Synthetic Control with Disaggregated Data Settings

Lea Bottmer *

Stanford University †

[Click here for latest version]

October 4, 2025

Abstract

The synthetic control estimator is widely used to evaluate aggregate-level policies, but researchers increasingly face settings with rich, disaggregated data (e.g., county-level outcomes within states). Existing approaches incorporate disaggregated data by estimating separate synthetic controls for each disaggregated treated unit, enlarging the donor pool with disaggregated control units, or both. While such strategies can improve fit, they also amplify noise, and there is little guidance on how to navigate these trade-offs and the choice of aggregation. This paper develops a general framework for synthetic control with disaggregated data that nests the classical synthetic control estimator and other existing approaches. Within this framework, I propose a multi-level SC (mlSC) estimator that formalizes the choice of aggregation levels as a data-driven regularization problem. The estimator regularizes flexibly toward the classical synthetic control solution while exploiting variation contained in the disaggregated data. In simulations calibrated to four empirical settings, mlSC consistently outperforms or matches existing approaches. Two applications—Minnesota's e-cigarette tax and minimum wage effects on teen employment—illustrate its practical value.

*Email: lbottmer@stanford.edu

†I thank Guido Imbens and Jann Spiess for their mentorship. I thank Dmitry Arkhangelsky, Matt Brown, Dante Domenella, Avi Feller, Lihua Lei, Helena Roy, Jason Weitze and seminar participants at Stanford University for their valuable comments and suggestions. All errors are my own. *Disclaimer:* Researchers' own analyses calculated (or derived) based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

1 Introduction

Synthetic control (SC) estimators have become a central tool for policy evaluation, particularly for assessing the impact of *aggregate-level* interventions, like state-wide policies, on a single, *aggregated* unit, like a state (see Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015). Athey and Imbens (2019) crown the SC estimator to be "arguably the most important innovation in the evaluation literature in the last fifteen years." Its appeal lies in constructing a simple and interpretable counterfactual for the treated unit. The counterfactual is a convex combination of units from a donor pool of similar, untreated units that approximates the treated unit's outcome path in the absence of the intervention. This approach is especially valuable in panel data settings with a small number of aggregated units, where traditional methods that rely on asymptotics for a large number of units are less reliable (Doudchenko and Imbens, 2016).

Yet little guidance exists on whether and how the SC estimator should incorporate data measured at levels below the treatment assignment, hereafter termed *disaggregated* data. For instance, state-wide policies are commonly studied using county, metropolitan, or even individual-level data (see, e.g., Card and Krueger, 1994; Dube and Zipperer, 2015; Deng and Zheng, 2023).[1] In these settings, researchers must decide whether to apply SC to aggregated outcomes, to disaggregated outcomes, or to a combination of both, while still targeting the treatment effect at the aggregate level. This choice is consequential. Using disaggregated data can change estimated effects and, in turn, the economic conclusions drawn from a study. It can also improve precision by exploiting additional cross-sectional variation. This advantage is magnified in typical classical SC settings, where the small number of time periods and aggregated units means that the gains from incorporating disaggregated data can be substantial. At the same time, disaggregation also introduces risks of overfitting and non-uniqueness when the donor pool greatly exceeds the number of pre-treatment periods (Abadie and L'Hour, 2021; Pouliot and Xie, 2022). Thus, the key question is how to exploit

[1] Another way of viewing disaggregated data is temporal disaggregation. For a discussion around the SC estimator and temporal disaggregation, see Sun et al. (2024).

the additional information in disaggregated data while maintaining credible and stable estimates.

In this paper, I develop a general framework for synthetic control estimators that incorporates disaggregated data and introduce the multi-level SC (mlSC) estimator. The framework nests existing approaches, ranging from full aggregation (classical SC) to full disaggregation of treated and/or control units. Building on this framework, I propose the mlSC estimator that formalizes the choice of aggregation as a data-adaptive penalization scheme. The estimator leverages the flexibility of SC weights to incorporate disaggregated control unit information while regularizing toward the classical SC solution. In simulations calibrated to four real data sets, mlSC performs well relative to classical SC and its disaggregated variants. In two empirical applications, I demonstrate that my estimator adapts to the data structure in a fully data-driven manner, delivering precise and credible treatment effect estimates without an a priori aggregation decision.

Applied researchers commonly implement several variants of the SC estimator to incorporate disaggregated data, even when the primary goal remains estimating the aggregate treatment effect. These range from full aggregation to full disaggregation of the treated unit, the control units, or both. Each variant corresponds to a special case within my general framework. Disaggregating the treated unit typically entails constructing a synthetic control for each disaggregated treated unit separately (Abadie and L'Hour, 2021), while disaggregating the control units expands the donor pool to include all disaggregated control units.

The value of disaggregating the treated unit alone lies not in improving aggregate-level precision, but in enabling the study of alternative estimands. Disaggregation of the treated unit reduces flexibility and adds noise, since replicating the aggregate outcome is simpler than replicating each disaggregated treated unit. When the estimand is at the aggregate level, it is therefore natural to target the aggregate directly. Simulations confirm that treated-unit disaggregation often worsens out-of-sample performance. However, it allows researchers to study alternative estimands such as treatment effect heterogeneity, which are impossible to investigate when the treated unit remains fully aggregated.

In contrast to the treated unit, incorporating disaggregated control data can substantially improve aggregate-level estimation precision, especially in low-noise settings. Simulations show that disaggregating the controls is the primary driver of better out-of-sample performance for SC estimators using disaggregated data. Expanding the donor pool increases estimator flexibility and provides more opportunities to find suitable matches for the (aggregated or disaggregated) treated unit (see, e.g., Hanushek et al., 2023; Kreif et al., 2016). However, disaggregated controls are inherently noisier than their aggregated counterparts, raising the risk of overfitting and high-dimensionality problems when the number of control units exceeds pre-treatment periods (Pouliot and Xie, 2022). Thus, aggregate-level precision benefits most when the disaggregated control data contain meaningful signal relative to noise.

I propose the multi-level SC (mlSC) estimator that addresses both disaggregation dimensions in a data-driven way through a hierarchical penalization approach. Given the asymmetry in gains, I focus on the variant that leaves the treated unit aggregated while disaggregating the controls. The estimator augments the SC objective with a penalty term that shrinks the disaggregated-control SC toward the classical SC solution, reflecting the hierarchical data structure. The magnitude of the penalty balances flexibility against the risk of overfitting: a large penalty favors the classical SC, while a small penalty leverages disaggregated control information when pre-treatment fit gains outweigh noise. The penalty parameter can be chosen via cross-validation over time or a theoretically motivated heuristic, letting the data determine the optimal level of aggregation.

In semi-synthetic simulations calibrated to four empirical data sets, I demonstrate that my mlSC estimator outperforms the classical SC estimator and generally outperforms or matches the performance of the variants that uses only disaggregated control units. The simulations model outcomes with a hierarchical linear factor structure and assign treatment based on observed policies to reflect realistic interventions. I further show that the performance gains from using disaggregated data depend critically on the noise level in the data. The disaggregated data helps most in settings when noise level is low. The mlSC estimator gains most over the classical SC and the naive SC variants at moderate noise levels, where it can extract the additional signal from disaggregated units without overfitting the noise.

To illustrate its practical utility, I apply the mlSC to two empirical studies where state-level policies were evaluated using disaggregated data (see Deng and Zheng, 2023; Callaway and Sant'Anna, 2021). In these examples, I demonstrate the implementation of the mlSC estimator and emphasize its ability to adaptively select appropriate levels of aggregation based on the characteristics of the underlying data. The original authors made opposing choices: Deng and Zheng (2023) aggregate their grocery store-level data to the state level, while Callaway and Sant'Anna (2021) use county-level data directly. While the mlSC estimator replicates these decisions, it generates different treatment effects than both the classical SC and disaggregated-control SC, highlighting the impact of data-driven aggregation..

Related work. This paper contributes to the literature on the synthetic control (SC) estimator by intersecting two recent prominent research directions; see Abadie (2021) for a review. The first direction addresses the common challenge of imperfect pre-treatment fit, typically by debiasing the estimator or relaxing its constraints to permit extrapolation (e.g., Doudchenko and Imbens, 2016; Ferman and Pinto, 2021; Ben-Michael et al., 2021; Kellogg et al., 2021; Abadie and L'Hour, 2021). Notably, these methods still presuppose a single aggregation level. The second direction leverages disaggregated data, but generally to extend the SC model to new settings, such as estimating heterogeneous effects or analyzing treatments assigned at the disaggregate level (e.g., Robbins et al., 2017; Abadie and L'Hour, 2021; Shi et al., 2022; Sun et al., 2024). In contrast, my central contribution is to formalize and analyze the use of disaggregated data to directly address imperfect pre-treatment fit within the *classical* SC setting, i.e., estimating an aggregate-level effect with a limited number of aggregate units, an approach that has seen informal use in applied work (e.g., Kreif et al., 2016; Deng and Zheng, 2023; Hanushek et al., 2023).

2 General Set-Up for Incorporating Disaggregated Data

To set up the discussion of the choices for the SC estimator with disaggregated data, I introduce the standard potential outcomes framework and adapt it for disaggregated data. Treatment assignment is assumed to be at the aggregate level. Moreover, the estimand of interest is the aggregate level effect, which is simply the population weighted average of the treatment effects at the disaggregate level.

I adopt the standard Rubin potential outcomes framework (see, e.g., Rubin, 1974; Imbens and Rubin, 2015). Suppose I observe a panel of $S+1$ aggregated units, e.g. *s*tates, over T time periods. For each aggregated unit, I observe C_s disaggregated units, e.g. *c*ounties. Recall from the introduction that disaggregated data refers to any unit-level below the level of treatment assignment. Let Y_{sct} denote the disaggregated outcome for disaggregated unit c in aggregated unit s at time t and Y_{st} be the aggregated outcome. Assume that each aggregated unit is a weighted average of its disaggregated components, i.e.

$$Y_{st} = \sum_{c=1}^{C_s} v_{sc}\, Y_{sct},$$

where v_{sc} denote the aggregation weights. It is important that $\sum_{c=1}^{C_s} v_{sc} = 1$ to retain interpretability of the synthetic control at all aggregation stages. One example of weights could be simple averages, e.g. $v_{sc} = \frac{1}{C_s}$.

Let $Y_{sct}(0)$ and $Y_{st}(0)$ denote the potential outcomes in absence of treatment and $Y_{sct}(1)$ and $Y_{st}(1)$ the potential outcome in presence of treatment, for the disaggregated and aggregated outcome respectively. Assume that treatment is binary and assigned at the aggregated level. As in many SC settings, attention is restricted to a single aggregated treated unit, which is assigned treatment in period T_0 that never turns off [2], i.e.

$$W_{sct} = \begin{cases} 1, \ \forall c = 1, ..., C_s, \ s \text{ treated} \\ \\ 0 \text{ else} \end{cases} .$$

The binary treatment indicator can be equivalently defined at the aggregate level, W_{st}. Overall, the observed outcomes on the disaggregate level can be written as

$$Y_{sct} = \begin{cases} Y_{sct}(1) \text{ if } W_{sct} = 1 \\ \\ Y_{sct}(0) \text{ if } W_{sct} = 0 \end{cases} .$$

The observed outcomes at the aggregate level follow as expected. The setting in this paper focuses on the setting without covariates, similarly to Ferman and Pinto (2021).

The estimand of interest is the average treatment effect on the aggregated treated unit. Without loss of generality, assume the aggregated treated unit is $s = 0$. The Stable Unit Treatment Value Assumption (SUTVA) is assumed to hold at the disaggregate level, so there is no interference between disaggregated units. Then, I can decompose the treatment effect at the aggregate level as

$$\begin{aligned} \tau &= \frac{1}{T - T_0 + 1} \sum_{t=T_0+1}^{T} \tau_{0t} \\ &= \frac{1}{T - T_0 + 1} \sum_{t=T_0+1}^{T} \left(Y_{0t}(1) - Y_{0t}(0)\right) \\ &= \frac{1}{T - T_0 + 1} \sum_{t=T_0+1}^{T} \sum_{c'=1}^{C_0} v_{0c'} \left(Y_{0c't}(1) - Y_{0c't}(0)\right) \end{aligned}$$

[2]In the typical SC setting, only a single aggregated unit is treated. If there are multiple treated units, they are either (1) aggregated to a single treated unit or (2) a synthetic control is found separately for each of them(see, e.g., Abadie and L'Hour, 2021). Oftentimes, multiple treated units arise in settings with staggered adoptions which can be accommodated in the synthetic control framework (see, e.g., Ben-Michael et al., 2022; Athey and Imbens, 2022).

3 Incorporating Disaggregated Data: Choices for the SC Estimator

When disaggregated data is available, applied researchers face a common choice in the SC setting: at what level of aggregation should the estimator operate? This choice arises in two places: (1) outcome for the treated unit and (2) outcomes for the control units. These two dimensions yield four possible data configurations, each with distinct implications for estimator performance. This section provides a structured discussion of these choices and their practical relevance. I illustrate the intuition behind the main trade-offs involved through a stylized example.

3.1 Synthetic Control Estimator and The Choice of Aggregation

The goal of the SC estimator is to obtain a precise estimate of the treatment effect on the treated by accurately approximating the unobserved aggregated treated unit's potential outcome in absence of treatment, $Y_{0t}(0)$:[3]

$$\hat{\tau}_{0t} = Y_{0t} - \hat{Y}_{0t}(0).$$

As introduced by Alberto Abadie and co-authors (see, e.g., Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015), the classical SC estimator uses data aggregated to the level of treatment. The estimator estimates the unobserved potential outcome as follows:

$$\hat{Y}_{0t}(0) = \sum_{s=1}^{S} \omega^s \, Y_{st} \; \forall \; t \geq T_0,$$

[3] Recall that the estimand is the treatment effect on the treated, hence $Y_{0t}(1)$ is observed.

where ω^s is chosen s.t.

$$\begin{aligned} &\arg\min_{\omega^s \in R^S} \sum_{t=1}^{T_0} (Y_{0t} - \sum_{s=1}^{S} \omega^s Y_{st})^2 \\ &s.t. \sum_{s=1}^{S} \omega^s = 1, \quad \omega^s \geq 0 \; \forall \; s = 1, ..., S. \end{aligned} \tag{3.1}$$

Intuitively, the estimator constructs a synthetic version of the aggregated treated unit using a convex combination of a donor pool of units similar to the aggregated treated unit that did not experience the same intervention. Selecting an appropriate donor pool is a crucial step in this estimation process, as it determines the quality of the counterfactual and the credibility of the resulting treatment effect estimates (see, e.g., Abadie et al., 2010; Doudchenko and Imbens, 2016; Ferman and Pinto, 2021).

Disaggregated data can be incorporated into the classical SC estimator in two key ways: (1) disaggregating the outcomes for the treated unit, Y_{0t}, and (2) disaggregating the outcomes for the control units, Y_{st}. A natural first approach to disaggregating the treated unit is to construct separate SC estimators for each disaggregated treated unit (e.g., each county within a treated state) (see, e.g., Abadie and L'Hour, 2021; Ben-Michael et al., 2022). Conversely, disaggregating control units involves expanding the donor pool to include all disaggregated control units, rather than just the aggregated outcomes.

For expositional clarity, this paper focuses on a simplified case with two levels of aggregation. In practice, however, many more choices are possible. If individual-level data are available, one could aggregate outcomes at various intermediate levels—households, cities, counties, metropolitan areas, or states.

Focusing on two aggregation levels yields four possible data configurations, three of which are commonly used in practice (see Table 1).[4] The classical SC estimator corresponds to the aggregated treated/aggregated control quadrant. Notably, some researchers choose to aggregate their data to the level of treatment even when disaggregated data is available

[4]Within the SC paradigm, if I am interested in the average treatment effect on the treated aggregated unit, it makes sense to directly target the overall aggregated unit in-sample if that is my objective out-of-sample, especially when the control units are not disaggregated. More details can be found in Section 4.

(see, e.g., Deng and Zheng, 2023). Moreover, Pac et al. (2019, NBER paper version) state that "to employ my primary synthetic control model estimation method, the unit of observation must be the same as the level of the policy change. Accordingly, I aggregate individuals into state-year cells based on the family's state of residence at the survey date and the year of the child's birth".

Table 1: Synthetic Control with Disaggregated Data in Practice

		Control Units	
		Aggregated	**Disaggregated**
	Aggregated	Classical SC Estimator e.g. Deng and Zheng (2023)	e.g. Kreif et al. (2016)
Treated Unit			
	Disaggregated	*not common*	e.g. Hanushek et al. (2023)

Kreif et al. (2016) implement an SC estimator that corresponds to the top-right quadrant (i.e. aggregated treated/disaggregated controls): they re-evaluate a pay-for-performance (P4P) initiative which affects several hospitals in the North West, but excludes hospitals in other regions. Following the recommendation from Abadie et al. (2010), they aggregate the treated hospitals into a treated North West region. For the control hospitals, however, they choose not to aggregate them into the nine control regions as "[t]his would leave insufficient power to detect whether there was a statistically significant treatment effect" (Kreif et al., 2016).

Hanushek et al. (2023) go one step further by also disaggregating the treated units as in the bottom right quadrant (i.e. disaggregated treated/disaggregated controls). They evaluate the comprehensive teacher and principal evaluation and compensation systems in the Dallas Independent School District by building a synthetic control for each school in the school district separately using control schools. They choose to conduct their analysis at the school level instead of the district level because it "recognizes the substantial variation in school quality within districts and dampens the impact of the reform efforts or challenges of other districts" (Hanushek et al., 2023).

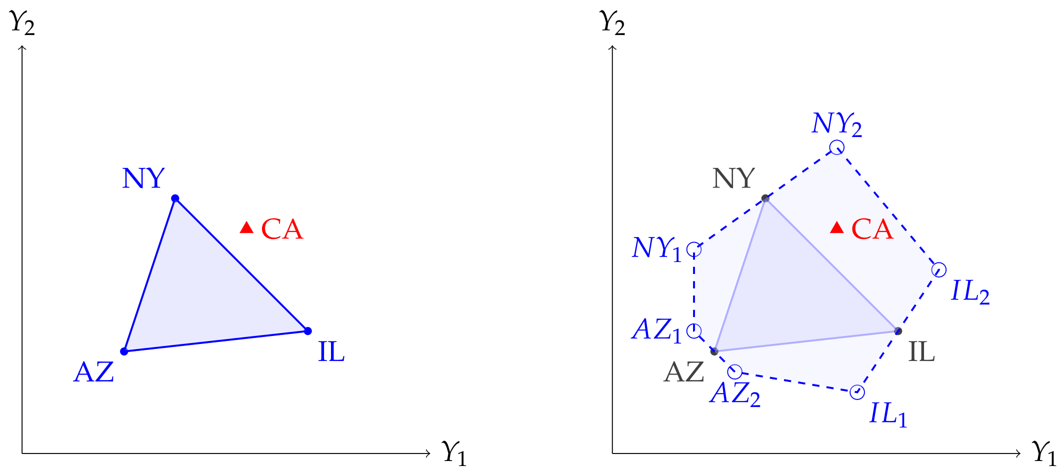
While applied researchers often face the task of selecting an appropriate level of aggregation, there is little formal guidance to inform this decision. The examples discussed illustrate some of the key trade-offs researchers consider. Aggregating treated units can help align the analysis with the level at which the policy is implemented or reduce idiosyncratic noise in outcomes. In contrast, disaggregation enables researchers to exploit more granular variation. Ultimately, the choice of aggregation level is a balancing act, where researchers must weigh the benefits and drawbacks of using disaggregated data.

3.2 The Choice of Aggregation: A Stylized Example

In this subsection, I preview the trade-off for the two dimensions of disaggregation I consider in this paper through a simple, stylized example. There are two key aspects to the trade-off: flexibility and overfitting due to additional noise in the estimation procedure.

First, I focus on the flexibility of the estimator. Consider an example with four states that serve as the aggregated units (say Arizona, AZ; California, CA; New York, NY; Illinois, IL) and three time periods ($t \in \{1, 2, 3\}$), where treatment at time period $t = 3$ is assigned to a single aggregated unit, CA. Moreover, each state consists of two counties, representing the disaggregate level of analysis. my goal is to estimate the treatment effect on California in period 3. In this setting, I compare the two dimensions of disaggregation. Recall that, when I disaggregate the control units, my SC estimator uses all counties as donors. When I disaggregate the treated unit, my SC estimator finds a synthetic control for each treated county separately.

Disaggregation of the control states. Figure 1, panel (a) shows the convex hull formed by the control states in two pre-treatment periods (shaded blue area). The SC estimator assigns convex weights to these three states, allowing it to perfectly match any treated unit that lies within this convex hull. Since CA lies outside the convex hull, the SC estimator cannot perfectly replicate CA's pre-treatment outcomes. In contrast, panel (b) displays the convex hull formed by the control counties, which are the disaggregated units, in the same two periods. With access to six counties, the SC estimator now has more flexibility, assigning weights across a larger donor pool. As a result, it is able to construct a synthetic CA that exactly matches the observed pre-treatment outcomes, something that was not possible using the aggregated state-level data.[5]



(a) Pre-treatment outcomes Y_t. Convex hull spanned by outcomes on state-level.

(b) Pre-treatment outcomes Y_t. Convex hull spanned by outcomes on county-level.

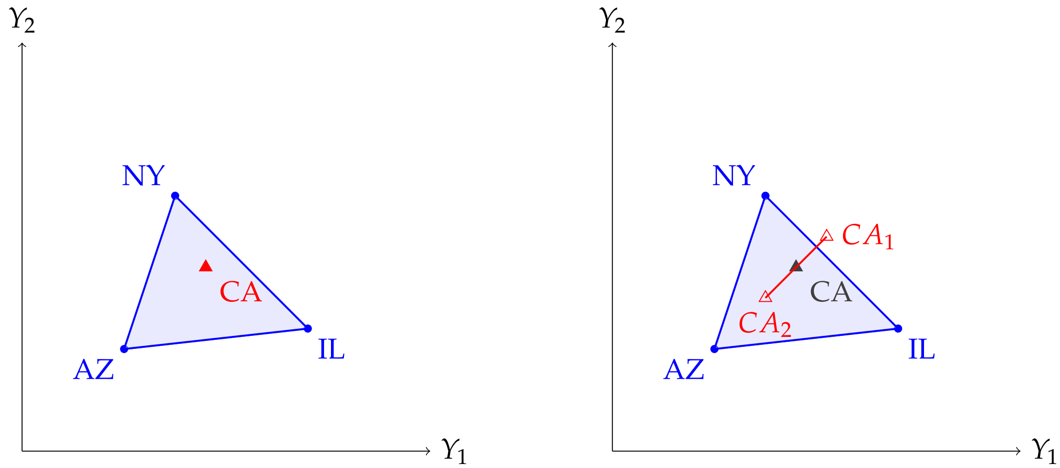
Figure 1: Disaggregating the control units. Four-aggregated unit, six-disaggregated unit, three-period example. Treated aggregated unit (red triangle): CA.

