Supplementary Material to *Sharp Closed-Form Bounds for Interference Contamination in Linear ATT Designs*

Contents

S1 Theoretical Infrastructure

This section specializes the umbrella results of the main text to the canonical linear ATT estimators and records the point-identification corollaries. Throughout the supplement I write results of the main text by their main-text number (for example, Theorem 4.9 of the main text); results and equations that carry an S prefix belong to this supplement.

S1.1 Specialization to Canonical Estimators

Corollary S1 (Difference-in-Differences)**.** *Under uniform weights $w_j = |\mathcal{C}|^{-1}$, $W_{\exp} = \rho$ and $\mathcal{B}^{\mathrm{DiD}} = \rho \cdot n_1^{-1} \sum_{j \in \mathcal{C}_1} [\Delta Y_j(0,1) - \Delta Y_j(0,0)]$, the exposure fraction times the unweighted average spillover.*

Corollary S2 (Synthetic Control)**.** *Under data-driven weights concentrating on a subset of donors, $\mathcal{B}^{\mathrm{SC}} = \sum_{j \in \mathcal{C}_1} w_j [\Delta Y_j(0,1) - \Delta Y_j(0,0)]$. When donors receiving large weights are disproportionately exposed, $W_{\exp}$ can approach unity and $\mathcal{B}^{SC}$ can substantially exceed $\mathcal{B}^{DiD}$.*

Corollary S3 (Staggered Adoption)**.** *Under cohort-time aggregation weights (Callaway and Sant'Anna, 2021; Sun and Abraham, 2021), $\mathcal{B}^{\mathrm{Stag}} = \sum_{j \in \mathcal{C}_1} w_j [\Delta Y_j(0,1) - \Delta Y_j(0,0)]$, where the weight structure inherits the specific aggregation scheme and the exposure set $\mathcal{C}_1$ evolves as treatment diffuses across cohorts.*

S1.2 Specializations of the Umbrella Theorem

Corollary S4 (Point Identification Under Symmetric Weights)**.** *If $w_j = \bar{w}$ for all $j \in \mathcal{C}_1$, the Level 2 region collapses to $\mathcal{B}(w) = W_{\exp} \cdot \psi$. Standard DiD satisfies this condition, yielding $\mathcal{B}^{\mathrm{DiD}} = \rho \cdot \psi$.*

Corollary S5 (Point Identification Under Conditional Homogeneity)**.** *If $\delta_j = \tau(x_k)$ for all $j \in S_k$ and each k, the Level 3 region collapses to $\mathcal{B}(w) = \sum_{k=1}^{K} W_k \, \tau(x_k)$ for any weight vector w.*

S2 Doubly Robust Construction

This section gives the construction behind the doubly robust estimator of Section 3 of the main text. The main text records the two nuisance models and the estimator $\hat{\psi}_{\mathrm{DR}}$ in its equation (22). Here I write the exposure propensity in product form, record the exposed-control regression and the conditional detector that supply the Level 3 constraints, and record the influence-function representation that underlies the double robustness and the semiparametric efficiency of Proposition 3.3 of the main text.

S2.1 Nuisance Functions in Product Form

Two nuisance models underpin the estimator. The outcome regression among unexposed controls, $\mu_0(X)$ in equation (20) of the main text, is the conditional mean outcome change

among controls not exposed to nearby treatment, and by Theorem 3.2(iv) of the main text it also equals the conditional mean of the missing counterfactual for exposed controls at the same covariate value. The exposure propensity, π_j in equation (21) of the main text, is the probability that control j is exposed given the covariates of the units in its exposure neighborhood $\mathcal{N}_j = \{k \in \mathcal{U} : d_{jk} < r, k \neq j\}$. Under Assumption 3 of the main text this propensity is assembled from the individual treatment propensities $e(X_k) = \mathbb{P}(A_k = 1 \mid X_k)$ through

$$\pi_j = 1 - \prod_{k \in \mathcal{N}_j} (1 - e(X_k)), \tag{S1}$$

since $E_j = 1$ if and only if at least one unit in $\mathcal{N}_j$ is treated and, conditional on covariates, the treatments are independent.

I also use the outcome regression among exposed controls, $\mu_1(X) = \mathbb{E}[\Delta Y_j \mid E_j = 1, X_j = X, A_j = 0]$. Under exposure ignorability the conditional spillover effect is the contrast $\tau(X) = \mu_1(X) - \mu_0(X)$, so its plug-in estimate $\hat{\tau}(X) = \hat{\mu}_1(X) - \hat{\mu}_0(X)$ provides the conditional spillover estimates that enter the finest level of the partial identification analysis as the Level 3 constraints of the main text.

S2.2 Influence Function and Semiparametric Efficiency

The estimator $\hat{\psi}_{\mathrm{DR}}$ of equation (22) of the main text is the empirical analog of the efficient influence function for ψ under the nonparametric model of Assumptions 1 through 5 of the main text, following the semiparametric efficiency theory of Robins et al. (1994), Bang and Robins (2005), and Kennedy (2023). Its estimation error splits into the influence function evaluated at the true nuisance parameters plus a remainder that is the product of the two nuisance errors, $R_n(\hat{\mu}_0, \hat{\pi}) = O_p(\|\hat{\mu}_0 - \mu_0\| \cdot \|\hat{\pi} - \pi\|)$. This product structure is the source of the double robustness of Proposition 3.3 of the main text. Consistency of either nuisance model alone drives the remainder to zero, and when both estimators converge at $o_p(n^{-1/4})$ rates the remainder is $o_p(n^{-1/2})$, so $\hat{\psi}_{\mathrm{DR}}$ inherits the influence-function representation that attains the semiparametric efficiency bound. I give the full decomposition and the two consistency conditions in the proof of Proposition 3.3 of the main text, which appears in Section S3. See Kennedy (2023) and Chernozhukov et al. (2018) for the general theory.

S3 Proofs

Statements referenced in the proofs. The proofs in this section invoke the results below, and I restate each by its main-text number and name so that the section can be read on its own. The maintained assumptions are Assumption 1 (Exposure Ignorability), Assumption 5 (Covariate Overlap), and Assumption 6 (Feasibility) of the main text. The feasible sets are Definition 4.1 and Definition 4.3 of the main text, the Level 1 and Level 3 feasible sets. The identification results are Theorem 3.2 (Identification of the Spillover Mean and the Counterfactual Constraint) and Proposition 3.3 (Double Robustness) of the main text. The bounding results are Lemma 4.4 (Nesting of Feasible Sets), Corollary 4.5 (Nesting of Identification Regions), Proposition 4.6 (Point Identification Under Symmetric Weights), Proposition 4.7 (Partial Identification Under Asymmetric Weights), Theorem 4.9 (Sorting

Structure of the Level 2 Solution), and Theorem 4.10 (Width of the Level 2 Identification Region) of the main text. The estimation results are Lemma 5.1 (Regularity of the Bound Mapping) and Proposition 5.2 (Consistency of Plug-in Bound Estimators) of the main text. The bounding linear programs and the cumulative-capacity functions are those of the main text, and the proofs refer to them by the equation numbers under which they appear there.

S3.1 Preparatory and Component Results

Lemma S6 (KKT Conditions for the Level 2 Maximization)**.** *A feasible vector $\delta^* \in \mathcal{F}_2^\delta$ solves* (35) *if and only if there exist $\lambda^* \in \mathbb{R}$, $\mu_j^{*\mathrm{L}} \geq 0$, and $\mu_j^{*\mathrm{U}} \geq 0$ for each $j \in \mathcal{C}_1$ satisfying:*

(a) Stationarity: $w_j - \dfrac{\lambda^*}{n_1} + \mu_j^{*\mathrm{L}} - \mu_j^{*\mathrm{U}} = 0$ *for all* $j \in \mathcal{C}_1$.

(b) Complementary slackness: $\mu_j^{*\mathrm{L}} \cdot \delta_j^* = 0$ *and* $\mu_j^{*\mathrm{U}} \cdot (d_j - \delta_j^*) = 0$ *for all* $j \in \mathcal{C}_1$.

(c) Primal feasibility: $0 \leq \delta_j^* \leq d_j$ *for all* j, $(1/n_1)\sum_j \delta_j^* = \psi$.

(d) Dual feasibility: $\mu_j^{*\mathrm{L}} \geq 0$, $\mu_j^{*\mathrm{U}} \geq 0$ *for all* $j \in \mathcal{C}_1$.

Full proof of Lemma S6. Form the Lagrangian of the upper-bound program (35), with multiplier $\lambda \in \mathbb{R}$ for the mean constraint and $\mu_j^{\mathrm{L}}, \mu_j^{\mathrm{U}} \geq 0$ for the box constraints $\delta_j \geq 0$ and $\delta_j \leq d_j$:

$$\mathcal{L}(\delta, \lambda, \mu^{\mathrm{L}}, \mu^{\mathrm{U}}) = \sum_{j \in \mathcal{C}_1} w_j \delta_j + \lambda \left(\psi - \frac{1}{n_1} \sum_{j \in \mathcal{C}_1} \delta_j \right) + \sum_{j \in \mathcal{C}_1} \mu_j^{\mathrm{L}} \delta_j + \sum_{j \in \mathcal{C}_1} \mu_j^{\mathrm{U}} (d_j - \delta_j). \tag{S2}$$

Differentiating the Lagrangian (S2) with respect to δ_j term by term:

$$\frac{\partial}{\partial \delta_j} \left[\sum_{j \in \mathcal{C}_1} w_j \delta_j \right] = w_j, \tag{S3}$$

$$\frac{\partial}{\partial \delta_j} \left[\lambda \left(\psi - \frac{1}{n_1} \sum_{j \in \mathcal{C}_1} \delta_j \right) \right] = -\frac{\lambda}{n_1}, \tag{S4}$$

$$\frac{\partial}{\partial \delta_j} \left[\sum_{j \in \mathcal{C}_1} \mu_j^{\mathrm{L}} \delta_j \right] = \mu_j^{\mathrm{L}}, \tag{S5}$$

$$\frac{\partial}{\partial \delta_j} \left[\sum_{j \in \mathcal{C}_1} \mu_j^{\mathrm{U}} (d_j - \delta_j) \right] = -\mu_j^{\mathrm{U}}. \tag{S6}$$

Summing (S3)–(S6):

$$\frac{\partial \mathcal{L}}{\partial \delta_j} = w_j - \frac{\lambda}{n_1} + \mu_j^{\mathrm{L}} - \mu_j^{\mathrm{U}}. \tag{S7}$$

Setting (S7) equal to zero at the optimum $(\delta^*, \lambda^*, \mu^{*\mathrm{L}}, \mu^{*\mathrm{U}})$ yields condition (a).

For complementary slackness (b): the constraint $\delta_j \geq 0$ can be written as $g_j^{\mathrm{L}}(\delta) = -\delta_j \leq 0$ with multiplier $\mu_j^{*\mathrm{L}} \geq 0$. The KKT complementarity condition requires $\mu_j^{*\mathrm{L}} \cdot g_j^{\mathrm{L}}(\delta^*) = 0$, that is:

$$\mu_j^{*\mathrm{L}} \cdot (-\delta_j^*) = 0 \quad \Longleftrightarrow \quad \mu_j^{*\mathrm{L}} \cdot \delta_j^* = 0. \tag{S8}$$

Similarly, the constraint $\delta_j \leq d_j$ can be written as $g_j^{\mathrm{U}}(\delta) = \delta_j - d_j \leq 0$ with multiplier $\mu_j^{*\mathrm{U}} \geq 0$. Complementarity requires:

$$\mu_j^{*\mathrm{U}} \cdot (\delta_j^* - d_j) = 0 \quad \Longleftrightarrow \quad \mu_j^{*\mathrm{U}} \cdot (d_j - \delta_j^*) = 0. \tag{S9}$$

Conditions (c) and (d) restate primal and dual feasibility directly. That the KKT conditions are necessary follows from the linear independence constraint qualification, which is satisfied automatically when all constraints are linear (Bertsekas, 1999, Proposition 3.3.1). That they are sufficient follows from the linearity of both the objective and the constraints: for linear programs, every KKT point is a global optimum (Bertsimas and Tsitsiklis, 1997, Theorem 4.3). □

Proposition S7 (Optimal Structure of the Level 2 Solution)**.** *Let δ^* solve* (35) *with associated multiplier λ^*. Then for each $j \in \mathcal{C}_1$:*

$$\delta_j^* = \begin{cases} d_j & \text{if } w_j > \lambda^*/n_1, \\ 0 & \text{if } w_j < \lambda^*/n_1, \\ \text{any value in } [0, d_j] & \text{if } w_j = \lambda^*/n_1. \end{cases} \tag{S10}$$

Full proof of Proposition S7. Fix $j \in \mathcal{C}_1$. We treat each case determined by the sign of $w_j - \lambda^*/n_1$.

Consider first the case $w_j > \lambda^*/n_1$. Rearranging stationarity condition (a) to isolate the multiplier difference:

$$\mu_j^{*\mathrm{U}} - \mu_j^{*\mathrm{L}} = w_j - \frac{\lambda^*}{n_1}. \tag{S11}$$

Since $w_j > \lambda^*/n_1$ by assumption, the right-hand side of (S11) is strictly positive:

$$\mu_j^{*\mathrm{U}} - \mu_j^{*\mathrm{L}} > 0. \tag{S12}$$

By dual feasibility (d), $\mu_j^{*\mathrm{L}} \geq 0$. Combining with (S12):

$$\mu_j^{*\mathrm{U}} > \mu_j^{*\mathrm{L}} \geq 0, \tag{S13}$$

which in particular implies:

$$\mu_j^{*\mathrm{U}} > 0. \tag{S14}$$

Applying the complementary slackness condition (b) for the capacity constraint to (S14):

$$\mu_j^{*\mathrm{U}} > 0 \implies d_j - \delta_j^* = 0, \tag{S15}$$

which gives:

$$\delta_j^* = d_j. \tag{S16}$$

This establishes that units with weight above the threshold are assigned their maximum capacity.

Now consider the case $w_j < \lambda^*/n_1$. Rearranging stationarity (a) in the opposite direction:

$$\mu_j^{*\mathrm{L}} - \mu_j^{*\mathrm{U}} = \frac{\lambda^*}{n_1} - w_j. \tag{S17}$$

Since $\lambda^*/n_1 > w_j$ by assumption:

$$\mu_j^{*\mathrm{L}} - \mu_j^{*\mathrm{U}} > 0. \tag{S18}$$

By dual feasibility (d), $\mu_j^{*\mathrm{U}} \geq 0$. Combining with (S18):

$$\mu_j^{*\mathrm{L}} > \mu_j^{*\mathrm{U}} \geq 0, \tag{S19}$$

so:

$$\mu_j^{*\mathrm{L}} > 0. \tag{S20}$$

Applying complementary slackness (b) for the nonnegativity constraint to (S20):

$$\mu_j^{*\mathrm{L}} > 0 \implies \delta_j^* = 0. \tag{S21}$$

This establishes that units with weight below the threshold are assigned zero spillover.

Finally, consider the case $w_j = \lambda^*/n_1$. From stationarity (a):

$$\mu_j^{*\mathrm{U}} - \mu_j^{*\mathrm{L}} = w_j - \frac{\lambda^*}{n_1} = 0, \tag{S22}$$

so:

$$\mu_j^{*\mathrm{U}} = \mu_j^{*\mathrm{L}}. \tag{S23}$$

Denote this common value by $c \geq 0$. Suppose $c > 0$. Then $\mu_j^{*\mathrm{L}} = c > 0$, and complementary slackness (b) for the lower bound requires:

$$\mu_j^{*\mathrm{L}} > 0 \implies \delta_j^* = 0. \tag{S24}$$

Simultaneously, $\mu_j^{*\mathrm{U}} = c > 0$, and complementary slackness (b) for the capacity constraint requires:

$$\mu_j^{*\mathrm{U}} > 0 \implies d_j - \delta_j^* = 0 \implies \delta_j^* = d_j. \tag{S25}$$

Equations (S24) and (S25) together require $0 = d_j$, contradicting $d_j > 0$. Therefore:

$$c = 0 \quad \implies \quad \mu_j^{*\mathrm{U}} = \mu_j^{*\mathrm{L}} = 0. \tag{S26}$$

With both multipliers equal to zero, the complementary slackness conditions reduce to $0 \cdot \delta_j^* = 0$ and $0 \cdot (d_j - \delta_j^*) = 0$, which are satisfied for any $\delta_j^* \in [0, d_j]$:

$$\delta_j^* \in [0, d_j] \quad \text{(unrestricted within the box)}. \tag{S27}$$

This establishes that units at exactly the threshold weight have their value determined by the mean constraint rather than by the optimality conditions for that unit alone. □

Theorem S8 (Level 1 Sharp Bounds)**.** *Under outcome support $\Delta Y \in [a, b]$ and monotone nonnegative spillovers, the sharp bounds on the contamination are*

$$\mathcal{B}_1^{\mathrm{L}}(w) = 0, \qquad \mathcal{B}_1^{\mathrm{U}}(w) = \sum_{j \in \mathcal{C}_1} w_j \, d_j, \tag{S28}$$

where $d_j = \Delta Y_j^{\mathrm{obs}} - a$ is the unit-specific capacity of exposed control j. The width is $\mathrm{Width}_1(w) = \sum_{j \in \mathcal{C}_1} w_j \, d_j$*, and both bounds are attained by explicit configurations in $\mathcal{F}_1^{\delta}$.*

Proof. The proof proceeds in three steps: deriving the bounds, proving they are achieved, and verifying that no tighter bounds are possible.

Step 1: Deriving the bounds. The contamination is $\mathcal{B}(w) = \sum_{j \in \mathcal{C}_1} w_j \delta_j$, where each $\delta_j \in [0, d_j]$ under $\mathcal{F}_1^{\delta}$ and the weights satisfy $w_j \geq 0$ for all $j \in \mathcal{C}_1$.

For the lower bound, observe that each term $w_j \delta_j$ in the sum is nonnegative, since both $w_j \geq 0$ and $\delta_j \geq 0$. A sum of nonnegative terms is nonnegative, so $\mathcal{B}(w) = \sum w_j \delta_j \geq 0$ for every feasible (δ_j).

For the upper bound, observe that $\delta_j \leq d_j$ for each j. Since $w_j \geq 0$, multiplying both sides of the inequality by w_j preserves the direction: $w_j \delta_j \leq w_j d_j$. Summing over all $j \in \mathcal{C}_1$:

$$\mathcal{B}(w) = \sum_{j \in \mathcal{C}_1} w_j \delta_j \leq \sum_{j \in \mathcal{C}_1} w_j d_j.$$

Combining: $0 \leq \mathcal{B}(w) \leq \sum_{j \in \mathcal{C}_1} w_j d_j$ for every $(\delta_j) \in \mathcal{F}_1^{\delta}$.

Step 2: Achieving the bounds (attainment). We exhibit feasible configurations that attain each bound.

For the lower bound, set $\delta_j^{\mathrm{L}} = 0$ for all $j \in \mathcal{C}_1$. This configuration satisfies $\delta_j^{\mathrm{L}} \in [0, d_j]$ for every j (since $0 \leq d_j$ by the fact that $\Delta Y_j^{\mathrm{obs}} \geq a$ under outcome support). The resulting contamination is $\mathcal{B}(w) = \sum w_j \cdot 0 = 0$, which equals $\mathcal{B}_1^{\mathrm{L}}(w)$.

This configuration corresponds to zero spillover for every exposed control: each unit's clean counterfactual equals its observed outcome, $\Delta Y_j(0, 0) = \Delta Y_j^{\mathrm{obs}}$, meaning exposure has no effect on any control unit. This is the no-interference scenario.

For the upper bound, set $\delta_j^{\mathrm{U}} = d_j$ for all $j \in \mathcal{C}_1$. This configuration satisfies $\delta_j^{\mathrm{U}} \in [0, d_j]$ trivially. The resulting contamination is $\mathcal{B}(w) = \sum w_j d_j$, which equals $\mathcal{B}_1^{\mathrm{U}}(w)$.

This configuration corresponds to maximal spillover for every exposed control: each unit's clean counterfactual is pushed to the lower support bound, $\Delta Y_j(0, 0) = a$, meaning the entire gap between the observed outcome and the lower support is attributable to spillover. This is the worst-case scenario in which interference has the largest possible effect on every exposed control.

Step 3: Sharpness. Since both bounds are attained by feasible configurations (Step 2), no tighter bounds are possible over $\mathcal{F}_1^{\delta}$. Any proposed lower bound $\ell > 0$ is violated by the configuration $\delta_j^{\mathrm{L}} = 0$, which produces $\mathcal{B}(w) = 0 < \ell$. Any proposed upper bound $u < \sum w_j d_j$ is violated by the configuration $\delta_j^{\mathrm{U}} = d_j$, which produces $\mathcal{B}(w) = \sum w_j d_j > u$. Therefore the bounds in (S28) are the tightest possible under $\mathcal{F}_1^{\delta}$. □

Theorem S9 (Sharp Level 2 Bounds on the Contamination)**.** *Let $\hat{\tau}(w)$ be a linear ATT estimator with weight vector w. Sort the exposed controls by weight in increasing order, $w_{(1)} < w_{(2)} < \cdots < w_{(n_1)}$, with corresponding capacities $d_{(1)}, \ldots, d_{(n_1)}$. Under outcome support $\Delta Y \in [a, b]$, monotone nonnegative spillovers, exposure ignorability (Assumptions 1–5), and the feasibility condition $0 \leq \psi \leq \bar{d}$:*

(I) The sharp upper bound on the contamination is

$$\mathcal{B}_2^{\mathrm{U}}(w) = \sum_{j=k^*+1}^{n_1} w_{(j)}\, d_{(j)} \;+\; w_{(k^*)}\Big[n_1\psi - C(k^*{+}1)\Big], \tag{S29}$$

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

(II) The sharp lower bound on the contamination is

$$\mathcal{B}_2^{\mathrm{L}}(w) = \sum_{j=1}^{k_*-1} w_{(j)}\, d_{(j)} \;+\; w_{(k_*)}\Big[n_1\psi - D(k_*{-}1)\Big], \tag{S30}$$

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

(III) The sharp identification region for the contamination at Level 2 is

$$\mathcal{B}(w) \in \left[\mathcal{B}_2^{\mathrm{L}}(w),\;\; \mathcal{B}_2^{\mathrm{U}}(w)\right]. \tag{S31}$$

Both bounds are attained: there exist feasible configurations in $\mathcal{F}_2^{\delta}$ achieving $\mathcal{B}_2^{\mathrm{U}}(w)$ and $\mathcal{B}_2^{\mathrm{L}}(w)$ exactly.

Full proof of Theorem S9. We derive parts (I) and (II) by direct computation from the optimal solutions in Theorem 4.9, then establish the attainment claim.

For the upper bound, the contamination at the optimal solution δ^* from Theorem 4.9(I) is

$$\mathcal{B}_2^{\mathrm{U}}(w) = \sum_{j=1}^{n_1} w_{(j)}\, \delta^*_{(j)}. \tag{S32}$$

Partitioning the sum into the three groups defined by the threshold k^* and substituting the values of $\delta^*_{(j)}$ from (39):

$$\mathcal{B}_2^{\mathrm{U}}(w) = \underbrace{\sum_{j=1}^{k^*-1} w_{(j)} \cdot 0}_{=\,0} \;+\; w_{(k^*)}\, \delta^*_{(k^*)} \;+\; \sum_{j=k^*+1}^{n_1} w_{(j)}\, d_{(j)}. \tag{S33}$$

The first sum vanishes because $\delta^*_{(j)} = 0$ for all $j < k^*$. Substituting the value of the threshold unit from (S118), $\delta^*_{(k^*)} = n_1\psi - C(k^*{+}1)$:

$$\mathcal{B}_2^{\mathrm{U}}(w) = w_{(k^*)}\Big[n_1\psi - C(k^*{+}1)\Big] \;+\; \sum_{j=k^*+1}^{n_1} w_{(j)}\, d_{(j)}. \tag{S34}$$

Rearranging into the form stated in (S29):

$$\mathcal{B}_2^{\mathrm{U}}(w) = \sum_{j=k^*+1}^{n_1} w_{(j)}\, d_{(j)} \;+\; w_{(k^*)}\Big[n_1\psi - C(k^*{+}1)\Big]. \tag{S35}$$

This establishes part (I).

We now derive an equivalent representation that separates the threshold-weight contribution from the weight premium. Expanding the threshold unit's term in (S34):

$$w_{(k^*)}\Big[n_1\psi - C(k^*{+}1)\Big] = w_{(k^*)}\, n_1\psi \;-\; w_{(k^*)}\, C(k^*{+}1). \tag{S36}$$

Substituting the definition $C(k^*{+}1) = \sum_{j=k^*+1}^{n_1} d_{(j)}$:

$$w_{(k^*)}\, C(k^*{+}1) = w_{(k^*)} \sum_{j=k^*+1}^{n_1} d_{(j)} = \sum_{j=k^*+1}^{n_1} w_{(k^*)}\, d_{(j)}. \tag{S37}$$

Combining (S34), (S36), and (S37):

$$\begin{aligned}\mathcal{B}_2^{\mathrm{U}}(w) &= w_{(k^*)}\, n_1\psi \;-\; \sum_{j=k^*+1}^{n_1} w_{(k^*)}\, d_{(j)} \;+\; \sum_{j=k^*+1}^{n_1} w_{(j)}\, d_{(j)} \\ &= w_{(k^*)}\, n_1\psi \;+\; \sum_{j=k^*+1}^{n_1} \Big(w_{(j)} - w_{(k^*)}\Big)\, d_{(j)}. \end{aligned} \tag{S38}$$

In this representation, the first term $w_{(k^*)}\, n_1\psi$ is the contamination that would obtain if every unit at or above the threshold received spillover at the threshold weight $w_{(k^*)}$. The second term $\sum_{j>k^*}(w_{(j)} - w_{(k^*)})\, d_{(j)}$ is the additional contamination arising from the fact that units above the threshold have weights strictly exceeding $w_{(k^*)}$, such that each such unit contributes a premium of $(w_{(j)} - w_{(k^*)})\, d_{(j)} > 0$. This decomposition will be used in the width analysis that follows.

For the lower bound, the contamination at the optimal solution δ^{**} from Theorem 4.9(II) is

$$\mathcal{B}_2^{\mathrm{L}}(w) = \sum_{j=1}^{n_1} w_{(j)}\, \delta^{**}_{(j)}. \tag{S39}$$

Partitioning into the three groups defined by k_* and substituting the values from (40):

$$\mathcal{B}_2^{\mathrm{L}}(w) = \sum_{j=1}^{k_*-1} w_{(j)}\, d_{(j)} \;+\; w_{(k_*)}\, \delta^{**}_{(k_*)} \;+\; \underbrace{\sum_{j=k_*+1}^{n_1} w_{(j)} \cdot 0}_{=\,0}. \tag{S40}$$

The last sum vanishes because $\delta^{**}_{(j)} = 0$ for all $j > k_*$. Substituting the threshold unit's value from (S145), $\delta^{**}_{(k_*)} = n_1\psi - D(k_*{-}1)$:

$$\mathcal{B}_2^{\mathrm{L}}(w) = \sum_{j=1}^{k_*-1} w_{(j)}\, d_{(j)} \;+\; w_{(k_*)}\Big[n_1\psi - D(k_*{-}1)\Big]. \tag{S41}$$

This establishes part (II).

Deriving the equivalent representation for the lower bound by the same sequence of operations. Expanding the threshold term in (S41):

$$w_{(k_*)}\Big[n_1\psi - D(k_*-1)\Big] = w_{(k_*)}\, n_1\psi \;-\; w_{(k_*)}\, D(k_*-1). \tag{S42}$$

Substituting the definition $D(k_*-1) = \sum_{j=1}^{k_*-1} d_{(j)}$ and distributing:

$$w_{(k_*)}\, D(k_*-1) = w_{(k_*)} \sum_{j=1}^{k_*-1} d_{(j)} = \sum_{j=1}^{k_*-1} w_{(k_*)}\, d_{(j)}. \tag{S43}$$

Combining (S41), (S42), and (S43):

$$\begin{aligned}
\mathcal{B}_2^{\mathrm{L}}(w) &= \sum_{j=1}^{k_*-1} w_{(j)}\, d_{(j)} \;+\; w_{(k_*)}\, n_1\psi \;-\; \sum_{j=1}^{k_*-1} w_{(k_*)}\, d_{(j)} \\
&= w_{(k_*)}\, n_1\psi \;+\; \sum_{j=1}^{k_*-1} \Big(w_{(j)} - w_{(k_*)}\Big)\, d_{(j)} \\
&= w_{(k_*)}\, n_1\psi \;-\; \sum_{j=1}^{k_*-1} \Big(w_{(k_*)} - w_{(j)}\Big)\, d_{(j)}.
\end{aligned} \tag{S44}$$

The first term is again a threshold-weight contribution, and the second (now subtracted) captures the discount from the fact that low-weight units have weights strictly below $w_{(k_*)}$.

For part (III), the identification region $[\mathcal{B}_2^{\mathrm{L}}(w), \mathcal{B}_2^{\mathrm{U}}(w)]$ is the range of contamination values achievable over $\mathcal{F}_2^{\delta}$, which is an interval because the objective $\sum w_j\delta_j$ is continuous on the connected compact set $\mathcal{F}_2^{\delta}$ (the intersection of a box with a hyperplane).

The attainment claim follows directly from Theorem 4.9: the optimal solutions δ^* and δ^{**} constructed there are feasible elements of $\mathcal{F}_2^{\delta}$ (verified in the proof of Theorem 4.9) that achieve $\mathcal{B}_2^{\mathrm{U}}(w)$ and $\mathcal{B}_2^{\mathrm{L}}(w)$ respectively. □

Theorem S10 (Sharp Level 3 Bounds on the Contamination)**.** *Under the conditions of Theorem S9, with K covariate strata and within-stratum feasibility $0 \le \tau(x_k) \le \bar{d}_k$ for each k:*

(I) The sharp upper bound is

$$\mathcal{B}_3^{\mathrm{U}}(w) = \sum_{k=1}^{K} \left[\sum_{j=k_k^*+1}^{n_k} w_{k,(j)}\, d_{k,(j)} \;+\; w_{k,(k_k^*)}\Big[n_k\tau(x_k) - C_k(k_k^*+1)\Big] \right]. \tag{S45}$$

(II) The sharp lower bound is

$$\mathcal{B}_3^{\mathrm{L}}(w) = \sum_{k=1}^{K} \left[\sum_{j=1}^{k_{*,k}-1} w_{k,(j)}\, d_{k,(j)} \;+\; w_{k,(k_{*,k})}\Big[n_k\tau(x_k) - D_k(k_{*,k}-1)\Big] \right]. \tag{S46}$$

(III) The sharp identification region at Level 3 is $\mathcal{B}(w) \in [\mathcal{B}_3^{\mathrm{L}}(w),\ \mathcal{B}_3^{\mathrm{U}}(w)]$.

Both bounds are attained.

Full proof of Theorem S10. By Proposition S16, the Level 3 upper bound decomposes as:

$$\mathcal{B}_3^{\mathrm{U}}(w) = \sum_{k=1}^{K} \mathcal{B}_{3,k}^{\mathrm{U}}(w), \tag{S47}$$

where each $\mathcal{B}_{3,k}^{\mathrm{U}}(w)$ solves the stratum sub-problem (S76). This sub-problem is a bounded-variable LP with n_k decision variables, box constraints $\delta_j \in [0, d_j]$ for $j \in S_k$, and a single equality constraint $(1/n_k)\sum_{j\in S_k} \delta_j = \tau(x_k)$. This is identical in structure to the Level 2 program (35), with n_1 replaced by n_k, ψ replaced by $\tau(x_k)$, and the index set restricted to S_k.

Applying Theorem S9(I) within stratum k: sort the units in S_k by weight, $w_{k,(1)} < \cdots < w_{k,(n_k)}$, with capacities $d_{k,(1)}, \ldots, d_{k,(n_k)}$. The stratum-level upper bound is:

$$\mathcal{B}_{3,k}^{\mathrm{U}}(w) = \sum_{j=k_k^*+1}^{n_k} w_{k,(j)}\, d_{k,(j)} \ + \ w_{k,(k_k^*)}\Big[n_k\tau(x_k) - C_k(k_k^*+1)\Big], \tag{S48}$$

where k_k^* is the within-stratum threshold index satisfying $C_k(k_k^*+1) < n_k\tau(x_k) \le C_k(k_k^*)$, the stratum-level analog of the Level 2 threshold condition. This exists and is unique by the argument in Theorem 4.9, applied to the strictly decreasing sequence $C_k(1) > C_k(2) > \cdots > C_k(n_k+1) = 0$ with $C_k(1) = n_k\bar{d}_k \ge n_k\tau(x_k) \ge 0 = C_k(n_k+1)$ (by the within-stratum feasibility condition).

Summing (S48) across $k = 1, \ldots, K$:

$$\mathcal{B}_3^{\mathrm{U}}(w) = \sum_{k=1}^{K} \left[\sum_{j=k_k^*+1}^{n_k} w_{k,(j)}\, d_{k,(j)} \ + \ w_{k,(k_k^*)}\Big[n_k\tau(x_k) - C_k(k_k^*+1)\Big] \right]. \tag{S49}$$

This is exactly (S45), establishing part (I). The derivation of part (II) is identical, applying Theorem S9(II) within each stratum to obtain the stratum-level lower bounds and summing.

For attainment: within each stratum k, the sorting solution from Theorem 4.9 provides a feasible $\delta^{*,(k)} \in \mathcal{F}_{2,k}^{\delta}$ achieving $\mathcal{B}_{3,k}^{\mathrm{U}}(w)$. Concatenating these across strata, define $\delta^* \in \mathbb{R}^{n_1}$ by $\delta_j^* = \delta_j^{*,(k)}$ for $j \in S_k$. This vector is feasible in $\mathcal{F}_3^{\delta}$:

$$\delta_j^* \in [0, d_j] \ \ \forall\, j \quad (\text{since } \delta_j^{*,(k)} \in [0, d_j] \ \ \forall\, j \in S_k), \tag{S50}$$

$$\frac{1}{n_k}\sum_{j\in S_k} \delta_j^* = \frac{1}{n_k}\sum_{j\in S_k} \delta_j^{*,(k)} = \tau(x_k) \quad \text{for each } k \quad (\text{since } \delta^{*,(k)} \in \mathcal{F}_{2,k}^{\delta}). \tag{S51}$$

The objective value at δ^* is:

$$\sum_{j\in\mathcal{C}_1} w_j\delta_j^* = \sum_{k=1}^{K}\sum_{j\in S_k} w_j\delta_j^{*,(k)} = \sum_{k=1}^{K} \mathcal{B}_{3,k}^{\mathrm{U}}(w) = \mathcal{B}_3^{\mathrm{U}}(w). \tag{S52}$$

The attaining configuration is therefore feasible and achieves the bound. The argument for the lower bound is identical with $\delta^{**,(k)}$ replacing $\delta^{*,(k)}$. □

Proposition S11 (Sharpness of the Level 2 Bounds)**.** *The bounds* $\mathcal{B}_2^{\mathrm{U}}(w)$ *and* $\mathcal{B}_2^{\mathrm{L}}(w)$ *in Theorem S9 are sharp:*

(i) ***Duality certificate.*** *The optimal value of the dual linear program associated with* (35) *equals* $\mathcal{B}_2^{\mathrm{U}}(w)$*, and analogously for the lower bound.*

(ii) ***Attainment.*** *The configurations* δ^* *and* δ^{**} *from Theorem 4.9 are feasible elements of* $\mathcal{F}_2^{\delta}$ *achieving the bounds.*

(iii) ***Constructive DGP.*** *There exists a joint distribution of potential outcomes satisfying outcome support, monotone nonnegative spillovers, and the mean spillover constraint, under which* $\mathcal{B}(w) = \mathcal{B}_2^{\mathrm{U}}(w)$*. An analogous DGP exists for the lower bound.*

Full proof of Proposition S11, Part (i): LP Duality. We derive and solve the dual of the upper bound program (35). The argument for the lower bound is analogous.

