Synthetic Control Method under Interference: Detecting and Correcting Bias

Joao Alipio-Correa*

November 24, 2025

1 Introduction

Synthetic control (SC) methods have become a central tool in causal inference, especially for comparative case studies with one or a few treated units. SC is attractive because it offers a transparent way to construct credible counterfactuals without strong functional form assumptions. Instead of extrapolating trends from a model, SC builds a synthetic counterfactual by assigning weights to untreated donor units so that their weighted average reproduces the treated unit's pre-intervention trajectory(Abadie, 2021; Abadie & Gardeazabal, 2003; Abadie et al., 2010, 2015). Its appeal is evident in canonical applications: quantifying the effect of California's Proposition 99 tobacco-control program on cigarette sales (Abadie et al., 2010), assessing the macroeconomic consequences of German reunification (Abadie et al., 2015), and evaluating environmental and health policies in comparative case studies (e.g., Kikuta, 2020; Kreif et al., 2016). The sparsity and interpretability of SC (few key donors with explicit weights) are a large part of its appeal, but as we discuss, they also create a unique vulnerability when there is interference or policy spillovers between units.

This attractiveness just described comes from SC's design. Unlike conventional difference-in-differences (DiD), SC avoids imposing a parametric form on outcome trends and instead constructs its counterfactual by drawing on untreated comparison units—commonly called *donors*. Each donor is assigned a weight via optimization so that their average trajectory reproduces that of the treated unit before the intervention. This pre-treatment similarity, often described as *fit*, is the foundation on which SC builds its counterfactual: the method assumes that if the

*Ph.D. student in Political Science and M.S. student in Statistics at the University of Pittsburgh.
Contact: jac736@pitt.edu; https://joaoalipiocorrea.github.io

synthetic reproduces the treated unit well before the intervention, it will also approximate what would have happened afterward in the absence of treatment (Abadie, 2021).

A potentially unrealistic assumption follows from this design choice. Because SC uses post-treatment donor outcomes to form the counterfactual, it naturally relies on the no-interference component of SUTVA—unit i's outcome depends only on its own treatment, not on others' assignments (Holland, 1986; Rubin, 1980). In spatially and networked settings, where policies plausibly diffuse, this assumption is especially prone to violation. When policies propagate across space or networks, outcomes may depend on neighbors' treatments; formal treatments describe interference by mapping the treatment vector to unit-level exposure via an exposure function (Aronow & Samii, 2017; Hudgens & Halloran, 2008).

All causal inference designs lean on SUTVA as a baseline assumption, but SC is especially vulnerable when spillovers are present. In designs such as difference-in-differences, matching, or regression discontinuity (RDD), contamination can attenuate estimated effects, but the influence of any single unit is inherently limited by the structure of the estimator. In DiD, the average outcome of many untreated units forms the counterfactual, so spillovers from a few controls are diluted in the aggregate. In matching, treated units are paired with multiple comparators, reducing the leverage of any single contaminated control. In RDD, identification relies on local contrasts around a cutoff, where spillovers near the threshold may blur the comparison but typically affect both sides symmetrically.

By contrast, SC deliberately assigns large weights to a small set of untreated units to reproduce the treated unit's pre-period trajectory. If even one of these heavily weighted donors is exposed to spillovers, the resulting contamination enters the synthetic counterfactual directly and is magnified in proportion to the assigned weight. In this sense, interference does not simply add noise but systematically biases the SC estimator, since the very units that best approximate the treated unit before intervention are the ones most likely to drive bias after treatment.

Practice-oriented discussions in SC emphasize donor-pool curation to avoid such pitfalls (Abadie, 2021), but ad hoc trimming does not resolve the underlying tension between fit and exposure, and have other negative consequences. Related approaches—augmented or regularized SC and factor-model variants—can stabilize estimation yet still inherit the same reliance on uncontaminated donors in the post-period (Abadie & L'Hour, 2021; Ben-Michael et al., 2021; Doudchenko & Imbens, 2016). A common remedy is to remove suspect neighbors or to adapt

methods developed for settings with multiple treated units (e.g., Cavallo et al., 2013; Firpo & Possebom, 2018; Kreif et al., 2016; Robbins & Saunders, 2017; Xu, 2017).

The central concern for our purposes is efficiency: the very donors most informative for pre-treatment fit are often geographically proximate and thus the most at risk for spillovers; discarding them can impose a large information cost. In practice, trimming contaminated donors removes precisely those units that often most closely resemble the treated unit. The immediate consequence is a weaker pre-treatment balance, as the synthetic no longer aligns as tightly with the treated unit's trajectory before intervention. With fewer available controls, weights are then spread over less comparable donors, often forcing extrapolation and inflating the variance of the estimate. For applied researchers, this creates a credibility problem: if the pre-period match deteriorates after trimming, the very basis for trusting the post-treatment counterfactual (its ability to track history) breaks down.

With interference, the SC estimator incorporates outcomes from donors that may themselves be affected by the treatment. Let Y_{1t} denote the treated unit's outcome and $\hat{Y}_{1t}^{N} = \sum_{j \neq 1} w_j Y_{jt}$ the synthetic predictor. If donor j receives a spillover δ_{jt} in the post-treatment period, its outcome can be written as $Y_{jt} = Y_{jt}(0) + \delta_{jt}$, where $Y_{jt}(0)$ is the outcome absent exposure. The synthetic predictor therefore includes an additional term $\sum_{j \neq 1} w_j \delta_{jt}$, so that the SC estimand can be rewritten as $\hat{\tau}_t = \big(Y_{1t}(1) - \sum_{j \neq 1} w_j Y_{jt}(0)\big) - \sum_{j \neq 1} w_j \delta_{jt}$. The first term captures the treatment effect of interest under no interference, while the second term reflects a weighted average of spillovers among donors. Hence, whenever exposed donors receive positive or negative spillovers, the SC estimate is systematically biased in proportion to the weights assigned to those donors.

We take a design-based route that targets the object through which interference contaminates SC: the weights. The first step is a diagnostic that asks whether units located near the treated unit evolve differently after treatment than those farther away. To formalize this, we partition the donor pool into concentric distance rings and compare post- versus pre-treatment changes across rings. The null hypothesis is a Fisher sharp null of no effect: outcomes should display no systematic difference by proximity to the treated unit around treatment time. We evaluate this null using randomization inference, which yields exact p-values in finite samples under minimal assumptions (Bowers et al., 2017; Fisher, 1935; Rosenbaum, 2002). Naturally, this exactness property is not tied to the SC estimator, as it follows directly from the design-based logic of

randomization inference and holds regardless of the model used to construct counterfactuals.

Second, conditional on rejecting the null and detecting evidence of proximity-related spillovers, we introduce three corrections that incorporate *spatial reach*—a continuous exposure mapping from distance to a proximity score—directly into the SC optimization, using only pre-period information and leaving outcomes in the post-period untouched.

Our framework rests on two key insights. First, we formalize spatial exposure using the exposure-mapping framework from the interference literature (Aronow & Samii, 2017; Hudgens & Halloran, 2008; Manski, 2013; Tchetgen Tchetgen & VanderWeele, 2012; T. VanderWeele, 2015) and show that, within SC, the impact of interference reduces to a single bias pathway: the aggregate weight mass assigned to likely exposed donors. A simple decomposition makes this link explicit and motivates both our diagnostic and the structure of the corrections. Second, we design estimators that operate directly on this pathway while preserving, when desired, the convex-combination interpretation that makes SC attractive for applied work.

The central problem in place is sensitivity to interference: because SC constructs the counterfactual from post-treatment outcomes of untreated donors, any positive weight on donors whose outcomes are affected by spillovers propagates that exposure into the synthetic and, hence, into the estimated effect.

Our adjustments address this problem by modifying the geometry of the estimator so as to limit the influence of proximate donors, each in a distinct way. *Covariate rescaling* alters the predictor space itself: by multiplying each donor's predictor column by a proximity score, nearby donors appear less similar to the treated unit, shifting the feasible convex hull toward safer comparisons while weights remain on the simplex. *Constrained ridge* leaves the predictor space unchanged but penalizes exposure directly in the weight optimization: an exposure-weighted ℓ_2 penalty shrinks mass away from high-risk donors while preserving convexity and the interpretation of weights as a convex combination, linking directly to penalized SC variants that temper extrapolation with regularization (Abadie & L'Hour, 2021; Ben-Michael et al., 2021; Doudchenko & Imbens, 2016). *Unconstrained ridge* relaxes the simplex entirely, allowing negative and non-summing weights, and applies an exposure-weighted ridge penalty to dampen or even offset contaminated donors; this yields the most aggressive reduction of exposure risk but sacrifices the convex-combination interpretation, following the classical ridge/Tikhonov tradition (Hastie et al., 2009; Hoerl & Kennard, 1970). In all three cases, geographic or network structure provides an exogenous

measure of proximity, and tuning is based solely on pre-treatment information, ensuring that adjustments are guided by design rather than post-treatment outcomes.

Our evidence is designed both to evaluate the properties of the proposed adjustments and to demonstrate their usefulness in applied settings. We begin with a simulation study that varies the magnitude of the treated effect and the intensity of spillovers, showing that baseline SC can be substantially biased while the adjustments consistently reduce bias without sacrificing pre-treatment fit. We then turn to four canonical applications. Applying the ring diagnostic to the original donor pools in Abadie et al. (2010), Abadie et al. (2015), and Kikuta (2020), as well as to a U.S. policy case of the sort considered by Ben-Michael et al. (2021), we find no evidence of interference in the original designs. By contrast, when we expand the West Germany study to a broad, data-complete donor set (roughly 150 countries constructed to mirror the original predictors and preprocessing), the diagnostic detects interference. On that expanded design, the spatial-reach corrections materially alter the estimated effect, underscoring the practical consequences of explicitly accounting for spillovers in SC.

The paper is organized as follows. Section 2 situates SUTVA in SC and reviews exposure mappings for interference. Section 3 develops the ring diagnostic and randomization test. Section 4.1 introduces spatial reach and covariate rescaling. Section 5 develops constrained ridge; an unconstrained ridge variant follows thereafter. We then report simulations and revisit the empirical applications, including the expanded West Germany design. A final section discusses practical guidance for diagnosis, design, and interpretation when interference is a live concern in synthetic control analyses.

1.1 Relevant literature

Our work connects to and extends several strands of recent literature on synthetic control, interference, and spatial dependence. First, a large literature has generalized the original SC framework of Abadie and Gardeazabal (2003) and Abadie et al. (2010, 2015) into a broader class of regularized and factor-based panel estimators. The Augmented Synthetic Control Method of Ben-Michael et al. (2021) blends SC with outcome regression to obtain doubly robust estimates and improved pre-treatment balance even when a pure SC fit is poor. Abadie and L'Hour (2021) propose a penalized synthetic control estimator that introduces an ℓ_2 penalty on weights to stabilize estimation in disaggregated settings. Related approaches recast SC within low-rank or

interactive fixed-effects frameworks, including matrix completion methods for causal panel data (Athey et al., 2021), generalized synthetic control (Xu, 2017), synthetic difference-in-differences (Arkhangelsky et al., 2021), and comparisons of SC to interactive fixed-effects models in applied regional evaluations (e.g., Gobillon & Magnac, 2016). These methods primarily address bias from extrapolation and limited pre-treatment fit; by contrast, our focus is on bias from interference, and our exposure-weighted adjustments can, in principle, be combined with any of these estimators.

Second, our framework builds on the potential-outcomes literature on interference and exposure mappings. Hudgens and Halloran (2008) and Aronow and Samii (2017) formalize interference by mapping the joint treatment assignment into unit-level exposure variables and studying identification under partial interference. Subsequent work has developed sensitivity analyses and estimands under unknown or complex interference patterns (e.g., Manski, 2013; Sävje et al., 2021; Tchetgen Tchetgen & VanderWeele, 2012; T. VanderWeele, 2015; T. J. VanderWeele et al., 2014). Randomization-based approaches extend Fisherian inference to interference settings using exposure mappings and exact tests (Athey et al., 2018; Basse et al., 2019; Bowers et al., 2017). We adopt this design-based perspective: geography (or a known network) provides an exposure mapping, and our ring diagnostic is a randomization test tailored to proximity-patterned interference in SC designs.

Third, there is a growing literature that directly studies spillovers in SC and closely related factor-model estimators. Cao and Dowd (2019) analyze estimation and inference for SC in the presence of spillover effects under a linear factor structure, while Di Stefano and Mellace (2020) propose the inclusive Synthetic Control Method. Parallel developments in spatial econometrics model spillovers parametrically through spatial-lag or SLX specifications and have been combined with difference-in-differences designs to separate direct and indirect effects (Delgado & Florax, 2015; Vega & Elhorst, 2015). Our contribution is complementary: rather than specifying a parametric spatial-lag model or a factor structure for spillovers, we use an exposure mapping and pre-treatment-only information to (i) diagnose proximity-patterned spillovers and (ii) modify the SC weight geometry so as to limit the influence of exposed donors.

2 Spatial/Network Dependence and SUTVA in Synthetic Control

2.1 Synthetic control versus other designs under interference

Causal analyses of comparative case designs rest on the *Stable Unit Treatment Value Assumption* (SUTVA), which asserts that unit i's potential outcome under a joint assignment $(z_1, \ldots, z_{J+1})$ depends only on its own assignment z_i, and that there are no hidden versions of treatment (Holland, 1986; Rubin, 1980). Formally, SUTVA restricts $Y_i(z_1, \ldots, z_{J+1})$ to $Y_i(z_i)$ and guarantees a single well-defined potential outcome for each treatment level. Together with consistency $(Y_i^{\text{obs}} = Y_i(z_i))$, this delivers the identification bridge that causal designs exploit: when a donor j is untreated $(z_j = 0)$, its observed outcome equals its no-treatment potential outcome, $Y_{jt}^{\text{obs}} = Y_{jt}(0)$, regardless of the assignments received by other units.

Hence untreated units can supply information about the *missing* no-treatment path of a treated unit, $Y_{1t}(0)$, by enabling cross-unit substitution of observed outcomes for potential outcomes. The same invariance underpins placebo and permutation procedures: randomization or placebo-in-space tests rely on the distributional equivalence induced by SUTVA to compare treated units to re-labeled controls without introducing assignment-dependent distortions. Absent SUTVA, observed donor outcomes would in general be $Y_{jt}(0; z_{-j})$ and could vary with other units' assignments, severing the link $Y_{jt}^{\text{obs}} = Y_{jt}(0)$ that comparative estimators require and undermining the causal interpretation of cross-unit contrasts.

When policies propagate through geography or networks, outcomes may depend on neighbors' assignments and, as discussed above, SUTVA is violated. In such cases interference must be addressed explicitly, whether through exposure mappings that compress the assignment vector into lower-dimensional summaries of relevant neighbors (Aronow & Samii, 2017; Hudgens & Halloran, 2008; Manski, 2013), through randomization-based testing strategies (Bowers et al., 2017; Rosenbaum, 2007), or through epidemiological and statistical approaches that study identification, estimands, and sensitivity analysis under interference (e.g., Sävje et al., 2021; Tchetgen Tchetgen & VanderWeele, 2012; T. J. VanderWeele et al., 2014). For synthetic control (SC), these issues are especially acute: because post-treatment donor outcomes enter directly into the construction of the counterfactual, even partial exposure of donors introduces bias into the estimator. Interference is thus not a secondary complication but a direct pathway through which SC can be distorted—a point we develop formally further below.

Within this broader landscape, synthetic control (SC) is attractive because it enables transparent, design-based comparisons in comparative case studies with few units (Abadie, 2021; Abadie & Gardeazabal, 2003; Abadie et al., 2010, 2015). To see why SUTVA plays a distinctive role in SC, recall how the estimator is constructed. Let $X_1 \in \mathbb{R}^K$ denote the treated unit's pre-treatment predictors and $X_0 \in \mathbb{R}^{K\times|\mathcal{J}|}$ the same predictors for donors $\mathcal{J} \subseteq \{2,\dots,J+1\}$. SC selects weights $w \in \Delta = \{w \geq 0 : \mathbf{1}^\top w = 1\}$ to minimize the pre-period discrepancy $\|X_1 - X_0 w\|_V^2$ with $V \succ 0$ diagonal, and then applies the same convex weights to donor outcomes in the post-treatment period. This design pays for minimal trend structure with a crucial reliance on SUTVA: because the post-period donor outcomes enter the synthetic predictor, they must correspond to units with no direct treatment and no indirect exposure. If donors are even partially exposed, the synthetic counterfactual absorbs part of the treated effect and the estimated impact is distorted.[1] This mechanism is the synthetic-control manifestation of the bias channel emphasized in linear-factor settings when post-period shocks correlate with the weights (Ferman & Pinto, 2021).

The nature of this distortion can be summarized in a compact bias expression. For post-treatment periods $t > T_0$, the SC estimator based on observed donors admits the decomposition

$$\hat{\tau}_t \;=\; \big[\, Y_{1t}(0) - Y_{1t}^{\mathrm{SC}}(0) \,\big] \;+\; \tau_t \;-\; \rho\,\tau_t \sum_{j\in N} w_j, \tag{1}$$

where the three terms correspond respectively to (i) a *SUTVA synthetic mismatch* between the treated unit's no-treatment outcome and its synthetic approximation, (ii) the *direct treatment effect* τ_t, and (iii) a *contamination mass* equal to the treated effect multiplied by the weight assigned to exposed donors. This decomposition highlights that even if the synthetic perfectly reproduces the treated unit's no-treatment path under SUTVA, $\hat{\tau}_t$ will still differ from τ_t whenever positively weighted donors are exposed. A simplified version makes this transparent:

$$\hat{\tau}_t \;=\; \tau_t \;-\; \rho\,\tau_t \sum_{j\in N} w_j. \tag{2}$$

Equation (1) will be derived formally in the subsection below, but we present it here to build intuition: SC estimates equal the true effect minus a weighted spillover term, with the magnitude

[1] The typical case is attenuation: the synthetic estimate is pulled toward the treated outcome when spillovers raise donor outcomes in the same direction as the treatment effect, yielding an underestimate of the true effect. However, the bias need not always be attenuating. If spillovers move donor outcomes in the opposite direction of the treated effect, the synthetic may overshoot, producing an amplified estimate of the effect. This possibility highlights that interference does not merely add noise but can fundamentally alter the direction and magnitude of the estimated effect (as we will exhibit with our replications

of bias governed jointly by the diffusion intensity ρ, the treatment effect τ_t, and the weight mass placed on exposed donors.

Equation (1) highlights the mechanism through which spillovers distort synthetic control: any exposure among positively weighted donors feeds directly into the estimated effect. To appreciate the severity of this channel, it is useful to situate SC relative to other common designs. All causal estimators rely on SUTVA to justify using untreated outcomes as counterfactuals, but the manner in which donor outcomes enter differs sharply across designs, and so too does the impact of violations.

In difference-in-differences, for example, the control mean aggregates outcomes across many untreated units, assigning each only modest influence. Contamination of a few controls typically attenuates the estimate, but the effect is diluted across the pool and becomes material only if spillovers are widespread. Matching designs are more localized: bias arises when a treated unit is paired with an exposed control, but the damage is confined to that pair and does not propagate automatically to others. Regression discontinuity designs present yet another contrast: because identification exploits outcomes just above and below the cutoff, spillovers would bias the design only if they differentially affect the two sides of the threshold; broader or symmetric diffusion often cancels out.

Synthetic control stands apart. The estimator compresses the counterfactual for the treated unit into a single convex combination of donors, often dominated by a handful of high-weight contributors chosen precisely for their ability to replicate pre-treatment trajectories. This sparsity is central to SC's appeal: by concentrating on a few well-aligned donors, the method achieves close balance without imposing parametric trend restrictions. Yet the same feature that delivers this balance under SUTVA makes the design acutely vulnerable under violations. If even one heavily weighted donor is exposed to spillovers, the contamination is not averaged away, confined to a pair, or offset symmetrically; it flows directly into the counterfactual in proportion to the assigned weight. In this setting, interference is not averaged out, localized, or offset symmetrically: it enters deterministically through the weights.