Disaggregating the treated unit. In Figure 2, panel (a), the treated aggregated unit CA now lies inside the convex hull formed by the state-level controls, enabling a perfect match and synthetic control. In panel (b), I disaggregate the treated state. In this case, a separate SC estimator is constructed for each treated county. While one county (CA_2) still lies inside the convex hull and can be matched perfectly, the other one (CA_1) does not. As a result, disaggregation of the treated state prevents us from replicating CA's average pre-treatment outcomes exactly. [6]

Together, these examples highlight the asymmetry in how disaggregation affects the performance of the SC estimator. Disaggregating control units expands the convex hull, increasing flexibility and fit. Disaggregating the treated unit, in contrast, may reduce flexibility by making each component harder to replicate. These opposing forces make the performance

[5]If CA were already inside the convex hull formed by the control states, the disaggregation would result in no change. If the states are disaggregated differently such that the convex hull formed by them does not include CA, it would at least find a closer match in terms of pre-treatment outcomes to CA than the states.

[6]This follows from Jensen's inequality: it is generally easier to fit the average of components than to fit each component individually. If both treated counties remain within the convex hull, disaggregation would not affect replicability.



(a) Pre-treatment outcomes. Convex hull spanned by outcomes on state-level.

(b) Pre-treatment outcomes. Convex hull spanned by outcomes on state-level.

Figure 2: Disaggregating the treated unit. Four-aggregated unit, six-disaggregated unit, three-period example. Treated aggregated unit (red triangle): CA.

of a fully disaggregated estimator ambiguous and highly data-dependent.

The second part of the trade-off concerns the role of noise in the estimation procedure. While closely matching pre-treatment outcomes is essential for constructing a good synthetic control, those outcomes typically contain idiosyncratic random noise. Disaggregated units, like counties, are noisier than their aggregated counterparts, like states, due to the hierarchical structure of the data. The expansion of the convex hull in Figure 1 could simply be driven by this idiosyncratic noise rather than systematic variation, causing overfitting to pre-treatment noise and poor post-treatment performance. Thus, it becomes crucial to distinguish between genuine signal and random noise in the disaggregated outcomes when deciding whether and how to incorporate disaggregated data into the estimator.

Ultimately, these examples underscore that the choice of aggregation level is both data- and context-dependent. It is shaped not only by the geometric relationship between the treated unit(s) and the control units in outcome space but also by the balance between noise and information in the disaggregated data. When disaggregation expands the donor pool in a way that captures meaningful variation, it can improve estimation. But when it primarily amplifies noise, it can worsen performance.

4 A General SC Estimator For All Aggregation Levels

In this section, I introduce a general class of SC estimators for disaggregated data, characterized by the weight matrix over which the estimator optimizes. This class provides a unifying framework accommodating any combination of aggregation and disaggregation. In particular, different restrictions on the weight matrix recover the four cases of full aggregation and full disaggregation for treated and control units discussed in Section 3, including the classical SC estimator as a special case.

The disaggregated general SC estimator (dGSC) is defined by a weight matrix, $W_{c'sc} \in \mathbb{R}^{C_0 \times \sum_{s=1}^{S} C_s}$, where c' refers to treated disaggregated unit c', s and c refer to a control disaggregated unit c contained in the aggregated unit s. This weight matrix has dimension $C_0 \times \sum_{s=1}^{S} C_s$, where C_0 is the number of disaggregated units contained in treated aggregated unit 0 and $\sum_{s=1}^{S} C_s$ is the total number of disaggregated control units. For treated aggregated unit 0, the estimator has the form

$$\hat{\tau}_{0t}^{dGSC} = \sum_{c'=1}^{C_0} v_{sc'} \left(Y_{0c't} - \sum_{s=1}^{S} \sum_{c=1}^{C_s} W_{c'sc} Y_{sct} \right),$$

where the weights $W_{c'sc}$ are chosen to solve the following optimization problem:

$$\arg \min_{W_{c'sc} \in \mathcal{R}^0} \sum_{t=1}^{T_0} \sum_{c'=1}^{C_0} v_{0c'} \left(Y_{0c't} - \sum_{s=1}^{S} \sum_{c=1}^{C_s} W_{c'sc} Y_{sct}\right)^2, \tag{4.1}$$

where $\mathcal{R}^0 = \{W_{c'sc} \mid \sum_{s=1}^{S} \sum_{c=1}^{C_s} W_{c'sc} = 1 \, \forall c' \quad \text{and} \quad W_{c'sc} \geq 0 \, \forall c', c, s \neq 0\}$. $\mathcal{R}^0$ incorporates the convexity constraints on the weights following the classical SC estimator restrictions in Equation 3.1. These two restrictions limit extrapolation and enhance interpretation through sparsity. Moreover, they serve as an implicit regularizer in (close to) high-dimensional settings. The dGSC estimator differs from other SC estimators with disaggregated data such as distributional SC (see, e.g., Chen, 2020; Gunsilius, 2023) in one important aspect: I focus solely on matching the mean as opposed to the entire distribution which matches the treated and control units also on higher-order moments.

The four estimators introduced in Section 3 are special cases in this general class that can be obtained by additional restrictions in $\mathcal{R}^0$ on the possible set of weight matrices. A complete overview of the estimators can be found in Table 2. Note that the dGSC estimator as given in Equation 4.1 corresponds to the case, where both the treated and the control units are disaggregated.

The dGSC-AA[7] estimator, which aggregates treated and control units, corresponds to

$$\mathcal{R}^{dGSC-AA} = \{W_{c'sc} \in \mathcal{R}^0 \,|\, \frac{W_{c'sc_1}}{W_{c'sc_2}} = \frac{v_{sc_1}}{v_{sc_2}} \,\forall c_1, c_2 \in C_s, \forall s \text{ and } W_{c'_1sc} = W_{c'_2sc} \,\forall c'_1, c'_2 \in C_0, \forall s, c\}.$$

Overall, these restrictions force all weights for disaggregated units c within an aggregated control unit s to be equal. Moreover, the weights also have to be the same for each disaggregated treated unit c'.

Remark The dGSC-AA optimization problem enforces multiple types of equality constraints on the full weight matrix $W_{c'sc}$: (a) within-aggregated control units equality, (b) across disaggregated treated units equality and (c) convexity of the weights. An alternative characterization of the optimization problem, given linear formulations of the constraints in $\mathcal{R}^{dGSC-AA}$, is given by its Lagrangian

$$\begin{aligned}\mathcal{L}^{dGSC-AA}(W, \Lambda, \Gamma, \{\lambda_{c'}\}, M) = &\sum_{t=1}^{T_0} \sum_{c'=1}^{C_0} v_{0c'} \, (Y_{0c't} - \sum_{s=1}^{S} \sum_{c=1}^{C_s} W_{c'sc} Y_{sct})^2 \\ &+ \sum_{c',s,c} \Lambda_{c'sc}(W_{c'sc} - v_{sc} \sum_{c=1}^{C_s} W_{c'sc]}) + \sum_{c',s,c} \Gamma_{c'sc}(W_{c'sc} - \sum_{c''=1}^{C_0} v_{0c''} W_{c''sc}) \\ &+ \sum_{c'=1}^{C_0} \lambda_{c'}(1 - \sum_{s=1}^{S} \sum_{c=1}^{C_s} W_{c'sc}) - \mu_{c'sc} W_{c'sc},\end{aligned}$$

where $\Lambda_{c'sc}$ and $\Gamma_{c'sc}$ are Lagrange multipliers enforcing the additional constraints in $\mathcal{R}^{dGSC-AA}$ and $\lambda_{c'}$ and $\mu_{c'sc}$ enforce the convexity constraints in $\mathcal{R}^0$. This formulation imposes the constraints exactly for the dGSC-AA estimator. I can similarly characterize the other special

[7] dGSC-AA: dGSC-**A**ggregate **A**ggregate. The naming convention lists the aggregation level of the treated unit first and then the aggregation level of the control units.

cases that follow.

The SC estimator that corresponds to only disaggregating the control units (dGSC-AD) is given by

$$\mathcal{R}^{dGSC-AD} = \{W_{c'sc} \in \mathcal{R}^0 \mid W_{c'_1sc} = W_{c'_2sc} \; \forall c'_1, c'_2 \in C_0\}.$$

This restriction enforces equality of weights across all disaggregated treated units, while allowing weights for disaggregated control units to vary freely within the constraints of $\mathcal{R}^0$. Note that the dGSC-AA estimator is a special case of dGSC-AD, obtained by imposing one additional restriction on the weight matrix. To address potential non-uniqueness in high-dimensional settings, I include a small L_2 penalty (see, e.g., Shen et al., 2023; Spiess et al., 2023).

The SC estimator that corresponds to only disaggregating the treated unit (dGSC-DA) is given by

$$\mathcal{R}^{dGSC-DA} = \{W_{c'sc} \in \mathcal{R}^0 | \frac{W_{c'sc_1}}{W_{c'sc_2}} = \frac{v_{sc_1}}{v_{sc_2}} \; \forall c_1, c_2 \in C_s\}.$$

This restriction enforces equal weights across all disaggregated units within each aggregated control unit while allowing each disaggregated treated unit to have its own synthetic control.

Table 2: dGSC estimator comparison.

	dGSC-AA (Classical SC)	**dGSC-AD**	**dGSC-DA**	**dGSC**
Uniform weights for all c within aggregated control unit s	Yes	No	Yes	No
Uniform weights for all c' within aggregated treated unit 0	Yes	Yes	No	No

Estimators: dGSC-AA: aggregated data for treated and control; dGSC-AD: aggregated data for treated, disaggregated data for control; dGSC-DA: disaggregated data for treated, aggregated data for control; dGSC: disaggregated data for treated and control

The optimization problem for the dGSC estimator differs from the classical SC estimator in Equation 3.1 in two key ways: it uses a weight matrix instead of a vector, and it sums over disaggregated treated units outside the squared pre-treatment error. Despite these differences, the dGSC class is flexible enough to recover all four combinations of full aggregation and disaggregation for treated and control units introduced in Section 3.

Proposition 1. *For all $\mathcal{R} \subseteq \mathcal{R}^{dGSC-AD}$, the dGSC optimization problem in Equation 4.1 is equivalent to the classical SC optimization problem in Equation 3.1, differing only in the selection of units included in the donor pool.*

Appendix A.1 contains the proof. The key insight is that, once equal weights are imposed across disaggregated treated units, the additional summation over these units outside the squared pre-treatment error does not change the solution.

Corollary 1. *Imposing $\mathcal{R}^{dGSC-AA}$ for the dGSC estimator, which additionally restricts all weights within each aggregated control unit to be equal, recovers the classical SC estimator.*

Remark The objective function of the classical SC estimator is commonly motivated as the sample analogue of the expected out-of-sample mean-squared error for the aggregate treatment effect on the treated unit. Proposition 1 shows that, within the dGSC class, this interpretation holds only if the weights for all disaggregated units within the treated aggregate are constrained to be equal, thus keeping the treated unit at the aggregate level. This restriction ensures that the dGSC estimator is targeting the same out-of-sample loss as the classical SC estimator for the aggregate-level treatment effect. A key implication of this result is that, when the estimand remains the standard treatment effect on the treated aggregated unit, the potential gains from disaggregation primarily arise from disaggregating the control units rather than the treated unit itself. These insights align with the findings of Arkhangelsky et al. (2021) and Ben-Michael et al. (2022).

Remark Another widely used estimator in panel data settings is the difference-in-differences (DiD) estimator (Ashenfelter and Card, 1984; Card, 1990; Card and Krueger, 1994). Following Doudchenko and Imbens (2016), the DiD estimator can be expressed within the general dGSC framework by modifying the optimization problem in Equation 4.1 in two ways: (1) adding an intercept term, μ^{did}, to capture level differences across treated and control units, and (2) constraining all weights to be uniform across the control units for each treated unit. Under these restrictions, the DiD estimator using aggregated data, DiD (aggregate), assigns weight $\hat{w}_s^{did,agg} = \frac{1}{S}$ to each aggregated control unit and estimates the intercept $\hat{\mu}^{did}$ based on the aggregate treated unit. The DiD estimator using disaggregated data, DiD (disaggregate), assigns weight $\hat{w}_{sc}^{did,disagg} = \frac{1}{\sum_s C_s}$ to each disaggregated control unit and estimates the intercept as the simple average across all disaggregated treated units. This average may or may not correspond exactly to the aggregated treated outcome, depending on the population weights v_{sc} used in the analysis. While the intercept adds flexibility, the uniform (non–data-driven) weights, proportional to the number of control units, limit the estimator's ability to fit pre-treatment outcomes and can result in substantial efficiency loss. Furthermore, the credibility of the DiD estimator relies on the parallel trends assumption, which often fails in empirical applications (Athey and Imbens, 2006; Freyaldenhoven et al., 2019; Kahn-Lang and Lang, 2020; Arkhangelsky et al., 2021; Ghanem et al., 2022; Roth, 2022; Rambachan and Roth, 2023; Arkhangelsky and Hirshberg, 2023). Because DiD weights are not data-driven, the level of aggregation has a smaller impact on estimation precision than in synthetic control, where weights are explicitly fitted to outcome trends. Further discussion of disaggregation for the DiD estimator is provided in Appendix H.

5 Leveraging All Aggregation Levels: The Multi-Level SC Estimator

In this section, I introduce the multi-level SC estimator (mlSC), which leverages all levels of aggregation to select an estimator within the general dGSC framework in a data-driven way. The mlSC estimator thus reframes the *a priori* choice of aggregation as a penalization problem by introducing hierarchical penalty terms. This structure captures the hierarchical nature of the data and allows the estimator to recover any of the four cases or any intermediate mixture, depending on the strength of the penalization. In this paper, I focus on the version that keeps the treated unit at the aggregate level, while allowing flexible aggregation among the control units. I then discuss two practical approaches for obtaining a feasible estimator.

5.1 Multi-Level SC Estimator

The mlSC estimator builds on the Lagrangian formulation of the dGSC-AA, dGSC-AD and dGSC-DA estimators from Section 4. Unlike those estimators, which fix the level of aggregation for units a priori by imposing hard equality constraints on the weights, the mlSC replaces these constraints with quadratic penalty terms. This soft penalization allows the data to determine the appropriate level of aggregation in a data-driven way.

In this framework, the two penalty parameters, λ_1 and λ_2, correspond to the linear formulations of the constraints in $\mathcal{R}^{dGSC-AA}$ and control how strongly the estimator is pulled toward full aggregation. Fixing the aggregation level a priori is equivalent to setting these parameters at specific values, which may unnecessarily restrict the estimator. In contrast, because the SC estimator inherently combines information across multiple control units to predict the missing counterfactual outcome $\hat{Y}_{0T}(0)$ for treatment effect estimation, selecting λ_1 and λ_2 from the data can improve out-of-sample performance and help approximate the optimal penalties λ_1^* and λ_2^*.

Formally, the mlSC estimator is defined as:

$$\begin{aligned} \arg\min_{W_{c'sc}\in\mathcal{R}^0} \; & \sum_{t=1}^{T_0}\sum_{c'=1}^{C_0} v_{0c'}\,(Y_{0c't}-\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{c'sc}Y_{sct})^2 \\ & + \lambda_1 \times \sum_{c'=1}^{C_0}\sum_{s=1}^{S}\sum_{c=1}^{C_s}(W_{c'sc}-v_{sc}W_{c's,\cdot})^2 \\ & + \lambda_2 \times \sum_{c'=1}^{C_0}\sum_{s=1}^{S}\sum_{c=1}^{C_s}(W_{c'sc}-\bar{W}_{\cdot,sc})^2, \end{aligned} \tag{5.1}$$

where $W_{c's,\cdot} = \sum_{c=1}^{C_s} W_{c'sc}$ is the aggregate weight in aggregated control unit s for disaggregated treated unit c' and $\bar{W}_{\cdot,sc} = \sum_{c''=1}^{C_0} v_{0c''}W_{c''sc}$ is the average weight for disaggregated control unit c in aggregated control unit s.

The first penalty discourages deviations from the assigned aggregate weights of the dGSC-AA estimator, while the second controls variation across disaggregated treated units.

Because the optimal values λ_1^*, λ_2^* depend on unobservable quantities such as out-of-sample prediction error, the penalty parameters must be estimated from the data, effectively treating aggregation as a tuning-parameter problem rather than a fixed modeling choice.

The mlSC estimator nests all four SC variants introduced in Section 4 as limiting cases. As $\lambda_1, \lambda_2 \to \infty$, it reduces to the classical SC estimator based solely on aggregated data. Setting $\lambda_1 \to \infty, \lambda_2 = 0$ recovers the dGSC-DA estimator, $\lambda_1 = 0, \lambda_2 \to \infty$ yields the dGSC-AD estimator, and setting both penalties to zero recovers the fully disaggregated dGSC estimator.

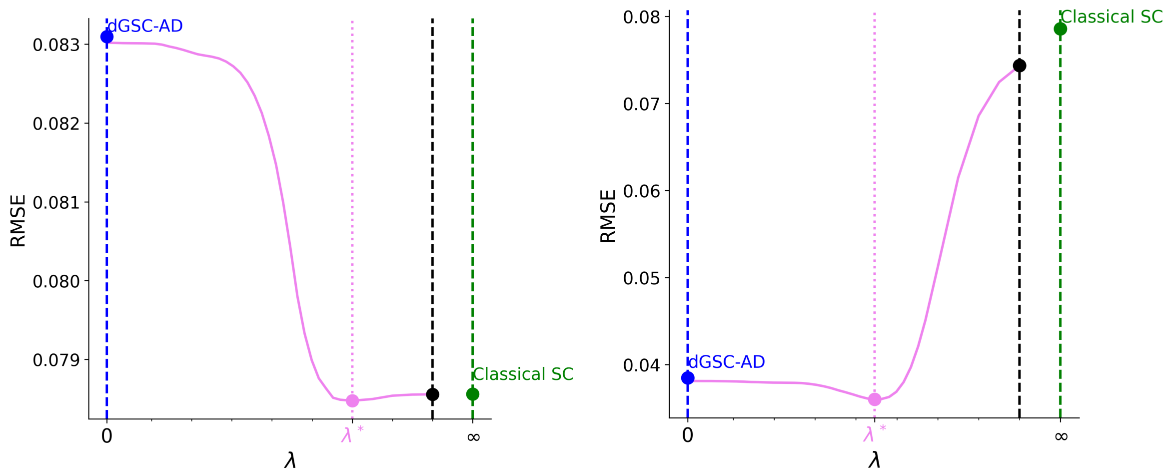
In contrast, the mlSC penalty is motivated by selecting the optimal aggregation level itself. Its ridge-type form ensures that even with generalized population weights, the estimator converges to the standard SC variants as the penalties grow large. Alternative penalty forms, for example, those based on conditional variance terms—are possible, but they do not guarantee convergence to the classical SC estimator in the large-penalty limit.

While the full mlSC framework allows flexible penalization across both treated and control units, in this paper I focus on the case where the treated unit remains aggregated. This corresponds to setting $\lambda_2 = \infty$, enforcing aggregation across treated units while allowing the estimator to flexibly learn the optimal degree of aggregation among control units through λ_1. Under this restriction, the mlSC optimization problem simplifies to:

$$
\begin{aligned}
\arg \min_{\omega_{sc} \in \mathbb{R}^{\sum_{s=1}^{S} C_s}} \quad & \sum_{t=1}^{T_0} (Y_{0t} - \sum_{s=1}^{S} \sum_{c=1}^{C_s} \omega_{sc} Y_{sct})^2 \\
& + \lambda_1 * \sigma_y^2 * \sum_{s=1}^{S} \sum_{c=1}^{C_s} (\omega_{sc} - v_{sc} w^s)^2 \\
s.t. \quad & \sum_{s=1}^{S} \sum_{c=1}^{C_s} \omega_{sc} = 1 \quad \text{and} \quad \omega_{sc} \geq 0 \; \forall c, s \neq 0,
\end{aligned}
\tag{5.2}
$$

where $w^s = \sum_{c=1}^{C_s} \omega_{sc}$ and the weight matrix $W_{c'sc}$ collapses to a vector, $\omega_{sc} \in \mathbb{R}^{\sum_{s=1}^{S} C_s}$. The additional scaling factor σ_y^2 ensures the penalty is on the same scale as the loss function while the penalty parameter λ is scale invariant.

This specification serves as my preferred estimator going forward. Figure 3 illustrates how the (oracle) mlSC adapts the penalty parameter λ^* to the characteristics of the data in a real-world application.



(a) Classical SC estimator dominates the dGSC-AD estimator.

(b) dGSC-AD estimator dominates the classical SC estimator.

Figure 3: RMSE as a function of λ for the oracle mlSC estimator using semi-synthetic data based on a subset of units for log wages data. Based on $S_{sim} = 1000$ simulation runs. For more details, see Section 6. Dashed blue line refers to the dGSC-AD estimator, dashed green line to the classical SC estimator and the dashed black line to the end of the λ-grid used in the optimization procedure.

5.2 Penalty Parameter

To obtain a feasible mlSC estimator in practice, I propose two approaches for selecting the penalty parameter, λ_1: cross-validation over time and a model-based heuristic. Estimating λ is necessary because the true counterfactual outcome for the aggregated treated unit post-treatment is unobserved.