The primal is a bounded-variable linear program:

$$\max_{\delta} \sum_{j \in \mathcal{C}_1} w_j \, \delta_j \quad \text{s.t.} \quad \frac{1}{n_1} \sum_{j \in \mathcal{C}_1} \delta_j = \psi, \quad 0 \le \delta_j \le d_j \ \ \forall \, j. \tag{S53}$$

Using the Lagrangian (S2) with multipliers $\lambda \in \mathbb{R}$, $\mu_j^{\mathrm{L}} \ge 0$, $\mu_j^{\mathrm{U}} \ge 0$, the dual function is obtained by maximizing the Lagrangian over the unconstrained δ. Collecting terms in (S2) by δ_j:

$$\mathcal{L} = \sum_{j \in \mathcal{C}_1} \left(w_j - \tfrac{\lambda}{n_1} + \mu_j^{\mathrm{L}} - \mu_j^{\mathrm{U}} \right) \delta_j \ + \ \lambda \psi \ + \ \sum_{j \in \mathcal{C}_1} \mu_j^{\mathrm{U}} \, d_j. \tag{S54}$$

Maximizing over $\delta_j \in \mathbb{R}$ (unconstrained, since the box constraints are absorbed into the Lagrangian): if the coefficient of δ_j is nonzero, the supremum is $+\infty$. The dual function is therefore finite only when every coefficient vanishes:

$$g(\lambda, \mu^{\mathrm{L}}, \mu^{\mathrm{U}}) = \begin{cases} \lambda \psi + \displaystyle\sum_{j \in \mathcal{C}_1} \mu_j^{\mathrm{U}} \, d_j & \text{if } w_j - \tfrac{\lambda}{n_1} + \mu_j^{\mathrm{L}} - \mu_j^{\mathrm{U}} = 0 \ \ \forall \, j, \\ +\infty & \text{otherwise.} \end{cases} \tag{S55}$$

The dual problem minimizes g over $\lambda \in \mathbb{R}$, $\mu_j^{\mathrm{L}} \ge 0$, $\mu_j^{\mathrm{U}} \ge 0$:

$$\min_{\lambda, \, \mu^{\mathrm{L}}, \, \mu^{\mathrm{U}}} \ \lambda \psi + \sum_{j \in \mathcal{C}_1} \mu_j^{\mathrm{U}} \, d_j \quad \text{s.t.} \quad w_j - \tfrac{\lambda}{n_1} + \mu_j^{\mathrm{L}} - \mu_j^{\mathrm{U}} = 0 \ \ \forall \, j, \quad \mu_j^{\mathrm{L}} \ge 0, \ \mu_j^{\mathrm{U}} \ge 0 \ \ \forall \, j. \tag{S56}$$

We now eliminate μ_j^{L} and μ_j^{U} by solving the stationarity constraint. From the equality constraint in (S56), solving for μ_j^{U}:

$$\mu_j^{\mathrm{U}} = w_j - \frac{\lambda}{n_1} + \mu_j^{\mathrm{L}}. \tag{S57}$$

Since $\mu_j^{\mathrm{U}} \ge 0$ and $\mu_j^{\mathrm{L}} \ge 0$, and since μ_j^{L} does not appear in the dual objective (only μ_j^{U} appears, multiplied by $d_j > 0$), the minimization over μ_j^{L} sets it as small as possible. When

$w_j - \lambda/n_1 \geq 0$, setting $\mu_j^{\mathrm{L}} = 0$ yields $\mu_j^{\mathrm{U}} = w_j - \lambda/n_1 \geq 0$, which is feasible. When $w_j - \lambda/n_1 < 0$, we need $\mu_j^{\mathrm{L}} \geq \lambda/n_1 - w_j > 0$ to ensure $\mu_j^{\mathrm{U}} \geq 0$; setting $\mu_j^{\mathrm{L}} = \lambda/n_1 - w_j$ gives $\mu_j^{\mathrm{U}} = 0$, minimizing the objective. In both cases:

$$\mu_j^{\mathrm{U}} = \max\big(w_j - \tfrac{\lambda}{n_1},\ 0\big) = \big(w_j - \tfrac{\lambda}{n_1}\big)^+, \tag{S58}$$

$$\mu_j^{\mathrm{L}} = \max\big(\tfrac{\lambda}{n_1} - w_j,\ 0\big) = \big(\tfrac{\lambda}{n_1} - w_j\big)^+, \tag{S59}$$

where $(x)^+ = \max(x, 0)$ denotes the positive part. Substituting (S58) into the dual objective (S56), the dual reduces to a one-dimensional minimization over the single scalar λ:

$$\min_{\lambda \in \mathbb{R}} \ \lambda\, \psi \ + \sum_{j \in \mathcal{C}_1} d_j \left(w_j - \tfrac{\lambda}{n_1}\right)^+. \tag{S60}$$

This is the dual of our bounded-variable LP, reduced from $2n_1 + 1$ dual variables to a single variable λ. The objective in (S60) is convex in λ (as a sum of a linear function and positive-part functions, which are convex), so its minimum is attained.

We now evaluate the dual at $\lambda^* = n_1\, w_{(k^*)}$, the optimal primal multiplier from (S129). At this value, $\lambda^*/n_1 = w_{(k^*)}$, and the positive parts become:

$$\big(w_{(j)} - w_{(k^*)}\big)^+ = \begin{cases} w_{(j)} - w_{(k^*)} & \text{if } j > k^* \ \ (\text{since } w_{(j)} > w_{(k^*)}), \\ 0 & \text{if } j \leq k^* \ \ (\text{since } w_{(j)} \leq w_{(k^*)}). \end{cases} \tag{S61}$$

Substituting (S61) into the reduced dual objective (S60):

$$\begin{aligned} g(\lambda^*) &= n_1\, w_{(k^*)}\, \psi \ + \sum_{j \in \mathcal{C}_1} d_{(j)} \left(w_{(j)} - w_{(k^*)}\right)^+ \\ &= n_1\, w_{(k^*)}\, \psi \ + \sum_{j=1}^{k^*} d_{(j)} \cdot 0 \ + \sum_{j=k^*+1}^{n_1} d_{(j)} \left(w_{(j)} - w_{(k^*)}\right) \\ &= w_{(k^*)}\, n_1 \psi \ + \sum_{j=k^*+1}^{n_1} \left(w_{(j)} - w_{(k^*)}\right) d_{(j)}. \end{aligned} \tag{S62}$$

Comparing (S62) to the decomposed primal optimal value (S38):

$$\mathcal{B}_2^{\mathrm{U}}(w) = w_{(k^*)}\, n_1 \psi + \sum_{j=k^*+1}^{n_1} \left(w_{(j)} - w_{(k^*)}\right) d_{(j)}, \tag{S63}$$

we see that $g(\lambda^*) = \mathcal{B}_2^{\mathrm{U}}(w)$. This establishes that the dual objective at λ^* equals the primal optimal value.

By weak duality, for any dual-feasible $(\lambda, \mu^{\mathrm{L}}, \mu^{\mathrm{U}})$ and any primal-feasible δ:

$$\sum_{j \in \mathcal{C}_1} w_j\, \delta_j \ \leq \ g(\lambda, \mu^{\mathrm{L}}, \mu^{\mathrm{U}}). \tag{S64}$$

Since the primal is feasible (Assumption 6) and bounded ($\mathcal{F}_2^\delta$ is compact), strong duality holds for linear programs (Bertsimas and Tsitsiklis, 1997, Theorem 4.4):

$$\max_{\delta \in \mathcal{F}_2^\delta} \sum_j w_j \delta_j \;=\; \min_{\lambda \in \mathbb{R}} \; \Big[\lambda \psi + \sum_j d_j \big(w_j - \tfrac{\lambda}{n_1}\big)^+ \Big]. \tag{S65}$$

The left-hand side is $\mathcal{B}_2^{\mathrm{U}}(w)$ by definition. The right-hand side is at most $g(\lambda^*) = \mathcal{B}_2^{\mathrm{U}}(w)$ (since λ^* is dual-feasible). Combined with weak duality (the minimum is at least as large as the maximum), equality holds:

$$\mathcal{B}_2^{\mathrm{U}}(w) = g(\lambda^*) = \min_{\lambda} \; g(\lambda). \tag{S66}$$

This is the sharpness certificate: the dual optimal value certifies that no primal-feasible δ can achieve an objective value exceeding $\mathcal{B}_2^{\mathrm{U}}(w)$. Any proposed bound $\tilde{\mathcal{B}} < \mathcal{B}_2^{\mathrm{U}}(w)$ would violate the weak duality inequality (S64) at the dual solution $(\lambda^*, \mu^{*\mathrm{L}}, \mu^{*\mathrm{U}})$.

The same argument applied to the minimization program (36) yields a dual certificate for $\mathcal{B}_2^{\mathrm{L}}(w)$. □

Full proof of Proposition S11, Part (iii): Constructive DGP. We construct a data-generating process for the upper bound; the lower bound is analogous with δ^{**} replacing δ^*.

Define the potential outcomes for each exposed control $j \in \mathcal{C}_1$ as:

$$\Delta Y_j(0,1) = \Delta Y_j^{\mathrm{obs}}, \tag{S67}$$

$$\Delta Y_j(0,0) = \Delta Y_j^{\mathrm{obs}} - \delta_j^*, \tag{S68}$$

where δ_j^* is the optimal allocation from Theorem 4.9(I). We verify that this DGP satisfies each model assumption.

The outcome support restriction requires $\Delta Y_j(0,0) \in [a, b]$. From (S68) and the fact that $\delta_j^* \in [0, d_j]$ (established in Theorem 4.9):

$$\begin{aligned} \Delta Y_j(0,0) &= \Delta Y_j^{\mathrm{obs}} - \delta_j^* \\ &\geq \Delta Y_j^{\mathrm{obs}} - d_j \\ &= \Delta Y_j^{\mathrm{obs}} - (\Delta Y_j^{\mathrm{obs}} - a) \\ &= a. \end{aligned} \tag{S69}$$

For the upper support bound:

$$\begin{aligned} \Delta Y_j(0,0) &= \Delta Y_j^{\mathrm{obs}} - \delta_j^* \\ &\leq \Delta Y_j^{\mathrm{obs}} - 0 \\ &= \Delta Y_j^{\mathrm{obs}} \\ &\leq b. \end{aligned} \tag{S70}$$

Combining (S69) and (S70):

$$\Delta Y_j(0,0) \in [a,b] \quad \text{for all } j \in \mathcal{C}_1. \tag{S71}$$

This confirms the outcome support restriction.

Monotone nonnegative spillovers require $\Delta Y_j(0,1) \geq \Delta Y_j(0,0)$. From (S67) and (S68):

$$\begin{aligned} \Delta Y_j(0,1) - \Delta Y_j(0,0) &= \Delta Y_j^{\text{obs}} - (\Delta Y_j^{\text{obs}} - \delta_j^*) \\ &= \delta_j^* \\ &\geq 0, \end{aligned} \tag{S72}$$

since $\delta_j^* \geq 0$ for all $j \in \mathcal{C}_1$ by the box constraints in $\mathcal{F}_2^\delta$. This confirms monotonicity.

The mean spillover constraint requires $(1/n_1)\sum_{j\in\mathcal{C}_1}[\Delta Y_j(0,1) - \Delta Y_j(0,0)] = \psi$. Using (S72):

$$\begin{aligned} \frac{1}{n_1}\sum_{j\in\mathcal{C}_1}\big[\Delta Y_j(0,1) - \Delta Y_j(0,0)\big] &= \frac{1}{n_1}\sum_{j\in\mathcal{C}_1}\delta_j^* \\ &= \psi, \end{aligned} \tag{S73}$$

where the last equality holds because $\delta^* \in \mathcal{F}_2^\delta$ satisfies the mean constraint by construction (Theorem 4.9).

Finally, the contamination under this DGP equals the upper bound. Substituting the constructed potential outcomes into $\mathcal{B}(w) = \sum_{j\in\mathcal{C}_1} w_j[\Delta Y_j(0,1) - \Delta Y_j(0,0)]$:

$$\begin{aligned} \mathcal{B}(w) &= \sum_{j\in\mathcal{C}_1} w_j\,\delta_j^* \\ &= \mathcal{B}_2^{\text{U}}(w), \end{aligned} \tag{S74}$$

by the definition of $\mathcal{B}_2^{\text{U}}(w)$ as the optimal value of (35) achieved at δ^*.

The construction for the lower bound is identical with δ^{**} from Theorem 4.9(II) replacing δ^*: set $\Delta Y_j(0,0) = \Delta Y_j^{\text{obs}} - \delta_j^{**}$, and the same verification establishes support, monotonicity, mean constraint, and $\mathcal{B}(w) = \mathcal{B}_2^{\text{L}}(w)$. □

Corollary S12 (Recovery of Point Identification Under Symmetric Weights)**.** *When all exposed controls receive equal weight $w_{(j)} = \bar{w}$, both bounds reduce to $\mathcal{B}_2^{\text{U}}(w) = \mathcal{B}_2^{\text{L}}(w) = W_{\text{exp}} \cdot \psi$, recovering Proposition 4.6.*

Proof. Under uniform weights, $\sum_{j=k^*+1}^{n_1} \bar{w}\, d_{(j)} + \bar{w}[n_1\psi - C(k^*+1)] = \bar{w}\,[C(k^*+1) + n_1\psi - C(k^*+1)] = \bar{w}\, n_1\psi = W_{\text{exp}} \cdot \psi$. The same calculation applies to the lower bound. □

Proposition S13 (Tightening Under Conditional Constraints)**.** *The identification region for $\mathcal{B}(w)$ at Level 3 is weakly contained in the identification region at Level 2:*

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

with strict inclusion whenever the conditional detection estimand $\tau(x)$ varies across strata and the weight vector assigns unequal weights to exposed controls in different strata.

Proof. The weak inclusion follows immediately from Corollary 4.5, since $\mathcal{F}_3 \subseteq \mathcal{F}_2$ (Lemma 4.4).

For strict inclusion, it suffices to show that the adversary's maximum under $\mathcal{F}_2$ exceeds the maximum under $\mathcal{F}_3$. At Level 2, the adversary can redistribute counterfactual mass freely across the entire exposed control group, subject to the global mean M. At Level 3, the adversary faces K separate mean constraints, one per stratum. Consider two strata S_1 and S_2 with $\tau(x_1) \neq \tau(x_2)$, and suppose there exist exposed controls $j_H \in S_1$ and $j_L \in S_2$ with $w_{j_H} > w_{j_L}$.

Under $\mathcal{F}_2$, the adversary can reduce c_{j_H} by ϵ and increase c_{j_L} by ϵ (a cross-stratum transfer preserving the global mean). Under $\mathcal{F}_3$, this transfer is infeasible: it changes the mean of stratum S_1 by $-\epsilon/n_1$ (now below its required value $\bar{\Delta Y}^{\text{obs}}_{S_1} - \tau(x_1)$) and the mean of stratum S_2 by $+\epsilon/n_2$ (now above its required value). The cross-stratum transfer is blocked by the stratum constraints. The adversary under $\mathcal{F}_3$ is therefore more constrained than under $\mathcal{F}_2$, and the achievable range of $\sum w_j c_j$ is weakly narrower.

To verify strict narrowing, observe that the cross-stratum transfer described above changes the weighted sum by $\epsilon(w_{j_L} - w_{j_H}) \neq 0$ under $\mathcal{F}_2$ (since $w_{j_H} \neq w_{j_L}$). This transfer is feasible under $\mathcal{F}_2$ but infeasible under $\mathcal{F}_3$. Hence the maximum of $\sum w_j c_j$ over $\mathcal{F}_2$ is achieved at a point not in $\mathcal{F}_3$, and the maximum over $\mathcal{F}_3$ is strictly smaller. By symmetry, the minimum over $\mathcal{F}_3$ is strictly larger. The inclusion is strict. □

Corollary S14 (Point Identification Under Conditional Homogeneity)**.** *Suppose that, within each covariate stratum S_k, the counterfactual outcome changes are homogeneous: $\Delta Y_j(0,0) = \Delta Y_{j'}(0,0)$ for all $j, j' \in S_k$ and each $k = 1, \ldots, K$. Then $\mathcal{F}_3$ is a singleton, and $\mathcal{B}(w)$ is point-identified for any weight vector w.*

Proof. Under conditional homogeneity, all exposed controls within stratum S_k share the same clean counterfactual value, say c_k^*. The stratum mean constraint requires $n_k^{-1} \sum_{j \in S_k} c_j = \bar{\Delta Y}^{\text{obs}}_{S_k} - \tau(x_k)$. Since all c_j in the stratum are equal to c_k^*, the constraint becomes $c_k^* = \bar{\Delta Y}^{\text{obs}}_{S_k} - \tau(x_k)$. This uniquely determines c_k^* for each stratum. Therefore the counterfactual vector is uniquely determined: $c_j = \bar{\Delta Y}^{\text{obs}}_{S_k} - \tau(x_k)$ for all $j \in S_k$ and each k. The feasible set $\mathcal{F}_3$ contains exactly one element.

Since the feasible set is a singleton, the optimization problems $\max_{(c_j) \in \mathcal{F}_3} \sum w_j c_j$ and $\min_{(c_j) \in \mathcal{F}_3} \sum w_j c_j$ yield the same value:

$$\sum_{j \in \mathcal{C}_1} w_j c_j = \sum_{k=1}^{K} \sum_{j \in S_k} w_j \left(\bar{\Delta Y}^{\text{obs}}_{S_k} - \tau(x_k) \right) = \sum_{k=1}^{K} W_k \left(\bar{\Delta Y}^{\text{obs}}_{S_k} - \tau(x_k) \right),$$

where $W_k = \sum_{j \in S_k} w_j$ is the total weight on stratum k. The contamination is therefore

$$\mathcal{B}(w) = S_{\text{obs}}(w) - \sum_{k=1}^{K} W_k \left(\bar{\Delta Y}^{\text{obs}}_{S_k} - \tau(x_k) \right) = \sum_{k=1}^{K} W_k \, \tau(x_k),$$

a known constant depending only on the estimator weights aggregated by stratum and the identified conditional detection estimands. This holds for any weight vector w, regardless of its symmetry or asymmetry. □

Proposition S15 (Qualitative Characterization of the Identification Region)**.** *Let $\hat{\tau}(w)$ be a linear ATT estimator with weight vector w, and let $\mathcal{F}_1 \supseteq \mathcal{F}_2 \supseteq \mathcal{F}_3$ be the feasible sets from Definitions 4.1–4.3. The identification regions for $\mathcal{B}(w)$ satisfy:*

(i) $[\mathcal{B}_3^{\mathrm{L}}, \mathcal{B}_3^{\mathrm{U}}] \subseteq [\mathcal{B}_2^{\mathrm{L}}, \mathcal{B}_2^{\mathrm{U}}] \subseteq [\mathcal{B}_1^{\mathrm{L}}, \mathcal{B}_1^{\mathrm{U}}]$ *(progressive tightening).*

(ii) At Level 2, the width $\mathcal{B}_2^{\mathrm{U}}(w) - \mathcal{B}_2^{\mathrm{L}}(w) = 0$ *if and only if* w_j *is constant across* $j \in \mathcal{C}_1$ *(symmetric weights).*

(iii) At Level 3, under conditional homogeneity within each covariate stratum, $\mathcal{B}_3^{\mathrm{U}}(w) - \mathcal{B}_3^{\mathrm{L}}(w) = 0$ *for any weight vector* w*.*

(iv) The width at each level is non-increasing in the number of constraints: adding stratum-level constraints weakly reduces the width, with strict reduction when the additional constraints are non-redundant and the weight vector exhibits cross-stratum asymmetry.

Proof. Part (i) is Corollary 4.5. Part (ii) combines Propositions 4.6 and 4.7. Part (iii) is Corollary S14. Part (iv) follows from $\mathcal{F}_3 \subseteq \mathcal{F}_2$, with strict inclusion under the stated conditions (Proposition S13). □

Proposition S16 (Stratum Decomposition of the Level 3 Program)**.** *The Level 3 sharp bounds decompose into K independent stratum contributions:*

$$\mathcal{B}_3^{\mathrm{U}}(w) = \sum_{k=1}^{K} \mathcal{B}_{3,k}^{\mathrm{U}}(w), \qquad \mathcal{B}_3^{\mathrm{L}}(w) = \sum_{k=1}^{K} \mathcal{B}_{3,k}^{\mathrm{L}}(w), \tag{S75}$$

where each stratum contribution solves the within-stratum LP:

$$\mathcal{B}_{3,k}^{\mathrm{U}}(w) = \max_{\{\delta_j\}_{j \in S_k}} \sum_{j \in S_k} w_j \, \delta_j \quad s.t. \quad \delta_j \in [0, d_j] \;\; \forall \, j \in S_k, \quad \frac{1}{n_k} \sum_{j \in S_k} \delta_j = \tau(x_k). \tag{S76}$$

Full proof of Proposition S16. Any feasible $\delta \in \mathcal{F}_3^{\delta}$ can be written as the concatenation of its stratum components $\delta^{(k)} = \{\delta_j\}_{j \in S_k}$ for $k = 1, \ldots, K$. The box constraints require $\delta_j \in [0, d_j]$ for each j, which is equivalent to requiring $\delta^{(k)} \in \mathcal{F}_{1,k}^{\delta}$ where

$$\mathcal{F}_{1,k}^{\delta} = \{(\delta_j)_{j \in S_k} : \delta_j \in [0, d_j] \;\; \forall \, j \in S_k\} \tag{S77}$$

is the box for stratum k. The stratum mean constraints require $(1/n_k) \sum_{j \in S_k} \delta_j = \tau(x_k)$ for each k, which restricts $\delta^{(k)}$ alone. Defining the stratum-level feasible set

$$\mathcal{F}_{2,k}^{\delta} = \left\{ (\delta_j)_{j \in S_k} \in \mathcal{F}_{1,k}^{\delta} : \frac{1}{n_k} \sum_{j \in S_k} \delta_j = \tau(x_k) \right\}, \tag{S78}$$

the global feasible set is the Cartesian product:

$$\mathcal{F}_3^{\delta} = \mathcal{F}_{2,1}^{\delta} \times \mathcal{F}_{2,2}^{\delta} \times \cdots \times \mathcal{F}_{2,K}^{\delta}. \tag{S79}$$

The objective is separable across strata, since $\sum_{j\in\mathcal{C}_1} w_j\delta_j = \sum_{k=1}^{K}\sum_{j\in S_k} w_j\delta_j$, and by (S79) the feasible set is a Cartesian product. Substituting into the maximization:

$$\begin{aligned}
\mathcal{B}_3^{\mathrm{U}}(w) &= \max_{\delta\in\mathcal{F}_3^\delta} \sum_{j\in\mathcal{C}_1} w_j\,\delta_j \\
&= \max_{\delta\in\mathcal{F}_3^\delta} \sum_{k=1}^{K}\sum_{j\in S_k} w_j\,\delta_j \\
&= \max_{\delta^{(1)}\in\mathcal{F}_{2,1}^\delta,\,\ldots,\,\delta^{(K)}\in\mathcal{F}_{2,K}^\delta} \quad \sum_{k=1}^{K}\sum_{j\in S_k} w_j\,\delta_j \\
&= \sum_{k=1}^{K} \max_{\delta^{(k)}\in\mathcal{F}_{2,k}^\delta} \sum_{j\in S_k} w_j\,\delta_j \\
&= \sum_{k=1}^{K} \mathcal{B}_{3,k}^{\mathrm{U}}(w). \qquad \text{(S80)}
\end{aligned}$$

The fourth equality (the max inside the sum) is valid because the feasible set is the Cartesian product (S79) and each summand depends only on the variables of its own stratum. Formally: for any function $F(\delta^{(1)},\ldots,\delta^{(K)}) = \sum_{k=1}^{K} f_k(\delta^{(k)})$ and product domain $\mathcal{D} = \mathcal{D}_1\times\cdots\times\mathcal{D}_K$:

$$\max_{\delta\in\mathcal{D}}\; F(\delta) = \sum_{k=1}^{K} \max_{\delta^{(k)}\in\mathcal{D}_k}\; f_k(\delta^{(k)}), \qquad \text{(S81)}$$

since the maximizer sets each component independently to its own maximum. This is the standard separability property of additive optimization over product domains (Bertsekas, 1999, Proposition 1.2.1). The argument for the lower bound is identical with max replaced by min. □

Proposition S17 (Tightening from Level 2 to Level 3)**.** *For any weight vector w, $\mathcal{B}_3^{\mathrm{L}}(w) \geq \mathcal{B}_2^{\mathrm{L}}(w)$ and $\mathcal{B}_3^{\mathrm{U}}(w) \leq \mathcal{B}_2^{\mathrm{U}}(w)$. The tightening is strict whenever there exist strata with $\tau(x_k) \neq \tau(x_{k'})$ and the weight vector assigns different weights to at least one pair of exposed controls in different strata.*

Full proof of Proposition S17. We prove the upper bound inequality; the lower bound follows by the same argument.

The weak inequality $\mathcal{B}_3^{\mathrm{U}}(w) \leq \mathcal{B}_2^{\mathrm{U}}(w)$ follows from $\mathcal{F}_3^\delta \subseteq \mathcal{F}_2^\delta$ (Lemma 4.4):

$$\mathcal{B}_3^{\mathrm{U}}(w) = \max_{\delta\in\mathcal{F}_3^\delta} \sum w_j\delta_j \;\leq\; \max_{\delta\in\mathcal{F}_2^\delta} \sum w_j\delta_j = \mathcal{B}_2^{\mathrm{U}}(w), \qquad \text{(S82)}$$

since maximizing over a subset yields a weakly smaller value.

For strict tightening, we proceed in two steps: first, we show that the Level 2 optimal solution δ^* does not belong to $\mathcal{F}_3^\delta$ under the stated conditions; second, we use the uniqueness

of the Level 2 optimum to conclude that no other point in $\mathcal{F}_2^\delta$ achieves the same objective value, so the Level 3 maximum is strictly smaller.

We begin by showing $\delta^* \notin \mathcal{F}_3^\delta$. The Level 2 optimal δ^* from Theorem 4.9(I) is constructed by sorting all n_1 exposed controls by weight globally and assigning $\delta_j^* = d_j$ for units above the global threshold $\lambda^*/n_1 = w_{(k^*)}$, $\delta_j^* = 0$ for units below the threshold, and $\delta_{(k^*)}^*$ equal to the residual. The within-stratum mean at this allocation is:

$$\bar{\delta}_k^* = \frac{1}{n_k} \sum_{j \in S_k} \delta_j^*. \tag{S83}$$

For δ^* to belong to $\mathcal{F}_3^\delta$, we would need $\bar{\delta}_k^* = \tau(x_k)$ for every $k = 1, \ldots, K$. The global sorting assigns δ_j^* based on where w_j falls relative to the global threshold, without regard to stratum membership. Consider two strata S_k and $S_{k'}$ with $\tau(x_k) \neq \tau(x_{k'})$ and units $i \in S_k$, $j \in S_{k'}$ with $w_i \neq w_j$. Without loss of generality, suppose $w_i > w_j$. If $w_i > \lambda^*/n_1 > w_j$ — that is, unit i is above the global threshold and unit j is below — the global allocation assigns $\delta_i^* = d_i$ and $\delta_j^* = 0$. The global sorting's allocation to stratum $S_{k'}$ is determined entirely by how the global threshold intersects $S_{k'}$: units in $S_{k'}$ above the threshold receive their full capacity, units below receive zero. The resulting stratum mean $\bar{\delta}_{k'}^*$ is determined by the proportion of stratum $S_{k'}$'s units that fall above the global threshold and their capacities. There is no mechanism in the global sorting to match this mean to $\tau(x_{k'})$, since the threshold is determined by the global budget $n_1\psi$, not the stratum budgets.

To verify concretely that $\bar{\delta}_{k'}^* \neq \tau(x_{k'})$: the global mean constraint gives $\sum_k (n_k/n_1)\bar{\delta}_k^* = \psi = \sum_k (n_k/n_1)\tau(x_k)$. If $\bar{\delta}_k^* = \tau(x_k)$ held for all strata, this would be satisfied trivially. But the global sorting assigns spillover based on weight rank, not stratum identity. When $\tau(x_k)$ varies across strata — meaning different strata have different mean spillover requirements — and the weight ordering does not perfectly align with the stratum structure, the global allocation produces stratum means that deviate from the required $\tau(x_k)$. Specifically, strata containing a disproportionate share of high-weight units receive more spillover than $n_k\tau(x_k)$ from the global allocation, while strata with disproportionately low-weight units receive less. Since $\tau(x_k) \neq \tau(x_{k'})$ and there exist cross-stratum weight differences, this misalignment occurs: at least one stratum has $\bar{\delta}_k^* \neq \tau(x_k)$. Therefore $\delta^* \notin \mathcal{F}_3^\delta$.

We now use the uniqueness of the Level 2 optimum to conclude strict tightening. Under distinct weights (our standing assumption), the Level 2 LP (35) has a unique optimal solution: Theorem 4.9 constructs a unique vertex of $\mathcal{F}_2^\delta$ at which the optimum is achieved, and with distinct weights no other vertex achieves the same objective value (since the objective coefficients w_j are all different, the objective function is not constant on any edge of the polytope). Therefore δ^* is the unique maximizer:

$$\sum_{j \in \mathcal{C}_1} w_j \delta_j < \sum_{j \in \mathcal{C}_1} w_j \delta_j^* = \mathcal{B}_2^{\mathrm{U}}(w) \quad \text{for all } \delta \in \mathcal{F}_2^\delta \setminus \{\delta^*\}. \tag{S84}$$

Since $\delta^* \notin \mathcal{F}_3^\delta$, every $\delta \in \mathcal{F}_3^\delta \subseteq \mathcal{F}_2^\delta$ belongs to $\mathcal{F}_2^\delta \setminus \{\delta^*\}$, and therefore:

$$\mathcal{B}_3^{\mathrm{U}}(w) = \max_{\delta \in \mathcal{F}_3^\delta} \sum w_j \delta_j < \mathcal{B}_2^{\mathrm{U}}(w). \tag{S85}$$

This establishes strict tightening for the upper bound. The same argument applies to the lower bound with δ^{**} replacing δ^*. □

Proposition S18 (Width Reduction via Jensen's Inequality)**.** *Let $\mathcal{W}_2(w)$ and $\mathcal{W}_3(w)$ denote the widths of the Level 2 and Level 3 identification regions.*

(i) The Level 3 width is bounded by:

$$\mathcal{W}_3(w) \;\le\; 2\sum_{k=1}^{K} n_k \max_{j\in S_k}\big|w_j - \bar{w}_k\big| \cdot \min\big(\tau(x_k),\; \bar{d}_k - \tau(x_k)\big), \tag{S86}$$

where $\bar{w}_k = n_k^{-1}\sum_{j\in S_k} w_j$ is the mean weight within stratum k.

(ii) Under homogeneous capacities ($d_j = d$ for all j), the total budget flexibility satisfies:

$$\sum_{k=1}^{K} \frac{n_k}{n_1} \min\big(\tau(x_k),\; d - \tau(x_k)\big) \;\le\; \min(\psi,\; d - \psi), \tag{S87}$$

with strict inequality whenever some strata have $\tau(x_k) < d/2$ and others have $\tau(x_{k'}) > d/2$.

Full proof of Proposition S18. For part (i): each stratum sub-problem is a Level 2 bounded-variable LP with n_k units, box constraints $\delta_j \in [0, d_j]$ for $j \in S_k$, and mean constraint $(1/n_k)\sum_{j\in S_k}\delta_j = \tau(x_k)$. Applying Theorem 4.10(iii) within stratum k, with the global quantities replaced by their stratum-level analogs ($n_1 \to n_k$, $\psi \to \tau(x_k)$, $\bar{d} \to \bar{d}_k$, $\bar{w} \to \bar{w}_k$):

$$\mathcal{W}_{3,k}(w) \le 2 \max_{j\in S_k}\big|w_j - \bar{w}_k\big| \cdot n_k \min\big(\tau(x_k),\; \bar{d}_k - \tau(x_k)\big). \tag{S88}$$

Summing (S88) across $k = 1, \ldots, K$ and using the width decomposition $\mathcal{W}_3(w) = \sum_{k=1}^{K}\mathcal{W}_{3,k}(w)$ from Proposition S16:

$$\begin{aligned}\mathcal{W}_3(w) &= \sum_{k=1}^{K}\mathcal{W}_{3,k}(w) \\ &\le \sum_{k=1}^{K} 2 \max_{j\in S_k}\big|w_j - \bar{w}_k\big| \cdot n_k \min\big(\tau(x_k),\; \bar{d}_k - \tau(x_k)\big).\end{aligned} \tag{S89}$$

This establishes (S86).

For part (ii): define the budget-flexibility function $f\colon [0,d] \to \mathbb{R}$ by:

$$f(t) = \min(t,\; d - t). \tag{S90}$$

We verify that f is concave. For any $t_1, t_2 \in [0,d]$ and $\alpha \in [0,1]$:

$$\begin{aligned}f(\alpha t_1 + (1-\alpha)t_2) &= \min\big(\alpha t_1 + (1-\alpha)t_2,\; d - \alpha t_1 - (1-\alpha)t_2\big) \\ &= \min\big(\alpha t_1 + (1-\alpha)t_2,\; \alpha(d - t_1) + (1-\alpha)(d - t_2)\big).\end{aligned} \tag{S91}$$

By the concavity of the min operation (the minimum of two affine functions is concave), or directly: the minimum of two convex combinations is at least the convex combination of the minima:

$$\begin{aligned}\min\big(\alpha t_1 + (1-\alpha)t_2,\ \alpha(d-t_1) + (1-\alpha)(d-t_2)\big) &\geq \alpha \min(t_1, d-t_1) + (1-\alpha)\min(t_2, d-t_2) \\ &= \alpha f(t_1) + (1-\alpha) f(t_2). \end{aligned} \tag{S92}$$

The inequality in (S92) follows from the general property that for any reals $a_1, a_2, b_1, b_2 \geq 0$ and $\alpha \in [0,1]$:

$$\min(\alpha a_1 + (1-\alpha)a_2,\ \alpha b_1 + (1-\alpha)b_2) \geq \alpha \min(a_1, b_1) + (1-\alpha)\min(a_2, b_2). \tag{S93}$$

This is verified by cases: if $\min(a_1, b_1) = a_1$ and $\min(a_2, b_2) = a_2$, then the right-hand side is $\alpha a_1 + (1-\alpha)a_2 \leq$ the left-hand side; the other cases are analogous. Combining (S91)–(S92) establishes that f is concave on $[0, d]$.

Now apply Jensen's inequality to the discrete probability measure π on $[0, d]$ that places mass n_k/n_1 on $\tau(x_k)$ for $k = 1, \ldots, K$. The expectation under π is:

$$\mathbb{E}_\pi[\tau] = \sum_{k=1}^{K} \frac{n_k}{n_1} \tau(x_k) = \psi, \tag{S94}$$

where the last equality uses $\psi = \sum_k (n_k/n_1)\tau(x_k)$. Jensen's inequality for the concave function f gives:

$$\mathbb{E}_\pi[f(\tau)] \leq f(\mathbb{E}_\pi[\tau]), \tag{S95}$$

which is:

$$\sum_{k=1}^{K} \frac{n_k}{n_1} \min\big(\tau(x_k),\ d - \tau(x_k)\big) \leq \min(\psi,\ d - \psi). \tag{S96}$$

This establishes (S87).