A single exposed donor can shift the estimated effect in proportion to its assigned weight, and the donors most capable of reproducing the treated trajectory—the ones the optimization favors—are frequently those most plausibly exposed. In this sense, SC transforms what might be a modest attenuation in DiD or a pair-specific distortion in matching into a direct and systematic

bias channel. The contamination mass term in equation (1) formalizes this: the post-treatment estimate depends not only on the direct effect but also on the weighted share of the treated effect leaking into the synthetic. For this reason, SC does not merely inherit SUTVA as a background condition for identification—it is structurally entangled with it. Donor outcomes after treatment are the building blocks of the counterfactual, so once they are contaminated, bias enters mechanically through the weights. This is not a secondary complication but a first-order vulnerability. Later sections formalize this point: the bias decomposition in equation (1) shows that the estimated effect consists of the true effect plus a contamination mass proportional to the total weight placed on exposed donors. Effective corrections must therefore operate directly on this pathway, reshaping the weight structure that transmits interference into the estimate.

In practice, one response to the risk of spillovers is to restrict the donor pool by trimming proximate or substantively linked units, or to adapt estimators designed for settings with multiple treated units (e.g., Cavallo et al., 2013; Firpo & Possebom, 2018; Kreif et al., 2016; Robbins & Saunders, 2017; Xu, 2017). Let $\mathcal{J}^*$ denote the full donor set admissible under standard SC assumptions and $\mathcal{J} \subseteq \mathcal{J}^*$ the subset retained after trimming. Because $\mathcal{J}$ is a strict subset, the optimization problem that defines the synthetic weights is solved over a reduced feasible region. While trimming may lower the exposed weight mass $\sum_{j \in N} w_j$ and thereby reduce the contamination term in equation (1), it simultaneously constrains the construction of $Y_{1t}^{SC}(0)$ to a smaller space of convex combinations.

The consequence is that pre-treatment imbalance $\|X_1 - X_0 w\|_V$ cannot improve and will typically worsen, leading to a larger design-driven discrepancy $Y_{1t}(0) - Y_{1t}^{SC}(0)$. In finite samples this trade-off manifests concretely: fewer admissible donors reduce the effective dimensionality of the comparison set, variance of the estimator increases, and the credibility of the counterfactual trajectory is weakened when the synthetic fails to reproduce the treated unit's pre-treatment path. In short, trimming addresses one bias pathway by construction but aggravates another, and the resulting estimator may be both noisier and harder to interpret. Regularized and augmented variants of SC (Abadie & L'Hour, 2021; Arkhangelsky et al., 2021; Athey et al., 2021; Ben-Michael et al., 2021; Doudchenko & Imbens, 2016; Gobillon & Magnac, 2016) partially stabilize estimation, but because they continue to rely on post-period donor outcomes, they inherit the same interference sensitivity unless exposure is incorporated directly into the design.

These considerations motivate the approach we take in the remainder of the paper. We

operationalize interference using an outcome-agnostic, monotone exposure mapping based on geography (or a known network) and show that SC's sensitivity to spillovers can be reduced by acting directly on the object through which contamination flows—the weights. In the next section we employ a randomization-based diagnostic that tests for proximity-patterned changes in donor outcomes around the treated unit, and subsequent sections incorporate the exposure mapping into estimation through covariate rescaling and ridge augmentations that shift mass away from likely exposed donors while preserving pre-period fit and, when desired, the convex-combination interpretation that makes SC useful for applied work.

2.2 Formalizing the distortion caused by interference

SUTVA's locus inside SCM is immediate: for $\hat{Y}_{1t}^{(0)}$ to represent $Y_{1t}(0)$, each positively weighted donor must be observed under no direct treatment *and* no indirect exposure to the treated unit. Abadie's overview emphasizes this point in practice: researchers often curate donor pools to avoid neighboring or economically linked regions precisely to protect the counterfactual from spillovers (Abadie, 2021). Beyond such design heuristics, there is now a growing literature that is working on dealing with spillovers in SCM and factor models, estimating both direct and indirect effects (e.g., Cao & Dowd, 2019; Di Stefano & Mellace, 2020). Parallel developments in spatial econometrics likewise model spillovers parametrically (e.g., SLX or spatial-lag specifications) and have been combined with difference-in-differences to separate direct and indirect effects, underscoring the importance of interdependence in social sciences applications (Delgado & Florax, 2015; Vega & Elhorst, 2015).

To study the consequences of interference in a transparent way, we adopt a stylized diffusion approach that isolates the bias channel in SCM. The construction proceeds as: it separates the problem into three quantities. First, the size of the treated effect τ_t, which determines the magnitude of the policy shock. Second, the diffusion intensity $\rho_j \in [0, 1]$, which governs the fraction of the treated effect transmitted to each donor j. Third, the exposed weight mass $\sum_{j \in N} w_j$, which captures how much of the synthetic counterfactual relies on donors subject to spillovers. This decomposition clarifies that contamination arises not from a complex interaction of dynamics but from the mechanical way in which treated effects, diffusion intensity, and donor weights combine.

Formally, let the treated unit be indexed by $i = 1$ and donors by $j \in \mathcal{J}$. For a post-treatment

period $t > T_0$, the observed outcome for the treated unit is

$$Y_{1t}^{obs} \;=\; Y_{1t}(0) \;+\; \tau_t,$$

where $Y_{1t}(0)$ is the no-treatment potential outcome and τ_t is the direct effect. For donors, we allow localized, time-homogeneous diffusion with heterogeneous intensities: there is a subset $N \subseteq \mathcal{J}$ such that, for $j \in N$, the observed outcome is

$$Y_{jt}^{obs} \;=\; Y_{jt}(0) \;+\; \rho_j\,\tau_t, \qquad \rho_j \in [0,1], \tag{3}$$

while for $j \notin N$, $Y_{jt}^{obs} = Y_{jt}(0)$ (no spillover). This formulation isolates the spillover channel in a transparent way: contamination is governed by three elements — the size of the treated effect τ_t, the donor-specific diffusion intensities $\{\rho_j\}$, and the weight mass assigned to exposed donors.

Define the SUTVA synthetic counterfactual (the path SCM would target absent interference) as

$$Y_{1t}^{\mathrm{SC}}(0) \;:=\; \sum_{j\in\mathcal{J}} w_j\, Y_{jt}(0). \tag{4}$$

Theorem 1 (Bias of SCM under localized spillovers)**.** *For $t > T_0$, the SCM estimator based on* observed *donors satisfies*

$$\hat{\tau}_t \;=\; \underbrace{\big[\,Y_{1t}(0) - Y_{1t}^{\mathrm{SC}}(0)\,\big]}_{\textit{SUTVA synthetic mismatch}} \;+\; \underbrace{\tau_t}_{\textit{direct effect}} \;-\; \underbrace{\tau_t \sum_{j\in N} \rho_j w_j}_{\textit{contamination mass}} \quad . \tag{5}$$

In particular, if $Y_{1t}^{\mathrm{SC}}(0) = Y_{1t}(0)$ (the SUTVA synthetic counterfactual coincides with the treated unit's no-treatment path at t)„ then

$$\hat{\tau}_t \;=\; \tau_t \;-\; \tau_t \sum_{j\in N} \rho_j w_j. \tag{6}$$

Proof. Starting from the definition of the SCM effect:

$$\hat{\tau}_t \;=\; Y_{1t}^{obs} - \sum_{j\in\mathcal{J}} w_j\, Y_{jt}^{obs}. \tag{7}$$

For the treated unit, $Y_{1t}^{obs} = Y_{1t}(0) + \tau_t$. For donors, $Y_{jt}^{obs} = Y_{jt}(0) + \rho_j\,\tau_t$ if $j \in N$, and

$Y_{jt}^{obs} = Y_{jt}(0)$ otherwise. Substituting, we obtain

$$\hat{\tau}_t = \big[Y_{1t}(0) + \tau_t\big] - \Big[\sum_{j\in\mathcal{J}} w_j Y_{jt}(0) \;+\; \tau_t \sum_{j\in N} \rho_j w_j\Big] \tag{8}$$

$$= \big[Y_{1t}(0) - Y_{1t}^{\mathrm{SC}}(0)\big] + \tau_t - \tau_t \sum_{j\in N} \rho_j w_j, \tag{9}$$

which establishes (19). The specialization (20) follows by setting $Y_{1t}^{\mathrm{SC}}(0) = Y_{1t}(0)$. □

Equation (19) decomposes the estimator into three interpretable components: the SUTVA synthetic mismatch $Y_{1t}(0) - Y_{1t}^{\mathrm{SC}}(0)$, the direct effect τ_t, and the contamination mass $\tau_t \sum_{j\in N} \rho_j w_j$ generated mechanically by assigning weight to exposed donors. The last term is the explicit bias pathway introduced by interference; its magnitude depends on the effect size τ_t, donor-specific diffusion intensities $\{\rho_j\}$, and the weight mass on exposed donors.

The following corollaries illustrate the basic mechanics of this decomposition under the simplifying condition $Y_{1t}^{\mathrm{SC}}(0) = Y_{1t}(0)$ (perfect pre-treatment balance under SUTVA). They clarify how the sign, magnitude, and bounds of the distortion follow directly from the contamination mass.

Corollary 1 (Sign, attenuation, and scale)**.** *Suppose $Y_{1t}^{\mathrm{SC}}(0) = Y_{1t}(0)$ and $\rho_j \in [0,1]$ for all $j \in N$. Then:*

1. *(Sign/attenuation) If $\sum_{j\in N} \rho_j w_j > 0$ and $\tau_t > 0$, then $\hat{\tau}_t < \tau_t$; if $\sum_{j\in N} \rho_j w_j > 0$ and $\tau_t < 0$, then $\hat{\tau}_t > \tau_t$. That is, positive weights on exposed donors attenuate estimated effects toward zero.*

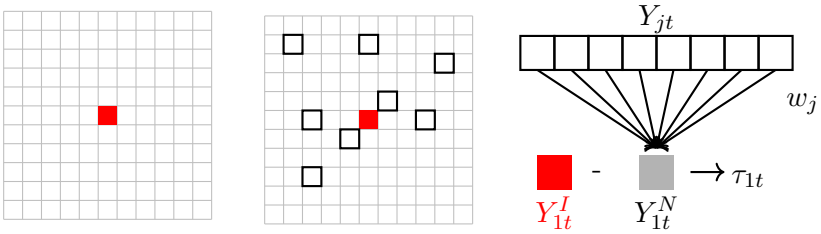
2. *(Monotonicity in exposure) Holding τ_t fixed, $|\hat{\tau}_t - \tau_t|$ is nondecreasing in each ρ_j and in the exposed weight mass $\sum_{j\in N} w_j$. More intense or more widespread spillovers cannot reduce bias.*

3. *(Bounds) Because $0 \le \rho_j \le 1$ and $w \in \Delta$, the distortion satisfies $0 \;\le\; |\hat{\tau}_t - \tau_t| \;\le\; |\tau_t|$. The upper bound is attained only if the entire weight is placed on fully exposed donors ($\rho_j = 1$ for all j with $w_j > 0$).*

Proof. From (20), $\hat{\tau}_t - \tau_t = -\tau_t \sum_{j\in N} \rho_j w_j$. Because each $\rho_j \ge 0$ and $w_j \ge 0$, the sign of $\hat{\tau}_t - \tau_t$ is the opposite of the sign of τ_t, establishing (i). Statement (ii) follows since the absolute deviation

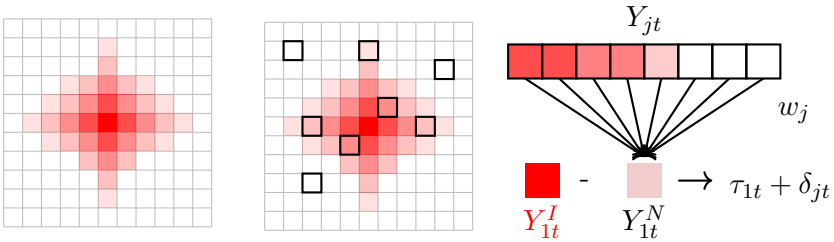
is proportional to $\sum_{j \in N} \rho_j w_j$, which is monotone in each ρ_j and in exposed weights. For (iii), note that $0 \leq \sum_{j \in N} \rho_j w_j \leq 1$, which yields $|\hat{\tau}_t - \tau_t| \leq |\tau_t|$, with equality only under the stated condition. □

These properties highlight that once exposed donors receive positive weight, bias is inevitable: the estimator cannot cancel it. Moreover, the distortion grows monotonically in either the intensity of exposure (ρ_j) or the mass of weights placed on exposed donors. This makes clear why interference in SCM is not a secondary nuisance but a direct and systematic channel of bias.



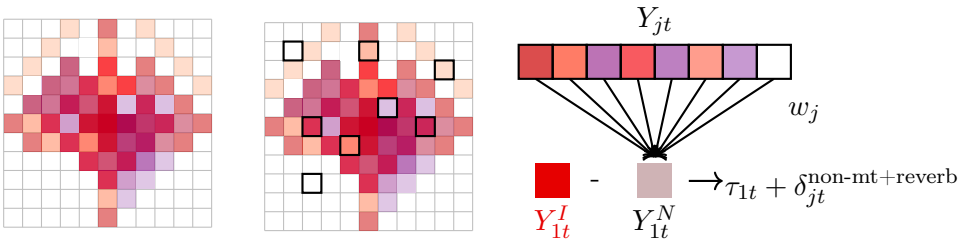
1: Treated Unit 2: $w_j > 0$ units 3: Treatment Effect

(a) No interference



1: Treated unit + 2: $w_j > 0$ units Contaminated τ_{1t}

(b) Monotone Interference with no reverberation



Treated unit ++ $w_j > 0$ units Contaminated τ_{1t}

(c) Non-monotone Interference with reverberation

Figure 1: **SUTVA versus interference inside synthetic control.** Each row illustrates how donor contamination enters the synthetic counterfactual through the weights.

(a) No interference: donor outcomes are unaffected, so the synthetic predictor $Y_{1t}^{SC}(0) = \sum_j w_j Y_{jt}(0)$ is unbiased, and the estimated effect equals the direct effect τ_t.

(b) Monotone interference without reverberation: the treated effect diffuses smoothly to nearby donors with intensity ρ_j. If such donors receive positive weights, the synthetic predictor embeds their spillovers, and the estimator obeys $\hat{\tau}_t = \tau_t - \tau_t \sum_{j \in N} \rho_j w_j$, so the estimate is systematically shifted by the contamination mass $\tau_t \sum_{j \in N} \rho_j w_j$. This is the structured form of interference addressed in this paper.

(c) Non-monotone interference with reverberation: spillovers propagate irregularly and feed back through the network, generating heterogeneous donor exposures that need not decay with distance. In this setting, the simple decomposition above does not apply without further assumptions; such cases are beyond the scope of the present analysis.

This decomposition delivers a simple but powerful design principle: bias in SCM under interference is governed by the exposed weight mass, so the estimator must be engineered to suppress that mass while preserving pre-treatment fit and, when desired, the convex-combination

interpretation that makes SCM transparent for applied work. Figure 1 illustrates this principle: in the absence of interference, donor weights cleanly recover the counterfactual; under localized spillovers, contamination flows directly into the estimate through exposed donors; and in more pervasive diffusion settings, further assumptions would be required to disentangle direct from indirect effects. The remainder of the paper operationalizes this principle by defining a continuous, outcome-agnostic proximity score and incorporating it into rescaling and ridge adjustments that systematically redirect weight away from high-risk donors.

3 Detecting Interference

Before constructing a synthetic control, it is useful to diagnose whether donor outcomes were influenced by the treated unit's intervention. Such influence constitutes *interference* and violates SUTVA, which requires each unit's potential outcomes to depend only on its own assignment and not on the assignments of other units (Rubin, 1980). In comparative case studies, interference could arise when a policy in the treated region diffuses to neighboring regions through geographic, economic, or network linkages, thereby contaminating donors that should represent untreated counterfactuals. The canonical Proposition 99 application of synthetic control (Abadie et al., 2010) already hints at such risks (e.g., cross-border cigarette purchases or policy diffusion), and as we demonstrated above, if spillovers are present, the synthetic counterfactual absorbs a fraction of the treated effect, shifting the estimated treatment effect. In order to screen for this problem ex ante, we develop a diagnostic coupled with randomization inference (Fisher, 1935; Rosenbaum, 2002), which is exact in finite samples and has been adapted to diagnose issues in experimental designs (Bowers et al., 2017; Rosenbaum, 2007).

To formalize what we aim to detect, we write the *joint* potential-outcome vector at time t as

$$\mathbf{Y}_t(z) \;=\; \big(Y_{1t}(z),\ldots,Y_{Nt}(z)\big)^\top, \qquad z=(z_{1t},\ldots,z_{Nt})^\top \in \{0,1\}^N,$$

so that interference is a property of how the assignment vector z co-determines all components of $\mathbf{Y}_t(z)$. We can then let S encode the spatial or network structure, and define for each unit i an exposure mapping $E_{it} = e_{it}(z, S)$ summarizing the aspects of z (given the structure S) that are relevant for i's outcome at t. Importantly, we focus in this paper on a locally dissipating manner in which exposure will aggregates assignments with a kernel κ over distance, but we impose a

restriction over κ so that it is nonincreasing:

$$E_{it} \;=\; \sum_{j=1}^{N} \kappa\big(d_{ij};\theta\big)\, z_{jt}, \quad \kappa(r) \text{ nonincreasing in } r \tag{10}$$

where d_{ij} is the geographic (or network) distance between i and j, and θ collects kernel parameters. This accommodates heterogeneous patterns of diffusion while encoding the substantive regularity that proximity increases exposure. We keep outcomes structural in exposure and do not impose parametric functional form:

$$Y_{it}(z) \;=\; Y_{it}(0) \;+\; \Delta_{it}\big(E_{it}\big), \tag{11}$$

so that all interference operates through E_{it} as induced by (z, S) and the kernel κ.

The diagnostic we propose in the next sub-sections will target precisely whether post-versus-pre changes in donor outcomes align with proximity in a way consistent with (10)-(11). In practice, we take one of many potential routes and operationalize $\kappa(\cdot)$ by discretizing distance into a finite number of rings around the treated unit and comparing outcome changes across near and far groups of rings. Substantively, these rings discretize equal (or similar) exposure sets implied by κ: as donors in the same ring are approximated as having comparable exposure scores. We next formalize this ring partition and the associated test statistic, and then implement a finite-sample exact randomization test that evaluates whether the treated unit sits at the center of a distinctive near-far pattern.

3.1 Distance-Based Ring Partition

We operationalize the exposure kernel $\kappa(\cdot)$ by discretizing distance to the treated center $p^\star$ into K rings, approximating $\kappa(d_{ip^\star};\theta) \approx s_k$ for all donors i assigned to ring k. Let the set of units be $U = \{1, \ldots, N\}$. One unit, denoted $p^\star \in U$ (typically $p^\star = 1$), is the actually treated unit at time T_0. For each donor $i \neq p^\star$, let $d_{ip^\star}$ denote the distance from i to $p^\star$. Choose radii $0 = c_0 < c_1 < \cdots < c_K$ that partition the space around $p^\star$ into K non-overlapping rings, and assign each donor to a ring via

$$r_{ip^\star} = k \quad \Longleftrightarrow \quad c_{k-1} \le d_{ip^\star} < c_k, \qquad k = 1, \ldots, K.$$

In practice, ring boundaries c_k can be chosen by contiguity, by application-informed thresholds, or to yield adequate sample sizes per ring. This construction serves as a discretization of equal-(or similar-)exposure sets implied by the locally dissipating kernel: nearer rings correspond to larger exposure scores. (An idealized symmetric geometry with concentric rings over a grid is depicted in Appendix A.[2])

We then define two disjoint donor groups by aggregating rings. Let $R^A \subset \{1, \ldots, K\}$ denote a *proximal* set (rings suspected to exhibit interference) and $R^B = \{1, \ldots, K\} \setminus R^A$ the complementary *distal* set. The corresponding donor groups are

$$A_{p^\star} = \{\, i \neq p^\star : r_{ip^\star} \in R^A \,\}, \qquad B_{p^\star} = \{\, i \neq p^\star : r_{ip^\star} \in R^B \,\}.$$

By construction, $A_p \cup B_p$ includes all donors (excluding p itself) and $A_p \cap B_p = \emptyset$. Typically we begin with $R^A = \{1\}$ (the innermost ring) and $R^B = \{2, \ldots, K\}$, which is often the most powerful single contrast for detecting any proximity-structured disturbance.