Cross-Validation over Time. The first approach estimates $\hat{\lambda}^*$ via leave-t_{cv}-out cross-validation for the aggregated treated unit. In practice, I select pre-treatment periods immediately preceding treatment, $T_0 - t_{cv}$, as the holdout set.[8] The rationale for using periods just before treatment is that they are likely to resemble the post-treatment periods, making them a good proxy for out-of-sample prediction. Formally, the cross-validated penalty is chosen

[8]It is not strictly necessary to use periods right before treatment; one could choose older periods or multiple non-contiguous periods as well.

as

$$\hat{\lambda}^* = \arg\min_{\lambda} \sum_{t=T_0-t_cv}^{T_0} \hat{\tau}_{0t}(\lambda)^2,$$

where $\hat{\tau}_{0t}(\lambda) = Y_{0t} - \sum_{s=1}^{S}\sum_{c=1}^{C_s} \omega^*_{sc}(\lambda) Y_{sct}$ denotes the estimated treatment effect in the hold-out pre-treatment periods and the weights $\omega^*_{sc}(\lambda)$ are obtained by solving Equation 5.2. This approach performs well when T_0 is sufficiently large, providing enough pre-treatment periods for cross-validation. [9]

Heuristic for λ. An practical alternative to cross-validation over time, especially when only a few pre-treatment periods are available, is a model-based heuristic. The heuristic is derived from the optimal λ^* in a stylized hierarchical random effects model under a simple scenario with $T = S = C_s = 2$. It is given by

$$\hat{\lambda} = 2\,\frac{\hat{\sigma}^2_{\varepsilon}}{\hat{\sigma}^2_{y}},$$

where $\hat{\sigma}^2_{\varepsilon}$ is the estimated variance of the error term and $\hat{\sigma}^2_{y}$ is the estimated variance of the outcome Y.[10] Dividing by $\hat{\sigma}^2_{y}$ ensures the heuristic is scale-invariant. Intuitively, this approach imposes a larger penalty when the data are noisier. For further derivation and justification of this heuristic, see Appendix B.

6 Simulation Results

I evaluate the performance of the disaggregated generalized SC (dGSC) estimators and the proposed multi-level synthetic control (mlSC) estimator through a series of semi-synthetic simulations based on four empirical datasets. This design allows me to compare the estimators under assignment mechanisms that resemble realistic policy interventions. Across these simulations, I find that disaggregating the control units is the main driver of performance

[9] Alternatively, one could perform cross-validation over units. This requires an additional exchangeability assumption across units and is computationally expensive.

[10] Several approaches can be used to estimate these variances. Here, I use a simplified hierarchical latent factor model as described in Appendix G, taking the average estimated variance across all other aggregated units except for the treated unit. The same procedure is applied to estimate $\hat{\sigma}^2_{y}$.

gains in the dGSC estimators. By contrast, disaggregating the treated unit yields, at most, modest improvements over either the classical SC or the dGSC-AD estimator. The benefits of control unit disaggregation are particularly pronounced in settings with low noise levels. Overall, the oracle mlSC consistently achieves the lowest estimation error, while the feasible mlSC, using the heuristic or cross-validation over time, outperforms the classical SC and generally matches or exceeds the performance of dGSC-AD. The mlSC provides the greatest improvement over either estimator in settings where the noise level is such that the classical SC and dGSC-AD perform roughly equally.

6.1 Simulation Set-Up

I follow the simulation framework of Arkhangelsky et al. (2021) to evaluate estimator performance. In particular, I create semi-synthetic placebo studies using four real-world datasets: state-level unemployment rates and weekly log wages from the U.S. Bureau of Labor Statistics (BLS), smoking rates from the Behavioral Risk Factor Surveillance System (BRFSS), and log GDP from the Penn World Table. Details on data construction and dataset descriptions are provided in Appendix C. The simulation design has two main components: outcome construction and treatment assignment.

For the outcomes, I assume a hierarchical latent factor model:

$$Y_{sct} = \underbrace{\alpha_s'\beta_t}_{L_{st}^{agg}} + \underbrace{\eta_{sc}'\beta_t}_{L_{sct}^{disagg}} + \varepsilon_{sct}, \quad \varepsilon_{sct} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2),$$

where $L^{agg}st$ captures the aggregated systematic component and $L^{disagg}sct$ captures deviations at the disaggregate level. For the simulations, I assume that aggregated units are simple averages of its disaggregated units, i.e. $v_{sc} = \frac{1}{C_s}$. Latent components are obtained from a rank-three factor model fit to the observed data:

$$L = \arg \min_{L:rank(L)=3} \sum_{sct} (Y_{sct} - L_{sct})^2,$$

with Y_{sct} denoting the observed outcome. Aggregated and disaggregated components are then defined as:

$$L^{agg}_{st} = \frac{1}{C_s}\sum_{c=1}^{C_s} L_{sct}$$

$$L^{disagg}_{sct} = L_{sct} - L^{state}_{st} \; \forall s$$

Residuals $e_{sct} = Y_{sct} - L_{sct}$ are used to estimate the error variance σ^2_{ε}. I evaluate estimator performance using root mean squared error (RMSE), appropriate for the null-effect setting where $\tau = 0$.

Unlike in randomized designs (see, e.g., Bertrand et al., 2004; Bottmer et al., 2024), treatment assignment is modeled to reflect the observational settings where SC estimators are typically applied. For the three U.S. state-level outcomes, treatment is assigned following historical adoption patterns of minimum wage and gun control laws (Arkhangelsky et al., 2021). For the international data, countries are grouped into six continents, with assignment based on financial market development and industrialization. [11] In all cases, treatment is assigned to a single aggregated treated unit ($N_{tr} = 1$) in a way that correlates with the systematic components, mimicking policy decisions that respond to latent economic conditions.

Formally, the treatment indicator is

$$W_{sct} = D_s \mathbb{1}_{t>T_0}, \qquad D_s \sim Bernoulli(\pi_s), \quad \pi_s = \frac{\exp\{\phi(\alpha_s + \bar{\eta}_s)\}}{1 + \exp\{\phi(\alpha_s + \bar{\eta}_s)\}},$$

where ϕ is estimated via logistic regression of observed treatment adoption on the systematic aggregated component.

6.2 Performance of the dGSC Estimators

I first examine the value of disaggregation in SC estimators. As a benchmark, I also report results for difference-in-differences (DiD) estimators, using both aggregated and disaggre-

[11] The systematic components explain roughly 14–30% of variation in treatment status.

gated data—a common alternative to SC methods when disaggregated data is available. In the main text, I present results using financial market–based assignment for the international data and minimum wage laws for the state-level data. Results for alternative treatment assignments, including random assignment, are provided in Appendix D. Table 3 reports the RMSEs and bias across the two main designs for the classical SC, dGSC-AD, dGSC-DA, dGSC, and DiD estimators.

Disaggregation of Control Units. First, I focus on disaggregating only the control units. The results show that disaggregation improves out-of-sample performance for all but one data set (unemployment rate). For the Penn table $log(GDP)$ data, which contains few aggregated units, a substantial portion of the improvement comes from decreased bias, suggesting that imperfect pre-treatment fit played a significant role in the performance. Differences in performance across data sets are largely driven by the underlying noise level in the disaggregated data. To further explore the impact of noise on estimator performance, I artificially increase the noise level and analyze the results in Section 6.4.

Disaggregation of the Treated Unit. Next, I examine the effect of disaggregating the treated unit. The dGSC estimator, which fully disaggregates both the treated and control units, performs comparably to the classical SC or dGSC-AD estimator, and in some cases slightly outperforms them, with improvements ranging from 3% to 5%. However, for most data sets, the dGSC estimator performs similarly to the better of the classical SC and dGSC-AD estimators. The dGSC-DA estimator, which disaggregates only the treated unit, consistently yields the poorest performance. These results indicate that disaggregating the treated unit alone provides limited benefits; the primary gains from disaggregation arise from disaggregating the control units.

Benchmark DiD Estimators. Finally, focusing on the DiD estimators[12], I find that they perform poorly across most data sets and assignment processes, especially relative to the dGSC-AD estimator. DiD estimators exhibit substantial bias for outcomes such as the unemployment and smoking rates. Incorporating disaggregated data into DiD offers minimal

[12] DiD (aggregate) uses aggregated data only, while DiD (disaggregate) uses disaggregated data only.

improvements, which are small relative to the gains achieved by dGSC estimators. This pattern reinforces that the main advantage of disaggregated data is realized within SC-type estimators, where weights are flexibly and data-drivenly fitted to outcome trends, unlike in DiD.

Table 3: Simulation results for all dGSC and DiD estimators: RMSE and Bias.

	Classical SC	**dGSC-AD**	**dGSC-DA**	**dGSC**	**DiD (aggregate)**	**DiD (disaggregate)**
			RMSE			
Assn.: Financial markets						
Penn table $log(GDP)$	0.384	0.035	0.409	0.052	0.152	0.153
Assn.: Min. wage						
Unemployment rate	0.089	0.095	0.139	0.084	0.135	0.134
Log wages	0.124	0.038	0.253	0.037	0.128	0.127
Smoking rate	0.153	0.056	0.217	0.063	0.295	0.315
			Bias			
Assn.: Financial markets						
Penn table $log(GDP)$	0.209	-0.000	0.204	0.013	0.020	0.033
Assn.: Min. wage						
Unemployment rate	0.004	-0.003	0.053	-0.000	0.018	0.012
Log wages	0.030	0.003	0.119	0.002	0.006	-0.004
Smoking rate	-0.070	-0.012	-0.126	-0.031	-0.156	-0.196

All results are based on $S_{sim} = 1,000$ simulation runs. $N_{tr} = 1$ and $T_{post} = 1$. Outcomes are normalized to have mean zero and unit variance. *Estimators:* classical SC estimator: aggregated data for treated and control; dGSC-AD: aggregated data for treated, disaggregated data for control; dGSC-DA: disaggregated data for treated, aggregated data for control; dGSC: disaggregated data for treated and control; DiD (aggregate): difference-in-differences using aggregated data; DiD (disaggregate): difference-in-differences using disaggregated data.

6.3 Performance of the Multi-Level SC Estimator

I next evaluate the performance of the mlSC estimator. Table 4 compares the classical SC, dGSC-AD, mlSC with heuristic penalty selection (as derived in Section 5.2), and the oracle mlSC (trained using post-treatment outcomes) in terms of RMSE and bias for the two main designs. Three key conclusions emerge from this analysis.

First, the oracle mlSC establishes a performance frontier. The oracle mlSC generally outperforms both the classical SC and dGSC-AD estimators. For the Penn Table $log(GDP)$ and log wages data, the oracle mlSC performs comparably to dGSC-AD, indicating that full disaggregation of control units is the most effective approach for these data sets. For the other data sets, the oracle mlSC achieves gains over the better of the classical SC and dGSC-ADestimators, ranging from 9% to 18%.

Second, the feasible mlSC estimators closely track the oracle frontier. Both the heuristic and cross-validated mlSC estimators perform near the oracle benchmark. When disaggregation meaningfully improves fit, the heuristic mlSC outperforms the classical SC and dGSC-AD estimators for all but one data set, where performance is essentially equivalent. The cross-validated mlSC exhibits similar patterns, performing on par with dGSC-AD for the Penn Table $log(GDP)$ and log wages data, and outperforming the dGSC estimators for the remaining two data sets.

Thirdly, the oracle mlSC generally exhibits reduced bias relative to the classical SC estimator. Compared to dGSC-AD, however, there is no consistent ranking in bias, reflecting that bias reductions from mlSC depend on the underlying data structure and the degree of disaggregation.

6.4 Performance of Estimators under Different Noise Regimes

Building on the stylized example in Section 3.2 and motivated by the varying magnitude of performance improvements across data sets, I next examine how the benefit of using disaggregated data depends on its noise level. Using the same setup, I estimate a rank-three factor model and the idiosyncratic error variance as before, though for the state-level data sets I use a subset of the data (see Appendix E for details). I then inflate the estimated error variance by a multiplier m, so that in each simulated dataset, the variance equals $m \cdot \hat{\sigma}_{\varepsilon}^2$. Increasing m raises the relative magnitude of idiosyncratic noise compared to the signal, which enlarges

Table 4: Simulation results for four real data sets: RMSE and Bias.

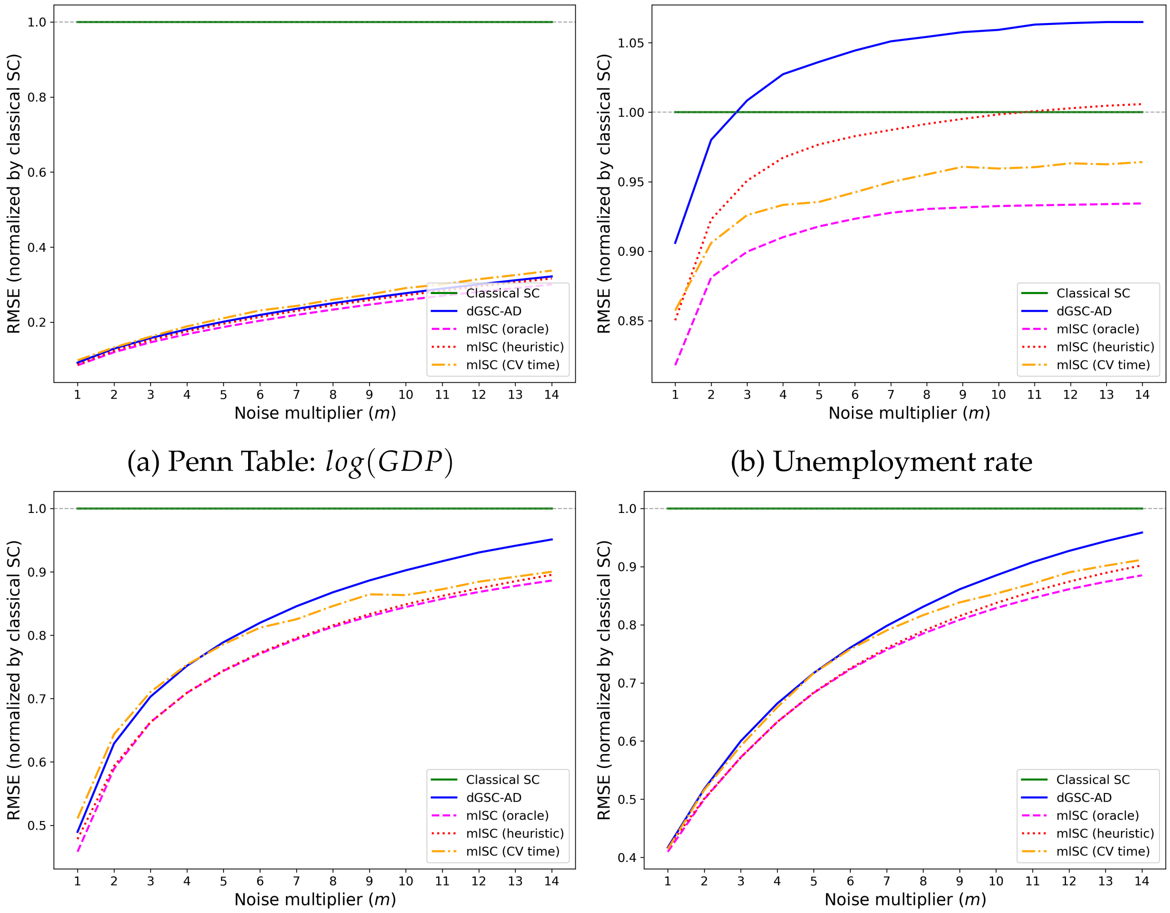
	mlSC (oracle)	mlSC (heuristic)	mlSC (CV time)	Classical SC	dGSC-AD
RMSE					
Assn.: Financial markets					
Penn table $log(GDP)$	0.034	0.035	0.036	0.384	0.035
Assn.: Min. wage					
Unemployment rate	0.081	0.083	0.083	0.089	0.095
Log wages	0.035	0.035	0.040	0.124	0.038
Smoking rate	0.046	0.046	0.049	0.153	0.056
Bias					
Assn.: Financial markets					
Penn table $log(GDP)$	0.002	0.000	0.005	0.209	-0.000
Assn.: Min. wage					
Unemployment rate	-0.001	-0.001	-0.005	0.004	-0.003
Log wages	0.002	0.002	0.004	0.030	0.003
Smoking rate	-0.009	-0.009	-0.010	-0.070	-0.012

All results are based on $S_{sim} = 1,000$ simulation runs. $N_{tr} = 1$ and $T_{post} = 1$. Outcomes are normalized to have mean zero and unit variance. In **bolt**: RMSE closest to oracle. *Estimators:* mlSC: multi-level SC estimator; classical SC estimator: aggregated data for treated and control; dGSC-AD: aggregated data for treated, disaggregated data for control.

the convex hull due to noise rather than meaningful variation, thereby increasing the risk of overfitting. Consequently, I expect the classical SC estimator to outperform dGSC-ADwhen the noise level is high.

Figure 4 reports results across different values of m for all four data sets. For the state-level data sets, I focus on the minimum-wage-law assignment, as alternative assignments produced similar patterns. As m increases, the classical SC estimator improves relative to dGSC-AD, and for the unemployment rate data set, it even surpasses dGSC-AD. The point at which the two lines cross depends not only on the absolute noise level, but also on the ratio of signal contained in the disaggregated data to the noise. Appendix C provides a decomposition of the aggregate, disaggregate, and noise components. The Penn Table data has the smallest noise component, while the smoking rate and log wages data have comparable noise levels. The unemployment rate exhibits substantially higher noise, explaining why the crossing occurs at a relatively low m; for the other data sets, the crossing point is not observed even at the largest multiplier considered ($m = 14$).

Across all noise scenarios, the oracle mlSC estimator consistently outperforms both the classical SC and dGSC-AD estimators. The largest gains occur when the classical SC and dGSC-AD estimators perform similarly, highlighting the ability of mlSC to adaptively leverage disaggregation depending on the signal-to-noise tradeoff.



(a) Penn Table: $log(GDP)$

(b) Unemployment rate

(c) Log wages

(d) Smoking rate

Figure 4: Varying noise multiplier $m \cdot \sigma_\varepsilon^2$ for data sets.

Estimators: mlSC: multi-level SC estimator; classical SC estimator: aggregated data for treated and control; dGSC-AD: aggregated data for treated, disaggregated data for control. Subset of states for the CPS unemployment rate and log wages and BRFSS smoking rate.

7 Theoretical Results

Work in Progress-section!

This section develops a formal framework to study the trade-off between the classical SC and the dGSC-AD estimator and the properties of the mlSC estimator. I introduce a hierarchical linear factor model that extends the standard latent factor model of Abadie et al. (2010) to (i) explicitly represent within–aggregate heterogeneity and (ii) allow random factor loadings. The model clarifies when disaggregation improves or worsens mean–squared error (MSE) relative to aggregation, and provides the building blocks for analyzing the multi-level

SC (mlSC) estimator.

7.1 Hierarchical Latent Factor Model

I adopt the standard linear latent factor model typically used to justify synthetic control methods (see, e.g. Abadie et al., 2010) to incorporate the disaggregate level data and mimic the hierarchical structure of the data.

Assumption 1 (Potential outcomes)**.** *Potential outcomes for disaggregated units c within aggregated units s at time t are given by*

$$\begin{aligned} Y_{sct}(0) &= \mu_{sc}'\beta_t + \varepsilon_{sct} = (\alpha_s + \eta_{sc})'\,\beta_t + \varepsilon_{sct} \\ Y_{sct}(1) &= \tau_{sct} + Y_{sct}(0); \end{aligned}$$

where $\beta_t \in \mathbb{R}^R$ denotes the R unknown latent factors, specific to each time period t. $\alpha_s \in \mathbb{R}^R$ denotes their unknown aggregate loadings, $\eta_{sc} \in \mathbb{R}^R$ the disaggregated deviations and ε_{sct} is an idiosyncratic error term. The factor loadings are time-invariant.

The potential outcomes for the aggregate level can be derived by taking weighted averages. I treat the factors and the unknown factor loadings as random instead of fixed as is mostly assumed in the literature, with the exception of Imbens and Viviano (2023) and Athey and Imbens (2025). The treatment assignment is still assumed to be fixed. For part of my theoretical analysis, I will condition on the random draws for the factor loadings since, for each observed sample, those will be fixed as well, even when I increase the number of time periods. Assumption 2 summarizes these assumptions.

Assumption 2 (Stochastic components)**.** *Factor loadings and factors are random:*

$$\alpha_s \overset{i.i.d.}{\sim} (0, \sigma_\alpha^2 I_R), \qquad \eta_{sc} \overset{i.i.d.}{\sim} (0, \sigma_\eta^2 I_R), \qquad \beta_t \overset{i.i.d.}{\sim} (0, \sigma_\beta^2 I_R), \qquad \varepsilon_{sct} \overset{i.i.d.}{\sim} (0, \sigma_\varepsilon^2).$$

Remark Note that this framework allows for a degree of agnosticism about the error term, e.g. instead of being purely stochastic, it can include a systematic component γ_{sc}, e.g. $\varepsilon_{sct} =$

$\gamma_{sc} + u_{sct}$. This interpretation absorbs nonlinear functions of the loadings $f_t(\alpha_s + \eta_{sc})$ into the "error." If γ_{sc} carries signal, e.g. is time-invariant and predictive out-of-sample, fitting this component is not purely overfitting, which favors disaggregated estimators.

7.2 Characterizing the Trade-Off under the Hierarchical Linear Latent Factor Model

In this section, I analyze the finite-sample mean-squared error (MSE) of the classical SC and the dGSC-AD estimator in the hierarchical linear latent factor model. I first show that the trade-off between the two estimators is driven by the increase in flexibility vs overfitting as opposed to the standard textbook bias-variance trade-off. Then, I use this result to provide insight into when disaggregation may be preferred to aggregation. For expositional purposes, I focus on the case of simple averages, i.e. $v_{sc} = \frac{1}{C_s}$ in this subsection.

7.2.1 Mean–Squared Error Decomposition

Assumption 3 ensures that factors (in additions to loadings) are stable across time, hence the same weight discrepancies that drives in-sample misfit also determines out-of-sample bias. This assumption justifies the transfer from in-sample to out-of-sample bias I will use in the following lemma.

Assumption 3 (Factor–structure stability)**.** *The post-treatment factor β_T is drawn from the same distribution as the pre-treatment factors $\{\beta_t : t \leq T_0\}$ and is independent of the idiosyncratic shocks ε_{sct}, $\beta_t \perp \varepsilon_{sct}$ for all s, c, t.*

Denote the vector of donor factor loadings by $M_{disagg} = \begin{pmatrix} \mu_{11} & \mu_{12} & \dots & \mu_{SC_S} \end{pmatrix}$ for the disaggregated weights and $M_{agg} = \begin{pmatrix} \mu_1 & \mu_2 & \dots & \mu_S \end{pmatrix}$ for the aggregated weights. For any SC-style estimator $\hat{w}$ solving Equation 4.1 define the post-treatment prediction error

$$E := Y_{0T}(0) - \hat{Y}_{0T}(0) = Y_{0T}(0) - Y'_{-0T}\hat{w},$$

where Y_{-0T} denotes the pre-treatment outcomes of all donor units and does not include the treated unit 0. Lemma 1 shows a decomposition of the out-of-sample MSEs of the classical SC and the dGSC-AD estimator.