For the strict inequality condition: f is piecewise linear, equal to t on $[0, d/2]$ and to $d - t$ on $[d/2, d]$. Jensen's inequality for a concave function is strict unless the function is affine over the support of the measure. Since f is affine on each of the two pieces, equality obtains whenever all $\tau(x_k)$ lie on the same piece — that is, all $\tau(x_k) \leq d/2$ or all $\tau(x_k) \geq d/2$. When some strata have $\tau(x_k) < d/2$ and others have $\tau(x_{k'}) > d/2$, the measure straddles the kink point and the inequality is strict. □

Corollary S19 (Point Identification Under Conditional Homogeneity)**.** *Suppose that within each stratum S_k, the spillover effects are homogeneous: $\delta_j = \tau(x_k)$ for all $j \in S_k$. Then the Level 3 identification region collapses to*

$$\mathcal{B}_3^{\mathrm{U}}(w) = \mathcal{B}_3^{\mathrm{L}}(w) = \sum_{k=1}^{K} W_k\, \tau(x_k), \tag{S97}$$

where $W_k = \sum_{j \in S_k} w_j$. This holds for any weight vector w.

Full proof of Corollary S19. Under conditional homogeneity, $\delta_j = \tau(x_k)$ for all $j \in S_k$. The stratum mean constraint requires:

$$\frac{1}{n_k} \sum_{j \in S_k} \delta_j = \tau(x_k). \tag{S98}$$

Since all δ_j in the stratum are equal to $\tau(x_k)$, the left-hand side of (S98) is:

$$\frac{1}{n_k} \sum_{j \in S_k} \tau(x_k) = \frac{1}{n_k} \cdot n_k \cdot \tau(x_k) = \tau(x_k), \tag{S99}$$

which equals the right-hand side. The stratum feasible set $\mathcal{F}^{\delta}_{2,k}$ contains exactly one element: $\delta_j = \tau(x_k)$ for all $j \in S_k$. Since the feasible set is a singleton, the maximum and minimum of $\sum_{j \in S_k} w_j \delta_j$ over $\mathcal{F}^{\delta}_{2,k}$ coincide:

$$\mathcal{B}^{\mathrm{U}}_{3,k}(w) = \mathcal{B}^{\mathrm{L}}_{3,k}(w) = \sum_{j \in S_k} w_j \, \tau(x_k) = \tau(x_k) \sum_{j \in S_k} w_j = W_k \, \tau(x_k). \tag{S100}$$

The within-stratum width is therefore:

$$\mathcal{W}_{3,k}(w) = \mathcal{B}^{\mathrm{U}}_{3,k}(w) - \mathcal{B}^{\mathrm{L}}_{3,k}(w) = 0 \quad \text{for each } k. \tag{S101}$$

Summing across strata using Proposition S16:

$$\mathcal{B}^{\mathrm{U}}_{3}(w) = \sum_{k=1}^{K} \mathcal{B}^{\mathrm{U}}_{3,k}(w) = \sum_{k=1}^{K} W_k \, \tau(x_k), \tag{S102}$$

$$\mathcal{B}^{\mathrm{L}}_{3}(w) = \sum_{k=1}^{K} \mathcal{B}^{\mathrm{L}}_{3,k}(w) = \sum_{k=1}^{K} W_k \, \tau(x_k). \tag{S103}$$

Therefore $\mathcal{B}^{\mathrm{U}}_{3}(w) = \mathcal{B}^{\mathrm{L}}_{3}(w) = \sum_k W_k \, \tau(x_k)$, and the width is zero:

$$\mathcal{W}_3(w) = 0. \tag{S104}$$

Since this holds for any weight vector w, the contamination is point-identified: $\mathcal{B}(w) = \sum_k W_k \, \tau(x_k)$. □

S3.2 Proofs of Main-Text Results

Proof. Partition the weighted sum of control outcomes into contributions from unexposed and exposed controls:

$$\sum_{j \in \mathcal{C}} w_j \, \Delta Y_j^{\mathrm{obs}} = \sum_{j \in \mathcal{C}_0} w_j \, \Delta Y_j(0,0) + \sum_{j \in \mathcal{C}_1} w_j \, \Delta Y_j(0,1). \tag{S105}$$

For each $j \in \mathcal{C}_1$, write $\Delta Y_j(0,1) = \Delta Y_j(0,0) + [\Delta Y_j(0,1) - \Delta Y_j(0,0)]$ to obtain

$$\sum_{j \in \mathcal{C}_1} w_j \, \Delta Y_j(0,1) = \sum_{j \in \mathcal{C}_1} w_j \, \Delta Y_j(0,0) + \sum_{j \in \mathcal{C}_1} w_j \, [\Delta Y_j(0,1) - \Delta Y_j(0,0)] \, . \tag{S106}$$

Substituting (S106) into (S105) and combining the $\Delta Y_j(0,0)$ terms yields

$$\sum_{j\in\mathcal{C}} w_j\,\Delta Y_j^{\text{obs}} = \sum_{j\in\mathcal{C}} w_j\,\Delta Y_j(0,0) + \sum_{j\in\mathcal{C}_1} w_j\,[\Delta Y_j(0,1) - \Delta Y_j(0,0)]\,. \tag{S107}$$

The result follows from $\hat{\tau}(w) = \Delta\bar{Y}_{\mathcal{T}} - \sum_{j\in\mathcal{C}} w_j \Delta Y_j^{\text{obs}}$. □

Proof. Part (i) follows from substituting $\Delta Y_j(0,1) = \Delta Y_j^{\text{obs}}$ for $j \in \mathcal{C}_1$ in (8). Part (ii) is immediate from the form of (8): the sum is zero if and only if all summands vanish (which requires either $w_j = 0$ or $\Delta Y_j(0,1) = \Delta Y_j(0,0)$ for each $j \in \mathcal{C}_1$) or non-zero summands cancel. Part (iii): under monotone nonnegative spillovers, each term $w_j[\Delta Y_j(0,1) - \Delta Y_j(0,0)]$ is nonnegative since $w_j \geq 0$, so the sum is nonnegative. □

Proof. Part (i): Under $\mathcal{F}_0$, for each $j \in \mathcal{C}_1$ the counterfactual $\Delta Y_j(0,0) \in [a,b]$, so $\Delta Y_j^{\text{obs}} - \Delta Y_j(0,0) \in [\Delta Y_j^{\text{obs}} - b,\, \Delta Y_j^{\text{obs}} - a]$. Since $w_j \geq 0$, summing over $j \in \mathcal{C}_1$ preserves the inequalities. Part (ii): Under $\mathcal{F}_1$, each $\Delta Y_j(0,0) \in [a, \Delta Y_j^{\text{obs}}]$, so $\Delta Y_j^{\text{obs}} - \Delta Y_j(0,0) \in [0,\, \Delta Y_j^{\text{obs}} - a]$. The lower bound is attained by setting $c_j = \Delta Y_j^{\text{obs}}$ (no spillover, feasible since $\Delta Y_j^{\text{obs}} \leq \Delta Y_j^{\text{obs}}$). The upper bound is attained by setting $c_j = a$ (maximal spillover, feasible since $a \leq \Delta Y_j^{\text{obs}}$ by support). □

Proof. We prove each part in sequence.

Part (i). For exposed controls ($E_j = 1$), the observed outcome change equals the exposed potential outcome: $\Delta Y_j^{\text{obs}} = \Delta Y_j(0,1)$. Thus

$$\mathbb{E}[\Delta Y_j \mid E_j = 1, X_j = x, A_j = 0] = \mathbb{E}[\Delta Y_j(0,1) \mid E_j = 1, X_j = x, A_j = 0]. \tag{S108}$$

For unexposed controls ($E_j = 0$), $\Delta Y_j^{\text{obs}} = \Delta Y_j(0,0)$, so

$$\mathbb{E}[\Delta Y_j \mid E_j = 0, X_j = x, A_j = 0] = \mathbb{E}[\Delta Y_j(0,0) \mid E_j = 0, X_j = x, A_j = 0]. \tag{S109}$$

By exposure ignorability (Assumption 1), $\Delta Y_j(0,0) \perp\!\!\!\perp E_j \mid X_j, \{X_k\}_{k\in\mathcal{N}_j}, A_j = 0$, which implies

$$\mathbb{E}[\Delta Y_j(0,0) \mid E_j = 0, X_j = x, A_j = 0] = \mathbb{E}[\Delta Y_j(0,0) \mid E_j = 1, X_j = x, A_j = 0]. \tag{S110}$$

The right-hand side of (S110) is the conditional mean of the *unobserved* clean counterfactual for exposed controls with covariates x. Exposure ignorability equates it to the conditional mean of the *observed* clean outcome for unexposed controls with the same covariates—the content of part (iv), which we record separately for its role in the partial identification analysis.

Taking the difference of (S108) and (S109), and substituting (S110):

$$\begin{aligned}
&\mathbb{E}[\Delta Y_j \mid E_j = 1, X_j = x, A_j = 0] - \mathbb{E}[\Delta Y_j \mid E_j = 0, X_j = x, A_j = 0] \\
&\quad = \mathbb{E}[\Delta Y_j(0,1) \mid E_j = 1, X_j = x, A_j = 0] - \mathbb{E}[\Delta Y_j(0,0) \mid E_j = 1, X_j = x, A_j = 0] \\
&\quad = \mathbb{E}[\Delta Y_j(0,1) - \Delta Y_j(0,0) \mid E_j = 1, X_j = x, A_j = 0] \\
&\quad = \tau(x),
\end{aligned} \tag{S111}$$

where the last equality uses exposure ignorability again: since $\Delta Y_j(0,0) \perp\!\!\!\perp E_j \mid X_j, A_j = 0$ and $\Delta Y_j(0,1)$ is observed only for $E_j = 1$, the conditional expectation $\mathbb{E}[\Delta Y_j(0,1) - \Delta Y_j(0,0) \mid E_j = 1, X_j = x, A_j = 0] = \mathbb{E}[\Delta Y_j(0,1) - \Delta Y_j(0,0) \mid X_j = x, A_j = 0] = \tau(x)$.

Part (ii). Integrate the result of part (i) over X_j conditional on $A_j = 0$:

$$\begin{aligned}\psi &= \mathbb{E}[\tau(X_j) \mid A_j = 0] \\ &= \mathbb{E}\left[\mathbb{E}[\Delta Y_j \mid E_j = 1, X_j, A_j = 0] - \mathbb{E}[\Delta Y_j \mid E_j = 0, X_j, A_j = 0] \mid A_j = 0\right].\end{aligned}$$

Part (iii). For exposed controls, $\Delta Y_j^{\text{obs}} = \Delta Y_j(0,1)$, so

$$\mathbb{E}[\Delta Y_j^{\text{obs}} \mid E_j = 1, A_j = 0] = \mathbb{E}[\Delta Y_j(0,1) \mid E_j = 1, A_j = 0].$$

By definition of $\psi_{\text{ATT}} = \mathbb{E}[\Delta Y_j(0,1) - \Delta Y_j(0,0) \mid E_j = 1, A_j = 0]$ and the result of part (i) integrated over exposed controls (which yields $\psi_{\text{ATT}} = \psi$ under exposure ignorability; see Remark S2 below):

$$\begin{aligned}\psi &= \mathbb{E}[\Delta Y_j(0,1) - \Delta Y_j(0,0) \mid E_j = 1, A_j = 0] \\ &= \mathbb{E}[\Delta Y_j^{\text{obs}} \mid E_j = 1, A_j = 0] - \mathbb{E}[\Delta Y_j(0,0) \mid E_j = 1, A_j = 0].\end{aligned}$$

Rearranging:

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

The left-hand side is the population mean of the unobserved counterfactual outcome changes for exposed controls—the mean of the unidentified vector $\{\Delta Y_j(0,0)\}_{j \in \mathcal{C}_1}$ from Section 2. The right-hand side consists entirely of identified quantities: the observed mean outcome change among exposed controls and the detection estimand ψ.

Part (iv). From (S110):

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

where the second equality uses $\Delta Y_j^{\text{obs}} = \Delta Y_j(0,0)$ for unexposed controls. □

Proof. Write $\hat{\psi}_{\text{DR}} = \psi + (\hat{\psi}_{\text{DR}} - \psi)$. The remainder decomposes as

$$\begin{aligned}\hat{\psi}_{\text{DR}} - \psi = \frac{1}{n_1} \sum_{j \in \mathcal{C}_1} [\Delta Y_j - \mu_0(X_j) - \psi] - \frac{1}{n_1} \sum_{j \in \mathcal{C}_0} \frac{\pi_j}{1 - \pi_j} [\Delta Y_j - \mu_0(X_j)] \\ + R_n(\hat{\mu}_0, \hat{\pi}),\end{aligned}$$

where the first two sums constitute the influence function evaluated at the true nuisance parameters, and the remainder satisfies $R_n(\hat{\mu}_0, \hat{\pi}) = O_p(\|\hat{\mu}_0 - \mu_0\| \cdot \|\hat{\pi} - \pi\|)$. Under condition (a), the residuals $\Delta Y_j - \mu_0(X_j)$ have conditional mean $\tau(X_j)$ for exposed and zero for unexposed controls; the second sum vanishes in probability regardless of $\hat{\pi}$. Under condition (b), the reweighting aligns the unexposed covariate distribution with the exposed distribution, making the estimator consistent regardless of $\hat{\mu}_0$. When both converge at $o_p(n^{-1/4})$ rates, $R_n = o_p(n^{-1/2})$, and $\hat{\psi}_{\text{DR}}$ inherits the influence function representation that achieves the efficiency bound. See Kennedy (2023) and Chernozhukov et al. (2018) for the general theory. □

Proof. Part (i) restates Theorem 3.2(iii) in sample form: $\psi = \mathbb{E}[\Delta Y_j^{\text{obs}} \mid E_j = 1, A_j = 0] - \mathbb{E}[\Delta Y_j(0,0) \mid E_j = 1, A_j = 0]$ implies $\mathbb{E}[\Delta Y_j(0,0) \mid E_j = 1, A_j = 0] = \mathbb{E}[\Delta Y_j^{\text{obs}} \mid E_j = 1, A_j = 0] - \psi$. Replacing population expectations with sample averages yields (23). Consistency of the right-hand side follows from the law of large numbers for $\bar{\Delta Y}_{\mathcal{C}_1}^{\text{obs}}$ and consistency of $\hat{\psi}_{\text{DR}}$ (Proposition 3.3). Part (ii) restates Theorem 3.2(iv): under exposure ignorability, $\mathbb{E}[\Delta Y_j(0,0) \mid X_j = x, E_j = 1, A_j = 0] = \mathbb{E}[\Delta Y_j \mid E_j = 0, X_j = x, A_j = 0] = \mu_0(x)$. Consistency of $\hat{\mu}_0(x)$ for $\mu_0(x)$ follows from standard nonparametric regression results under regularity conditions on the covariate distribution and the outcome surface. □

Proof. The inclusion $\mathcal{F}_2 \subseteq \mathcal{F}_1$ is immediate: $\mathcal{F}_2$ imposes all constraints of $\mathcal{F}_1$ plus an equality constraint.

For $\mathcal{F}_3 \subseteq \mathcal{F}_2$: let $(c_j)_{j \in \mathcal{C}_1} \in \mathcal{F}_3$, so that $n_k^{-1} \sum_{j \in S_k} c_j = \bar{\Delta Y}_{S_k}^{\text{obs}} - \tau(x_k)$ for each k. The global mean decomposes as

$$\frac{1}{n_1} \sum_{j \in \mathcal{C}_1} c_j = \sum_{k=1}^{K} \frac{n_k}{n_1} \left(\bar{\Delta Y}_{S_k}^{\text{obs}} - \tau(x_k) \right) = \bar{\Delta Y}_{\mathcal{C}_1}^{\text{obs}} - \sum_{k=1}^{K} \frac{n_k}{n_1} \tau(x_k) = \bar{\Delta Y}_{\mathcal{C}_1}^{\text{obs}} - \psi, \tag{S112}$$

where the last equality uses $\psi = \sum_k (n_k/n_1)\tau(x_k)$ by the law of iterated expectations. Hence $(c_j) \in \mathcal{F}_2$. The inclusion is generally strict: if $\tau(x_1) \neq \tau(x_2)$ for some strata, the stratum constraints are not implied by the global constraint. □

Proof. For any function f and sets $A \subseteq B$: $\min_B f \leq \min_A f$ and $\max_B f \geq \max_A f$. Applying this to $f(c) = \sum w_j c_j$ with $\mathcal{F}_k \subseteq \mathcal{F}_{k-1}$ from Lemma 4.4, the upper bounds are non-increasing and the lower bounds are non-decreasing across levels, yielding the stated inclusion. □

Proof. The weighted sum of the unobserved counterfactuals under symmetric weights is

$$\sum_{j \in \mathcal{C}_1} w_j \, c_j = \bar{w} \sum_{j \in \mathcal{C}_1} c_j = \bar{w} \cdot n_1 \cdot \frac{1}{n_1} \sum_{j \in \mathcal{C}_1} c_j.$$

The first equality follows from $w_j = \bar{w}$ for all j: a constant can be factored out of the sum. The second equality multiplies and divides by n_1. The quantity $n_1^{-1} \sum_{j \in \mathcal{C}_1} c_j$ is the mean of the counterfactual vector, which is constrained by $\mathcal{F}_2$ to equal $M = \bar{\Delta Y}_{\mathcal{C}_1}^{\text{obs}} - \psi$. Therefore:

$$\sum_{j \in \mathcal{C}_1} w_j \, c_j = \bar{w} \cdot n_1 \cdot M = W_{\text{exp}} \cdot \left(\bar{\Delta Y}_{\mathcal{C}_1}^{\text{obs}} - \psi \right).$$

This expression does not depend on the individual c_j values—only on their mean, which is pinned down by ψ. Any two configurations (c_j) and (c'_j) in $\mathcal{F}_2$ produce the same weighted

sum, because they share the same mean. The weighted sum is therefore a constant over $\mathcal{F}_2$, and the identification region for $\mathcal{B}(w)$ collapses to a single point:

$$\begin{aligned}\mathcal{B}(w) &= S_{\text{obs}}(w) - W_{\text{exp}}\left(\bar{\Delta Y}^{\text{obs}}_{\mathcal{C}_1} - \psi\right) \\ &= \bar{w}\sum_{j\in\mathcal{C}_1}\Delta Y_j^{\text{obs}} - \bar{w}\cdot n_1\cdot\bar{\Delta Y}^{\text{obs}}_{\mathcal{C}_1} + W_{\text{exp}}\cdot\psi \\ &= W_{\text{exp}}\cdot\bar{\Delta Y}^{\text{obs}}_{\mathcal{C}_1} - W_{\text{exp}}\cdot\bar{\Delta Y}^{\text{obs}}_{\mathcal{C}_1} + W_{\text{exp}}\cdot\psi \\ &= W_{\text{exp}}\cdot\psi.\end{aligned}$$

The first line substitutes $S_{\text{obs}}(w) = \sum w_j\Delta Y_j^{\text{obs}} = \bar{w}\sum\Delta Y_j^{\text{obs}}$ and the constant value of the weighted sum. The cancellation in the third line leaves only $W_{\text{exp}}\cdot\psi$, the total weight on exposed controls multiplied by the average spillover effect. Since this quantity involves no unobserved counterfactuals, the contamination is point-identified. □

Proof. It suffices to exhibit two configurations in $\mathcal{F}_2$ that produce different values of $\sum w_jc_j$. Since $\mathcal{F}_2$ contains more than one element, there exist distinct $(c_j), (c'_j) \in \mathcal{F}_2$ with $c_k \neq c'_k$ for some $k \in \mathcal{C}_1$. Both configurations satisfy the same mean constraint, $n_1^{-1}\sum c_j = n_1^{-1}\sum c'_j = M$, so $\sum(c_j - c'_j) = 0$—the differences sum to zero.

Now consider the difference in the weighted sums:

$$\sum_{j\in\mathcal{C}_1} w_jc_j - \sum_{j\in\mathcal{C}_1} w_jc'_j = \sum_{j\in\mathcal{C}_1} w_j(c_j - c'_j).$$

This is a weighted sum of mean-zero differences. We claim it is generically nonzero when the weights are not all equal. To see this, suppose for contradiction that $\sum w_j(c_j - c'_j) = 0$ for *every* pair of distinct configurations in $\mathcal{F}_2$. Let $\Delta_j = c_j - c'_j$, so $\sum\Delta_j = 0$. The condition $\sum w_j\Delta_j = 0$ for all such Δ means w is orthogonal to every mean-zero vector in $\mathbb{R}^{n_1}$ that is consistent with the box constraints. But the set of feasible mean-zero perturbations spans an (n_1-1)-dimensional subspace (when $n_1 \geq 2$ and the box constraints are non-degenerate), so w can be orthogonal to all of them only if w is proportional to the constant vector $(1, \ldots, 1)$, that is, all weights are equal. Since by assumption the weights are not all equal, there exists at least one feasible perturbation Δ with $\sum w_j\Delta_j \neq 0$.

Now let us construct such a perturbation explicitly. Let $j_H, j_L \in \mathcal{C}_1$ with $w_{j_H} > w_{j_L}$. Choose $\epsilon > 0$ small enough that both $c_{j_H} - \epsilon$ and $c_{j_L} + \epsilon$ remain in their respective box constraints. Define $c_j^{(1)} = c_j$ for $j \notin \{j_H, j_L\}$, $c_{j_H}^{(1)} = c_{j_H} - \epsilon$, and $c_{j_L}^{(1)} = c_{j_L} + \epsilon$. This new configuration satisfies the box constraints (by choice of ϵ) and the mean constraint (since $-\epsilon + \epsilon = 0$ preserves the mean). The change in the weighted sum is

$$\sum w_jc_j^{(1)} - \sum w_jc_j = -\epsilon\, w_{j_H} + \epsilon\, w_{j_L} = \epsilon(w_{j_L} - w_{j_H}) < 0,$$

since $w_{j_H} > w_{j_L}$. Therefore the two feasible configurations produce different weighted sums, which means they produce different contamination values. The identification region has strictly positive width. □

Full proof of Theorem 4.9. We prove part (I); part (II) is derived from (I) at the end.

By Proposition S7, at the optimum δ^* with multiplier λ^*, every unit with $w_{(j)} > \lambda^*/n_1$ satisfies $\delta^*_{(j)} = d_{(j)}$, and every unit with $w_{(j)} < \lambda^*/n_1$ satisfies $\delta^*_{(j)} = 0$. Since the weights are sorted as $w_{(1)} < w_{(2)} < \cdots < w_{(n_1)}$, the threshold λ^*/n_1 partitions the index set $\{1, \ldots, n_1\}$ into a lower segment assigned to zero and an upper segment assigned to capacity. Denote by k^* the index of the unit at the partition boundary. Then:

$$\delta^*_{(j)} = d_{(j)} \quad \text{for all } j > k^*, \tag{S113}$$

$$\delta^*_{(j)} = 0 \qquad \text{for all } j < k^*. \tag{S114}$$

The value of the threshold unit $\delta^*_{(k^*)}$ is determined by the mean constraint. Substituting (S113)–(S114) into the constraint $(1/n_1)\sum_{j=1}^{n_1} \delta^*_{(j)} = \psi$:

$$\frac{1}{n_1}\left[\sum_{j=1}^{k^*-1} 0 \;+\; \delta^*_{(k^*)} \;+\; \sum_{j=k^*+1}^{n_1} d_{(j)}\right] = \psi. \tag{S115}$$

The first sum vanishes and the last sum equals $C(k^*{+}1)$ by definition (37), giving:

$$\frac{1}{n_1}\Big[\delta^*_{(k^*)} + C(k^*{+}1)\Big] = \psi. \tag{S116}$$

Multiplying both sides by n_1:

$$\delta^*_{(k^*)} + C(k^*{+}1) = n_1\psi. \tag{S117}$$

Solving for $\delta^*_{(k^*)}$:

$$\delta^*_{(k^*)} = n_1\psi - C(k^*{+}1). \tag{S118}$$

This is the residual spillover that the threshold unit must absorb after the higher-weight units have been filled to capacity.

For the solution to be feasible, $\delta^*_{(k^*)}$ must lie in the interval $[0, d_{(k^*)}]$. We verify each bound. For the lower bound, $\delta^*_{(k^*)} \geq 0$ requires:

$$n_1\psi - C(k^*{+}1) \geq 0 \quad \Longleftrightarrow \quad n_1\psi \geq C(k^*{+}1). \tag{S119}$$

For the upper bound, $\delta^*_{(k^*)} \leq d_{(k^*)}$ requires:

$$n_1\psi - C(k^*{+}1) \leq d_{(k^*)}. \tag{S120}$$

Adding $C(k^*{+}1)$ to both sides:

$$n_1\psi \leq d_{(k^*)} + C(k^*{+}1). \tag{S121}$$

By definition (37), $d_{(k^*)} + C(k^*{+}1) = d_{(k^*)} + \sum_{\ell=k^*+1}^{n_1} d_{(\ell)} = \sum_{\ell=k^*}^{n_1} d_{(\ell)} = C(k^*)$, so:

$$n_1\psi \leq C(k^*). \tag{S122}$$

Combining (S119) and (S122), the feasibility requirement is:

$$C(k^*+1) \le n_1\psi \le C(k^*). \tag{S123}$$

The left inequality is strict: if $n_1\psi = C(k^*+1)$, then $\delta^*_{(k^*)} = 0$ by (S118), which would place unit (k^*) in the zero group. The effective threshold would then shift to k^*-1, contradicting the definition of k^* as the boundary index. Therefore:

$$C(k^*+1) < n_1\psi \le C(k^*). \tag{S124}$$

We now show that such an index k^* exists and is unique. The function $C(k)$ is strictly decreasing in k, as established above:

$$C(1) > C(2) > \cdots > C(n_1) > C(n_1+1) = 0. \tag{S125}$$

At the endpoints:

$$C(1) = n_1\bar{d} \ge n_1\psi \quad \text{(by feasibility: } \psi \le \bar{d}\text{)}, \tag{S126}$$

$$C(n_1+1) = 0 \le n_1\psi \qquad \text{(by feasibility: } \psi \ge 0\text{)}. \tag{S127}$$

Since $C(\cdot)$ is a strictly decreasing integer-indexed function with $C(1) \ge n_1\psi$ and $C(n_1+1) \le n_1\psi$, there is exactly one index $k^* \in \{1, \ldots, n_1\}$ at which $n_1\psi$ crosses below $C(k^*)$:

$$C(k^*+1) < n_1\psi \le C(k^*). \tag{S128}$$

This is precisely the threshold condition in the theorem statement.

It remains to verify that the constructed δ^* satisfies all KKT conditions. Set the mean constraint multiplier to:

$$\lambda^* = n_1 \cdot w_{(k^*)}. \tag{S129}$$

For units above the threshold ($j > k^*$), set:

$$\mu^{*\mathrm{L}}_{(j)} = 0, \tag{S130}$$

$$\mu^{*\mathrm{U}}_{(j)} = w_{(j)} - w_{(k^*)}. \tag{S131}$$

Since $w_{(j)} > w_{(k^*)}$ for $j > k^*$, we have $\mu^{*\mathrm{U}}_{(j)} > 0$, so dual feasibility holds. Checking stationarity (a) using (S129):

$$\begin{aligned} w_{(j)} - \frac{\lambda^*}{n_1} + \mu^{*\mathrm{L}}_{(j)} - \mu^{*\mathrm{U}}_{(j)} &= w_{(j)} - w_{(k^*)} + 0 - \big(w_{(j)} - w_{(k^*)}\big) \\ &= 0. \end{aligned} \tag{S132}$$

Checking complementary slackness (b) with $\delta^*_{(j)} = d_{(j)}$:

$$\mu^{*\mathrm{L}}_{(j)} \cdot \delta^*_{(j)} = 0 \cdot d_{(j)} = 0, \tag{S133}$$

$$\mu^{*\mathrm{U}}_{(j)} \cdot \big(d_{(j)} - \delta^*_{(j)}\big) = \mu^{*\mathrm{U}}_{(j)} \cdot \big(d_{(j)} - d_{(j)}\big) = 0. \tag{S134}$$

For units below the threshold ($j < k^*$), set:

$$\mu^{*\mathrm{U}}_{(j)} = 0, \tag{S135}$$

$$\mu^{*\mathrm{L}}_{(j)} = w_{(k^*)} - w_{(j)}. \tag{S136}$$

Since $w_{(j)} < w_{(k^*)}$ for $j < k^*$, we have $\mu^{*\mathrm{L}}_{(j)} > 0$, so dual feasibility holds. Checking stationarity (a):

$$\begin{aligned} w_{(j)} - \frac{\lambda^*}{n_1} + \mu^{*\mathrm{L}}_{(j)} - \mu^{*\mathrm{U}}_{(j)} &= w_{(j)} - w_{(k^*)} + \big(w_{(k^*)} - w_{(j)}\big) - 0 \\ &= 0. \end{aligned} \tag{S137}$$

Checking complementary slackness (b) with $\delta^*_{(j)} = 0$:

$$\mu^{*\mathrm{L}}_{(j)} \cdot \delta^*_{(j)} = \mu^{*\mathrm{L}}_{(j)} \cdot 0 = 0, \tag{S138}$$

$$\mu^{*\mathrm{U}}_{(j)} \cdot \big(d_{(j)} - \delta^*_{(j)}\big) = 0 \cdot d_{(j)} = 0. \tag{S139}$$

For the threshold unit ($j = k^*$), set:

$$\mu^{*\mathrm{L}}_{(k^*)} = 0, \qquad \mu^{*\mathrm{U}}_{(k^*)} = 0. \tag{S140}$$

Checking stationarity (a):

$$\begin{aligned} w_{(k^*)} - \frac{\lambda^*}{n_1} + 0 - 0 &= w_{(k^*)} - w_{(k^*)} \\ &= 0. \end{aligned} \tag{S141}$$

Checking complementary slackness (b):

$$\mu^{*\mathrm{L}}_{(k^*)} \cdot \delta^*_{(k^*)} = 0 \cdot \delta^*_{(k^*)} = 0, \tag{S142}$$

$$\mu^{*\mathrm{U}}_{(k^*)} \cdot \big(d_{(k^*)} - \delta^*_{(k^*)}\big) = 0 \cdot \big(d_{(k^*)} - \delta^*_{(k^*)}\big) = 0. \tag{S143}$$

Primal feasibility was established by construction: box constraints hold for all three groups, and the mean constraint is satisfied by (S118). Dual feasibility holds since all multipliers are nonnegative by (S130)–(S131), (S135)–(S136), and (S140). Since KKT conditions are sufficient for optimality in a linear program (Lemma S6), the vector δ^* defined in (39) is optimal for the maximization (35).

For part (II), observe that the minimization $\min_{\delta \in \mathcal{F}_2^\delta} \sum w_j \delta_j$ is equivalent to $-\max_{\delta \in \mathcal{F}_2^\delta} \sum(-w_j)\delta_j$. Applying part (I) with weights $\tilde{w}_j = -w_j$ reverses the sorted order:

$$\tilde{w}_{(1)} = -w_{(n_1)} \;<\; \tilde{w}_{(2)} = -w_{(n_1-1)} \;<\; \cdots \;<\; \tilde{w}_{(n_1)} = -w_{(1)}. \tag{S144}$$

The maximization of $\sum \tilde{w}_j \delta_j$ assigns $\delta_j = d_j$ to units with the largest $\tilde{w}_j$ (the smallest w_j) and $\delta_j = 0$ to units with the smallest $\tilde{w}_j$ (the largest w_j). Translating back to the original weight ordering: the minimization of $\sum w_j \delta_j$ assigns $\delta^{**}_{(j)} = d_{(j)}$ for $j < k_*$ (low-weight units

at capacity), $\delta^{**}_{(j)} = 0$ for $j > k_*$ (high-weight units at zero), and the threshold unit absorbs the residual:

$$\delta^{**}_{(k_*)} = n_1\psi - \sum_{\ell=1}^{k_*-1} d_{(\ell)}. \tag{S145}$$

The threshold condition $\sum_{\ell=1}^{k_*-1} d_{(\ell)} < n_1\psi \le \sum_{\ell=1}^{k_*} d_{(\ell)}$ follows by the same feasibility and uniqueness argument applied to the cumulative capacity from the bottom. □

Full proof of Theorem 4.10. We first establish the inner-product representation, then prove each part in turn.

The width is the difference of the sharp bounds:

$$\mathcal{W}(w) = \mathcal{B}_2^{\mathrm{U}}(w) - \mathcal{B}_2^{\mathrm{L}}(w) = \sum_{j=1}^{n_1} w_{(j)}\,\delta^{*}_{(j)} - \sum_{j=1}^{n_1} w_{(j)}\,\delta^{**}_{(j)}. \tag{S146}$$

Combining the two sums:

$$\mathcal{W}(w) = \sum_{j=1}^{n_1} w_{(j)}\big(\delta^{*}_{(j)} - \delta^{**}_{(j)}\big) = \sum_{j=1}^{n_1} w_{(j)}\,\Delta_{(j)}. \tag{S147}$$

Since both δ^* and δ^{**} belong to $\mathcal{F}_2^{\delta}$, they share the same mean:

$$\frac{1}{n_1}\sum_{j=1}^{n_1} \delta^{*}_{(j)} = \frac{1}{n_1}\sum_{j=1}^{n_1} \delta^{**}_{(j)} = \psi, \tag{S148}$$

so their difference sums to zero:

$$\sum_{j=1}^{n_1} \Delta_{(j)} = n_1\psi - n_1\psi = 0. \tag{S149}$$

Decomposing $w_{(j)} = \bar{w} + \tilde{w}_{(j)}$ in (S147) and applying (S149):

$$\begin{aligned}
\mathcal{W}(w) &= \sum_{j=1}^{n_1} \big(\bar{w} + \tilde{w}_{(j)}\big)\,\Delta_{(j)} \\
&= \bar{w}\sum_{j=1}^{n_1} \Delta_{(j)} + \sum_{j=1}^{n_1} \tilde{w}_{(j)}\,\Delta_{(j)} \\
&= \bar{w}\cdot 0 + \sum_{j=1}^{n_1} \tilde{w}_{(j)}\,\Delta_{(j)} \\
&= \sum_{j=1}^{n_1} \tilde{w}_{(j)}\,\Delta_{(j)}.
\end{aligned} \tag{S150}$$

This establishes the inner-product representation (41).

For part (i), the "if" direction is immediate: if $w_j = \bar{w}$ for all $j \in \mathcal{C}_1$, then $\tilde{w}_{(j)} = 0$ for every j, and:

$$\mathcal{W}(w) = \sum_{j=1}^{n_1} 0 \cdot \Delta_{(j)} = 0. \tag{S151}$$

For the "only if" direction, suppose the weights are not all equal. We show $\mathcal{W}(w) > 0$ when $0 < \psi < \bar{d}$. The width is the difference between the maximum and minimum of the linear function $f(\delta) = \sum_{j \in \mathcal{C}_1} w_j \delta_j$ over the polytope $\mathcal{F}_2^\delta$:

$$\mathcal{W}(w) = \max_{\delta \in \mathcal{F}_2^\delta} f(\delta) - \min_{\delta \in \mathcal{F}_2^\delta} f(\delta). \tag{S152}$$

This difference is zero if and only if f is constant on $\mathcal{F}_2^\delta$. A linear function $f(\delta) = w^\top \delta$ is constant on a convex polytope if and only if the gradient w is orthogonal to every feasible direction of the polytope (Bertsekas, 1999, Proposition 2.1.2). The polytope $\mathcal{F}_2^\delta$ is the intersection of the box $\prod_j [0, d_j]$ with the hyperplane $\{(1/n_1) \sum \delta_j = \psi\}$. When $0 < \psi < \bar{d}$, this intersection has nonempty interior relative to the hyperplane, and its feasible directions include all vectors of the form

$$v = \alpha(e_i - e_j), \quad i, j \in \mathcal{C}_1, \;\; i \neq j, \;\; \alpha > 0 \text{ sufficiently small}, \tag{S153}$$

where e_i is the i-th standard basis vector. These unit-transfer directions satisfy $\sum_\ell v_\ell = 0$ (they preserve the mean constraint) and, for α small enough, remain in the box. The gradient w is orthogonal to $e_i - e_j$ if and only if:

$$w^\top (e_i - e_j) = w_i - w_j = 0. \tag{S154}$$

For f to be constant on $\mathcal{F}_2^\delta$, we need (S154) to hold for all pairs $i \neq j$ in $\mathcal{C}_1$, which requires:

$$w_i = w_j \quad \text{for all } i, j \in \mathcal{C}_1. \tag{S155}$$

Since the weights are not all equal by assumption, there exist i, j with $w_i \neq w_j$, so f is not constant on $\mathcal{F}_2^\delta$, and:

$$\mathcal{W}(w) = \max f - \min f > 0. \tag{S156}$$

This establishes both the "only if" direction of part (i) and part (ii).