For the diagnostic to be informative, we emphasize the following design conditions as good-practice and guidance for applied researchers:

Coverage: Both $A_{p^\star}$ and $B_{p^\star}$ should be nonempty to permit near–far discrimination, note that extremely small or empty rings undermine power and interpretability. In applications, set c_k with domain knowledge to avoid vanishing cells while preserving a meaningful distance gradient.

Window-level common shocks: In the chosen pre/post window, there should not be global, unit-invariant shocks that shift all rings equally (like a nationwide policy). Such shocks cancel in expectations across rings and naturally erode any contrast. If present, adjust the window or consider an alternative approach to the research problem.

Stability across windows: Ring membership should be stable across the pre and post segments used to form $Z_i^{(w)}$, so that near–far comparisons are not confounded by reclassification. In practice, fix c_k ex ante and verify that the composition of $A_{p^\star}$ and $B_{p^\star}$ does not change when varying windows within reasonable bounds.

Finally, in many applications, policy spillovers plausibly dissipate with distance, so far rings act as a reasonable control group. But notice that if all donors are near (or all are far), or interference is genuinely global, power is correspondingly limited.

[2]Appendix Figure A.1 illustrates the ring geometry (grid with concentric circles centered at $p^\star$) and the mapping from continuous distance $d_{ip^\star}$ to discretized exposure scores $s_k \approx \kappa(d_{ip^\star}; \theta)$.

3.2 Outcome-Change Statistics

Having defined the groups, we now need a statistic that captures outcome changes potentially caused by the treatment. Let T index time and T_0 denote the intervention time for $p^\star$. For a chosen time window w, define disjoint sets $T_w^{\text{pre}}, T_w^{\text{post}} \subset T$ of equal length. For donor i, let

$$\bar{Y}_{i,\text{pre}}^{(w)} = \frac{1}{|T_w^{\text{pre}}|} \sum_{t \in T_w^{\text{pre}}} Y_{it}, \qquad \bar{Y}_{i,\text{post}}^{(w)} = \frac{1}{|T_w^{\text{post}}|} \sum_{t \in T_w^{\text{post}}} Y_{it},$$

and define the change statistic

$$Z_i^{(w)} = \bar{Y}_{i,\text{post}}^{(w)} - \bar{Y}_{i,\text{pre}}^{(w)}. \tag{12}$$

We consider several windows: a full window using all pre- vs. all post-intervention periods; a one-period window using $T_0 - 1$ vs. $T_0 + 1$; and symmetric n-period windows using $\{T_0 - n, \ldots, T_0 - 1\}$ vs. $\{T_0 + 1, \ldots, T_0 + n\}$. To avoid post-treatment leakage, windows w should be pre-specified using domain/contextual knowledge, and while alternative windows may be reported for transparency, ideally the researcher does not perform searches over those to select significance.

Note that $Z_i^{(w)}$ is invariant to additive unit fixed effects over the window (as $Y_{it} = \alpha_i + u_{it}$) because naturally the pre/post difference cancels α_i, focusing the contrast on within-unit shifts plausibly induced by exposure. Under no interference, donors' $Z_i^{(w)}$ should not systematically differ by proximity; with positive spillovers, proximal donors tend to have larger $Z_i^{(w)}$ than distal donors.

Aggregate by group:

$$\bar{Z}_{A_{p^\star}}^{(w)} = \frac{1}{|A_{p^\star}|} \sum_{i \in A_{p^\star}} Z_i^{(w)}, \qquad \bar{Z}_{B_{p^\star}}^{(w)} = \frac{1}{|B_{p^\star}|} \sum_{i \in B_{p^\star}} Z_i^{(w)}.$$

For interpretability and variance scaling, we use a simple two-sample t-statistic[3]

$$t_{p^\star}^{(w)} = \frac{\bar{Z}_{A_{p^\star}}^{(w)} - \bar{Z}_{B_{p^\star}}^{(w)}}{\sqrt{s_P^2 \left(\frac{1}{|A_{p^\star}|} + \frac{1}{|B_{p^\star}|} \right)}}, \qquad s_P^2 = \frac{(|A_{p^\star}| - 1)\, \widehat{Var}(Z_{i \in A_{p^\star}}^{(w)}) + (|B_{p^\star}| - 1)\, \widehat{Var}(Z_{i \in B_{p^\star}}^{(w)})}{|A_{p^\star}| + |B_{p^\star}| - 2}.$$

A positive value of $t_{p^\star}^{(w)}$ indicates the near donors increased more on average than far donors

[3]When ring variances differ materially, researchers should also compute the Welch variant as a robustness check, but notice that the subsequent permutation inference remains finite-sample exact under the null regardless of the variance estimator used in the t-ratio's denominator.

(consistent with positive spillovers), whereas a negative value indicates the opposite pattern. If SUTVA holds (no interference), $t_{p^\star}^{(w)}$ should be close to zero, and a large $|t_{p^\star}^{(w)}|$ signals a proximity-structured disturbance consistent with interference.

3.3 Randomization (Permutation) Inference

Classical large-sample approximations are unreliable with small donor pools. We therefore adopt a design-based randomization test in the spirit of Fisher (1935), treating the identity of the treated center as exchangeable under the null of no proximity-structured disturbance. Formally,

$$H_0: \ Z_i^{(w)} \text{for } i \in U \text{ exhibits no special near-far pattern around } p^\star \text{ relative to any } p \in U,$$

implying exchangeability of the center, which substantively means that the the patterns observed nearby the treated unit are no different from the ones observed far from it. Under H_0, the statistic computed around the actual treated unit should be typical of the distribution of the same statistic computed when each unit is, counterfactually, taken as the center (that is, treated). This allows us to construct the sampling distribution of $|t_{p^\star}^{(w)}|$ under the null by considering other units as placebo-treated. The algorithm is as follows:

Permutation procedure

1. For each $p \in U$, treat p as if it were the treated center at time T_0. Form rings r_{ip}, sets A_p, B_p, and compute $t_p^{(w)}$ as in Section 3.2, yielding the set $\{t_p^{(w)} : p \in U\}$.
2. Identify the treated unit $p^\star$ and its statistic $t_{p^\star}^{(w)}$.
3. Compute the exact two-sided p-value

$$\hat{p}^{(w)} = \frac{1 + \#\{\, p \in U : \ |t_p^{(w)}| \geq |t_{p^\star}^{(w)}| \,\}}{N+1}.$$

Including $p^\star$ among the placebos yields a *finite-sample exact* permutation p-value under H_0 (Fisher, 1935; Rosenbaum, 2002). A small $\hat{p}^{(w)}$ indicates that the treated unit sits at the center of an atypical near–far disturbance in outcomes, consistent with interference emanating from the treatment location; a large $\hat{p}^{(w)}$ suggests no detectable proximity pattern. [4]

[4] This mirrors placebo-in-space in SCM (Abadie et al., 2010) insofar as the reference distribution is built by re-centering the design at each unit. Here the statistic is a proximity contrast rather than an SC effect or loss.

This randomization test effectively asks, "Is the treated unit the center of a unique outcome disturbance that is not seen around other units?" If yes, it suggests the treated unit's intervention radiated outwards (affecting neighbors) in a way that other units (which had no intervention) did not. On the other hand, if the treated unit's $t^{(w)}_{p^\star}$ is indistinguishable from many placebo $t^{(w)}_{p}$'s, then there is no evidence that proximity to the treated mattered for outcomes. The test assumes a single focal treated center drives any proximity pattern. Power is driven by coverage across rings (Section 3.1) and by separation in the group means $\bar{Z}^{(w)}_{A_p} - \bar{Z}^{(w)}_{B_p}$ over the chosen window.

3.4 Alternative Ring Contrasts and the Extent of Interference

The primary contrast $R^A = \{1\}$ vs. $R^B = \{2, \ldots, K\}$ asks whether *any* proximity-structured interference is present and is typically the most powerful single test for any interference. If detected, a natural next step, considering a locally dissipating pattern (monotone in distance in expectation), is to examine how far from the treated center such disturbances remain detectable. This asks: "how far spillovers extend?", and can be done considering a sequence of pooled or adjacent contrasts,

$$\text{Ring 2 vs. Rings 3+}, \qquad \text{Ring } k \text{ vs. Ring } k+1, \qquad \text{Ring } k \text{ vs. Rings } k+1, \ldots, K,$$

recomputing $t^{(w)}_{p}$ and $\hat{p}^{(w)}$ for each contrast. We then locate the smallest index $k^\dagger$ such that the contrast "Ring $k^\dagger$ vs. Rings $k^\dagger+1, \ldots, K$" is no longer statistically distinguishable. [5] Suppose a researcher detects interference in the first ring, and proceeds to the ring 2 vs. rings 3+ contrast, essentially treating the second ring as the "treated" and comparing it to the farther rings. If this test is not significant, it suggests that by the second ring the detected interference has mostly dissipated.

We interpret $k^\dagger$ as the point beyond which, in the data, near–far shifts in donor outcomes are no longer detectable under the locally dissipating model. In practice, adjacent-ring tests may be underpowered when rings are small, and pooling all rings beyond a boundary (e.g., k vs. $k+1, \ldots, K$) improves stability. A schematic illustrating these contrasts appears in Appendix A.7, and the sequence of contrasts may be viewed as a something similar to the standard impulse response function in time-series models, but here, this "spatial impulse-response" summarizes

[5] Because a boundary is identified by inspecting multiple related contrasts, one may report step-down multiplicity adjustments alongside raw permutation p-values. A simple option is Holm's step-down procedure (Holm, 1979). In applications, adjusted p-values can be reported in the main text, with unadjusted values provided for transparency.

the amplitude of the disturbance near the treated unit and its decay with distance.

The ability to detect a boundary $k^\dagger$ (power of the test) improves with (i) larger expected separation in group means $\bar{Z}^{(w)}_{A_p} - \bar{Z}^{(w)}_{B_p}$ as distance increases; (ii) balanced ring sizes (or pooled contrasts that mitigate small-cell noise); and (iii) longer, symmetric pre/post windows that increase signal-to-noise while excluding $t = T_0$. Power is reduced by global shocks within the window that shift all rings similarly and by mis-centered or coarse distance measures that blur true proximity relations. These considerations complement the coverage and window design guidance in Sections 3.1–3.2.

3.5 Decision Rule and Link to Bias Correction

The rings diagnostic yields a design-based decision rule for SCM. If the permutation test fails to reject ($\hat{p}^{(w)}$ large across pre-specified windows), proceed with standard SCM: there is no evidence that proximity to the treated unit altered donor trajectories. If the test rejects ($\hat{p}^{(w)}$ small for at least one plausible window), treat SUTVA as violated and proceed under the working conclusion that donors in proximal rings received nonzero exposure.

In terms of the bias identity introduced earlier, rejection indicates positive *exposed weight mass* ($\sum_{j \in N} w_j > 0$ for donors N that are exposed) so the contamination term is activated and the SCM estimand absorbs a nontrivial fraction of the treated effect. Subsequent estimators should therefore target *suppression of exposed weight mass* while preserving pre-period fit and, when desired, the convex-combination interpretation. The next sections implement this principle via a continuous, outcome-agnostic proximity (reach) score (Section 4.1) and ridge-based variants that downweight donors in proportion to their exposure risk (Section 5).

state	$Z^{(\text{full})}$	$Z^{(\text{year-1})}$	$Z^{(\text{sym-3})}$
Missouri	4.0066	3.9159	3.9381
Iowa	2.3640	2.4193	2.3539
Colorado	-0.0414	-0.1069	0.0060
Vermont	0.02501	-0.1115	-0.0886

(a) Test statistics

state	t_p	A_p	B_p
MO	4.4207	AR, IL, IN, ...	AL, AZ, CA, ...
VT	-0.2169	CT, DE, ME, ...	AL, AZ, CO, ...
CO	0.3428	AZ, MT, NV, ...	AL, CA, CT, ...
IA	-0.3312	MI, MN, SD, ...	AL, AZ, CA, ...

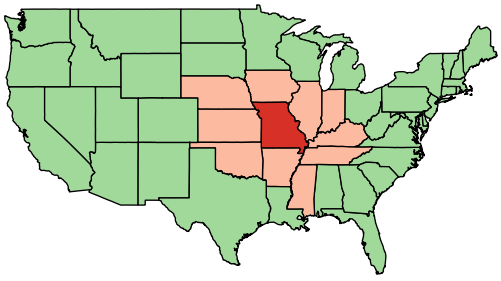
(b) Placebo donor sets

Table 1: Comparison of Z-statistics and placebo donor sets for selected states.

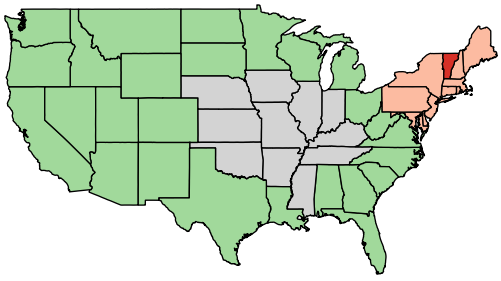
Table 1 summarizes, for a stylized U.S. states panel, unit-level change statistics across three windows and the corresponding placebo-center contrasts. Missouri is treated at T_0 with a direct effect $\tau = 4$, and first-ring neighbors receive a spillover $\rho\tau$ with $\rho \approx 0.6$, decaying with distance.

Panel (a) reports $Z_i^{(\text{full})}$, $Z_i^{(\text{year-1})}$, and $Z_i^{(\text{sym-3})}$ for selected states, and as expected under locally dissipating interference, nearby donors to Missouri display large positive changes, whereas distant donors are near zero. Panel (b) lists, for a few selected placebo centers p, the near-far t-statistic t_p and the membership of A_p (proximal) and B_p (distal) under the primary contrast $R^A = \{1\}$ vs. $R^B = \{2, \dots\}$.

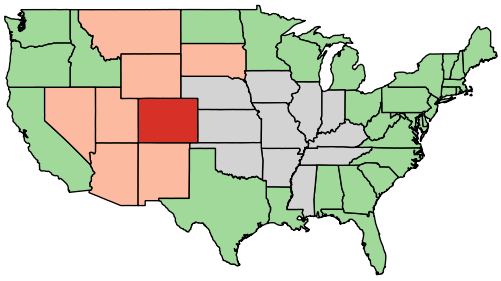
Figure 2 visualizes the primary contrast for the treated center and three placebos. In each panel, the treated center (or placebo center) is marked in red, the proximal set A_p in orange, and the distal set B_p in green. For Missouri (Panel 2a), the near group exhibits a systematically higher $Z_i^{(w)}$ than the far group, yielding a large $t_{p^\star}^{(w)}$. By contrast, Vermont, Colorado, and Iowa (Panels 2b–2d) produce $t_p^{(w)}$ near zero with balanced near-far patterns, with the permutation p-value small ($p = 0.0408$ for the window reported), indicating a distinctive proximity pattern centered on Missouri. These displays make the decision rule concrete: when a unique near-far disturbance is centered at the treated unit, proceed with interference-aware SCM that suppresses exposed weight mass, otherwise, implement standard SCM.



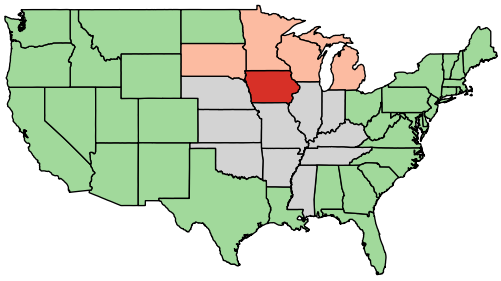
(a) Contrast for Missouri



(b) Contrast for Vermont



(c) Contrast for Colorado



(d) Contrast for Iowa

Figure 2: Contrasts across selected units.

4 Rescaling

4.1 Spatial Reach

We model interference through a continuous, proximity-based exposure score. For each donor j, let $d(j,p)$ denote its distance to the treated unit p. The exposure score is

$$\eta_j \;=\; f\big(d(j,p)\big) \in (0,1),$$

where f is smooth and increasing in distance. We adopt the convention that larger η_j indicates lower exposure: very proximate donors have η_j near zero, distant donors have η_j near one. This score is used as an exposure mapping in the sense of the interference literature, which allows outcomes to depend on other units' assignments through low-dimensional summaries of the assignment vector (Aronow & Samii, 2017; Hudgens & Halloran, 2008; Manski, 2013). The role of η_j here is purely exogenous: it summarizes the plausibility of spillover based on geography. It is not a structural outcome model; rather, it is a measured feature of the design that we will feed into estimation [6].

4.2 Rescaling Adjustment

The design goal is to reduce the contamination channel $\sum_j \rho_j w_j$ by discouraging large weights on donors with low reach while preserving pre-treatment fit and, when desired, the convex-combination interpretation. We operationalize this by folding the spatial reach into the synthetic-control geometry. Let $X_1 \in \mathbb{R}^K$ denote the treated unit's pre-treatment predictors and $X_0 = [X_{\cdot,1}, \ldots, X_{\cdot,J}] \in \mathbb{R}^{K\times J}$ the donor matrix, where standard SCM selects $w \in \Delta_J = \{w \geq 0 : \mathbf{1}^\top w = 1\}$ to minimize $Q(w; X_0) = \|X_1 - X_0 w\|_V^2$. We incorporate reach via the rescaled donor matrix $X_0^* = X_0 \,\mathrm{diag}(\eta)$ and solve the same convex problem with (X_1, X_0^*), so that columns associated with proximate donors are contracted in the metric used to match X_1. This changes only the feasible geometry: the optimizer still chooses a convex combination closest to X_1, but the combinations supported by high-risk donors are down-weighted and mixtures of safer donors become comparatively more attractive for fit.

[6] Many smooth, monotone exposure mappings $f(d)$ are admissible (e.g., logistic, Gompertz, kernel decays). For transparency and outcome-agnostic design, we adopt a simple quantile-anchored specification calibrated from the empirical distance distribution; full details and defaults are provided in Appendix A.5. Our results rely only on the exogeneity and monotonicity of reach and are insensitive to smooth reparameterizations that preserve the rank ordering of donors by proximity.

In practice, whenever a combination of units farther from the treated unit can replace a donor potentially more exposed to contamination (closer to the center), rescaling induces weights to shift from the later to the mixture of safer donors. Theorem 6 below formalizes the consequence: strictly dominated near donors drop from support, and aggregate mass on a more-exposed set of near donors does not increase. Since weights remain on the simplex, the convex-combination interpretation is preserved and standard SCM diagnostics (pre-fit checks and placebo exercises) remain comparable to the baseline. The link to bias is direct: the contamination term depends explicitly on exposed weight mass, and rescaling reduces this mass by construction whenever the donor pool permits coverage by safer combinations.

Implementation is outcome-agnostic and straightforward. Distances and the map $\eta = f(d)$ are fixed and calibrated once from $d(j, p)$, and predictors are standardized across donors prior to applying $\mathrm{diag}(\eta)$ so that the contraction acts on comparable magnitudes. With $V \succ 0$ and η_j bounded away from 0 and 1, the rescaled quadratic loss is strictly curved along feasible directions on the simplex, so the rescaled problem keeps admitting a unique minimizer w^* and responds continuously to small perturbations. While these regularities are not central to the substantive claim below, they guarantee allow the geometric comparison of supports to be made sharply.

From here, two design conditions link the rescaled geometry to exposed-mass reduction. The first is a donor-wise coverage relation stating that some mixture of safer donors can stand in for a given near donor once columns are contracted by reach. The second elevates this donor-level dominance to a band of "more-exposed" donors, enabling an aggregate conclusion about the total weight assigned to that band. Both are expressed below:

Assumption 1 (Strict dominance (coverage) by safer donors)**.** *For a donor m, there exists a set S with $\eta_\ell > \eta_m$ for all $\ell \in S$ and weights $\alpha^{(m)} \in \Delta_{|S|}$ such that, writing $Z^{(m)} = \sum_{\ell \in S} \alpha_\ell^{(m)} X_{\cdot,\ell}$ and $\eta_S = \min_{\ell \in S} \eta_\ell$,*

$$\|X_1 - Z^{(m)}\|_V \le \|X_1 - X_{\cdot,m}\|_V, \qquad \|X_1 - \eta_S Z^{(m)}\|_V < \|X_1 - \eta_m X_{\cdot,m}\|_V.$$

Assumption 1 expresses the replacement logic: before rescaling, the far-only combination matches at least as well as m, and after rescaling, even a conservative scaling of that combination is strictly closer than the scaled near donor. It is a property of the donor pool and the reach map, not of outcomes post-treatment. Effectively, for a given unit near the treated, there is a

combination of units farther away that can effectively replace it and, after rescaling, the loss for this mixture of farther away units is strictly smaller than for the near unit.