Lemma 1 (MSE decomposition)**.** *Under Assumptions 1–3,*

$$E = \underbrace{(\mu_0 - M'w^*)'\beta_T}_{\textit{oracle bias}} + \underbrace{(M'(w^* - \hat{w}))'\beta_T}_{\textit{estimation bias}} + \underbrace{(\varepsilon_{0T} - \varepsilon'_{-0,T}\hat{w})}_{\textit{post-treatment noise}}.$$

Conditional on $\{\alpha_s, \eta_{sc}\}$ *the expected MSE is*

$$\mathbb{E}[E^2|\alpha_s, \eta_{sc}] = \underbrace{\|\mu_0 - M'w^*\|^2\sigma_\beta^2}_{\text{Oracle bias}} + \underbrace{\mathbb{E}\left[\|M'(w^* - \hat{w})\|^2\right]\sigma_\beta^2}_{\text{Estimation bias}} + \underbrace{\mathbb{E}[(\varepsilon_{0T} - \varepsilon'_{-0,T}\hat{w})^2]}_{\text{Variance}},$$

where cross–terms vanish by mean-zero and independence assumptions between β_T *and* ε_{scT}*.*

The proof is given in Appendix A.2. The MSE can be decomposed into three components. *Oracle bias* captures the imperfect factor replication even with optimal oracle weights and remains if the treated unit lies outside the donor convex hull. *Estimation bias* arises from fitting $\hat{w}$ on noisy pre-treatment data and generally does not vanish as $T_0 \to \infty$ when $\sigma_\varepsilon^2 > 0$ (Ferman and Pinto, 2021). Lastly, *post-treatment variance* reflects new shocks in period T and grows with the ℓ_2-norm of the estimated weights. Note that any trade-off between aggregation and disaggregation has to operate through one of these channels.

7.2.2 Flexibility vs Overfitting Trade-Off for the Classical SC and dGSC-AD Estimator

In this subsection, I show that the statistical trade-off between the classical SC estimator and the dGSC-AD estimator does not align with a textbook bias–variance trade-off. In a standard bias–variance framework, the variance component is typically monotone in model complexity, but that monotonicity need not hold here. While for any given weights w_{sc} and $w_s = \sum_{c=1}^{C_s} w_{sc}$, the weight vector norm is larger for the disaggregated vector w_{sc} than the weighted norm for the aggregated vector w_s, the weight vector I consider in the trade-off come from two separate optimization problems. This observation means that positive weights might be given to more units, suggesting that the overall norm of the weight vector might decrease. Instead, the relevant trade-off concerns the flexibility of the estimator—captured by the oracle bias—and the risk of overfitting, which can be further decomposed into two noise–driven components: estimation bias and post-treatment variance.

While I cannot make sharp statements about the relative size of the post-treatment variance across estimators, I can characterize the other two components: the dGSC-AD estimator weakly dominates the classical SC in terms of oracle bias, but its greater flexibility increases the potential estimation error. I formalize this reasoning in the next proposition.

Proposition 2 (General switching (crossing) condition)**.** *Let $\bar{w}_{agg}$ be the weights that minimize the difference between the aggregated factor loadings and $\bar{w}_{disagg}$ for the disaggregated factor loadings, i.e. solving Equation 4.1 with the correct additional constraints imposed. Suppose that $\|\hat{w}_{\text{agg}}\|_2^2 \neq \|\hat{w}_{\text{disagg}}\|_2^2$. Since the estimated weights depend on the variance noise σ^{*2}, i.e. $\hat{w}_{\cdot}(\sigma^{*2})$, the crossing point satisfies the implicit equation*

$$B_{\text{agg}}(\sigma^{*2}) - B_{\text{disagg}}(\sigma^{*2}) = \sigma^{*2}\big(\|\hat{w}_{\text{disagg}}(\sigma^{*2})\|_2^2 - \sum_{s=1}^{S} \hat{w}^{*2}_{\text{agg}}(\sigma^{*2})\frac{1}{C_s}\big),$$

where $B_{\cdot}(\sigma^{*2}) = \text{OracleBias}(\hat{w}_{\cdot}(\sigma^{*2}))^2 + \text{EstimationBias}(\hat{w}_{\cdot}(\sigma^{*2}))^2$ includes oracle and estimation bias terms.

No closed-form solution exists in Proposition 2 because the optimal weights themselves depend on σ_ε^2. Nevertheless, the implicit equation can be solved numerically or approximated by simulation to quantify the noise level at which disaggregation begins to outperform aggregation as I demonstrate in Section 6.4.

If the aggregated donor hull cannot closely reproduce the treated factor loading μ_0 but the richer disaggregated donor set can (i.e., $\text{OracleBias}(\hat{w}_{\text{disagg}}) \ll \text{OracleBias}(\hat{w}_{\text{agg}})$), the oracle bias gap is large and disaggregation remains advantageous even when the noise variance is moderately high. This oracle gap grows with between–aggregate heterogeneity (σ_α^2) and within–aggregate heterogeneity (σ_η^2), because both make the aggregate convex hull less capable of spanning the treated unit's loading, especially compared to the disaggregate one that can take advantage of the additional heterogeneity.

Conversely, estimating weights on noisy pre–treatment outcomes can introduce a persistent estimation bias at the disaggregated level, since the higher–dimensional weight vector is more susceptible to overfitting. This estimation component depends on the size of the error variance and on the concentration of the limiting weights, and can tilt the comparison back toward aggregation.

The scaling factor $\|\hat{w}_{\text{disagg}}\|_2^2 - \sum_{s=1}^{S} \hat{w}_{\text{agg}}^{*2} \frac{1}{C_s}$ in front of the variance captures the relative amplification of idiosyncratic shocks. If the disaggregated weights are more diffuse (larger ℓ_2–norm), the variance penalty of disaggregation is might be larger than the aggregated one, lowering the switching point and making aggregation preferable at a given noise level.

For fixed (deterministic) oracle weights, the implicit equation in Proposition 2 can be turned into an exact threshold (see Corollary 2): disaggregation dominates whenever the idiosyncratic noise variance σ_ε^2 is smaller than the bias reduction scaled by the difference in the concentration of the weights. Even when only disaggregation achieves a perfect synthetic control (so the aggregate oracle has positive bias), aggregated data may still be preferred if the reduction in bias is outweighed by the increase in variance. This threshold justifies why a data-driven penalty in mlSC targets the same boundary without any a priori choices of aggregation.

Corollary 2 (Oracle switching (crossing) condition)**.** *Let $\bar{w}_{\text{agg}}^*$ and $\bar{w}_{\text{dis}}^*$ denote oracle weights based on aggregated and disaggregated donor sets, and let $\hat{w}^{\text{agg}}, \hat{w}^{\text{dis}}$ be their sample estimates. Suppose that $\|\bar{w}^{\text{agg},*}\|_2^2 \neq \|\bar{w}^{\text{dis},*}\|_2^2$. Under Assumptions 1-3, disaggregation improves MSE relative to aggregation iff*

$$\sigma_\varepsilon^2 \; < \; s^* \; := \; \frac{\Delta_{\text{oracle}}^*}{\|\bar{w}_{\text{dis}}^*\|_2^2 - \sum_{s=1}^{S} \bar{w}_{\text{agg}}^{*2} \frac{1}{C_s}}, \tag{7.1}$$

where $\Delta_{oracle}^* = \|\mu_0 - M' w_{agg}^*\|^2 \sigma_\beta^2 - \|\mu_0 - M' w_{disagg}^*\|^2 \sigma_\beta^2$. If the denominator is negative the inequality reverses sign accordingly.

7.3 Guarantees for mlSC Estimator

I first show that under the standard assumptions of the classical SC estimator, the mlSC estimator always recovers the classical SC solution. This demonstrates that, when the classical SC estimator is favorable in practice—i.e., when pre-treatment fit is already good—the mlSC estimator returns the same estimates without any loss of efficiency or bias. To formalize this result, I introduce two assumptions. The first is the standard no-anticipation assumption, and the second is perfect pre-treatment fit, which is used in Abadie et al. (2010) to derive bias properties of the classical SC estimator.

Assumption 4 (Perfect pre-treatment fit for classical SC)**.** *There exist weights w^s such that $Y_{0t} = \sum_{s=1}^{S} w^s Y_{st}$ for all $t \leq T_0$, where w^s minimizes the classical SC objective in Equation 3.1.*

Proposition 3 (Recovery of Classical SC)**.** *Under Assumption 4, the mlSC and classical SC optimization problems have identical solutions, $\hat{w}^{mlSC} = \hat{w}^{SC}$, implying $\hat{\tau}^{mlSC} = \hat{\tau}^{SC}$.*

The proof, provided in Appendix A.4, relies on reformulating the mlSC objective in terms of aggregate-specific parameters and deviations, weighted appropriately. Overall, this observation reinforces that the mlSC estimator is a *superset* of the classical SC estimator. Perfect pre-treatment fit is crucial here, as it ensures that the optimization over aggregate-specific weights and deviations separates cleanly, so the mlSC penalties do not alter the solution.

Proposition 3 highlights an important property of the mlSC framework: it inherits all favorable properties of the classical SC estimator whenever the standard assumptions from Abadie et al. (2010) hold. Consequently, when pre-treatment fit is already excellent using aggregated data, the mlSC estimator will naturally select the classical SC solution. Conversely, if pre-treatment fit is poor, the mlSC estimator can gain by partially or fully disaggregating the control units, exploiting any distinguishable signal in the disaggregated data, without violating the classical SC simplex constraints on the weights. In other words, the potential value of mlSC arises precisely in settings where the classical SC assumptions are relaxed in practice—namely, when perfect pre-treatment fit is unattainable.

8 Two Applications of the Multi-Level SC Estimator: Revisiting Minnesota's e-Cigarette Tax and Iowa's Minimum Wage Increase

I illustrate the practical use of the mlSC estimator in two empirical settings. The goal of these applications is not to provide new causal estimates, but rather to demonstrate how the estimator adapts to the trade-offs between aggregated and disaggregated data, and how applied researchers can leverage this flexibility. In each example, I compare county- and state-level analyses and benchmark mlSC against classical SC, dGSC-AD, and standard DiD estimators. The first application revisits Deng and Zheng (2023), and the second follows Callaway and Sant'Anna (2021).

8.1 Minnesota's e-Cigarette Tax

I revisit Deng and Zheng (2023), who study the effects of Minnesota's 2013 cigarette and e-cigarette sales tax increases. In the paper, the authors focus on the e-cigarette sales tax on e-cigarettes and cigarette sales and prices. Following Amato et al. (2015), I focus on the impact of the cigarette sales tax on cigarette sales instead. The policy increased Minnesota's cigarette sales tax to \$1.75 per pack (from \$1.60 to \$3.35 per pack) in July 2013. Treatment is thus assigned at the state level.

My analysis uses NielsenIQ Retail Scanner Data (provided through Kilts Center for Marketing, University of Chicago), which records weekly sales at participating grocery stores. I construct outcomes at both the county- and state-level, taking average units sold per grocery store. To focus on the incremental effect of Minnesota's July 2013 tax, I restrict the donor pool to states without similar policy changes in 2013, yielding $S = 47$ control states covering 1378 counties.[13] I use all data in 2013, thus $T = 52$ weeks in total, with $T_0 = 26$ pre-treatment weeks.

[13] Note that this data set includes substantially less counties than the total number of counties in the US.

Figure 5 shows the deviation of counterfactual predictions from observed outcomes, $\Delta = Y_{MN} - \hat{Y}_{MN}$, for all estimators. The pre-treatment fit (Figure 5b) highlights that the DiD estimators exhibit poor pre-sample fit, signaling potential parallel trends violations. Classical SC improves fit over DiD, and the dGSC-AD estimator achieves nearly perfect pre-treatment fit. The two feasible mlSC estimators adapt between these two dGSC extremes, looking closer to the dGSC-AD than classical SC estimator though. Overall, using disaggregated data reduces the pre-treatment RMSE by 97.7-99.9% compared to the classical SC.[14]

Weight patterns further illustrate the trade-off: the dGSC-AD estimator assigns positive weight to 35 states, with a total of 120 counties (average $\bar{C}_s = 3.43$ per state), while classical SC assigns weight to only 11 states covering 288 counties, concentrating mostly in the Midwest. For more details on the weights for the two estimators, see Appendix I.0.1.

Table 5 summarizes estimated treatment effects. While all methods suggest a reduction in sales, magnitudes vary substantially, emphasizing sensitivity of the SC estimators to the aggregation level used. The cross-validation curve in Figure 6 guides the choice of λ: mlSC selects a penalty between the extremes, indicating that some aggregation helps stabilize noisy county-level outcomes, while still exploiting disaggregated information. The classical SC estimator beats the dGSC-AD estimator in the hold out pre-treatment data set. This finding echoes the approach taken in Deng and Zheng (2023). However, this application also shows how aggregation choice can change the estimated effect of the cigarette sales tax increase, which can lead to changes in policy implications.

8.2 Iowa's Minimum Wage Increase

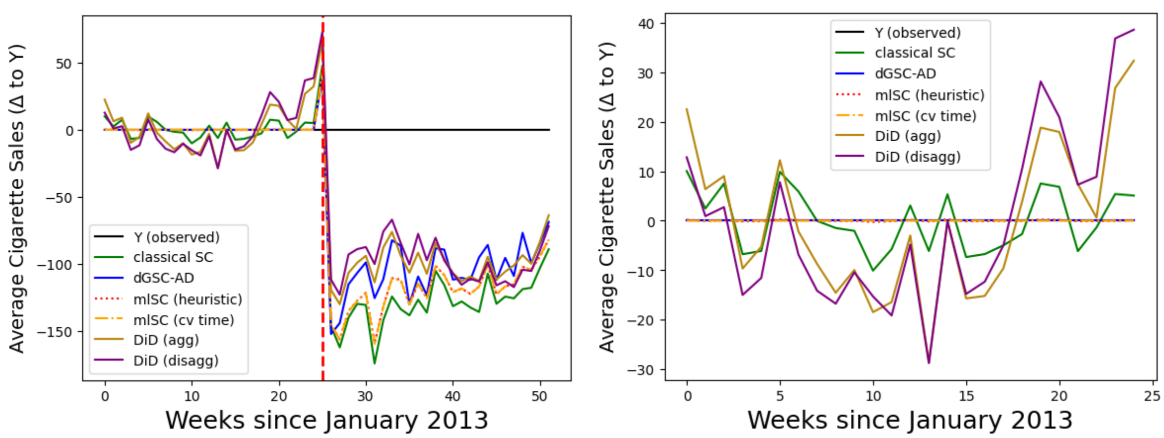
Following Callaway and Sant'Anna (2021), I also apply my method to evaluate the effect of a minimum wage increase on teen employment. Unlike their multi-state design, I focus on a single treated state, namely Iowa, during the period from 2001:Q2-2007:Q2. During this time, the federal minimum wage remained fixed at \$5.15 per hour. In 2007:Q2, Iowa implemented a state-level minimum wage increase to \$6.20, providing a clean setting to study the local

[14] The DiD estimators, on the contrary, have an increased pre-treatment RMSE comapred to the classical SC (approximately 2.5-2.8 times higher).

Table 5: Estimated treatment effects for MN from July-December 2013. Average cigarette sales per grocery store is $\bar{Y} = 816.48$.

Estimator	Estimated Treatment Effect
classical SC	-122.52
dGSC-AD	-99.66
mlSC (heuristic)	-113.30 ($\hat{\lambda} = 0.0463$)
mlSC (cv time; $t_{cv} = 5$)	-113.41 ($\hat{\lambda} = 0^{+}$)
DiD (aggregate)	-94.12
DiD (disaggregate)	-91.85

Notes: $\hat{\lambda} = 0^{+}$ means that $\hat{\lambda}$ is positive but very small. *Estimators:* classical SC estimator: aggregated data for treated and control; mlSC: multi-level SC estimator with penalty parameter estimate $\hat{\lambda}$; DiD (aggregate): difference-in-differences using aggregated data; DiD (disaggregate): difference-in-differences using disaggregated data.



(a) All time periods

(b) Pre-treatment fit

Figure 5: Treated State: MN in July 2013: Outcome Data and Predicted Values from SC and DiD estimators. Outcomes are given relative to $Y_{MN,t}$ ($\Delta = Y_{MN,t} - \hat{Y}_{MN,t}$)

Estimators: classical SC estimator: aggregated data for treated and control; mlSC: multi-level SC estimator with penalty parameter estimate $\hat{\lambda}$; DiD (aggregate): difference-in-differences using aggregated data; DiD (disaggregate): difference-in-differences using disaggregated data.

labor market impact of this policy shift. Hence, treatment is assigned at the state level. While the authors use a difference-in-differences set-up and leave the outcome data at the county-level to estimate the aggregate effect, I explore whether I should have aggregated the data, given the interest in the treatment effect of the policy on Iowa as a whole.

My outcome variable is the teen employment rate. The outcome is measured quarterly using data from the Quarterly Workforce Indicators (QWI) compiled by Dube and Zipperer

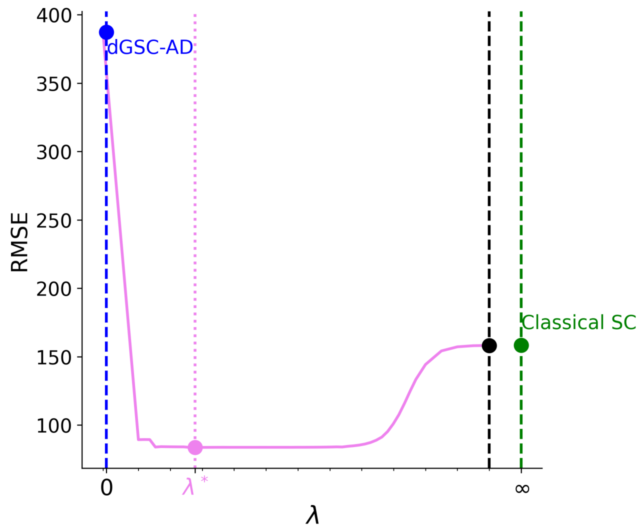


Figure 6: Total RMSE curve as a function of λ for the $t_{cv} = 5$ hold out pre-treatment periods preceding treatment.

Estimators: classical SC estimator: aggregated data for treated and control; dGSC-AD: aggregated data for treated, disaggregated data for control

(2015), which is given at the county-level.[15] I restrict the donor pool to control states that did not increase their minimum wage above the federal level during the sample period and for which county-level data is available. This results in a donor pool of $S = 13$ control states, comprising a total of 1141 counties, observed over $T = 25$ quarters ($T_0 = 24$ pre-treatment periods).

Figure 7 plots the counterfactual paths of my proposed mlSC estimators and the benchmark estimators in terms of deviations from the observed outcome Y (see Figure 7a for all pre-treatment periods and Figure 7b for pre-treatment only). Focusing on the pre-treatment fit, I see large differences in pre-treatment fit. The classical SC estimator's pre-treatment fit is relatively poor and even worse than the DiD estimators' fit. The dGSC-AD estimator's fit is almost perfect in this application as well. The feasible mlSC estimators' fits lie close to the dGSC-AD estimator, but in between the two dGSC estimators again, as expected.

Figure 7 shows deviations of predicted outcomes from observed values for all estimators.

[15]The teen employment rate is computed as the number of employed teens divided by the teen population in each county. I winsorize the variable at the 0.5% level to reduce the influence of extreme outliers and counteract measurement error.

Pre-treatment fit (Figure 7b) is poor for classical SC and DiD, with the classical SC performing the worst. Because Iowa's teenage employment rate exceeds that of most donor states in many quarters, the convexity constraints of classical SC make it difficult to match Iowa's trajectory. This is precisely the type of setting where disaggregation can provide substantial gains. The dGSC-AD and mlSC estimators achieve near-perfect pre-treatment fit, reducing the pre-treatment RMSE by 99.7-99.9% compared to the classical SC.[16]

In terms of weights, the dGSC-AD places positive weight on 11 states, with a total of 115 counties ($\bar{C}_s = 10.45$ per state)[17] , while classical SC assigns weight to two states only. However, those two states contain 133 counties in total, suggesting that, despite the state-level aggregation, the number of contributing counties in both approaches is roughly similar. For more details on the weights for the two estimators, see Appendix I.0.2.

Table 6 presents estimated treatment effects. All estimators predict negative effects on teen employment, but magnitudes vary, just like in the first application. The mlSC cross-validation curve (Figure 8) again guides the choice of λ: it selects a $\hat{\lambda}^*$ close to dGSC-ADin terms of RMSE, suggesting that disaggregated control data is particularly informative here, while some aggregation might still improve stability. Overall, I find that the dGSC-AD estimator beats the classical SC estimator for the hold out pre-treatment data set. This application illustrates how aggregation choice can change the estimated effect of increasing minimum wage laws in Iowa, a question of direct relevance to labor economics.

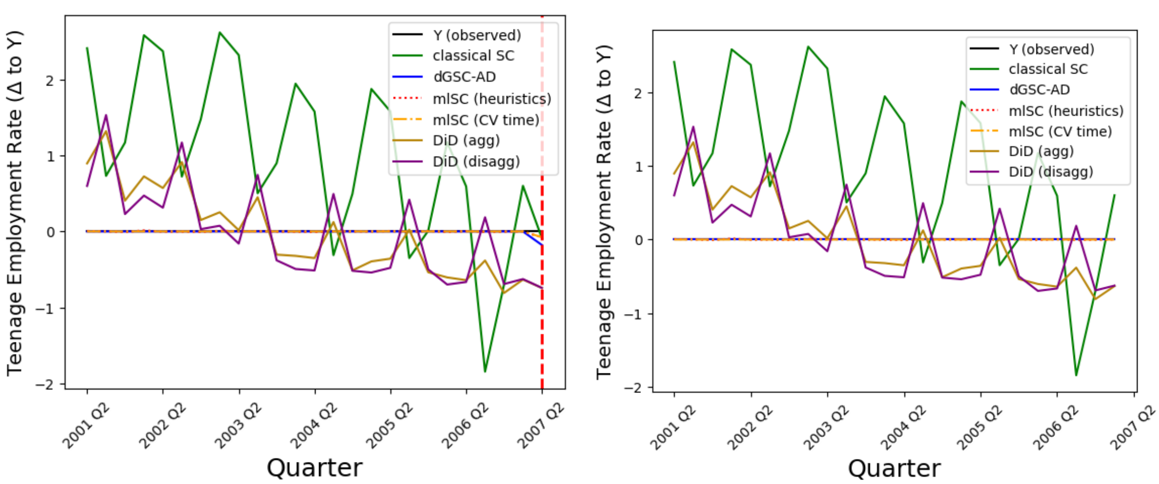
[16]The DiD estimators reduce the pre-treatment RMSE by around 60% compared to the classical SC.

