}(w)$. For any vector $y \in \mathbf{R}^{K-1}$ and a closed convex subset $C \subset \mathbf{R}^{K-1}$, the projection of y onto C along $\|\cdot\|_{\hat{V}(w)}$ is denoted by $\Pi_{\hat{V}(w)}(y | C)$. Then, we can write $\hat{d}(w)$ as follows:

$$\begin{aligned} \hat{d}(w) &= |J_0[B_2 \hat{V}^{-1}(w) B_2' (\hat{\varphi}(w) - \hat{\lambda}(w))]| \\ &= |J_0[B_2 \hat{V}^{-1}(w) (Y_n(w) - \Pi_{\hat{V}(w)}(Y_n(w) | \Lambda(w)))]| \\ &= |J_0[B_2 \hat{V}^{-1}(w) \Pi_{\hat{V}(w)}(Y_n(w) | \Lambda^\circ(w, \hat{V}(w)))]|. \end{aligned}$$

It suffices to show that for each sequence $P_n \in \mathcal{P}$ and each sequence $w_n \in \mathbb{W}_{P_n}$,

$$\lim_{n \rightarrow \infty} P_n \left\{ T_{\text{bf}}(w_n) > \hat{c}_{1-\kappa, \text{bf}}(w_n) \right\} \leq \kappa.$$

We apply Lemma 3.1 of Canen and Song (2025) by setting $L = 0$,

$$Y_n = Y_n(w_n), \quad \hat{\Omega}_n = \hat{V}(w_n), \quad \Omega_n = V_{P_n}(w_n), \quad \text{and} \quad \mu_n(w_n) = \sqrt{n} B_2' \varphi_{P_n}(w_n).$$

For this, we check Assumption 3.1 of Canen and Song (2025). First, note that

$$(C.1) \quad \hat{V}(w_n) - V_{P_n}(w_n) = o_P(1),$$

by Assumption 4.1 and the Law of Large Numbers, and

$$Z_n := \hat{\Omega}_n^{-1/2}(Y_n - \mu_n) = \hat{V}^{-1/2}(w_n) \sqrt{n} B'_2(\hat{\varphi}(w_n) - \varphi_{P_n}(w_n)) \rightarrow_d N(0, I_{K-1}),$$

as $n \rightarrow \infty$, by (C.1) and the Central Limit Theorem applied to independent random variables, together with the condition (4.6) in Assumption 4.2. Furthermore, by Assumption 5.1, for some constants $C, c > 0$,

$$\lambda_{\min}(V_{P_n}(w_n)) > c \text{ and } \|V_{P_n}(w_n)\| < C.$$

Thus, Assumption 3.1 in Canen and Song (2025) is satisfied, and the desired result follows from their Lemma 3.1. ■

Proof of Theorem 5.2: We consider the first statement. By the SMC at λ , $\theta_t(\lambda, w^*(\lambda)) = \theta_t^*$. Let

$$\tilde{\psi}_{ij,t} = \frac{n}{n_t} \frac{1\{G_{i,t} = j\}}{\hat{p}_{j,t}} (y_{i,t} - \mu_{j,t}).$$

Recall $\hat{p}_{j,t} = n_{j,t}/n_t$. Note that

$$\begin{aligned} (C.2) \quad \sqrt{n}(\hat{\theta}_t(w) - \theta_t^*) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\tilde{\psi}_{i0,t} - \sum_{j=1}^K \tilde{\psi}_{ij,t} w_j \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\psi_{i0,t} - \sum_{j=1}^K \psi_{ij,t} w_j \right) + R_n(w), \end{aligned}$$

where

$$\begin{aligned} R_n(w) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{n}{n_t} 1\{G_{i,t} = 0\} \left(\frac{1}{\hat{p}_{0,t}} - \frac{1}{p_{0,t}} \right) (y_{i,t} - \mu_{0,t}) \\ &\quad + \sum_{j=1}^K \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{n}{n_t} 1\{G_{i,t} = j\} \left(\frac{1}{\hat{p}_{j,t}} - \frac{1}{p_{j,t}} \right) (y_{i,t} - \mu_{j,t}) w_j. \end{aligned}$$

Using standard arguments, we can show that

$$\sup_{w \in \Delta_{K-1}} |R_n(w)| = o_p(1),$$

as $n \rightarrow \infty$. Using similar arguments, we can show that

$$\sup_{w \in \Delta_{K-1}} |\hat{\sigma}_t^2(w) - \sigma_t^2(w)| = o_p(1),$$

where $\sigma^2(w) = \frac{1}{n} \sum_{i=1}^n \mathbf{E}_{P_n} \left[(\psi_{i0,t} - \sum_{j=1}^K \psi_{ij,t} w_j)^2 \right]$.

Thus, we find that

$$(C.3) \quad \sup_{w \in \Delta_{K-1}} P_n \left\{ \left| \frac{\sqrt{n}(\hat{\theta}_t(w) - \theta_t^*)}{\hat{\sigma}_t(w)} \right| > z_{1-\beta(\alpha, \kappa)} \right\} \leq \alpha - \kappa + o(1),$$

by the Central Limit Theorem applied to the asymptotic linear representation in (C.2).