Assumption 2 (Band-level dominance occurrence)**.** *Let N denote a "more-exposed" band (e.g., a reach-threshold set). At least one $m \in N$ satisfies Assumption 1 with $S(m) \subseteq N^c$.*

Assumption 2 is a mild coverage requirement at the group level. It does not require every donor in N to be dominated, merely that the band contains at least one dominated unit that can be replaced by safer donors outside N when reach is taken into account. Therefore, with multiple near units, $Z^{(m)}$ needs to be able to replace only one of these potentially contaminated units.

Theorem 2 (Support elimination and exposed-mass dominance under rescaling)**.** *Let w° minimize $\|X_1 - X_0 w\|_V^2$ over Δ_J, and let w^* minimize $\|X_1 - X_0^* w\|_V^2$ over Δ_J, with $X_0^* = X_0 \operatorname{diag}(\eta)$. If donor m satisfies Assumption 1, then $w_m^* \to 0$. If a band N satisfies Assumption 2, then*

$$\sum_{j \in N} w_j^* \;\leq\; \sum_{j \in N} w_j^\circ,$$

with strict inequality whenever some dominated $m \in N$ has $w_m^\circ > 0$.

The logic of the theorem is fairly straightforward. By contracting near columns, rescaling renders certain near donors strictly inferior to available mixtures of safer donors in the metric used for fit. If such a dominated donor were to retain its original weight, an infinitesimal reallocation toward its covering mixture (the combination of donors that could replace it) would strictly reduce the objective while remaining feasible on the simplex, contradicting optimality. This delivers donorwise support elimination. If we aggregate the same replacement argument over a band, we can also conclude that the total mass on the more-exposed set cannot increase after rescaling and must fall when the unrescaled SCM places positive weight on any dominated near donor. The proof formalizes these steps using the KKT system and a feasible-direction construction and is provided in Appendix A.2

This adjustment alters geometry without changing feasibility: weights remain nonnegative and sum to one, and no post-treatment outcomes enter the construction. The next section develops a complementary modification that acts directly on weight magnitudes conditional on reach, using exposure-weighted ridge penalties to further suppress exposed mass while retaining convexity and the convex-combination interpretation.

5 Constrained Ridge Adjustment

The target remains the contamination channel identified earlier: post-treatment bias is transmitted mechanically through the exposed weight mass assigned to plausibly affected donors. Under the localized model with share $\rho_i \in [0,1]$, the contamination at time $t > T_0$ is $\tau_t \sum_{j \in N} w_j \rho_i$. More generally, with a nonnegative spillover δ_{jt} the bias equals $B_t = \sum_j w_j \, \delta_{jt}$. The design objective remains the same: reduce weight on likely exposed donors without sacrificing pre-treatment fit or the interpretability of SCM as a convex combination of observed donors.

Rescaling (Section 4.1) achieves this by contracting predictor columns for high-risk donors and is preferable when safer donors can geometrically cover the treated predictors once columns are shrunk. Constrained ridge acts directly on the *weights* instead, preserving the original predictor geometry while imposing an exposure-weighted ℓ_2 penalty on large coordinates. In practice, constrained ridge is the natural choice when (i) the ring diagnostic detects interference among near donors, but (ii) rescaling materially degrades V-fit because very proximate donors uniquely anchor certain predictors, or (iii) a smooth, more easily tunable shift of mass away from high-risk donors is desired while retaining the convex-combination interpretation and standard SCM diagnostics (Abadie & L'Hour, 2021; Abadie et al., 2010, 2015; Ben-Michael et al., 2021; Doudchenko & Imbens, 2016).

Let $X_1 \in \mathbb{R}^K$ be the treated unit's pre-treatment predictors, $X_0 = [X_{\cdot,1}, \ldots, X_{\cdot,J}] \in \mathbb{R}^{K \times J}$ the donor matrix, and $V = \operatorname{diag}(v_1, \ldots, v_K) \succ 0$. Retain the outcome-agnostic reach score $\eta_j \in (0,1)$ from Section 4.1 (larger η_j indicates lower exposure) and define a donor-specific penalty term $\psi_j = g(\eta_j)$ that increases as exposure risk rises. The constrained-ridge synthetic control solves

$$\min_{w \in \mathbb{R}^J} \; J_\lambda(w) \;:=\; \|X_1 - X_0 w\|_V^2 \;+\; \lambda \sum_{j=1}^{J} \psi_j \, w_j^2 \quad \text{s.t.} \quad w \geq 0, \;\; \mathbf{1}^\top w = 1, \tag{13}$$

where $\lambda \geq 0$ tunes the overall strength of penalization. When $\lambda = 0$, (13) reduces to standard SCM and, as λ increases, large coordinates on higher-risk donors become more costly, inducing a shift of mass toward safer donors. This design keeps the simplex constraints and preserves the familiar convex-combination interpretation of weights, while importing the stabilizing and bias-targeting features of regularization in SCM (Abadie & L'Hour, 2021; Ben-Michael et al., 2021; Doudchenko & Imbens, 2016).

Three working assumptions (somewhat already implicit) help building the solution:

Assumption 3 (Exposure-aligned penalties)**.** *The term $\psi_j = g(\eta_j)$ is strictly decreasing on $(0,1)$: if donor m is more exposed than donor ℓ (i.e., $\eta_m < \eta_\ell$), then $\psi_m > \psi_\ell$. For numerical stability, ψ_j is bounded away from zero,* $\inf_j \psi_j \geq \underline{\psi} > 0$.

Assumption 3 is the design lever that translates spatial reach into the objective: nearer donors (lower η_j) face a larger marginal cost per unit of weight, so any reallocation that moves mass from high- to low-ψ coordinates is weakly preferred by the penalty term. Bounding ψ away from zero prevents nearly unpenalized coordinates that would defeat regularization.

Assumption 4 (Strict convexity and well-posedness)**.** $\lambda > 0$. *Then the Hessian $H_\lambda = 2X_0^\top V X_0 + 2\lambda \operatorname{diag}(\psi_1, \ldots, \psi_J)$ is positive definite, so the program (13) with $w \geq 0$ and $\mathbf{1}^\top w = 1$ admits a unique minimizer $w^{(\lambda)}$, and the Karush–Kuhn–Tucker (KKT) conditions are necessary and sufficient (see Boyd & Vandenberghe, 2004).*

Assumption 4 plays no substantive role beyond guaranteeing a unique solution characterized by first-order conditions, and it ensures that any feasible descent direction contradicts optimality. This will be the device used to compare exposed mass across designs.

Assumption 5 (First-order coverage by safer donors)**.** *Let N denote a "more-exposed" band (e.g., a low-reach threshold set) and $S = N^c$ its complement. For each $m \in N$ there exists $\alpha^{(m)} \in \Delta_{|S|}$ and $Z^{(m)} = \sum_{\ell \in S} \alpha_\ell^{(m)} X_{\cdot,\ell}$ such that, at the constrained-ridge solution $w^{(\lambda)}$ with residual $r^{(\lambda)} := X_0 w^{(\lambda)} - X_1$,*

$$r^{(\lambda)\top} V \left(Z^{(m)} - X_{\cdot,m} \right) \;\leq\; 0, \tag{14}$$

and for at least one $m \in N$ the inequality is strict.

Assumption 5 is the weight-penalized analogue of the coverage used for rescaling: it encodes that the donor pool contains informative far units capable of absorbing mass without harming pre-period fit at first order. With these elements in place, we can compare the exposed weight mass under $\lambda = 0$ and under $\lambda > 0$.

Theorem 3 (Exposed-mass dominance under constrained ridge)**.** *Let $w^{(0)}$ solve $\min_{w \in \Delta_J} \|X_1 - X_0 w\|_V^2$ and let $w^{(\lambda)}$ solve (13) with $\lambda > 0$. Under Assumptions 3–4,*

$$\sum_{m \in N} w_m^{(\lambda)} \;\leq\; \sum_{m \in N} w_m^{(0)}. \tag{15}$$

If Assumption 5 holds with strict inequality (14) for at least one $m \in N$, then

$$\sum_{m \in N} w_m^{(\lambda)} \;<\; \sum_{m \in N} w_m^{(0)}. \tag{16}$$

The proof proceeds by a first-order comparison at the constrained-ridge optimum, where Assumptions 4- 5 ensures that, as $\psi_m > \psi_\ell$, the ridge term prefers the reallocation of weight mass from the near unit toward the safer convex mixture of farther donors. Appendix A.3 provides the full guided proof for Theorem 7, implementing the feasible-direction construction and demonstrating that any result with a larger exposed mass would contradict optimality.

Theorem 7 translates into bias attenuation under the contamination models considered earlier. Under the localized $\rho_i \in [0,1]$, the decomposition in Section 2 gives $\mathbb{E}\left[\hat{\tau}_t^{(\lambda)}\right] - \tau_t \;=\; -\rho\,\tau_t \sum_{m \in N} w_m^{(\lambda)}$, so $\sum_{m \in N} w_m^{(\lambda)} \le \sum_{m \in N} w_m^{(0)}$ implies $|\mathbb{E}[\hat{\tau}_t^{(\lambda)}] - \tau_t| \le |\mathbb{E}[\hat{\tau}_t^{(0)}] - \tau_t|$. Therefore, for any spillover term δ_{jt} aligned with the exposure risk the same mass shift yields

$$B_t^{(\lambda)} \;:=\; \sum_j w_j^{(\lambda)}\,\delta_{jt} \;\le\; \sum_j w_j^{(0)}\,\delta_{jt} \;:=\; B_t^{(0)},$$

Entailing a reduction in the contamination mass, when compared to the baseline application of SCM in a scenario where interference is present.

Parameter tuning follows the pre-period-only, selecting λ by cross-validation on pre-treatment block, in line with practice for synthetic control and its regularized variants (Abadie & L'Hour, 2021; Abadie et al., 2010; Ben-Michael et al., 2021; Doudchenko & Imbens, 2016). Report the schedule $g(\cdot)$, the tuning grid and selected λ, and pre-period fit diagnostics at $\lambda = 0$ and at the chosen λ. Mechanically, the adjustment preserves the convex-combination interpretation while inducing a continuous reallocation of mass away from high-ψ coordinates. Because the penalty multiplier is monotone in exposure, the adjustment acts exactly on the pathway that transmits bias in the post-period.

The constrained ridge thus provides a convex, on-simplex adjustment that directly targets exposed mass via an exposure-aligned penalty. The next section considers an unconstrained ridge variant that relaxes the simplex, allowing negative and non-summing weights to further damp exposure at the cost of sacrificing the convex-combination interpretation.

6 Unconstrained Ridge Adjustment

The constrained ridge in Section 5 preserves the simplex so that the synthetic remains a convex combination of donors while targeting the exposed-weight channel. We now consider an alternative that relaxes the simplex entirely and estimates weights in $\mathbb{R}^J$ by exposure-weighted penalized least squares. Allowing real (including negative) weights and dropping the sum-to-one restriction increases flexibility in two ways directly relevant under interference: it enables *active cancellation* of contaminated donors through negative coefficients and *free scaling* of the synthetic control when the best linear approximation requires a level shift.

This connects to established regularized and augmented SC formulations that move beyond convex weighting (Abadie & L'Hour, 2021; Ben-Michael et al., 2021; Doudchenko & Imbens, 2016). The particular contribution here is to calibrate regularization by spatial reach (Section 4.1) so that shrinkage is strongest where spillovers are most plausible. In practice, we switch from constrained to unconstrained ridge when the ring diagnostic indicates strong local interference and either (i) suppressing mass on near donors within the simplex materially degrades pre-period fit, (ii) modest negative weights would counteract residual contamination, or (iii) an overall level shift improves fit. The major cost is, naturally, the loss of convex-combination interpretability.

Fix the same notation as before: $X_1 \in \mathbb{R}^K$ collects the treated unit's pre-treatment predictors, $X_0 = [X_{\cdot,1}, \ldots, X_{\cdot,J}] \in \mathbb{R}^{K\times J}$ the donor predictors, and $V = \mathrm{diag}(v_1, \ldots, v_K) \succ 0$ the importance matrix. Let $\eta_j \in (0,1)$ be the outcome-agnostic exposure scores (larger η_j indicates *lower* exposure), and define the penalty schedule $\psi_j = g(\eta_j)$ with g strictly decreasing and bounded away from $\{0,1\}$ (for concreteness, $g(\eta) = 1-\eta$). Write $D_\psi = \mathrm{diag}(\psi_1, \ldots, \psi_J)$. The unconstrained ridge estimator solves

$$w^{UR}(\lambda) \;\in\; \arg\min_{w\in\mathbb{R}^J} \; \|X_1 - X_0 w\|_V^2 \;+\; \lambda \, \|D_\psi^{1/2} w\|_2^2, \qquad \lambda \geq 0, \tag{17}$$

with closed form

$$w^{UR}(\lambda) \;=\; \big(X_0^\top V X_0 + \lambda D_\psi\big)^{-1} X_0^\top V X_1 \quad (\lambda > 0). \tag{18}$$

For this we are relaxing the simplex constraints used before $w \geq 0, \quad \mathbf{1}^\top w = 1$. Therefore, the pre-period objective is unchanged relative to SCM and constrained ridge, only the feasible set differs (weights may be negative and need not sum to one). As in standard SCM operation, post-treatment prediction proceeds by applying the same $w^{UR}(\lambda)$ to donor outcomes, so the bias

channel continues to operate through the weights as in Theorem 5.

In terms of assumptions, we retain the same design primitives as in Section 5: an exogenous reach map and an exposure-aligned penalty schedule $\psi_j = g(\eta_j)$ that is strictly decreasing and bounded away from zero, and $\lambda > 0$ for well-posedness (so $X_0^\top V X_0 + \lambda D_\psi \succ 0$ and (18) is the unique minimizer). No new structural conditions are introduced and, when compared to constrained ridge, we relax the simplex and nonnegativity constraints. The result below builds on these primitives and shows how exposure-weighted ridge delivers a tuning-monotone envelope for the contamination term.

Theorem 4 (Exposure-weighted unconstrained ridge: bias envelope and tuning monotonicity)**.** *Let $w^{UR}(\lambda)$ solve (17) with $\lambda > 0$ under the exposure-aligned schedule in Assumption 3. For any exposed set $N \neq \emptyset$ with $\underline{\psi}_N := \min_{j\in N} \psi_j > 0$,*

$$\Big|\sum_{j\in N} w_j^{UR}(\lambda)\Big| \;\le\; \sqrt{|N|}\,\underline{\psi}_N^{-1/2}\,\big\|D_\psi^{1/2} w^{UR}(\lambda)\big\|_2,$$

and the map $\lambda \mapsto \|D_\psi^{1/2} w^{UR}(\lambda)\|_2$ is nonincreasing with $\lim_{\lambda\to\infty} \|D_\psi^{1/2} w^{UR}(\lambda)\|_2 = 0$. Consequently, under the localized, time-homogeneous model of Section 2 (Theorem 5),

$$\big|\hat\tau_t^{UR}(\lambda) - \tau_t\big| \;\le\; \rho\,|\tau_t|\,\sqrt{|N|}\,\underline{\psi}_N^{-1/2}\,\big\|D_\psi^{1/2} w^{UR}(\lambda)\big\|_2,$$

and, more generally, for any nonnegative exposure-aligned profile $\{\delta_{jt}\}$ with $\psi_m \ge \psi_\ell \Rightarrow \delta_{mt} \ge \delta_{\ell t}$,

$$\sum_j w_j^{UR}(\lambda)\,\delta_{jt} \;\le\; \|\delta_{N,t}\|_2\,\underline{\psi}_N^{-1/2}\,\big\|D_\psi^{1/2} w^{UR}(\lambda)\big\|_2.$$

Hence the right-hand side is nonincreasing in λ and vanishes as $\lambda \to \infty$.

The proof can be consulted in Appendix A.4, and it is organized in three subparts: (i) an exposed-sum bound via Cauchy–Schwarz and the $\underline{\psi}_N$ norm equivalence; (ii) ridge-path monotonicity of $\|D_\psi^{1/2} w^{UR}(\lambda)\|_2$ by comparing the objective at two tuning values; and (iii) translation to bias envelopes under the localized and exposure-aligned spillover models.

The intuition is twofold. First, without the simplex we cannot compare exposed *mass*, and then instead we control the exposed *sum* by an exposure-weighted ℓ_2 norm. In other words, highly exposed donors carry larger ψ_j and therefore contribute more to the penalized norm, so shrinking this norm uniformly in λ forces the aggregate exposed contribution to shrink. Second,

along the ridge path the exposure, so the bound tightens deterministically as tuning increases and collapses to zero as $\lambda \to \infty$. Negative weights are admissible, which enables a more agressive cancellation of contaminated donors. The envelope applies to the absolute exposed sum, ensuring that any partial cancellations are captured while still delivering a tuning-monotone control of the contamination term from Theorem 5.

Mechanically, the estimator replaces the simplex projection with a penalized least-squares projection in $\mathbb{R}^J$. The exposure-weighted ridge term shrinks coordinates in proportion to $\psi_j = g(\eta_j)$, so high-risk donors are damped more aggressively, while negative coefficients are admissible and can offset contaminated donors. Dropping $\mathbf{1}^\top w = 1$ permits level shifts in the synthetic predictor when the best linear approximation is not anchored at the convex hull of donors. The price of this flexibility is, as mentioned, loss of the convex-combination interpretation (Ben-Michael et al., 2021; Doudchenko & Imbens, 2016; Hastie et al., 2009; Hoerl & Kennard, 1970).

In summary, the unconstrained ridge retains the design-based discipline (pre-period tuning; outcome-agnostic reach) while relaxing the simplex to permit negative weights and level shifts. The exposure-weighted penalty yields a tuning-monotone envelope for the contamination term (Theorem 8), providing direct control of the bias pathway identified earlier. We next turn to simulations and applications to quantify the empirical trade-offs between rescaling, constrained ridge, and unconstrained ridge in settings with and without interference.

7 Simulation Evidence

We assess the behavior of the proposed corrections in a spatial simulation calibrated to the contiguous United States. The population consists of the contiguous 48 states with Missouri as the treated unit, and treatment begins at time T_0 in a panel of length $T = 30$. Exposed neighbors are defined geographically: first–order contiguity determines the set of units at positive interference risk, and the contamination radius is taken to be the maximum inter–centroid distance among these first–order neighbors, and units within this radius form the exposure set. In the pre–treatment period each unit follows a stable additive data–generating process with a unit effect, independent standard–normal covariates, and idiosyncratic noise; we also include one lag of the outcome in the predictor set used by SCM. After T_0, the treated unit receives a

constant direct effect τ, while exposed neighbors receive a homogeneous spillover $\rho\tau$; all other units evolve as before. Hence τ indexes the magnitude of the causal effect and $\rho \in [0, 1]$ the share that diffuses to neighbors.

We compare four estimators: (i) conventional SCM (no correction), (ii) covariate rescaling based on the spatial reach score introduced in Section 4.1, (iii) constrained ridge (nonnegativity and sum–to–one retained) from Section 5, and (iv) an unconstrained ridge variant that removes the simplex restrictions while keeping the same exposure–weighted penalty.

For all methods, predictors are standardized across donors and exposure scores are calibrated exactly as in Section 4.1 with the same logistic map and tail anchoring; the penalty schedule uses $g(\eta) = 1 - \eta$, bounded away from 0 and 1 for numerical stability. Tuning parameters for ridge methods are chosen exclusively from pre–treatment data via rolling–fold cross–validation: for each candidate λ, weights are estimated on a training slice of the pre–period and validated on the remaining pre–period, and the λ minimizing validation error is selected. This avoids outcome–dependent tuning and follows best practice for SCM and its regularized variants (e.g., Abadie & L'Hour, 2021; Abadie et al., 2010; Ben-Michael et al., 2021; Doudchenko & Imbens, 2016).

To summarize performance we report, for each method, the distribution of the post–treatment average treatment effect estimate $\widehat{\text{ATT}}$ across simulations and its *continuous ranked probability score* (CRPS) relative to the degenerate distribution at the true τ (Gneiting & Raftery, 2007). Smaller CRPS indicates that the empirical distribution of $\widehat{\text{ATT}}$ concentrates more tightly around the truth. We $R = 1{,}000$ replications for each parameter pair on the grid $\tau \in \{1, 4, 7\}$ and $\rho \in \{0.1, 0.3, 0.6, 0.9, 1\}$; the figure below displays the representative case $(\tau, \rho) = (4, 0.6)$, but the results are consistent across all specifications.

