Sharp Closed-Form Bounds for Interference Contamination in Linear ATT Designs

Joao Alipio-Correa*

July 3, 2026

Abstract

Difference-in-differences, synthetic control, and staggered-adoption estimators build the treated counterfactual by taking a weighted average of control outcomes. When treatment spills over onto nearby controls, those outcomes carry part of the treatment effect, the counterfactual is contaminated, and the treatment-effect estimate is biased by an amount that depends on spillovers the researcher never observes. I show that the bias any such linear estimator inherits is a weighted sum of the individual spillovers on exposed controls, with weights equal to the estimator's own control weights. A doubly robust procedure identifies the average spillover effect among exposed controls under an exposure-ignorability condition, and I treat that average as a constraint on the unobserved counterfactuals rather than as a test of no interference. The constraint yields sharp closed-form bounds on the bias that solve a sorting problem in near-linear time, with width equal to an inner product between the estimator's centered weights and the sorted spillover capacities. Uniform weights point-identify the bias once the average spillover is known, whereas dispersed weights leave it genuinely partially identified. A Monte Carlo study confirms these predictions. I revisit two Brazilian programs, Rio de Janeiro's police pacification under difference-in-differences and Maricá's municipal basic income under synthetic control.

Keywords: causal inference, interference, partial identification, spillovers, doubly robust estimation

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

1 Introduction

Difference-in-differences, synthetic control, and staggered-adoption designs are the standard tools for evaluating policies that were not randomly assigned. Each forms the treated counterfactual in the same way, as a weighted average of control-group outcomes, weighted equally in two-way fixed-effects difference-in-differences (Callaway and Sant'Anna, 2021; de Chaisemartin and D'Haultfœuille, 2020), data-driven in synthetic control (Abadie et al., 2010; Abadie, 2021), and aggregated across cohorts and periods in modern staggered designs (Sun and Abraham, 2021; Borusyak et al., 2024). This construction reads the control outcomes as clean measurements of the untreated trajectory. They are not clean when treatment spills over onto the controls. A pacification program displaces crime into the neighborhoods next to the ones it treats. A cash-transfer program raises commerce in towns that border the treated municipality. In each case the controls closest to the treatment absorb part of its effect, the aggregated counterfactual is contaminated, and the estimate is biased by an amount that depends on spillovers the researcher never observes, with a sign that is not guaranteed either way. Interference of this kind is pervasive in spatial and network settings (Aronow and Samii, 2017; Sävje et al., 2021; Leung, 2022), yet the applied researcher who reaches for one of these estimators has no way to gauge how badly a given estimate is contaminated.

These estimators are usually compared on efficiency, on robustness to violations of parallel trends, and on their handling of treatment-effect heterogeneity. Under interference a different feature separates them, namely the shape of the weights they place on the control units. Difference-in-differences spreads weight evenly across all controls, synthetic control concentrates it on a handful of donors, and staggered designs induce uneven cohort-time weights. This paper argues that the shape of these weights governs whether the interference bias can be recovered at all. When the weights are uniform, knowing the average spillover pins the bias down exactly. When the weights are concentrated, the same average is consistent with

a whole range of biases, because heavily weighted units can absorb a disproportionate share of the aggregate spillover without moving its mean. Identification of the interference bias is therefore an estimator-specific property, controlled by how dispersed the estimator's weights are.

I make this precise in three steps. First, I show that the bias any linear estimator of the average effect on the treated inherits under interference, which I call the "contamination," is a weighted sum of the unobserved unit-level spillovers on exposed controls, with weights equal to the estimator's own control weights (Theorem 2.2). This decomposition makes the bias a well-defined partial-identification target, since the target is now a linear function of a bounded, unobserved vector. Second, I identify the mean of that vector. Under an exposure-ignorability condition, a doubly robust procedure consistently estimates the average spillover on the exposed controls, ψ. Rather than use ψ only to test the null of no interference, I use it as a constraint, so that the mean of the unobserved counterfactuals is pinned down by ψ and the observed data, and the configurations in which every spillover is simultaneously extreme are ruled out. Third, this mean constraint, together with outcome-support and monotonicity restrictions, defines a feasible set, and the sharp bounds on the contamination are the solutions to linear programs over it. The programs have a closed-form solution through a "sorting" characterization, in which the adversary that maximizes the bias fills the highest-weight controls to their spillover capacity first, until the aggregate budget ψ is spent.

The sorting solution makes the role of the weights visible. The width of the identification region is an inner product between the estimator's centered weight vector and the sorted spillover capacities (Theorem 4.10), so it vanishes when the weights are uniform and grows as they concentrate. Two regimes follow. Standard two-way fixed-effects difference-in-differences places equal weight on every control and point-identifies the contamination,

which equals $\rho\,\psi$ for ρ the fraction of exposed controls, so the detection estimate by itself removes the bias. Synthetic control, staggered designs, synthetic difference-in-differences (Arkhangelsky et al., 2021), and matching or inverse-probability estimators (Abadie and Imbens, 2006; Hirano et al., 2003) place uneven weight on the controls and face genuine partial identification, with an interval whose width is available in closed form. Standard difference-in-differences is the exceptional case in which the average spillover point-identifies the contamination. For most estimators in applied use the contamination remains partially identified, and the bounds developed here give the sharpest available statement about its range. When conditional detection estimates $\tau(x)$ are available, they tighten the bounds by confining the reallocation of spillover to within-stratum transfers, with the gain governed by a Jensen inequality on the budget-flexibility function.

The bounds are sharp. I certify this two ways, through linear-programming duality and through constructive configurations of the counterfactual vector that attain each endpoint. They are also cheap to compute, since the sorting solution runs in $O(n_1 \log n_1)$ time. For inference I develop plug-in estimators, establish the Lipschitz regularity of the bound mapping and the asymptotic normality it implies, and construct Imbens–Manski confidence intervals with influence-function standard errors that remain valid under spatial clustering. A sensitivity analysis over the exposure radius separates the identification boundary from the inferential one.

A Monte Carlo study across eleven data-generating configurations confirms the theory: the difference-in-differences region collapses to a point while the synthetic-control region does not, the doubly robust detector tolerates misspecification of either nuisance model, and the Imbens–Manski intervals hold their nominal coverage. Two applications to Brazilian electoral politics carry both regimes onto real data. Rio de Janeiro installed Pacifying Police Units across its favelas on a staggered schedule, and a difference-in-differences analysis of

their effect on left-bloc vote share point-identifies the contamination, which I net out of the reported effect. Maricá financed an unconditional municipal basic income from oil royalties, and a synthetic-control analysis of its effect on incumbent vote share leaves the contamination partially identified. The sharp bounds show that the qualitative conclusion survives even the worst-case spillover profile.

The framework connects two literatures that have developed largely apart. One is the detection of spillovers by semiparametric, doubly robust estimation (Robins et al., 1994; Kennedy, 2023; Chernozhukov et al., 2018). The other is the partial identification of causal parameters under incomplete information (Manski, 1990; Horowitz and Manski, 1995; Molinari, 2020). The "detection-as-constraint" principle joins them. An identified moment becomes an equality constraint on a partial-identification feasible set, and the trimming proportion of Lee (2009) and the sensitivity anchor of Bonvini and Kennedy (2022) enter their analyses the same way, though there the constraint is imposed by assumption and here it is obtained from identification. The linear-programming formulation parallels the honest sensitivity analysis of Rambachan and Roth (2023) for violations of parallel trends, a related threat to panel designs that I compare against in the supplement. A fast-growing literature treats interference in difference-in-differences and synthetic control directly (Butts, 2021; Xu, 2023; Di Stefano and Mellace, 2024; Mealli and Viviens, 2025), usually by modifying the estimator or discarding contaminated units. The approach here is complementary. It takes any linear estimator as given, with its weights fixed by the researcher's design, and returns sharp estimator-specific bounds on the bias those weights induce.

The remainder of the paper is organized as follows. Section 2 sets up the potential-outcomes framework under interference and proves the contamination decomposition. Section 3 develops the detection step, states the exposure-ignorability condition and the doubly robust estimator of the average spillover, and turns that estimate into a moment constraint.

Section 4 derives the sharp bounds, the sorting characterization, and the width theorem, and reads off the point-identification and partial-identification regimes. Section 5 gives the estimation and inference theory, including the Imbens–Manski intervals and the exposure-radius sensitivity analysis. Section 6 reports the Monte Carlo study, and Section 7 presents the two Brazilian applications. Section 8 discusses what the results imply for choosing among estimators when interference is a concern.

2 Framework and Contamination Decomposition

I consider N units indexed by $i \in \mathcal{U} = \{1, \ldots, N\}$, each observed once before treatment and once after. Treatment status is $A_i \in \{0, 1\}$, so that $\mathcal{T} = \{i : A_i = 1\}$ collects the treated units and $\mathcal{C} = \mathcal{U} \setminus \mathcal{T}$ the controls, and each unit carries pre-treatment covariates $X_i \in \mathbb{R}^p$. Under interference the outcome of a unit can respond to the whole treatment vector $\boldsymbol{A} = (A_1, \ldots, A_N)$ rather than to A_i alone. Following Aronow and Samii (2017), I summarize the relevant part of that vector through an exposure mapping $E_j = g(\boldsymbol{A}_{-j}, j) \in \{0, 1\}$, an indicator for whether the treatments of the other units place unit j in an exposed state. The mapping that motivates this paper is spatial, $E_j = \mathbb{1}\{\min_{k \in \mathcal{T}} d_{jk} < r\}$, so that a control counts as exposed when a treated unit sits within distance r of it. Exposure partitions the controls into an exposed set $\mathcal{C}_1 = \{j \in \mathcal{C} : E_j = 1\}$ of size n_1 and an unexposed set $\mathcal{C}_0 = \{j \in \mathcal{C} : E_j = 0\}$, and I write $\rho = n_1/|\mathcal{C}|$ for the exposed fraction. The top row of Figure 1 shows how the radius r carves $\mathcal{C}_1$ out of the controls.

Because own treatment a and exposure e each range over $\{0, 1\}$, every unit has four potential outcomes, and I write $\Delta Y_i(a, e) = Y_{i,\text{post}}(a, e) - Y_{i,\text{pre}}(a, e)$ for the pre-to-post change under each. I impose no anticipation, so that pre-treatment outcomes carry no (a, e) argument, and under SUTVA (Rubin, 1980) the exposure argument would be idle,

since SUTVA sets $\Delta Y_i(a,1) = \Delta Y_i(a,0)$. Interference is the failure of that equality. A control is never directly treated, so which of its two changes it reveals depends on E_j. An unexposed control hands over the clean change $\Delta Y_j(0,0)$, and an exposed control reveals instead $\Delta Y_j(0,1)$, its change under exposure but not treatment. The gap between those two is the exposed control's individual spillover effect, which no single unit reveals, since a control is exposed or not but never both. This is the fundamental problem of causal inference (Holland, 1986), with geographic proximity in the role of treatment assignment, and gathering the clean changes $\Delta Y_j(0,0)$ of the exposed controls leaves a vector that no data reveal. In symbols, the four potential outcomes of a unit, the observed change of a control, the individual spillover effect, and the vector left unidentified are

$$Y_{i,t}(a,e), \quad a \in \{0,1\}, \quad e \in \{0,1\}, \quad t \in \{\text{pre}, \text{post}\}. \tag{1}$$