First, take any sequence $P_n \in \mathcal{P}$. Note that

$$\begin{aligned}
P_n\{\theta_t^* \in C_{1-\alpha}^{\text{bf}}\} &= P_n \left\{ \inf_{w \in \tilde{C}_{1-\kappa}} \left| \frac{\sqrt{n}(\hat{\theta}_t(w) - \theta_t^*)}{\hat{\sigma}_t(w)} \right| \leq z_{1-\beta(\alpha, \kappa)} \right\} \\
&\geq \inf_{w \in \mathbb{W}_{P_n}} P_n \left\{ w \in \tilde{C}_{1-\kappa}, \left| \frac{\sqrt{n}(\hat{\theta}_t(w) - \theta_t^*)}{\hat{\sigma}_t(w)} \right| \leq z_{1-\beta(\alpha, \kappa)} \right\} \\
&\geq 1 - \sup_{w \in \mathbb{W}_{P_n}} P_n \{w \notin \tilde{C}_{1-\kappa}\} - \sup_{w \in \mathbb{W}_{P_n}} P_n \left\{ \left| \frac{\sqrt{n}(\hat{\theta}_t(w) - \theta_t^*)}{\hat{\sigma}_t(w)} \right| > z_{1-\beta(\alpha, \kappa)} \right\} \\
&\geq 1 - \kappa - (\alpha - \kappa) + o(1) = 1 - \alpha + o(1),
\end{aligned}$$

by Theorem C.1 and (C.3). ■

Appendix D. Additional Simulation Results

TABLE D.1. Results for Monte Carlo Simulations with Different Group Sizes.

Parameters			MAD		Coverage Probability		CI Length	
K	T	n	SCD	CSDID	SCD	CSDID	SCD	CSDID
<i>Scenario A: PTA and SMC hold</i>								
10	50	1500	0.088	0.147	1.000	0.999	1.552	1.218
10	50	3000	0.062	0.106	1.000	0.998	1.030	0.868
10	100	1500	0.082	0.148	1.000	1.000	1.475	1.284
10	100	3000	0.059	0.109	0.999	0.997	0.971	0.917
30	50	1500	0.139	0.235	1.000	0.997	2.497	1.801
30	50	3000	0.094	0.172	1.000	0.997	1.865	1.305
30	100	1500	0.131	0.226	1.000	0.999	2.316	1.893
30	100	3000	0.093	0.160	1.000	0.999	1.718	1.366
<i>Scenario B: PTA fails but SMC holds</i>								
10	50	1500	0.089	1.623	0.998	-	0.985	-
10	50	3000	0.065	1.625	0.997	-	0.672	-
10	100	1500	0.085	1.596	0.996	-	0.963	-
10	100	3000	0.062	1.592	0.995	-	0.657	-
30	50	1500	0.142	1.981	0.999	-	1.605	-
30	50	3000	0.098	1.964	0.996	-	1.110	-
30	100	1500	0.133	1.890	0.997	-	1.434	-
30	100	3000	0.093	1.886	0.994	-	0.964	-
<i>Scenario C: PTA holds but SMC fails</i>								
10	50	1500	1.039	0.147	-	0.999	-	1.218
10	50	3000	1.034	0.106	-	0.998	-	0.869
10	100	1500	1.009	0.148	-	1.000	-	1.284
10	100	3000	1.009	0.109	-	0.997	-	0.917
30	50	1500	1.282	0.235	-	0.997	-	1.802
30	50	3000	1.294	0.172	-	0.997	-	1.306
30	100	1500	1.162	0.226	-	0.999	-	1.894
30	100	3000	1.149	0.160	-	0.999	-	1.366

Notes: This table presents the simulation results for Scenarios A, B, and C described in subsection 5.4, when there is one large group in the donor pool. More specifically, we set $p = [0.8/K, \dots, 0.8/K, 0.2]'$ for $K = 10$, and $p = [0.925/K, \dots, 0.925/K, 0.075]'$ for $K = 30$. The table reports Mean Absolute Deviation (MAD), Coverage Probability, and Confidence Interval (CI) Length for our proposed method of Synthetic Control with Differencing (SCD) and the Difference-in-Differences estimator proposed by Callaway and Sant'Anna (2021) (CSDID). Inference for SCD is conducted using Algorithm 1. Inference results are only reported for cases where the target parameter is identified. The number of MC simulations is 1,000.

TABLE D.2. Results for Monte Carlo Simulations with Different Time Factors.