For part (iii), we apply Hölder's inequality with the ℓ^∞–ℓ^1 pairing to the inner-product representation (S150):

$$\mathcal{W}(w) = \left| \sum_{j=1}^{n_1} \tilde{w}_{(j)} \, \Delta_{(j)} \right| \leq \max_{1 \leq j \leq n_1} |\tilde{w}_{(j)}| \cdot \sum_{j=1}^{n_1} |\Delta_{(j)}|. \tag{S157}$$

We bound $\sum |\Delta_{(j)}|$ by exploiting the zero-sum property (S149). Decompose $\Delta_{(j)} = \Delta_{(j)}^+ - \Delta_{(j)}^-$ into positive and negative parts, where $\Delta_{(j)}^+ = \max(\Delta_{(j)}, 0)$ and $\Delta_{(j)}^- = \max(-\Delta_{(j)}, 0)$. From (S149):

$$\sum_{j=1}^{n_1} \Delta_{(j)}^+ = \sum_{j=1}^{n_1} \Delta_{(j)}^-, \tag{S158}$$

so:

$$\sum_{j=1}^{n_1} |\Delta_{(j)}| = \sum_{j=1}^{n_1} \Delta_{(j)}^{+} + \sum_{j=1}^{n_1} \Delta_{(j)}^{-} = 2 \sum_{j=1}^{n_1} \Delta_{(j)}^{+}. \tag{S159}$$

The positive part $\Delta_{(j)}^{+} = (\delta_{(j)}^{*} - \delta_{(j)}^{**})^{+}$ satisfies $\Delta_{(j)}^{+} \leq \delta_{(j)}^{*}$ (since $\delta_{(j)}^{**} \geq 0$). Summing, the total of the maximizer's allocation is $n_1\psi$ (the total spillover budget):

$$\sum_{j=1}^{n_1} \Delta_{(j)}^{+} \leq \sum_{j=1}^{n_1} \delta_{(j)}^{*} = n_1\psi. \tag{S160}$$

Similarly, $\Delta_{(j)}^{+} \leq d_{(j)} - \delta_{(j)}^{**}$ (since $\delta_{(j)}^{*} \leq d_{(j)}$). Summing, the positive shift also cannot exceed the remaining room at each unit (the gap between capacity and the minimizer's allocation):

$$\sum_{j=1}^{n_1} \Delta_{(j)}^{+} \leq \sum_{j=1}^{n_1} \left(d_{(j)} - \delta_{(j)}^{**}\right) = n_1\bar{d} - n_1\psi = n_1(\bar{d} - \psi). \tag{S161}$$

Combining (S160) and (S161):

$$\sum_{j=1}^{n_1} \Delta_{(j)}^{+} \leq n_1 \min(\psi,\ \bar{d} - \psi). \tag{S162}$$

Now that we bounded the second quantity, we can plug in everything into Hölder's inequality, substituting (S159) and (S162) into (S157):

$$\mathcal{W}(w) \leq \max_j |\tilde{w}_{(j)}| \cdot 2n_1 \min(\psi,\ \bar{d} - \psi). \tag{S163}$$

This establishes part (iii). The quantity $\max_j |\tilde{w}_{(j)}|$ is the weight dispersion: how far the most extreme weight deviates from the average. For uniform weights, this is zero, so width is zero and the bounds collapse. For non-uniform weights, we recover bounds. The factor $2n_1 \min(\psi,\ \bar{d} - \psi)$ is the budget flexibility: how much total spillover the adversary can redistribute. The quantity is maximized when $\psi = \bar{d}/2$, when the average spillover is in the middle of the feasible range, providing maximum room to move things around. It is zero when $\psi = 0$, as no spillovers exist to redistribute, or $\psi = \bar{d}$, when every unit is already at capacity and there is no room to shift anything. The width is small whenever either factor is small.

For part (iv), suppose $d_j = d$ for all $j \in \mathcal{C}_1$. We first show that under this condition the allocation differences $\Delta_{(j)}$ do not depend on the weight vector and are nondecreasing in j.

With homogeneous capacities, the cumulative capacity from the top is $C(k) = (n_1 - k + 1)d$, which depends only on k, n_1, and d. The threshold condition for the upper bound, $C(k^*+1) < n_1\psi \leq C(k^*)$, becomes:

$$(n_1 - k^*)d < n_1\psi \leq (n_1 - k^* + 1)d, \tag{S164}$$

which determines k^* as a function of $n_1\psi/d$ alone — it does not involve the weights. Analogously, the threshold for the lower bound is $k_* = \lceil n_1\psi/d \rceil$, and from (S164):

$$k^* = n_1 - \lceil n_1\psi/d \rceil + 1 = n_1 - k_* + 1. \tag{S165}$$

Since both thresholds depend only on n_1, ψ, and d, the optimal allocations δ^* and δ^{**} from Theorem 4.9 — and therefore the differences $\Delta_{(j)} = \delta^*_{(j)} - \delta^{**}_{(j)}$ — are independent of w.

We now show that $\Delta_{(j)}$ is nondecreasing in j under homogeneous capacities. Consider the three regions determined by k_* and $k^* = n_1 - k_* + 1$, noting that $k_* \le k^*$ when $k_* \le (n_1+1)/2$ (that is, when $\psi \le d/2$, the generic case; the argument for $k_* > k^*$ is analogous).

For $j < k_*$: the upper-bound allocation gives $\delta^*_{(j)} = 0$ (below the upper threshold) and the lower-bound allocation gives $\delta^{**}_{(j)} = d$ (below the lower threshold, hence at capacity). Therefore:

$$\Delta_{(j)} = 0 - d = -d \quad \text{for all } j < k_*. \tag{S166}$$

For $k_* \le j \le k^*$: both allocations assign intermediate or zero values. At the boundaries, $\Delta_{(k_*)}$ involves the lower-bound residual and $\Delta_{(k^*)}$ involves the upper-bound residual. The residual values lie in $[0, d]$, so $\Delta_{(j)} \in [-d, d]$ for these indices. Crucially, since $\delta^*_{(j)}$ transitions from 0 to d (passing through its residual) and $\delta^{**}_{(j)}$ transitions from d to 0 (passing through its residual) as j increases through this region, $\Delta_{(j)}$ transitions from values near $-d$ to values near d:

$$\Delta_{(k_*)} \in [-d, 0], \qquad \Delta_{(k^*)} \in [0, d]. \tag{S167}$$

For $j > k^*$: the upper-bound allocation gives $\delta^*_{(j)} = d$ (above the upper threshold, at capacity) and the lower-bound allocation gives $\delta^{**}_{(j)} = 0$ (above the lower threshold, at zero). Therefore:

$$\Delta_{(j)} = d - 0 = d \quad \text{for all } j > k^*. \tag{S168}$$

Combining (S166)–(S168): the sequence $\Delta_{(1)}, \ldots, \Delta_{(n_1)}$ moves from $-d$ (for small j) through intermediate values in $[-d, d]$ to d (for large j), and is nondecreasing throughout:

$$\Delta_{(1)} \le \Delta_{(2)} \le \cdots \le \Delta_{(n_1)}. \tag{S169}$$

With Δ independent of w and nondecreasing, the width (S150) becomes a linear function of the sorted weight vector:

$$\mathcal{W}(w) = \sum_{j=1}^{n_1} \Delta_{(j)}\, w_{(j)}, \tag{S170}$$

with coefficients $\Delta_{(1)} \le \Delta_{(2)} \le \cdots \le \Delta_{(n_1)}$. By a standard result in the theory of majorization (Marshall et al., 2011, Proposition C.1), a function of the form $\phi(x) = \sum_{j=1}^{n} c_{(j)} x_{(j)}$, where $c_{(1)} \le \cdots \le c_{(n)}$ and $x_{(1)} \le \cdots \le x_{(n)}$ denotes the sorted values of x, is Schur-convex. Applying this to (S170):

$$w' \succ w \quad \Longrightarrow \quad \mathcal{W}(w') \ge \mathcal{W}(w), \tag{S171}$$

where $\succ$ denotes the majorization partial order (that is, w' is a mean-preserving spread of w on the exposed controls). This establishes part (iv). □

Proof. Part (I) is Theorem S8. Parts (II) and (III) are Theorems S9 and S10. Part (IV) combines Corollary 4.5 with Propositions 4.7 and S17. Part (V) is Theorem 4.10(i). Sharpness follows from Proposition S11 and the attainment results in each theorem. □

Full proof of Lemma 5.1. We prove the result for the upper bound; the lower bound follows by an analogous argument.

The breakpoints of the threshold index k^* as a function of ψ occur at the values $\psi = C(k)/n_1$ for $k = 1, \ldots, n_1$, which are the points where $n_1\psi$ equals a cumulative capacity value. Denote these breakpoints in increasing order as $\psi_1 < \psi_2 < \cdots < \psi_{n_1}$, where $\psi_j = C(n_1 - j + 1)/n_1$ (the cumulative capacities $C(1) > C(2) > \cdots > C(n_1)$ map to increasing ψ values). For ψ in the open interval (ψ_{j-1}, ψ_j) (with $\psi_0 = 0$ and $\psi_{n_1} = \bar{d}$ by convention), the threshold index k^* is constant, equal to some value that we denote k_j^*.

On each such interval, the bound formula from Theorem S9(I) gives:

$$\mathcal{B}_2^{\mathrm{U}}(w;\psi) = \underbrace{\sum_{\ell=k_j^*+1}^{n_1} w_{(\ell)}\, d_{(\ell)}}_{A_j \text{ (constant in } \psi)} + w_{(k_j^*)}\Big[n_1\psi - C(k_j^*{+}1)\Big]. \tag{S172}$$

The first term $A_j = \sum_{\ell=k_j^*+1}^{n_1} w_{(\ell)} d_{(\ell)}$ does not depend on ψ (it involves only observed capacities and weights for the fixed set of above-threshold units). The second term is linear in ψ. Therefore $\mathcal{B}_2^{\mathrm{U}}(w;\psi)$ is affine in ψ on each interval, with slope:

$$\left.\frac{\partial \mathcal{B}_2^{\mathrm{U}}}{\partial \psi}\right|_{\text{piece } j} = n_1\, w_{(k_j^*)}. \tag{S173}$$

This establishes part (i) (piecewise linearity) and part (ii) (slopes).

For part (iii), observe that as ψ increases across successive intervals, the threshold index k^* decreases: a larger spillover budget $n_1\psi$ exceeds more cumulative capacities $C(k)$, pushing k^* to smaller values. Since the weights are sorted as $w_{(1)} < w_{(2)} < \cdots < w_{(n_1)}$, a smaller k^* gives a smaller weight $w_{(k^*)}$ and hence a smaller slope $n_1 w_{(k^*)}$. The slopes are therefore decreasing across successive linear pieces:

$$\text{piece } j \text{ slope} = n_1\, w_{(k_j^*)} \;>\; n_1\, w_{(k_{j+1}^*)} = \text{piece } (j+1) \text{ slope}, \tag{S174}$$

since $k_j^* > k_{j+1}^*$ and the weights are strictly increasing. A piecewise linear function with decreasing slopes is concave.

For the continuity claim in part (iv), we verify that the bound values match at each breakpoint $\psi = C(k^*)/n_1$. Approaching from the left (where the threshold is k^* and the adjusting unit (k^*) has value $\delta_{(k^*)} = n_1\psi - C(k^*{+}1)$), at $n_1\psi = C(k^*)$:

$$\begin{aligned} \delta_{(k^*)} &= C(k^*) - C(k^*{+}1) \\ &= d_{(k^*)}. \end{aligned} \tag{S175}$$

So the adjusting unit reaches its full capacity. The bound value from the left is:

$$\begin{aligned} \mathcal{B}_2^{\mathrm{U}}(w; C(k^*)/n_1^-) &= \sum_{\ell=k^*+1}^{n_1} w_{(\ell)}\, d_{(\ell)} + w_{(k^*)}\, d_{(k^*)} \\ &= \sum_{\ell=k^*}^{n_1} w_{(\ell)}\, d_{(\ell)}. \end{aligned} \tag{S176}$$

Approaching from the right (where the threshold shifts to k^*-1 and the adjusting unit (k^*-1) has value $\delta_{(k^*-1)} = n_1\psi - C(k^*)$), at $n_1\psi = C(k^*)$:

$$\begin{aligned}\delta_{(k^*-1)} &= C(k^*) - C(k^*) \\ &= 0. \end{aligned} \tag{S177}$$

So the new adjusting unit has zero contribution. The bound value from the right is:

$$\begin{aligned}\mathcal{B}_2^{\mathrm{U}}(w; C(k^*)/n_1^+) &= \sum_{\ell=k^*}^{n_1} w_{(\ell)}\, d_{(\ell)} + w_{(k^*-1)} \cdot 0 \\ &= \sum_{\ell=k^*}^{n_1} w_{(\ell)}\, d_{(\ell)}. \end{aligned} \tag{S178}$$

Comparing (S176) and (S178): the left and right limits are equal, confirming continuity at the breakpoint. The mechanism is transparent: the threshold jump (from k^* to k^*-1) is exactly compensated by the adjusting unit's value moving from its full capacity to zero, leaving the total bound value unchanged.

For the Lipschitz constant: on each linear piece, the absolute slope is $n_1\, w_{(k_j^*)} \le n_1\, w_{(n_1)}$, since $w_{(k_j^*)} \le w_{(n_1)}$ for any threshold index. The maximum slope is $n_1\, w_{(n_1)}$, achieved on the first piece where $k^* = n_1$. For any $\psi, \psi' \in [0, \bar{d}]$, the piecewise linear function satisfies:

$$\begin{aligned}\left|\mathcal{B}_2^{\mathrm{U}}(w;\psi') - \mathcal{B}_2^{\mathrm{U}}(w;\psi)\right| &\le \max_j\ n_1\, w_{(k_j^*)}\, |\psi' - \psi| \\ &= n_1\, w_{(n_1)}\, |\psi' - \psi|. \end{aligned} \tag{S179}$$

The first inequality holds because $\mathcal{B}_2^{\mathrm{U}}(w;\cdot)$ is continuous and piecewise linear with absolute slope bounded by $n_1\, w_{(n_1)}$ on each piece. For any $\psi, \psi' \in [0, \bar{d}]$ with $\psi < \psi'$, the derivative exists almost everywhere and satisfies $|\mathrm{d}\mathcal{B}_2^{\mathrm{U}}/\mathrm{d}\psi| \le n_1\, w_{(n_1)}$, so by the fundamental theorem of calculus for absolutely continuous functions:

$$\left|\mathcal{B}_2^{\mathrm{U}}(w;\psi') - \mathcal{B}_2^{\mathrm{U}}(w;\psi)\right| = \left|\int_{\psi}^{\psi'} \frac{\mathrm{d}\mathcal{B}_2^{\mathrm{U}}}{\mathrm{d}t}\,\mathrm{d}t\right| \le \int_{\psi}^{\psi'} \left|\frac{\mathrm{d}\mathcal{B}_2^{\mathrm{U}}}{\mathrm{d}t}\right| \mathrm{d}t \le n_1\, w_{(n_1)}\,(\psi' - \psi). \tag{S180}$$

This establishes part (iv).

The argument for the lower bound is analogous: $\mathcal{B}_2^{\mathrm{L}}(w;\psi)$ is piecewise linear with slopes $n_1\, w_{(k_*)}$ that increase as ψ increases (since k_* increases with ψ, and larger k_* gives larger $w_{(k_*)}$). The function is convex, continuous at each breakpoint by the same compensating argument, and Lipschitz with the same constant $n_1\, w_{(n_1)}$ (the maximum slope, achieved at the last piece where $k_* = n_1$). □

Full proof of Proposition 5.2. For part (i): by Lemma 5.1(iv), the mapping $\psi \mapsto \mathcal{B}_2^{\mathrm{U}}(w;\psi)$ is Lipschitz-continuous on $[0, \bar{d}]$ with constant $L = n_1 w_{(n_1)}$. Therefore:

$$\left|\hat{\mathcal{B}}_2^{\mathrm{U}}(w) - \mathcal{B}_2^{\mathrm{U}}(w)\right| = \left|\mathcal{B}_2^{\mathrm{U}}(w;\hat{\psi}_{\mathrm{DR}}) - \mathcal{B}_2^{\mathrm{U}}(w;\psi)\right| \le n_1\, w_{(n_1)}\left|\hat{\psi}_{\mathrm{DR}} - \psi\right|. \tag{S181}$$

By Proposition 3.3, $\hat{\psi}_{\mathrm{DR}} \xrightarrow{p} \psi$ under the stated conditions. Combining with (S181): for any $\epsilon > 0$,

$$\begin{aligned} P\Big(\big|\hat{\mathcal{B}}_2^{\mathrm{U}} - \mathcal{B}_2^{\mathrm{U}}\big| > \epsilon\Big) &\leq P\Big(n_1\, w_{(n_1)}\, \big|\hat{\psi}_{\mathrm{DR}} - \psi\big| > \epsilon\Big) \\ &= P\Big(\big|\hat{\psi}_{\mathrm{DR}} - \psi\big| > \frac{\epsilon}{n_1\, w_{(n_1)}}\Big) \\ &\to 0 \end{aligned} \tag{S182}$$

as $n \to \infty$, since $\hat{\psi}_{\mathrm{DR}} \xrightarrow{p} \psi$. This establishes $\hat{\mathcal{B}}_2^{\mathrm{U}}(w) \xrightarrow{p} \mathcal{B}_2^{\mathrm{U}}(w)$. The same argument applies to the lower bound, with the Lipschitz constant $n_1\, w_{(n_1)}$ from Lemma 5.1(iv).

For part (ii): by Proposition S16, the Level 3 bounds decompose as $\mathcal{B}_3^{\mathrm{U}}(w) = \sum_{k=1}^{K} \mathcal{B}_{3,k}^{\mathrm{U}}(w)$, where each stratum contribution $\mathcal{B}_{3,k}^{\mathrm{U}}(w)$ is a Level 2 bound within stratum k with detection estimand $\tau(x_k)$. By Lemma 5.1 applied within stratum k, the mapping $\tau(x_k) \mapsto \mathcal{B}_{3,k}^{\mathrm{U}}(w; \tau(x_k))$ is Lipschitz with constant $n_k\, w_{k,(n_k)}$, where $w_{k,(n_k)}$ is the largest weight within stratum k. Therefore:

$$\begin{aligned} \big|\hat{\mathcal{B}}_3^{\mathrm{U}} - \mathcal{B}_3^{\mathrm{U}}\big| &= \left|\sum_{k=1}^{K} \Big[\mathcal{B}_{3,k}^{\mathrm{U}}(w; \hat{\tau}(x_k)) - \mathcal{B}_{3,k}^{\mathrm{U}}(w; \tau(x_k))\Big]\right| \\ &\leq \sum_{k=1}^{K} \big|\mathcal{B}_{3,k}^{\mathrm{U}}(w; \hat{\tau}(x_k)) - \mathcal{B}_{3,k}^{\mathrm{U}}(w; \tau(x_k))\big| \\ &\leq \sum_{k=1}^{K} n_k\, w_{k,(n_k)}\, \big|\hat{\tau}(x_k) - \tau(x_k)\big|. \end{aligned} \tag{S183}$$

Since $\hat{\tau}(x_k) \xrightarrow{p} \tau(x_k)$ for each $k = 1, \ldots, K$ (a finite number of strata), the right-hand side of (S183) converges to zero in probability:

$$\sum_{k=1}^{K} n_k\, w_{k,(n_k)}\, \big|\hat{\tau}(x_k) - \tau(x_k)\big| \xrightarrow{p} 0, \tag{S184}$$

which establishes $\hat{\mathcal{B}}_3^{\mathrm{U}}(w) \xrightarrow{p} \mathcal{B}_3^{\mathrm{U}}(w)$. The argument for the lower bounds at both levels is identical. □

S4 Estimation and Inference Details

This section collects the estimation and inference details that Section 5 of the main text defers. I record in full the weight-concentration remark that governs when the delta-method standard error is trustworthy, treat the non-generic parameter values at which the bound mapping has a kink, derive the Level 3 covariance from the per-stratum influence functions, and set out the complete implementation recipe.

S4.1 Weight Concentration and Standard-Error Calibration

The generic-case delta-method standard error of Section 5 of the main text is well calibrated for near-uniform weights and can fail for concentrated weights. The following remark records the pattern in full.

Remark S1 (Weight concentration and standard-error calibration)**.** *The delta-method standard error of the generic-case approximation is well calibrated for uniform weights. For difference-in-differences the ratio of the analytic standard error to the Monte Carlo standard deviation falls in* $[1.01, 1.10]$ *across all simulation configurations. For staggered weights the lower-bound ratio is similarly well calibrated, between* 0.98 *and* 1.08*, but the upper-bound ratio ranges from* 0.72 *to* 1.06*, with undercalibration most pronounced under large-spillover configurations. For highly concentrated weights, as in synthetic control, the local slope* $n_1 w_{(k^*)}$ *can be dominated by a single unit's weight, and the delta-method standard error can substantially understate sampling variability. The simulation evidence in Section 6 of the main text confirms this pattern. For such designs the identification-region width dominates estimation uncertainty and the Imbens–Manski confidence interval remains correctly sized, but a practitioner using concentrated weights should read the pointwise standard error with caution. I recommend that formal inference, meaning confidence intervals and hypothesis tests based on the generic-case approximation, be restricted to difference-in-differences and staggered-adoption weights, where the delta method is well calibrated. For synthetic-control weights, where a single donor dominates the weight vector, the delta-method standard error is uninformative, and the bounds are best read as point estimates of the identification region rather than as the basis for a confidence interval. Formal inference under concentrated weights requires joint asymptotics of* $(\hat{\psi}, \hat{d})$ *that account for the effect of weight estimation on the capacity vector, a problem I leave to future work.*

S4.2 The Kink Case

At the non-generic values of ψ where $n_1\psi$ coincides with a cumulative capacity $C(k)$, the bound mapping has a kink. There it is Hadamard directionally differentiable but not differentiable, and the limiting distribution is non-standard (Fang and Santos, 2019). This set of values has Lebesgue measure zero in the parameter space, and the concavity of the upper bound ensures that the generic-case normal approximation, taken along the steeper slope, gives a conservative confidence interval even near a kink. I therefore recommend the generic-case approximation of Section 5 of the main text as the primary inference tool.

S4.3 Level 3 Covariance

For the Level 3 bounds the stratum decomposition gives, in the generic case,

$$\sqrt{n}\big(\hat{\mathcal{B}}_3^{\mathrm{U}} - \mathcal{B}_3^{\mathrm{U}}\big) = \sum_{k=1}^{K} s_k \cdot \sqrt{n}\big(\hat{\tau}(x_k) - \tau(x_k)\big) + o_p(1),$$

where $s_k = n_k w_{k,(k_k^*)}$ is the stratum-level slope. The asymptotic variance is

$$\sigma_{3,\mathrm{U}}^2 = \boldsymbol{s}'\,\Sigma_\tau\,\boldsymbol{s}, \qquad \Sigma_{\tau,k\ell} = \mathrm{Cov}\big(\hat{\tau}(x_k),\, \hat{\tau}(x_\ell)\big), \tag{S185}$$

and the lower bound uses the slopes $s_k = n_k w_{k,(k_{*,k})}$. When the per-stratum doubly robust estimator uses the full set of unexposed controls $\mathcal{C}_0$ for the inverse-propensity correction in every stratum, as is natural in the interference setting where $\mathcal{C}_0$ is the reference group for all strata, the off-diagonal terms $\Sigma_{\tau,k\ell}$ for $k \neq \ell$ are in general nonzero, because the shared $\mathcal{C}_0$ contributions induce positive cross-stratum covariance. Under correct specification of at least one nuisance model this dependence is $o(n^{-1})$ and the variance reduces to $\sum_{k=1}^{K} s_k^2 \, \sigma_{\tau_k}^2$, but in finite samples the cross-stratum terms can be substantial.

I recommend estimating Σ_τ from the per-unit influence-function contributions. For stratum k, the influence function of the per-stratum doubly robust estimator $\hat{\tau}(x_k)$ evaluated at unit $j \in \mathcal{C}$ is

$$\varphi_k(Z_j) = \begin{cases} n_k^{-1}\big(\Delta Y_j - \hat{\mu}_0(X_j) - \hat{\tau}(x_k)\big) & \text{if } j \in S_k, \\ -n_k^{-1} \dfrac{\hat{\pi}_j}{1-\hat{\pi}_j}\big(\Delta Y_j - \hat{\mu}_0(X_j)\big) & \text{if } j \in \mathcal{C}_0, \\ 0 & \text{otherwise.} \end{cases} \tag{S186}$$

The covariance estimator is $\hat{\Sigma}_{\tau,k\ell} = \frac{n_{\mathcal{C}}}{n_{\mathcal{C}}-1} \sum_{j \in \mathcal{C}} \varphi_k(Z_j)\, \varphi_\ell(Z_j)$, and the standard error of the Level 3 upper bound is $\hat{\sigma}_{3,\mathrm{U}} = (\boldsymbol{s}'\, \hat{\Sigma}_\tau\, \boldsymbol{s})^{1/2}$. The unexposed-unit case in (S186) is the source of the cross-stratum covariance, because each unit $j \in \mathcal{C}_0$ contributes to the influence function of every stratum, scaled by n_k^{-1}. Omitting these terms, as one would by estimating the within-stratum variance from within-stratum residuals alone, can understate the variance by a factor of three to eight in realistic sample sizes.

S4.4 Implementation Recipe

The complete pipeline takes treatment assignments, outcomes, spatial locations, and covariates as input, and returns the contamination bounds with confidence intervals and a sensitivity analysis.

1. Specify the ATT estimator and extract its weight vector w on the control group. Specify the outcome-support lower bound a.

2. Choose an exposure radius r, or a grid of radii for the sensitivity analysis. Classify controls into exposed $\mathcal{C}_1(r)$ and unexposed $\mathcal{C}_0(r)$, and compute the capacities $d_j = \Delta Y_j^{\mathrm{obs}} - a$.

3. Estimate $\hat{\psi}_{\mathrm{DR}}(r)$ by the doubly robust procedure. For parametric nuisance models such as logit and linear regression the standard plug-in estimator suffices, and for flexible estimators such as random forests or the LASSO use cross-fitting (Chernozhukov et al., 2018). Check feasibility, that $0 \leq \hat{\psi}_{\mathrm{DR}} \leq \bar{d}$.

4. Compute the Level 2 bounds from Theorem 4.11(II) of the main text: sort the exposed controls by weight, form the cumulative capacities, find the threshold indices, and evaluate the closed-form expressions. The computational cost is $O(n_1 \log n_1)$.

5. *(Optional, Level 3.)* Construct K covariate strata on $\mathcal{C}_1$, estimate $\hat{\tau}(x_k)$ for each stratum, and compute the Level 3 bounds from Theorem 4.11(III) of the main text.

6. Construct the confidence intervals. For Level 2, compute the influence-function values $\varphi(Z_j)$ of $\hat{\psi}_{\mathrm{DR}}$ from Section 5 of the main text, estimate $\hat{\sigma}_\psi = (n_{\mathcal{C}}^{-1} \sum_{j \in \mathcal{C}} \varphi(Z_j)^2)^{1/2} / \sqrt{n_{\mathcal{C}}}$, and apply the delta method to obtain $\hat{\sigma}_{\mathrm{U}} = |n_1 w_{(k^*)}| \, \hat{\sigma}_\psi$ and $\hat{\sigma}_{\mathrm{L}} = |n_1 w_{(k_*)}| \, \hat{\sigma}_\psi$. For Level 3, compute the stratum-specific influence functions $\varphi_k(Z_j)$ from (S186), estimate the full covariance $\hat{\Sigma}_\tau$, and form $\hat{\sigma}_{3,\mathrm{U}} = (\boldsymbol{s}' \hat{\Sigma}_\tau \boldsymbol{s})^{1/2}$. Solve the Imbens–Manski critical-value equation of the main text for $c_n(\alpha)$ and report the interval $\mathrm{CI}_{1-\alpha}$.

7. *(Optional.)* Repeat over the radius grid $r_1 < \cdots < r_m$ and display the sensitivity band.

S5 Extended Discussion

This section collects two remarks that the main text refers to but does not prove and a comparison of the Level 1 bounds with the Horowitz-Manski benchmark.

Remark S2 (Equivalence of ATE and ATT of Exposure)**.** *The detection estimand ψ averages over all controls, while $\mathcal{B}(w)$ involves exposed controls specifically. Let $\psi_{\mathrm{ATT}} = \mathbb{E}[\Delta Y_j(0,1) - \Delta Y_j(0,0) \mid E_j = 1, A_j = 0]$. Under exposure ignorability, $\psi = \psi_{\mathrm{ATT}}$: the conditional distribution of spillover effects given X_j does not depend on E_j. This ensures that ψ, which averages over the entire control group and is more precisely estimable, can be used directly to constrain the contamination.*

Remark S3 (Tied weights)**.** *Theorem 4.9 assumes distinct weights for uniqueness of the optimal allocation; we now formalize the observation above regarding tied weights. When $w_j = w_{j'}$ for some $j \neq j'$, the LP objective $\sum_{j \in \mathcal{C}_1} w_j \delta_j$ is invariant to how spillover is redistributed among units sharing the same weight, so the sharp bounds $\mathcal{B}_2^U(w)$ and $\mathcal{B}_2^L(w)$ remain unique even though the optimal allocation δ^* is not. The sorting characterization extends by breaking ties in the weight ordering arbitrarily: the threshold index k^* may depend on the chosen tie-breaking rule, but the closed-form expressions* (39)–(40) *yield the same bound values under any permutation of tied units. Under DiD, all exposed controls receive equal weight $w_j = 1/n_1$, so every feasible allocation achieves the same objective value and the bounds collapse to the point $\bar{w} \cdot n_1 \psi$. For synthetic control and staggered designs, exact ties among the continuous-valued weights are measure-zero events.*

Proposition S20 (Comparison to Horowitz-Manski Bounds)**.** *Under the same outcome support $\Delta Y \in [a, b]$ and contamination fraction $\rho = n_1/|\mathcal{C}|$, the Horowitz-Manski bounds on the contamination from a fraction ρ of corrupted observations are*

$$\mathcal{B}^{\mathrm{HM}}(w) \in [0, \;\; W_{\exp} \cdot (b - a)] \, . \tag{S187}$$

The Level 1 bounds satisfy $\mathcal{B}_1^{\mathrm{U}}(w) \leq \mathcal{B}^{\mathrm{U,HM}}(w)$, with strict inequality whenever $\Delta Y_j^{\mathrm{obs}} < b$ for some $j \in \mathcal{C}_1$ with $w_j > 0$.

Proof. In the Horowitz-Manski framework, the contaminated observations (exposed controls, in our setting) have outcomes drawn from an unknown distribution with support $[a, b]$. Without exploiting the relationship between the contaminated and clean outcomes, the worst-case

spillover effect for each unit is $b - a$—the full range of the outcome support. Applying this uniform bound:

$$\mathcal{B}^{\mathrm{U,HM}}(w) = \sum_{j \in \mathcal{C}_1} w_j (b - a) = W_{\exp} \cdot (b - a).$$

In our setting, the interference structure provides additional information. For each exposed control j, we observe $\Delta Y_j^{\mathrm{obs}} = \Delta Y_j(0, 1)$, and we know that under monotone nonnegative spillovers the clean counterfactual satisfies $\Delta Y_j(0, 0) \leq \Delta Y_j^{\mathrm{obs}}$. The maximum spillover for unit j is therefore $\Delta Y_j^{\mathrm{obs}} - a = d_j$, which is weakly smaller than $b - a$ because $\Delta Y_j^{\mathrm{obs}} \leq b$ by outcome support. The Level 1 upper bound is

$$\mathcal{B}_1^{\mathrm{U}}(w) = \sum_{j \in \mathcal{C}_1} w_j d_j = \sum_{j \in \mathcal{C}_1} w_j (\Delta Y_j^{\mathrm{obs}} - a).$$

Since $d_j = \Delta Y_j^{\mathrm{obs}} - a \leq b - a$ for each j, and $w_j \geq 0$, we have $w_j d_j \leq w_j (b - a)$ term by term, so $\mathcal{B}_1^{\mathrm{U}}(w) \leq \mathcal{B}^{\mathrm{U,HM}}(w)$.

For strict inequality: if $\Delta Y_j^{\mathrm{obs}} < b$ for some j with $w_j > 0$, then $w_j d_j < w_j (b - a)$ for that unit, and the sum is strictly smaller. □

S6 Monte Carlo: Full Results

These simulations establish three results that the theory alone does not make obvious. First, the weight structure of the ATT estimator is the dominant determinant of how informative the bounds are. Difference-in-differences achieves exact point identification of the contamination in every configuration, while synthetic-control weights with comparable sample sizes yield identification regions no tighter than the support-only benchmark. This confirms the width theorem (Theorem 4.11(iv) of the main text) quantitatively and shows that the three-way split into point identification, informative partial identification, and uninformative partial identification is not an artifact of limiting arguments but appears sharply at sample sizes as small as $N = 500$. Second, the doubly robust estimator $\hat{\psi}_{\mathrm{DR}}$ is well calibrated and genuinely doubly robust. Under single-model misspecification the bias is negligible and $\sqrt{N}$-consistency holds, and under double misspecification the estimator is $\sqrt{N}$-inconsistent, which shows that the DR property is structural rather than an artifact of the DGP. Third, the Imbens–Manski confidence intervals achieve at least nominal coverage in every cell of the design, including configurations where the SE calibration is poor, because the partial-identification uncertainty dominates the estimation uncertainty for all non-uniform weight types.

I evaluate the finite-sample performance of the framework through an extensive Monte Carlo study that addresses four questions. First, do the bounds tighten progressively from Level 1 through Level 3 as the theory predicts, and does the weight structure govern the degree of tightening? Second, are the Imbens–Manski confidence intervals correctly sized? Third, does the doubly robust estimator protect against single-model misspecification? Fourth, how do the proposed bounds compare with Manski worst-case bounds and with the naive strategy of ignoring interference? Every cell uses 5,000 replications, which fixes the Monte Carlo standard error on a coverage rate at 0.0031.