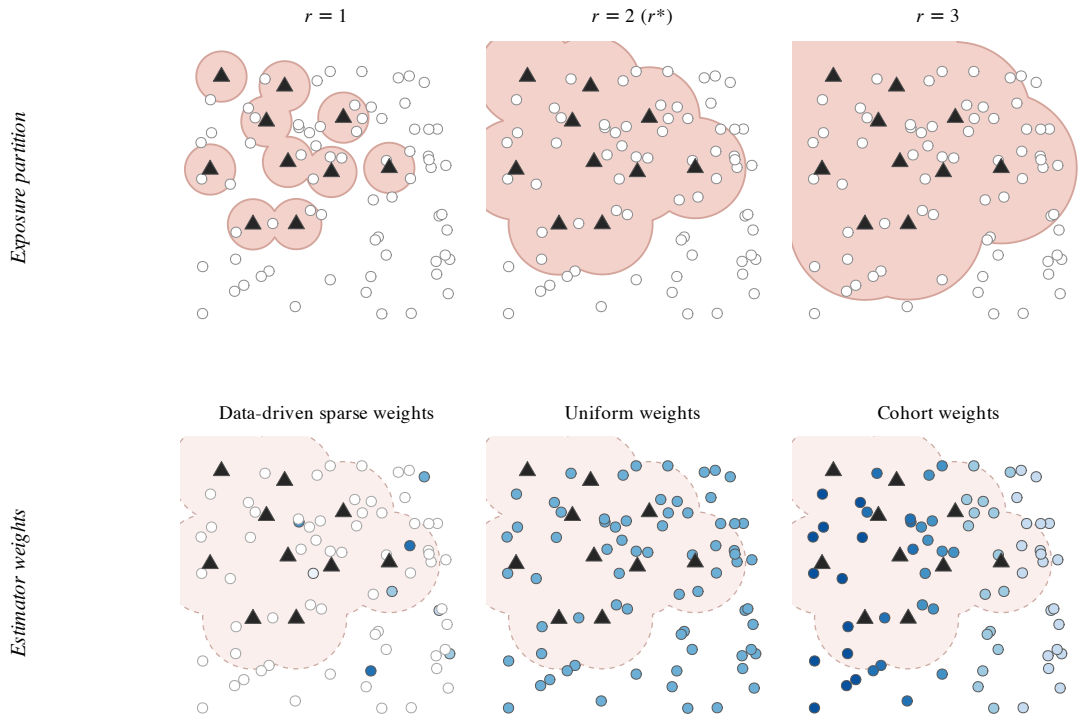
$$\Delta Y_j^{\text{obs}} = E_j \cdot \Delta Y_j(0,1) + (1 - E_j) \cdot \Delta Y_j(0,0). \tag{2}$$

$$\Delta Y_j(0,1) - \Delta Y_j(0,0) \tag{3}$$

$$\{\Delta Y_j(0,0)\}_{j \in \mathcal{C}_1}, \tag{4}$$

A useful feature of this setting is that the missing counterfactuals are attached to known units. Because E_j is a function of the observed treatments and distances, the partition into $\mathcal{C}_1$ and $\mathcal{C}_0$ is observed, so I know exactly which controls carry the unobserved clean change $\Delta Y_j(0,0)$, whereas the sample-selection and unmeasured-confounding analyses that constrain a latent counterfactual the same way cannot say which units carry it (Lee, 2009; Bonvini and Kennedy, 2022). The bottom row of Figure 1 overlays the weight vectors of three standard estimators on these same units and shows how much of each one's weight lands inside $\mathcal{C}_1$. Those weights are fixed by the design the researcher has already chosen, and I write a generic estimator in terms of them.

Figure 1: Exposure partition and weight vectors of three canonical linear ATT estimators, illustrating how weight mass landing inside the exposed region $\mathcal{C}_1$ drives the contamination $\mathcal{B}(w)$.



Note: A single realization of $N = 80$ units in a 10×10 spatial window (fixed seed) with $|\mathcal{T}| = 10$ treated units (triangles) and 70 controls (circles); the point layout is identical across all six panels. The top row varies the exposure radius r and shows how $E_j = \mathbb{1}\{\min_{k\in\mathcal{T}} d_{jk} < r\}$ partitions the controls: the coral region is the union $\bigcup_{k\in\mathcal{T}} B(k, r)$ of radius-r discs, so its interior is the exposed control set $\mathcal{C}_1 = \{j \in \mathcal{C} : E_j = 1\}$ and its complement within the window is $\mathcal{C}_0$. The bottom row fixes the exposure polygon at a reference radius r^* (redrawn with reduced opacity) and overlays the weight vector w_j on an eight-step blue ramp, with unweighted controls in white. Columns display the three canonical weight families of Definition 2.1: sparse donor weights (synthetic control), uniform weights (DiD), and cohort-block weights (staggered DiD). The schematic uses unitless distances; no estimand is computed.

Definition 2.1 (Linear ATT Estimator)**.** *A linear ATT estimator takes the form*

$$\hat{\tau}(w) = \Delta\bar{Y}_{\mathcal{T}} - \sum_{j\in\mathcal{C}} w_j \, \Delta Y_j^{\text{obs}}, \tag{5}$$

where $\Delta\bar{Y}_{\mathcal{T}} = |\mathcal{T}|^{-1} \sum_{i\in\mathcal{T}} \Delta Y_i^{\text{obs}}$ *is the average outcome change among treated units, and* $w = (w_j)_{j\in\mathcal{C}}$ *is a vector of nonnegative weights satisfying* $\sum_{j\in\mathcal{C}} w_j = 1$.

The weight vector w encodes the design the researcher has chosen, and every result below holds for whatever w that choice produces. Under interference this estimator no longer targets a clean effect. The bias it takes on is the object I study, and I call that bias the contamination. The following theorem gives its exact form.

Theorem 2.2 (Contamination Decomposition)**.** *Let $\hat{\tau}(w)$ be a linear ATT estimator as in Definition 2.1. Under interference with exposure mapping $E_j = g(\boldsymbol{A}_{-j}, j)$, the estimator admits the decomposition*

$$\hat{\tau}(w) = \tau^{\mathrm{SUTVA}}(w) - \mathcal{B}(w), \tag{6}$$

where

$$\tau^{\mathrm{SUTVA}}(w) = \Delta\bar{Y}_{\mathcal{T}} - \sum_{j \in \mathcal{C}} w_j \, \Delta Y_j(0,0) \tag{7}$$

is the quantity the estimator would target under SUTVA, and the contamination bias is

$$\mathcal{B}(w) = \sum_{j \in \mathcal{C}_1} w_j \left[\Delta Y_j(0,1) - \Delta Y_j(0,0)\right]. \tag{8}$$

Only exposed controls contribute to $\mathcal{B}(w)$; for $j \in \mathcal{C}_0$, $\Delta Y_j^{\mathrm{obs}} = \Delta Y_j(0,0)$ and the contribution is zero.

The decomposition splits the estimator into two pieces. The first, $\tau^{\mathrm{SUTVA}}(w)$, is the effect the estimator was built to recover, the value it would return if every control outcome were a clean measurement of the untreated trajectory. The second, $\mathcal{B}(w)$, is the price interference exacts, a weighted sum of the individual spillover effects on the exposed controls that runs over the estimator's own weights. Unexposed controls drop out, because for them the observed change already equals the clean change. How exposed an estimator is to interference therefore turns on the weight it places on the controls that spillovers have reached, which is why the identification question that follows is estimator-specific.

All proofs are collected in the Supplement.

Proposition 2.3 (Equivalent Representations and Sign)**.** *The contamination $\mathcal{B}(w)$ admits the following properties.*

(i) ***Observed–counterfactual decomposition.*** *Let $W_{\mathrm{exp}} = \sum_{j \in \mathcal{C}_1} w_j$ denote the total weight on exposed controls. Then*

$$\mathcal{B}(w) = \sum_{j \in \mathcal{C}_1} w_j \, \Delta Y_j^{\mathrm{obs}} \; - \; \sum_{j \in \mathcal{C}_1} w_j \, \Delta Y_j(0,0), \tag{9}$$

where the first sum is computable from the data and the weight vector, and the second sum depends on the unobserved counterfactual vector $\{\Delta Y_j(0,0)\}_{j\in\mathcal{C}_1}$.

(ii) ***Vanishing conditions.*** $\mathcal{B}(w) = 0$ *if and only if at least one of the following holds: (a) no control unit is exposed,* $\mathcal{C}_1 = \emptyset$*; (b) the estimator places zero weight on all exposed controls,* $w_j = 0$ *for all* $j \in \mathcal{C}_1$*; or (c) the weighted average spillover effect among exposed controls is exactly zero,* $\sum_{j\in\mathcal{C}_1} w_j[\Delta Y_j(0,1) - \Delta Y_j(0,0)] = 0$.

(iii) ***Sign under monotone spillovers.*** *Suppose* $\Delta Y_j(0,1) \geq \Delta Y_j(0,0)$ *for all* $j \in \mathcal{C}_1$ *(monotone nonnegative spillovers).*[1] *Then* $\mathcal{B}(w) \geq 0$*, and* $\hat{\tau}(w) \leq \tau^{\text{SUTVA}}(w)$*: the estimator is weakly downward-biased relative to its SUTVA target. Under monotone nonpositive spillovers,* $\mathcal{B}(w) \leq 0$ *and the bias reverses.*

$$\mathcal{B}(w) = W_{\text{exp}} \cdot \left(\bar{\Delta Y}_w^{(1)} - \bar{C}_w\right), \tag{10}$$

Written this way, the contamination is the total exposed weight W_{exp} times the gap between two weighted averages over the exposed controls, the observed one $\bar{\Delta Y}_w^{(1)} = W_{\text{exp}}^{-1} \sum_{j\in\mathcal{C}_1} w_j \Delta Y_j(0,1)$, which the data deliver, and the counterfactual one $\bar{C}_w = W_{\text{exp}}^{-1} \sum_{j\in\mathcal{C}_1} w_j \Delta Y_j(0,0)$, which they do not. This is how the contamination model of Horowitz and Manski (1995) appears once it is expressed through the estimator's weights, and the Supplement develops the connection in full. Everything unknown in $\mathcal{B}(w)$ therefore sits in the single scalar $\bar{C}_w$, or equivalently in the vector $\{\Delta Y_j(0,0)\}_{j\in\mathcal{C}_1}$ from which it is built, so a range for that vector maps directly into a range for the contamination. Two restrictions confine the vector. When outcome changes are known to lie in a support $[a,b]$, each unobserved $\Delta Y_j(0,0)$ is confined to $[a,b]$, and I call the resulting feasible set $\mathcal{F}_0$. When spillovers are in addition monotone nonnegative, so that exposure can only have raised each exposed control's outcome change, the clean change cannot exceed the observed one and the upper limit tightens to ΔY_j^{obs}, a bound that differs from unit to unit and gives the smaller feasible set $\mathcal{F}_1$. Each feasible set for the vector translates into an interval for the contamination.

Proposition 2.4 (Identification Region)**.** *Let* $\hat{\tau}(w)$ *be a linear ATT estimator. Under interference:*

[1] "Monotone nonnegative" is a sign restriction on the direction of spillovers: exposure weakly increases the outcome change for each unit. It does not refer to monotonicity in distance, dosage, or any other continuous variable.

(i) Under outcome support $\Delta Y \in [a, b]$ alone, the contamination satisfies

$$\mathcal{B}(w) \in \left[\sum_{j \in \mathcal{C}_1} w_j (\Delta Y_j^{\text{obs}} - b),\ \sum_{j \in \mathcal{C}_1} w_j (\Delta Y_j^{\text{obs}} - a) \right]. \tag{11}$$

(ii) Under outcome support and monotone nonnegative spillovers,

$$\mathcal{B}(w) \in \left[0,\ \sum_{j \in \mathcal{C}_1} w_j (\Delta Y_j^{\text{obs}} - a) \right]. \tag{12}$$

The lower bound is achieved when $\Delta Y_j(0,0) = \Delta Y_j^{\text{obs}}$ for all $j \in \mathcal{C}_1$ (zero spillovers). The upper bound is achieved when $\Delta Y_j(0,0) = a$ for all $j \in \mathcal{C}_1$ (maximal spillovers).

These bounds are sharp, since the configurations named in the proposition attain the endpoints, but they can be wide, because support and monotonicity are all they use. The detection step of Section 3 adds information about the missing vector. Under an exposure-ignorability condition a doubly robust procedure identifies the average spillover ψ on the exposed controls, which fixes the mean of the unobserved vector and rules out the configurations in which every spillover is simultaneously extreme. This constraint shrinks $\mathcal{F}_1$ to a smaller set $\mathcal{F}_2$. When conditional spillover estimates are available they impose the same kind of constraint stratum by stratum and shrink the set further to $\mathcal{F}_3$. Section 4 solves the resulting bounds in closed form and reads their width off the estimator's weights, so that the feasible sets nest as $\mathcal{F}_0 \supseteq \mathcal{F}_1 \supseteq \mathcal{F}_2 \supseteq \mathcal{F}_3$ and the contamination interval tightens at each step.

The form the contamination takes for difference-in-differences, synthetic control, and staggered adoption is recorded in Corollaries S1 through S3 of the Supplement.

The exposure mapping E_j is itself a modeling choice, and the radius r that defines it is seldom known with certainty. Supplement S9 studies how the bounds respond as r is varied, and separates the range of contaminations that the identification assumptions allow from the sampling uncertainty around it.

3 Detection: Identification and Doubly Robust Estimation

Theorem 2.2 leaves the bias of any linear ATT estimator a function of the unobserved counterfactual vector $\{\Delta Y_j(0,0)\}_{j\in\mathcal{C}_1}$, and under outcome support and monotonicity alone Proposition 2.4 constrains that vector only loosely. The region narrows once the data are made informative about the missing counterfactuals, and the control group itself supplies the leverage. For an unexposed control, $\Delta Y_j(0,0)$ is observed directly. For an exposed control it is counterfactual. If whether a control sits near a treated unit is as good as random once I condition on covariates, then the unexposed controls report what the exposed controls' outcome changes would have been absent any spillover, and identification proceeds by treating exposure as a treatment applied to the control group.

Identifying the mean of $\{\Delta Y_j(0,0)\}_{j\in\mathcal{C}_1}$ from the observed data rests on five conditions, of which the first is substantive and the remaining four are supporting or regularity conditions.