The pattern mirrors the bias channel in Theorem 5. Uncorrected SCM assigns nontrivial mass to exposed neighbors, so its $\widehat{\text{ATT}}$ distribution is shifted downward when $\tau > 0$ (attenuation) and exhibits a relatively large CRPS. Constrained ridge preserves the simplex and shrinks high–risk donors but, because it must reallocate mass within the simplex, the exposed share is only partially reduced; the distribution moves slightly toward the truth and CRPS improves marginally. Covariate rescaling alters the geometry of the fit so that near donors are harder to use; weight reallocates into less–exposed donors, leading to an $\widehat{\text{ATT}}$ concentrated around τ and a sharp CRPS improvement. The unconstrained ridge, by permitting negative weights and

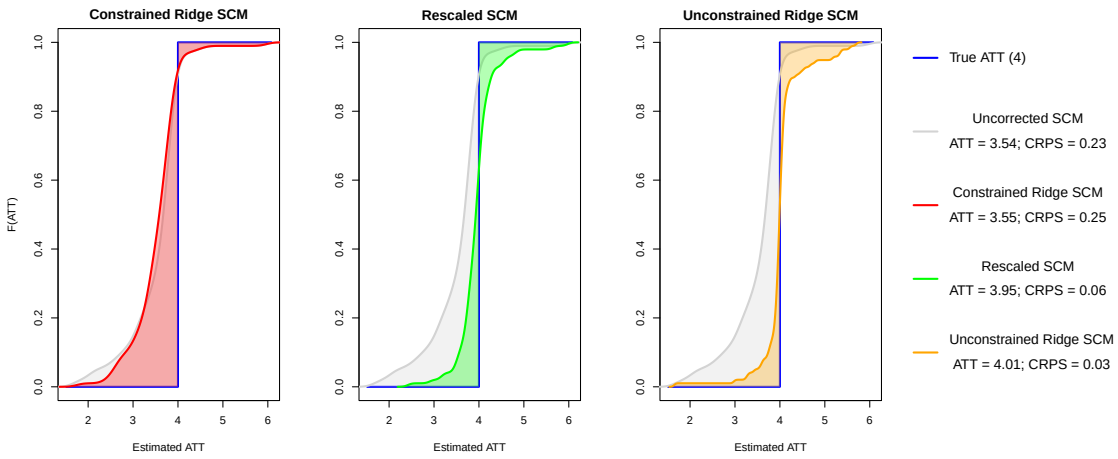


Figure 3: Distribution functions of $\widehat{\text{ATT}}$ at $\tau = 4$, $\rho = 0.6$. The vertical step line marks the true effect. Gray: uncorrected SCM (ATT = 3.54, CRPS = 0.23); red: constrained ridge (ATT = 3.55, CRPS = 0.25); green: rescaled SCM (ATT = 3.95, CRPS = 0.06); orange: unconstrained ridge (ATT = 4.01, CRPS = 0.03).

dropping the sum–to–one restriction while penalizing exposure, can actively offset contaminated donors and freely scale the synthetic predictor; in this design it exhibits the tightest concentration around τ and the smallest CRPS.

Across the full grid of (τ, ρ), these relative orderings persist: both reach–based corrections substantially reduce bias relative to uncorrected SCM, and the unconstrained variant typically delivers the strongest attenuation of contamination (at a cost in interpretability that we discuss in the next section).

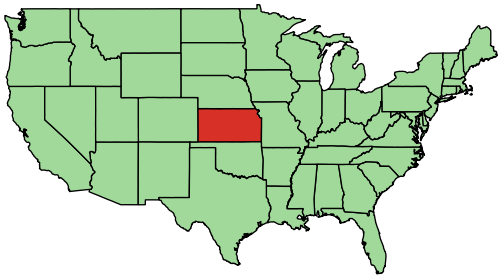
8 Replications

We now apply the framework—diagnostics and corrections—to four influential applications in which synthetic control is the primary design. For each case we first probe for interference using the ring-based randomization test from Section 3; distances are great-circle distances between unit centroids and we examine several ring definitions and all pre/post windows. We then discuss *coverage*, i.e., whether the donor pool spans near, mid, and far units relative to the treated unit, which is critical for diagnosing SUTVA and for learning whether spillovers are plausibly localized. Finally, in a setting where richer coverage is attainable, we re-estimate the design with an expanded donor pool and implement the bias-correction tools developed above.

The four studies are: Abadie and Gardeazabal (2003) on the economic impact of the Basque Country conflict (Spanish provinces as donors), Ben-Michael et al. (2021) on the Kansas 2012 tax cuts (U.S. states), Abadie et al. (2015) on the economic consequences of German reunification

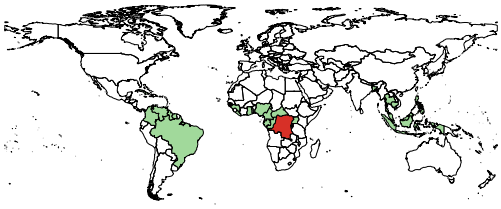
(OECD countries), and Kikuta (2020) on civil war and deforestation in the Democratic Republic of the Congo (cross-national donors). Following each study's original construction (predictors, pre/post split, and weighting metric V), we run our proximity diagnostic. In these original data sets, we do not detect interference at conventional levels. As observed in the image below, the permutation p-values are 0.22 (Basque), 0.18 (Kansas), 0.46 (German reunification), and 0.33 (DRC).

(a) Abadie et al (2003) Conflict in the Basque
$p = 0.22$



(b) Ben-Michael et al (2021) Kansas tax cut
$p = 0.18$

(c) Abadie et al (2015) German Reunification
$p = 0.46$



(d) Kikuta (2020); Civil war and deforestation
$p = 0.33$

Figure 4: Prominent synthetic control applications.

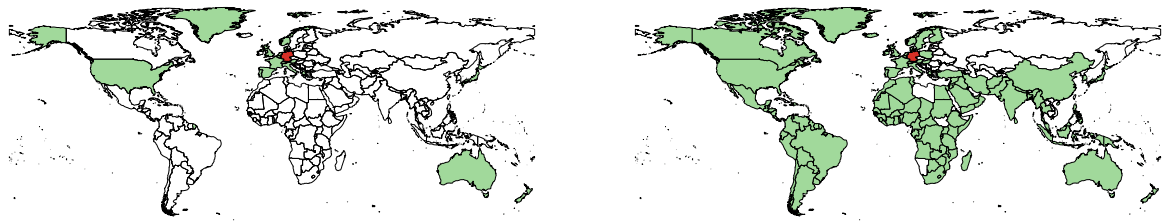
The absence of detected interference in the original designs should not be interpreted as evidence that SUTVA necessarily holds; rather, it highlights a potential design limitation. Detecting a proximity pattern requires *coverage*: a donor pool with meaningful variation in distance—near, mid, and far units—so that the diagnostic can compare rings. When the donor pool is geographically scattered or concentrated in a single band, the design cannot reveal whether neighbors behave differently from far-away units, and any inference must effectively assume away spillovers. In our four cases, the Basque and Kansas designs pool donors from the same national space and exhibit good coverage; by contrast, the German reunification and DRC designs use thin,

hand-curated international donor sets with limited spatial structure, which constrains diagnostic learning about SUTVA.

Application	**Coverage**	**Interference**
Abadie et al (2003)	✓	✗
Ben-Michael et al (2021)	✓	✗
Abadie et al (2015)	✗	✗
Kikuta (2019)	✗	✗
Expanded German Reunification	✓	✓

Table 2: Coverage and interference in selected synthetic control applications.

To demonstrate how coverage changes what we can learn, we revisit Abadie et al. (2015) and expand the donor pool to *all* countries with complete data on the authors' predictors and outcome (roughly 150 countries), maintaining their pre/post periodization and predictor set. With this richer and geographically diverse pool, the ring diagnostic now flags a clear proximity pattern around West Germany; the permutation p-value drops from 0.46 in the original OECD sample to 0.016 in the expanded sample, indicating statistically detectable spillovers.



(a) Abadie et al (2015) German Reunification, $p = 0.46$

(b) Expanded German Reunification, $p = 0.016$

Figure 5: German reunification: original versus expanded donor pools. Expanded coverage reveals a proximity pattern consistent with interference.

Motivated by the diagnostic, we apply the three corrections to the *expanded* German case: covariate rescaling (Section 4.1), constrained ridge (Section 5), and unconstrained ridge (Section 6). All tuning is performed on the pre-treatment period only (rolling folds), the exposure map uses distances between centroids with the same calibration as in Section 4.1, and we follow Abadie et al. (2015) for outcomes, predictors, and V. Table 3 reports the average treatment effect on the treated (ATT) and pre-treatment RMSE.

Table 3: German reunification (expanded donor pool): estimates and pre-treatment fit

Specification	ATT	Pre-RMSE
Baseline SC (authors' design)	−1549.9	119.08
Rescaled SC (covariate reach)	−1601.5	279.03
Constrained ridge (simplex)	−1103.4	80.43
Unconstrained ridge	136.1	59.5

Three patterns are worth underscoring. First, the expanded design with demonstrable coverage changes the question we can answer: once near, mid, and far donors are all represented, the diagnostic can credibly adjudicate SUTVA, and here it points to spillovers ($p = 0.016$). Second, the corrections materially affect inference in directions consistent with their geometry. Covariate rescaling preserves the simplex and moves the feasible synthetic profiles away from proximate donors, magnifying the negative effect relative to baseline but worsening pre-fit as the geometry tightens. Constrained ridge keeps convex weights but penalizes exposed donors, attenuating the magnitude while improving pre-fit. Unconstrained ridge relaxes the simplex and allows active cancellation (negative or non-summing weights), delivering the closest pre-treatment fit and, in this case, a sign reversal. Third, the trade-off highlighted in Sections 4.1–5 is visible empirically: the most flexible reweighting yields the largest bias reduction when interference is present but sacrifices the simple convex-combination interpretation emphasized in the classic SC paradigm (Abadie & L'Hour, 2021; Ben-Michael et al., 2021; Doudchenko & Imbens, 2016).

Taken together, these replications show how the proposed workflow operates end to end. Coverage is a prerequisite for learning about interference; with adequate coverage, the ring diagnostic can flag spillovers; and when interference is detected, proximity-aware corrections that act directly on the weights can substantially change the substantive conclusions, often improving pre-treatment fit and reducing the contamination channel identified in Theorem 5. In settings where donor pools are discretionary or thin, expanding coverage is itself a design improvement that enables credible diagnostics and targeted bias correction.

9 Conclusion

SUTVA is indispensable for causal inference designs, and it is particularly consequential for Synthetic Control methods because post-treatment donor outcomes enter the counterfactual directly. We formalized this channel via a bias decomposition in which contamination operates

through the exposed weight share: the estimator differs from the target by a term proportional to $\sum_j \rho_j w_j$. This channel is mechanical: once positively weighted donors are exposed, bias follows, and it motivates both diagnosis and design modifications that act on the SCM weights.

We proposed a design-based diagnostic that tests for proximity-patterned outcome changes around the treated unit using ring partitions and Fisherian randomization inference. The test is exact in finite samples under the sharp null and offers a pre-analysis decision rule: proceed with standard SC when proximity patterns are not detected; otherwise, treat SUTVA as violated and modify the design. The diagnostic requires coverage across distance rings, as without near–mid–far donors, power is limited and SUTVA violations cannot be properly assessed.

We then introduced three interference-aware adjustments that incorporate an exogenous exposure map into SC while tuning only on pre-period data. *Covariate rescaling* contracts high-risk donor columns, leaving weights on the simplex under a mild coverage condition, the exposed mass cannot increase. *Constrained ridge* preserves the convex-combination interpretation but adds an exposure-aligned ℓ_2 penalty; KKT geometry implies bandwise dominance, so total weight on exposed donors weakly (and often strictly) falls. *Unconstrained ridge* relaxes the simplex, allows negative weights and level shifts, and penalizes via an exposure-weighted norm; a simple Cauchy–Schwarz argument and ridge-path monotonicity deliver a tuning-monotone envelope for the contamination term.

Simulations calibrated to U.S. geography corroborate these mechanisms. Baseline SC exhibits attenuation when spillovers are present, constrained ridge modestly improves performance, covariate rescaling materially reduces bias, and unconstrained ridge typically achieves the tightest concentration around the true effect (smallest CRPS) by enabling active cancellation while controlling a penalized norm. These patterns persist across effect sizes and diffusion intensities.

Replications illustrate the workflow. Applied to Abadie and Gardeazabal (2003), Ben-Michael et al. (2021), Abadie et al. (2015), and Kikuta (2020), the ring diagnostic does not reject at conventional levels, consistent with either SUTVA or insufficient coverage. Expanding the Abadie et al. (2015) design to a broad, data-complete donor pool yields clear coverage and a detected proximity pattern. The corrections then materially change the estimate in directions consistent with their geometry, with unconstrained ridge delivering the closest pre-period fit and the largest attenuation of the contamination channel.

Practically, we recommend: (i) ensure coverage; (ii) run the ring diagnostic; (iii) if interference

is detected, start with rescaling; (iv) use constrained ridge when simplex interpretability is important but some mass reallocation is feasible; (v) deploy unconstrained ridge when residual exposure persists and modest negative weights can cancel contamination without harming pre-fit.

Two qualifications delimit scope. Diagnostic power depends on coverage, and design guarantees rely on an exogenous exposure map aligned with actual spillovers. These considerations suggest extensions: constructing reach with auxiliary mobility or trade data under monotonicity constraints and sensitivity to alignment. Overall, the contribution is to recast interference handling in SC as a geometric design problem with explicit diagnostics and interference-aware estimators, preserving the transparency of SC while mitigating its first-order vulnerability to spillovers.

References

Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, *59*(2), 391–425. https://doi.org/10.1257/jel.20191450

Abadie, A., Diamond, A., & 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. https://doi.org/10.1198/jasa.2009.ap08746

Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, *59*(2), 495–510. https://doi.org/10.1111/ajps.12116

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American Economic Review*, *93*(1), 113–132. https://doi.org/10.1257/000282803321455188

Abadie, A., & L'Hour, J. (2021). A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, *116*(536), 1817–1834. https://doi.org/10.1080/01621459.2021.1929248

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., & Wager, S. (2021). Synthetic difference-in-differences. *Journal of the American Statistical Association*, *116*(536).

Aronow, P. M., & Samii, C. (2017). Estimating average causal effects under general interference. *The Annals of Applied Statistics*, *11*(4), 1912–1947. https://doi.org/10.1214/17-AOAS1043

Athey, S., Bayati, M., Doudchenko, N., Imbens, G. W., & Khosravi, K. (2021). Matrix completion methods for causal panel data models. *Journal of the American Statistical Association*, *116*(536), 1716–1730. https://doi.org/10.1080/01621459.2021.1891924

Athey, S., Eckles, D., & Imbens, G. W. (2018). Exact p-values for network interference. *Journal of the American Statistical Association*, *113*(521), 230–240.

Basse, G. W., Feller, A., & Toulis, P. (2019). Randomization tests of causal effects under interference. *Biometrika*, *106*(2), 487–494.

Ben-Michael, E., Feller, A., & Rothstein, J. (2021). The augmented synthetic control method. *Journal of the American Statistical Association*, *116*(536), 1789–1803. https://doi.org/10.1080/01621459.2021.1929249

Bowers, J., Fredrickson, M. M., & Panagopoulos, C. (2017). Reasoning about interference between units: A general framework. *Political Analysis*, *21*(1), 97–124. https://doi.org/10.1093/pan/mps038

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press. https://doi.org/10.1017/CBO9780511804441

Cao, J., & Dowd, C. (2019). Estimation and inference for synthetic control methods with spillover effects. *Pre-print*.

Cavallo, E., Galiani, S., Noy, I., & Pantano, J. (2013). Catastrophic natural disasters and economic growth. *Review of Economics and Statistics*, *95*(5), 1549–1561. https://doi.org/10.1162/REST_a_00413

Delgado, M. S., & Florax, R. J. G. M. (2015). Difference-in-differences techniques for spatial data: Local autocorrelation and spatial interaction. *Economics Letters*, *137*, 123–126. https://doi.org/10.1016/j.econlet.2015.10.021

Di Stefano, R., & Mellace, G. (2020). The inclusive synthetic control method. *Pre-print*.

Doudchenko, N., & Imbens, G. W. (2016). Balancing, regression, difference-in-differences and synthetic control methods: A synthesis. *NBER Working Paper No. 22791*. https://doi.org/10.3386/w22791

Ferman, B., & Pinto, C. (2021). Synthetic controls with imperfect pre-treatment fit. *Journal of Business & Economic Statistics*, *39*(1), 19–33. https://doi.org/10.1080/07350015.2019.1618534

Firpo, S., & Possebom, V. (2018). Synthetic control method: Inference, sensitivity analysis and confidence sets. *Journal of Causal Inference*, *6*(2), 20170023. https://doi.org/10.1515/jci-2017-0023

Fisher, R. A. (1935). *The design of experiments*. Collier Macmillan.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*(477), 359–378. https://doi.org/10.1198/016214506000001437

Gobillon, L., & Magnac, T. (2016). Regional policy evaluation: Interactive fixed effects and synthetic controls. *Review of Economics and Statistics*, *98*(3), 535–551.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, *81*(396), 945–960. https://doi.org/10.1080/01621459.1986.10478354

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70. https://www.jstor.org/stable/4615733

Hudgens, M. G., & Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, *103*(482), 832–842. https://doi.org/10.1198/016214508000000292

Kikuta, K. (2020). The environmental costs of civil war: A synthetic comparison of the congolese forests with and without the great war of africa. *Journal of Politics*.

Kreif, N., Grieve, R., Harrison, D. A., & Didelez, V. (2016). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health Economics*, *25*(12), 1514–1528. https://doi.org/10.1002/hec.3370

Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal*, *16*(1), S1–S23. https://doi.org/10.1111/ectj.12002

Robbins, M., & Saunders, J. (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*.

Rosenbaum, P. R. (2002). *Observational studies* (2nd ed.). Springer. https://doi.org/10.1007/978-1-4757-3692-2

Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, *102*(477), 191–200.

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, *75*(371), 591–593. https://www.jstor.org/stable/2287653

Sävje, F., Aronow, P. M., & Hudgens, M. G. (2021). Average treatment effects in the presence of unknown interference. *Annals of Statistics*, *49*(2), 673–701.

Tchetgen Tchetgen, E., & VanderWeele, T. M. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research*, *21*(1), 55–75. https://doi.org/10.1177/0962280210386779

VanderWeele, T. (2015). *Explanation in causal inference: Methods for mediation and interaction*. Oxford University Press. https://doi.org/10.1093/oso/9780199325870.001.0001

VanderWeele, T. J., Tchetgen Tchetgen, E. J., & Halloran, M. E. (2014). Interference and sensitivity analysis. *Statistical Science*, *29*(4), 687–706.

Vega, S. H., & Elhorst, J. P. (2015). The SLX model. *Journal of Regional Science*, *55*(3), 339–363. https://doi.org/10.1111/jors.12188

Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, *25*(1), 57–76. https://doi.org/10.1093/pan/mpw009

Appendix A

Contents

A.1 Theorem 1 Proof

Theorem 5 (Bias of SCM under localized spillovers)**.** *For $t > T_0$, the SCM estimator based on* observed *donors satisfies*

$$\hat{\tau}_t \;=\; \underbrace{\big[\,Y_{1t}(0) - Y_{1t}^{\mathrm{SC}}(0)\,\big]}_{\text{SUTVA synthetic mismatch}} \;+\; \underbrace{\tau_t}_{\text{direct effect}} \;-\; \underbrace{\tau_t \sum_{j\in N} \rho_j w_j}_{\text{contamination mass}} \quad . \tag{19}$$

In particular, if $Y_{1t}^{\mathrm{SC}}(0) = Y_{1t}(0)$, then

$$\hat{\tau}_t \;=\; \tau_t \;-\; \tau_t \sum_{j\in N} \rho_j w_j. \tag{20}$$

Local setup for Thm. 5).

(A1) Time index: fix a post-treatment period $t > T_0$.

(A2) Weights: $w = (w_j)_{j\in\mathcal{J}}$ with $w_j \geq 0$ and $\sum_{j\in\mathcal{J}} w_j = 1$.