S6.1 Design

S6.1.1 Data-Generating Process

Each replication generates N units on a unit square with coordinates drawn uniformly. A fraction ρ of units are assigned to treatment through a logistic model conditional on covariates $X_i = (X_{i1}, X_{i2})' \sim \mathcal{N}(0, I_2)$. Control units within distance $r = 1$ of any treated unit are classified as exposed ($\mathcal{C}_1$), and the remainder are unexposed ($\mathcal{C}_0$). Potential outcomes under no exposure follow $\Delta Y_j(0,0) = \mu_0(X_j) + \varepsilon_j$ with $\varepsilon_j \sim \mathcal{N}(0, \sigma^2)$. Spillover effects are additive, $\Delta Y_j(0,1) = \Delta Y_j(0,0) + \tau(X_j)$, where $\tau(X_j)$ governs the unit-level spillover magnitude. In all configurations $\tau(X_j) \geq 0$, which satisfies the monotone nonnegative spillover assumption. Since the Normal outcome distribution has unbounded support, the analyst must specify a finite lower bound a on the support of ΔY. In applied work a would be determined from the natural lower bound of the outcome variable, for example zero for non-negative outcomes or a substantive floor on outcome changes. In the DGP I calibrate $a = -1$ so that essentially no clean counterfactuals fall below a. Under Config A, $\mathbb{P}(\Delta Y_j(0,0) < -1) < 0.005$, so the support restriction is effectively non-binding at the truth. The sensitivity analysis in Section S6.6 varies a to assess the consequences of conservative and aggressive choices.

I compute four weight types for each replication. Equal weights (DiD) assign $w_j = 1/|\mathcal{C}|$ to all controls. Sparse synthetic control (SC sparse) concentrates weight on approximately 20 donor units through a Dirichlet draw. Dense synthetic control (SC dense) spreads weight across roughly 500 donors. Staggered DiD assigns near-uniform weights with small perturbations that mimic cohort-time aggregation. Table 5 summarizes the weight structure under Config A at $N = 2000$. DiD yields an HHI of 0.0002 and $\max|\tilde{w}| \approx 0$, while SC sparse produces an HHI of 0.028 and $\max|\tilde{w}| = 0.118$, a 140-fold difference in concentration that drives the divergence in bound widths.

S6.1.2 Configurations

I employ eleven configurations, labeled A through K, that vary along three dimensions: the interference structure (ρ, $\psi/\bar{d}$), the treatment-effect heterogeneity (τ_x), and the specification of the nuisance models (μ_0, π).

- **Baseline (A):** $\rho = 0.30$, $\psi/\bar{d} = 0.25$, linear μ_0 and π. Reference case for all results.
- **Exposure fraction (B, C):** $\rho \in \{0.10, 0.50\}$. Tests bound behavior under sparse and dense exposure. Supplement only.
- **Spillover magnitude (D, E):** $\psi/\bar{d} \in \{0.05, 0.50\}$. Config D probes the detection boundary, and Config E is the maximum-width regime where the Hölder bound is tightest.
- **Heterogeneous spillovers (F):** $\tau_x = 2.0$. The only configuration where Level 3 stratification provides tightening beyond Level 2.
- **Misspecification factorial (G, H, I):** Config G misspecifies μ_0 (nonlinear outcome, linear fit), Config H misspecifies π (interaction in the true propensity, main-effects fit), and Config I misspecifies both. With Config A, these form the 2×2 DR factorial.
- **Null (J):** $\psi = 0$. No interference. This validates size control, since the bounds should collapse and the detection test should reject at the nominal rate.
- **High capacity heterogeneity (K):** $\sigma = 2.5$. Produces highly dispersed capacities

d_j and probes the limits of the Schur-convexity result (Theorem 4.11(iv) of the main text) when the homogeneity condition is violated.

Table 4 summarizes the parameter values for all eleven configurations.

The default sample size is $N = 2000$, and I report $N \in \{500, 1000, 2000\}$ for coverage and SE calibration.

S6.2 Progressive Tightening of the Bounds

The central prediction of the theory is that the identification region narrows as assumptions accumulate. Level 1 uses support restrictions alone, Level 2 adds the sorting characterization that exploits the mean constraint from $\hat{\psi}_{\mathrm{DR}}$, and Level 3 refines further through covariate stratification. Table 6 reports median bound widths across configurations and weight types at $N = 2000$.

Three patterns are immediate. First, DiD achieves point identification at Level 2 in every configuration. The median L_2 width equals 0.000, consistent with the theoretical result that uniform weights eliminate the partial-identification problem once ψ is known ($\mathcal{B}^{\mathrm{DiD}} = \rho \cdot \psi$). The tightening from L_1 to L_2 is 100% in all cases.

Second, staggered DiD achieves substantial but incomplete tightening. Under the baseline (Config A), the L_2 width is 0.378 (IQR 0.046), a 70.6% reduction from L_1. The tightening ranges from 62.5% in the large-spillover regime (Config E) to 98.4% under the null (Config J). The degree of tightening tracks the theory. The near-uniform weights of staggered DiD produce a small $\max|\tilde{w}|$ of 0.0009 (Table 5), which confines the identification region close to the point-identified DiD solution.

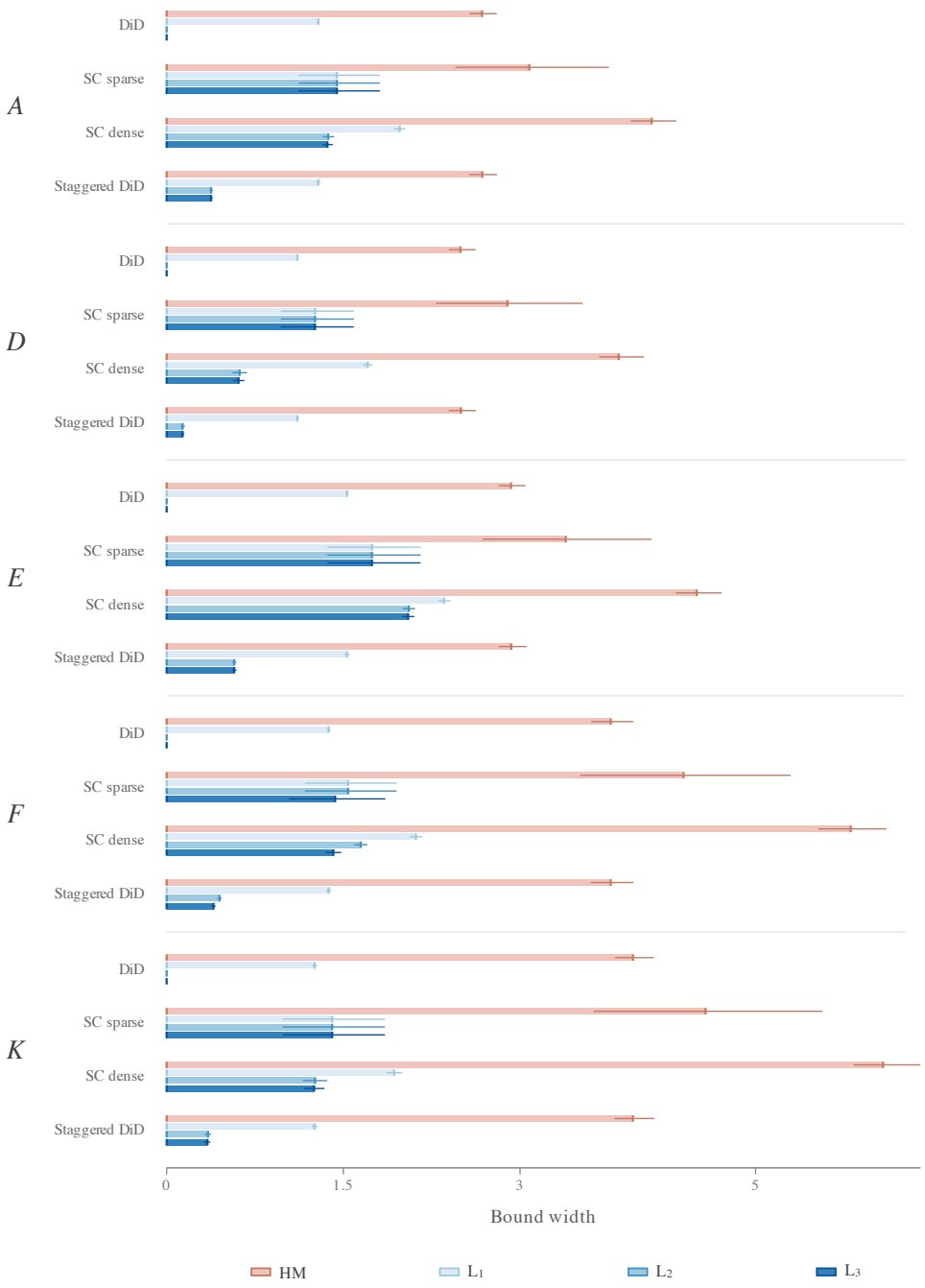
Third, SC sparse weights show negligible tightening from L_1 to L_2 in most configurations. Under Config A, the L_2 width is 1.447 (IQR 0.673), identical to L_1 to three decimal places, a 0.0% reduction. The concentration of weight on a small number of donors (HHI = 0.028, $\max|\tilde{w}| = 0.118$) means the adversary keeps almost full freedom to redistribute spillovers even after the mean constraint binds. This confirms the width theorem, since the identification problem is severe when the weight vector is concentrated.

The contamination itself is economically meaningful. Under the baseline, the true contamination is 25.4% of the ATT for DiD and staggered weights and 27.3% for SC sparse. Under Config E (large spillovers), it reaches 49.8–56.1% of the ATT. Table 79 documents that contamination exceeds 10% of the ATT in 100% of replications under Configs A, E, and F for DiD and staggered weights. For SC sparse weights the rate is 98.8% under Config A and 91.3% under Config F, which reflects the greater variability in contamination magnitude when weight is concentrated on few donors.

Table 7 reports the actual bound endpoints and confidence intervals, which makes the identification regions concrete. Under Config A with staggered weights, the estimated L_2 region is $[0.144, 0.523]$ with a 95% Imbens–Manski CI of $(0.117, 0.581)$, which brackets the true contamination of 0.254. The CI width relative to the ATT is 0.464, informative enough to bound the direction and approximate magnitude of the bias though not to point-identify it.

For Config F (heterogeneous spillovers), Level 3 stratification provides additional tightening. With staggered weights, the median L_3 width is 0.397 (IQR 0.057), an 11.9% reduction from the L_2 width of 0.451. Table 56 documents that the tightening increases with the

Figure 1: Empirical progressive tightening of the identification region. Median bound widths with interquartile whiskers, for five DGP configurations and four weight schemes, comparing the Horowitz–Manski support-only bound with the refinement levels L_1, L_2, and L_3, realizing the ranking by weight concentration predicted by the width theorem (Theorem 4.10 of the main text).



Note: Rows are grouped into five configuration blocks (Configs A, D, E, F, K of Table 4), each containing four weight-scheme rows (DiD, SC sparse, SC dense, Staggered DiD). Within each row, the top bar labelled HM is the Horowitz–Manski support-only bound $W_{\exp} \cdot d_{\max}$, and the three darker-blue bars labelled L_1, L_2, L_3 are the widths of the sharp Level-k identification regions $\mathcal{W}_k(w) = \mathcal{B}_k^{\mathrm{U}}(w) - \mathcal{B}_k^{\mathrm{L}}(w)$, ordered widest first. Bar lengths are medians across 5,000 Monte Carlo replications at $N = 5000$, and thin whiskers report the empirical interquartile range ($q_{0.25}$–$q_{0.75}$), suppressed for rows that collapse to zero width. DiD rows show no bar body at L_2 or L_3 because uniform weights yield exact point

number of strata K. At $K = 10$ the L_3 width falls to 0.381, a 15.5% reduction from L_2. Coverage stays at or above 0.999 across all values of K, which confirms that the finer stratification does not compromise inferential validity. The Jensen-predicted minimum tightening (Proposition S18) provides a conservative lower bound of 1.5% at $K = 10$ against the observed 15.5%. The gap reflects that the Jensen bound captures only the contribution of between-stratum heterogeneity to the budget-flexibility function, while the actual tightening also benefits from within-stratum homogeneity of capacities.

S6.3 Coverage and Standard Error Calibration

Table 8 reports coverage of the 95% Imbens–Manski confidence intervals and the ratio of analytic standard errors to Monte Carlo standard deviations across three sample sizes.

Coverage exceeds the nominal 95% level in every cell of the table. For DiD weights under Config A, coverage is 0.982 at $N = 500$ and 0.984 at $N = 2000$, close to the nominal level without substantial overcoverage. Staggered weights show mild overcoverage (1.000 across all N for Config A), consistent with the conservative nature of the Imbens–Manski construction for partially identified parameters. The interval covers the entire identification region, so when the true parameter lies in the interior the coverage exceeds the nominal rate.

The SE calibration ratios confirm that the analytic standard errors from the influence-function representation are well calibrated. For DiD weights, the SE ratio falls in $[1.01, 1.10]$ across all configurations and sample sizes, close to unity with a slight upward bias that produces conservative inference. For staggered weights, the upper-bound SE ratio ranges from 0.72 to 1.06, with undercalibration most pronounced under Config E (large spillovers, SE ratio as low as 0.72) and mild overcalibration under Config D (small spillovers, SE ratio up to 1.06). The lower-bound ratio is 0.98–1.08. Part of the apparent undercalibration for staggered weights reflects a conditional-versus-unconditional distinction. The analytic SE conditions on the realized weight vector, while the Monte Carlo SD averages over treatment-assignment realizations that produce different weight vectors. The definitive calibration test is CI coverage rather than the SE ratio. The remaining undercalibration of the upper-bound SE for staggered weights under large-spillover configurations does not compromise coverage, because the Imbens–Manski CI absorbs the partial-identification uncertainty, which dominates the estimation uncertainty in these cases. SC dense weights show intermediate SE calibration (upper-bound ratio 0.43–0.97), consistent with their intermediate position between staggered and SC sparse in weight concentration.

I use analytic standard errors throughout rather than the cluster bootstrap, and the rationale is empirical. The SE ratios in Table 8 show that the influence-function-based analytic SEs, which treat contributions as independent, match the Monte Carlo variability. The doubly robust influence function absorbs the dominant spatial dependence through the outcome model and exposure propensity, and the residual dependence among units separated by more than $2r$ is negligible at the sample sizes considered.

Two exceptions merit comment. For SC sparse weights, the SE ratio for the upper bound deteriorates sharply under Config D (small spillovers), from 0.73 at $N = 500$ to 0.06 at $N = 2000$. This happens because the L_2 upper bound for SC sparse sits near the L_1 boundary, which places the bound mapping close to a kink where the local slope, and hence the delta-method variance, approaches zero while the true sampling variability stays

positive. Coverage nonetheless remains at 1.000 because the CI width is dominated by the identification-region width ($L_2 \approx 1.270$), which dwarfs the estimation uncertainty. For Config K (high noise, $\sigma = 2.5$) the same pattern appears. DiD SE ratios stay well calibrated (1.01–1.02), while the SC sparse upper-bound ratios are essentially zero. In practice the SC sparse SE ratio is uninformative, because the sorting characterization provides no tightening for these weights and the CI reduces to the L_1 region augmented by a negligible estimation margin.

S6.4 Double Robustness

The doubly robust property guarantees consistency of $\hat{\psi}_{\text{DR}}$ when at least one of the outcome model $\mu_0(x)$ or the exposure propensity $\pi(x)$ is correctly specified. Table 9 evaluates this guarantee through a 2×2 misspecification factorial using SC sparse weights.

Under correct specification of both models (Config A), the DR estimator has bias below 0.002 in absolute value and RMSE 0.079 at $N = 2000$. When μ_0 alone is misspecified (Config G), the bias stays negligible at 0.000 with RMSE 0.092, because the correct propensity score protects the estimator. When π alone is misspecified (Config H), the bias is -0.001 with RMSE 0.097, because the correct outcome model protects the estimator. In both single-misspecification cases the $\sqrt{N}$-scaled bias stays bounded (Tables 18 and 19), which confirms $\sqrt{N}$-consistency.

The cross-fitted estimator $\hat{\psi}_{\text{CF}}$ provides an additional check. Under single misspecification (Configs G and H), it carries larger RMSE than the standard DR estimator, 0.253 against 0.092 for Config G and 0.333 against 0.097 for Config H, both at $N = 2000$. The sample-splitting cost inflates variance without reducing bias once the Donsker condition is already satisfied by the linear nuisance models. Table 71 confirms that under correct specification (Config A) both estimators have comparable bias, while the cross-fitted variant has roughly triple the standard deviation, 0.233 against 0.094 at $N = 2000$. I recommend the standard DR estimator when the nuisance models are parametric and cross-fitting when flexible estimators such as random forests are used.

Config I, where both models are misspecified, warrants careful discussion. The DR estimator bias at $N = 2000$ is -0.004 with RMSE 0.243, which is small. The scaled RMSE RMSE $\times \sqrt{N}$ grows from 10.2 at $N = 200$ to 11.0 at $N = 5{,}000$ (Table 20), which indicates that the RMSE does not shrink at the $\sqrt{N}$ rate. The estimator is $\sqrt{N}$-inconsistent, since it converges to a limit close to but not equal to the true ψ. The upper-bound bias is -0.025 at $N = 2000$, and coverage is 0.982, slightly below the off-diagonal cells but still above 95%. The mild coverage degradation, rather than a dramatic failure, reflects the specific DGP, where the misspecification is moderate (an omitted interaction term in the propensity and a linear approximation to a mildly nonlinear outcome) and produces a small but nonvanishing bias. A more severe misspecification would produce more visible degradation.

S6.5 Comparison with Existing Methods

S6.5.1 Manski Bounds

Table 10 compares the proposed bounds against the Manski worst-case benchmark, which uses only the outcome support restriction without the sorting characterization or the mean constraint. The Manski width is $W_{\text{exp}} \times d_{\max}$, the widest interval consistent with the support.

Under Config A, the Manski width is 2.538 for DiD and 2.921 for SC sparse. The proposed L_2 bounds reduce these to 0.000 for DiD, a 100% tightening, and to 1.447 for SC sparse, a 49.8% tightening. For staggered weights, the L_2 width of 0.378 is an 85.2% reduction from the Manski bound of 2.537. Under Config D (small spillovers), where the contamination is only 7.7% of the ATT, the proposed bounds achieve 94.4% tightening for staggered weights. Even for SC sparse, where the sorting characterization cannot tighten the bounds, the L_1 bounds provide a 53.5% reduction over Manski by using unit-specific capacities rather than the uniform worst-case capacity.

S6.5.2 Inferential Properties and the Cost of Ignoring Interference

Table 11 quantifies what happens when a researcher ignores interference against what happens when the researcher uses the proposed framework.

The plugin point estimate $\hat{B}_{\text{plugin}} = W_{\text{exp}} \times \hat{\psi}$ is nearly unbiased across all configurations, with plugin bias below 0.002 in absolute value. For DiD weights the plugin falls inside the estimated L_2 bounds in 100% of replications, consistent with point identification. For SC sparse and staggered weights the plugin containment rate is also 100% under Configs A and E, which confirms that the estimated identification region is conservative.

The detection test rejects the null of no interference ($H_0 \colon \psi = 0$) in 100% of replications under Configs A and E and in 83% under Config D. The lower detection rate under Config D reflects the small spillover magnitude ($\psi/\bar{d} = 0.05$), where the signal-to-noise ratio is low and the test lacks power at $N = 2000$. This is the detection boundary and not a failure of the method, since the bounds stay valid whether or not the test rejects.

The practical cost of ignoring interference appears in the "Naive CI fails" column. Under all three configurations shown, the standard confidence interval for the ATT, constructed under the false assumption of no interference, fails to cover the true ATT in 100% of replications for every weight type. The contamination bias exceeds what a SUTVA-assuming interval can absorb.

S6.6 Sensitivity Analysis

Two analyst choices govern the bounds, the support lower bound a and the exposure radius r. Neither is identified from the data, and both require substantive justification. I evaluate the sensitivity to each.

Support bound a. The support lower bound enters through the unit-specific capacities $d_j = \Delta Y_j^{\text{obs}} - a$. A more conservative (lower) a inflates capacities and widens the identification region, while an aggressive (higher) a tightens the region but risks violating the support assumption. Table 70 reports bound widths across $a \in \{-3, -2, -1, -0.5, 0\}$ for Config A, whose DGP calibration is $a = -1$. The pattern confirms the theory, since conservative

choices ($a < -1$) widen the bounds but hold coverage at 1.000 across all values tested. For staggered weights, the L_2 width increases from 0.378 at $a = -1$ to 0.602 at $a = -3$, a 59% widening. At the aggressive end ($a = 0$) the L_2 width shrinks to 0.259, but on average 9.7 units have negative capacity ($d_j < 0$), which indicates that the support assumption is violated. Coverage stays at 1.000 even at $a = 0$ in this DGP, but the analyst should treat an aggressive a with caution, because violations of the support assumption invalidate the theoretical guarantees.

Exposure radius r. Table 77 shows that for staggered weights the L_2 width is stable across $r \in [0.5, 2.25]$, ranging from 0.373 to 0.380, then collapses to zero at $r = 3.0$ when all controls become exposed and $\mathcal{C}_0$ vanishes. The CI length instead increases sharply, from 0.462 at $r = 0.5$ to 3.108 at $r = 2.75$, which reflects the loss of unexposed controls needed to estimate ψ. The breakdown point, the radius at which the CI becomes uninformative, arrives before the identification region itself widens, which highlights a tension between the coverage of the bounds and the statistical precision of the detection estimand.

S6.7 Additional Tables

The full per-configuration tables are collected in Section S8.

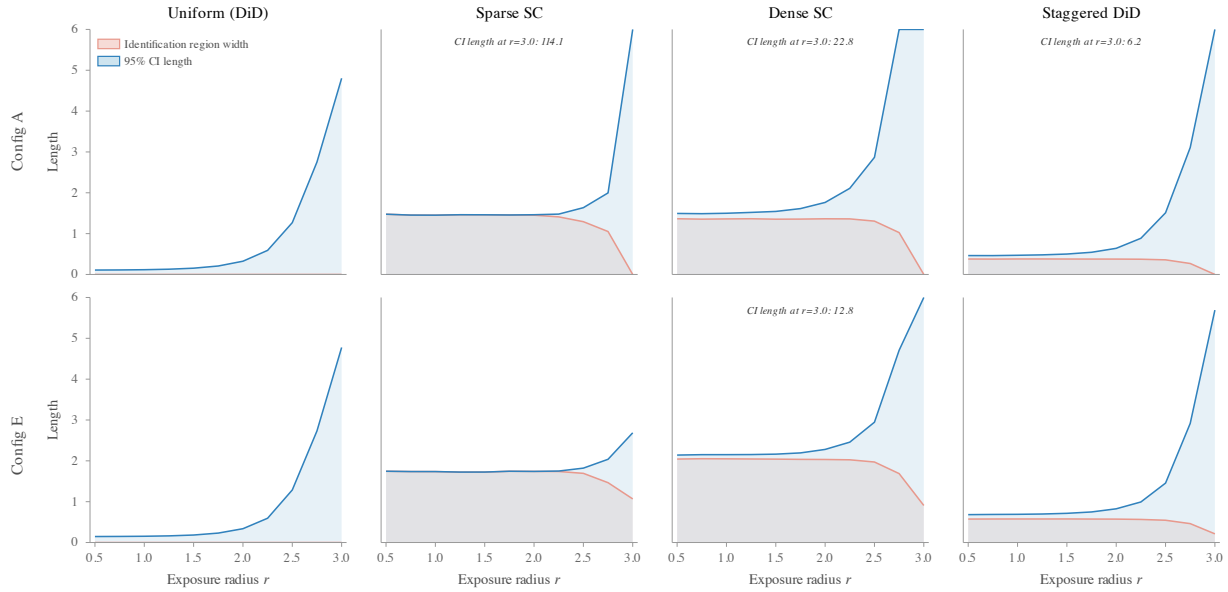
S7 Application Details

The main text carries each application result in full but defers four supporting pieces to keep the section within budget. This section supplies them. Section S7.1 shows the two exposure maps that define the exposed control set in each design. Section S7.2 gives the full reconciliation of the two reporting conventions for the point-identified UPP contamination, the simplex convention that produces the headline +0.0450 and the estimator-faithful convention that produces +0.0434. Section S7.3 develops the second Maricá outcome, vote concentration, at the length the main text gives the first. Section S7.4 reports the broader left-bloc replication that reproduces the detection pattern at $t = 4.17$. Section S7.5 records the exposure-mapping menu behind the reported cells, the full grid of radius, k-nearest-neighbor, and contiguity specifications with the selection rule that picks the reported radius. Throughout I report results for the local donor pool, the pool the main text uses, and I write results of the main text by their main-text number.

S7.1 Exposure Maps for the Two Designs

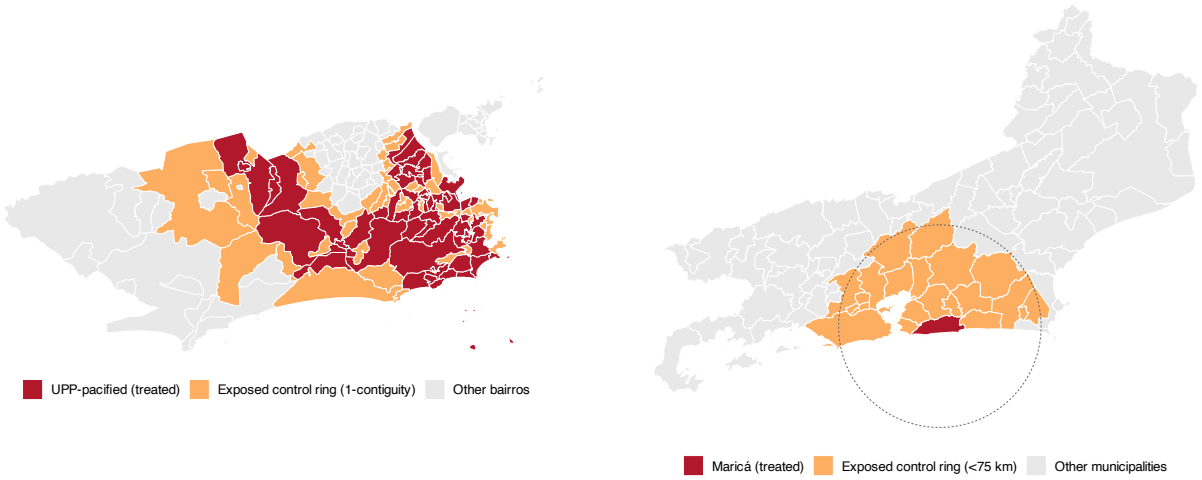
Each application defines exposure geographically, and the maps in Figure 3 show the resulting partition of the control units into the exposed set $\mathcal{C}_1$ and the unexposed set $\mathcal{C}_0$. In the UPP study exposure is first-order spatial contiguity: a control bairro is exposed if it shares a polygon border with at least one pacified bairro (Figure 3, panel a). This queen-contiguity rule uses only the bairro geometry and the treatment locations, both fixed before any outcome is examined, and it splits the 100 control bairros into $n_1 = 40$ exposed and $n_0 = 60$ unexposed, so $\rho = 0.40$. In the Maricá study exposure is proximity: a donor municipality is exposed if its centroid lies within the reported radius of Maricá (Figure 3, panel b, drawn at

Figure 2: Sensitivity of the Level-2 identification-region width and the 95% Imbens–Manski confidence-interval length to the exposure radius $r \in [0.5, 3.0]$, across Configs A and E and four weight schemes. The identification region stays stable while the confidence interval fans out, which locates the breakdown radius as an inferential rather than an identification boundary.



Note: Rows are Configs A (baseline) and E (large-spillover, $\psi/\bar{d} = 0.50$) of Table 4. Columns are the four weight schemes of Table 5, namely DiD ($w_j = |\mathcal{C}|^{-1}$), SC sparse, SC dense, and Staggered DiD (cohort-time weights following (Callaway and Sant'Anna, 2021; Sun and Abraham, 2021)). The coral curve is the median Level-2 width $\mathcal{W}_2(w) = \mathcal{B}_2^{\mathrm{U}}(w) - \mathcal{B}_2^{\mathrm{L}}(w)$, and the blue curve is the median 95% Imbens–Manski confidence-interval length. Both are aggregated across 5,000 Monte Carlo replications at $N = 2000$, and the blue curve contains the coral curve pointwise. The y-axis is linear and capped at 6, and panels exceeding the cap display italic annotations reporting the uncapped peak value at $r = 3.0$, so the cap is a display device. The weight-dispersion ranking across columns matches the width theorem (Theorem 4.10 of the main text). DiD yields essentially zero-width regions, whereas SC sparse drives the confidence-interval length well past the cap as $r \to 3$. Exact numerical values appear in Table 77 in Section S8.

the 75-kilometer radius used for the incumbent outcome). The concentric structure makes the radius a genuine analyst choice, which is why Section S7.5 reports the full menu of radii rather than a single specification.



(a) UPP: queen-contiguity exposure.

(b) Maricá: 75-km proximity exposure.

Figure 3: Exposure maps for the two designs. Panel (a): the treated (pacified) bairros of Rio de Janeiro and the control bairros that border them, the $n_1 = 40$ contiguity-exposed controls that make up $\mathcal{C}_1$ under the queen-contiguity rule. Panel (b): Maricá and the ring of donor municipalities within roughly 75 kilometers, the 21 exposed donors that make up $\mathcal{C}_1$ for the incumbent-share outcome. In both designs the partition into $\mathcal{C}_1$ and $\mathcal{C}_0$ is fixed ex ante from geography, not chosen to produce detection.

S7.2 The Two Reporting Conventions for the UPP Contamination

Under the uniform difference-in-differences weights, Proposition 4.6 of the main text collapses the Level 2 region to the point $\mathcal{B}_2(w) = \rho\hat{\psi} = +0.0053$, and the de-contaminated target is $\tau^{\text{SUTVA}}(w) = \hat{\tau}(w) + \mathcal{B}_2(w)$. The main text reports $+0.0450$ and states the estimator-faithful figure $+0.0434$ in one sentence. I give the full reconciliation here, because the two numbers answer two different questions and both are worth stating precisely.

The simplex convention treats each exposed control as projecting in full onto the differenced outcome the estimator uses. Under this convention every unit of exposed-control weight carries its full share of the spillover, the contamination is $\rho\hat{\psi} = 0.40 \times 0.0133 = +0.0053$, and the de-contaminated effect is

$$\tau^{\text{SUTVA}}(w) \;=\; 0.0397 + 0.0053 \;=\; +0.0450.$$

This is the convention in which the point-identification result of Proposition 4.6 is stated, and it is the headline the main text carries. About 13% of the naive four-point effect is identified spillover on this reading.

The estimator-faithful convention asks a narrower question: of the contamination the simplex convention attributes, how much does the Callaway–Sant'Anna estimator actually inherit through the specific differenced path it forms? On this panel the Callaway–Sant'Anna control-side loading that lands on the ΔY path the estimator differences sums to 0.69 of a unit rather than a full unit, so the bias the estimator carries is

$$0.69 \times \mathcal{B}_2(w) \;=\; 0.69 \times 0.0053 \;=\; +0.0037,$$

and the estimator-faithful de-contaminated effect is $0.0397 + 0.0037 = +0.0434$, about 9% of the naive effect. I confirm this figure by injecting the estimated spillover directly into the estimator and re-forming the ATT, which returns the same +0.0434 up to rounding. The two conventions agree on sign, on order of magnitude, and on the substantive reading: the naive difference-in-differences estimate is biased toward zero because the exposed controls are themselves lifted by the program, and de-contaminating moves the effect upward, so pacification matters more once the spillover in the controls is netted out. I report +0.0450 as the headline because it is the number the point-identification theory delivers, and I record +0.0434 here as the tighter estimator-specific reading.

S7.3 Maricá: Vote Concentration in Full

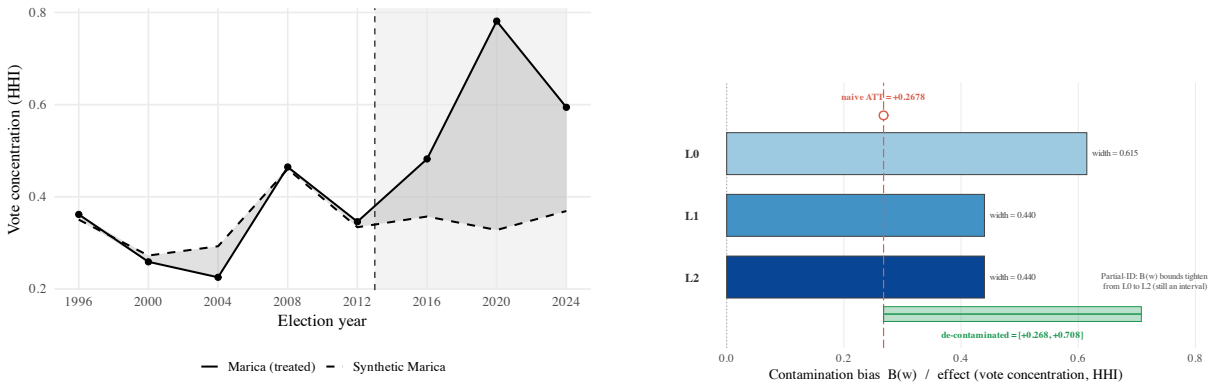
The main text summarizes the second Maricá outcome in a single paragraph. I develop it here at the length the incumbent-share outcome receives in the main text, because it is the cleanest demonstration that contamination is a property of the design rather than of one outcome.

Vote concentration is the Herfindahl share of the leading bloc, a measure of how lopsided local competition is: it rises when the vote consolidates around the strongest bloc and falls when it fragments. The synthetic control for Maricá tracks the real municipality closely before 2013, with a pre-treatment root-mean-square prediction error of 0.032, among the tightest of any outcome (Figure 4). The post-treatment gap is large, $\hat{\tau}(w) = +0.268$: the program does not only raise incumbent support, it concentrates competition around the incumbent bloc.

Exposure is proximity within 100 kilometers of Maricá, which splits the donor pool into $n_1 = 40$ exposed and $n_0 = 46$ unexposed municipalities, so $\rho = 0.47$. The doubly robust estimator detects positive spillover among the exposed donors, $\hat{\psi} = +0.059$ with a standard error of 0.020 and $t = 3.00$, and the Imbens–Manski interval $[0, +0.440]$ for the contamination signals genuine spillover. Because the synthetic control weights concentrate on a handful of donors, Proposition 4.7 of the main text applies and the contamination is only partially identified. Here the mean constraint at Level 2 adds nothing to the monotonicity restriction at Level 1, so the two regions coincide,

$$\mathcal{B}_1(w) \;=\; \mathcal{B}_2(w) \;=\; [\,0,\; +0.440\,],$$

inside the Manski baseline $[0, +0.615]$ (Figure 4, panel b). The coincidence is not a defect: the higher tier binds only when the weighted mean it imposes cuts into the box of feasible spillover allocations, and for this outcome it does not, so Level 2 returns the Level 1 answer. De-contaminating gives $\tau^{\mathrm{SUTVA}}(w) = [+0.268, +0.708]$, signed upward for the same reason



(a) Synthetic control fit.

(b) Nested contamination regions.

Figure 4: Maricá under synthetic control, vote concentration. Panel (a): the real municipality (solid) and its synthetic counterpart (dashed) track closely before 2013 and separate sharply afterward, with the post-treatment gap $\hat{\tau}(w) = +0.268$. Panel (b): the nested contamination regions. Adding monotonicity tightens the support-only Manski baseline $[0, +0.615]$ to the Level 1 region $[0, +0.440]$, but the mean constraint adds nothing further, so $\mathcal{B}_1(w) = \mathcal{B}_2(w) = [0, +0.440]$. Under the concentrated donor weights the region is an interval, not a point.

as the incumbent outcome: the exposed donors are lifted by the program and depress the synthetic counterfactual, biasing the measured gap toward zero. The sign of the correction is settled and its magnitude is bounded, and the concentration result stands beside the incumbent result as a second face of the same contaminated design.