Assumption 1 (Exposure Ignorability)**.** *Conditional on pre-treatment covariates and the spatial neighborhood structure, exposure is independent of potential outcomes under no exposure:*

$$\Delta Y_j(0,0) \perp\!\!\!\perp E_j \mid X_j, \{X_k\}_{k\in\mathcal{N}_j}, A_j = 0.$$

Assumption 2 (Treatment Ignorability)**.** *Treatment assignment is independent of potential outcomes conditional on covariates:*

$$A_i \perp\!\!\!\perp (\Delta Y_i(a,e))_{a,e} \mid X_i.$$

Assumption 3 (Conditional Independence of Treatment Assignments)**.** *Conditional on covariates, treatment assignments are independent across units:*

$$A_i \perp\!\!\!\perp A_j \mid X_i, X_j \quad \text{for } i \neq j.$$

Assumption 4 (Positivity for Exposure)**.** *Every control unit has a bounded probability of being in either exposure group:*

$$\eta < \mathbb{P}\left(E_j = 1 \mid X_j, \{X_k\}_{k\in\mathcal{N}_j}, A_j = 0\right) < 1-\eta \quad \text{for some } \eta > 0 \text{ and all } j \in \mathcal{C}.$$

Assumption 5 (Covariate Overlap)**.** *The covariate distributions of exposed and unexposed controls have common support:*

$$supp(X_j \mid E_j = 1, A_j = 0) \subseteq supp(X_j \mid E_j = 0, A_j = 0).$$

Exposure ignorability (Assumption 1) is the substantive requirement. It asks that, once I condition on a control's own covariates and those of the units in its exposure neighborhood, whether the control sits near a treated unit carry no information about its outcome change under no exposure. The condition holds by construction when treatment is randomized with known probabilities and is credible under observable adoption rules such as population or fiscal thresholds, and it fails when an unobserved local shock both draws treatment to an area and moves the outcome trajectories of nearby controls, as in conflict zones where violence drives adoption and decline together. Assumptions 2 and 3 give one sufficient foundation, since they make a control's exposure a function of its neighbors' treatment statuses alone. Positivity and overlap (Assumptions 4 and 5) are the usual regularity conditions that keep the inverse-propensity weights bounded and ensure every exposed covariate profile has unexposed counterparts. I take up the plausibility of exposure ignorability, and a falsifiable balance check, in the Supplement.

Definition 3.1 (Detection Estimand)**.** *The detection estimand is the average causal effect of exposure on outcome changes among control units:*

$$\psi = \mathbb{E}\left[\Delta Y_j(0,1) - \Delta Y_j(0,0) \mid A_j = 0\right]. \tag{13}$$

The conditional detection estimand at covariate value x is

$$\tau(x) = \mathbb{E}\left[\Delta Y_j(0,1) - \Delta Y_j(0,0) \mid X_j = x, A_j = 0\right], \tag{14}$$

so that $\psi = \mathbb{E}[\tau(X_j) \mid A_j = 0]$.

The two estimands are contrasts of the unobserved $\Delta Y_j(0,0)$ and so are not directly computable, yet under the five conditions each equals an observable contrast between exposed and unexposed controls, and through those contrasts the data pin down the mean of the counterfactual vector on which the contamination depends.

Theorem 3.2 (Identification of the Spillover Mean and the Counterfactual Constraint)**.** *Under Assumptions 1–5:*

(i) The conditional detection estimand $\tau(x)$ is identified by the observable contrast

$$\tau(x) = \mathbb{E}[\Delta Y_j \mid E_j = 1, X_j = x, A_j = 0] - \mathbb{E}[\Delta Y_j \mid E_j = 0, X_j = x, A_j = 0]. \tag{15}$$

(ii) The detection estimand ψ is identified by the integrated contrast

$$\psi = \mathbb{E}\left[\mathbb{E}[\Delta Y_j \mid E_j = 1, X_j, A_j = 0] - \mathbb{E}[\Delta Y_j \mid E_j = 0, X_j, A_j = 0] \mid A_j = 0\right]. \quad (16)$$

(iii) The population mean of the unobserved counterfactual outcome changes for exposed controls is identified:[2]

$$\mathbb{E}\left[\Delta Y_j(0,0) \mid E_j = 1, A_j = 0\right] = \mathbb{E}\left[\Delta Y_j^{\text{obs}} \mid E_j = 1, A_j = 0\right] - \psi. \quad (17)$$

(iv) The conditional mean of the unobserved counterfactuals at covariate value x is identified:

$$\mathbb{E}\left[\Delta Y_j(0,0) \mid X_j = x, E_j = 1, A_j = 0\right] = \mathbb{E}\left[\Delta Y_j \mid E_j = 0, X_j = x, A_j = 0\right]. \quad (18)$$

$$\frac{1}{n_1} \sum_{j \in \mathcal{C}_1} \Delta Y_j(0,0) \approx \bar{\Delta Y}_{\mathcal{C}_1}^{\text{obs}} - \psi, \quad (19)$$

Parts (i) and (ii) identify the spillover mean and its covariate-specific version by the logic that identifies an average treatment effect under unconfoundedness. Parts (iii) and (iv) feed the bounds. Part (iii) fixes the mean of the missing counterfactuals: once ψ is known, the average of $\{\Delta Y_j(0,0)\}_{j \in \mathcal{C}_1}$ equals the observed exposed mean less ψ, and the sample analog in (19) is the mean constraint I carry into Section 4. Part (iv) refines this covariate by covariate, so that when the covariates fully determine $\Delta Y_j(0,0)$ the counterfactuals are pinned down one by one and the contamination is point-identified for every weight vector.

Estimating the bounds reduces to estimating ψ, and I estimate it with a doubly robust estimator built on two nuisance models. The first is the outcome regression among unexposed controls, $\mu_0(X)$ in (20), which by part (iv) also equals the conditional mean of the missing counterfactuals for exposed controls at the same covariate value. The second is the exposure propensity π_j in (21), the probability that a control is exposed given the covariates of the units in its exposure neighborhood. Supplement S2 records the exposed-control regression μ_1, the conditional detector $\hat{\tau} = \hat{\mu}_1 - \hat{\mu}_0$, and the product formula that assembles π_j from unit-level treatment propensities. The estimator $\hat{\psi}_{\text{DR}}$ in (22) averages the outcome-model residual

[2]The accounting identity $\mathbb{E}[\Delta Y_j(0,0) \mid E_j = 1, A_j = 0] = \mathbb{E}[\Delta Y_j^{obs} \mid E_j = 1, A_j = 0] - \psi_{ATT}$ holds by definition of $\psi_{ATT} = \mathbb{E}[\Delta Y_j(0,1) - \Delta Y_j(0,0) \mid E_j = 1, A_j = 0]$. Equation (17) substitutes the average detection estimand ψ for ψ_{ATT}; the two coincide under exposure ignorability (Assumption 1), as established in Remark S2. Using ψ rather than ψ_{ATT} is deliberate: ψ averages over the full control group and is estimated more precisely.

over the exposed controls and corrects it with the same residual over the unexposed controls, reweighted by $\hat{\pi}_j/(1-\hat{\pi}_j)$ so the correction is drawn from the exposed group's covariate distribution. It is consistent for ψ whenever either nuisance model is correctly specified, and attains the semiparametric efficiency bound when both converge quickly enough, as Proposition 3.3 states.

$$\mu_0(X) = \mathbb{E}[\Delta Y_j \mid E_j = 0, X_j = X, A_j = 0], \tag{20}$$

$$\pi_j = \mathbb{P}\left(E_j = 1 \mid \{X_k\}_{k\in\mathcal{N}_j}, A_j = 0\right), \tag{21}$$

$$\hat{\psi}_{\mathrm{DR}} = \frac{1}{n_1}\sum_{j\in\mathcal{C}_1}\left[\Delta Y_j - \hat{\mu}_0(X_j)\right] - \frac{1}{n_1}\sum_{j\in\mathcal{C}_0}\frac{\hat{\pi}_j}{1-\hat{\pi}_j}\left[\Delta Y_j - \hat{\mu}_0(X_j)\right]. \tag{22}$$

Proposition 3.3 (Double Robustness)**.** *Under Assumptions 1–5, the estimator $\hat{\psi}_{\mathrm{DR}}$ in (22) is consistent for ψ if either:*

(a) the outcome model $\hat{\mu}_0(X)$ is consistent for $\mu_0(X)$, regardless of whether $\hat{\pi}_j$ is consistent for π_j; or

(b) the exposure propensity $\hat{\pi}_j$ is consistent for π_j, regardless of whether $\hat{\mu}_0(X)$ is consistent for $\mu_0(X)$.

When both nuisance models are consistent at rates $o_p(n^{-1/4})$, $\hat{\psi}_{\mathrm{DR}}$ is $\sqrt{n}$-consistent and asymptotically normal, with variance achieving the semiparametric efficiency bound for ψ under the nonparametric model defined by Assumptions 1–5. The proof, which follows from the product-rate condition on the influence function remainder, is in Appendix S3.2.

Proposition 3.4 (Detection Estimates as Feasible Set Constraints)**.** *Under the conditions of Theorem 3.2 and consistency of $\hat{\psi}_{\mathrm{DR}}$:*

(i) The estimated mean constraint

$$\frac{1}{n_1}\sum_{j\in\mathcal{C}_1}\Delta Y_j(0,0) = \bar{\Delta Y}^{\mathrm{obs}}_{\mathcal{C}_1} - \psi, \tag{23}$$

where $\bar{\Delta Y}^{\mathrm{obs}}_{\mathcal{C}_1} = n_1^{-1}\sum_{j\in\mathcal{C}_1}\Delta Y_j^{\mathrm{obs}}$, determines the population mean of the unobserved counterfactuals from observable quantities and ψ. The right-hand side is consistently estimated by $\bar{\Delta Y}^{\mathrm{obs}}_{\mathcal{C}_1} - \hat{\psi}_{\mathrm{DR}}$.

(ii) The estimated conditional constraint at covariate value x,

$$\mathbb{E}\left[\Delta Y_j(0,0) \mid X_j = x, E_j = 1, A_j = 0\right] = \mu_0(x), \tag{24}$$

determines the conditional mean of the unobserved counterfactuals from the outcome model for unexposed controls, consistently estimated by $\hat{\mu}_0(x)$.

Proposition 3.4 connects the detection framework to the partial identification analysis. Adding the estimated mean constraint to the feasible set $\mathcal{F}_1$ of Proposition 2.4 produces the strictly smaller $\mathcal{F}_2$, the conditional constraints shrink it further to $\mathcal{F}_3$, and Section 4 solves for the sharp bounds on the contamination over these sets.

4 Sharp Bounds on the Contamination

The contamination is a weighted sum of the spillovers that treatment sends onto the exposed controls (Theorem 2.2), all of them unobserved. The detection step of Section 3 recovers their average ψ and, within covariate strata, the conditional averages $\tau(x_k)$, but it leaves the individual spillovers free to vary around those averages. This section answers how far the contamination can move once its average is fixed. I read the problem as a contest between the researcher, who has already fixed the estimator weights w, and an adversary, who assigns the unobserved spillovers to drive the contamination as high or as low as the maintained assumptions permit. The reachable values form the sharp identification region, and its endpoints solve two optimization problems over a feasible set of counterfactual vectors. I describe that feasible set in two equivalent coordinate systems, the counterfactual levels used in the definitions just below and the spillovers used for the optimization, and pass between them once both are in hand. I split the observed part of the estimator from the unobserved part and write the contamination as

$$\mathcal{B}(w) = \underbrace{\sum_{j \in \mathcal{C}_1} w_j \, \Delta Y_j^{\text{obs}}}_{S_{\text{obs}}(w)} \; - \; \sum_{j \in \mathcal{C}_1} w_j \, \Delta Y_j(0,0), \tag{25}$$

where $S_{\text{obs}}(w)$ is a known constant computable from the data. Bounding $\mathcal{B}(w)$ is equivalent to optimizing the weighted sum of unobserved counterfactuals over the feasible set $\mathcal{F}$, that is, over all vectors $(c_j)_{j \in \mathcal{C}_1}$ of counterfactual outcome changes $c_j = \Delta Y_j(0,0)$ satisfying the constraints at the chosen refinement level:

$$\mathcal{B}^{\text{U}}(w) = S_{\text{obs}}(w) - \min_{(c_j) \in \mathcal{F}} \sum_{j \in \mathcal{C}_1} w_j \, c_j, \qquad \mathcal{B}^{\text{L}}(w) = S_{\text{obs}}(w) - \max_{(c_j) \in \mathcal{F}} \sum_{j \in \mathcal{C}_1} w_j \, c_j. \tag{26}$$

Definition 4.1 (Level 1 Feasible Set)**.** *Under outcome support $\Delta Y \in [a, b]$ and monotone nonnegative spillovers ($\Delta Y_j(0,1) \geq \Delta Y_j(0,0)$ for all $j \in \mathcal{C}_1$), the feasible set is*

$$\mathcal{F}_1 = \left\{ (c_j)_{j \in \mathcal{C}_1} \in \mathbb{R}^{n_1} : c_j \in \left[a,\ \Delta Y_j^{\text{obs}} \right] \text{ for all } j \in \mathcal{C}_1 \right\}. \tag{27}$$

Definition 4.2 (Level 2 Feasible Set)**.** *Under the conditions of Definition 4.1 and Assumptions 1–5 (enabling identification of ψ), the feasible set is*

$$\mathcal{F}_2 = \left\{ (c_j)_{j \in \mathcal{C}_1} \in \mathcal{F}_1 : \frac{1}{n_1} \sum_{j \in \mathcal{C}_1} c_j = \bar{\Delta Y}_{\mathcal{C}_1}^{\text{obs}} - \psi \right\}, \tag{28}$$

where $\bar{\Delta Y}_{\mathcal{C}_1}^{\text{obs}} = n_1^{-1} \sum_{j \in \mathcal{C}_1} \Delta Y_j^{\text{obs}}$ and ψ is the detection estimand from Theorem 3.2(iii).