[17]"Substantial" is defined as any weight above an equal weighting scheme, so $w_{sc} = \frac{1}{1141} = 0.009$

Table 6: Estimated treatment effects for IA in 2007 Q2. Outcome: Teen Employment Rate (in %). Average teen employment rate is 14.57%

Estimator	Estimated Treatment Effect
classical SC	-0.089
dGSC-AD	-0.179
mlSC (heuristic)	-0.077($\hat{\lambda} = 0.4855$)
mlSC (cv time; $t_{cv} = 4$)	-0.075 ($\hat{\lambda} = 0.0001$)
DiD (aggregate)	-0.744
DiD (disaggregate)	-0.743

Estimators: classical SC estimator: aggregated data for treated and control; mlSC: multi-level SC estimator with penalty parameter estimate $\hat{\lambda}$; DiD (aggregate): difference-in-differences using aggregated data; DiD (disaggregate): difference-in-differences using disaggregated data.



(a) All time periods

(b) Pre-treatment fit

Figure 7: Treated State: Iowa in 2007 Q2: Outcome Data and Predicted Values from SC and DiD estimators. Outcomes are given relative to $Y_{IA,t}$ ($\Delta = Y_{IA,t} - \hat{Y}_{IA,t}$)

Estimators: classical SC estimator: aggregated data for treated and control; mlSC: multi-level SC estimator with penalty parameter estimate $\hat{\lambda}$; DiD (aggregate): difference-in-differences using aggregated data; DiD (disaggregate): difference-in-differences using disaggregated data.

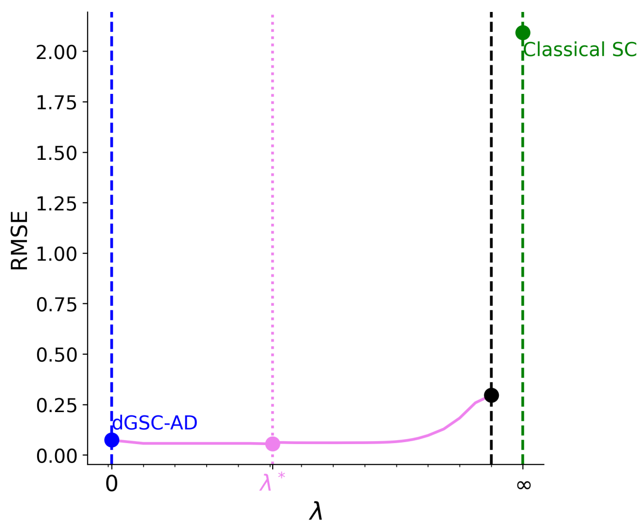


Figure 8: Total RMSE curve as a function of λ for the $t_{cv} = 5$ hold out pre-treatment periods preceding treatment.

Estimators: classical SC estimator: aggregated data for treated and control; dGSC-AD: aggregated data for treated, disaggregated data for control

9 Extending the Setting: Disaggregating the Treated Unit

While earlier sections show that disaggregating the treated unit offers little gain in estimator performance when the target is an aggregate-level effect, such disaggregation substantially broadens the scope of possible analyses. For example, Bottmer (2025) demonstrates that disaggregating the treated unit allows for generalization of treatment effect estimates in classical SC settings, particularly when the setting is limited to a single observation. In this paper, I further show that disaggregating the treated unit enables the study of treatment effect heterogeneity, aligning with the goals of the distributional synthetic control literature (see, e.g., Chen, 2020; Gunsilius, 2023). Thus, whereas the value of disaggregating control units derives mainly from a statistical trade-off, the value of disaggregating the treated unit lies in expanding the set of estimands, not in improving conventional predictive performance.

9.1 Set-Up

In panel data setting considered in this paper, disaggregating the treated unit makes it possible to estimate unit-level treatment effects rather than only the aggregate average. Unlike the distributional synthetic control literature, which characterizes the distribution of effects across units (e.g., Chen, 2020; Gunsilius, 2023), this framework allows direct inference at the disaggregate level. Such granularity can reveal heterogeneous responses and facilitate mechanism analysis, rather than recovering only an aggregate-level distribution.

As shown in Section 2, the aggregate effect of interest can be written as the average of the unit-level effects,

$$\tau_{0t} = \sum_{c'=1}^{C_0} \tau_{0c't}.$$

The dGSC estimator solves for synthetic control weights separately for each treated subunit, yielding

$$\hat{\tau}_{0t}^{dGSC} = \frac{1}{C_0} \sum_{c'=1}^{C_0} \hat{\tau}_{0c't}, \quad \hat{\tau}_{0c't} = Y_{0c't} - \sum_{s=1}^{S} \sum_{c=1}^{C_s} \hat{W}_{c'sc} Y_{sct}.$$

Thus, running the dGSC estimator directly provides estimates of county-level treatment effects.

9.2 Simulation Results for Multi-Level SC Estimator with Fully Disaggregated Data

I next extend the simulation study of Section 6 to evaluate the mlSC estimator with both hierarchical penalties. I report results for the oracle mlSC (trained on post-treatment outcomes) and the cross-validated version. Results for the heuristic penalty do not translate directly to this two-penalty setting as, in the very stylized model, disaggregation of the treated unit plays no role since choosing λ_2 is irrelevant for the expected out-of-sample MSE (see Appendix J for more details).

to do if simulations will ever finish?

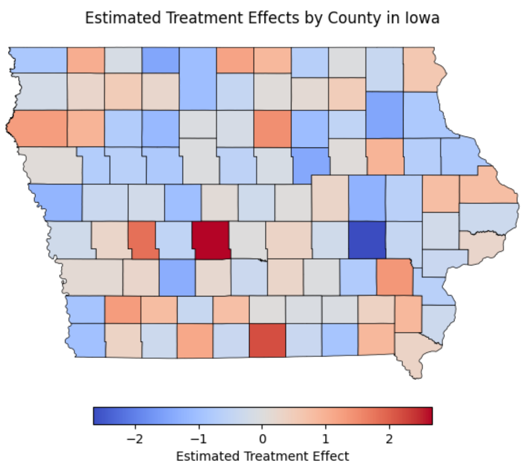
9.3 Disaggregating the Treated Unit in Iowa's Minimum Wage Increase

Finally, I revisit the second empirical application, the effect of Iowa's increase in minimum wage on teenage employment, allowing for county-level heterogeneity. Figure 9 shows the estimated county-level effects. Panel (a) maps the estimates, which display no clear geographic clustering, while panel (b) plots their distribution. The dGSC estimator yields a state-level effect of $\hat{\tau}^{dGSC} = -0.061$, closely matching the classical SC estimate. closely matching the classical SC estimate. Consistent with this modest average effect, the histogram is tightly centered near zero, though a few counties exhibit more extreme responses.

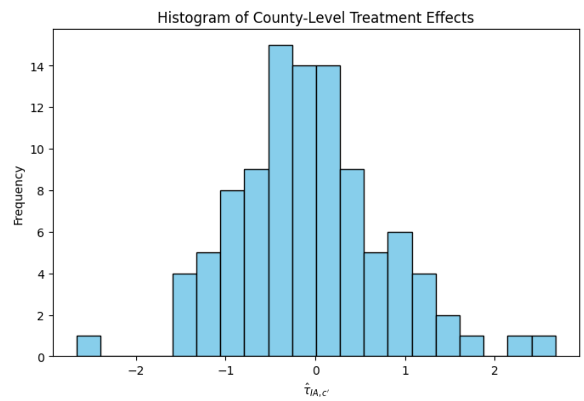
Table 7: Simulation results for four real data sets: RMSE and Bias. Subset of states: $S = 20$

	mlSC (oracle)	**mlSC (CV time)**	**Classical SC**	**dGSC**
RMSE				
Assn.: Financial markets				
Penn table $log(GDP)$				
Assn.: Min. wage				
Unemployment rate				
Log wages				
Smoking rate				
Bias				
Assn.: Financial markets				
Penn table $log(GDP)$				
Assn.: Min. wage				
Unemployment rate				
Log wages				
Smoking rate				

All results are based on $S_{sim} = 1,000$ simulation runs. $N_{tr} = 1$ and $T_{post} = 1$. Outcomes are normalized to have mean zero and unit variance. In **bolt**: RMSE closest to oracle. *Estimators:* mlSC: multi-level SC estimator with two penalty terms; classical SC estimator: aggregated data for treated and control; dGSC: disaggregated data for treated and control.



(a) Map of Iowa's counties including the estimated treatment effects



(b) Histogram of county-level treatment effects

Figure 9: Treated state: Iowa in 2007 Q2: Estimated treatment effects by county.

10 Conclusion

In this paper, I investigate the value of disaggregated data in synthetic control applications and provide guidance for applied researchers working with multiple levels of aggregation. I develop a framework that nests different SC variants used in practice and transforms the choice of aggregation from an a priori modeling decision into a data-driven optimization problem. I introduce the multi-level SC estimator, which implements this approach by incorporating disaggregated data directly into the analysis and leveraging the additional variation in control units to improve aggregate treatment effect estimation. While disaggregating the treated unit does not necessarily enhance estimation accuracy, it enables researchers to examine new objects of interest, such as treatment effect heterogeneity. Together, these results demonstrate how researchers can use disaggregated data to improve estimation precision and broaden the range of policy-relevant estimands.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of economic literature*, 59(2):391–425.

Abadie, A., 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(490):493–505.

Abadie, A., Diamond, A., and Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2):495–510.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American economic review*, 93(1):113–132.

Abadie, A. and L'Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116(536):1817–1834.

Amato, M. S., Boyle, R. G., and Brock, B. (2015). Higher price, fewer packs: evaluating a tobacco tax increase with cigarette sales data. *American journal of public health*, 105(3):e5–e8.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.

Arkhangelsky, D. and Hirshberg, D. (2023). Large-sample properties of the synthetic control method under selection on unobservables. *arXiv preprint arXiv:2311.13575*.

Ashenfelter, O. C. and Card, D. (1984). Using the longitudinal structure of earnings to estimate the effect of training programs.

Athey, S. and Imbens, G. (2025). Identification of average treatment effects in nonparametric panel models. *arXiv preprint arXiv:2503.19873*.

Athey, S. and Imbens, G. W. (2006). Identification and inference in nonlinear difference-in-differences models. *Econometrica*, 74(2):431–497.

Athey, S. and Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11(1):685–725.

Athey, S. and Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, 226(1):62–79.

Baum, C. L. and Ruhm, C. J. (2016). The effects of paid family leave in california on labor market outcomes. *Journal of Policy Analysis and Management*, 35(2):333–356.

Ben-Michael, E., Feller, A., and Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, 116(536):1789–1803.

Ben-Michael, E., Feller, A., and Rothstein, J. (2022). Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(2):351–381.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275.

Bottmer, L. (2025). From policy evaluation to generalizing treatment effects: A framework using synthetic control. Unpublished manuscript.

Bottmer, L., Imbens, G. W., Spiess, J., and Warnick, M. (2024). A design-based perspective on synthetic control methods. *Journal of Business & Economic Statistics*, 42(2):762–773.

Callaway, B. and Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, 225(2):200–230.

Card, D. (1990). The impact of the mariel boatlift on the miami labor market. *Ilr Review*, 43(2):245–257.

Card, D. and Krueger, A. B. (1994). American economic association. *The American Economic Review*, 84(4):772–793.

Chen, Y.-T. (2020). A distributional synthetic control method for policy evaluation. *Journal of Applied Econometrics*, 35(5):505–525.

Deng, X. and Zheng, Y. (2023). Estimating the effects of e-cigarette taxes: a generalized synthetic control approach. *Applied Economics*, pages 1–16.

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

Dube, A. and Zipperer, B. (2015). *Pooling multiple case studies using synthetic controls: An application to minimum wage policies*, volume 8944. SSRN.

Dwyer-Lindgren, L., Mokdad, A. H., Srebotnjak, T., Flaxman, A. D., Hansen, G. M., and Murray, C. J. (2014). Cigarette smoking prevalence in us counties: 1996-2012. *Population health metrics*, 12:1–13.

Ferman, B. and Pinto, C. (2021). Synthetic controls with imperfect pretreatment fit. *Quantitative Economics*, 12(4):1197–1221.

Freyaldenhoven, S., Hansen, C., and Shapiro, J. M. (2019). Pre-event trends in the panel event-study design. *American Economic Review*, 109(9):3307–3338.

Ghanem, D., Sant'Anna, P. H., and Wüthrich, K. (2022). Selection and parallel trends. *arXiv preprint arXiv:2203.09001*.

Gunsilius, F. F. (2023). Distributional synthetic controls. *Econometrica*, 91(3):1105–1117.

Hanushek, E. A., Luo, J., Morgan, A. J., Nguyen, M., Ost, B., Rivkin, S. G., and Shakeel, A. (2023). The effects of comprehensive educator evaluation and pay reform on achievement. Technical report, National Bureau of Economic Research.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press.

Imbens, G. W. and Viviano, D. (2023). Identification and inference for synthetic controls with confounding. *arXiv preprint arXiv:2312.00955*.

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(3):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(536):1804–1816.

Kreif, N., Grieve, R., Hangartner, D., Turner, A. J., Nikolova, S., and Sutton, M. (2016). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health economics*, 25(12):1514–1528.

Neumark, D. and Wascher, W. (2001). Minimum wages and training revisited. *Journal of Labor Economics*, 19(3):563–595.

Pac, J. E., Bartel, A. P., Ruhm, C. J., and Waldfogel, J. (2019). Paid family leave and breastfeeding: Evidence from california. Technical report, National Bureau of Economic Research.

Pouliot, G. A. and Xie, Z. (2022). Degrees of freedom and information criteria for the synthetic control method. *arXiv preprint arXiv:2207.02943*.

Rambachan, A. and Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, 90(5):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(517):109–126.

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3):305–322.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Shen, D., Ding, P., Sekhon, J., and Yu, B. (2023). Same root different leaves: Time series and cross-sectional methods in panel data. *Econometrica*, 91(6):2125–2154.

Shi, C., Sridhar, D., Misra, V., and Blei, D. (2022). On the assumptions of synthetic control methods. In *International Conference on Artificial Intelligence and Statistics*, pages 7163–7175. PMLR.

Spiess, J., Venugopal, A., et al. (2023). Double and single descent in causal inference with an application to high-dimensional synthetic control. *Advances in Neural Information Processing Systems*, 36:63642–63659.

Sun, L., Ben-Michael, E., and Feller, A. (2024). Temporal aggregation for the synthetic control method. In *AEA Papers and Proceedings*, volume 114, pages 614–617. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Appendices

Appendix A Proofs

A.1 Proof of Proposition 1

Let $\mathcal{R} \subseteq \mathcal{R}^{dGSC-AD}$. I show that the optimization problem for the dGSC estimators imposing $\mathcal{R} \subseteq \mathcal{R}^{dGSC-AD}$ is equivalent to the classical SC objective function. I start with the optimization problem for the dGSC estimators. Note that, by enforcing the constraints on the weights given in $\mathcal{R}$, the weights $W_{c'sc}$ become equivalent for each disaggregated treated unit c', making the weight matrix independent of c', thus $W_{c'sc} = W_{\cdot sc}$. I will use this observation in the rewriting of the optimization problem:

$$
\begin{aligned}
&\arg\min_{W_{\cdot sc}\in\mathcal{R}} \sum_{t=1}^{T_0}\sum_{c'=1}^{C_0} v_{0c'}(Y_{0c't} - \sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct})^2 \\
&= \arg\min_{W_{\cdot sc}\in\mathcal{R}} \sum_{t=1}^{T_0}\sum_{c'=1}^{C_0} v_{0c'}\left(Y_{0c't}^2 - 2\cdot Y_{0c't}\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct} + (\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct})^2\right) \\
&= \arg\min_{W_{\cdot sc}\in\mathcal{R}} \sum_{t=1}^{T_0}\left(\sum_{c'=1}^{C_0} v_{0c'}\; Y_{0c't}^2 - \sum_{c'=1}^{C_0} 2\cdot v_{0c'}\; Y_{0c't}\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct} + \sum_{c'=1}^{C_0} v_{0c'}\; (\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct})^2\right) \\
&= \arg\min_{W_{\cdot sc}\in\mathcal{R}} \sum_{t=1}^{T_0}\left(\sum_{c'=1}^{C_0} v_{0c'}\; Y_{0c't}^2 - 2\cdot\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct}\cdot \underbrace{\sum_{c'=1}^{C_0} v_{0c'}Y_{0c't}}_{Y_{0t}} + (\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct})^2 \underbrace{\sum_{c'=1}^{C_0} v_{0c'}}_{=1}\right) \\
&= \arg\min_{W_{\cdot sc}\in\mathcal{R}} \sum_{t=1}^{T_0}\left(-2\cdot\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct}\cdot Y_{0t} + (\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct})^2\right) \\
&= \arg\min_{W_{\cdot sc}\in\mathcal{R}} \sum_{t=1}^{T_0}\left(-2\cdot\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct}\cdot\; Y_{0t} + (\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct})^2\right) \\
&= \arg\min_{W_{\cdot sc}\in\mathcal{R}} \sum_{t=1}^{T_0}\left(-2\cdot\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct}\cdot\; Y_{0t} + (\sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct})^2 + Y_{0t}^2\right) \\
&= \arg\min_{W_{\cdot sc}\in\mathcal{R}} \sum_{t=1}^{T_0}(Y_{0t} - \sum_{s=1}^{S}\sum_{c=1}^{C_s} W_{\cdot sc}Y_{sct})^2.
\end{aligned}
$$

This optimization problem is, up to a reindexing of donor units, equivalent to the classical SC problem given in Equation 3.1, where the treated unit is kept at the aggregate level.

A.2 Proof of Lemma 1

Under Assumption 1, the post-treatment error can be written as

$$E = (\mu_0'\beta_T + \varepsilon_{0T}) - (M'\hat{w})'\beta_T - \varepsilon_{-0,T}'\hat{w}.$$

Add and subtract the oracle synthetic control $M'w^*$ to obtain

$$E = (\mu_0 - M'w^*)'\beta_T + (M'(w^* - \hat{w}))'\beta_T + (\varepsilon_{0T} - \varepsilon_{-0,T}'\hat{w}).$$

This gives the decomposition in the statement. Then, conditional on $\{\alpha_s, \eta_{sc}\}$, β_T is mean zero with variance σ_β^2 and is independent of ε_{0T} and $\varepsilon_{-0,T}$. Thus,

$$\mathbb{E}[E^2|\mu] = \sigma_\beta^2\|\mu_0 - M'w^*\|^2 + \sigma_\beta^2\mathbb{E}[\|M'(w^* - \hat{w})\|^2] + \mathbb{E}[(\varepsilon_{0T} - \varepsilon_{-0,T}'\hat{w})^2],$$

since cross–terms have zero expectation.

A.3 Proof of Proposition 2

Apply Lemma 1 to the two estimators of interest, classical SC and dGSC-AD estimator:

- $\hat{w}_{\text{agg}}$: weight vector for the classical SC: $MSE_{agg} = \sigma_\beta^2\|\mu_0 - M'w_{\text{agg}}^*\|^2 + \sigma_\beta^2\mathbb{E}[\|M'(w_{\text{agg}}^* - \hat{w}_{\text{agg}})\|^2] + \sigma_\varepsilon^2(\frac{1}{C_0} + \sum_{s=1}^S \hat{w}_{\text{agg}}^2\frac{1}{C_s})$
- $\hat{w}_{\text{dis}}$: weight vector for the dGSC-ADestimator: $MSE_{disagg} = \sigma_\beta^2\|\mu_0 - M'w_{\text{disagg}}^*\|^2 + \sigma_\beta^2\mathbb{E}[\|M'(w_{\text{disagg}}^* - \hat{w}_{\text{disagg}})\|^2] + \sigma_\varepsilon^2(\frac{1}{C_0} + \|\hat{w}_{\text{dis}}\|_2^2)$

Define the difference in MSE

$$\Delta\text{MSE} = \underbrace{\Big(\text{OracleBias}(\hat{w}_{\text{agg}})^2 - \text{OracleBias}(\hat{w}_{\text{dis}})^2\Big)}_{\Delta_{\text{oracle}}}$$

$$+ \underbrace{\left(\text{EstimationBias}(\hat{w}_{\text{agg}})^2 - \text{EstimationBias}(\hat{w}_{\text{dis}})^2\right)}_{\Delta_{\text{est}}}$$

$$+ \sigma_\varepsilon^2 \Big(\sum_{s=1}^{S} \hat{w}_{\text{agg}}^2 \frac{1}{C_s} - \|\hat{w}_{\text{dis}}\|_2^2 \Big).$$

The implicit switching point is when this equation is set to 0, thus resulting in the following equation

$$\delta_{oracle} + \Delta_{est} + \sigma^{*2} \Big(\sum_{s=1}^{S} \hat{w}_{\text{agg}}^2 \frac{1}{C_s} - \|\hat{w}_{\text{dis}}\|_2^2 \Big) = 0.$$

Defining $B. = \text{OracleBias}(\hat{w}.)^2 + \text{EstimationBias}(\hat{w}.)$ and rearranging yields the result in Proposition 2. This expression cannot be further simplified as the estimated weights depend on the error variance.