Parameters			MAD		Coverage Probability		CI Length	
K	T	n	SCD	CSDID	SCD	CSDID	SCD	CSDID
<i>Scenario A: PTA and SMC hold</i>								
10	50	1500	0.079	0.147	0.999	0.998	1.472	1.166
10	50	3000	0.059	0.104	1.000	0.999	0.964	0.829
10	100	1500	0.079	0.147	1.000	0.999	1.403	1.232
10	100	3000	0.054	0.100	1.000	0.998	0.905	0.872
30	50	1500	0.128	0.229	1.000	0.999	2.449	1.766
30	50	3000	0.094	0.163	1.000	1.000	1.830	1.275
30	100	1500	0.130	0.228	0.999	0.998	2.277	1.867
30	100	3000	0.090	0.157	1.000	0.999	1.721	1.351
<i>Scenario B: PTA fails but SMC holds</i>								
10	50	1500	0.091	1.914	1.000	-	1.003	-
10	50	3000	0.068	1.908	0.996	-	0.683	-
10	100	1500	0.095	1.861	0.995	-	0.969	-
10	100	3000	0.062	1.858	0.997	-	0.654	-
30	50	1500	0.138	2.082	1.000	-	1.593	-
30	50	3000	0.103	2.074	0.997	-	1.069	-
30	100	1500	0.139	2.005	0.994	-	1.383	-
30	100	3000	0.096	1.982	0.991	-	0.926	-
<i>Scenario C: PTA holds but SMC fails</i>								
10	50	1500	1.173	0.147	-	0.998	-	1.165
10	50	3000	1.176	0.104	-	0.999	-	0.830
10	100	1500	1.179	0.147	-	0.999	-	1.233
10	100	3000	1.186	0.100	-	0.999	-	0.872
30	50	1500	1.358	0.229	-	0.999	-	1.770
30	50	3000	1.399	0.163	-	1.000	-	1.275
30	100	1500	1.332	0.228	-	0.998	-	1.867
30	100	3000	1.354	0.157	-	0.999	-	1.352

Notes: This table presents the simulation results for Scenarios A, B, and C described in subsection 5.4, modifying the distribution of time factors as follows: $F_t \sim N(\xi\sqrt{t}, 0.5^2 \cdot I_3)$, where $\xi = [0.01, 0.02, 0.04]'$. The table reports Mean Absolute Deviation (MAD), Coverage Probability, and Confidence Interval (CI) Length for our proposed method of Synthetic Control with Differencing (SCD) and the Difference-in-Differences estimator proposed by Callaway and Sant'Anna (2021) (CSDID). Inference for SCD is conducted using Algorithm 1. Inference results are only reported for cases where the target parameter is identified. The number of MC simulations is 1,000.

TABLE D.3. Results for Monte Carlo Simulations with Different Factor Loadings.

Parameters			MAD		Coverage Probability		CI Length	
K	T	n	SCD	CSDID	SCD	CSDID	SCD	CSDID
<i>Scenario A: PTA and SMC hold</i>								
10	50	1500	0.079	0.158	0.999	1.000	1.810	1.274
10	50	3000	0.059	0.116	1.000	0.998	1.224	0.907
10	100	1500	0.079	0.160	1.000	1.000	1.731	1.342
10	100	3000	0.054	0.112	1.000	1.000	1.165	0.955
30	50	1500	0.128	0.260	1.000	0.998	2.965	1.954
30	50	3000	0.094	0.178	1.000	1.000	2.245	1.414
30	100	1500	0.130	0.249	1.000	0.998	2.795	2.069
30	100	3000	0.090	0.175	1.000	0.999	2.130	1.488
<i>Scenario B: PTA fails but SMC holds</i>								
10	50	1500	0.094	1.918	0.999	-	1.202	-
10	50	3000	0.070	1.906	1.000	-	0.833	-
10	100	1500	0.093	1.863	0.998	-	1.165	-
10	100	3000	0.065	1.858	0.997	-	0.804	-
30	50	1500	0.141	2.084	1.000	-	1.952	-
30	50	3000	0.101	2.074	0.999	-	1.350	-
30	100	1500	0.138	2.009	0.996	-	1.728	-
30	100	3000	0.098	1.980	0.999	-	1.192	-
<i>Scenario C: PTA holds but SMC fails</i>								
10	50	1500	1.159	0.158	-	1.000	-	1.276
10	50	3000	1.155	0.116	-	0.998	-	0.909
10	100	1500	1.134	0.160	-	1.000	-	1.342
10	100	3000	1.128	0.112	-	1.000	-	0.955
30	50	1500	1.339	0.260	-	0.998	-	1.956
30	50	3000	1.341	0.178	-	1.000	-	1.414
30	100	1500	1.230	0.249	-	0.998	-	2.070
30	100	3000	1.213	0.175	-	0.999	-	1.490

Notes: This table presents the simulation results for Scenarios A, B, and C described in subsection 5.4, modifying the distribution of individual factor loadings as follows: $\Lambda_i|G_i \sim t_5(m_{G_i}, I_3)$. The table reports Mean Absolute Deviation (MAD), Coverage Probability, and Confidence Interval (CI) Length for our proposed method of Synthetic Control with Differencing (SCD) and the Difference-in-Differences estimator proposed by [Callaway and Sant'Anna \(2021\)](#) (CSDID). Inference for SCD is conducted using Algorithm 1. Inference results are only reported for cases where the target parameter is identified. The number of MC simulations is 1,000.