Definition 4.3 (Level 3 Feasible Set)**.** *Let $\{S_1, \ldots, S_K\}$ be a partition of $\mathcal{C}_1$ into K covariate strata, with $|S_k| = n_k$ and associated covariate values x_k. Under the conditions of Definition 4.1 and Assumptions 1–5 (enabling identification of $\tau(x_k)$ for each k), the feasible set is*

$$\mathcal{F}_3 = \left\{ (c_j)_{j \in \mathcal{C}_1} \in \mathcal{F}_1 : \frac{1}{n_k} \sum_{j \in S_k} c_j = \bar{\Delta Y}_{S_k}^{\text{obs}} - \tau(x_k) \ \text{ for each } k = 1, \ldots, K \right\}, \tag{29}$$

where $\bar{\Delta Y}_{S_k}^{\text{obs}} = n_k^{-1} \sum_{j \in S_k} \Delta Y_j^{\text{obs}}$ and $\tau(x_k)$ is the conditional detection estimand from Theorem 3.2(i).

The counterfactual levels are convenient for stating the assumptions, but the optimization is cleaner in the spillovers themselves. Write $\delta_j = \Delta Y_j^{\text{obs}} - \Delta Y_j(0,0)$ for the spillover on exposed control j and $d_j = \Delta Y_j^{\text{obs}} - a$ for its capacity, so that the contamination is $\mathcal{B}(w) = \sum_{j \in \mathcal{C}_1} w_j \delta_j$ and the three feasible sets become

$$\mathcal{F}_1^{\delta} = \left\{ (\delta_j)_{j \in \mathcal{C}_1} : \delta_j \in [0,\ d_j] \text{ for all } j \in \mathcal{C}_1 \right\}, \tag{30}$$

$$\mathcal{F}_2^{\delta} = \left\{ (\delta_j) \in \mathcal{F}_1^{\delta} : \frac{1}{n_1} \sum_{j \in \mathcal{C}_1} \delta_j = \psi \right\}, \tag{31}$$

$$\mathcal{F}_3^{\delta} = \left\{ (\delta_j) \in \mathcal{F}_1^{\delta} : \frac{1}{n_k} \sum_{j \in S_k} \delta_j = \tau(x_k) \text{ for each } k = 1, \ldots, K \right\}. \tag{32}$$

The box $\delta_j \in [0, d_j]$ carries nonnegativity and the outcome-support bound together, the Level 2 constraint fixes the average spillover at the detection estimand ψ, and the Level 3 constraints fix the average within each stratum at $\tau(x_k)$.

Adding constraints can only shrink the feasible set, so the three levels are nested and their identification regions nested in turn.

Lemma 4.4 (Nesting of Feasible Sets)**.** $\mathcal{F}_3 \subseteq \mathcal{F}_2 \subseteq \mathcal{F}_1$.

Corollary 4.5 (Nesting of Identification Regions)**.** *For any weight vector w,*

$$\left[\mathcal{B}_3^{\mathrm{L}}(w),\ \mathcal{B}_3^{\mathrm{U}}(w)\right] \subseteq \left[\mathcal{B}_2^{\mathrm{L}}(w),\ \mathcal{B}_2^{\mathrm{U}}(w)\right] \subseteq \left[\mathcal{B}_1^{\mathrm{L}}(w),\ \mathcal{B}_1^{\mathrm{U}}(w)\right],$$

where $\mathcal{B}_k^{\mathrm{U}}(w) = S_{\mathrm{obs}}(w) - \min_{(c_j)\in\mathcal{F}_k} \sum w_j c_j$ *and* $\mathcal{B}_k^{\mathrm{L}}(w) = S_{\mathrm{obs}}(w) - \max_{(c_j)\in\mathcal{F}_k} \sum w_j c_j$ *for* $k = 1, 2, 3$.

Which level to report is dictated by the weakest assumption the design can defend, Level 1 when exposure ignorability is itself in doubt, Level 2 when it is credible but the covariates are too coarse for stratum-level detection, and Level 3 when abundant covariates and adequate stratum sizes make the conditional estimands reliable. The supplement develops the full trade-off.

The feasible set is common to every estimator, so they differ only in the objective: the weights each places on the exposed controls and how those weights meet the mean constraint. Fixing the average spillover couples the individual spillovers, since raising one above ψ forces another below it, and whether that coupling binds the contamination depends on the weights. Equal weights leave the weighted sum untouched, while unequal weights let the adversary shift spillover mass toward the heavily weighted controls and move the contamination without changing its mean, and the next two propositions make the two regimes precise, first point identification under symmetric weights and then genuine partial identification under asymmetric weights.

Proposition 4.6 (Point Identification Under Symmetric Weights)**.** *Suppose the weight vector w assigns equal weight to all exposed controls: $w_j = \bar{w}$ for all $j \in \mathcal{C}_1$, for some constant $\bar{w} > 0$. Then under the Level 2 feasible set $\mathcal{F}_2$, the contamination $\mathcal{B}(w)$ is point-identified:*

$$\mathcal{B}(w) = W_{\mathrm{exp}} \cdot \psi, \tag{33}$$

where $W_{\mathrm{exp}} = \sum_{j\in\mathcal{C}_1} w_j = n_1\bar{w}$ is the total weight on exposed controls.

Proposition 4.7 (Partial Identification Under Asymmetric Weights)**.** *Suppose the weight vector w assigns unequal weights to at least two exposed controls: there exist $j, j' \in \mathcal{C}_1$ with $w_j \neq w_{j'}$. Suppose further that $\mathcal{F}_2$ contains more than one element (which holds whenever $n_1 \geq 2$ and the box constraints are not all degenerate). Then the identification region for*

$\mathcal{B}(w)$ has strictly positive width:

$$\mathcal{B}_2^{\mathrm{U}}(w) - \mathcal{B}_2^{\mathrm{L}}(w) > 0.$$

This dichotomy sorts the standard estimators. Two-way fixed-effects difference-in-differences weights every control equally and point-identifies the contamination at $\rho\psi$ for ρ the fraction of exposed controls, whereas synthetic control (Abadie et al., 2010), staggered designs (Sun and Abraham, 2021; Borusyak et al., 2024), synthetic difference-in-differences (Arkhangelsky et al., 2021), and matching or inverse-probability estimators (Abadie and Imbens, 2006; Hirano et al., 2003) place uneven weight on the controls and face genuine partial identification, and the supplement works out the per-estimator weights and region widths.

Definition 4.8 (Bounding Linear Programs)**.** *For $k \in \{1, 2, 3\}$, the sharp upper and lower bounds on $\mathcal{B}(w)$ are*

$$\mathcal{B}_k^{\mathrm{U}}(w) = \max_{(\delta_j) \in \mathcal{F}_k^{\delta}} \sum_{j \in \mathcal{C}_1} w_j \, \delta_j, \qquad \mathcal{B}_k^{\mathrm{L}}(w) = \min_{(\delta_j) \in \mathcal{F}_k^{\delta}} \sum_{j \in \mathcal{C}_1} w_j \, \delta_j. \tag{34}$$

The identification region for the contamination at Level k is $[\mathcal{B}_k^{\mathrm{L}}(w),\ \mathcal{B}_k^{\mathrm{U}}(w)]$.

At Level 2 the upper bound solves

$$\begin{aligned} \mathcal{B}_2^{\mathrm{U}}(w) \ = \ \max_{\delta \in \mathbb{R}^{n_1}} \quad & \sum_{j \in \mathcal{C}_1} w_j \, \delta_j \\ \text{subject to} \quad & 0 \le \delta_j \le d_j \quad \text{for all } j \in \mathcal{C}_1, \\ & \frac{1}{n_1} \sum_{j \in \mathcal{C}_1} \delta_j = \psi. \end{aligned} \tag{35}$$

The sharp lower bound is the corresponding minimization,

$$\begin{aligned} \mathcal{B}_2^{\mathrm{L}}(w) \ = \ \min_{\delta \in \mathbb{R}^{n_1}} \quad & \sum_{j \in \mathcal{C}_1} w_j \, \delta_j \\ \text{subject to} \quad & 0 \le \delta_j \le d_j \quad \text{for all } j \in \mathcal{C}_1, \\ & \frac{1}{n_1} \sum_{j \in \mathcal{C}_1} \delta_j = \psi. \end{aligned} \tag{36}$$

The decision vector is the unobserved spillover profile, the objective is the contamination itself, the box constraints carry monotonicity and support, and the single equality constraint ties the average spillover to the detection estimand ψ. Both programs are feasible exactly when that average lies in the admissible range, which the next assumption records.

Assumption 6 (Feasibility)**.** *The detection estimand satisfies* $0 \leq \psi \leq \bar{d}$*, where* $\bar{d} = n_1^{-1} \sum_{j \in \mathcal{C}_1} d_j$ *is the average unit-specific capacity. For Level 3, the stratum-specific estimands satisfy* $0 \leq \tau(x_k) \leq \bar{d}_k$ *for each* k*, where* $\bar{d}_k = n_k^{-1} \sum_{j \in S_k} d_j$*.*

A detection estimate outside $[0, \bar{d}]$ leaves the mean constraint unsatisfiable and points to a violated monotonicity or support assumption rather than to a valid input. Within the admissible range, the Karush-Kuhn-Tucker conditions for these bounded-variable programs, collected in the supplement, reduce the optimum to a single threshold on the weights. Every exposed control weighted above the threshold is filled to its capacity d_j, every one below it is held at zero, and at most one control at the threshold takes an interior value pinned by the mean constraint. Sorting the controls by increasing weight, $w_{(1)} < \cdots < w_{(n_1)}$ with capacities $d_{(1)}, \ldots, d_{(n_1)}$, turns this rule into a greedy fill from the top of the order downward, and to describe it I track the capacity remaining above and below each rank. Define the cumulative capacity from the top:

$$C(k) = \sum_{\ell=k}^{n_1} d_{(\ell)}, \qquad k = 1, \ldots, n_1 + 1, \tag{37}$$

with $C(n_1 + 1) = 0$, and from the bottom:

$$D(k) = \sum_{\ell=1}^{k} d_{(\ell)}, \qquad k = 0, 1, \ldots, n_1, \tag{38}$$

with $D(0) = 0$. Both are strictly monotone, with $D(n_1) = C(1) = n_1 \bar{d} \geq n_1 \psi$ by feasibility.