A.4 Proof of Proposition 3

Step 1: Reparameterization

The mlSC estimator can be rewritten in terms of deviations from the aggregated outcome and aggregated weights. Define the aggregated weight for state s from the mlSC weight as $w_s^{agg} = \sum_{c=1}^{C_s} w_{sc}$ and the deviations from the average county-level weight within each state as $u_{sc} = w_{sc} - v_{sc} w_s^{agg}$. Similarly, define deviations from the aggregated outcome as $Z_{sct} = Y_{sct} - Y_{st}$. By construction, $\sum_c u_{sc} = \sum_c v_{sc} Z_{sct} = 0$. Using these definitions, the synthetic control estimate can be rewritten as

$$\sum_{s=1}^{S} \sum_{c=1}^{C_s} w_{sc} Y_{sct} = \sum_{s=1}^{S} [w_s^{agg} Y_{st} + \sum_{c=1}^{C_s} u_{sc} Z_{sct}] \quad \& \quad \sum_{s=1}^{S} \sum_{c=1}^{C_s} (w_{sc} - v_{sc} w_s^{agg})^2 = \sum_{s=1}^{S} \sum_{c=1}^{C_s} u_{sc}^2.$$

With this reparametrization, the mlSC optimization problem becomes:

$$\begin{aligned}
\arg \min_{w_s \in \mathbb{R}^S, u_{sc} \in \mathbb{R}^{\sum_{c=1}^{C_s} C_s}} \quad & \sum_{t=1}^{T_0} \Big(Y_{0t} - \sum_{s=1}^{S} \underbrace{w_s}_{=w_s^{agg}} Y_{st} - \sum_{s=1}^{S} \sum_{c=1}^{C_s} u_{sc} Z_{sct}\Big)^2 \\
& + \lambda_1 \sum_{s=1}^{S} \sum_{c=1}^{C_s} u_{sc}^2 \qquad \qquad (A.1) \\
s.t. \quad & \sum_{s=1}^{S} \underbrace{w_s}_{=w_s^{agg}} = 1, \sum_{s=1}^{S} \sum_{c=1}^{C_s} u_{sc} = 0 \quad \text{and} \quad w_s \geq 0 \; \forall s \neq 0,
\end{aligned}$$

where w_s is a vector for aggregate specific weights and u_{sc} are the within-aggregate deviations. The objective can be decomposed into three components: (i) a term depending only on w_s, (ii) a term depending only on u_{sc}, and (iii) a cross-term.

Step 2: Candidate solution

Consider the candidate solution $w_s = w_s^*$, which achieves perfect pre-treatment fit under Assumption 4, and $u_{sc} = 0 \; \forall s, c$, i.e., all counties in a state have equal weights. I now proceed to show that this candidate solution achieves the unique minimum for this optimization problem.

Step 3: Check constraints

The candidate solution satisfies all constraints: (1) $\sum_{s=1}^{S} w_s^* = 1, w_s^* \geq 0$ hold because w_s^* is the solution to the classical SC problem, Equation 3.1, which imposes exactly those constraints and (2) $\sum_{s=1}^{S} \sum_{c=1}^{C_s} u_{sc} = \sum_{s=1}^{S} \sum_{c=1}^{C_s} 0 = 0$, hence also satisfying the constraint on the u_{sc}s.

Step 4: Evaluate the objective at the candidate

Let $L(w_s, u_{sc})$ be the mlSC objective function. Substituting $u_{sc} = 0 \; \forall s, c$ gives

$$L(w_s^*, 0) = \sum_{t=1}^{T_0} (Y_{0t} - \sum_{s=1}^{S} w_s^* Y_{st})^2 + 0 = \sum_{t=1}^{T_0} (Y_{0t} - \sum_{s=1}^{S} w_s^* Y_{st})^2.$$

Under perfect pre-treatment fit (Assumption 4), $L(w_s^*, 0) = 0$. The general mlSC objective can be decomposed into two terms: (1) the pre-treatment fit, $(Y_{0t} - \sum_{s=1}^{S} w_s Y_{st} - \sum_{s=1}^{S} \sum_{c=1}^{C_s} u_{sc} Z_{sct})^2$, and (2) the penalty, $\lambda_1 \sum_{s=1}^{S} \sum_{c=1}^{C_s} u_{sc}^2$. Since both terms are quadratic and $\lambda_1 \geq 0$, this implies that the objective function for any candidate solution w, u is positive, $L(w, u) \geq 0$. Hence, since $L(w_s^*, 0) = 0$, the candidate solution obtains the minimum that the objective function can take on. Moreover, since both terms are convex, the minimum is unique.

Step 5: Connection to classical SC

With $u_{sc} = 0$, the mlSC optimization problem reduces to

$$\begin{aligned} \arg \min_{w_s \in \mathbb{R}^S, u_{sc} \in \mathbb{R}^{\sum_{c=1}^{C_s} C_s}} \quad & \sum_{t=1}^{T_0} \left(Y_{0t} - \sum_{s=1}^{S} w_s Y_{st}\right)^2 \\ s.t. \quad & \sum_{s=1}^{S} w_s = 1, \quad \text{and} \quad w_s \geq 0 \; \forall s \neq 0, \end{aligned} \tag{A.2}$$

which is exactly equivalent to the classical SC optimization problem (Equation 3.1).

Appendix B Derivation of the Optimal λ^* in a Stylized Model

This appendix derives the optimal penalty parameter λ^* in a stylized setting with two aggregated units ($S = 2$) and two disaggregated units within each aggregate $C_s = 2$). Aggregated unit 0 is treated and aggregated unit 1 serves as the donor unit. Both aggregated outcomes are simple averages of their disaggregated components.

Step 1: Set-Up and Reparameterization

Let the two disaggregated units in the donor aggregated unit 1 receive weights $(w, 1 - w)$ since the convexity constraint requires that the weights sum to 1. I ignore the positivity constraints for now. When using only aggregated data, the solution places equal weight on each disaggregated donor unit, $w = \frac{1}{2}$). Write the deviation from the aggregate-only solution as

$$w = \frac{1}{2} - \Delta, \quad 1 - w = \frac{1}{2} + \Delta.$$

The mlSC objective can then be expressed as

$$\arg\min_{w} \, (Y_{01} - wY_{111} - (1-w)Y_{121})^2 + \lambda \left((w - \frac{1}{2})^2 + (1 - w - \frac{1}{2})^2 \right)$$

$$\iff \arg\min_{\Delta} (Y_{01} - (\frac{1}{2} - \Delta)Y_{111} - (\frac{1}{2} + \Delta)Y_{121})^2 + \lambda 2\, \Delta^2$$

$$\iff \arg\min_{\Delta} (Y_{01} - Y_{11} - \Delta(Y_{121} - Y_{111}))^2 + \lambda 2\, \Delta^2$$

$$\iff \arg\min_{\Delta} (\bar{d}_1 - \Delta(d_{11}))^2 + \lambda 2\, \Delta^2,$$

where $\bar{d}_t = Y_{0t} - Y_{1,t}$ and $d_{st} = Y_{s2t} - Y_{s1t}$ are defined as the between-aggregated units and within-aggregated unit differences.

Step 2: First-Order Condition

Differentiating and solving for Δ gives

$$-2(\bar{d}_1 - \Delta d_{11}) \cdot d_{11} + 4\lambda\Delta = 0$$

$$\iff \bar{d}_1 d_{11} - \Delta d_{11}^2 - 2\lambda\Delta = 0$$

$$\iff \Delta(d_{11}^2 + 2\lambda) = \bar{d}_1\, d_{11}$$

$$\iff \hat{\Delta} = \frac{\bar{d}_1 d_{11}}{d_{11}^2 + 2\lambda}$$

Step 3: Out-of-Sample Mean-Squared Error

The optimal λ minimizes the expected out-of-sample mean squared error (MSE),

$$\mathbb{E}[MSE] = \mathbb{E}[(\bar{d}_2 - \hat{\Delta}\, d_{12})^2] = \mathbb{E}[(\bar{d}_2 - \frac{\bar{d}_1 d_{11}}{d_{11}^2 + 2\lambda}\, d_{12})^2].$$

To make further progress, I condition on information at period $t = 1$: $\mathcal{F}_1$.

Step 4: Data-Generating Process and Conditional Distributions Assume

$$Y_{sct} = \alpha_s + \eta_{sc} + \varepsilon_{sct},$$

where $\alpha_s \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\alpha^2), \eta_{sc} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\eta^2)$ and $\varepsilon_{sct} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$.

Define shorthand:

$$\bar{a} = 2\,\sigma_\alpha^2 + \sigma_\eta^2 + \sigma_\varepsilon^2, \bar{b} = 2\sigma_\alpha^2 + \sigma_\eta^2, a = 2\,\sigma_\eta^2 + 2\,\sigma_\varepsilon^2, b = 2\,\sigma_\eta^2.$$

The relevant conditional distributions needed for the expected MSE are

$$\bar{d}_2?\bar{d}_1, \quad \text{and} \quad d_{12|d_{11}}.$$

Using Normality and independence, I find that

$$\begin{pmatrix} \bar{d}_2 \\ \bar{d}_1 \end{pmatrix} \sim \mathcal{N}\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2\,\sigma_\alpha^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 & 2\sigma_\alpha^2 + \sigma_\eta^2 \\ 2\sigma_\alpha^2 + \sigma_\eta^2 & 2\,\sigma_\alpha^2 + \sigma_\eta^2 + \sigma_\varepsilon^2 \end{pmatrix} \Big).$$

Similarly,

$$\begin{pmatrix} d_{12} \\ d_{11} \end{pmatrix} \sim \mathcal{N}\Big(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2\,\sigma_\eta^2 + 2\,\sigma_\varepsilon^2 & 2\,\sigma_\eta^2 \\ 2\sigma_\eta^2 & 2\,\sigma_\eta^2 + 2\,\sigma_\varepsilon^2 \end{pmatrix} \Big).$$

Combining these, the following conditional distributions are then

$$\bar{d}_2|\bar{d}_1 \sim \mathcal{N}(\bar{d}_1 \frac{\bar{b}}{\bar{a}}, \frac{\bar{a}^2 - \bar{b}^2}{\bar{a}})$$

$$d_{12}|d_{11} \sim \mathcal{N}(d_{11} \frac{b}{a}, \frac{a^2 - b^2}{a}).$$

Step 5: Expected Conditional MSE

Conditional on $\mathcal{F}_1$,

$$\mathbb{E}[(\bar{d}_2 - \frac{\bar{d}_1\, d_{11}}{d_{11}^2 + 2\,\lambda}\, d_{12})^2|\mathcal{F}_1] = \underbrace{(\bar{d}_1 \frac{\bar{b}}{\bar{a}} - \frac{\bar{d}_1 d_{11}}{d_{11}^2 + 2\lambda} d_{11} \frac{b}{a})^2}_{\mathbb{E}[\cdot|\mathcal{F}_1]^2} \\ + \underbrace{\frac{\bar{a}^2 - \bar{b}^2}{\bar{a}} + \Big(\frac{\bar{d}_1 d_{11}}{d_{11}^2 + 2\lambda}\Big)^2 \frac{a^2 - b^2}{a}}_{Var(\cdot|\mathcal{F}_1)}$$

$$= \bar{d}_1^2\Bigg((\frac{\bar{b}}{\bar{a}} - \frac{d_{11}^2}{d_{11}^2 + 2\lambda}\frac{b}{a})^2 + \frac{d_{11}^2}{(d_{11}^2 + 2\lambda)^2}\frac{a^2 - b^2}{a}\Bigg) + c$$
$$= c + c' \cdot \frac{1}{(d_{11}^2 + 2\lambda)^2}\Bigg((\frac{\bar{b}}{\bar{a}} \cdot (d_{11}^2 + 2\lambda) - d_{11}^2\frac{b}{a})^2 + \frac{a^2 - b^2}{a}\, d_{11}^2\Bigg),$$

where c and c' are constants independent of λ. Thus,

$$\mathbb{E}[(\bar{d}_2 - \frac{\bar{d}_1\, d_{11}}{d_{11}^2 + 2\,\lambda}\, d_{12})^2|\mathcal{F}_1] = c + c' \cdot \frac{1}{(d_{11}^2 + 2\,\lambda)^2}\, \Bigg((d_{11}^2(\frac{\bar{b}}{\bar{a}} - \frac{b}{a}) + 2\,\frac{\bar{b}}{\bar{a}}\,\lambda)^2 + \frac{a^2 - b^2}{a} d_{11}^2 \Bigg) \tag{B.1}$$

Step 6: Optimal λ^*

Define

$$u(x) = \frac{1}{(d_{11}^2 + 2\lambda)^2} \quad u'(x) = -\frac{4}{(d_{11}^2 + 2\lambda)^3}$$
$$v(x) = (d_{11}^2(\frac{\bar{b}}{\bar{a}} - \frac{b}{a}) + 2\,\frac{\bar{b}}{\bar{a}}\lambda)^2 + \frac{a^2 - b^2}{a} d_{11}^2 \quad v'(x) = 4\frac{\bar{b}}{\bar{a}}(d_{11}^2(\frac{\bar{b}}{\bar{a}} - \frac{b}{a}) + 2\,\frac{\bar{b}}{\bar{a}}\lambda)$$

Differentiating with respect to λ and solving yields

$$\begin{aligned}
\frac{\partial \mathbb{E}[MSE|\mathcal{F}_1]}{\partial \lambda} &= -\frac{4c}{(d_{11}^2 + 2\lambda)^3}\Bigg\{ (d_{11}^2\, \underbrace{(\frac{\bar{b}}{\bar{a}} - \frac{b}{a})}_{=x} + 2\,\frac{\bar{b}}{\bar{a}}\lambda)^2 + \frac{a^2 - b^2}{a} d_{11}^2 \Bigg\} \\
&\quad + c \cdot \frac{4}{(d_{11}^2 + 2\lambda)^2}\frac{\bar{b}}{\bar{a}}(d_{11}^2(\frac{\bar{b}}{\bar{a}} - \frac{b}{a}) + 2\,\frac{\bar{b}}{\bar{a}}\lambda) = 0 \\
\iff_{\lambda \neq -\frac{1}{2}d_{11}^2} &\ -\Bigg\{ (d_{11}^2 x + 2\,\frac{\bar{b}}{\bar{a}}\lambda)^2 + \frac{a^2 - b^2}{a} d_{11}^2 \Bigg\} \\
&\quad + (d_{11}^2 + 2\lambda)\,\frac{\bar{b}}{\bar{a}}(d_{11}^2(\frac{\bar{b}}{\bar{a}} - \frac{b}{a}) + 2\,\frac{\bar{b}}{\bar{a}}\lambda) = 0 \\
\iff &\ -\Bigg\{ d_{11}^4 x^2 + 2d_{11}^2 x 2\frac{\bar{b}}{\bar{a}}\lambda + 4(\frac{\bar{b}}{\bar{a}})^2\lambda^2 + \frac{a^2 - b^2}{a} d_{11}^2 \Bigg\} \\
&\quad + \frac{\bar{b}}{\bar{a}}(d_{11}^2 + 2\lambda) d_{11}^2 x + 2(\frac{\bar{b}}{\bar{a}})^2\lambda(d_{11}^2 + 2\lambda) = 0 \\
&\quad - d_{11}^4 x^2 - 2d_{11}^2 x 2\frac{\bar{b}}{\bar{a}}\lambda - 4(\frac{\bar{b}}{\bar{a}})^2\lambda^2 - \frac{a^2 - b^2}{a} d_{11}^2 + \frac{\bar{b}}{\bar{a}} d_{11}^4 x \\
&\quad + \frac{\bar{b}}{\bar{a}} d_{11}^2 x 2\lambda + 2(\frac{\bar{b}}{\bar{a}})^2\lambda d_{11}^2 + 4(\frac{\bar{b}}{\bar{a}})^2\lambda^2 = 0 \\
\iff &\ - d_{11}^4 x^2 - \frac{a^2 - b^2}{a} d_{11}^2 + \frac{\bar{b}}{\bar{a}} d_{11}^4 x
\end{aligned}$$

$$
\begin{aligned}
&+\lambda(-4d_{11}^2 x\frac{\bar{b}}{\bar{a}}+2\frac{\bar{b}}{\bar{a}}d_{11}^2 x+2(\frac{\bar{b}}{\bar{a}})^2 d_{11}^2)\\
&+\lambda^2(4(\frac{\bar{b}}{\bar{a}})^2-4(\frac{\bar{b}}{\bar{a}})^2)=0\\
\iff &d_{11}^4 x^2-\frac{a^2-b^2}{a}d_{11}^2+\frac{\bar{b}}{\bar{a}}d_{11}^4 x\\
&+\lambda(-4d_{11}^2 x\frac{\bar{b}}{\bar{a}}+2\frac{\bar{b}}{\bar{a}}d_{11}^2 x+2(\frac{\bar{b}}{\bar{a}})^2 d_{11}^2)=0
\end{aligned}
$$

Rearranging yields

$$
\begin{aligned}
\lambda &= \frac{d_{11}^4 x^2+\frac{1^2-b^2}{a}d_{11}^2-\frac{\bar{b}}{\bar{a}}d_{11}^4 x}{2(\frac{\bar{b}}{\bar{a}})^2 d_{11}^2-2d_{11}^2 x\frac{\bar{b}}{\bar{a}}}\\
&= \frac{d_{11}^4 x(x-\frac{\bar{b}}{\bar{a}})+\frac{a^2-b^2}{a}d_{11}^2}{2\frac{\bar{b}}{\bar{a}}d_{11}^2(\frac{\bar{b}}{\bar{a}}-x)}\\
&= \frac{\frac{a^2-b^2}{a}d_{11}^2-\frac{b}{a}d_{11}^4 x}{2\frac{\bar{b}}{\bar{a}}d_{11}^2\frac{b}{a}}\\
&= \frac{\frac{a^2-b^2}{a}-\frac{b}{a}d_{11}^2 x}{2\frac{\bar{b}}{\bar{a}}\frac{b}{a}}\\
&= \frac{a^2-b^2-bd_{11}^2(\frac{\bar{b}}{\bar{a}}-\frac{b}{a})}{2\ \frac{\bar{b}}{\bar{a}}\ b}
\end{aligned}
$$

Overall, the optimal λ is

$$
\lambda^* = \frac{a^2-b^2-b\ d_{11}^2\ (\frac{\bar{b}}{\bar{a}}-\frac{b}{a})}{2\ \frac{\bar{b}}{\bar{a}}\ b}.
$$

Note that

$$
\begin{aligned}
\frac{\bar{b}}{\bar{a}}-\frac{b}{a} &= \frac{2\sigma_\alpha^2+\sigma_\eta^2}{2\ \sigma_\alpha^2+\sigma_\eta^2+\sigma_\varepsilon^2}-\frac{2\sigma_\eta^2}{2\ \sigma_\eta^2+2\ \sigma_\varepsilon^2}\\
&= \frac{(2\sigma_\alpha^2+\sigma_\eta^2)(2\sigma_\eta^2+2\sigma_\varepsilon^2)-4\sigma_\eta^2\sigma_\alpha^2-2\sigma_\eta^2\sigma_\varepsilon^2}{(2\sigma_\alpha^2+\sigma_\eta^2+\sigma_\varepsilon^2)(2\sigma_\eta^2+2\sigma_\varepsilon^2)}\\
&= \frac{4\sigma_\alpha^2\sigma_\eta^2+4\sigma_\alpha^2\sigma_\varepsilon^2+2\sigma_\eta^4+2\sigma_\eta^2\sigma_\varepsilon^2-4\sigma_\eta^2\sigma_\alpha^2-2\sigma_\eta^4-2\sigma_\eta^2\sigma_\varepsilon^2}{(2\sigma_\alpha^2+\sigma_\eta^2+\sigma_\varepsilon^2)(2\sigma_\eta^2+2\sigma_\varepsilon^2)}\\
&= \frac{4\sigma_\alpha^2\sigma_\varepsilon^2}{(2\sigma_\alpha^2+\sigma_\eta^2+\sigma_\varepsilon^2)(2\sigma_\eta^2+2\sigma_\varepsilon^2)}
\end{aligned}
$$

Substituting in the variances for $a, b, \bar{a}, \bar{b}$ for the rest of the expression yields

$$\lambda^* = \frac{8\ \sigma_\eta^2 \sigma_\varepsilon^2 + 4\ \sigma_\varepsilon^2 - 2\ \sigma_\eta^2\ d_{11}^2 \frac{4\ \sigma_\alpha^2 \sigma_\varepsilon^2}{(2\sigma_\alpha^2 + \sigma_\eta^2 + \sigma_\varepsilon^2)(2\sigma_\eta^2 + 2\sigma_\varepsilon^2)}}{4\ \sigma_\eta^2\ \frac{2\ \sigma_\alpha^2 + \sigma_\eta^2}{2\ \sigma_\alpha^2 + \sigma_\eta^2 + \sigma_\varepsilon^2}} = \frac{2\ \sigma_\varepsilon^2 + \frac{\sigma_\varepsilon^2}{\sigma_\eta^2} - 2\ \ d_{11}^2 \frac{\sigma_\alpha^2 \sigma_\varepsilon^2}{(2\sigma_\alpha^2 + \sigma_\eta^2 + \sigma_\varepsilon^2)(2\sigma_\eta^2 + 2\sigma_\varepsilon^2)}}{\frac{2\ \sigma_\alpha^2 + \sigma_\eta^2}{2\ \sigma_\alpha^2 + \sigma_\eta^2 + \sigma_\varepsilon^2}}.$$

Step 7: Practical Approximation

The leading term $2\sigma_\varepsilon^2$ in the numerator often dominates, particularly when within-aggregate heterogeneity (σ_η^2) is large. The remaining terms involve factor variances and cross-terms (such as $\sigma_\varepsilon^2/\sigma_\eta^2$ and the d_{11} interaction) that reflect higher-order structure and are typically smaller or more difficult to estimate reliably. For practical implementation, using $2\sigma_\varepsilon^2$ provides a conservative and easy-to-compute proxy for the penalty parameter.

Appendix C Placebo Data Set Construction

The four real-world data sets used in my simulation study come from three publicly available repositories. Unemployment rates and weekly log wages are obtained directly from the Bureau of Labor Statistics (BLS) website, smoking rates from the Behavioral Risk Factor Surveillance System (BRFSS), and log GDP from the Penn table $log(GDP)$ World Table. The BRFSS and Penn World Table data are accessed through the supplementary materials of Dwyer-Lindgren et al. (2014) and Arkhangelsky et al. (2021), respectively.