S7.4 The Broader Left Bloc: A Replication

The main text notes that the pattern reproduces for the broader left bloc at $t = 4.17$. I record the full result here. The left-bloc share aggregates the vote of the left parties rather than the single incumbent party, and the synthetic control tracks it with a looser pre-treatment fit than the other two outcomes, a root-mean-square prediction error of 0.170 driven by a single pre-2013 swing, but it follows the trajectory and separates durably after treatment (Figure 5). The post-treatment gap is $\hat{\tau}(w) = +0.45$.

At the 100-kilometer radius the exposure split is again $n_1 = 40$ and $n_0 = 46$, so $\rho = 0.47$. The doubly robust estimator returns the strongest spillover signal of the three outcomes, $\hat{\psi} = +0.232$ with a standard error of 0.056 and $t = 4.17$, and the Imbens–Manski interval $[0, +0.680]$ signals genuine spillover. As with vote concentration the mean constraint does not bind beyond monotonicity, so $\mathcal{B}_1(w) = \mathcal{B}_2(w) = [0, +0.680]$ inside the wide Manski baseline $[0, +1.362]$, and de-contaminating gives $\tau^{\mathrm{SUTVA}}(w) = [+0.45, +1.13]$. The interval is wider than for the other two outcomes, reflecting both the looser pre-treatment fit and the larger dispersion of the left-bloc outcome, which is why I treat the left bloc as a replication of the pattern rather than as a headline number. The pattern it replicates is exactly the one the two headline outcomes establish: detected positive spillover, a correction signed upward, and partial rather than point identification under the concentrated donor weights.

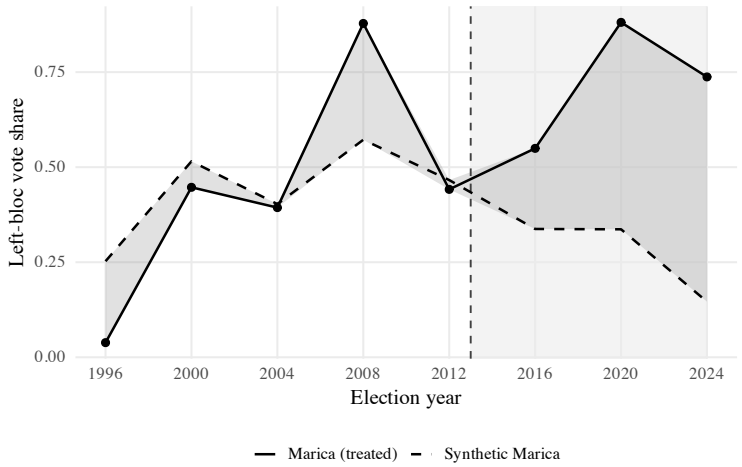


Figure 5: Maricá under synthetic control, left-bloc share. The real municipality (solid) and its synthetic counterpart (dashed) separate after 2013, with the post-treatment gap $\hat{\tau}(w) = +0.45$. The pre-treatment fit is looser than for the incumbent-share and vote-concentration outcomes, so the left bloc is reported as a replication of the detection-and-correction pattern rather than as a headline estimate.

S7.5 The Exposure-Mapping Menu

The reported cells fix an exposure specification, and this section records the full menu behind that choice so the reader can see it is disciplined rather than selected for detection. I build the menu from three families of exposure mappings evaluated on the same donor geometry. The first is a set of radius rings at $25, 50, 75, 100, 150, 200$, and 250 kilometers. The second is a set of k-nearest-neighbor graphs at $k \in \{3, 5, 8, 12, 20\}$. The third is first- and second-order queen contiguity. A specification enters the analysis only if it leaves enough units on both sides of the exposure boundary to estimate the spillover, which I operationalize as at least four exposed and four unexposed donors and an exposed fraction $\rho \in [0.10, 0.75]$. Table 1 shows the menu and the selection rule for the local pool.

Table 2 reports the detection and bounds for the retained local specifications across the three outcomes. Detection is the test that the doubly robust spillover $\hat{\psi}$ is significantly positive, and the Level 2 region and the Imbens–Manski interval are for the contamination $\mathcal{B}(w)$. Two features of the menu deserve comment. First, detection is specification-dependent: the natural radius detects for each outcome, at 75 kilometers for incumbent share and at 100 kilometers for the two broader outcomes, while narrower rings leave the spillover estimate small and insignificant. This is the reason the main text pairs the detection analysis with the sensitivity-to-radius curve of Section 5 of the main text rather than presenting a single specification as decisive. Second, the two k-nearest-neighbor specifications carry uninformative standard errors: restricting the donor set to a k-NN graph makes the synthetic control re-fit concentrate its weight on a single donor, and under that concentration the delta-method standard error is not reliable, exactly the failure mode flagged in the main text's remark on weight concentration. The Level 2 identification region, which does not depend on the standard error, remains well defined for those specifications, but the confidence interval and

Table 1: Exposure-mapping menu and selection rule, Maricá local donor pool. Each row is one exposure specification. A specification is retained only if it leaves at least four donors on each side of the boundary and an exposed fraction $\rho \in [0.10, 0.75]$. The retained radius specifications supply the reported cells, at 75 kilometers for incumbent share and at 100 kilometers for vote concentration and left-bloc share. The state and metropolitan pools are screened by the same rule.

Specification	Family	n_1	n_0	ρ	Status
25 km	radius	3	83	0.03	dropped ($n_1 < 4$)
50 km	radius	10	76	0.12	retained
75 km	radius	21	65	0.24	retained (incumbent cell)
100 km	radius	40	46	0.47	retained (concentration, left-bloc cells)
150 km	radius	86	0	1.00	dropped ($n_0 = 0$)
200 km	radius	86	0	1.00	dropped ($n_0 = 0$)
250 km	radius	86	0	1.00	dropped ($n_0 = 0$)
$k = 3$	k-NN	3	83	0.03	dropped ($n_1 < 4$)
$k = 5$	k-NN	5	81	0.06	dropped ($\rho < 0.10$)
$k = 8$	k-NN	8	78	0.09	dropped ($\rho < 0.10$)
$k = 12$	k-NN	12	74	0.14	retained
$k = 20$	k-NN	20	66	0.23	retained
queen, 1st order	contiguity	1	85	0.01	dropped ($n_1 < 4$)
queen, 2nd order	contiguity	2	84	0.02	dropped ($n_1 < 4$)

the detection verdict for them should be read as artifacts of the delta method rather than as evidence against spillover. The radius specifications, where the donor weights stay spread, give the reliable comparison, and the natural radius is the cell I carry to the main text.

Table 2: Detection and bounds across retained exposure specifications, Maricá local donor pool. $\hat{\psi}$ is the doubly robust spillover estimate on the exposed donors with its delta-method standard error. The column "detected" marks $\hat{\psi}$ significantly positive at the 5% level. $\mathcal{B}_2(w)$ is the Level 2 contamination region, and the last column is its Imbens–Manski confidence interval. Reported cells in bold.

Outcome	Spec.	n_1	n_0	ρ	$\hat{\psi}$ (SE)	Detected	$\mathcal{B}_2(w)$	IM CI
	50 km	10	76	0.12	+0.014 (0.052)	no	$[0, +0.028]$	$[0, +0.220]$
	$k = 12$	12	74	0.14	+0.050 (10.8)†	no	$[0, +0.120]$	$[0, +51.2]^\dagger$
Incumbent share	**75 km**	**21**	**65**	**0.24**	**+0.089 (0.042)**	**yes**	$[\mathbf{0}, \mathbf{+0.387}]$	$[\mathbf{0}, \mathbf{+0.568}]$
	$k = 20$	20	66	0.23	+0.103 (6.09)†	no	$[0, +0.410]$	$[0, +29.9]^\dagger$
	100 km	40	46	0.47	−0.003 (0.034)	no	$[0, 0]$	$[0, +0.620]$
	50 km	10	76	0.12	+0.032 (0.029)	no	$[0, +0.066]$	$[0, +0.163]$
	$k = 12$	12	74	0.14	+0.043 (5.92)†	no	$[0, +0.105]$	$[0, +28.6]^\dagger$
Vote concentration	75 km	21	65	0.24	+0.007 (0.023)	no	$[0, +0.035]$	$[0, +0.256]$
	$k = 20$	20	66	0.23	+0.012 (3.41)†	no	$[0, +0.061]$	$[0, +33.5]^\dagger$
	100 km	**40**	**46**	**0.47**	**+0.059 (0.020)**	**yes**	$[\mathbf{0}, \mathbf{+0.440}]$	$[\mathbf{0}, \mathbf{+0.440}]$
	50 km	10	76	0.12	+0.064 (0.056)	no	$[0, +0.015]$	$[0, +0.037]$
	$k = 12$	12	74	0.14	+0.084 (11.8)†	no	$[0, +0.020]$	$[0, +0.020]^\dagger$
Left-bloc share	75 km	21	65	0.24	−0.091 (0.046)	no	$[0, 0]$	$[0, +0.633]$
	$k = 20$	20	66	0.23	−0.085 (6.80)†	no	$[0, 0]$	$[0, +89.0]^\dagger$
	100 km	**40**	**46**	**0.47**	**+0.232 (0.056)**	**yes**	$[\mathbf{0}, \mathbf{+0.680}]$	$[\mathbf{0}, \mathbf{+0.680}]$

† Under the k-nearest-neighbor specifications the synthetic control re-fit concentrates its weight on a single donor, so the delta-method standard error and any quantity that depends on it (the confidence interval and the detection verdict) are uninformative. The Level 2 region does not depend on the standard error and remains well defined.

Table 3 consolidates the three Maricá outcomes at their reported specifications, collecting the pre-treatment fit, the naive effect, the detection statistics, the nested regions, and the de-contaminated interval in one place. The three columns share a single mechanism: a concentrated donor-weight vector, a doubly robust detection step that finds positive spillover, and a Level 2 region that stays an interval because Proposition 4.7 of the main text governs the asymmetric-weight case. They differ in how tightly the mean constraint binds, coinciding with Level 1 for the two broader outcomes and cutting inside it for incumbent share, and in the width of the resulting correction. In every column the correction is signed upward, so the program's effect on incumbent strength and on the concentration of competition is at least as large as the naive synthetic control reports and possibly larger.

Table 3: The three Maricá outcomes at their reported exposure specifications, local donor pool. The naive effect is the synthetic control post-treatment gap $\hat{\tau}(w)$. The estimate $\hat{\psi}$ is the doubly robust spillover with its standard error. L0, L1, and L2 are the Manski baseline, the Level 1 region under monotonicity, and the Level 2 region under the added mean constraint. The de-contaminated target is $\tau^{\text{SUTVA}}(w) = \hat{\tau}(w) + \mathcal{B}_2(w)$. All three partially identify under the concentrated donor weights.

	Incumbent share	Vote concentration	Left-bloc share
Exposure radius	75 km	100 km	100 km
$(n_1,\, n_0)$	(21, 65)	(40, 46)	(40, 46)
ρ	0.24	0.47	0.47
Pre-treatment RMSPE	0.054	0.032	0.170
Naive effect $\hat{\tau}(w)$	+0.59	+0.268	+0.45
$\hat{\psi}$ (SE)	+0.089 (0.042)	+0.059 (0.020)	+0.232 (0.056)
t	2.12	3.00	4.17
IM CI for $\mathcal{B}(w)$	$[0, +0.568]$	$[0, +0.440]$	$[0, +0.680]$
L0 (Manski)	$[0, +0.733]$	$[0, +0.615]$	$[0, +1.362]$
L1	$[0, +0.458]$	$[0, +0.440]$	$[0, +0.680]$
L2 $= \mathcal{B}_2(w)$	$[0, +0.387]$	$[0, +0.440]$	$[0, +0.680]$
De-contaminated $\tau^{\text{SUTVA}}(w)$	$[+0.59, +0.98]$	$[+0.268, +0.708]$	$[+0.45, +1.13]$

Note: For incumbent share the mean constraint binds and L2 cuts inside L1. For vote concentration and left-bloc share it does not, so $\mathcal{B}_1(w) = \mathcal{B}_2(w)$. The wider left-bloc region reflects the looser pre-treatment fit and the larger outcome dispersion, so the left bloc is reported as a replication of the pattern rather than as a headline estimate.

S8 Supplement Tables

This section collects the full-detail simulation and application tables that Sections 6 and 7 of the main text report only in summary. The first block gives the design grid and the full-sample tables behind the main-text figures. The per-configuration panels that follow report the doubly robust estimator, the bound endpoints, coverage, detection, and nesting checks for all eleven designs A–K. The diagnostic tables document computation, sensitivity, and distributional checks, and the application scorecards give the full detection-and-bounds ladders for the two empirical studies.

S8.1 Monte Carlo: Design Grid and Full-Sample Tables

Table 4 lists the parameters of all eleven designs. The remaining tables in this block report, at the default sample size, the weight structure of the four estimators, the bound widths across every configuration (Table 6), the bound endpoints, coverage and standard-error calibration, the double-robustness factorial, and the width and inferential comparisons against the support-only benchmark.

Table 4: DGP Configuration Parameters

Config	ρ	$\psi/\bar{d}$	τ_x	μ_0	π	σ	Paper	Purpose
A	0.30	0.25	0	Linear	Linear	1.0	Main	Baseline
B	0.10	0.25	0	Linear	Linear	1.0	Supp.	Low exposure
C	0.50	0.25	0	Linear	Linear	1.0	Supp.	High exposure
D	0.30	0.05	0	Linear	Linear	1.0	Main	Small spillovers
E	0.30	0.50	0	Linear	Linear	1.0	Main	Large spillovers
F	0.30	0.25	2.0	Linear	Linear	1.0	Main	Heterogeneous $\rightarrow L_3$
G	0.30	0.25	0	Nonlinear	Linear	1.0	Main	DR: wrong μ_0
H	0.30	0.25	0	Linear	Nonlinear	1.0	Main	DR: wrong π
I	0.30	0.25	0	Nonlinear	Nonlinear	1.0	Main	DR: both wrong
J	0.30	0.00	0	Linear	Linear	1.0	Main	Null (no interference)
K	0.30	0.25	0	Linear	Linear	2.5	Main	High capacity heterogeneity

Note: ρ: treatment probability (fraction treated). $\psi/\bar{d}$: spillover-to-capacity ratio governing contamination severity. τ_x: coefficient on X_1 in the spillover function $\tau(X) = \tau_{\text{base}} + \tau_x X_1$; nonzero values generate heterogeneous spillovers. μ_0, π: functional form of the outcome model and exposure propensity, respectively; "Nonlinear" indicates the true model contains interaction or nonlinear terms while the estimator fits a linear specification. σ: outcome noise standard deviation. Support lower bound $a = -1$ and exposure radius $r = 1$ in all configurations. "Supp." indicates results appear in the supplement only.

Table 5: Weight Structure Summary (Config A, $N = 2000$)

Weight	$n_{\neq 0}$	W_{exp}	HHI(C_1)	$\max \lvert\tilde{w}\rvert$
DiD	1453	0.300	0.0002	0.0000
SC sparse	20	0.344	0.0281	0.1180
SC dense	567	0.462	0.0016	0.0094
Staggered	1453	0.300	0.0003	0.0009

Note:
$n_{\neq 0}$: number of control units receiving nonzero weight. W_{exp}: total weight on exposed controls (exposure mass); larger values amplify contamination. HHI(C_1): Herfindahl index on exposed-control weights; higher values indicate weight concentration. $\max \lvert\tilde{w}\rvert$: maximum deviation of any weight from the uniform mean; governs the Hölder width bound (Theorem 4.4(iii)). Config A at $N = 2000$. DiD assigns uniform weights (low concentration), while SC concentrates on few donors (high HHI).

Table 6: Bound Width by Refinement Level (All Configurations)

DGP	Weight	B^{true}	B/τ	L_1	L_2	L_3	CI_width/ATT	Tight
A								
A	DiD	0.254	0.254	1.288 (0.031)	0.000 (0.000)	—	0.111	10
A	SC sparse	0.273	0.273	1.447 (0.673)	1.447 (0.673)	—	1.447	
A	Staggered	0.254	0.254	1.287 (0.056)	0.378 (0.046)	—	0.464	7
D								
D	DiD	0.077	0.077	1.110 (0.026)	0.000 (0.000)	—	0.103	10
D	SC sparse	0.067	0.067	1.273 (0.612)	1.270 (0.615)	—	1.279	
D	Staggered	0.076	0.076	1.109 (0.047)	0.134 (0.062)	—	0.260	8
E								
E	DiD	0.498	0.498	1.532 (0.039)	0.000 (0.000)	—	0.132	10
E	SC sparse	0.561	0.561	1.744 (0.797)	1.744 (0.797)	—	1.744	
E	Staggered	0.497	0.497	1.530 (0.065)	0.573 (0.059)	—	0.673	6
F								
F	DiD	0.342	0.342	1.377 (0.044)	0.000 (0.000)	0.000 (0.000)	0.137	10
F	SC sparse	0.348	0.348	1.511 (0.802)	1.511 (0.802)	1.383 (0.864)	1.511	
F	Staggered	0.342	0.342	1.376 (0.068)	0.451 (0.056)	0.397 (0.057)	0.554	6
J								
J	DiD	0.000	0.000	1.034 (0.024)	0.000 (0.000)	—	0.100	10
J	SC sparse	0.000	0.000	1.176 (0.557)	0.063 (1.017)	—	5.556	5
J	Staggered	0.000	0.000	1.034 (0.044)	0.001 (0.028)	—	0.153	9
K								
K	DiD	0.223	0.223	1.257 (0.063)	0.000 (0.000)	—	0.258	10
K	SC sparse	0.240	0.240	1.400 (0.893)	1.399 (0.893)	—	1.400	
K	Staggered	0.223	0.223	1.255 (0.084)	0.348 (0.087)	—	0.569	7

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

Table 7: Bound Endpoints by Refinement Level

DGP	Weight	B^{true}	L_1	L_2	CI_2	L_3	CI_3
A							
A	DiD	0.254	[0.000, 1.288]	[0.253, 0.253]	**(0.197, 0.310)**	—	—
A	SC sparse	0.273	[0.000, 1.447]	[0.000, 1.447]	**(0.000, 1.447)**	—	—
A	Staggered	0.254	[0.000, 1.287]	[0.144, 0.523]	**(0.117, 0.581)**	—	—
D							
D	DiD	0.077	[0.000, 1.110]	[0.076, 0.076]	**(0.024, 0.128)**	—	—
D	SC sparse	0.067	[0.000, 1.273]	[0.000, 1.270]	**(0.000, 1.279)**	—	—
D	Staggered	0.076	[0.000, 1.109]	[0.043, 0.177]	**(0.018, 0.280)**	—	—
E							
E	DiD	0.498	[0.000, 1.532]	[0.498, 0.498]	**(0.432, 0.564)**	—	—
E	SC sparse	0.561	[0.000, 1.744]	[0.000, 1.744]	**(0.000, 1.744)**	—	—
E	Staggered	0.497	[0.000, 1.530]	[0.282, 0.858]	**(0.250, 0.924)**	—	—
F							
F	DiD	0.342	[0.000, 1.377]	[0.344, 0.344]	**(0.275, 0.413)**	[0.345, 0.345]	**(0.230, 0.4**
F	SC sparse	0.348	[0.000, 1.511]	[0.000, 1.511]	**(0.000, 1.511)**	[0.000, 1.383]	**(0.000, 1.5**
F	Staggered	0.342	[0.000, 1.376]	[0.195, 0.646]	**(0.162, 0.716)**	[0.197, 0.595]	**(0.141, 0.7**
J							
J	DiD	0.000	[0.000, 1.034]	[0.000, 0.000]	**(-0.046, 0.055)**	—	—
J	SC sparse	0.000	[0.000, 1.176]	[0.000, 0.063]	**(0.000, 5.556)**	—	—
J	Staggered	0.000	[0.000, 1.034]	[0.000, 0.001]	**(-0.025, 0.127)**	—	—
K							
K	DiD	0.223	[0.000, 1.257]	[0.223, 0.223]	**(0.093, 0.352)**	—	—
K	SC sparse	0.240	[0.000, 1.400]	[0.000, 1.399]	**(0.000, 1.400)**	—	—
K	Staggered	0.223	[0.000, 1.255]	[0.126, 0.477]	**(0.064, 0.636)**	—	—

Note:
Bound endpoints shown as $[B_L, B_U]$ intervals; Imbens–Manski confidence intervals shown in bol
$(\text{CI}_L, \text{CI}_U)$. B^{true}: true contamination bias. L_1: identified set from support restrictions only. L_2: i
tified set after applying the sorting characterization (Theorem 4.3); for DiD, this collapses to a point.
95

Table 8: Coverage and SE Calibration by Sample Size

		$N = 500$					$N = 1000$					$N = 2000$				
DGP	Weight	Cov L2	Cov L3	CI len	SE rat U	SE rat L	Cov L2	Cov L3	CI len	SE rat U	SE rat L	Cov L2	Cov L3	CI len	SE rat U	SE rat L
A	DiD	0.982	1.000	0.209	1.04	1.04	0.977	1.000	0.154	1.03	1.03	0.984	1.000	0.111	1.05	1.05
A	SC sparse	1.000	1.000	1.441	—	—	1.000	1.000	1.462	—	—	1.000	1.000	1.447	—	—
A	SC dense	1.000	1.000	1.608	0.71	—	1.000	1.000	1.548	0.7	—	1.000	1.000	1.505	0.72	—
A	Staggered	1.000	1.000	0.547	0.81	1	1.000	1.000	0.499	0.79	0.99	1.000	1.000	0.464	0.79	1.01
D	DiD	0.981	0.998	0.191	1.09	1.09	0.980	0.999	0.142	1.07	1.07	0.972	1.000	0.103	1.07	1.07
D	SC sparse	1.000	0.995	1.492	0.73	—	1.000	0.998	1.301	0.32	—	1.000	1.000	1.279	0.06	—
D	SC dense	1.000	0.999	1.095	1.09	—	1.000	1.000	0.955	1.03	—	1.000	1.000	0.855	1	—
D	Staggered	0.999	0.999	0.375	1.06	1.08	1.000	1.000	0.309	1.05	1.07	1.000	1.000	0.260	1.04	1.06
E	DiD	0.994	1.000	0.253	1.08	1.08	0.996	1.000	0.184	1.1	1.1	0.996	1.000	0.132	1.09	1.09
E	SC sparse	1.000	1.000	1.733	—	—	1.000	1.000	1.727	—	—	1.000	1.000	1.744	—	—
E	SC dense	1.000	1.000	2.209	0.43	0.19	1.000	1.000	2.174	0.43	0.12	1.000	1.000	2.141	0.43	0.04
E	Staggered	1.000	1.000	0.762	0.72	1	1.000	1.000	0.714	0.76	1	1.000	1.000	0.673	0.78	0.98
F	DiD	0.996	1.000	0.264	1.06	1.06	0.996	1.000	0.191	1.09	1.09	0.996	1.000	0.137	1.08	1.08
F	SC sparse	1.000	0.999	1.528	—	—	1.000	1.000	1.538	—	—	1.000	1.000	1.511	—	—
F	SC dense	1.000	1.000	1.883	0.6	0.07	1.000	1.000	1.813	0.61	—	1.000	1.000	1.780	0.6	—
F	Staggered	1.000	1.000	0.653	0.78	1.02	1.000	1.000	0.594	0.8	1.04	1.000	1.000	0.554	0.81	1.03
K	DiD	0.984	0.997	0.479	1.02	1.02	0.976	0.999	0.356	1.01	1.01	0.972	1.000	0.258	1.01	1.01
K	SC sparse	0.995	0.978	1.513	0.57	—	0.994	0.992	1.412	0.18	—	0.995	0.994	1.400	0.01	—
K	SC dense	1.000	0.998	1.956	0.97	0.13	1.000	1.000	1.738	0.9	0.04	1.000	1.000	1.603	0.88	—
K	Staggered	1.000	0.998	0.778	0.97	1.01	1.000	1.000	0.660	0.94	1	1.000	1.000	0.569	0.92	1

Note:
Cov L_2, Cov L_3: empirical coverage of the 95

Table 9: Double Robustness: 2×2 Misspecification Factorial

Config	N	Bias($\hat{\psi}$)	RMSE($\hat{\psi}$)	Bias($\hat{\psi}_{CF}$)	RMSE($\hat{\psi}_{CF}$)	Bias(UB)	Cov(L_2)
μ_0 correct, *pi* correct							
A	500	-0.0015	0.1527	—	—	0.0000	1.000
A	2000	-0.0006	0.0791	—	—	0.0000	1.000
μ_0 wrong, *pi* correct							
G	500	0.0061	0.1810	0.0016	0.4977	-0.0007	1.000
G	2000	0.0004	0.0920	-0.002	0.2526	0.0000	1.000
μ_0 correct, *pi* wrong							
H	500	0.0012	0.1844	0.0052	0.6147	-0.0010	1.000
H	2000	-0.0011	0.0974	-0.0027	0.3329	0.0000	1.000
μ_0 wrong, *pi* wrong							
I	500	-0.0081	0.4586	-0.0128	1.551	-0.1800	0.988
I	2000	-0.0040	0.2429	-0.0138	0.8475	-0.0252	0.982

Note:
Double robustness validation under the 2×2 misspecification factorial. Config A: both μ_0 and π correctly specified. Config G: μ_0 misspecified (wrong outcome model), π correct. Config H: μ_0 correct, π misspecified (wrong propensity score). Config I: both misspecified. Bias($\hat{\psi}$), RMSE($\hat{\psi}$): bias and root mean squared error of the DR spillover estimator. Bias($\hat{\psi}_{CF}$), RMSE($\hat{\psi}_{CF}$): same for the cross-fitted variant. "—" indicates cross-fitting was not run for that configuration. Bias(UB): bias of the upper bound plugin estimator. Cov(L_2): coverage of the 95

Table 10: Bound Width Comparisons ($N = 2000$)

DGP	Weight	Manski worst-case	L_1 width	L_2 width	Tightening HM $\to$ L_2 (%)
A					
A	DiD	2.538	1.288	0.000	100.0
A	SC sparse	2.921	1.447	1.447	49.8
A	Staggered	2.537	1.287	0.378	85.2
D					
D	DiD	2.367	1.110	0.000	100.0
D	SC sparse	2.744	1.273	1.270	53.5
D	Staggered	2.367	1.109	0.134	94.4
E					
E	DiD	2.793	1.532	0.000	100.0
E	SC sparse	3.212	1.744	1.744	45.2
E	Staggered	2.792	1.530	0.573	79.6

Note:
Manski worst-case: $W_{\exp} \times d_{\max}$, the widest possible identified set without using the sorting characterization. L_1 width: identified set width using support restrictions only (Assumption 3.1). L_2 width: identified set width after applying the sorting characterization (Theorem 4.3). Tightening HM $\to$ L_2: percentage reduction from the Manski worst-case to the L_2 bound, measuring how much the sorting characterization narrows the identified set. DiD achieves point identification (L_2 width $approx 0$), yielding near-100

Table 11: Inferential Comparisons with Competitor Methods ($N = 2000$)

DGP	Weight	Plugin bias	Plugin in L_2 bounds	Plugin in oracle	Test rejects	Naive CI fails	Conclusion flips
A							
A	DiD	-0.0002	1.000	0.000	1.00	Yes	0.0%
A	SC sparse	0.0001	1.000	1.000	1.00	Yes	0.0%
A	Staggered	-0.0002	1.000	1.000	1.00	Yes	0.0%
D							
D	DiD	-0.0006	0.998	0.000	0.83	Yes	0.0%
D	SC sparse	-0.0019	0.998	0.998	0.83	Yes	0.1%
D	Staggered	-0.0006	0.998	0.911	0.83	Yes	0.0%
E							
E	DiD	-0.0003	1.000	0.000	1.00	Yes	0.0%
E	SC sparse	-0.0019	1.000	1.000	1.00	Yes	0.0%
E	Staggered	-0.0002	1.000	1.000	1.00	Yes	0.0%

Note:
Plugin bias: $\hat{B}_{\text{plugin}} - B^{\text{true}}$, where $\hat{B}_{\text{plugin}} = W_{\text{exp}} \times \hat{\psi}$. Plugin in L_2 bounds: fraction of replications where the plugin falls inside the estimated L_2 interval. Plugin in oracle: fraction where the plugin falls inside the oracle L_2 interval (known nuisance parameters). For DiD, the oracle interval is a single point $[B^{\text{true}}, B^{\text{true}}]$, so the plugin (a noisy point estimate) falls inside with probability zero — this is expected and reflects point identification, not estimator failure. Test rejects: rejection rate of the Wald test $H_0! : \psi = 0$ at $\alpha = 0.05$. Naive CI fails: fraction of replications where the standard (SUTVA-assuming) CI fails to cover the true ATT, measuring the cost of ignoring interference. Conclusion flips: fraction where the naive ATT estimate is significant but the bias-corrected interval includes zero. $N = 2000$, 5,000 replications.

S8.2 Table S1a: DR Estimator Performance (Configs A–K)

Bias, standard deviation, RMSE, and $\sqrt{N}$-scaled RMSE of the doubly robust spillover estimator $\hat{\psi}_{\text{DR}}$ across all configurations and sample sizes. Bounded $\sqrt{N}$-scaled RMSE confirms the parametric rate under correct or single-model specification.

Table 12: DR Estimator Performance: Config A: Baseline

Weight	N	B	Bias($\hat{\psi}$)	Bias$\sqrt{N}$	SD($\hat{\psi}$)	RMSE($\hat{\psi}$)	RMSE$\sqrt{N}$
DiD	200	5000	-0.0035	-0.049	0.2403	0.2403	3.398
DiD	500	5000	-0.0015	-0.034	0.1528	0.1527	3.415
DiD	1000	5000	-0.0011	-0.034	0.1144	0.1143	3.616
DiD	2000	5000	-0.0006	-0.027	0.0791	0.0791	3.539
DiD	5000	5000	-0.0004	-0.029	0.0512	0.0512	3.617
SC sparse	200	5000	-0.0035	-0.049	0.2403	0.2403	3.398
SC sparse	500	5000	-0.0015	-0.034	0.1528	0.1527	3.415
SC sparse	1000	5000	-0.0011	-0.034	0.1144	0.1143	3.616
SC sparse	2000	5000	-0.0006	-0.027	0.0791	0.0791	3.539
SC sparse	5000	5000	-0.0004	-0.029	0.0512	0.0512	3.617
SC dense	200	5000	-0.0035	-0.049	0.2403	0.2403	3.398
SC dense	500	5000	-0.0015	-0.034	0.1528	0.1527	3.415
SC dense	1000	5000	-0.0011	-0.034	0.1144	0.1143	3.616
SC dense	2000	5000	-0.0006	-0.027	0.0791	0.0791	3.539
SC dense	5000	5000	-0.0004	-0.029	0.0512	0.0512	3.617
Staggered	200	5000	-0.0035	-0.049	0.2403	0.2403	3.398
Staggered	500	5000	-0.0015	-0.034	0.1528	0.1527	3.415
Staggered	1000	5000	-0.0011	-0.034	0.1144	0.1143	3.616
Staggered	2000	5000	-0.0006	-0.027	0.0791	0.0791	3.539
Staggered	5000	5000	-0.0004	-0.029	0.0512	0.0512	3.617

Note:
Bias($\hat{\psi}$): mean bias of the DR spillover estimator across replications. Bias$\sqrt{N}$: bias scaled by $\sqrt{N}$; bounded values confirm $\sqrt{N}$-consistency. SD($\hat{\psi}$): Monte Carlo standard deviation. RMSE($\hat{\psi}$): root mean squared error. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms the parametric rate. All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000,$ 5,000 replications per cell.

Table 13: DR Estimator Performance: Config B: No spillover

Weight	N	B	Bias($\hat{\psi}$)	Bias$\sqrt{N}$	SD($\hat{\psi}$)	RMSE($\hat{\psi}$)	RMSE$\sqrt{N}$
DiD	500	5000	-0.0142	-0.316	0.4833	0.4834	10.809
DiD	2000	5000	0.0092	0.409	0.2567	0.2568	11.485
SC sparse	500	5000	-0.0142	-0.316	0.4833	0.4834	10.809
SC sparse	2000	5000	0.0092	0.409	0.2567	0.2568	11.485
SC dense	500	5000	-0.0142	-0.316	0.4833	0.4834	10.809
SC dense	2000	5000	0.0092	0.409	0.2567	0.2568	11.485
Staggered	500	5000	-0.0142	-0.316	0.4833	0.4834	10.809
Staggered	2000	5000	0.0092	0.409	0.2567	0.2568	11.485

Note:
Bias($\hat{\psi}$): mean bias of the DR spillover estimator across replications. Bias$\sqrt{N}$: bias scaled by $\sqrt{N}$; bounded values confirm $\sqrt{N}$-consistency. SD($\hat{\psi}$): Monte Carlo standard deviation. RMSE($\hat{\psi}$): root mean squared error. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms the parametric rate. All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

Table 14: DR Estimator Performance: Config C: Uniform spillover

Weight	N	B	Bias($\hat{\psi}$)	Bias$\sqrt{N}$	SD($\hat{\psi}$)	RMSE($\hat{\psi}$)	RMSE$\sqrt{N}$
DiD	500	5000	-1e-04	-0.002	0.1130	0.1129	2.525
DiD	2000	5000	-6e-04	-0.028	0.0576	0.0576	2.576
SC sparse	500	5000	-1e-04	-0.002	0.1130	0.1129	2.525
SC sparse	2000	5000	-6e-04	-0.028	0.0576	0.0576	2.576
SC dense	500	5000	-1e-04	-0.002	0.1130	0.1129	2.525
SC dense	2000	5000	-6e-04	-0.028	0.0576	0.0576	2.576
Staggered	500	5000	-1e-04	-0.002	0.1130	0.1129	2.525
Staggered	2000	5000	-6e-04	-0.028	0.0576	0.0576	2.576

Note:
Bias($\hat{\psi}$): mean bias of the DR spillover estimator across replications. Bias$\sqrt{N}$: bias scaled by $\sqrt{N}$; bounded values confirm $\sqrt{N}$-consistency. SD($\hat{\psi}$): Monte Carlo standard deviation. RMSE($\hat{\psi}$): root mean squared error. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms the parametric rate. All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

Table 15: DR Estimator Performance: Config D: Strong spillover

Weight	N	B	Bias($\hat{\psi}$)	Bias$\sqrt{N}$	SD($\hat{\psi}$)	RMSE($\hat{\psi}$)	RMSE$\sqrt{N}$
DiD	200	5000	-0.0045	-0.063	0.2427	0.2428	3.433
DiD	500	5000	0.0030	0.066	0.1555	0.1555	3.477
DiD	1000	5000	-0.0017	-0.054	0.1114	0.1114	3.523
DiD	2000	5000	-0.0021	-0.094	0.0801	0.0801	3.583
DiD	5000	5000	0.0005	0.036	0.0512	0.0512	3.623
SC sparse	200	5000	-0.0045	-0.063	0.2427	0.2428	3.433
SC sparse	500	5000	0.0030	0.066	0.1555	0.1555	3.477
SC sparse	1000	5000	-0.0017	-0.054	0.1114	0.1114	3.523
SC sparse	2000	5000	-0.0021	-0.094	0.0801	0.0801	3.583
SC sparse	5000	5000	0.0005	0.036	0.0512	0.0512	3.623
SC dense	200	5000	-0.0045	-0.063	0.2427	0.2428	3.433
SC dense	500	5000	0.0030	0.066	0.1555	0.1555	3.477
SC dense	1000	5000	-0.0017	-0.054	0.1114	0.1114	3.523
SC dense	2000	5000	-0.0021	-0.094	0.0801	0.0801	3.583
SC dense	5000	5000	0.0005	0.036	0.0512	0.0512	3.623
Staggered	200	5000	-0.0045	-0.063	0.2427	0.2428	3.433
Staggered	500	5000	0.0030	0.066	0.1555	0.1555	3.477
Staggered	1000	5000	-0.0017	-0.054	0.1114	0.1114	3.523
Staggered	2000	5000	-0.0021	-0.094	0.0801	0.0801	3.583
Staggered	5000	5000	0.0005	0.036	0.0512	0.0512	3.623

Note:
Bias($\hat{\psi}$): mean bias of the DR spillover estimator across replications. Bias$\sqrt{N}$: bias scaled by $\sqrt{N}$; bounded values confirm $\sqrt{N}$-consistency. SD($\hat{\psi}$): Monte Carlo standard deviation. RMSE($\hat{\psi}$): root mean squared error. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms the parametric rate. All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000,$ 5,000 replications per cell.