Theorem 4.9 (Sorting Structure of the Level 2 Solution)**.** *Assume the weights* $\{w_j\}_{j \in \mathcal{C}_1}$ *are distinct. Sort them in increasing order* $w_{(1)} < \cdots < w_{(n_1)}$ *with associated capacities* $d_{(1)}, \ldots, d_{(n_1)}$*.*

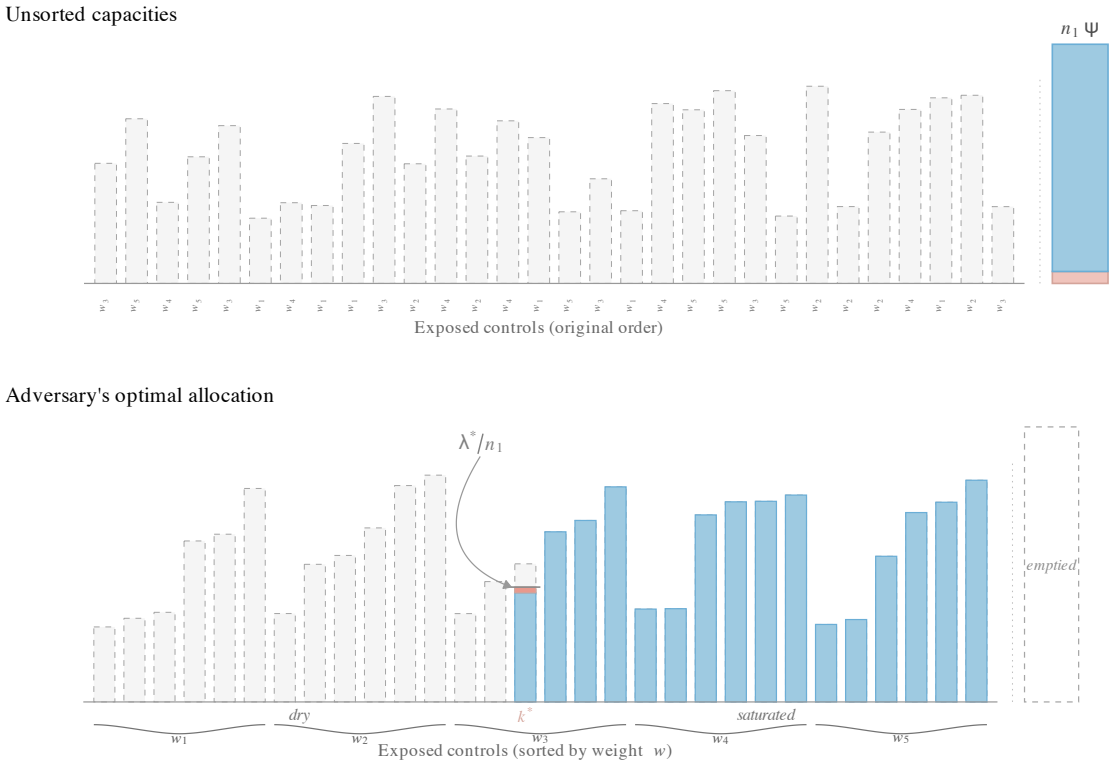
(I) For the maximization (35)*, there exists a unique index* $k^* \in \{1, \ldots, n_1\}$ *such that the optimal solution is*

$$\delta^*_{(j)} = \begin{cases} 0 & \text{if } j < k^*, \\ n_1\psi - C(k^*{+}1) & \text{if } j = k^*, \\ d_{(j)} & \text{if } j > k^*, \end{cases} \tag{39}$$

where k^* *is the unique index satisfying* $C(k^*{+}1) < n_1\psi \leq C(k^*)$*.*

(II) For the minimization (36)*, there exists a unique index* $k_* \in \{1, \ldots, n_1\}$ *such that the*

Figure 2: Water-filling characterization of the Level 2 adversarial allocation (Theorem 4.9): sorting exposed controls by weight w_j converts the spillover budget $n_1\psi$ into the greedy allocation δ_j^*, with units above the dual threshold λ^*/n_1 saturated, units below dry, and the threshold unit k^* absorbing the residual.



Note: The top panel displays 30 exposed controls in their original (unsorted) order with capacities $d_j \sim \text{Uniform}(0.30, 1.00)$; the column on the right depicts the spillover budget $n_1\psi$ that the adversary must redistribute across the units. The bottom panel reorders the same units by ascending weight (five cohorts with weights $\{0.02, 0.05, 0.10, 0.18, 0.30\}$, six units per cohort) and shows the optimum: greedy allocation from the highest-weight unit downward, equivalent by the rearrangement inequality (Hardy et al., 1952) to the co-monotone pairing that maximizes $\sum_{j\in\mathcal{C}_1} w_j\delta_j$. Exactly one interior $\delta_j^* \in (0, d_j)$ occurs at the threshold unit k^*, consistent with the extreme-point structure of a bounded-variable linear program with one equality constraint; the short horizontal mark locates the dual multiplier $\lambda^*/n_1 = w_{(k^*)}$ on the weight axis (Theorem 4.9; (37)). Within-cohort ties are broken arbitrarily per Remark S3 without affecting $\mathcal{B}_2^{\mathrm{U}}(w)$ or $\mathcal{B}_2^{\mathrm{L}}(w)$. The figure is a stylized illustration of the upper-bound allocation (Theorem 4.9(I)); the counter-monotone lower-bound allocation is not depicted.

optimal solution is

$$\delta^{**}_{(j)} = \begin{cases} d_{(j)} & \text{if } j < k_*, \\ n_1\psi - D(k_*-1) & \text{if } j = k_*, \\ 0 & \text{if } j > k_*, \end{cases} \tag{40}$$

where k_* is the unique index satisfying $D(k_*-1) < n_1\psi \leq D(k_*)$.

The two solutions are one greedy construction run in opposite directions. For the upper bound the adversary spends a fixed budget of $n_1\psi$ on the highest-weight controls first and fills each to capacity until the budget is exhausted, and the threshold control k^* absorbs the remainder. For the lower bound the budget goes to the lowest-weight controls instead. This is the discrete rearrangement inequality of Hardy et al. (1952): a weighted sum under a fixed total is largest when large spillovers meet large weights and smallest when they meet small weights. The same construction is a water-filling problem. The dual multiplier $\lambda^*/n_1 = w_{(k^*)}$ is the water level, the controls above it are saturated and those below are dry, as Figure 2 shows for the upper bound. When several controls carry the same weight, the division of spillover among them is arbitrary and leaves both bounds unchanged, so the distinct-weight assumption costs nothing and the supplement records the tie-breaking.

Both bounds are sharp. Linear-programming duality certifies that no feasible spillover configuration exceeds them, and explicit data-generating processes attain each endpoint under all maintained assumptions (Proposition S11). The gap between the two bounds is the width of the identification region. It measures the uncertainty about the contamination that remains after the detection estimate, and the next theorem gives it in closed form.

Theorem 4.10 (Width of the Level 2 Identification Region)**.** *Under the conditions of Theorem S9, the width of the identification region at Level 2 is*

$$\mathcal{W}(w) \;\equiv\; \mathcal{B}_2^{\mathrm{U}}(w) - \mathcal{B}_2^{\mathrm{L}}(w) \;=\; \sum_{j=1}^{n_1} \tilde{w}_{(j)}\, \Delta_{(j)}, \tag{41}$$

where $\tilde{w}_{(j)} = w_{(j)} - \bar{w}$ and $\Delta_{(j)} = \delta^*_{(j)} - \delta^{**}_{(j)}$. This representation implies:

(i) $\mathcal{W}(w) = 0$ if and only if w_j is constant across $j \in \mathcal{C}_1$.

(ii) $\mathcal{W}(w) > 0$ *whenever at least two exposed controls receive different weights and* $0 < \psi < \bar{d}$.

(iii) $\mathcal{W}(w) \leq 2 \max_{j \in \mathcal{C}_1} |\tilde{w}_{(j)}| \cdot n_1 \min(\psi,\ \bar{d} - \psi)$.

(iv) When the unit-specific capacities are homogeneous ($d_j = d$ *for all* $j \in \mathcal{C}_1$)*, the width* $\mathcal{W}(w)$ *is a Schur-convex function of the weight vector: if* w' *majorizes* w*, then* $\mathcal{W}(w') \geq \mathcal{W}(w)$.

Equation (41) writes the width as an inner product of two mean-zero vectors, the centered weights $\tilde{w}$ and the allocation shifts $\Delta_{(j)} = \delta^*_{(j)} - \delta^{**}_{(j)}$. Only the uneven part of the weighting enters, since the constant part is annihilated by the zero-sum shifts, which makes the width a property of the estimator's weights rather than of the data alone. An estimator that treats the exposed controls alike has $\tilde{w} = 0$ and zero width however much spillover the adversary could move, and one that concentrates weight on a few controls has large width because the shifts on those controls are magnified. Part (iii) bounds the width by the largest weight deviation times the redistributable budget $2n_1 \min(\psi, \bar{d} - \psi)$, and part (iv) turns this into a Schur-convex ordering under equal capacities, so that among estimators with the same total exposed weight the more dispersed ones face the wider region.

Conditioning on covariates tightens the region further. Stratum-level detection estimands $\tau(x_k)$ replace the single mean constraint with K within-stratum constraints. These confine the adversary to within-stratum transfers and split the problem into K independent copies of the Level 2 program (supplement). The gain over Level 2 follows from Jensen's inequality applied to the concave budget-flexibility function $\min(t, \bar{d} - t)$, so strata whose spillovers are very small or very large leave little room for redistribution. These bounds are reliable only when the strata are large enough to estimate $\tau(x_k)$ well, roughly $n_k \geq 100$ to 150 exposed controls in my simulations. Below that, the researcher should coarsen the strata or report the Level 2 bounds, which need only $\hat{\psi}_{\mathrm{DR}}$.

I now collect the sharp bounds at all three levels in a single statement.

Theorem 4.11 (Sharp Bounds on the Contamination Under Interference)**.** *Let $\hat{\tau}(w)$ be a linear ATT estimator with weight vector $w \geq 0$, and let $\mathcal{C}_1$ denote the set of n_1 exposed controls with unit-specific capacities $d_j = \Delta Y_j^{\text{obs}} - a$ and mean capacity $\bar{d} = n_1^{-1} \sum d_j$. Sort the exposed controls by weight in increasing order: $w_{(1)} \leq \cdots \leq w_{(n_1)}$, with corresponding capacities $d_{(1)}, \ldots, d_{(n_1)}$.*

Under outcome support $\Delta Y \in [a, b]$ and monotone nonnegative spillovers, the contamination $\mathcal{B}(w) = \sum_{j \in \mathcal{C}_1} w_j \delta_j$ satisfies:

(I) ***Level 1*** *(support and monotonicity). $\mathcal{B}(w) \in \big[0,\ \sum_{j \in \mathcal{C}_1} w_j\, d_j\big]$.*

(II) ***Level 2*** *(adding the mean constraint). Under exposure ignorability (Assumptions 1–5) and $0 \leq \psi \leq \bar{d}$,*

$$\mathcal{B}_2^{\mathrm{U}}(w) = \sum_{j=k^*+1}^{n_1} w_{(j)}\, d_{(j)} + w_{(k^*)} \Big[n_1 \psi - C(k^*{+}1) \Big], \tag{42}$$

$$\mathcal{B}_2^{\mathrm{L}}(w) = \sum_{j=1}^{k_*-1} w_{(j)}\, d_{(j)} + w_{(k_*)} \Big[n_1 \psi - D(k_*{-}1) \Big], \tag{43}$$

where $C(k)$, $D(k)$ are the cumulative capacities and k^*, k_* are the unique threshold indices.

(III) ***Level 3*** *(adding conditional constraints). With K covariate strata and $0 \leq \tau(x_k) \leq \bar{d}_k$,*

$$\mathcal{B}_3^{\mathrm{U}}(w) = \sum_{k=1}^{K} \left[\sum_{j=k_k^*+1}^{n_k} w_{k,(j)}\, d_{k,(j)} + w_{k,(k_k^*)} \Big[n_k \tau(x_k) - C_k(k_k^*{+}1) \Big] \right], \tag{44}$$

$$\mathcal{B}_3^{\mathrm{L}}(w) = \sum_{k=1}^{K} \left[\sum_{j=1}^{k_{*,k}-1} w_{k,(j)}\, d_{k,(j)} + w_{k,(k_{*,k})} \Big[n_k \tau(x_k) - D_k(k_{*,k}{-}1) \Big] \right]. \tag{45}$$

(IV) ***Nesting.*** *$[\mathcal{B}_3^{\mathrm{L}},\ \mathcal{B}_3^{\mathrm{U}}] \subseteq [\mathcal{B}_2^{\mathrm{L}},\ \mathcal{B}_2^{\mathrm{U}}] \subseteq [0,\ \sum w_j d_j]$, with strict inclusion from Level 1 to Level 2 when $0 < \psi < \bar{d}$, and from Level 2 to Level 3 when $\tau(x_k)$ varies across strata and the weights exhibit cross-stratum asymmetry.*

(V) ***Width.*** *The Level 2 width satisfies $\mathcal{W}_2(w) = \sum_{j=1}^{n_1} \tilde{w}_{(j)}\, \Delta_{(j)}$ and equals zero if and only if all exposed controls receive equal weight.*

All bounds are sharp: attained by feasible configurations (Theorems S8–S10), certified by LP duality (Proposition S11), and achievable by explicit data-generating processes consistent with all maintained assumptions.

Theorem 4.11 is the researcher's working toolkit. For any linear estimator, compute the weights it places on the exposed controls, read the capacities d_j off the observed outcomes,

estimate ψ and, where the design supports stratum-level detection, the $\tau(x_k)$ by the doubly robust procedure, then evaluate the closed forms at the highest level the maintained assumptions allow. The computation needs no simulation or numerical optimization, since sorting and the threshold search run in $O(n_1 \log n_1)$ time, and it returns the familiar no-interference answer when $\psi = 0$. Two specializations collapse the region to a point, uniform weights at Level 2 and within-stratum homogeneity at Level 3, and both are recorded in the supplement. The ordering that survives when neither special case holds is the next result.

Corollary 4.12 (Estimator Ranking by Weight Dispersion)**.** *Under homogeneous capacities, the Level 2 width is Schur-convex in the weight vector: among estimators with the same $W_{\exp}$, those with more concentrated weights on exposed controls face the widest identification region.*

Throughout I impose nonnegative spillovers, the leading case when interference is suspected. The same sorting characterization extends to signed spillovers by widening each box to the full support $[a, b]$, at the cost of a nonzero Level 1 lower bound, an extension I take up in Section 8 and develop in the supplement.