Data set size. The data sets fall into two categories based on the level of aggregation. The first three data sets—unemployment, log wages, and smoking rates—are county-level panels, where U.S. states serve as their aggregate units. The fourth data set, from the Penn World Table, is country-level, with continents as the aggregates. These categories differ in the size of their aggregates, the disaggregate-to-aggregate ratios, and the length of the time dimension. Table 8 summarizes key size characteristics.

Data set construction. The aggregated outcomes are computed as weighted averages of their constituent disaggregated units, with population weights being fixed to v_{sc}. This ensures internal consistency between aggregated and disaggregated data and matches the

Table 8: Details on size of data sets used for simulation study

Data set	N_s	N_c	T
Penn table $log(GDP)$ Table	6	111	48
Unemployment rate	50	3127	31
Log wages	50	3106	31
Smoking rate	50	3126	17

paper's theoretical set-up. Counties with incomplete time series are dropped, slightly reducing the total number of counties from the full set of 3,144.

Factor model characteristics. For each full data set, I estimate a hierarchical linear factor model with rank 3 and report the relative size of the aggregate factor component, the disaggregate factor component, and the idiosyncratic error. Table 9 reports results for the full samples; Table 10 shows the same quantities for the restricted samples ($S = 20$) used in the simulation experiments.

Table 9: Details on size of the components estimated using the $rank = 3$ factor model as used for the simulation study in Section 6.

Data set	$\|\|L^{agg}\|\|_F / \sqrt{(\sum_s C_s) * T}$	$\|\|L^{disagg}\|\|_F / \sqrt{(\sum_s C_s) * T}$	$\hat{\sigma}_\varepsilon$
Penn table $log(GDP)$ Table	0.754	0.653	0.075
Unemployment rate	0.692	0.650	0.314
Log wages	0.861	0.499	0.100
Smoking rate	0.713	0.691	0.120

Table 10: Details on size of the components estimated using the $rank = 3$ factor model as used for the simulation study in Section 6. Subset of states $S = 20$.

Data set	$\|\|L^{agg}\|\|_F / \sqrt{(\sum_s C_s) * T}$	$\|\|L^{disagg}\|\|_F / \sqrt{(\sum_s C_s) * T}$	$\hat{\sigma}_\varepsilon$
Penn table $log(GDP)$ Table	0.754	0.653	0.075
Unemployment rate	0.589	0.741	0.323
Log wages	0.833	0.547	0.086
Smoking rate	0.554	0.825	0.109

Appendix D Simulations with Other Treatment Assignments

In this section, I report the RMSE and bias of the dGSC, DiD, and mlSC estimators for the CPS and BRFSS data sets under two additional treatment assignments. The first assigns treatment based on gun control laws, and the second uses a random assignment. The results closely mirror those in the main text and reinforce the paper's main conclusions.

Table 11: Simulation results for all dGSC and DiD estimators: RMSE and Bias.

	Classical SC	**dGSC-AD**	**dGSC-DA**	**dGSC**	**DiD (aggregate)**	**DiD (disaggregate)**
			RMSE			
Assn.: Open carry						
Unemployment rate	0.067	0.073	0.132	0.058	0.137	0.138
Log wages	0.069	0.026	0.117	0.026	0.138	0.138
Smoking rate	0.098	0.039	0.160	0.038	0.214	0.208
Assn.: Random						
Unemployment rate	0.078	0.082	0.121	0.073	0.135	0.134
Log wages	0.069	0.024	0.126	0.025	0.133	0.134
Smoking rate	0.106	0.041	0.161	0.044	0.216	0.221
			Bias			
Assn.: Open carry						
Unemployment rate	-0.004	-0.005	0.046	-0.007	-0.032	-0.037
Log wages	-0.006	0.000	-0.044	-0.002	-0.004	-0.014
Smoking rate	0.011	0.000	0.056	0.002	0.042	-0.000
Assn.: Random						
Unemployment rate	-0.000	-0.004	0.038	-0.004	0.002	-0.002
Log wages	0.001	-0.002	-0.020	-0.002	-0.006	-0.016
Smoking rate	-0.011	-0.006	0.011	-0.008	-0.010	-0.052

All results are based on $S_{sim} = 1,000$ simulation runs. $N_{tr} = 1$ and $T_{post} = 1$. Outcomes are normalized to have mean zero and unit variance. *Estimators:* classical SC estimator: aggregated data for treated and control; dGSC-AD: aggregated data for treated, disaggregated data for control; dGSC-DA: disaggregated data for treated, aggregated data for control; dGSC: disaggregated data for treated and control; DiD (aggregate): difference-in-differences using aggregated data; DiD (disaggregate): difference-in-differences using disaggregated data.

Table 12: Simulation results for four real data sets: RMSE and Bias.

	mlSC (oracle)	**mlSC (heuristic)**	**mlSC (CV time)**	**Classical SC**	**dGSC-AD**
RMSE					
Assn.: Open carry					
Unemployment rate	0.053	0.056	0.055	0.067	0.073
Log wages	0.022	0.022	0.024	0.069	0.026
Smoking rate	0.030	0.031	0.031	0.098	0.039
Assn.: Random					
Unemployment rate	0.070	0.070	0.068	0.078	0.082
Log wages	0.022	0.022	0.025	0.069	0.024
Smoking rate	0.034	0.034	0.036	0.106	0.041
Bias					
Assn.: Open carry					
Unmployment rate	-0.003	-0.003	-0.004	-0.004	-0.005
Log wages	-0.000	-0.000	-0.001	-0.006	0.000
Smoking rate	0.002	0.001	0.003	0.011	0.000
Assn.: Random					
Unemployment rate	-0.001	-0.001	-0.003	-0.000	-0.004
Log wages	-0.001	-0.001	-0.002	0.001	-0.002
Smoking rate	-0.004	-0.004	-0.004	-0.011	-0.006

All results are based on $S_{sim} = 1,000$ simulation runs. $N_{tr} = 1$ and $T_{post} = 1$. Outcomes are normalized to have mean zero and unit variance. In **bolt**: RMSE closest to oracle. *Estimators:* mlSC: multi-level SC estimator; classical SC estimator: aggregated data for treated and control; dGSC-AD: aggregated data for treated, disaggregated data for control.

Appendix E Simulations with Smaller Data Settings

n this section, I examine the robustness of my results to changes in the relative sizes of the aggregated units S and their disaggregated units C_s. Specifically, I restrict the sample to the smallest $S = 20$ states which yields an average of $C_s \approx 20$ counties per state. Tables 13 and 14 report the RMSE and bias results.

Across all designs, disaggregation becomes more valuable in this smaller setting. Notably, the dGSC-AD estimator now outperforms classical SC for the unemployment rate data set. However, the magnitude of the gains still varies across designs, indicating that relative sizes of aggregates and disaggregates are only one of several factors driving the benefits of disaggregated data. The oracle mlSC estimator delivers further improvements over dGSC-AD, and the feasible versions continue to track the oracle closely. Finally, the DiD estimators perform substantially worse than the best dGSC estimators in every scenario, and, consistent with earlier findings, disaggregation offers no RMSE improvement for DiD.

Table 13: Simulation results for all dGSC and DiD estimators: RMSE and Bias. Subset of states ($S = 20$).

	Classical SC	dGSC-AD	dGSC-DA	dGSC	DiD (aggregate)	DiD (disaggregate)
			RMSE			
Assn.: Financial markets						
Penn Table: $log(GDP)$	0.384	0.035	0.409	0.052	0.152	0.153
Assn.: Min. wage						
Unemployment rate	0.144	0.130	0.195	0.122	0.272	0.281
Log wages	0.079	0.038	0.166	0.041	0.189	0.200
Smoking rate	0.237	0.099	0.348	0.153	0.354	0.362
Assn.: Open carry						
Unemployment rate	0.117	0.108	0.164	0.099	0.204	0.198
Log wages	0.055	0.031	0.158	0.032	0.128	0.125
Smoking rate	0.184	0.063	0.276	0.085	0.238	0.234
Assn.: Random						
Unemployment rate	0.123	0.109	0.173	0.105	0.219	0.219
Log wages	0.059	0.033	0.158	0.033	0.142	0.145
Smoking rate	0.183	0.067	0.270	0.096	0.249	0.245
			Bias			
Assn.: Financial markets						
Penn Table: $log(GDP)$	0.209	-0.000	0.204	0.013	0.020	0.033
Assn.: Min. wage						
Unemployment rate	-0.006	-0.008	0.038	-0.006	-0.071	-0.118
Log wages	-0.010	-0.005	0.031	-0.011	-0.044	-0.093
Smoking rate	-0.119	-0.038	-0.078	-0.058	-0.191	-0.222
Assn.: Open carry						
Unemployment rate	0.003	0.000	0.015	0.001	0.028	-0.022
Log wages	0.008	-0.002	0.087	-0.003	0.027	-0.024
Smoking rate	-0.014	-0.007	-0.033	-0.008	0.009	-0.026
Assn.: Random						
Unemployment rate	0.001	0.001	0.022	0.001	-0.002	-0.051
Log wages	0.005	-0.002	0.078	-0.004	0.008	-0.043
Smoking rate	-0.027	-0.009	-0.053	-0.014	-0.001	-0.036

All results are based on $S_{sim} = 1,000$ simulation runs. $N_{tr} = 1$ and $T_{post} = 1$. Outcomes are normalized to have mean zero and unit variance. *Estimators:* classical SC estimator: aggregated data for treated and control; dGSC-AD: aggregated data for treated, disaggregated data for control; dGSC-DA: disaggregated data for treated, aggregated data for control; dGSC: disaggregated data for treated and control; DiD (aggregate): difference-in-differences using aggregated data; DiD (disaggregate): difference-in-differences using disaggregated data.

Table 14: Simulation results for four real data sets: RMSE and Bias. Subset of states ($S = 20$).

	mlSC (oracle)	mlSC (heuristic)	mlSC (CV time)	Classical SC	dGSC-AD
RMSE					
Assn.: Financial markets					
Penn Table: $log(GDP)$	0.034	0.035	0.036	0.384	0.035
Assn.: Min. wage					
Unemployment rate	0.117	0.122	0.123	0.144	0.130
Log wages	0.036	0.038	0.040	0.079	0.038
Smoking rate	0.097	0.099	0.098	0.237	0.099
Assn.: Open carry					
Unemployment rate	0.093	0.096	0.094	0.117	0.108
Log wages	0.028	0.029	0.030	0.055	0.031
Smoking rate	0.061	0.061	0.063	0.184	0.063
Assn.: Random					
Unemployment rate	0.098	0.100	0.101	0.123	0.109
Log wages	0.030	0.030	0.032	0.059	0.033
Smoking rate	0.065	0.066	0.066	0.183	0.067
Bias					
Penn Table: $log(GDP)$	0.002	0.000	0.005	0.209	-0.000
Assn.: Min. wage					
Unemployment rate	-0.012	-0.009	-0.011	-0.006	-0.008
Log wages	-0.006	-0.008	-0.007	-0.010	-0.005
Smoking rate	-0.036	-0.038	-0.041	-0.119	-0.038
Assn.: Open carry					
Unemployment rate	0.001	0.003	-0.000	0.003	0.000
Log wages	-0.002	-0.003	-0.002	0.008	-0.002
Smoking rate	-0.006	-0.006	-0.005	-0.014	-0.007
Assn.: Random					
Unemployment rate	-0.003	-0.002	-0.004	0.001	0.001
Log wages	-0.002	-0.003	-0.002	0.005	-0.002
Smoking rate	-0.008	-0.008	-0.011	-0.027	-0.009

All results are based on $S_{sim} = 1,000$ simulation runs. $N_{tr} = 1$ and $T_{post} = 1$. Outcomes are normalized to have mean zero and unit variance. In **bolt**: RMSE closest to oracle. *Estimators:* mlSC: multi-level SC estimator; classical SC estimator: aggregated data for treated and control; dGSC-AD: aggregated data for treated, disaggregated data for control.

Appendix F Simulations with AR(2) Error Term

In this section, I assess the robustness of my results to an alternative error structure in the semi-synthetic data sets. Specifically, I model the error terms ε_{sct} as following an AR(2) process, as in Arkhangelsky et al. (2021), consistent with evidence in Angrist and Pischke (2009) and Bertrand et al. (2004). This modification introduces additional structure in the errors, which the disaggregated data can exploit.

The main findings remain robust under this alternative specification. Notably, the benefits of incorporating disaggregated data are even larger under the AR(2) error structure, reflecting the greater information content captured at the disaggregate level.

Table 15: Simulation results for all dGSC and DiD estimators: RMSE and Bias.

	Classical SC	dGSC-AD	dGSC-DA	dGSC	DiD (aggregate)	DiD (disaggregate)
RMSE						
Assn.: Financial Markets						
Penn table $log(GDP)$	0.382	0.016	0.407	0.047	0.150	0.152
Assn.: Minimum wage						
Unemployment rate	0.092	0.086	0.145	0.081	0.132	0.131
Log wages	0.116	0.000	0.247	0.006	0.123	0.122
Smoking rate	0.157	0.073	0.222	0.080	0.293	0.313
Assn.: Open carry						
Unemployment rate	0.078	0.074	0.137	0.069	0.147	0.147
Log wages	0.066	0.000	0.115	0.009	0.138	0.138
Smoking rate	0.102	0.052	0.165	0.053	0.219	0.214
Assn.: Random						
Unemployment rate	0.090	0.079	0.132	0.078	0.137	0.136
Log wages	0.065	0.000	0.124	0.008	0.131	0.132
Smoking rate	0.109	0.056	0.165	0.060	0.216	0.220
Bias						
Assn.: Financial Markets						
Penn table $log(GDP)$	0.207	0.001	0.202	0.015	0.021	0.034
Assn.: Minimum wage						
Unemployment rate	0.006	0.025	0.057	0.021	0.019	0.013
Log wages	0.026	0.000	0.117	0.001	0.005	-0.005
Smoking rate	-0.069	-0.019	-0.127	-0.042	-0.154	-0.194
Assn.: Open carry						
Unemployment rate	-0.006	-0.009	0.044	-0.008	-0.032	-0.037
Log wages	-0.006	0.000	-0.044	-0.000	-0.004	-0.014
Smoking rate	0.008	-0.001	0.055	0.000	0.041	-0.002
Assn.: Random						
Unemployment rate	0.002	0.005	0.037	0.003	0.001	-0.004
Log wages	0.001	0.000	-0.019	0.000	-0.005	-0.015
Smoking rate	-0.009	-0.006	0.014	-0.009	-0.009	-0.051

All results are based on $S_{sim} = 1,000$ simulation runs. $N_{tr} = 1$ and $T_{post} = 1$. Outcomes are normalized to have mean zero and unit variance. *Estimators:* classical SC estimator: aggregated data for treated and control; dGSC-AD: aggregated data for treated, disaggregated data for control; dGSC-DA: disaggregated data for treated, aggregated data for control; dGSC: disaggregated data for treated and control; DiD (aggregate): difference-in-differences using aggregated data; DiD (disaggregate): difference-in-differences using disaggregated data.

Table 16: Simulation results for four real data sets: RMSE and Bias.

	mlSC (oracle)	**mlSC (heuristic)**	**mlSC (CV time)**	**Classical SC**	**dGSC-AD**
			RMSE		
Assn.: Financial markets					
Penn table $log(GDP)$	0.016	0.018	0.013	0.382	0.016
Assn.: Minimum wage					
Unemployment rate	0.071	0.073	0.076	0.092	0.086
Log wages	0.000	0.003	0.000	0.116	0.000
Smoking rate	0.059	0.059	0.061	0.157	0.073
Assn.: Open carry					
Unemployment rate	0.056	0.057	0.058	0.078	0.074
Log wages	0.000	0.002	0.000	0.066	0.000
Smoking rate	0.042	0.043	0.042	0.102	0.052
Assn.: Random					
Unemployment rate	0.066	0.066	0.067	0.090	0.079
Log wages	0.000	0.002	0.000	0.065	0.000
Smoking rate	0.046	0.047	0.046	0.109	0.056
			Bias		
Assn.: Financial markets					
Penn table $log(GDP)$	0.001	0.001	0.002	0.207	0.001
Assn.: Minimum wage					
Unemployment rate	0.015	0.019	0.010	0.006	0.025
Log wages	-0.000	-0.000	0.000	0.026	0.000
Smoking rate	-0.013	-0.012	-0.011	-0.069	-0.019
Assn.: Open carry					
Unmployment rate	-0.004	-0.004	-0.007	-0.006	-0.009
Log wages	0.000	-0.000	0.000	-0.006	0.000
Smoking rate	-0.001	-0.001	0.001	0.008	-0.001
Assn.: Random					
Unemployment rate	0.004	0.005	0.003	0.002	0.005
Log wages	-0.000	0.000	0.000	0.001	0.000
Smoking rate	-0.004	-0.004	-0.004	-0.009	-0.006

All results are based on $S_{sim} = 1{,}000$ simulation runs. $N_{tr} = 1$ and $T_{post} = 1$. Outcomes are normalized to have mean zero and unit variance. In **bold**: RMSE closest to oracle. *Estimators:* mlSC: multi-level SC estimator; classical SC estimator: aggregated data for treated and control; dGSC-AD: aggregated treated, disaggregated control.

Appendix G Variance Decomposition

In this section, I outline a simple variance decomposition to estimate the noise variance $\hat{\sigma}^2_\varepsilon$used for the heuristic $\hat{\lambda}$ n Section 5.2. I employ a simplified version of the random effects model from Section 7, omitting the time factors:

$$Y_{sct} = \alpha_s + \eta_{sc} + \varepsilon_{sct},$$

where $\alpha_s \overset{\text{i.i.d.}}{\sim} (0, \sigma^2_\alpha), \eta_{sc} \overset{\text{i.i.d.}}{\sim} (0, \sigma^2_\eta)$ and $\varepsilon_{sct} \overset{\text{i.i.d.}}{\sim} (0, \sigma^2_\varepsilon)$.

Under this model, the variance components are estimated as follows:

$$\begin{aligned}
\hat{\mu}_{sc} &= \frac{1}{T_0} \sum_{t=1}^{T_0} Y_{sct} \\
\hat{\alpha}_s &= \frac{1}{C_s} \sum_{c=1}^{C_s} \hat{\mu}_{sc} \\
\hat{\eta}_{sc} &= \hat{\mu}_{sc} - \hat{\alpha}_s \\
Var(\eta_{sc}) &\approx \frac{1}{C_s} \sum_{c=1}^{C_s} \hat{\eta}^2_{sc} \\
Var(\epsilon_{sct}) &\approx \frac{1}{C_s} \sum_{c=1}^{C_s} \frac{1}{T_0} \sum_{t=1}^{T_0} (Y_{sct} - \hat{\mu}_{sc})^2 \\
Var(\alpha_s) &\approx \hat{s}^2_\alpha,
\end{aligned}$$

where $\hat{s}$ is the sample standard deviation. All calculations use pre-treatment data only.

Appendix H Aggregation for the Difference-in-Differences Estimator

This appendix examines the role of disaggregation in difference-in-differences (DiD) settings. Using disaggregated data when treatment occurs at the aggregate level is common in applied work (e.g., Card and Krueger, 1994; Neumark and Wascher, 2001; Baum and Ruhm, 2016).

While the main text briefly discussed reasons to prefer synthetic control over DiD and disaggregation in DiD estimators, here I formally investigate how disaggregation affects the DiD estimator.

For consistency with the broader DiD literature, I assume a two-way fixed effects specification for the latent factor model:

$$
\begin{aligned}
Y_{sct}(0) &= \alpha_s + \eta_{sc} + \beta_t + \varepsilon_{sct} \\
Y_{sct}(1) &= Y_{sct}(0) + \tau_{sct}
\end{aligned}
$$

with $\varepsilon_{sct} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$. Let $v_{sc} = \frac{1}{C_s} \; \forall s$, so aggregated units are simple averages of their disaggregated units.

In the setup of this paper, the DiD estimator reduces to a simple difference in averages because only a single unit receives treatment at a given time. Accordingly, we can frame the problem as a two-group, two-period design: the two groups are (i) the treated unit—either a single aggregate unit for the aggregated estimator or all its disaggregated sub-units for the disaggregated estimator—and (ii) the control group; the two periods are pre-treatment, $\bar{t} = 1$ ($t = 1, ..., T_0$) and post-treatment $\bar{t} = 2$ ($t = T_0 + 1, ..., T$). Let

$$
\bar{Y}_{sc1} = \frac{1}{T_0} \sum_{t=1}^{T_0} Y_{sct} \quad \text{and} \quad \bar{Y}_{sc2} = \frac{1}{T - T_0 + 1} \sum_{t=T_0+1}^{T} Y_{sct}
$$

denote the pre- and post-treatment averages, respectively. The time averages for the aggregated outcomes are similarly defined, $\bar{Y}_{s2}$ and $\bar{Y}_{s1}$.

With this notation, the two DiD estimators—one using aggregate data and one using disaggregated data—are defined as follows:

$$
\begin{aligned}
\hat{\tau}^{DiD,agg} &= \left[\bar{Y}_{02} - \bar{Y}_{01}\right] - \left[\frac{1}{S}\sum_{s=1}^{S}(\bar{Y}_{s2} - \bar{Y}_{s1})\right] \\
&= \left[\sum_{c'=1}^{C_0} \frac{1}{C_0}(\bar{Y}_{0c'2} - \bar{Y}_{0c'1})\right] - \left[\sum_{s=1}^{S}\sum_{c=1}^{C_s} \frac{1}{S}\frac{1}{C_s}(\bar{Y}_{sc2} - \bar{Y}_{sc1})\right]
\end{aligned}
$$

$$\hat{\tau}^{DiD,disagg} = \left[\sum_{c'=1}^{C_0} \frac{1}{C_0}(\bar{Y}_{0c'2} - \bar{Y}_{0c'1})\right] - \left[\sum_{s=1}^{S}\sum_{c=1}^{C_s} \frac{1}{\sum_{s'=1}^{S} C_{s'}}(\bar{Y}_{sc2} - \bar{Y}_{sc1})\right]$$

Note that the first part of the DiD estimator is identical for both the aggregate and disaggregate versions. Consequently, the difference between the two estimators reduces to

$$\hat{\tau}^{DiD,agg} - \hat{\tau}^{DiD,disagg} = \sum_{s=1}^{S}\sum_{c=1}^{C_s}\left(\frac{1}{\sum_{s'=1}^{S} C_{s'}} - \frac{1}{S}\frac{1}{C_s}\right)(\bar{Y}_{sc2} - \bar{Y}_{sc1}).$$

This expression shows that if each aggregate unit contains the same number of disaggregated units, $C_s = C\ \forall s$, the two estimators coincide. Otherwise, the estimators differ.