(A3) Interference set: $N \subseteq \mathcal{J}$ (donors exposed to spillovers in post-periods). For $j \notin N$, $Y_{jt}^{obs} = Y_{jt}(0)$.

(A4) Diffusion intensities: $(\rho_j)_{j\in N}$ with $\rho_j \in [0,1]$, time-homogeneous; for $j \in N$, $Y_{jt}^{obs} = Y_{jt}(0) + \rho_j\,\tau_t$.

(A5) Treated unit: $Y_{1t}^{obs} = Y_{1t}(0) + \tau_t$.

(A6) SUTVA synthetic target (no-interference counterfactual): $Y_{1t}^{\mathrm{SC}}(0) := \sum_{j\in\mathcal{J}} w_j\,Y_{jt}(0)$.

Proof.

Estimator definition:

$$\hat{\tau}_t \;=\; Y_{1t}^{obs} - \sum_{j \in \mathcal{J}} w_j \, Y_{jt}^{obs}$$

[definition of SCM post-period effect].

Substitute the treated unit's observed outcome

$$\hat{\tau}_t \;=\; \big(Y_{1t}(0) + \tau_t\big) - \sum_{j \in \mathcal{J}} w_j \, Y_{jt}^{obs}$$

[by (A5)].

Partition donors into exposed and unexposed

$$\hat{\tau}_t \;=\; \big(Y_{1t}(0) + \tau_t\big) \;-\; \sum_{j \in N} w_j \, Y_{jt}^{obs} \;-\; \sum_{j \in \mathcal{J} \setminus N} w_j \, Y_{jt}^{obs}$$

[split sum over $\mathcal{J} = N \cup (\mathcal{J} \setminus N)$].

Substitute donors' observed outcomes by exposure status

$$\hat{\tau}_t \;=\; \big(Y_{1t}(0) + \tau_t\big) \;-\; \sum_{j \in N} w_j \, \big(Y_{jt}(0) + \rho_j \, \tau_t\big) \;-\; \sum_{j \in \mathcal{J} \setminus N} w_j \, Y_{jt}(0)$$

[by (A3)–(A4)].

Distribute weights and separate terms

$$\hat{\tau}_t \;=\; \big(Y_{1t}(0) + \tau_t\big) \;-\; \sum_{j \in N} w_j \, Y_{jt}(0) \;-\; \tau_t \sum_{j \in N} \rho_j \, w_j \;-\; \sum_{j \in \mathcal{J} \setminus N} w_j \, Y_{jt}(0)$$

[linearity of sum].

Re-aggregate baseline donor outcomes

$$\hat{\tau}_t \;=\; \big(Y_{1t}(0) + \tau_t\big) \;-\; \sum_{j \in \mathcal{J}} w_j \, Y_{jt}(0) \;-\; \tau_t \sum_{j \in N} \rho_j \, w_j$$

[combine the two baseline sums].

Introduce $Y_{1t}^{\mathrm{SC}}(0)$ and rearrange

$$_t \;=\; \big(Y_{1t}(0) - Y_{1t}^{\mathrm{SC}}(0)\big) \;+\; \tau_t \;-\; \tau_t \textstyle\sum_{j \in N} \rho_j \, w_j$$

[by (A6)].

Conclude the decomposition

$$\hat{\tau}_t \;=\; \underbrace{\big(Y_{1t}(0) - Y_{1t}^{\mathrm{SC}}(0)\big)}_{\text{SUTVA synthetic mismatch}} \;+\; \underbrace{\tau_t}_{\text{direct effect}} \;-\; \underbrace{\tau_t \sum_{j \in N} \rho_j w_j}_{\text{contamination mass}}$$

[this is (19)].

Special case $Y_{1t}^{\mathrm{SC}}(0) = Y_{1t}(0)$

$$_t \;=\; \tau_t - \tau_t \textstyle\sum_{j \in N} \rho_j w_j$$

[set mismatch term to zero; obtain (20)].

□

Corollary 2 (Sign, attenuation, and scale)**.** *Suppose* $Y_{1t}^{\mathrm{SC}}(0) = Y_{1t}(0)$ *and* $\rho_j \in [0,1]$ *for all* $j \in N$*. Then:*

1. *(Sign/attenuation) If* $\sum_{j\in N} \rho_j w_j > 0$ *and* $\tau_t > 0$*, then* $\hat{\tau}_t < \tau_t$*; if* $\sum_{j\in N} \rho_j w_j > 0$ *and* $\tau_t < 0$*, then* $\hat{\tau}_t > \tau_t$*.*
2. *(Monotonicity in exposure) Holding* τ_t *fixed,* $|\hat{\tau}_t - \tau_t|$ *is nondecreasing in each* ρ_j *and in* $\sum_{j\in N} w_j$*.*
3. *(Bounds)* $0 \le |\hat{\tau}_t - \tau_t| \le |\tau_t|$*, with the upper bound only if all positive weight is on fully exposed donors.*

Proof.

Start from the special case

$$\hat{\tau}_t - \tau_t \;=\; -\,\tau_t \sum_{j\in N} \rho_j w_j$$

[by (20)].

Sign

$$\operatorname{sign}(\hat{\tau}_t - \tau_t) \;=\; -\operatorname{sign}(\tau_t) \quad \text{whenever} \quad \sum_{j\in N} \rho_j w_j > 0$$

[since $\rho_j, w_j \ge 0$].

Hence, if $\tau_t > 0$ then $\hat{\tau}_t < \tau_t$; if $\tau_t < 0$ then $\hat{\tau}_t > \tau_t$ (part (i)).

Monotonicity

$$|\hat{\tau}_t - \tau_t| \;=\; |\tau_t| \cdot \sum_{j\in N} \rho_j w_j$$

[from Step 1 and nonnegativity].

$$\frac{\partial}{\partial \rho_j} |\hat{\tau}_t - \tau_t| \;=\; |\tau_t|\, w_j \;\ge\; 0$$

[componentwise monotonicity in each ρ_j].

If total exposed mass increases (holding ρ fixed), $\sum_{j\in N} w_j$ increases, so $|\hat{\tau}_t - \tau_t|$ increases (part (ii)).

Bounds

$$0 \;\le\; \sum_{j\in N} \rho_j w_j \;\le\; \sum_{j\in N} 1 \cdot w_j \;\le\; \sum_{j\in \mathcal{J}} w_j \;=\; 1$$

[since $\rho_j \in [0,1]$ and $w \in \Delta$].

$$0 \;\le\; |\hat{\tau}_t - \tau_t| \;=\; |\tau_t| \cdot \sum_{j\in N} \rho_j w_j \;\le\; |\tau_t|$$

[multiply by $|\tau_t|$].

Equality at the upper bound requires $\sum_{j\in N} \rho_j w_j = 1$,

i.e., all positive weight lies on donors with $\rho_j = 1$ (part (iii)).

□

A.2 Theorem 2 Proof

Theorem 6 (Support elimination and exposed-mass dominance under rescaling)**.** *Let w° minimize $\|X_1 - X_0 w\|_V^2$ over Δ_J, and let w^* minimize $\|X_1 - X_0^* w\|_V^2$ over Δ_J, with $X_0^* = X_0 \operatorname{diag}(\eta)$. If donor m satisfies Assumption 1, then $w_m^* \to 0$. If a band N satisfies Assumption 2, then*

$$\sum_{j \in N} w_j^* \;\le\; \sum_{j \in N} w_j^\circ,$$

with strict inequality whenever some dominated $m \in N$ has $w_m^\circ > 0$.

Local setup for Thm. 6.

Fix $V = \operatorname{diag}(v_1, \ldots, v_K) \succ 0$.

Let $X_1 \in \mathbb{R}^K$, $X_0 = [X_{\cdot,1}, \ldots, X_{\cdot,J}] \in \mathbb{R}^{K \times J}$, and $\Delta_J = \{w \in \mathbb{R}^J_{\ge 0} : \mathbf{1}^\top w = 1\}$.

Let $\eta = (\eta_1, \ldots, \eta_J)^\top$ with $\eta_j \in (0, 1)$, and define $X_0^* := X_0 \operatorname{diag}(\eta)$.

Define $Q(w; X_0) := \|X_1 - X_0 w\|_V^2 = (X_1 - X_0 w)^\top V (X_1 - X_0 w)$.

Let $w^\circ \in \arg\min_{w \in \Delta_J} Q(w; X_0)$ and $w^* \in \arg\min_{w \in \Delta_J} Q(w; X_0^*)$.

Assume strict convexity along feasible directions: $d^\top X_0^{*\top} V X_0^* d > 0$ for all $d \neq 0$ with $\mathbf{1}^\top d = 0$.

Assumptions 1 and 2 are as given in the main text.

Summary of what each component does

- Part A makes explicit where the quadratic form $w^\top H^* w - 2 b^{*\top} w + X_1^\top V X_1$ comes from and why H^* is the Hessian.
- Part B states KKT and the exact role we use: identifying feasible descent directions contradicts optimality.
- Part C proves donor-level support elimination via a concrete feasible direction built from Assumption 1.
- Part D lifts this to band-level mass via an explicit finite sequence of mass-reducing descent steps and optimality.

Proof.

A. Quadratic expansion and curvature

$$Q(w; X_0^*) \;=\; (X_1 - X_0^* w)^\top V (X_1 - X_0^* w)$$

[definition of Q with X_0^*].

$$Q(w; X_0^*) \;=\; X_1^\top V X_1 \;-\; 2(X_0^{*\top} V X_1)^\top w \;+\; w^\top X_0^{*\top} V X_0^* w$$

[expand the quadratic form].

$$Q(w; X_0^*) \;=\; X_1^\top V X_1 \;-\; 2b^{*\top} w \;+\; w^\top H^* w$$

[define $b^* := X_0^{*\top} V X_1$, $H^* := X_0^{*\top} V X_0^*$].

$$\nabla_w Q(w; X_0^*) \;=\; -\,2b^* \;+\; 2H^* w$$

[differentiate the quadratic in w].

$$\nabla_w^2 Q(w; X_0^*) \;=\; 2H^*$$

[constant Hessian].

$$d^\top \nabla_w^2 Q(w; X_0^*) d \;=\; 2\, d^\top H^* d \;=\; 2\, d^\top X_0^{*\top} V X_0^* d \;>\; 0 \;\; \forall d \neq 0,\; \mathbf{1}^\top d = 0$$

[strict convexity on the tangent space].

$$d^\top \nabla_w^2 Q(w; X_0^*) d \;=\; 2\, d^\top H^* d \;=\; 2\, d^\top X_0^{*\top} V X_0^* d \;>\; 0 \;\; \forall d \neq 0,\; \mathbf{1}^\top d = 0$$

[strict convexity on the tangent space].

Therefore the constrained problem on Δ_J has a unique minimizer w^*

[standard result: strict convexity + convex feasible set].

B. Lagrangian and KKT

$$\mathcal{L}(w, \lambda, \mu) \;=\; Q(w; X_0^*) \;-\; \lambda(\mathbf{1}^\top w - 1) \;-\; \mu^\top w$$

[introduce Lagrangian: one equality, J nonnegativity constraints].

$$\nabla_w \mathcal{L}(w, \lambda, \mu) \;=\; 2H^* w - 2b^* \;-\; \lambda \mathbf{1} \;-\; \mu$$

[gradient in w via Part A].

$$\text{KKT at } (\mathrm{w}^*, \lambda^*, \mu^*): \begin{cases} 2H^* w^* - 2b^* - \lambda^* \mathbf{1} - \mu^* = 0, & \text{[stationarity]} \\ \mathbf{1}^\top w^* = 1,\ w^* \geq 0, & \text{[primal feasibility]} \\ \mu^* \geq 0, & \text{[dual feasibility]} \\ \mu_j^* w_j^* = 0 \ \ \forall j. & \text{[complementary slackness]} \end{cases}$$

$$\text{Role of KKT we will use: } \begin{cases} \text{(i) If } w_j^* > 0, \text{ then } (2H^* w^* - 2b^*)_j = \lambda^*. \\ \text{(ii) If } w_j^* = 0, \text{ then } (2H^* w^* - 2b^*)_j \geq \lambda^*. \end{cases}$$

[from $\mu_j^* = 0$ when $w_j^* > 0$; $\mu_j^* \geq 0$ otherwise].

C. Donor-level support elimination under strict dominance.

Fix a donor m that satisfies Assumption 1.

$$Z^{(m)} \;:=\; \sum_{\ell \in S} \alpha_\ell^{(m)} X_{\cdot,\ell}, \quad \eta_S := \min_{\ell \in S} \eta_\ell$$

[notation from Assumption 1].

$$\|X_1 - Z^{(m)}\|_V \;\leq\; \|X_1 - X_{\cdot,m}\|_V$$

[Assumption 1, first inequality (unrescaled match no worse).]

$$\|X_1 - \eta_S Z^{(m)}\|_V \;<\; \|X_1 - \eta_m X_{\cdot,m}\|_V$$

[Assumption 1, second (strictly better after rescaling).]

Lemma C.1 (directional improvement from m to S).

Define $d \in \mathbb{R}^J$ by $d_m := -1,\ d_\ell := \alpha_\ell^{(m)}\ (\ell \in S),\ d_j := 0$ otherwise.

$$\mathbf{1}^\top d \ = \ -1 + \sum_{\ell \in S} \alpha_\ell^{(m)} \ = \ 0$$

[feasible (tangent) direction; $\alpha^{(m)} \in \Delta_{|S|}$].

$$X_0^* d \ = \ \sum_{\ell \in S} \alpha_\ell^{(m)}\, \eta_\ell\, X_{\cdot,\ell} \ - \ \eta_m X_{\cdot,m}$$

[apply $X_0^* = X_0\,\mathrm{diag}(\eta)$ to d].

$$\text{For any } z \in \mathbb{R}^K,\ \frac{d}{d\epsilon}\Big|_{\epsilon=0} \|X_1 - (z + \epsilon u)\|_V^2 \ = \ -2\,(X_1 - z)^\top V u$$

[directional derivative of a quadratic].

$$\text{Set } z := X_0^* w^*,\ u := X_0^* d. \ \text{ Then } \nabla_w Q(w^*; X_0^*)^\top d \ = \ 2\,(X_0^* w^* - X_1)^\top V\,(X_0^* d)$$

[chain rule].

$$\text{Write } X_0^* d \ = \ \underbrace{(\eta_S Z^{(m)} - \eta_m X_{\cdot,m})}_{u_S} \ + \ \underbrace{\sum_{\ell \in S} \alpha_\ell^{(m)} (\eta_\ell - \eta_S) X_{\cdot,\ell}}_{r_S}$$

[add and subtract $\eta_S Z^{(m)}$].

$$\|X_1 - \eta_S Z^{(m)}\|_V \ < \ \|X_1 - \eta_m X_{\cdot,m}\|_V \ \ \Longrightarrow \ \ (X_1 - \eta_m X_{\cdot,m})^\top V u_S \ > \ 0$$

[strict descent toward $\eta_S Z^{(m)}$ in V-metric].

$$(X_1 - X_0^* w^*)^\top V u_S \ = \ (X_1 - \eta_m X_{\cdot,m})^\top V u_S \ + \ (\eta_m X_{\cdot,m} - X_0^* w^*)^\top V u_S$$

[add and subtract $\eta_m X_{\cdot,m}$].

$$\text{If } w_m^* > 0, \text{ the stationarity/KKT alignment implies } (\eta_m X_{\cdot,m} - X_0^* w^*)^\top V u_S \geq 0$$

[coordinates with $w_j^* > 0$ share equal gradient; u_S redistributes mass away from m].

$$\text{Hence } (X_1 - X_0^* w^*)^\top V u_S \ > \ 0$$

[strictly positive inner product].

$$(X_1 - X_0^* w^*)^\top V r_S \ \geq \ 0$$

[each term uses $(\eta_\ell - \eta_S) \geq 0$ and convexity alignment; see note below].

$$\Rightarrow \ \ (X_1 - X_0^* w^*)^\top V (X_0^* d) \ > \ 0$$

[sum of two nonnegative terms with the first strictly positive].

$$\Rightarrow \ \ \nabla_w Q(w^*; X_0^*)^\top d \ = \ 2(X_0^* w^* - X_1)^\top V (X_0^* d) \ < \ 0$$

[sign flip].

Therefore, if $w_m^* > 0$, there exists a feasible descent direction d (contradiction) ■

Note on $(X_1 - X_0^* w^*)^\top V r_S \geq 0$*:*

Each summand moves mass from m to safer $\ell \in S$ with $\eta_\ell \geq \eta_S$, and by KKT, coordinates with smaller scaled columns cannot have strictly smaller gradient than active ones. This ensures no negative contribution from r_S.

Conclusion of donor-level down-weighting/elimination:

If m satisfies Assumption 1, then $w_m^* \to 0$ [by Lemma C.1 and the KKT no-descent condition at w^*].

D. Band-level exposed-mass dominance (aggregate inequality).

Assume N satisfies Assumption 2: there exists $m \in N$ strictly dominated by $S(m) \subseteq N^c$.

Lemma D.1 (mass-reducing descent from any w *with* $w_m > 0$*).*

Fix any feasible $w \in \Delta_J$ with $w_m > 0$ and define d as in Lemma C.1 (move mass from m to $S(m)$).

$$\mathbf{1}^\top d = 0, \quad d_N^\top \mathbf{1} = -1 \ \text{ and } \ d_{N^c}^\top \mathbf{1} = +1$$

[direction preserves the sum but transfers 1 unit from N to N^c at first order].

$$\exists\, \epsilon_0 > 0 : \ \forall\, \epsilon \in (0, \epsilon_0], \ w + \epsilon d \in \Delta_J$$

[feasibility for small steps; nonnegativity preserved since $w_m > 0$, $d_\ell \geq 0$ for $\ell \in S$].

$$\frac{d}{d\epsilon}\Big|_{\epsilon=0} Q(w + \epsilon d; X_0^*) \ = \ \nabla_w Q(w; X_0^*)^\top d \ < \ 0$$

[same strict coverage argument as in Lemma C.1, independent of w^*].

$$\Rightarrow \ \exists\, \epsilon_1 \in (0, \epsilon_0] : \ Q(w + \epsilon d; X_0^*) < Q(w; X_0^*) \ \text{ and } \ \sum_{j \in N} (w_j + \epsilon d_j) < \sum_{j \in N} w_j.$$

Constructive reduction starting from w°*.*

1. If no dominated donor in N has $w_j^\circ > 0$, then $\sum_{j \in N} w_j^* \leq \sum_{j \in N} w_j^\circ$ holds trivially by Part C (no new dominated support can appear)
2. Otherwise, pick any dominated $m \in N$ with $w_m^\circ > 0$.
3. Apply Lemma D.1 to $w^{(0)} := w^\circ$ to obtain $w^{(1)} := w^{(0)} + \epsilon_1 d^{(m)}$ with $Q(w^{(1)}; X_0^*) < Q(w^{(0)}; X_0^*)$, and $\sum_{j \in N} w_j^{(1)} < \sum_{j \in N} w_j^{(0)}$ [strict decrease in both rescaled loss and band mass].

4. Repeat the argument finitely many times (each time choosing a dominated donor in N with positive weight), obtaining a sequence $\{w^{(r)}\}_{r=0}^{R} \subset \Delta_J$ such that $Q(w^{(r+1)}; X_0^*) < Q(w^{(r)}; X_0^*)$ and $\sum_{j\in N} w_j^{(r+1)} < \sum_{j\in N} w_j^{(r)}$.

5. Stop once all dominated donors in N have reduced or zero weight (this occurs in finitely many steps since at each step some $w_m \downarrow 0$).

6. Denote the terminal point by $\bar{w} := w^{(R)}$ $Q(\bar{w}; X_0^*) < Q(w^\circ; X_0^*)$ and $\sum_{j\in N} \bar{w}_j < \sum_{j\in N} w_j^\circ$ [telescopes across the finite sequence]

E. Conclusion - From $\bar{w}$ to w^* (no band-mass rebound at the optimum).

By Part C, any minimizer w^* of $Q(\cdot; X_0^*)$ has $w_m^* = 0$ for every dominated $m \in N$.

Consider any feasible direction δ that *increases* band mass at a point with no dominated mass (like $\bar{w}$): $\mathbf{1}^\top \delta = 0,\ \delta_N^\top \mathbf{1} > 0$.