5 Estimation and Inference

The bounds in Theorem 4.11 are population objects that depend on the detection estimand ψ, the conditional estimands $\tau(x_k)$, and the unit-specific capacities d_j. Only the capacities are observed, so the remaining inputs must be estimated. I estimate the bounds by their empirical counterparts, evaluating the closed-form expressions of Theorem 4.11 at the doubly robust estimates. The estimate $\hat{\psi}_{\mathrm{DR}}$ replaces ψ, the estimate $\hat{\tau}(x_k)$ replaces each conditional estimand $\tau(x_k)$, and the threshold index $\hat{k}^*$ is recomputed from the estimated budget as the unique solution of $C(\hat{k}^*+1) < n_1\hat{\psi}_{\mathrm{DR}} \leq C(\hat{k}^*)$. The Level 1 bounds use only observed quantities and need no estimation. For the Level 2 upper bound the substitution yields the

plug-in estimator, and the lower-bound and Level 3 estimators are formed the same way:

$$\hat{\mathcal{B}}_2^{\mathrm{U}}(w) = \sum_{j=\hat{k}^*+1}^{n_1} w_{(j)}\, d_{(j)} \;+\; w_{(\hat{k}^*)}\Big[n_1\hat{\psi}_{\mathrm{DR}} - C(\hat{k}^*+1)\Big], \tag{46}$$

As the doubly robust estimate $\hat{\psi}_{\mathrm{DR}}$ varies, the threshold index $\hat{k}^*$ can jump from one control to the next, and the bound values must not jump with it, or the plug-in estimator in (46) would be ill-behaved at the thresholds where the detector crosses a cumulative capacity.

Lemma 5.1 (Regularity of the Bound Mapping)**.** *Fix the observed quantities $(w_{(j)}, d_{(j)})_{j=1}^{n_1}$ and view the Level 2 bounds as functions of ψ alone. Then:*

(i) Both mappings $\psi \mapsto \mathcal{B}_2^{\mathrm{U}}(w;\psi)$ and $\psi \mapsto \mathcal{B}_2^{\mathrm{L}}(w;\psi)$ are piecewise linear on $[0,\bar{d}]$, with breakpoints at $\psi = C(k)/n_1$ (upper) and $\psi = D(k)/n_1$ (lower).

(ii) On each linear piece, the slope of the upper bound is $n_1\, w_{(k^*)}$ and the slope of the lower bound is $n_1\, w_{(k_*)}$.

(iii) $\mathcal{B}_2^{\mathrm{U}}(w;\cdot)$ is concave and $\mathcal{B}_2^{\mathrm{L}}(w;\cdot)$ is convex on $[0,\bar{d}]$.

(iv) Both mappings are Lipschitz with constant $L = n_1 \cdot w_{(n_1)}$:

$$\big|\mathcal{B}_2^{\mathrm{U}}(w;\psi') - \mathcal{B}_2^{\mathrm{U}}(w;\psi)\big| \le n_1\, w_{(n_1)}\, |\psi' - \psi| \quad \textit{for all } \psi, \psi' \in [0,\bar{d}]. \tag{47}$$

Within each regime the bounds therefore move at a constant rate as ψ varies, and the Lipschitz constant $n_1 w_{(n_1)}$ transfers any estimation error in $\hat{\psi}_{\mathrm{DR}}$ to the plug-in bounds in fixed proportion, so the convergence rate of the detector carries over to the bounds.

Proposition 5.2 (Consistency of Plug-in Bound Estimators)**.** *Under the conditions of Proposition 3.3 and the feasibility condition $0 < \psi < \bar{d}$:*

(i) The Level 2 plug-in estimators are consistent: $\hat{\mathcal{B}}_2^{\mathrm{U}}(w) \xrightarrow{p} \mathcal{B}_2^{\mathrm{U}}(w)$ and $\hat{\mathcal{B}}_2^{\mathrm{L}}(w) \xrightarrow{p} \mathcal{B}_2^{\mathrm{L}}(w)$.

(ii) Under additionally consistent estimation of $\mu_0(x)$ within each stratum, the Level 3 plug-in estimators are consistent.

Consistency follows from the continuous mapping theorem once the detector estimates are themselves consistent, and with the bounds now pinned to their population values the next question is their sampling distribution in finite samples. At a generic ψ, where $n_1\psi$ avoids every cumulative capacity $C(k)$, the threshold index is locally constant and the bound map is locally affine with slope $n_1 w_{(k^*)}$, so the delta method gives:

$$\sqrt{n}\big(\hat{\mathcal{B}}_2^{\mathrm{U}}(w) - \mathcal{B}_2^{\mathrm{U}}(w)\big) \xrightarrow{d} N\Big(0,\ \big(n_1\, w_{(k^*)}\big)^2\, \sigma_\psi^2\Big), \tag{48}$$

The asymptotic variance is the square of the local slope $n_1 w_{(k^*)}$ times the variance of $\hat{\psi}_{\mathrm{DR}}$. Because both bounds are smooth functions of the same scalar $\hat{\psi}_{\mathrm{DR}}$, they are jointly

normal and perfectly correlated. Under symmetric weights the two local slopes coincide, so the two standard errors are equal, consistent with the point-identification result. This delta-method standard error tracks the Monte Carlo standard deviation closely when the weights are near uniform. For difference-in-differences the ratio of the two falls in $[1.01, 1.10]$ across the configurations of Section 6, and it stays close to one for staggered designs. When the weights concentrate on a few donors, as in synthetic control, the local slope is dominated by a single weight and the delta-method standard error understates the sampling variability. I therefore restrict formal delta-method inference to difference-in-differences and staggered weights, where the calibration holds, and treat the concentrated-weight bounds as estimates of the identification region alone. The full calibration analysis is deferred to the Supplement.

Because units share interference structure only within overlapping exposure neighborhoods, the doubly robust influence function already absorbs the dominant spatial dependence through the outcome model and the exposure propensity, and the residual dependence is negligible for units more than $2r$ apart. The analytic influence-function standard errors are therefore my default inference tool, and the simulations in Section 6 confirm that they track the Monte Carlo standard deviation. A cluster bootstrap over spatially separated blocks is available as a robustness check, but it is not the default, because single-linkage clustering yields too few blocks ($G < 30$) and inflates the standard errors by a factor of two to fourteen relative to the Monte Carlo truth. The full treatment, with alternative spatial and network-robust variance estimators, is in the Supplement.

$$\mathrm{CI}_{1-\alpha} = \left[\hat{\mathcal{B}}_2^{\mathrm{L}}(w) - c_n(\alpha)\, \hat{\sigma}_{\mathrm{L}}, \quad \hat{\mathcal{B}}_2^{\mathrm{U}}(w) + c_n(\alpha)\, \hat{\sigma}_{\mathrm{U}} \right], \tag{49}$$

where $c_n(\alpha)$ solves

$$\Phi\left(c_n + \frac{\hat{\mathcal{B}}_2^{\mathrm{U}} - \hat{\mathcal{B}}_2^{\mathrm{L}}}{\max(\hat{\sigma}_{\mathrm{L}},\ \hat{\sigma}_{\mathrm{U}})} \right) - \Phi(-c_n) = 1 - \alpha. \tag{50}$$

The critical value $c_n(\alpha)$ adapts to how informative the identification region is. When the region is wide relative to sampling error, c_n approaches the one-sided quantile $z_{1-\alpha}$ and

the interval places a one-sided margin outside each bound. When the region is narrow, c_n approaches the two-sided quantile $z_{1-\alpha/2}$ and the interval approaches the usual symmetric one. Under symmetric weights the region collapses to a point and (49) reduces to the standard normal interval $\widehat{W_{\exp}\psi} \pm z_{1-\alpha/2}\hat{\sigma}$, which confirms internal consistency with the point-identified case. The construction is self-contained, requiring only that I read the plug-in bounds off Theorem 4.11, form the analytic standard errors from the delta-method variance in (48), solve (50) for $c_n(\alpha)$ by bisection, and report the interval (49). Its uniform validity over the identification region follows from Imbens and Manski (2004, Proposition 1), and Stoye (2009) shows the length-coverage tradeoff to be optimal. The same steps deliver the Level 3 interval, now with the Level 3 bounds and their influence-function covariance.

The exposure radius r sets the boundary between exposed and unexposed controls, and the researcher must choose it even though the true spillover range is rarely known (Butts, 2021; Pollmann, 2023). I therefore report the bounds as a function of r. The sensitivity curve $r \mapsto [\mathcal{B}_2^{\mathrm{L}}(w;r), \mathcal{B}_2^{\mathrm{U}}(w;r)]$ is evaluated over a grid and displayed as a band. At $r = 0$ no control is exposed and the contamination is zero. As r grows the exposed set $\mathcal{C}_1(r)$ enlarges and the region typically widens. At the smallest radius that leaves no unexposed controls, $\psi(r)$ is no longer identified and the Level 2 bounds become undefined. A breakdown radius is the smallest r at which the region for $\mathcal{B}(w)$ grows wide enough to overturn the substantive reading of the ATT, in the spirit of the breakdown frontier of Rambachan and Roth (2023) for parallel-trends violations. Its construction is given in the Supplement.

The complete pipeline takes treatment assignments, outcomes, spatial locations, and covariates and returns the contamination bounds, a confidence interval, and the sensitivity band, at a cost of $O(n_1 \log n_1)$ set by the sort in the water-filling step. It runs in under half a second per replication at $N = 2000$, and the full seven-step recipe is given in the Supplement.

6 Monte Carlo

I evaluate the framework in finite samples with a Monte Carlo study built around three predictions of the theory. The first is that the estimator's weight structure is the dominant determinant of how informative the bounds are. Difference-in-differences point-identifies the contamination in every configuration, while concentrated synthetic-control weights leave an identification region no tighter than the support-only benchmark. The width theorem separates three regimes, point identification, informative partial identification, and uninformative partial identification, and all three are already present at $N = 500$. The second is that the doubly robust detector $\hat{\psi}_{\mathrm{DR}}$ tolerates misspecification of either nuisance model, so that it stays nearly unbiased when either the outcome regression or the exposure propensity is correctly specified and fails only when both are wrong. The third is that the Imbens–Manski confidence intervals hold at least their nominal coverage in every cell of the design. Each cell uses 5,000 replications, which fixes the Monte Carlo standard error on a coverage rate at 0.0031.

6.1 Design

Each replication places N units on the unit square with covariates $X_i \sim \mathcal{N}(0, I_2)$ and assigns treatment through a logistic propensity model. A control within distance $r = 1$ of a treated unit is classified as exposed, and the remaining controls as unexposed. The clean counterfactual outcome follows $\Delta Y_j(0,0) = \mu_0(X_j) + \varepsilon_j$, and spillovers are additive and nonnegative, $\Delta Y_j(0,1) = \Delta Y_j(0,0) + \tau(X_j)$ with $\tau(X_j) \geq 0$, which satisfies the monotone-spillover restriction. For each replication I compute four weighting schemes: equal weights for difference-in-differences, sparse and dense synthetic control, and near-uniform staggered weights that mimic cohort-time aggregation. Their concentration spans two orders of mag-

nitude, from a Herfindahl index of 0.0002 for difference-in-differences to 0.028 for sparse synthetic control. Eleven configurations, labeled A through K, vary the exposure fraction, the spillover magnitude, the heterogeneity of the unit-level spillover, and the specification of the two nuisance models. The full grid and every secondary result appear in the supplement.

6.2 Progressive Tightening of the Bounds

Figure 3 plots the median width of the identification region for three representative designs, and Table 1 reports the same widths numerically at $N = 2000$. Reading across each weight scheme, the bars shorten as constraints accumulate, from the support-only Horowitz–Manski benchmark through Level 1, then the mean constraint at Level 2, then the conditional constraints at Level 3. The ranking matches the width theorem. Uniform difference-in-differences weights collapse the region to a single point at Level 2 (a 100% reduction from Level 1) because they force the contamination to equal $\rho\psi$ once ψ is known. Staggered weights are nearly uniform, so they tighten by 70.6% at the baseline without reaching a point. Sparse synthetic-control weights barely move (a 0.0% reduction at the baseline) because a few dominant donors absorb the entire mean constraint and leave the adversary free to reallocate spillover among them. For the design with heterogeneous spillovers, the conditional constraints at Level 3 tighten the staggered-weight region by a further 11.9%, and this reaches 15.5% once ten strata are used. The gain reflects the between-stratum dispersion that the budget-flexibility function rewards (Proposition S18). Under the baseline the contamination runs between 25 and 27% of the treatment effect and reaches roughly half of it under large spillovers, so netting it out is large enough to change substantive conclusions.