Under the two-way fixed effects model, both estimators are unbiased. Therefore, any differences in out-of-sample performance in terms of mean squared error arise solely from differences in variance. Specifically, the MSE difference between the aggregate and disaggregate DiD estimator is determined entirely by the variance of the weighted control units.[18]

The control-unit variances for the two estimators are then

$$\begin{aligned}
&\textit{Control variance, aggregate}: \\
Var\Big(\Big[\sum_{s=1}^{S}\sum_{c=1}^{C_s}\frac{1}{S}\frac{1}{C_s}(\bar{Y}_{sc2} - \bar{Y}_{sc1})\Big]\Big) &= \frac{1}{S^2}\sum_{s=1}^{S}\sum_{c=1}^{C_s}\frac{1}{C_s^2}Var(\Delta\varepsilon_{sc}) \\
&= \frac{1}{S^2}\sum_{s=1}^{S}\sum_{c=1}^{C_s}\frac{1}{C_s}2\cdot\sigma_\varepsilon^2 \\
&= 2\cdot\sigma_\varepsilon^2\frac{1}{S^2}\sum_{s=1}^{S}\frac{1}{C_s}
\end{aligned}$$

and

$$\textit{Control variance, disaggregate}:$$

[18]Recall that the first term of the estimator is common to both versions, and, because the states are independent, no covariance terms appear in the variance.

$$Var(\Big[\sum_{s=1}^{S}\sum_{c=1}^{C_s}\frac{1}{\sum_{s'=1}^{S}C_{s'}}(\bar{Y}_{sc2}-\bar{Y}_{sc1})\Big]) = \frac{1}{(\sum_{s'}C_{s'})^2}\sum_{s=1}^{S}\sum_{c=1}^{C_s}Var(\Delta\varepsilon_{sc})$$
$$= 2\cdot\sigma_\varepsilon^2\frac{1}{\sum_{s'}C_{s'}}$$

Applying the arithmetic mean–harmonic mean inequality gives

$$\frac{S}{\sum_s\frac{1}{C_s}} \leq \frac{\sum_{s'}C_{s'}}{S}$$
$$\iff \frac{1}{S}\sum_s\frac{1}{C_s} \geq \frac{S}{\sum_{s'}C_{s'}}$$
$$\iff \frac{1}{S^2}\sum_s\frac{1}{C_s} \geq \frac{1}{\sum_{s'}C_{s'}}$$

Thus, under these assumptions, the disaggregated DiD estimator is (weakly) more efficient than the aggregated estimator. The efficiency gain arises because, when aggregate units contain different numbers of disaggregated units, the aggregate estimator effectively gives more weight to smaller units, increasing variance. Disaggregation mitigates this imbalance and reduces the overall variance of the estimator.

Remark Consider a generalized version of the DiD estimator in which the aggregation weights for disaggregated units differ from simple averages. The aggregated estimator then becomes

$$\hat{\tau}^{DiD,agg} = \Big[\sum_{c\in\mathcal{C}_0}v_{0c}(Y_{0c2}-Y_{0c1})\Big] - \Big[\sum_{s=1}^{S}\sum_{c=1}^{C_s}\frac{1}{S}v_{sc}(Y_{sc2}-Y_{sc1})\Big],$$

where v_{sc} denotes the weight for disaggregated unit c within aggregated unit s. The disaggregated estimator remains the same. This mismatch introduces a bias in the treated term of the disaggregated estimator relative to the true weighted effect:

$$\hat{\tau}^{DiD,agg} - \tau = \Bigg(\sum_{c\in\mathcal{C}_0}v_{0c}(\varepsilon_{0c2}-\varepsilon_{0c1}) - \sum_{s=1}^{S}\sum_{c=1}^{C_s}\frac{1}{S}\frac{1}{C_s}(\varepsilon_{sc2}-\varepsilon_{sc1})\Bigg)$$
$$\hat{\tau}^{DiD,disagg} - \tau = \sum_{c\in\mathcal{C}_0}(\frac{1}{C_0}-v_{0c})\tau_{0c2} + \Bigg(\sum_{c\in\mathcal{C}_0}\frac{1}{C_0}(\varepsilon_{0c2}-\varepsilon_{0c1}) - \sum_{s=1}^{S}\sum_{c=1}^{C_s}\frac{1}{\sum_{s'=1}^{S}C_{s'}}(\varepsilon_{sc2}-\varepsilon_{sc1})\Bigg)$$

The contribution of the control units to variance remains similar to the standard setting, since the disaggregated estimator still spreads weight approximately equally. As a result, the mean squared error of the disaggregated estimator now reflects both variance and this additional bias. Without further assumptions about the structure of the treatment effects within aggregates, we cannot determine in advance whether the aggregated or disaggregated estimator will perform better.

Remark When the two-way fixed effects model is violated, both aggregated and disaggregated DiD estimators can become biased, potentially misestimating the treatment effect. Consequently, the mean squared error for each estimator will reflect both a bias and a variance component. Intuitively, if the model misspecification is largely idiosyncratic across disaggregated units, the disaggregated estimator is likely to perform better because it can exploit within-aggregate variation. Conversely, if the misspecification is structured or correlated within aggregates (e.g., within states), the aggregated estimator may be more efficient and exhibit lower MSE.

Appendix I Additional Details for Empirical Applications

This section provides further information on the counties to which the dGSC-AD estimator assigns positive weight in the two empirical applications.

I.0.1 Minnesota's Cigarette Tax Increase

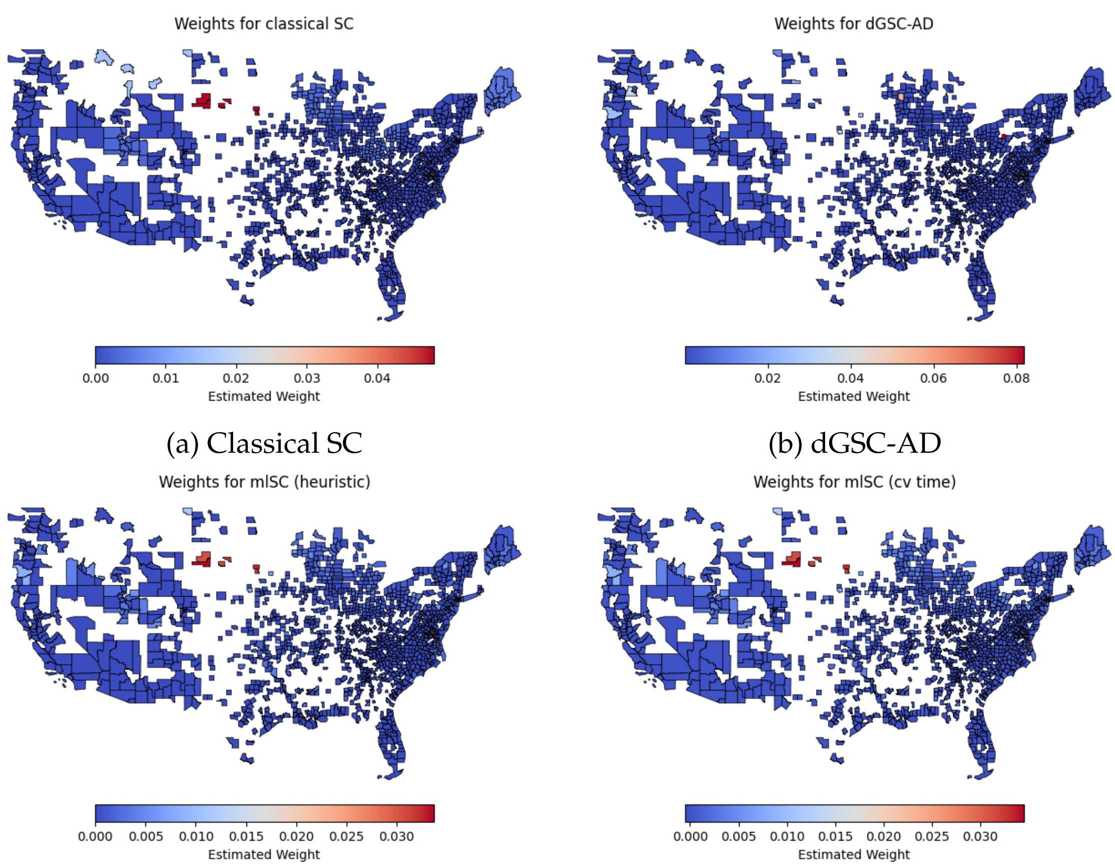
The dGSC-AD estimator selects 120 counties from 35 control states, distributed as follows: AL (3), AR (3), CA (1), CO (3), GA (2), IA (5), ID (2), IL (3), IN (6), KS (1), KY (5), MD (1), MI (6), MO (2), MS (3), MT (1), NC (7), ND (2), NE (2), NM (1), NV (1), NY (2), OH (9), OK (5), OR (6), PA (6), SC (3), SD (2), TN (1), UT (5), VA (8), VT (1), WI (9), WV (2), WY (1).

By contrast, the classical SC estimator using aggregated data assigns positive weight to only eleven states, with most weight concentrated in the Midwest: CT (0.04), IL (0.025), KY (0.01), ME (0.08), MT (0.08), OH (0.20), OR (0.002), RI (0.07), SD (0.29), UT (0.05), and WI (0.15).

Table 17 reports the weight vector norms for all SC-type estimators, and Figure 10 visualizes the estimated county-level weights. Overall, both the norms and the maps show that the feasible mlSC estimators distribute weights across a larger set of counties compared to the classical SC. In contrast, the dGSC-AD estimator concentrates weight on a small number of counties, leading to the highest vector norm.

Estimator	Weight Vector Norm
Classical SC	0.018
dGSC-AD	0.025
mlSC (heuristic)	0.009
mlSC (cv time)	0.009

Table 17: Weight vector norms for SC-type estimators (on county-level)



(a) Classical SC (b) dGSC-AD

(c) mlSC (heuristics) (d) mlSC (cv time)

Figure 10: Treated state: MN in July 2013: Weight vectors by control units' counties.

I.0.2 Iowa's Minimum Wage Increase

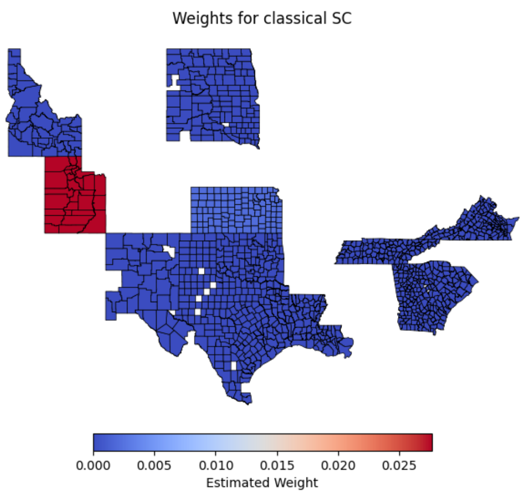
The dGSC-AD estimator assigns positive weight to 115 counties across 11 of the 13 control states[19]. The counties are distributed as follows: GA (8), ID (3), KS (29), LA (3), ND (7), OK (8), SC (1), SD (14), TN (4), TX (18), and VA (20). On average, $\bar{C}_s = 10.45$ counties per state receive positive weight. In contrast, the classical SC estimator assigns weight to only two states: UT (0.775) and KS (0.225).

Table 18 reports the weight vector norms for all SC-type estimators, and Figure 11 visualizes the estimated county-level weights. Similarly to the first application, both the norms and the maps show that the feasible mlSC estimators distribute weights across a larger set of counties compared to the classical SC. In contrast, the dGSC-AD estimator concentrates weight on a small number of counties, leading to the highest vector norm.

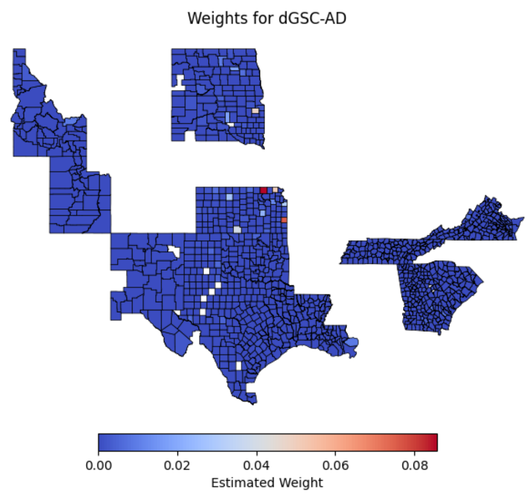
Estimator	**Weight Vector Norm**
Classical SC	0.022
dGSC-AD	0.030
mlSC (heuristic)	0.005
mlSC (cv time)	0.005

Table 18: Weight vector norms for SC-type estimators (on county-level)

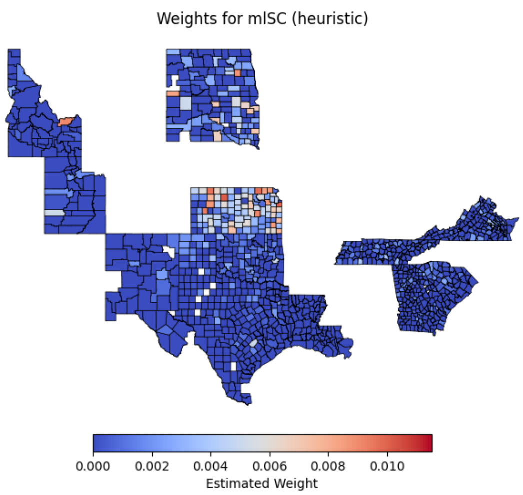
[19] Here, "substantial" weight is defined as any weight exceeding an equal weighting scheme, i.e., $w_{sc} = 1/1141 \approx 0.009$



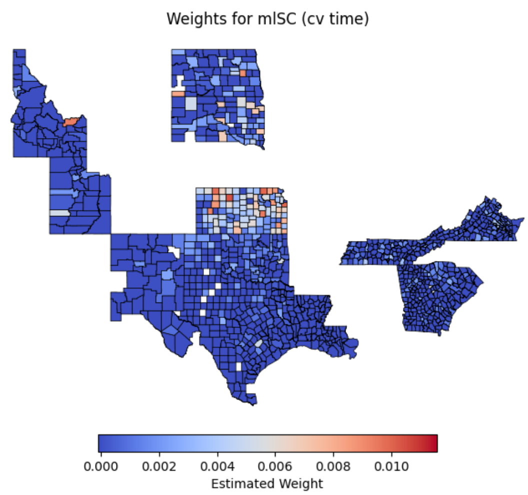
(a) Classical SC



(b) dGSC-AD



(c) mlSC (heuristics)



(d) mlSC (cv time)

Figure 11: Treated state: Iowa in 2007 Q2: Weight vectors by control units' counties.

Appendix J Derivation of Optimal λ_1^*, λ_2^* in a Stylized Model

I revisit the stylized example from Appendix B, now allowing for disaggregation of the treated unit. Recall that aggregated units are simple averages of their disaggregated units. Let the disaggregated donor unit weights for treated county 1 be $w_1 = \frac{1}{2} - \Delta_1$ and $1 - w_1 = \frac{1}{2} + \Delta_1$ and for treated county 2 be $w_2 = \frac{1}{2} - \Delta_2$ and $1 - w_2 = \frac{1}{2} + \Delta_2$.

Step 1: Reparameterization

The mlSC objective with two penalty terms (ignoring non-negativity constraints) is:

$$\begin{aligned}
&\arg\min_{w_1,w_2} \frac{1}{2}(Y_{011} - w_1 Y_{111} - (1-w_1)Y_{121})^2 + \frac{1}{2}(Y_{021} - w_2 Y_{111} - (1-w_2)Y_{121})^2 \\
&\qquad + \lambda_1 \Big((w_1 - \frac{1}{2})^2 + (1 - w_1 - \frac{1}{2})^2 + (w_2 - \frac{1}{2})^2 + (1 - w_2 - \frac{1}{2})^2 \Big) \\
&\qquad + \lambda_2 \Big((w_1 - w_2)^2 + (1 - w_1 - (1 - w_2))^2 \Big) \\
&\arg\min_{\Delta_1,\Delta_2} \frac{1}{2}(Y_{011} - (\frac{1}{2} - \Delta_1)Y_{111} - (\frac{1}{2} + \Delta_1)Y_{121})^2 + \frac{1}{2}(Y_{021} - (\frac{1}{2} - \Delta_2)Y_{111} - (\frac{1}{2} + \Delta_2)Y_{121})^2 \\
&\qquad + \lambda_1 \Big(2 * \Delta_1^2 + 2 * \Delta_2^2 \Big) + \lambda_2 \Big(2 * (\Delta_1 - \Delta_2)^2 \Big) \\
&\arg\min_{\Delta_1,\Delta_2} \frac{1}{2}(Y_{011} - Y_{11} - \Delta_1(Y_{121} - Y_{111}))^2 + \frac{1}{2}(Y_{021} - Y_{11} - \Delta_2(Y_{121} - Y_{111}))^2 \\
&\qquad + \lambda_1 \Big(2 * \Delta_1^2 + 2 * \Delta_2^2 \Big) + \lambda_2 \Big(2 * (\Delta_1 - \Delta_2)^2 \Big) \\
&\arg\min_{\Delta_1,\Delta_2} \frac{1}{2}(\bar{d}_{11} - \Delta_1 d_{11})^2 + \frac{1}{2}(\bar{d}_{21} - \Delta_2 d_{11})^2 + \lambda_1 \Big(2 * \Delta_1^2 + 2 * \Delta_2^2 \Big) + \lambda_2 \Big(2 * (\Delta_1 - \Delta_2)^2 \Big)
\end{aligned}$$

where I define $\bar{d}_{c't} = Y_{0c't} - Y_{1,t}$ and $d_{st} = Y_{s2t} - Y_{s1t}$ as the difference between treated disaggregated unit and control aggregated unit and within-aggregated unit differences.

Step 2: First-Order Conditions

The first-order conditions are

$$\begin{aligned}
\frac{\partial}{\partial \Delta_1} = 0 &\iff 0 = 2 * \frac{1}{2}(\bar{d}_{11} - \Delta_1 d_{11]})(-d_{11}) + 4\lambda_1\Delta_1 + 4\lambda_2(\Delta_2 - \Delta_1)(-1) \\
\frac{\partial}{\partial \Delta_1} = 0 &\iff 0 = 2 * \frac{1}{2}(\bar{d}_{21} - \Delta_2 d_{11]})(-d_{11}) + 4\lambda_1\Delta_2 + 4\lambda_2(\Delta_2 - \Delta_1)
\end{aligned}$$

The first-order conditions yield the linear system:

$$\begin{pmatrix} d_{11}^2 + 4\lambda_1 + 4\lambda_2 & -4\lambda_2 \\ -4\lambda_2 & d_{11}^2 + 4\lambda_2 + 4\lambda_2 \end{pmatrix} \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix} = \begin{pmatrix} \bar{d}_{11} d_{11} \\ \bar{d}_{21} d_{11} \end{pmatrix}$$

Solving this system gives

$$\begin{pmatrix} \hat{\Delta}_1 \\ \hat{\Delta}_2 \end{pmatrix} = \begin{pmatrix} \frac{(d_{11}^2+4\lambda_1+4\lambda_2)\bar{d}_{11}d_{11}+4\lambda_2\bar{d}_{21}d_{11}}{(d_{11}^2+4\lambda_1+8\lambda_2)*(d_{11}^2+4\lambda_2)} \\ \frac{4\lambda_2\bar{d}_{11}d_{11}+(d_{11}^2+4\lambda_1+4\lambda_2)\bar{d}_{21}d_{11}}{(d_{11}^2+4\lambda_1+8\lambda_2)*(d_{11}^2+4\lambda_2)} \end{pmatrix}$$

Then, I find that

$$\begin{aligned} \hat{\Delta}_1 + \hat{\Delta}_2 &= \frac{d_{11}(\bar{d}_{11}+\bar{d}_{21})}{d_{11}^2+4\lambda_1} \\ \hat{\Delta}_1 - \hat{\Delta}_2 &= \frac{d_{11}(\bar{d}_{11}+\bar{d}_{21})}{d_{11}^2+4\lambda_1+8\lambda_2} \end{aligned}$$

Step 3: Out-of-Sample Mean-Squared Error

The optimal λ_1 and $\lambda_2$minimize the expected out-of-sample error for the aggregated treated outcome:

$$\mathbb{E}[MSE] = \mathbb{E}[(\frac{1}{2}(\bar{d}_{12} - \hat{\Delta}_1 d_{12}) + \frac{1}{2}(\bar{d}_{22} - \hat{\Delta}_2 d_{12}))^2] = \mathbb{E}[(\bar{d}_2' - \frac{1}{2}(\hat{\Delta}_1 + \hat{\Delta}_2)d_{12})^2],$$

where $\bar{d}_t' = Y_{0t} - Y_{1t}$, the between aggregated unit difference.

Substituting the solutions for $\hat{\Delta}_1 + \hat{\Delta}_2$ yields

$$\begin{aligned} \mathbb{E}[(\bar{d}_2' - \frac{1}{2}(\hat{\Delta}_1 + \hat{\Delta}_2)d_{12})^2] &= \mathbb{E}[(\bar{d}_2' - \frac{1}{2}\frac{d_{11}(\bar{d}_{11}+\bar{d}_{21})}{d_{11}^2+4\lambda_1}d_{12})^2] \\ &= \mathbb{E}[(\bar{d}_2' - \frac{d_{11}\bar{d}_1}{d_{11}^2+4\lambda_1}d_{12})^2], \end{aligned}$$

which is, up to a scaling constant on λ_1, identical to the expected MSE when the treated unit is not aggregated. Notably, this expression is independent of λ_2, the penalty parameter governing heterogeneity across disaggregated treated units. Hence, if the goal is to predict the aggregated outcome, tuning λ_2 is irrelevant and the focus should be on λ_1.

When I depart from this very stylized world—for example by allowing population-weighted aggregation or control units of different sizes—disaggregation can matter. In those richer settings the aggregate prediction becomes a weighted linear combination of the county-level predictions, so the mapping from(Δ_1, Δ_2) to the aggregate prediction no longer reduces to a simple sum and λ_2 will generally enter the out-of-sample MSE.