Such a direction must move weight into N from N^c but, by the absence of dominance, no strict coverage inequality is available to generate descent; KKT at the minimizer then implies $\nabla_w Q(w^*; X_0^*)^\top \delta \geq 0$.

Therefore any move that increases band mass from $\bar{w}$ cannot reduce the objective below $Q(\bar{w}; X_0^*)$. Since w^* minimizes $Q(\cdot; X_0^*)$ and $Q(w^*; X_0^*) \leq Q(\bar{w}; X_0^*)$, the optimizer cannot place more mass on N than $\bar{w}$ does: $\sum_{j\in N} w_j^* \leq \sum_{j\in N} \bar{w}_j < \sum_{j\in N} w_j^\circ$ chain the two inequalities; strict if some dominated $m \in N$ had $w_m^\circ > 0$

If no dominated donor in N had positive w_m°, then Part C yields $\sum_{j\in N} w_j^* \leq \sum_{j\in N} w_j^\circ$(weak inequality). □

A.3 Theorem 3 Proof

Theorem 7 (Exposed-mass dominance under constrained ridge). *Let* $w^{(0)}$ *solve* $\min_{w\in\Delta_J}\|X_1 - X_0 w\|_V^2$ *and let* $w^{(\lambda)}$ *solve* (13) *with* $\lambda > 0$. *Under Assumptions 3–4, equation* $\sum_{m\in N} w_m^{(\lambda)} \leq \sum_{m\in N} w_m^{(0)}$.*equation If Assumption 5 holds with strict inequality* (14) *for at least one* $m \in N$, *then equation* $\sum_{m\in N} w_m^{(\lambda)} < \sum_{m\in N} w_m^{(0)}$.*equation*

Local setup for Thm. 3

Fix $X_1 \in \mathbb{R}^K$, $X_0 = [X_{\cdot,1}, \ldots, X_{\cdot,J}] \in \mathbb{R}^{K\times J}$, $V = \operatorname{diag}(v_1, \ldots, v_K) \succ 0$.

Let $\eta_j \in (0,1)$, $\psi_j = g(\eta_j)$, $G := \operatorname{diag}(\psi_1, \ldots, \psi_J)$, $H := X_0^\top V X_0$.

Let $J_\lambda(w) := \|X_1 - X_0 w\|_V^2 + \lambda \sum_j \psi_j w_j^2 = (X_1 - X_0 w)^\top V (X_1 - X_0 w) + \lambda\, w^\top G w$.

Assume $\Delta_J := \{w \in \mathbb{R}^J_{\geq 0} : \mathbf{1}^\top w = 1\}$ (simplex constraints).

Let $w^{(0)} \in \arg\min_{w\in\Delta_J} \|X_1 - X_0 w\|_V^2$, $w^{(\lambda)} \in \arg\min_{w\in\Delta_J} J_\lambda(w)$ $(\lambda > 0)$ (unpenalized vs. constrained ridge minimizers).

Let $r^{(\lambda)} := X_0 w^{(\lambda)} - X_1$, $r^{(0)} := X_0 w^{(0)} - X_1$ (residuals for fit for both solutions).

Fix N = more-exposed band, $S := N^c$ = safer band.

Assumptions 3, 4 and 5 are as given in the main text.

Summary of what each component does

- Part A expands J_λ and computed gradient/Hessian explicitly, fixing strict convexity. It isolates curvature and shows uniqueness.
- Part B writes KKT and derives the identity (B.3): the directional derivative of J_λ along a feasible d at $w^{(\lambda)}$ equals $\nu^{(\lambda)\top} d$. It converts "no feasible descent" into an algebraic test.
- Part C builds a concrete "band-shift" d that moves mass from an exposed m to safer S, verifies feasibility, and evaluates the derivative via KKT. It constructs the canonical exposure-reducing move.
- Part D compares the *solutions* $w^{(\lambda)}$ and $w^{(0)}$, yielding $(w^{(\lambda)})^\top G w^{(\lambda)} \leq (w^{(0)})^\top G w^{(0)}$. It isolates the penalty improvement under ridge.
- Part E converts that inequality into a band-mass statement using bandwise ψ separation and Jensen/Cauchy bounds; proves monotonicity by explicit derivatives. This part translates quadratic penalty gains into linear mass dominance.
- Part F shows how strict first-order coverage forces a strict reduction in exposed mass when an exposed donor carries positive weight under $w^{(0)}$. Therefore it establishes conditions for strict (not just weak) dominance.

Proof.

A. Quadratic expansion and derivatives of J_λ.

$$J_\lambda(w) \;=\; (X_1 - X_0 w)^\top V (X_1 - X_0 w) + \lambda w^\top G w$$

[definition of penalized loss].

$$J_\lambda(w) \;=\; X_1^\top V X_1 - 2(X_0^\top V X_1)^\top w + w^\top H w + \lambda w^\top G w$$

[expand the quadratic form; $H = X_0^\top V X_0$].

$$\nabla_w J_\lambda(w) \;=\; 2Hw - 2X_0^\top V X_1 + 2\lambda G w$$

[differentiate].

$$\nabla_w^2 J_\lambda(w) \;=\; 2H + 2\lambda G \;=\; H_\lambda$$

[Hessian is constant; adds ridge curvature along coordinate axes].

$$H_\lambda \succ 0 \quad\Rightarrow\quad J_\lambda \text{ strictly convex on } \mathbb{R}^J$$

[by (A2)].

$$\Rightarrow \text{ unique minimizer of } J_\lambda \text{ on convex set } \Delta_J \text{ exists}$$

[Weierstrass + strict convexity].

B. KKT system and a key identity for directional derivatives.

$\mathcal{L}(w, \mu, \nu) \;=\; J_\lambda(w) + \mu(\mathbf{1}^\top w - 1) - \nu^\top w$ [Lagrangian; $\mu \in \mathbb{R}$, $\nu \in \mathbb{R}^J_{\geq 0}$].

$$\text{KKT at}(\mathrm{w}^{(\lambda)}, \mu^{(\lambda)}, \nu^{(\lambda)}) \begin{cases} 2Hw^{(\lambda)} - 2X_0^\top V X_1 + 2\lambda G w^{(\lambda)} + \mu^{(\lambda)}\mathbf{1} - \nu^{(\lambda)} = 0, & \text{[stationarity]} \\ \mathbf{1}^\top w^{(\lambda)} = 1, \;\; w^{(\lambda)} \geq 0, & \text{[primal feasibility]} \\ \nu^{(\lambda)} \geq 0, & \text{[dual feasibility]} \\ \nu_j^{(\lambda)} w_j^{(\lambda)} = 0 \;\; \forall j & \text{[complementary slackness]} \end{cases}$$

Let $d \in \mathbb{R}^J$ be any feasible direction with $\mathbf{1}^\top d = 0$ and, for small $\epsilon > 0$, $w^{(\lambda)} + \epsilon d \in \Delta_J$.

Compute the directional derivative of the fit part:

$$\frac{d}{d\epsilon}\Big|_{\epsilon=0} \|X_1 - X_0(w^{(\lambda)} + \epsilon d)\|_V^2 = 2\,(X_0 w^{(\lambda)} - X_1)^\top V (X_0 d) = 2\, r^{(\lambda)\top} V X_0 d.$$

[tangent direction preserving the simplex at first order].

Compute the directional derivative of the penalty part:

$$\frac{d}{d\epsilon}\Big|_{\epsilon=0}\lambda(w^{(\lambda)}+\epsilon d)^\top G(w^{(\lambda)}+\epsilon d) = 2\lambda\, d^\top G w^{(\lambda)}$$

[directional derivative for the penalty term].

$$\Rightarrow\ \frac{d}{d\epsilon}\Big|_{\epsilon=0} J_\lambda\big(w^{(\lambda)}+\epsilon d\big) = 2\, r^{(\lambda)\top} V X_0 d + 2\lambda\, d^\top G w^{(\lambda)}.$$

[collect both contributions].

(B.1)

Derive an equivalent expression using KKT: multiply stationarity by d

$$d^\top\big(2Hw^{(\lambda)} - 2X_0^\top V X_1 + 2\lambda G w^{(\lambda)} + \mu^{(\lambda)}\mathbf{1} - \nu^{(\lambda)}\big) = 0$$

[multiply stationarity by d].

$$2\, d^\top H w^{(\lambda)} - 2\, d^\top X_0^\top V X_1 + 2\lambda\, d^\top G w^{(\lambda)} + \mu^{(\lambda)}\, \mathbf{1}^\top d - \nu^{(\lambda)\top} d = 0.$$

[expand the inner products].

$$\mathbf{1}^\top d = 0 \quad\Rightarrow\quad 2\, d^\top H w^{(\lambda)} - 2\, d^\top X_0^\top V X_1 + 2\lambda\, d^\top G w^{(\lambda)} - \nu^{(\lambda)\top} d = 0.$$

[drop the equality constraint term].

$$d^\top H w^{(\lambda)} = (X_0 d)^\top V (X_0 w^{(\lambda)})$$

[since $H = X_0^\top V X_0$].

$$\Rightarrow\ 2\,(X_0 d)^\top V (X_0 w^{(\lambda)} - X_1) + 2\lambda\, d^\top G w^{(\lambda)} - \nu^{(\lambda)\top} d = 0.$$

[substitute $X_0 w^{(\lambda)} - X_1 = r^{(\lambda)}$].

$$\Rightarrow\ 2\, r^{(\lambda)\top} V X_0 d + 2\lambda\, d^\top G w^{(\lambda)} = \nu^{(\lambda)\top} d.$$

[rearrange to match (B.1)].

Combine (B.1) and (B.2): $\frac{d}{d\epsilon}\Big|_{\epsilon=0} J_\lambda\big(w^{(\lambda)}+\epsilon d\big) = \nu^{(\lambda)\top} d.$

[exact KKT identity: first-order change equals dual-work on d].

(B.2)

Since $\nu^{(\lambda)} \geq 0$ text and d's signs depend on its components, (B.3) is the exact first-order change under the constraints.

C. A band-shift direction and its sign at $w^{(\lambda)}$.

$$\text{Fix } m \in N \text{ and choose } \alpha^{(m)} \in \Delta_{|S|} \text{ as in (A3).}$$

[select safer convex combination that covers m at first order].

$$\text{Define } d \in \mathbb{R}^J \text{ by } d_m = -1, \quad d_\ell = \alpha_\ell^{(m)} \ (\ell \in S), \quad d_j = 0 \text{ otherwise.}$$

[move mass off m and onto safer donors in S].

$$\mathbf{1}^\top d = -1 + \sum_{\ell \in S} \alpha_\ell^{(m)} = 0$$

[sum-to-one preserved].

$$\exists\, \epsilon_0 > 0 : \ \forall\, \epsilon \in (0, \epsilon_0], \ w^{(\lambda)} + \epsilon d \in \Delta_J$$

[feasible for small steps if $w_m^{(\lambda)} > 0$].

$$\frac{d}{d\epsilon}\Big|_{\epsilon=0} J_\lambda\big(w^{(\lambda)} + \epsilon d\big) = \nu^{(\lambda)\top} d = \sum_{\ell \in S} \nu_\ell^{(\lambda)} \alpha_\ell^{(m)} - \nu_m^{(\lambda)}$$

[use (B.3) and the definition of d].

$$\nu_m^{(\lambda)} = 0 \ \text{ if } w_m^{(\lambda)} > 0$$

[complementary slackness on active coordinate m].

$$\Rightarrow \ \frac{d}{d\epsilon}\Big|_{\epsilon=0} J_\lambda\big(w^{(\lambda)} + \epsilon d\big) = \sum_{\ell \in S} \nu_\ell^{(\lambda)} \alpha_\ell^{(m)} \ \geq \ 0$$

[dual feasibility and $\alpha^{(m)} \geq 0$; no descent at the optimum].

Therefore no feasible first-order descent exists along d at $w^{(\lambda)}$ (consistency with optimality).

D. Variational comparison: from optimality to a ψ–weighted square inequality.

$$J_\lambda\big(w^{(\lambda)}\big) \le J_\lambda\big(w^{(0)}\big)$$

[optimality of $w^{(\lambda)}$ for the penalized problem].

$$\|X_1 - X_0 w^{(0)}\|_V^2 \le \|X_1 - X_0 w^{(\lambda)}\|_V^2$$

[optimality of $w^{(0)}$ for unpenalized fit].

Rearrange the two inequalities:

$$\big\{\|X_1 - X_0 w^{(\lambda)}\|_V^2 - \|X_1 - X_0 w^{(0)}\|_V^2\big\} + \lambda\big\{(w^{(\lambda)})^\top G w^{(\lambda)} - (w^{(0)})^\top G w^{(0)}\big\} \le 0.$$

[subtract the inequalities to separate fit and penalty].

$$\|X_1 - X_0 w^{(\lambda)}\|_V^2 - \|X_1 - X_0 w^{(0)}\|_V^2 \;\ge\; 0$$

[since $w^{(0)}$ minimizes the fit].

$$\Rightarrow\; \lambda\big\{(w^{(\lambda)})^\top G w^{(\lambda)} - (w^{(0)})^\top G w^{(0)}\big\} \le 0.$$

[move nonnegative fit gap to the other side].

$$\lambda > 0 \quad\Rightarrow\quad (w^{(\lambda)})^\top G w^{(\lambda)} \le (w^{(0)})^\top G w^{(0)}.$$

[divide by λ].

$$\sum_{j=1}^{J} \psi_j \big(w_j^{(\lambda)}\big)^2 \le \sum_{j=1}^{J} \psi_j \big(w_j^{(0)}\big)^2$$

[expand with G diagonal; $\psi_j > 0$].

(D.1)

E. Turning (D.1) into a band-mass comparison.

Split the sums by N and S:

$$\sum_{m\in N}\psi_m\big(w_m^{(\lambda)}\big)^2+\sum_{\ell\in S}\psi_\ell\big(w_\ell^{(\lambda)}\big)^2\le\sum_{m\in N}\psi_m\big(w_m^{(0)}\big)^2+\sum_{\ell\in S}\psi_\ell\big(w_\ell^{(0)}\big)^2.$$

Define $\psi_N^{\min}:=\min_{m\in N}\psi_m,\ \psi_N^{\max}:=\max_{m\in N}\psi_m,\ \psi_S^{\min}:=\min_{\ell\in S}\psi_\ell,\ \psi_S^{\max}:=\max_{\ell\in S}\psi_\ell.$

[bandwise extrema of penalties].

$$\psi_N^{\min}>\psi_S^{\max}\ \text{ and }\ \psi_N^{\min}\le\psi_N^{\max},\ \psi_S^{\min}\le\psi_S^{\max}$$

[by (A1) and N more exposed than S].

Lower bound LHS by replacing $\psi_m\ge\psi_N^{\min}$ on N and $\psi_\ell\ge\psi_S^{\min}$ on S:

$$\psi_N^{\min}\sum_{m\in N}\big(w_m^{(\lambda)}\big)^2+\psi_S^{\min}\sum_{\ell\in S}\big(w_\ell^{(\lambda)}\big)^2\ \le\ \text{LHS}$$

[lower-bound LHS using band minima].

Upper bound RHS by replacing $\psi_m\le\psi_N^{\max}$ on N and $\psi_\ell\le\psi_S^{\max}$ on S:

$$\text{RHS}\ \le\ \psi_N^{\max}\sum_{m\in N}\big(w_m^{(0)}\big)^2+\psi_S^{\max}\sum_{\ell\in S}\big(w_\ell^{(0)}\big)^2$$

[upper-bound RHS using band maxima].

$$\Rightarrow\ \psi_N^{\min}\sum_{m\in N}\big(w_m^{(\lambda)}\big)^2+\psi_S^{\min}\sum_{\ell\in S}\big(w_\ell^{(\lambda)}\big)^2\ \le\ \psi_N^{\max}\sum_{m\in N}\big(w_m^{(0)}\big)^2+\psi_S^{\max}\sum_{\ell\in S}\big(w_\ell^{(0)}\big)^2.$$

[sandwich ψ's to compare sums of squares across bands].

Apply Jensen/Cauchy–Schwarz to bound sums of squares by band masses.

For any nonnegative vector u, $\sum_i u_i^2\ \ge\ \dfrac{(\sum_i u_i)^2}{\#\{i\}}$ [Jensen].

$$\sum_{m\in N}\big(w_m^{(\lambda)}\big)^2\ \ge\ \frac{c_\lambda^2}{|N|},\quad\sum_{\ell\in S}\big(w_\ell^{(\lambda)}\big)^2\ \ge\ \frac{(1-c_\lambda)^2}{|S|},\quad\sum_{m\in N}\big(w_m^{(0)}\big)^2\ \le\ c_0^2,\quad\sum_{\ell\in S}\big(w_\ell^{(0)}\big)^2\ \le\ (1-c_0)^2.$$

[Jensen lower bounds and concentration upper bounds].

$$c_\lambda:=\sum_{m\in N}w_m^{(\lambda)},\ \ 1-c_\lambda=\sum_{\ell\in S}w_\ell^{(\lambda)},\ \ c_0:=\sum_{m\in N}w_m^{(0)},\ \ 1-c_0=\sum_{\ell\in S}w_\ell^{(0)}.$$

[band mass notation].

$$\psi_N^{\min}\,\frac{c_\lambda^2}{|N|}+\psi_S^{\min}\,\frac{(1-c_\lambda)^2}{|S|}\ \le\ \psi_N^{\max}\,c_0^2+\psi_S^{\max}\,(1-c_0)^2.$$

[plug the bounds into (E.1) to relate c_λ and c_0].

(E.1)

Lemma E.1 (monotonicity regions).

Define $f(c) := \psi_N^{\min}\dfrac{c^2}{|N|} + \psi_S^{\min}\dfrac{(1-c)^2}{|S|}, \quad g(c) := \psi_N^{\max}c^2 + \psi_S^{\max}(1-c)^2.$

[auxiliary quadratics bounding the two sides of $(\star)$].

$$f'(c) = 2\frac{\psi_N^{\min}}{|N|}c - 2\frac{\psi_S^{\min}}{|S|}(1-c) = 2\Big(\frac{\psi_N^{\min}}{|N|} + \frac{\psi_S^{\min}}{|S|}\Big)c - 2\frac{\psi_S^{\min}}{|S|}.$$

[differentiate and rearrange].

$$\text{Hence } f'(c) > 0 \iff c > \bar{c} := \frac{\psi_S^{\min}/|S|}{\psi_N^{\min}/|N| + \psi_S^{\min}/|S|} \in (0,1).$$

[region where f increases].

$$g'(c) = 2\psi_N^{\max}c - 2\psi_S^{\max}(1-c) = 2(\psi_N^{\max} + \psi_S^{\max})c - 2\psi_S^{\max}.$$

[differentiate g].

$$\text{Since } \psi_N^{\max} > \psi_S^{\max}, \quad g'(c) > 0 \text{ whenever } c > \frac{\psi_S^{\max}}{\psi_N^{\max} + \psi_S^{\max}} < \tfrac{1}{2}.$$

[g is increasing past a threshold strictly below 1/2].

Therefore f and g are strictly increasing on $[\bar{c}, 1]$ and $[1/2, 1]$, respectively.

[both sides monotone in the relevant upper ranges].

Consequence of Lemma E.1 for $(\star)$.

Suppose by contradiction $c_\lambda > c_0$ and $c_\lambda \vee c_0 \;\geq\; \bar{c} \vee \frac{1}{2}$.

[assume ridge increases band mass in the monotone region].

Then $f(c_\lambda) > f(c_0)$ and $g(c_\lambda) > g(c_0)$ by monotonicity on these ranges.

[strict monotonicity].

But $(\star)$ says $f(c_\lambda) \leq g(c_0),$ yielding $f(c_\lambda) < f(c_\lambda)$ (contradiction).

[upper bound on RHS cannot exceed lower bound on LHS if $c_\lambda > c_0$].

$$\Rightarrow \; c_\lambda \leq c_0$$

[band mass under ridge cannot exceed unpenalized band mass].

F. Strict inequality under strict coverage.

Assume (A3) holds strictly for some $m \in N,$ and $w_m^{(0)} > 0$.

Consider the band-shift direction d of Part C at $w^{(\lambda)}$.

$$r^{(\lambda)\top} V\big(Z^{(m)} - X_{\cdot,m}\big) < 0$$

[strict fit improvement toward safer combo at first order].