6.3 Coverage and Double Robustness

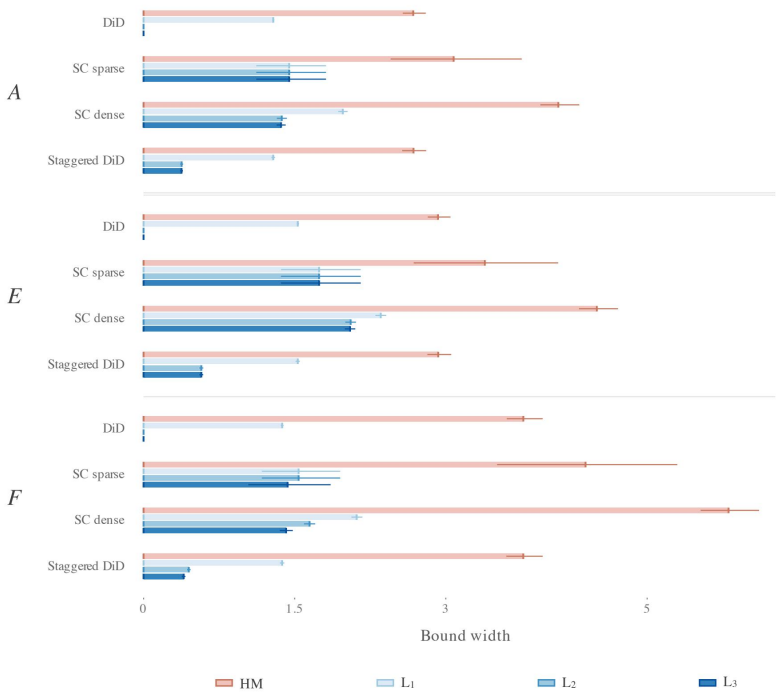
The detector behaves as Proposition 3.3 predicts. When either the outcome regression or the exposure propensity is correctly specified, $\hat{\psi}_{\mathrm{DR}}$ carries negligible bias, with root-mean-square error between 0.079 and 0.097 at $N = 2000$, and it loses $\sqrt{N}$-consistency only when both models are wrong, where the error rises to 0.243. Cross-fitting the detector roughly triples its standard deviation, from 0.094 to 0.233 at $N = 2000$, without lowering its bias, since sample splitting pays a variance cost that the linear nuisance models do not require. I therefore recommend the standard doubly robust estimator when the nuisance models are parametric and reserve cross-fitting for flexible learners such as random forests. The Imbens–Manski intervals hold at least nominal coverage in every cell, and for difference-in-differences weights the analytic influence-function standard error tracks the Monte Carlo standard deviation to within a ratio between 1.01 and 1.10. Finally, under every configuration the conventional ATT confidence interval, constructed on the no-interference assumption, fails to cover the true effect in 100% of replications for every weight scheme, because the contamination exceeds what a SUTVA interval can absorb.

Table 1: Bound width by refinement level, three representative designs.

DGP	Weight	B^{true}	B/τ	L_1	L_2	L_3	CI_width/ATT	Tight (%)
A (baseline)								
A	DiD	0.254	0.254	1.288 (0.031)	0.000 (0.000)	—	0.111	100.0
A	SC sparse	0.273	0.273	1.447 (0.673)	1.447 (0.673)	—	1.447	0.0
A	Staggered	0.254	0.254	1.287 (0.056)	0.378 (0.046)	—	0.464	70.6
E (large spillover)								
E	DiD	0.498	0.498	1.532 (0.039)	0.000 (0.000)	—	0.132	100.0
E	SC sparse	0.561	0.561	1.744 (0.797)	1.744 (0.797)	—	1.744	0.0
E	Staggered	0.497	0.497	1.530 (0.065)	0.573 (0.059)	—	0.673	62.5
F (heterogeneous spillover)								
F	DiD	0.342	0.342	1.377 (0.044)	0.000 (0.000)	0.000 (0.000)	0.137	100.0
F	SC sparse	0.348	0.348	1.511 (0.802)	1.511 (0.802)	1.383 (0.864)	1.511	0.0
F	Staggered	0.342	0.342	1.376 (0.068)	0.451 (0.056)	0.397 (0.057)	0.554	67.2

Note: Median bound widths (interquartile range in parentheses) over 5,000 replications at $N = 2000$; medians are used because widths are right-skewed. B^{true}: true contamination; B/τ: contamination relative to the ATT; L_1–L_3: identified-set widths under support only, the sorting characterization, and covariate stratification (Config F only); CI_width/ATT: Imbens–Manski interval width relative to the ATT; Tight (%): reduction from L_1 to L_2. The full grid of all configurations appears in Supplement Table S8.

Figure 3: Progressive tightening of the identification region, three representative designs.



Note: Median widths, with interquartile whiskers, over 5,000 replications at $N = 5000$, for the baseline (A), large-spillover (E), and heterogeneous-spillover (F) designs. Each weight scheme shows the support-only Horowitz–Manski benchmark and the sharp Level-k regions L_1 through L_3, widest first. Difference-in-differences collapses to a point at L_2, and sparse synthetic control barely moves. All five configurations appear in the Supplement, and the numerical values are collected in Table 1.

7 Applications

I apply the framework to two Brazilian policy evaluations, one drawn from each identification regime. The first evaluates Rio de Janeiro's Pacifying Police Units with a uniformly weighted difference-in-differences, the case in which Proposition 4.6 reduces the contamination to $\rho\hat{\psi}$ and the detection estimate alone de-contaminates the effect. The second evaluates Maricá's municipal basic income with synthetic control, where concentrated donor weights leave a genuine identification region and the sorting bounds of Section 4 give the sharpest available statement. I carry the synthetic control case through two outcomes, the level of incumbent support and the structure of local competition, so that the contamination appears as a property of the design and not of a single outcome. In all three results the doubly robust

procedure detects interference at conventional levels, so each partial identification analysis corrects a contamination the data confirm is present.

7.1 Police Pacification and the Local Vote

Between 2008 and 2014 Rio de Janeiro installed Pacifying Police Units (UPPs) in its favela communities on a staggered schedule, replacing armed control of territory with a permanent police presence. I ask whether pacification shifted the local left-bloc vote share, using 156 bairros, 55 of them treated and 101 control, in a staggered-adoption event study. Interference is immediate here, since a unit changes policing, commerce, and mobility in the blocks around it, so a control bairro bordering a pacified one is plausibly exposed even though it never received a unit of its own. Reading such neighbors as clean counterfactuals is the contamination the framework is built to correct.

Identification of the treatment effect rests on parallel trends, and the event study supports it: the pre-treatment leads are jointly small and individually insignificant, with $p = 0.84$ at lead 3 and $p = 0.51$ at lead 2, and the effect switches on only at adoption (Figure 4). The naive average effect on the treated is $\hat{\tau}(w) = +0.0397$, and the Callaway and Sant'Anna (2021) and Sun and Abraham (2021) estimators corroborate it at $+0.0377$, so pacification raises the left-bloc share by about four points. The question is whether that gap is the program's effect or partly the trace of spillover into the controls that measure it. I define exposure by first-order spatial contiguity to any treated bairro, which splits the control pool into $n_1 = 40$ exposed and $n_0 = 60$ unexposed units, so that $\rho = 0.40$.

The doubly robust estimator detects positive spillover among the exposed controls, $\hat{\psi} = +0.0133$ with a standard error of 0.0050 and $t = 2.66$. The implied contamination of the difference-in-differences estimate is $\mathcal{B}(w) = \rho\hat{\psi} = +0.0053$, and its Imbens–Manski interval, $[+0.0014, +0.0092]$, excludes zero, so pacification spills into the adjacent bairros and moves

their vote share in the same direction as the treated effect. An exact recovery of the Callaway–Sant'Anna implied control weights on this panel returns a uniform weight on every exposed control, so the symmetric-weight case of Proposition 4.6 applies and the nested regions collapse to a point. The support-only Manski baseline, $[-0.0295, +0.4269]$, and the Level 1 region under added monotonicity, $\mathcal{B}_1(w) = [0, +0.4269]$, carry no information, while the mean constraint at Level 2 pins the contamination at $\mathcal{B}_2(w) = \{+0.0053\} = \rho\hat{\psi}$ (Figure 4).

This contamination is positive, so the naive estimator is biased toward zero: the exposed controls are themselves lifted by the program, and their raised outcomes shrink the measured gap. De-contaminating recovers the SUTVA target, $\tau^{\text{SUTVA}}(w) = \hat{\tau}(w) + \mathcal{B}_2(w) = +0.0450$, so about 13% of the naive four-point effect is spillover that standard difference-in-differences nets out silently. I report $+0.0450$ under the paper's simplex convention, which treats the exposed-control weight as projecting in full onto the differenced outcome. On this panel the Callaway–Sant'Anna estimator places control mass 0.69 on the ΔY path it differences, so the bias it actually inherits is $0.69\,\mathcal{B}_2(w) = 0.0037$ and the estimator-faithful headline is $+0.0434$, which I confirm by injecting the estimated spillover into the estimator directly. Sign, order of magnitude, and the substantive reading hold under either convention. Pacification matters more once the spillover in the controls is netted out, and because the weights are uniform the correction is a single number rather than a range, so the detection estimate carries the entire identification content.

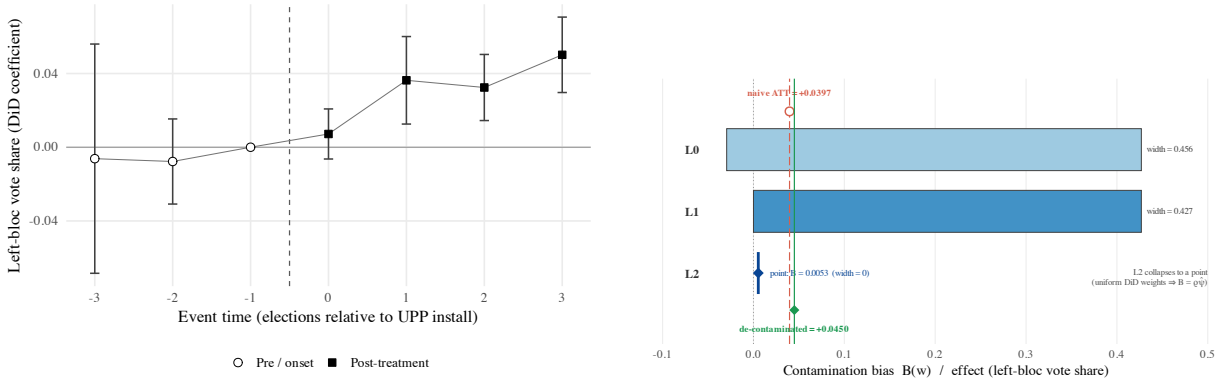
7.2 A Municipal Basic Income and Its Neighbors

In 2013 the municipality of Maricá launched the Renda Básica de Cidadania, an unconditional basic income paid in a local currency called the Mumbuca and financed by a sovereign fund of offshore oil royalties. I ask how the program affected the incumbent party's vote share. A single treated unit and a long panel make synthetic control the natural estimator.

Interference is at least as plausible here as in the pacification study, because the program turned Maricá into a regional economic engine, with oil-funded public employment, the circulating Mumbuca, expanded commerce, and metropolitan commuting, so its benefits reach the neighboring municipalities that make up the donor pool. A donor that absorbs those spillovers is a contaminated counterfactual.

Synthetic control identification rests on close pre-treatment fit, and that fit holds across local, state, and metropolitan donor pools (Figure 5). Incumbent-party share shows a large post-treatment gap, $\hat{\tau}(w) = +0.59$, a rise of 59 percentage points. Spillover is detected in the local donor pool, where exposure is defined by proximity within roughly 75 kilometers of Maricá. The doubly robust estimate is $\hat{\psi} = +0.089$, with a standard error of 0.042 and $t = 2.12$, over 21 exposed and 65 unexposed donors, so $\rho = 0.24$, and the Imbens–Manski interval $[0, +0.568]$ signals genuine spillover. The synthetic control weights concentrate on a handful of donors, so Proposition 4.7 applies and the contamination is only partially identified. The regions tighten steadily as information accrues, from the Manski baseline $[0, +0.733]$ to the Level 1 region $\mathcal{B}_1(w) = [0, +0.458]$ and the Level 2 region $\mathcal{B}_2(w) = [0, +0.387]$, yet Level 2 stays an interval rather than the singleton of the difference-in-differences case (Figure 5). De-contaminating gives $\tau^{\text{SUTVA}}(w) = [+0.59, +0.98]$, so the true incumbent effect is at least the naive estimate and possibly two-thirds larger, because the exposed donors are themselves lifted by the program and depress the synthetic counterfactual. The sign of the correction is settled even though its magnitude is not.