Table 16: DR Estimator Performance: Config E: Spatial decay

Weight	N	B	Bias($\hat{\psi}$)	Bias$\sqrt{N}$	SD($\hat{\psi}$)	RMSE($\hat{\psi}$)	RMSE$\sqrt{N}$
DiD	200	5000	0.0044	0.062	0.2432	0.2432	3.440
DiD	500	5000	-0.0020	-0.045	0.1582	0.1582	3.537
DiD	1000	5000	0.0036	0.113	0.1120	0.1121	3.544
DiD	2000	5000	-0.0011	-0.049	0.0813	0.0813	3.638
DiD	5000	5000	-0.0006	-0.045	0.0513	0.0513	3.625
SC sparse	200	5000	0.0044	0.062	0.2432	0.2432	3.440
SC sparse	500	5000	-0.0020	-0.045	0.1582	0.1582	3.537
SC sparse	1000	5000	0.0036	0.113	0.1120	0.1121	3.544
SC sparse	2000	5000	-0.0011	-0.049	0.0813	0.0813	3.638
SC sparse	5000	5000	-0.0006	-0.045	0.0513	0.0513	3.625
SC dense	200	5000	0.0044	0.062	0.2432	0.2432	3.440
SC dense	500	5000	-0.0020	-0.045	0.1582	0.1582	3.537
SC dense	1000	5000	0.0036	0.113	0.1120	0.1121	3.544
SC dense	2000	5000	-0.0011	-0.049	0.0813	0.0813	3.638
SC dense	5000	5000	-0.0006	-0.045	0.0513	0.0513	3.625
Staggered	200	5000	0.0044	0.062	0.2432	0.2432	3.440
Staggered	500	5000	-0.0020	-0.045	0.1582	0.1582	3.537
Staggered	1000	5000	0.0036	0.113	0.1120	0.1121	3.544
Staggered	2000	5000	-0.0011	-0.049	0.0813	0.0813	3.638
Staggered	5000	5000	-0.0006	-0.045	0.0513	0.0513	3.625

Note:
Bias($\hat{\psi}$): mean bias of the DR spillover estimator across replications. Bias$\sqrt{N}$: bias scaled by $\sqrt{N}$; bounded values confirm $\sqrt{N}$-consistency. SD($\hat{\psi}$): Monte Carlo standard deviation. RMSE($\hat{\psi}$): root mean squared error. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms the parametric rate. All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000,$ 5,000 replications per cell.

Table 17: DR Estimator Performance: Config F: Heterogeneous

Weight	N	B	Bias($\hat{\psi}$)	Bias$\sqrt{N}$	SD($\hat{\psi}$)	RMSE($\hat{\psi}$)	RMSE$\sqrt{N}$
DiD	200	5000	0.0007	0.010	0.2380	0.2380	3.366
DiD	500	5000	0.0009	0.020	0.1577	0.1577	3.526
DiD	1000	5000	-0.0027	-0.085	0.1113	0.1113	3.521
DiD	2000	5000	0.0023	0.104	0.0796	0.0797	3.562
DiD	5000	5000	-0.0008	-0.058	0.0507	0.0507	3.583
SC sparse	200	5000	0.0007	0.010	0.2380	0.2380	3.366
SC sparse	500	5000	0.0009	0.020	0.1577	0.1577	3.526
SC sparse	1000	5000	-0.0027	-0.085	0.1113	0.1113	3.521
SC sparse	2000	5000	0.0023	0.104	0.0796	0.0797	3.562
SC sparse	5000	5000	-0.0008	-0.058	0.0507	0.0507	3.583
SC dense	200	5000	0.0007	0.010	0.2380	0.2380	3.366
SC dense	500	5000	0.0009	0.020	0.1577	0.1577	3.526
SC dense	1000	5000	-0.0027	-0.085	0.1113	0.1113	3.521
SC dense	2000	5000	0.0023	0.104	0.0796	0.0797	3.562
SC dense	5000	5000	-0.0008	-0.058	0.0507	0.0507	3.583
Staggered	200	5000	0.0007	0.010	0.2380	0.2380	3.366
Staggered	500	5000	0.0009	0.020	0.1577	0.1577	3.526
Staggered	1000	5000	-0.0027	-0.085	0.1113	0.1113	3.521
Staggered	2000	5000	0.0023	0.104	0.0796	0.0797	3.562
Staggered	5000	5000	-0.0008	-0.058	0.0507	0.0507	3.583

Note:
Bias($\hat{\psi}$): mean bias of the DR spillover estimator across replications. Bias$\sqrt{N}$: bias scaled by $\sqrt{N}$; bounded values confirm $\sqrt{N}$-consistency. SD($\hat{\psi}$): Monte Carlo standard deviation. RMSE($\hat{\psi}$): root mean squared error. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms the parametric rate. All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000,$ 5,000 replications per cell.

Table 18: DR Estimator Performance: Config G: Wrong μ_0

Weight	N	B	Bias($\hat{\psi}$)	Bias$\sqrt{N}$	SD($\hat{\psi}$)	RMSE($\hat{\psi}$)	RMSE$\sqrt{N}$
DiD	200	5000	0.0068	0.097	0.2787	0.2787	3.942
DiD	500	5000	0.0061	0.136	0.1809	0.1810	4.046
DiD	1000	5000	0.0011	0.036	0.1290	0.1290	4.080
DiD	2000	5000	0.0004	0.019	0.0920	0.0920	4.115
DiD	5000	5000	0.0006	0.040	0.0590	0.0590	4.170
SC sparse	200	5000	0.0068	0.097	0.2787	0.2787	3.942
SC sparse	500	5000	0.0061	0.136	0.1809	0.1810	4.046
SC sparse	1000	5000	0.0011	0.036	0.1290	0.1290	4.080
SC sparse	2000	5000	0.0004	0.019	0.0920	0.0920	4.115
SC sparse	5000	5000	0.0006	0.040	0.0590	0.0590	4.170
SC dense	200	5000	0.0068	0.097	0.2787	0.2787	3.942
SC dense	500	5000	0.0061	0.136	0.1809	0.1810	4.046
SC dense	1000	5000	0.0011	0.036	0.1290	0.1290	4.080
SC dense	2000	5000	0.0004	0.019	0.0920	0.0920	4.115
SC dense	5000	5000	0.0006	0.040	0.0590	0.0590	4.170
Staggered	200	5000	0.0068	0.097	0.2787	0.2787	3.942
Staggered	500	5000	0.0061	0.136	0.1809	0.1810	4.046
Staggered	1000	5000	0.0011	0.036	0.1290	0.1290	4.080
Staggered	2000	5000	0.0004	0.019	0.0920	0.0920	4.115
Staggered	5000	5000	0.0006	0.040	0.0590	0.0590	4.170

Note:
Bias($\hat{\psi}$): mean bias of the DR spillover estimator across replications. Bias$\sqrt{N}$: bias scaled by $\sqrt{N}$; bounded values confirm $\sqrt{N}$-consistency. SD($\hat{\psi}$): Monte Carlo standard deviation. RMSE($\hat{\psi}$): root mean squared error. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms the parametric rate. All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000,$ 5,000 replications per cell.

Table 19: DR Estimator Performance: Config H: Wrong π

Weight	N	B	Bias($\hat{\psi}$)	Bias$\sqrt{N}$	SD($\hat{\psi}$)	RMSE($\hat{\psi}$)	RMSE$\sqrt{N}$
DiD	200	5000	0.0022	0.031	0.2792	0.2792	3.948
DiD	500	5000	0.0012	0.026	0.1845	0.1844	4.124
DiD	1000	5000	-0.0008	-0.024	0.1368	0.1368	4.326
DiD	2000	5000	-0.0011	-0.050	0.0974	0.0974	4.356
DiD	5000	5000	0.0004	0.028	0.0622	0.0622	4.399
SC sparse	200	5000	0.0022	0.031	0.2792	0.2792	3.948
SC sparse	500	5000	0.0012	0.026	0.1845	0.1844	4.124
SC sparse	1000	5000	-0.0008	-0.024	0.1368	0.1368	4.326
SC sparse	2000	5000	-0.0011	-0.050	0.0974	0.0974	4.356
SC sparse	5000	5000	0.0004	0.028	0.0622	0.0622	4.399
SC dense	200	5000	0.0022	0.031	0.2792	0.2792	3.948
SC dense	500	5000	0.0012	0.026	0.1845	0.1844	4.124
SC dense	1000	5000	-0.0008	-0.024	0.1368	0.1368	4.326
SC dense	2000	5000	-0.0011	-0.050	0.0974	0.0974	4.356
SC dense	5000	5000	0.0004	0.028	0.0622	0.0622	4.399
Staggered	200	5000	0.0022	0.031	0.2792	0.2792	3.948
Staggered	500	5000	0.0012	0.026	0.1845	0.1844	4.124
Staggered	1000	5000	-0.0008	-0.024	0.1368	0.1368	4.326
Staggered	2000	5000	-0.0011	-0.050	0.0974	0.0974	4.356
Staggered	5000	5000	0.0004	0.028	0.0622	0.0622	4.399

Note:
Bias($\hat{\psi}$): mean bias of the DR spillover estimator across replications. Bias$\sqrt{N}$: bias scaled by $\sqrt{N}$; bounded values confirm $\sqrt{N}$-consistency. SD($\hat{\psi}$): Monte Carlo standard deviation. RMSE($\hat{\psi}$): root mean squared error. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms the parametric rate. All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000,$ 5,000 replications per cell.

Table 20: DR Estimator Performance: Config I: Both wrong

Weight	N	B	Bias($\hat{\psi}$)	Bias$\sqrt{N}$	SD($\hat{\psi}$)	RMSE($\hat{\psi}$)	RMSE$\sqrt{N}$
DiD	200	5000	-0.0040	-0.057	0.7186	0.7185	10.161
DiD	500	5000	-0.0081	-0.182	0.4585	0.4586	10.254
DiD	1000	5000	0.0013	0.041	0.3394	0.3393	10.731
DiD	2000	5000	-0.0040	-0.178	0.2429	0.2429	10.863
DiD	5000	5000	-0.0006	-0.039	0.1562	0.1561	11.041
SC sparse	200	5000	-0.0040	-0.057	0.7186	0.7185	10.161
SC sparse	500	5000	-0.0081	-0.182	0.4585	0.4586	10.254
SC sparse	1000	5000	0.0013	0.041	0.3394	0.3393	10.731
SC sparse	2000	5000	-0.0040	-0.178	0.2429	0.2429	10.863
SC sparse	5000	5000	-0.0006	-0.039	0.1562	0.1561	11.041
SC dense	200	5000	-0.0040	-0.057	0.7186	0.7185	10.161
SC dense	500	5000	-0.0081	-0.182	0.4585	0.4586	10.254
SC dense	1000	5000	0.0013	0.041	0.3394	0.3393	10.731
SC dense	2000	5000	-0.0040	-0.178	0.2429	0.2429	10.863
SC dense	5000	5000	-0.0006	-0.039	0.1562	0.1561	11.041
Staggered	200	5000	-0.0040	-0.057	0.7186	0.7185	10.161
Staggered	500	5000	-0.0081	-0.182	0.4585	0.4586	10.254
Staggered	1000	5000	0.0013	0.041	0.3394	0.3393	10.731
Staggered	2000	5000	-0.0040	-0.178	0.2429	0.2429	10.863
Staggered	5000	5000	-0.0006	-0.039	0.1562	0.1561	11.041

Note:
Bias($\hat{\psi}$): mean bias of the DR spillover estimator across replications. Bias$\sqrt{N}$: bias scaled by $\sqrt{N}$; bounded values confirm $\sqrt{N}$-consistency. SD($\hat{\psi}$): Monte Carlo standard deviation. RMSE($\hat{\psi}$): root mean squared error. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms the parametric rate. All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000,$ 5,000 replications per cell.

Table 21: DR Estimator Performance: Config J: Small n_1/n_0

Weight	N	B	Bias($\hat{\psi}$)	Bias$\sqrt{N}$	SD($\hat{\psi}$)	RMSE($\hat{\psi}$)	RMSE$\sqrt{N}$
DiD	200	5000	0.0015	0.022	0.2444	0.2443	3.455
DiD	500	5000	0.0000	0.000	0.1557	0.1557	3.481
DiD	1000	5000	-0.0041	-0.129	0.1129	0.1130	3.573
DiD	2000	5000	-0.0004	-0.017	0.0809	0.0809	3.616
DiD	5000	5000	0.0005	0.033	0.0508	0.0508	3.590
SC sparse	200	5000	0.0015	0.022	0.2444	0.2443	3.455
SC sparse	500	5000	0.0000	0.000	0.1557	0.1557	3.481
SC sparse	1000	5000	-0.0041	-0.129	0.1129	0.1130	3.573
SC sparse	2000	5000	-0.0004	-0.017	0.0809	0.0809	3.616
SC sparse	5000	5000	0.0005	0.033	0.0508	0.0508	3.590
SC dense	200	5000	0.0015	0.022	0.2444	0.2443	3.455
SC dense	500	5000	0.0000	0.000	0.1557	0.1557	3.481
SC dense	1000	5000	-0.0041	-0.129	0.1129	0.1130	3.573
SC dense	2000	5000	-0.0004	-0.017	0.0809	0.0809	3.616
SC dense	5000	5000	0.0005	0.033	0.0508	0.0508	3.590
Staggered	200	5000	0.0015	0.022	0.2444	0.2443	3.455
Staggered	500	5000	0.0000	0.000	0.1557	0.1557	3.481
Staggered	1000	5000	-0.0041	-0.129	0.1129	0.1130	3.573
Staggered	2000	5000	-0.0004	-0.017	0.0809	0.0809	3.616
Staggered	5000	5000	0.0005	0.033	0.0508	0.0508	3.590

Note:
Bias($\hat{\psi}$): mean bias of the DR spillover estimator across replications. Bias$\sqrt{N}$: bias scaled by $\sqrt{N}$; bounded values confirm $\sqrt{N}$-consistency. SD($\hat{\psi}$): Monte Carlo standard deviation. RMSE($\hat{\psi}$): root mean squared error. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms the parametric rate. All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000,$ 5,000 replications per cell.

Table 22: DR Estimator Performance: Config K: Large n_1/n_0

Weight	N	B	Bias($\hat{\psi}$)	Bias$\sqrt{N}$	SD($\hat{\psi}$)	RMSE($\hat{\psi}$)	RMSE$\sqrt{N}$
DiD	200	5000	-0.0201	-0.284	0.6021	0.6024	8.519
DiD	500	5000	0.0021	0.047	0.3839	0.3839	8.584
DiD	1000	5000	-0.0014	-0.043	0.2765	0.2764	8.742
DiD	2000	5000	0.0005	0.024	0.1994	0.1993	8.914
DiD	5000	5000	-0.0016	-0.112	0.1280	0.1280	9.051
SC sparse	200	5000	-0.0201	-0.284	0.6021	0.6024	8.519
SC sparse	500	5000	0.0021	0.047	0.3839	0.3839	8.584
SC sparse	1000	5000	-0.0014	-0.043	0.2765	0.2764	8.742
SC sparse	2000	5000	0.0005	0.024	0.1994	0.1993	8.914
SC sparse	5000	5000	-0.0016	-0.112	0.1280	0.1280	9.051
SC dense	200	5000	-0.0201	-0.284	0.6021	0.6024	8.519
SC dense	500	5000	0.0021	0.047	0.3839	0.3839	8.584
SC dense	1000	5000	-0.0014	-0.043	0.2765	0.2764	8.742
SC dense	2000	5000	0.0005	0.024	0.1994	0.1993	8.914
SC dense	5000	5000	-0.0016	-0.112	0.1280	0.1280	9.051
Staggered	200	5000	-0.0201	-0.284	0.6021	0.6024	8.519
Staggered	500	5000	0.0021	0.047	0.3839	0.3839	8.584
Staggered	1000	5000	-0.0014	-0.043	0.2765	0.2764	8.742
Staggered	2000	5000	0.0005	0.024	0.1994	0.1993	8.914
Staggered	5000	5000	-0.0016	-0.112	0.1280	0.1280	9.051

Note:
Bias($\hat{\psi}$): mean bias of the DR spillover estimator across replications. Bias$\sqrt{N}$: bias scaled by $\sqrt{N}$; bounded values confirm $\sqrt{N}$-consistency. SD($\hat{\psi}$): Monte Carlo standard deviation. RMSE($\hat{\psi}$): root mean squared error. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms the parametric rate. All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000,$ 5,000 replications per cell.

S8.3 Table S1b: Bound Endpoint Estimation Performance (Configs A–K)

Bias and RMSE of the L_2 upper and lower bound plug-in estimators. Convergence of RMSE$\times$ $\sqrt{N}$ to a constant confirms $\sqrt{N}$-consistency of the bound endpoints.

Table 23: Bound Endpoint Performance: Config A: Baseline

Weight	N	Bias(UB)	RMSE(UB)	RMSE(UB)$\sqrt{N}$	Bias(LB)	RMSE(LB)	RMSE(LB)$\sqrt{N}$
DiD	200	-0.0010	0.0717	1.014	-1e-03	0.0717	1.014
DiD	500	-0.0005	0.0458	1.025	-5e-04	0.0458	1.025
DiD	1000	-0.0003	0.0343	1.085	-3e-04	0.0343	1.085
DiD	2000	-0.0002	0.0237	1.062	-2e-04	0.0237	1.062
DiD	5000	-0.0001	0.0153	1.085	-1e-04	0.0153	1.085
SC sparse	200	-0.0163	0.1166	1.649	0e+00	0.0000	0.000
SC sparse	500	0.0000	0.0026	0.058	0e+00	0.0000	0.000
SC sparse	1000	0.0000	0.0000	0.000	0e+00	0.0000	0.000
SC sparse	2000	0.0000	0.0000	0.000	0e+00	0.0000	0.000
SC sparse	5000	0.0000	0.0000	0.000	0e+00	0.0000	0.000
SC dense	200	-0.0310	0.2335	3.302	1e-04	0.0037	0.052
SC dense	500	-0.0123	0.1377	3.079	0e+00	0.0000	0.000
SC dense	1000	-0.0067	0.1007	3.185	0e+00	0.0000	0.000
SC dense	2000	-0.0035	0.0688	3.075	0e+00	0.0000	0.000
SC dense	5000	-0.0016	0.0440	3.111	0e+00	0.0000	0.000
Staggered	200	-0.0111	0.1116	1.578	-3e-04	0.0398	0.562
Staggered	500	-0.0046	0.0654	1.462	-2e-04	0.0255	0.571
Staggered	1000	-0.0023	0.0462	1.462	-2e-04	0.0193	0.611
Staggered	2000	-0.0007	0.0297	1.329	-1e-04	0.0135	0.602
Staggered	5000	-0.0002	0.0185	1.308	-1e-04	0.0088	0.619

Note:
Bias(UB), Bias(LB): mean bias of the L_2 upper and lower bound plugin estimators. RMSE(UB), RMSE(LB): root mean squared error of the bound endpoints. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms $\sqrt{N}$-consistency of the plugin bound estimators (Proposition 5.2). All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

Table 24: Bound Endpoint Performance: Config B: No spillover

Weight	N	Bias(UB)	RMSE(UB)	RMSE(UB)$\sqrt{N}$	Bias(LB)	RMSE(LB)	RMSE(LB)$\sqrt{N}$
DiD	500	-0.0003	0.0458	1.025	-3e-04	0.0458	1.025
DiD	2000	0.0009	0.0257	1.148	9e-04	0.0257	1.148
SC sparse	500	-0.1759	0.4383	9.800	0e+00	0.0000	0.000
SC sparse	2000	-0.0024	0.0488	2.183	0e+00	0.0000	0.000
SC dense	500	-0.0424	0.1898	4.243	5e-04	0.0044	0.099
SC dense	2000	-0.0092	0.0937	4.192	0e+00	0.0001	0.006
Staggered	500	-0.0116	0.0718	1.606	2e-04	0.0263	0.588
Staggered	2000	-0.0028	0.0384	1.718	5e-04	0.0146	0.653

Note:
Bias(UB), Bias(LB): mean bias of the L_2 upper and lower bound plugin estimators. RMSE(UB), RMSE(LB): root mean squared error of the bound endpoints. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms $\sqrt{N}$-consistency of the plugin bound estimators (Proposition 5.2). All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

Table 25: Bound Endpoint Performance: Config C: Uniform spillover

Weight	N	Bias(UB)	RMSE(UB)	RMSE(UB)$\sqrt{N}$	Bias(LB)	RMSE(LB)	RMSE(LB)$\sqrt{N}$
DiD	500	0.0000	0.0565	1.263	0e+00	0.0565	1.263
DiD	2000	-0.0003	0.0288	1.288	-3e-04	0.0288	1.288
SC sparse	500	0.0000	0.0000	0.000	0e+00	0.0000	0.000
SC sparse	2000	0.0000	0.0000	0.000	0e+00	0.0000	0.000
SC dense	500	-0.0088	0.1432	3.202	0e+00	0.0000	0.000
SC dense	2000	-0.0030	0.0717	3.206	0e+00	0.0000	0.000
Staggered	500	-0.0032	0.0760	1.700	0e+00	0.0314	0.703
Staggered	2000	-0.0005	0.0350	1.565	-2e-04	0.0163	0.731

Note:
Bias(UB), Bias(LB): mean bias of the L_2 upper and lower bound plugin estimators. RMSE(UB), RMSE(LB): root mean squared error of the bound endpoints. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms $\sqrt{N}$-consistency of the plugin bound estimators (Proposition 5.2). All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

Table 26: Bound Endpoint Performance: Config D: Strong spillover

Weight	N	Bias(UB)	RMSE(UB)	RMSE(UB)$\sqrt{N}$	Bias(LB)	RMSE(LB)	RMSE(LB)$\sqrt{N}$
DiD	200	0.0053	0.0624	0.882	0.0053	0.0624	0.882
DiD	500	0.0022	0.0441	0.986	0.0022	0.0441	0.986
DiD	1000	-0.0003	0.0328	1.038	-0.0003	0.0328	1.038
DiD	2000	-0.0006	0.0240	1.072	-0.0006	0.0240	1.072
DiD	5000	0.0002	0.0154	1.087	0.0002	0.0154	1.087
SC sparse	200	-0.1900	0.5279	7.466	0.0000	0.0000	0.000
SC sparse	500	-0.1512	0.3909	8.742	0.0000	0.0000	0.000
SC sparse	1000	-0.0423	0.2046	6.471	0.0000	0.0000	0.000
SC sparse	2000	-0.0054	0.0810	3.620	0.0000	0.0000	0.000
SC sparse	5000	0.0000	0.0000	0.000	0.0000	0.0000	0.000
SC dense	200	-0.0308	0.3544	5.012	0.0000	0.0000	0.000
SC dense	500	-0.0236	0.2616	5.849	0.0000	0.0000	0.000
SC dense	1000	-0.0219	0.1966	6.217	0.0000	0.0000	0.000
SC dense	2000	-0.0139	0.1405	6.282	0.0000	0.0000	0.000
SC dense	5000	-0.0030	0.0864	6.112	0.0000	0.0000	0.000
Staggered	200	0.0039	0.1355	1.916	0.0030	0.0340	0.481
Staggered	500	0.0023	0.1007	2.253	0.0012	0.0245	0.548
Staggered	1000	-0.0010	0.0766	2.423	-0.0002	0.0185	0.584
Staggered	2000	-0.0014	0.0563	2.518	-0.0004	0.0136	0.608
Staggered	5000	0.0004	0.0360	2.544	0.0001	0.0088	0.619

Note:
Bias(UB), Bias(LB): mean bias of the L_2 upper and lower bound plugin estimators. RMSE(UB), RMSE(LB): root mean squared error of the bound endpoints. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms $\sqrt{N}$-consistency of the plugin bound estimators (Proposition 5.2). All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

Table 27: Bound Endpoint Performance: Config E: Spatial decay

Weight	N	Bias(UB)	RMSE(UB)	RMSE(UB)$\sqrt{N}$	Bias(LB)	RMSE(LB)	RMSE(LB)$\sqrt{N}$
DiD	200	0.0013	0.0730	1.032	0.0013	0.0730	1.032
DiD	500	-0.0006	0.0475	1.061	-0.0006	0.0475	1.061
DiD	1000	0.0011	0.0336	1.063	0.0011	0.0336	1.063
DiD	2000	-0.0003	0.0244	1.091	-0.0003	0.0244	1.091
DiD	5000	-0.0002	0.0154	1.088	-0.0002	0.0154	1.088
SC sparse	200	-0.0004	0.0256	0.362	0.0000	0.0000	0.000
SC sparse	500	0.0000	0.0000	0.000	0.0000	0.0000	0.000
SC sparse	1000	0.0000	0.0000	0.000	0.0000	0.0000	0.000
SC sparse	2000	0.0000	0.0000	0.000	0.0000	0.0000	0.000
SC sparse	5000	0.0000	0.0000	0.000	0.0000	0.0000	0.000
SC dense	200	-0.0123	0.1296	1.833	0.0004	0.0028	0.039
SC dense	500	-0.0069	0.0779	1.742	0.0000	0.0005	0.010
SC dense	1000	-0.0013	0.0538	1.701	0.0000	0.0000	0.001
SC dense	2000	-0.0021	0.0391	1.748	0.0000	0.0001	0.003
SC dense	5000	-0.0009	0.0243	1.719	0.0000	0.0000	0.000
Staggered	200	-0.0028	0.0822	1.163	0.0015	0.0430	0.608
Staggered	500	-0.0022	0.0552	1.234	-0.0001	0.0276	0.617
Staggered	1000	0.0010	0.0400	1.265	0.0007	0.0195	0.617
Staggered	2000	-0.0005	0.0293	1.311	-0.0001	0.0141	0.629
Staggered	5000	-0.0002	0.0185	1.307	-0.0001	0.0088	0.623

Note:
Bias(UB), Bias(LB): mean bias of the L_2 upper and lower bound plugin estimators. RMSE(UB), RMSE(LB): root mean squared error of the bound endpoints. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms $\sqrt{N}$-consistency of the plugin bound estimators (Proposition 5.2). All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

Table 28: Bound Endpoint Performance: Config F: Heterogeneous

Weight	N	Bias(UB)	RMSE(UB)	RMSE(UB)$\sqrt{N}$	Bias(LB)	RMSE(LB)	RMSE(LB)$\sqrt{N}$
DiD	200	0.0002	0.0714	1.010	2e-04	0.0714	1.010
DiD	500	0.0003	0.0473	1.058	3e-04	0.0473	1.058
DiD	1000	-0.0008	0.0334	1.056	-8e-04	0.0334	1.056
DiD	2000	0.0007	0.0239	1.069	7e-04	0.0239	1.069
DiD	5000	-0.0002	0.0152	1.075	-2e-04	0.0152	1.075
SC sparse	200	-0.0039	0.0455	0.644	0e+00	0.0000	0.000
SC sparse	500	0.0000	0.0006	0.014	0e+00	0.0000	0.000
SC sparse	1000	0.0000	0.0000	0.000	0e+00	0.0000	0.000
SC sparse	2000	0.0000	0.0000	0.000	0e+00	0.0000	0.000
SC sparse	5000	0.0000	0.0000	0.000	0e+00	0.0000	0.000
SC dense	200	-0.0185	0.1870	2.645	0e+00	0.0012	0.017
SC dense	500	-0.0073	0.1141	2.551	0e+00	0.0000	0.000
SC dense	1000	-0.0057	0.0777	2.457	0e+00	0.0000	0.000
SC dense	2000	-0.0004	0.0547	2.448	0e+00	0.0000	0.000
SC dense	5000	-0.0014	0.0346	2.445	0e+00	0.0000	0.000
Staggered	200	-0.0045	0.0965	1.365	7e-04	0.0403	0.569
Staggered	500	-0.0008	0.0601	1.343	3e-04	0.0266	0.595
Staggered	1000	-0.0011	0.0406	1.285	-4e-04	0.0189	0.598
Staggered	2000	0.0008	0.0288	1.288	4e-04	0.0136	0.608
Staggered	5000	-0.0003	0.0182	1.290	-1e-04	0.0087	0.613

Note:
Bias(UB), Bias(LB): mean bias of the L_2 upper and lower bound plugin estimators. RMSE(UB), RMSE(LB): root mean squared error of the bound endpoints. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms $\sqrt{N}$-consistency of the plugin bound estimators (Proposition 5.2). All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

Table 29: Bound Endpoint Performance: Config G: Wrong μ_0

Weight	N	Bias(UB)	RMSE(UB)	RMSE(UB)$\sqrt{N}$	Bias(LB)	RMSE(LB)	RMSE(LB)$\sqrt{N}$
DiD	200	0.0022	0.0832	1.176	0.0022	0.0832	1.176
DiD	500	0.0018	0.0543	1.214	0.0018	0.0543	1.214
DiD	1000	0.0003	0.0387	1.224	0.0003	0.0387	1.224
DiD	2000	0.0001	0.0276	1.235	0.0001	0.0276	1.235
DiD	5000	0.0002	0.0177	1.251	0.0002	0.0177	1.251
SC sparse	200	-0.0258	0.1636	2.313	0.0000	0.0000	0.000
SC sparse	500	-0.0007	0.0202	0.452	0.0000	0.0000	0.000
SC sparse	1000	-0.0003	0.0185	0.586	0.0000	0.0000	0.000
SC sparse	2000	0.0000	0.0000	0.000	0.0000	0.0000	0.000
SC sparse	5000	0.0000	0.0000	0.000	0.0000	0.0000	0.000
SC dense	200	-0.0324	0.2800	3.959	0.0000	0.0005	0.008
SC dense	500	-0.0086	0.1636	3.658	0.0000	0.0000	0.000
SC dense	1000	-0.0063	0.1157	3.660	0.0000	0.0000	0.000
SC dense	2000	-0.0032	0.0796	3.560	0.0000	0.0000	0.000
SC dense	5000	-0.0010	0.0506	3.578	0.0000	0.0000	0.000
Staggered	200	-0.0089	0.1305	1.846	0.0015	0.0457	0.646
Staggered	500	-0.0026	0.0785	1.756	0.0010	0.0303	0.677
Staggered	1000	-0.0019	0.0524	1.658	0.0002	0.0218	0.688
Staggered	2000	-0.0005	0.0348	1.554	0.0001	0.0156	0.699
Staggered	5000	0.0002	0.0214	1.511	0.0001	0.0101	0.713

Note:
Bias(UB), Bias(LB): mean bias of the L_2 upper and lower bound plugin estimators. RMSE(UB), RMSE(LB): root mean squared error of the bound endpoints. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms $\sqrt{N}$-consistency of the plugin bound estimators (Proposition 5.2). All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

Table 30: Bound Endpoint Performance: Config H: Wrong π

Weight	N	Bias(UB)	RMSE(UB)	RMSE(UB)$\sqrt{N}$	Bias(LB)	RMSE(LB)	RMSE(LB)$\sqrt{N}$
DiD	200	0.0010	0.0826	1.168	1e-03	0.0826	1.168
DiD	500	0.0004	0.0552	1.234	4e-04	0.0552	1.234
DiD	1000	-0.0002	0.0410	1.298	-2e-04	0.0410	1.298
DiD	2000	-0.0003	0.0292	1.307	-3e-04	0.0292	1.307
DiD	5000	0.0001	0.0187	1.320	1e-04	0.0187	1.320
SC sparse	200	-0.0240	0.1401	1.981	0e+00	0.0000	0.000
SC sparse	500	-0.0010	0.0292	0.653	0e+00	0.0000	0.000
SC sparse	1000	0.0000	0.0023	0.072	0e+00	0.0000	0.000
SC sparse	2000	0.0000	0.0000	0.000	0e+00	0.0000	0.000
SC sparse	5000	0.0000	0.0000	0.000	0e+00	0.0000	0.000
SC dense	200	-0.0351	0.2744	3.880	1e-04	0.0069	0.097
SC dense	500	-0.0152	0.1686	3.769	0e+00	0.0002	0.004
SC dense	1000	-0.0093	0.1228	3.882	0e+00	0.0000	0.000
SC dense	2000	-0.0054	0.0851	3.804	0e+00	0.0000	0.000
SC dense	5000	-0.0014	0.0538	3.805	0e+00	0.0000	0.000
Staggered	200	-0.0111	0.1301	1.840	8e-04	0.0459	0.649
Staggered	500	-0.0059	0.0811	1.813	3e-04	0.0308	0.689
Staggered	1000	-0.0033	0.0569	1.801	-1e-04	0.0231	0.730
Staggered	2000	-0.0014	0.0375	1.678	-2e-04	0.0166	0.741
Staggered	5000	0.0000	0.0227	1.606	1e-04	0.0106	0.752

Note:
Bias(UB), Bias(LB): mean bias of the L_2 upper and lower bound plugin estimators. RMSE(UB), RMSE(LB): root mean squared error of the bound endpoints. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms $\sqrt{N}$-consistency of the plugin bound estimators (Proposition 5.2). All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

Table 31: Bound Endpoint Performance: Config I: Both wrong

Weight	N	Bias(UB)	RMSE(UB)	RMSE(UB)$\sqrt{N}$	Bias(LB)	RMSE(LB)	RMSE(LB)$\sqrt{N}$
DiD	200	0.0307	0.1689	2.389	0.0307	0.1689	2.389
DiD	500	0.0076	0.1212	2.711	0.0076	0.1212	2.711
DiD	1000	0.0040	0.0946	2.991	0.0040	0.0946	2.991
DiD	2000	-0.0005	0.0712	3.182	-0.0005	0.0712	3.182
DiD	5000	-0.0001	0.0468	3.308	-0.0001	0.0468	3.308
SC sparse	200	-0.2760	0.6072	8.587	0.0011	0.0339	0.479
SC sparse	500	-0.1800	0.4691	10.490	0.0000	0.0000	0.000
SC sparse	1000	-0.0846	0.3341	10.565	0.0000	0.0000	0.000
SC sparse	2000	-0.0252	0.1741	7.787	0.0000	0.0000	0.000
SC sparse	5000	-0.0012	0.0373	2.637	0.0000	0.0000	0.000
SC dense	200	-0.0950	0.5239	7.409	0.0092	0.0671	0.949
SC dense	500	-0.1045	0.4450	9.949	0.0012	0.0166	0.372
SC dense	1000	-0.0710	0.3600	11.383	0.0001	0.0017	0.055
SC dense	2000	-0.0500	0.2715	12.141	0.0000	0.0001	0.006
SC dense	5000	-0.0196	0.1656	11.712	0.0000	0.0000	0.000
Staggered	200	-0.0058	0.2381	3.368	0.0208	0.1055	1.492
Staggered	500	-0.0265	0.1932	4.319	0.0049	0.0694	1.552
Staggered	1000	-0.0242	0.1585	5.012	0.0024	0.0536	1.695
Staggered	2000	-0.0226	0.1226	5.483	-0.0002	0.0404	1.807
Staggered	5000	-0.0134	0.0789	5.576	-0.0001	0.0267	1.885

Note:
Bias(UB), Bias(LB): mean bias of the L_2 upper and lower bound plugin estimators. RMSE(UB), RMSE(LB): root mean squared error of the bound endpoints. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms $\sqrt{N}$-consistency of the plugin bound estimators (Proposition 5.2). All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

Table 32: Bound Endpoint Performance: Config J: Small n_1/n_0

Weight	N	Bias(UB)	RMSE(UB)	RMSE(UB)$\sqrt{N}$	Bias(LB)	RMSE(LB)	RMSE(LB)$\sqrt{N}$
DiD	200	0.0288	0.0523	0.740	0.0288	0.0523	0.740
DiD	500	0.0185	0.0332	0.742	0.0185	0.0332	0.742
DiD	1000	0.0128	0.0231	0.731	0.0128	0.0231	0.731
DiD	2000	0.0095	0.0171	0.763	0.0095	0.0171	0.763
DiD	5000	0.0062	0.0108	0.766	0.0062	0.0108	0.766
SC sparse	200	0.3872	0.6585	9.313	0.0000	0.0000	0.000
SC sparse	500	0.4219	0.6910	15.452	0.0000	0.0000	0.000
SC sparse	1000	0.4591	0.7408	23.427	0.0000	0.0000	0.000
SC sparse	2000	0.5169	0.8016	35.849	0.0000	0.0000	0.000
SC sparse	5000	0.5549	0.8472	59.908	0.0000	0.0000	0.000
SC dense	200	0.2154	0.3662	5.179	0.0000	0.0012	0.017
SC dense	500	0.1612	0.2728	6.099	0.0000	0.0000	0.000
SC dense	1000	0.1246	0.2111	6.674	0.0000	0.0000	0.000
SC dense	2000	0.1011	0.1696	7.584	0.0000	0.0000	0.000
SC dense	5000	0.0729	0.1210	8.556	0.0000	0.0000	0.000
Staggered	200	0.0677	0.1216	1.719	0.0156	0.0285	0.403
Staggered	500	0.0435	0.0785	1.755	0.0103	0.0184	0.412
Staggered	1000	0.0301	0.0546	1.725	0.0072	0.0130	0.412
Staggered	2000	0.0224	0.0401	1.793	0.0054	0.0097	0.433
Staggered	5000	0.0144	0.0254	1.793	0.0035	0.0062	0.437

Note:
Bias(UB), Bias(LB): mean bias of the L_2 upper and lower bound plugin estimators. RMSE(UB), RMSE(LB): root mean squared error of the bound endpoints. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms $\sqrt{N}$-consistency of the plugin bound estimators (Proposition 5.2). All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

Table 33: Bound Endpoint Performance: Config K: Large n_1/n_0

Weight	N	Bias(UB)	RMSE(UB)	RMSE(UB)$\sqrt{N}$	Bias(LB)	RMSE(LB)	RMSE(LB)$\sqrt{N}$
DiD	200	0.0093	0.1562	2.209	0.0093	0.1562	2.209
DiD	500	0.0033	0.1096	2.451	0.0033	0.1096	2.451
DiD	1000	-0.0001	0.0820	2.594	-0.0001	0.0820	2.594
DiD	2000	0.0002	0.0598	2.674	0.0002	0.0598	2.674
DiD	5000	-0.0005	0.0384	2.715	-0.0005	0.0384	2.715
SC sparse	200	-0.2471	0.5975	8.449	0.0000	0.0000	0.000
SC sparse	500	-0.0819	0.3303	7.385	0.0000	0.0000	0.000
SC sparse	1000	-0.0178	0.1730	5.472	0.0000	0.0000	0.000
SC sparse	2000	-0.0006	0.0260	1.164	0.0000	0.0000	0.000
SC sparse	5000	0.0000	0.0000	0.000	0.0000	0.0000	0.000
SC dense	200	-0.1302	0.5631	7.964	0.0016	0.0166	0.235
SC dense	500	-0.0739	0.4080	9.122	0.0001	0.0022	0.049
SC dense	1000	-0.0463	0.2947	9.319	0.0000	0.0004	0.014
SC dense	2000	-0.0222	0.2013	9.001	0.0000	0.0000	0.000
SC dense	5000	-0.0106	0.1241	8.777	0.0000	0.0000	0.000
Staggered	200	-0.0310	0.2472	3.496	0.0066	0.0887	1.255
Staggered	500	-0.0241	0.1804	4.034	0.0021	0.0616	1.377
Staggered	1000	-0.0204	0.1361	4.305	0.0000	0.0462	1.462
Staggered	2000	-0.0125	0.0957	4.281	0.0001	0.0339	1.518
Staggered	5000	-0.0064	0.0576	4.070	-0.0003	0.0219	1.549

Note:
Bias(UB), Bias(LB): mean bias of the L_2 upper and lower bound plugin estimators. RMSE(UB), RMSE(LB): root mean squared error of the bound endpoints. RMSE$\sqrt{N}$: RMSE scaled by $\sqrt{N}$; convergence to a constant confirms $\sqrt{N}$-consistency of the plugin bound estimators (Proposition 5.2). All 11 configurations, 4 weight types, $N \in 200, 500, 1000, 2000$, 5,000 replications per cell.