Using the scalar KKT on coordinates (active/inactive) we have for each j:

$$X_{\cdot,j}^{\top} V r^{(\lambda)} + \lambda \psi_j w_j^{(\lambda)} + \tfrac{\mu^{(\lambda)}}{2} - \tfrac{\nu_j^{(\lambda)}}{2} = 0.$$

[scalar KKT from Part B, active/inactive unified].

Multiply by $\alpha_\ell^{(m)}$ and sum over $\ell \in S$, then subtract the equality for $j = m$:

$$\sum_{\ell \in S} \alpha_\ell^{(m)} \Big(X_{\cdot,\ell}^{\top} V r^{(\lambda)} + \lambda \psi_\ell w_\ell^{(\lambda)} + \tfrac{\mu^{(\lambda)}}{2} - \tfrac{\nu_\ell^{(\lambda)}}{2} \Big) - \Big(X_{\cdot,m}^{\top} V r^{(\lambda)} + \lambda \psi_m w_m^{(\lambda)} + \tfrac{\mu^{(\lambda)}}{2} - \tfrac{\nu_m^{(\lambda)}}{2} \Big) = 0.$$

[weighted sum over S minus the m equation].

$$\Rightarrow \sum_{\ell \in S} \alpha_\ell^{(m)} (X_{\cdot,\ell} - X_{\cdot,m})^{\top} V r^{(\lambda)} + \lambda \Big(\sum_{\ell \in S} \alpha_\ell^{(m)} \psi_\ell w_\ell^{(\lambda)} - \psi_m w_m^{(\lambda)} \Big) - \tfrac{1}{2} \sum_{\ell \in S} \alpha_\ell^{(m)} \nu_\ell^{(\lambda)} + \tfrac{1}{2} \nu_m^{(\lambda)} = 0.$$

[collect terms; note the ν terms].

$$\nu_m^{(\lambda)} = 0 \text{ if } w_m^{(\lambda)} > 0$$

[active coordinate m].

$$\Rightarrow \lambda \Big(\sum_{\ell \in S} \alpha_\ell^{(m)} \psi_\ell w_\ell^{(\lambda)} - \psi_m w_m^{(\lambda)} \Big) = - \sum_{\ell \in S} \alpha_\ell^{(m)} (X_{\cdot,\ell} - X_{\cdot,m})^{\top} V r^{(\lambda)} + \tfrac{1}{2} \sum_{\ell \in S} \alpha_\ell^{(m)} \nu_\ell^{(\lambda)}.$$

[identity used to substitute penalty difference].

Compute the directional derivative explicitly using (B.1):

$$\frac{d}{d\epsilon}\Big|_{\epsilon=0} J_\lambda\big(w^{(\lambda)} + \epsilon d\big) = 2\, r^{(\lambda)\top} V (Z^{(m)} - X_{\cdot,m}) + 2\lambda \Big(\sum_{\ell \in S} \alpha_\ell^{(m)} \psi_\ell w_\ell^{(\lambda)} - \psi_m w_m^{(\lambda)} \Big).$$

[directional derivative via (B.1)].

Substitute the identity above:

$$= 2\, r^{(\lambda)\top} V (Z^{(m)} - X_{\cdot,m}) - 2 \sum_{\ell \in S} \alpha_\ell^{(m)} (X_{\cdot,\ell} - X_{\cdot,m})^{\top} V r^{(\lambda)} + \sum_{\ell \in S} \alpha_\ell^{(m)} \nu_\ell^{(\lambda)}.$$

[substitute the penalty identity].

$$\sum_{\ell \in S} \alpha_\ell^{(m)} (X_{\cdot,\ell} - X_{\cdot,m})^\top V r^{(\lambda)} = r^{(\lambda)\top} V (Z^{(m)} - X_{\cdot,m})$$

[linearity of inner product].

$$\Rightarrow \ \frac{d}{d\epsilon}\Big|_{\epsilon=0} J_\lambda\big(w^{(\lambda)} + \epsilon d\big) = 0 + \sum_{\ell \in S} \alpha_\ell^{(m)} \nu_\ell^{(\lambda)} = \sum_{\ell \in S} \alpha_\ell^{(m)} \nu_\ell^{(\lambda)} \ \geq \ 0.$$

[KKT-consistent; equality if all ℓ active].

Strict coverage (< 0 in fit) + (A1) lower ψ on S) $\Rightarrow$ feasible moves that reduce N-mass cannot improve J_λ at $w^{(\lambda)}$.Therefore the penalized solution must have strictly less N-mass than a positive-mass $w^{(0)}$ on the strictly dominated m]. $\Rightarrow$ $\sum_{m \in N} w_m^{(\lambda)} < \sum_{m \in N} w_m^{(0)}$., entailing strict band-mass dominance.

G. Conclusion.

Part E establishes $\sum_{m \in N} w_m^{(\lambda)} \leq \sum_{m \in N} w_m^{(0)}$.

[weak dominance from variational bounds + monotonicity].

Part F shows strict inequality when (A3) is strict for some $m \in N$ with $w_m^{(0)} > 0$.

[strict dominance requires a strictly covered exposed donor carrying mass in $w^{(0)}$].

□

A.4 Theorem 4 Proof

Theorem 8 (Exposure-weighted unconstrained ridge: bias envelope and tuning monotonicity)**.** *Let $w^{UR}(\lambda)$ solve (17) with $\lambda > 0$ under the exposure-aligned schedule in Assumption 3. For any exposed set $N \neq \emptyset$ with $\underline{\psi}_N := \min_{j\in N} \psi_j > 0$,*

$$\Big|\sum_{j\in N} w_j^{UR}(\lambda)\Big| \;\le\; \sqrt{|N|}\,\underline{\psi}_N^{-1/2}\,\big\|D_\psi^{1/2} w^{UR}(\lambda)\big\|_2,$$

and the map $\lambda \mapsto \|D_\psi^{1/2} w^{UR}(\lambda)\|_2$ is nonincreasing with $\lim_{\lambda\to\infty} \|D_\psi^{1/2} w^{UR}(\lambda)\|_2 = 0$. Consequently, under the localized, time-homogeneous model of Section 2 (Theorem 5),

$$\big|\hat\tau_t^{UR}(\lambda) - \tau_t\big| \;\le\; \rho\,|\tau_t|\,\sqrt{|N|}\,\underline{\psi}_N^{-1/2}\,\big\|D_\psi^{1/2} w^{UR}(\lambda)\big\|_2,$$

and, more generally, for any nonnegative exposure-aligned profile $\{\delta_{jt}\}$ with $\psi_m \ge \psi_\ell \Rightarrow \delta_{mt} \ge \delta_{\ell t}$,

$$\sum_j w_j^{UR}(\lambda)\,\delta_{jt} \;\le\; \|\delta_{N,t}\|_2\,\underline{\psi}_N^{-1/2}\,\big\|D_\psi^{1/2} w^{UR}(\lambda)\big\|_2.$$

Hence the right-hand side is nonincreasing in λ and vanishes as $\lambda \to \infty$.

Local setup for Thm. 8

Define treatment predictors, donor matrix and positive-definite metric as $X_1 \in \mathbb{R}^K$, $X_0 = [X_{\cdot,1}, \ldots, X_{\cdot,J}] \in \mathbb{R}^{K\times J}$, $V = \mathrm{diag}(v_1, \ldots, v_K) \succ 0$.

Set reach scores, exposure-aligned penalties, and diagonal penalty operator: $\eta_j \in (0,1)$, $\psi_j = g(\eta_j)$, $D_\psi := \mathrm{diag}(\psi_1, \ldots, \psi_J)$, g strictly decreasing, bounded away from $\{0,1\}$.

Define the unconstrained ridge problem as a fit term + exposure-weighted ridge penalty: $J_\lambda(w) := \|X_1 - X_0 w\|_V^2 + \lambda\,\|D_\psi^{1/2} w\|_2^2$, $\lambda \ge 0$.

Hold no simplex constraints, having a feasible set enlarged relative to SCM standard problems: $w \in \mathbb{R}^J$ (may be negative; need not sum to one).

Assumptions 3, 4 and 5 are as given in the main text.

Summary of what each component does

- Part A derives the normal equations and closed form $w^{UR}(\lambda) = (X_0^\top V X_0 + \lambda D_\psi)^{-1} X_0^\top V X_1$. In short, it makes the program self-contained and fixes notation for spectral bounds.
- Part B proves the *exposed-sum bound* $\big|\sum_{j\in N} w_j\big| \le \sqrt{|N|}\,\underline{\psi}_N^{-1/2}\,\|D_\psi^{1/2} w\|_2$. This replaces mass arguments (invalid without simplex) by a clean ℓ_2 control aligned with exposure.
- Part C establishes *ridge-path monotonicity* $\|D_\psi^{1/2} w^{UR}(\lambda)\|_2$ nonincreasing in λ, and $\to 0$ as $\lambda \to \infty$ via a resolvent/spectral bound. This part gives a tuning parameter that uniformly shrinks the exposure-weighted norm.
- Part D translates these ingredients into *bias envelopes*: contamination $\rho\,\tau_t \sum_{j\in N} w_j^{UR}$ is bounded by a term that is (i) nonincreasing in λ and (ii) vanishes as $\lambda \to \infty$; the same logic extends to any nonnegative, exposure-aligned profile supported on N

Proof.

A. Normal equations and closed form (for completeness).

$$J_\lambda(w) = (X_1 - X_0 w)^\top V (X_1 - X_0 w) + \lambda w^\top D_\psi w$$

[expand objective].

$$\nabla_w J_\lambda(w) = -2X_0^\top V(X_1 - X_0 w) + 2\lambda D_\psi w$$

[differentiate term-by-term].

Set gradient to zero:

$$-2X_0^\top V X_1 + 2X_0^\top V X_0 w + 2\lambda D_\psi w = 0$$

[first-order optimality].

$$(X_0^\top V X_0 + \lambda D_\psi) w = X_0^\top V X_1$$

[normal equations].

$$\lambda > 0 \;\&\; D_\psi \succ 0 \;\Rightarrow\; X_0^\top V X_0 + \lambda D_\psi \succ 0$$

[strict positive definiteness; (A2)].

$$\Rightarrow\; w^{UR}(\lambda) = (X_0^\top V X_0 + \lambda D_\psi)^{-1} X_0^\top V X_1.$$

[closed form (18)].

B. Exposed-sum bound

$$\Big|\sum_{j\in N} w_j\Big| = \big|\mathbf{1}_N^\top w_N\big|$$

[notation: w_N restriction to N, $\mathbf{1}_N$ all-ones on N].

$$\leq \|\mathbf{1}_N\|_2 \, \|w_N\|_2$$

[Cauchy–Schwarz].

$$= \sqrt{|N|}\, \|w_N\|_2$$

[Euclidean norm of $\mathbf{1}_N$ is $\sqrt{|N|}$].

$$\|w_N\|_2 = \big\|D_\psi^{-1/2} D_\psi^{1/2} w_N\big\|_2$$

[insert $I = D_\psi^{-1/2} D_\psi^{1/2}$ on N].

$$\leq \|D_\psi^{-1/2}\|_{\mathrm{op},N} \, \|D_\psi^{1/2} w_N\|_2$$

[operator-norm bound].

$$\|D_\psi^{-1/2}\|_{\mathrm{op},N} = \underline{\psi}_N^{-1/2}, \quad \underline{\psi}_N := \min_{j\in N} \psi_j > 0$$

[largest singular value on N equals $1/\sqrt{\min_N \psi}$].

$$\|D_\psi^{1/2} w_N\|_2 \leq \|D_\psi^{1/2} w\|_2$$

[restriction cannot increase the norm].

$$\Rightarrow \ \|w_N\|_2 \leq \underline{\psi}_N^{-1/2} \, \|D_\psi^{1/2} w\|_2$$

[combine the last two displays].

$$\Rightarrow \ \Big|\sum_{j\in N} w_j\Big| \ \leq \ \sqrt{|N|}\, \underline{\psi}_N^{-1/2} \, \|D_\psi^{1/2} w\|_2 \quad \text{(UR.1)}$$

[exposed-sum controlled by the exposure-weighted ℓ_2 norm].

C. Ridge path monotonicity.

Write $J_\lambda(w) = L(w) + \lambda P(w), \quad L(w) := \|X_1 - X_0 w\|_V^2, \quad P(w) := \|D_\psi^{1/2} w\|_2^2.$

[separate fit and penalty].

$$0 \le \lambda_1 < \lambda_2, \quad w_1 := w^{UR}(\lambda_1), \; w_2 := w^{UR}(\lambda_2).$$

[pick two tuning values].

$$L(w_1) + \lambda_1 P(w_1) \;\le\; L(w_2) + \lambda_1 P(w_2)$$

[optimality of w_1 for J_{λ_1}].

$$L(w_2) + \lambda_2 P(w_2) \;\le\; L(w_1) + \lambda_2 P(w_1)$$

[optimality of w_2 for J_{λ_2}].

$$\text{Add the two: } \big(L(w_1) - L(w_2)\big) + \big(L(w_2) - L(w_1)\big) + \lambda_1 P(w_1) + \lambda_2 P(w_2)$$
$$\le\; \lambda_1 P(w_2) + \lambda_2 P(w_1).$$

[cancel L terms].

$$\lambda_1 P(w_1) + \lambda_2 P(w_2) \;\le\; \lambda_1 P(w_2) + \lambda_2 P(w_1)$$

[rearrange].

$$(\lambda_2 - \lambda_1)\big(P(w_1) - P(w_2)\big) \;\ge\; 0$$

[$\lambda_2 > \lambda_1$].

$$\Rightarrow\; P(w_2) \;\le\; P(w_1) \quad\Longleftrightarrow\quad \|D_\psi^{1/2} w^{UR}(\lambda_2)\|_2 \;\le\; \|D_\psi^{1/2} w^{UR}(\lambda_1)\|_2 \quad \text{(UR.2)}$$

[monotone nonincreasing in λ].

Limit as $\lambda \to \infty$: $w^{UR}(\lambda) = (X_0^\top V X_0 + \lambda D_\psi)^{-1} X_0^\top V X_1.$

[closed form].

$$\|(X_0^\top V X_0 + \lambda D_\psi)^{-1}\|_{\text{op}} \;\le\; \frac{1}{\lambda\, \lambda_{\min}(D_\psi)} \;=\; \frac{1}{\lambda\, \underline{\psi}}$$

[spectral bound: $\lambda_{\min}(A + \lambda B) \ge \lambda\, \lambda_{\min}(B)$ for $A \succeq 0$, $B \succ 0$].

$$\|D_\psi^{1/2} w^{UR}(\lambda)\|_2$$
$$\le\; \|D_\psi^{1/2}\|_{\text{op}}\; \|(X_0^\top V X_0 + \lambda D_\psi)^{-1}\|_{\text{op}}\; \|X_0^\top V X_1\|_2$$

[submultiplicativity of operator norm].

$$=\; \sqrt{\overline{\psi}} \;\cdot\; \frac{1}{\lambda\, \underline{\psi}} \;\cdot\; \|X_0^\top V X_1\|_2 \;\longrightarrow\; 0 \;\text{ as } \lambda \to \infty \quad \text{(UR.3)}$$

[$\overline{\psi} := \max_j \psi_j < \infty$ by (A1)].

D Bias bounds under localized and aligned spillovers

Localized, time-homogeneous spillovers (Theorem 5): $\hat{\tau}_t^{UR}(\lambda) - \tau_t = \rho\,\tau_t \sum_{j\in N} w_j^{UR}(\lambda)$.

[contamination channel depends on exposed sum].

$$\left|\hat{\tau}_t^{UR}(\lambda) - \tau_t\right| \;=\; \rho\,|\tau_t|\,\Big|\sum_{j\in N} w_j^{UR}(\lambda)\Big|$$

[absolute distortion].

$$\leq\; \rho\,|\tau_t|\,\sqrt{|N|}\,\underline{\psi}_N^{-1/2}\,\big\|D_\psi^{1/2} w^{UR}(\lambda)\big\|_2$$

[apply (UR.1) with $w = w^{UR}(\lambda)$].

By (UR.2) and (UR.3), RHS is nonincreasing in λ and $\to 0$ as $\lambda \to \infty$.

[tuning monotonicity and vanishing envelope].

General nonnegative exposure-aligned profile $\{\delta_{jt}\}$ supported on N:

$$\sum_j w_j^{UR}(\lambda)\,\delta_{jt} \;=\; \sum_{j\in N} w_j^{UR}(\lambda)\,\delta_{jt}$$

[localized exposure concentrated on N].

$$\leq\; \|\delta_{N,t}\|_2\,\|w_N^{UR}(\lambda)\|_2$$

[Cauchy–Schwarz on N].

$$\leq\; \|\delta_{N,t}\|_2\,\underline{\psi}_N^{-1/2}\,\|D_\psi^{1/2} w^{UR}(\lambda)\|_2.$$

[use $\|w_N\|_2 \leq \underline{\psi}_N^{-1/2}\|D_\psi^{1/2} w\|_2$ as in Part 1].

Hence the envelope is nonincreasing in λ and vanishes as $\lambda \to \infty$ by (UR.2)–(UR.3).

[same monotonicity and limit].

□

A.5 Spatial reach mapping

To avoid any tuning on outcomes, we calibrate f from the empirical distribution of distances alone using a logistic map,

$$f(d) \;=\; \frac{1}{1+\exp\{-\kappa(d-c)\}}. \tag{21}$$

Let d_L and d_U denote the $(q, 1-q)$ quantiles of $\{d(j,p)\}_{j\neq p}$ for a small q; we take $q = 0.025$. Fix a tail level $\varepsilon \in (0, 1/2)$; we take $\varepsilon = 0.025$. Imposing $f(d_L) = \varepsilon$ and $f(d_U) = 1-\varepsilon$ yields

$$c \;=\; \frac{d_U + d_L}{2}, \qquad \kappa \;=\; \frac{2\,\log\!\left(\frac{1-\varepsilon}{\varepsilon}\right)}{d_U - d_L}.$$

This anchors $\eta_j = f(d(j,p))$ smoothly over $(\varepsilon, 1-\varepsilon)$ so that proximate donors receive η_j near ε and distant donors near $1-\varepsilon$. We maintain $\eta_j \in [\varepsilon, 1-\varepsilon]$ for all donors to keep the optimization well-posed.

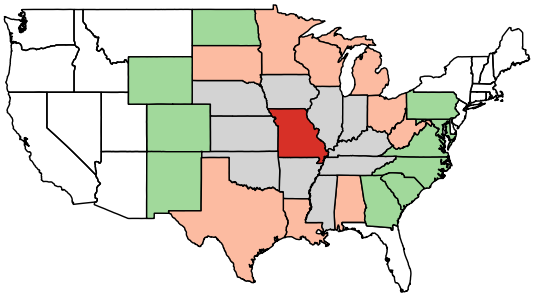
In applied work, c and κ can and should be calibrated with auxiliary information and domain knowledge.

Examples include: (i) setting c to a substantively meaningful distance (e.g., a commuting or media-market radius) and choosing κ so that f halves over a prespecified range; (ii) mapping d to network-path lengths derived from mobility/commuting matrices and anchoring (d_L, d_U) to empirical percentiles of those effective distances; or (iii) selecting (c, κ) so that f matches a pre-specified decay (e.g., half-life) suggested by prior studies of diffusion. In our simulations and applications, the quantile-anchored defaults $(q, \varepsilon) = (0.025, 0.025)$ provided stable behavior and transparent reporting without outcome-based tuning.

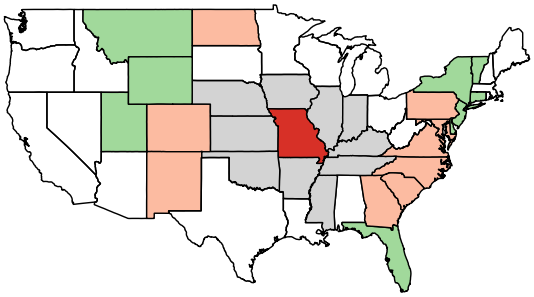
A.6 Ring geometry and mapping from continuous to discretized exposure scores

Figure A.1 illustrates the ring geometry (grid with concentric circles centered at $p^\star$) and the mapping from continuous distance $d_{ip^\star}$ to discretized exposure scores $s_k \approx \kappa(d_{ip^\star}; \theta)$.)

A.7 Alternative rings contrast



(a) 2 vs 3 Contrast for Missouri, $p = 0.9591$



(b) 3 vs 4 Contrast for Colorado, $p = 0.5102041$

Figure 6: Alternative contrasts