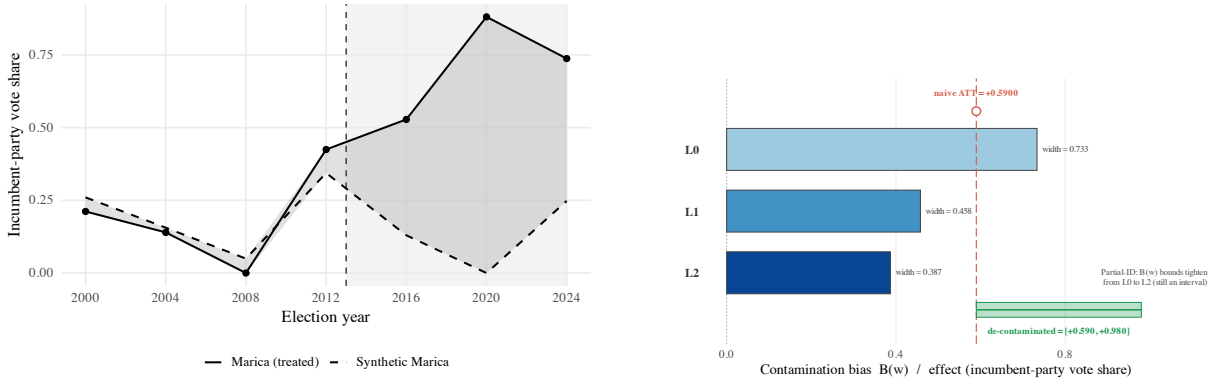
The same Maricá design contaminates a second outcome with a different structure. Vote concentration is the Herfindahl share of the leading bloc and measures how lopsided local competition is. It too shows a large gap with good pre-treatment fit, $\hat{\tau}(w) = +0.268$. At a 100-kilometer exposure radius the doubly robust estimate is $\hat{\psi} = +0.059$, with a standard error of 0.020 and $t = 3.0$, over 40 exposed and 46 unexposed donors, so $\rho = 0.47$, and



(a) Event-study estimates.

(b) Nested contamination regions.

Figure 4: UPP application under difference-in-differences. Panel (a) reports the staggered-adoption event study: the pre-treatment leads are jointly small and individually insignificant, and the effect switches on at adoption, supporting parallel trends. Panel (b) shows the nested contamination regions, where the support-only Manski baseline and the Level 1 region under added monotonicity are wide and the Level 2 mean constraint collapses to the point $\rho\hat{\psi} = +0.0053$, point-identifying the contamination under uniform weights.



(a) Synthetic control fit.

(b) Nested contamination regions.

Figure 5: Maricá application under synthetic control, incumbent-party share. Panel (a) shows the close pre-treatment fit and the large post-treatment gap ($\hat{\tau}(w) = +0.59$). Panel (b) shows the nested contamination regions tightening from the Manski baseline through Level 1 to Level 2, where under concentrated donor weights $\mathcal{B}_2(w) = [0, +0.387]$ remains an interval rather than a point.

the Imbens–Manski interval [0, +0.440] again signals spillover. Here the mean constraint adds nothing to the monotonicity restriction, so the Level 1 and Level 2 regions coincide at $\mathcal{B}_1(w) = \mathcal{B}_2(w) = [0, +0.440]$ inside the Manski baseline [0, +0.615], and de-contaminating gives $\tau^{\text{SUTVA}}(w) = [+0.268, +0.708]$, again signed upward. The program raises incumbent

Table 2: Detection and partial identification across the three results.

	UPP (DiD)	**Maricá (SC)**	
	point-identified	partially identified	
Outcome	left-bloc share	incumbent share	concentration
Naive effect $\hat{\tau}(w)$	+0.0397	+0.59	+0.268
Exposure	contiguity (1-ring)	proximity (∼75 km)	proximity (∼100 km)
(n_1, n_0)	(40, 60)	(21, 65)	(40, 46)
ρ	0.40	0.24	0.47
$\hat{\psi}$ (SE)	+0.0133 (0.0050)	+0.089 (0.042)	+0.059 (0.020)
t	2.66	2.12	3.0
IM CI for $\mathcal{B}(w)$	[+0.0014, +0.0092]	[0, +0.568]	[0, +0.440]
$\mathcal{B}_2(w)$	$\{+0.0053\} = \rho\hat{\psi}$	[0, +0.387]	[0, +0.440]
De-contaminated $\tau^{\mathrm{SUTVA}}(w)$	+0.0450	[+0.59, +0.98]	[+0.268, +0.708]

Note: $\hat{\psi}$ is the doubly robust spillover estimate with its Imbens–Manski (IM) interval; $\mathcal{B}_2(w)$ is the Level 2 contamination region; $\tau^{\mathrm{SUTVA}}(w) = \hat{\tau}(w) + \mathcal{B}_2(w)$. UPP point-identifies; both Maricá outcomes partially identify. For Maricá concentration the mean constraint does not bind, so $\mathcal{B}_2(w) = \mathcal{B}_1(w)$.

support and also concentrates competition around the incumbent bloc, and both effects are partially identified for the same reason. The full vote-concentration analysis, along with a replication for the broader left bloc that reproduces the pattern at $t = 4.2$, is in the supplement.

7.3 Two Regimes, One Diagnostic

Table 2 collects the three results, and the contrast in its last two rows is the framework's central claim made concrete. UPP and the two Maricá outcomes run through the same machinery of an exposure mapping, a doubly robust detection step, and the sorting bounds, and all three reject no-interference. They differ in the weight structure of the estimator, and that difference decides the identification regime: the uniform difference-in-differences weights point-identify the contamination at $\rho\hat{\psi}$ and collapse $\mathcal{B}_2(w)$ to a single number, while the concentrated synthetic control weights behind both Maricá outcomes leave $\mathcal{B}_2(w)$ a nondegenerate interval. The weight structure alone, and not the outcome, decides whether

the correction is a point or a range. The same diagnostic therefore delivers a sharp correction in one design and a sharp interval in the other, and it recovers in real policy evaluations the two identification regimes the theory predicts.

8 Discussion

These results make the dispersion of an estimator's control weights a criterion for choosing among panel estimators when interference is a concern. Estimators are usually chosen on efficiency, on robustness to violations of parallel trends, and on their handling of treatment-effect heterogeneity. Weight dispersion is a further consideration, and the decisive one under interference, because it governs whether the bias can be recovered at all. An estimator that spreads weight evenly across the controls, as standard difference-in-differences does, point-identifies its contamination once the average spillover ψ is known, since the bias equals $\rho\,\psi$ for ρ the exposure fraction and the detection estimate then removes it. An estimator that concentrates weight on a few controls, as synthetic control does and as staggered-adoption and synthetic difference-in-differences estimators do to a lesser degree, leaves the same contamination only partially identified, with an interval whose width grows with the dispersion of the weights among the exposed controls (Theorem 4.10, Corollary 4.12). The choice of estimator therefore bears on how informative the contamination bounds are, alongside its familiar consequences for efficiency and bias under SUTVA. An estimator selected for pre-treatment fit or for its handling of staggered timing can pay for that choice in a wider identification region, while the uniform weights of difference-in-differences deliver a sharper reading of the contamination even where the estimator is otherwise less attractive.

Before the framework developed here, a researcher who suspected interference was contaminating an estimate had three responses available: assume the spillover away under

SUTVA, retreat to a narrower comparison the design can still defend, or abandon the design. I add a fourth. The researcher keeps the estimator the data support and reports the usual point estimate together with a closed-form interval that says how much interference could be moving it, given the observed data and the single detection estimate ψ. The width of that interval is itself a property of the estimator's weights. It measures how much interference the reported point estimate can conceal, and that amount is small when the weights are even and large when they concentrate on a few donors.

Several limitations warrant mention. Exposure ignorability is maintained throughout, and the exposure-radius sensitivity analysis probes the exposure mapping but not ignorability itself, for which no formal sensitivity analysis is offered here. The monotone nonnegative spillover assumption restricts the analysis to interference of a single sign. The two-sided case leaves the sorting characterization intact, with counter-monotone pairing of weights and spillovers supplying the lower bound, and I develop it in the Supplement, not the main text. When a single donor dominates the weight vector, as in synthetic control, the delta-method standard errors are uninformative, so under concentrated weights the bounds are best read as point estimates of the identification region rather than as the basis for confidence intervals (Section 5).

The same limitations point to the natural next steps. A formal sensitivity analysis for exposure ignorability, built on an odds-ratio bound for the association between exposure and the potential outcomes, would complement the exposure-radius analysis, and if it were joined to the parallel-trends sensitivity framework of Rambachan and Roth (2023), a single analysis could carry both threats to a panel estimate at once. Valid inference under the concentrated weights of synthetic control is harder, since it calls for the joint asymptotics of the detection estimate and the estimated capacities that the delta method does not provide, and I leave it, together with data-driven stratum construction at Level 3, to future work.

References

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

Abadie, A., Diamond, A. and Hainmueller, J. (2010), 'Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program', *Journal of the American Statistical Association* **105**(490), 493–505.

Abadie, A. and Imbens, G. W. (2006), 'Large sample properties of matching estimators for average treatment effects', *Econometrica* **74**(1), 235–267.

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

Aronow, P. M. and Samii, C. (2017), 'Estimating average causal effects under general interference, with application to a social network experiment', *Annals of Applied Statistics* **11**(4), 1912–1947.

Bonvini, M. and Kennedy, E. H. (2022), 'Sensitivity analysis via the proportion of unmeasured confounding', *Journal of the American Statistical Association* **117**(539), 1540–1550.

Borusyak, K., Jaravel, X. and Spiess, J. (2024), 'Revisiting event-study designs: Robust and efficient estimation', *Review of Economic Studies* **91**(6), 3253–3285.

Butts, K. (2021), 'Difference-in-differences estimation with spatial spillovers', *arXiv preprint arXiv:2105.03737* .

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

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. and Robins, J. (2018), 'Double/debiased machine learning for treatment and structural parameters', *Econometrics Journal* **21**(1), C1–C68.

de Chaisemartin, C. and D'Haultfœuille, X. (2020), 'Two-way fixed effects estimators with heterogeneous treatment effects', *American Economic Review* **110**(9), 2964–2996.

Di Stefano, R. and Mellace, G. (2024), 'The inclusive synthetic control method', *arXiv preprint arXiv:2403.17624* . Revised November 2024.

Hardy, G. H., Littlewood, J. E. and Pólya, G. (1952), *Inequalities*, 2nd edn, Cambridge University Press, Cambridge.

Hirano, K., Imbens, G. W. and Ridder, G. (2003), 'Efficient estimation of average treatment effects using the estimated propensity score', *Econometrica* **71**(4), 1161–1189.

Holland, P. W. (1986), 'Statistics and causal inference', *Journal of the American Statistical Association* **81**(396), 945–960.

Horowitz, J. L. and Manski, C. F. (1995), 'Identification and robustness with contaminated and corrupted data', *Econometrica* **63**(2), 281–302.

Imbens, G. W. and Manski, C. F. (2004), 'Confidence intervals for partially identified parameters', *Econometrica* **72**(6), 1845–1857.

Kennedy, E. H. (2023), Semiparametric doubly robust targeted double machine learning: A review, *in* 'Handbook of Matching and Weighting Adjustments for Causal Inference', Chapman and Hall/CRC. arXiv:2203.06469.

Lee, D. S. (2009), 'Training, wages, and sample selection: Estimating sharp bounds on treatment effects', *Review of Economic Studies* **76**(3), 1071–1102.

Leung, M. P. (2022), 'Causal inference under approximate neighborhood interference', *Econometrica* **90**(1), 267–293.

Manski, C. F. (1990), 'Nonparametric bounds on treatment effects', *American Economic Review* **80**(2), 319–323.

Mealli, F. and Viviens, J. (2025), 'Difference-in-differences in the presence of unknown interference', *arXiv preprint arXiv:2512.21176* .

Molinari, F. (2020), Microeconometrics with partial identification, *in* S. N. Durlauf, L. P. Hansen, J. J. Heckman and R. L. Matzkin, eds, 'Handbook of Econometrics', Vol. 7, Elsevier, pp. 355–486.

Pollmann, M. (2023), 'Causal inference for spatial treatments', *arXiv preprint arXiv:2011.00373* . Revised 2023.

Rambachan, A. and Roth, J. (2023), 'A more credible approach to parallel trends', *Review of Economic Studies* **90**(5), 2555–2591.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994), 'Estimation of regression coefficients when some regressors are not always observed', *Journal of the American Statistical Association* **89**(427), 846–866.

Rubin, D. B. (1980), 'Randomization analysis of experimental data: The Fisher randomization test comment', *Journal of the American Statistical Association* **75**(371), 591–593.

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

Stoye, J. (2009), 'More on confidence intervals for partially identified parameters', *Econometrica* **77**(4), 1299–1315.

Sun, L. and Abraham, S. (2021), 'Estimating dynamic treatment effects in event studies with heterogeneous treatment effects', *Journal of Econometrics* **225**(2), 175–199.

Xu, R. (2023), 'Difference-in-differences with interference', *arXiv preprint arXiv:2306.12003* .