S8.4 Table S2: Coverage Panels (Configs A–K)

Empirical coverage of the 95% Imbens–Manski confidence intervals, endpoint-specific coverage, and CI length across all configurations and sample sizes.

Table 34: Coverage Panel: Config A: Baseline

Weight	N	Cov L_2	Cov L_3	Cov UB	Cov LB	CI length
DiD	200	0.979	0.999	0.986	0.993	0.312
DiD	500	0.982	1.000	0.989	0.993	0.209
DiD	1000	0.977	1.000	0.987	0.991	0.154
DiD	2000	0.984	1.000	0.992	0.992	0.111
DiD	5000	0.981	1.000	0.991	0.990	0.071
SC sparse	200	1.000	0.999	1.000	1.000	1.437
SC sparse	500	1.000	1.000	1.000	1.000	1.441
SC sparse	1000	1.000	1.000	1.000	1.000	1.462
SC sparse	2000	1.000	1.000	1.000	1.000	1.447
SC sparse	5000	1.000	1.000	1.000	1.000	1.446
SC dense	200	1.000	1.000	0.994	1.000	1.707
SC dense	500	1.000	1.000	0.990	1.000	1.608
SC dense	1000	1.000	1.000	0.984	1.000	1.548
SC dense	2000	1.000	1.000	0.985	1.000	1.505
SC dense	5000	1.000	1.000	0.981	1.000	1.458
Staggered	200	1.000	1.000	0.983	0.983	0.629
Staggered	500	1.000	1.000	0.987	0.980	0.547
Staggered	1000	1.000	1.000	0.985	0.977	0.499
Staggered	2000	1.000	1.000	0.983	0.978	0.464
Staggered	5000	1.000	1.000	0.976	0.975	0.431

Note:
Cov L_2: empirical coverage of the 95

Table 35: Coverage Panel: Config B: No spillover

Weight	N	Cov L_2	Cov L_3	Cov UB	Cov LB	CI length
DiD	500	0.994	1.000	0.999	0.995	0.225
DiD	2000	0.991	1.000	0.997	0.994	0.126
SC sparse	500	1.000	0.999	1.000	1.000	1.769
SC sparse	2000	1.000	1.000	1.000	1.000	1.480
SC dense	500	1.000	1.000	1.000	1.000	0.872
SC dense	2000	1.000	1.000	1.000	1.000	0.719
Staggered	500	1.000	1.000	1.000	0.991	0.320
Staggered	2000	1.000	1.000	0.998	0.984	0.230

Note:
Cov L_2: empirical coverage of the 95

Table 36: Coverage Panel: Config C: Uniform spillover

Weight	N	Cov L_2	Cov L_3	Cov UB	Cov LB	CI length
DiD	500	0.943	0.996	0.966	0.977	0.211
DiD	2000	0.942	0.999	0.967	0.975	0.108
SC sparse	500	1.000	1.000	1.000	1.000	1.460
SC sparse	2000	1.000	1.000	1.000	1.000	1.447
SC dense	500	1.000	1.000	0.955	1.000	2.185
SC dense	2000	1.000	1.000	0.949	1.000	2.096
Staggered	500	1.000	1.000	0.958	0.951	0.795
Staggered	2000	1.000	1.000	0.944	0.946	0.713

Note:
Cov L_2: empirical coverage of the 95

Table 37: Coverage Panel: Config D: Strong spillover

Weight	N	Cov L_2	Cov L_3	Cov UB	Cov LB	CI length
DiD	200	0.983	0.984	1.000	0.983	0.280
DiD	500	0.981	0.998	0.998	0.983	0.191
DiD	1000	0.980	0.999	0.992	0.988	0.142
DiD	2000	0.972	1.000	0.984	0.988	0.103
DiD	5000	0.971	1.000	0.986	0.986	0.067
SC sparse	200	1.000	0.964	1.000	1.000	1.757
SC sparse	500	1.000	0.995	1.000	1.000	1.492
SC sparse	1000	1.000	0.998	1.000	1.000	1.301
SC sparse	2000	1.000	1.000	1.000	1.000	1.279
SC sparse	5000	1.000	1.000	1.000	1.000	1.260
SC dense	200	1.000	0.985	1.000	1.000	1.319
SC dense	500	1.000	0.999	1.000	1.000	1.095
SC dense	1000	1.000	1.000	1.000	1.000	0.955
SC dense	2000	1.000	1.000	0.995	1.000	0.855
SC dense	5000	1.000	1.000	0.980	1.000	0.775
Staggered	200	0.996	0.985	1.000	0.961	0.480
Staggered	500	0.999	0.999	0.996	0.960	0.375
Staggered	1000	1.000	1.000	0.989	0.970	0.309
Staggered	2000	1.000	1.000	0.970	0.970	0.260
Staggered	5000	1.000	1.000	0.965	0.967	0.218

Note:
Cov L_2: empirical coverage of the 95

Table 38: Coverage Panel: Config E: Spatial decay

Weight	N	Cov L_2	Cov L_3	Cov UB	Cov LB	CI length
DiD	200	0.994	1	0.995	0.999	0.386
DiD	500	0.994	1	0.997	0.998	0.253
DiD	1000	0.996	1	0.997	0.999	0.184
DiD	2000	0.996	1	0.997	0.998	0.132
DiD	5000	0.994	1	0.996	0.998	0.084
SC sparse	200	1.000	1	1.000	1.000	1.671
SC sparse	500	1.000	1	1.000	1.000	1.733
SC sparse	1000	1.000	1	1.000	1.000	1.727
SC sparse	2000	1.000	1	1.000	1.000	1.744
SC sparse	5000	1.000	1	1.000	1.000	1.743
SC dense	200	1.000	1	0.997	1.000	2.280
SC dense	500	1.000	1	0.996	1.000	2.209
SC dense	1000	1.000	1	0.995	1.000	2.174
SC dense	2000	1.000	1	0.995	1.000	2.141
SC dense	5000	1.000	1	0.992	1.000	2.112
Staggered	200	1.000	1	0.988	0.995	0.846
Staggered	500	1.000	1	0.988	0.992	0.762
Staggered	1000	1.000	1	0.988	0.995	0.714
Staggered	2000	1.000	1	0.990	0.994	0.673
Staggered	5000	1.000	1	0.989	0.993	0.637

Note:
Cov L_2: empirical coverage of the 95

Table 39: Coverage Panel: Config F: Heterogeneous

Weight	N	Cov L_2	Cov L_3	Cov UB	Cov LB	CI length
DiD	200	0.996	1.000	0.997	0.999	0.403
DiD	500	0.996	1.000	0.997	0.999	0.264
DiD	1000	0.996	1.000	0.997	0.999	0.191
DiD	2000	0.996	1.000	0.997	0.999	0.137
DiD	5000	0.995	1.000	0.997	0.999	0.087
SC sparse	200	1.000	0.996	1.000	1.000	1.467
SC sparse	500	1.000	0.999	1.000	1.000	1.528
SC sparse	1000	1.000	1.000	1.000	1.000	1.538
SC sparse	2000	1.000	1.000	1.000	1.000	1.511
SC sparse	5000	1.000	1.000	1.000	1.000	1.540
SC dense	200	1.000	1.000	0.999	1.000	1.993
SC dense	500	1.000	1.000	0.998	1.000	1.883
SC dense	1000	1.000	1.000	0.998	1.000	1.813
SC dense	2000	1.000	1.000	0.996	1.000	1.780
SC dense	5000	1.000	1.000	0.994	1.000	1.732
Staggered	200	1.000	1.000	0.995	0.996	0.761
Staggered	500	1.000	1.000	0.994	0.994	0.653
Staggered	1000	1.000	1.000	0.993	0.995	0.594
Staggered	2000	1.000	1.000	0.994	0.995	0.554
Staggered	5000	1.000	1.000	0.991	0.994	0.514

Note:
Cov L_2: empirical coverage of the 95

Table 40: Coverage Panel: Config G: Wrong μ_0

Weight	N	Cov L_2	Cov L_3	Cov UB	Cov LB	CI length
DiD	200	0.975	0.999	0.983	0.993	0.351
DiD	500	0.979	1.000	0.989	0.989	0.236
DiD	1000	0.978	1.000	0.988	0.990	0.174
DiD	2000	0.986	1.000	0.993	0.993	0.126
DiD	5000	0.978	1.000	0.991	0.987	0.081
SC sparse	200	1.000	0.997	1.000	1.000	1.547
SC sparse	500	1.000	1.000	1.000	1.000	1.576
SC sparse	1000	1.000	1.000	1.000	1.000	1.554
SC sparse	2000	1.000	1.000	1.000	1.000	1.559
SC sparse	5000	1.000	1.000	1.000	1.000	1.573
SC dense	200	1.000	1.000	0.994	1.000	1.865
SC dense	500	1.000	1.000	0.991	1.000	1.751
SC dense	1000	1.000	1.000	0.987	1.000	1.688
SC dense	2000	1.000	1.000	0.987	1.000	1.630
SC dense	5000	1.000	1.000	0.979	1.000	1.579
Staggered	200	1.000	1.000	0.982	0.980	0.699
Staggered	500	1.000	1.000	0.988	0.974	0.597
Staggered	1000	1.000	1.000	0.989	0.974	0.547
Staggered	2000	1.000	1.000	0.987	0.975	0.505
Staggered	5000	1.000	1.000	0.975	0.974	0.469

Note:
Cov L_2: empirical coverage of the 95

Table 41: Coverage Panel: Config H: Wrong π

Weight	N	Cov L_2	Cov L_3	Cov UB	Cov LB	CI length
DiD	200	0.984	0.999	0.990	0.994	0.355
DiD	500	0.986	1.000	0.991	0.995	0.245
DiD	1000	0.983	1.000	0.991	0.992	0.184
DiD	2000	0.980	1.000	0.990	0.990	0.135
DiD	5000	0.982	1.000	0.990	0.992	0.087
SC sparse	200	1.000	0.999	1.000	1.000	1.439
SC sparse	500	1.000	1.000	1.000	1.000	1.456
SC sparse	1000	1.000	1.000	1.000	1.000	1.446
SC sparse	2000	1.000	1.000	1.000	1.000	1.455
SC sparse	5000	1.000	1.000	1.000	1.000	1.453
SC dense	200	1.000	1.000	0.998	1.000	1.784
SC dense	500	1.000	1.000	0.994	1.000	1.657
SC dense	1000	1.000	1.000	0.990	1.000	1.589
SC dense	2000	1.000	1.000	0.988	1.000	1.536
SC dense	5000	1.000	1.000	0.983	1.000	1.483
Staggered	200	1.000	1.000	0.991	0.982	0.677
Staggered	500	1.000	1.000	0.989	0.980	0.577
Staggered	1000	1.000	1.000	0.990	0.979	0.526
Staggered	2000	1.000	1.000	0.987	0.978	0.483
Staggered	5000	1.000	1.000	0.980	0.976	0.445

Note:
Cov L_2: empirical coverage of the 95

Table 42: Coverage Panel: Config I: Both wrong

Weight	N	Cov L_2	Cov L_3	Cov UB	Cov LB	CI length
DiD	200	0.985	0.963	1.000	0.985	0.792
DiD	500	0.989	0.991	1.000	0.989	0.570
DiD	1000	0.989	0.999	1.000	0.989	0.431
DiD	2000	0.983	1.000	0.997	0.986	0.314
DiD	5000	0.974	1.000	0.986	0.988	0.205
SC sparse	200	0.976	0.871	1.000	1.000	1.608
SC sparse	500	0.988	0.949	1.000	1.000	1.354
SC sparse	1000	0.984	0.969	1.000	1.000	1.162
SC sparse	2000	0.982	0.977	1.000	1.000	1.097
SC sparse	5000	0.982	0.981	1.000	1.000	1.081
SC dense	200	1.000	0.960	1.000	1.000	2.322
SC dense	500	1.000	0.991	1.000	1.000	1.863
SC dense	1000	1.000	0.999	1.000	1.000	1.609
SC dense	2000	1.000	1.000	1.000	1.000	1.403
SC dense	5000	1.000	1.000	0.999	1.000	1.236
Staggered	200	0.998	0.963	1.000	0.983	1.040
Staggered	500	1.000	0.991	1.000	0.978	0.807
Staggered	1000	1.000	0.999	1.000	0.976	0.665
Staggered	2000	1.000	1.000	0.999	0.972	0.545
Staggered	5000	1.000	1.000	0.982	0.971	0.445

Note:
Cov L_2: empirical coverage of the 95

Table 43: Coverage Panel: Config J: Small n_1/n_0

Weight	N	Cov L_2	Cov L_3	Cov UB	Cov LB	CI length
DiD	200	0.974	0.999	1	0.974	0.266
DiD	500	0.980	1.000	1	0.980	0.184
DiD	1000	0.985	1.000	1	0.985	0.137
DiD	2000	0.981	1.000	1	0.981	0.100
DiD	5000	0.983	1.000	1	0.983	0.064
SC sparse	200	1.000	0.999	1	1.000	2.306
SC sparse	500	1.000	1.000	1	1.000	3.097
SC sparse	1000	1.000	1.000	1	1.000	4.119
SC sparse	2000	1.000	1.000	1	1.000	5.556
SC sparse	5000	1.000	1.000	1	1.000	7.936
SC dense	200	1.000	1.000	1	1.000	1.249
SC dense	500	1.000	1.000	1	1.000	1.023
SC dense	1000	1.000	1.000	1	1.000	0.855
SC dense	2000	1.000	1.000	1	1.000	0.686
SC dense	5000	1.000	1.000	1	1.000	0.500
Staggered	200	0.951	0.997	1	0.951	0.415
Staggered	500	0.961	0.999	1	0.961	0.284
Staggered	1000	0.964	1.000	1	0.964	0.210
Staggered	2000	0.963	1.000	1	0.963	0.153
Staggered	5000	0.963	1.000	1	0.963	0.097

Note:
Cov L_2: empirical coverage of the 95

Table 44: Coverage Panel: Config K: Large n_1/n_0

Weight	N	Cov L_2	Cov L_3	Cov UB	Cov LB	CI length
DiD	200	0.987	0.985	1.000	0.987	0.707
DiD	500	0.984	0.997	0.998	0.986	0.479
DiD	1000	0.976	0.999	0.989	0.987	0.356
DiD	2000	0.972	1.000	0.984	0.988	0.258
DiD	5000	0.972	1.000	0.986	0.987	0.166
SC sparse	200	0.992	0.933	1.000	1.000	1.777
SC sparse	500	0.995	0.978	1.000	1.000	1.513
SC sparse	1000	0.994	0.992	1.000	1.000	1.412
SC sparse	2000	0.995	0.994	1.000	1.000	1.400
SC sparse	5000	0.996	0.996	1.000	1.000	1.405
SC dense	200	1.000	0.985	1.000	1.000	2.326
SC dense	500	1.000	0.998	1.000	1.000	1.956
SC dense	1000	1.000	1.000	1.000	1.000	1.738
SC dense	2000	1.000	1.000	0.996	1.000	1.603
SC dense	5000	1.000	1.000	0.985	1.000	1.480
Staggered	200	0.999	0.986	1.000	0.973	1.005
Staggered	500	1.000	0.998	0.998	0.972	0.778
Staggered	1000	1.000	1.000	0.991	0.972	0.660
Staggered	2000	1.000	1.000	0.980	0.967	0.569
Staggered	5000	1.000	1.000	0.985	0.971	0.488

Note:
Cov L_2: empirical coverage of the 95

S8.5 Table S3: Detection Rates (Configs A–K)

Fraction of replications where the Imbens–Manski CI lower bound exceeds zero (detection rate) and the Wald test rejection rate at $\alpha = 0.05$, across all configurations and sample sizes.

Table 45: Detection Power: Config A: Baseline

Weight	N	Detection rate	Test rejects
DiD	200	0.870	0.87
DiD	500	0.990	0.99
DiD	1000	1.000	1.00
DiD	2000	1.000	1.00
DiD	5000	1.000	1.00
SC sparse	200	0.000	0.87
SC sparse	500	0.000	0.99
SC sparse	1000	0.000	1.00
SC sparse	2000	0.000	1.00
SC sparse	5000	0.000	1.00
SC dense	200	0.000	0.87
SC dense	500	0.000	0.99
SC dense	1000	0.000	1.00
SC dense	2000	0.000	1.00
SC dense	5000	0.000	1.00
Staggered	200	0.922	0.87
Staggered	500	0.995	0.99
Staggered	1000	1.000	1.00
Staggered	2000	1.000	1.00
Staggered	5000	1.000	1.00

Note:
Detection rate: fraction of replications where the Imbens–Manski CI lower bound exceeds zero, indicating statistically detectable interference contamination. Test rejects: rejection rate of the Wald test $H_0! : \psi = 0$ at $\alpha = 0.05$. Under DGPs with nonzero spillovers, both statistics measure the ability to detect interference. Under Config B (no spillover), they measure size (false positive rate). SC weights show low detection rates because partial identification produces wide CIs. 5,000 replications per cell.

Table 46: Detection Power: Config B: No spillover

Weight	N	Detection rate	Test rejects
DiD	500	0.273	0.273
DiD	2000	0.787	0.787
SC sparse	500	0.000	0.273
SC sparse	2000	0.000	0.787
SC dense	500	0.000	0.273
SC dense	2000	0.000	0.787
Staggered	500	0.400	0.273
Staggered	2000	0.872	0.787

Note:
Detection rate: fraction of replications where the Imbens–Manski CI lower bound exceeds zero, indicating statistically detectable interference contamination. Test rejects: rejection rate of the Wald test $H_0! : \psi = 0$ at $\alpha = 0.05$. Under DGPs with nonzero spillovers, both statistics measure the ability to detect interference. Under Config B (no spillover), they measure size (false positive rate). SC weights show low detection rates because partial identification produces wide CIs. 5,000 replications per cell.

Table 47: Detection Power: Config C: Uniform spillover

Weight	N	Detection rate	Test rejects
DiD	500	1	1
DiD	2000	1	1
SC sparse	500	0	1
SC sparse	2000	0	1
SC dense	500	0	1
SC dense	2000	0	1
Staggered	500	1	1
Staggered	2000	1	1

Note:
Detection rate: fraction of replications where the Imbens–Manski CI lower bound exceeds zero, indicating statistically detectable interference contamination. Test rejects: rejection rate of the Wald test $H_0! : \psi = 0$ at $\alpha = 0.05$. Under DGPs with nonzero spillovers, both statistics measure the ability to detect interference. Under Config B (no spillover), they measure size (false positive rate). SC weights show low detection rates because partial identification produces wide CIs. 5,000 replications per cell.

Table 48: Detection Power: Config D: Strong spillover

Weight	N	Detection rate	Test rejects
DiD	200	0.180	0.182
DiD	500	0.345	0.345
DiD	1000	0.540	0.540
DiD	2000	0.830	0.830
DiD	5000	0.996	0.996
SC sparse	200	0.000	0.182
SC sparse	500	0.000	0.345
SC sparse	1000	0.000	0.540
SC sparse	2000	0.000	0.830
SC sparse	5000	0.000	0.996
SC dense	200	0.000	0.182
SC dense	500	0.000	0.345
SC dense	1000	0.000	0.540
SC dense	2000	0.000	0.830
SC dense	5000	0.000	0.996
Staggered	200	0.271	0.182
Staggered	500	0.466	0.345
Staggered	1000	0.658	0.540
Staggered	2000	0.897	0.830
Staggered	5000	0.999	0.996

Note:
Detection rate: fraction of replications where the Imbens–Manski CI lower bound exceeds zero, indicating statistically detectable interference contamination. Test rejects: rejection rate of the Wald test $H_0! : \psi = 0$ at $\alpha = 0.05$. Under DGPs with nonzero spillovers, both statistics measure the ability to detect interference. Under Config B (no spillover), they measure size (false positive rate). SC weights show low detection rates because partial identification produces wide CIs. 5,000 replications per cell.

Table 49: Detection Power: Config E: Spatial decay

Weight	N	Detection rate	Test rejects
DiD	200	0.995	0.995
DiD	500	1.000	1.000
DiD	1000	1.000	1.000
DiD	2000	1.000	1.000
DiD	5000	1.000	1.000
SC sparse	200	0.000	0.995
SC sparse	500	0.000	1.000
SC sparse	1000	0.000	1.000
SC sparse	2000	0.000	1.000
SC sparse	5000	0.000	1.000
SC dense	200	0.000	0.995
SC dense	500	0.000	1.000
SC dense	1000	0.000	1.000
SC dense	2000	0.000	1.000
SC dense	5000	0.000	1.000
Staggered	200	0.997	0.995
Staggered	500	1.000	1.000
Staggered	1000	1.000	1.000
Staggered	2000	1.000	1.000
Staggered	5000	1.000	1.000

Note:
Detection rate: fraction of replications where the Imbens–Manski CI lower bound exceeds zero, indicating statistically detectable interference contamination. Test rejects: rejection rate of the Wald test $H_0! : \psi = 0$ at $\alpha = 0.05$. Under DGPs with nonzero spillovers, both statistics measure the ability to detect interference. Under Config B (no spillover), they measure size (false positive rate). SC weights show low detection rates because partial identification produces wide CIs. 5,000 replications per cell.

Table 50: Detection Power: Config F: Heterogeneous

Weight	N	Detection rate	Test rejects
DiD	200	0.932	0.932
DiD	500	0.998	0.998
DiD	1000	1.000	1.000
DiD	2000	1.000	1.000
DiD	5000	1.000	1.000
SC sparse	200	0.000	0.932
SC sparse	500	0.000	0.998
SC sparse	1000	0.000	1.000
SC sparse	2000	0.000	1.000
SC sparse	5000	0.000	1.000
SC dense	200	0.000	0.932
SC dense	500	0.000	0.998
SC dense	1000	0.000	1.000
SC dense	2000	0.000	1.000
SC dense	5000	0.000	1.000
Staggered	200	0.960	0.932
Staggered	500	0.999	0.998
Staggered	1000	1.000	1.000
Staggered	2000	1.000	1.000
Staggered	5000	1.000	1.000

Note:
Detection rate: fraction of replications where the Imbens–Manski CI lower bound exceeds zero, indicating statistically detectable interference contamination. Test rejects: rejection rate of the Wald test $H_0!$: $\psi = 0$ at $\alpha = 0.05$. Under DGPs with nonzero spillovers, both statistics measure the ability to detect interference. Under Config B (no spillover), they measure size (false positive rate). SC weights show low detection rates because partial identification produces wide CIs. 5,000 replications per cell.

Table 51: Detection Power: Config G: Wrong μ_0

Weight	N	Detection rate	Test rejects
DiD	200	0.850	0.850
DiD	500	0.985	0.985
DiD	1000	0.999	0.999
DiD	2000	1.000	1.000
DiD	5000	1.000	1.000
SC sparse	200	0.000	0.850
SC sparse	500	0.000	0.985
SC sparse	1000	0.000	0.999
SC sparse	2000	0.000	1.000
SC sparse	5000	0.000	1.000
SC dense	200	0.000	0.850
SC dense	500	0.000	0.985
SC dense	1000	0.000	0.999
SC dense	2000	0.000	1.000
SC dense	5000	0.000	1.000
Staggered	200	0.903	0.850
Staggered	500	0.993	0.985
Staggered	1000	0.999	0.999
Staggered	2000	1.000	1.000
Staggered	5000	1.000	1.000

Note:
Detection rate: fraction of replications where the Imbens–Manski CI lower bound exceeds zero, indicating statistically detectable interference contamination. Test rejects: rejection rate of the Wald test $H_0! : \psi = 0$ at $\alpha = 0.05$. Under DGPs with nonzero spillovers, both statistics measure the ability to detect interference. Under Config B (no spillover), they measure size (false positive rate). SC weights show low detection rates because partial identification produces wide CIs. 5,000 replications per cell.

Table 52: Detection Power: Config H: Wrong π

Weight	N	Detection rate	Test rejects
DiD	200	0.778	0.778
DiD	500	0.966	0.966
DiD	1000	0.996	0.996
DiD	2000	1.000	1.000
DiD	5000	1.000	1.000
SC sparse	200	0.000	0.778
SC sparse	500	0.000	0.966
SC sparse	1000	0.000	0.996
SC sparse	2000	0.000	1.000
SC sparse	5000	0.000	1.000
SC dense	200	0.000	0.778
SC dense	500	0.000	0.966
SC dense	1000	0.000	0.996
SC dense	2000	0.000	1.000
SC dense	5000	0.000	1.000
Staggered	200	0.848	0.778
Staggered	500	0.981	0.966
Staggered	1000	0.998	0.996
Staggered	2000	1.000	1.000
Staggered	5000	1.000	1.000

Note:
Detection rate: fraction of replications where the Imbens–Manski CI lower bound exceeds zero, indicating statistically detectable interference contamination. Test rejects: rejection rate of the Wald test $H_0! : \psi = 0$ at $\alpha = 0.05$. Under DGPs with nonzero spillovers, both statistics measure the ability to detect interference. Under Config B (no spillover), they measure size (false positive rate). SC weights show low detection rates because partial identification produces wide CIs. 5,000 replications per cell.

Table 53: Detection Power: Config I: Both wrong

Weight	N	Detection rate	Test rejects
DiD	200	0.136	0.140
DiD	500	0.191	0.192
DiD	1000	0.309	0.310
DiD	2000	0.513	0.513
DiD	5000	0.879	0.879
SC sparse	200	0.000	0.140
SC sparse	500	0.000	0.192
SC sparse	1000	0.000	0.310
SC sparse	2000	0.000	0.513
SC sparse	5000	0.000	0.879
SC dense	200	0.000	0.140
SC dense	500	0.000	0.192
SC dense	1000	0.000	0.310
SC dense	2000	0.000	0.513
SC dense	5000	0.000	0.879
Staggered	200	0.171	0.140
Staggered	500	0.278	0.192
Staggered	1000	0.431	0.310
Staggered	2000	0.643	0.513
Staggered	5000	0.928	0.879

Note:
Detection rate: fraction of replications where the Imbens–Manski CI lower bound exceeds zero, indicating statistically detectable interference contamination. Test rejects: rejection rate of the Wald test $H_0! : \psi = 0$ at $\alpha = 0.05$. Under DGPs with nonzero spillovers, both statistics measure the ability to detect interference. Under Config B (no spillover), they measure size (false positive rate). SC weights show low detection rates because partial identification produces wide CIs. 5,000 replications per cell.

Table 54: Detection Power: Config J: Small n_1/n_0

Weight	N	Detection rate	Test rejects
DiD	200	0.026	0.046
DiD	500	0.020	0.040
DiD	1000	0.015	0.037
DiD	2000	0.019	0.037
DiD	5000	0.017	0.033
SC sparse	200	0.000	0.046
SC sparse	500	0.000	0.040
SC sparse	1000	0.000	0.037
SC sparse	2000	0.000	0.037
SC sparse	5000	0.000	0.033
SC dense	200	0.000	0.046
SC dense	500	0.000	0.040
SC dense	1000	0.000	0.037
SC dense	2000	0.000	0.037
SC dense	5000	0.000	0.033
Staggered	200	0.049	0.046
Staggered	500	0.039	0.040
Staggered	1000	0.036	0.037
Staggered	2000	0.037	0.037
Staggered	5000	0.037	0.033

Note:
Detection rate: fraction of replications where the Imbens–Manski CI lower bound exceeds zero, indicating statistically detectable interference contamination. Test rejects: rejection rate of the Wald test $H_0! : \psi = 0$ at $\alpha = 0.05$. Under DGPs with nonzero spillovers, both statistics measure the ability to detect interference. Under Config B (no spillover), they measure size (false positive rate). SC weights show low detection rates because partial identification produces wide CIs. 5,000 replications per cell.

Table 55: Detection Power: Config K: Large n_1/n_0

Weight	N	Detection rate	Test rejects
DiD	200	0.241	0.243
DiD	500	0.446	0.446
DiD	1000	0.673	0.673
DiD	2000	0.915	0.915
DiD	5000	0.999	0.999
SC sparse	200	0.000	0.243
SC sparse	500	0.000	0.446
SC sparse	1000	0.000	0.673
SC sparse	2000	0.000	0.915
SC sparse	5000	0.000	0.999
SC dense	200	0.000	0.243
SC dense	500	0.000	0.446
SC dense	1000	0.000	0.673
SC dense	2000	0.000	0.915
SC dense	5000	0.000	0.999
Staggered	200	0.322	0.243
Staggered	500	0.556	0.446
Staggered	1000	0.771	0.673
Staggered	2000	0.951	0.915
Staggered	5000	0.999	0.999

Note:
Detection rate: fraction of replications where the Imbens–Manski CI lower bound exceeds zero, indicating statistically detectable interference contamination. Test rejects: rejection rate of the Wald test $H_0! : \psi = 0$ at $\alpha = 0.05$. Under DGPs with nonzero spillovers, both statistics measure the ability to detect interference. Under Config B (no spillover), they measure size (false positive rate). SC weights show low detection rates because partial identification produces wide CIs. 5,000 replications per cell.

S8.6 Table S4: Level 3 Tightening via Stratification

Tightening from Level 2 to Level 3 bounds as a function of the number of strata K, for Config F (heterogeneous spillovers). The table reports both the observed tightening and the value predicted by Jensen's inequality.

Table 56: Level 3 Tightening via Stratification (Config F)

K	Weight	N	L_2 width	L_3 width	Actual (%)	Jensen pred (%)	Cov L_3
1	SC sparse	500	1.523	1.523	0.0	0.0	1.000
1	SC sparse	1000	1.503	1.503	0.0	0.0	1.000
1	SC sparse	2000	1.533	1.533	0.0	0.0	1.000
1	Staggered	500	0.454	0.453	-0.4	0.0	1.000
1	Staggered	1000	0.446	0.447	-0.1	0.0	1.000
1	Staggered	2000	0.448	0.448	-0.1	0.0	1.000
2	SC sparse	500	1.526	1.368	13.3	1.1	0.996
2	SC sparse	1000	1.503	1.403	8.8	0.5	0.997
2	SC sparse	2000	1.541	1.496	3.4	0.1	0.998
2	Staggered	500	0.454	0.392	13.0	1.1	1.000
2	Staggered	1000	0.449	0.388	13.5	0.5	1.000
2	Staggered	2000	0.448	0.389	13.1	0.1	1.000
3	SC sparse	500	1.514	1.311	14.5	2.0	1.000
3	SC sparse	1000	1.503	1.324	13.4	1.3	0.999
3	SC sparse	2000	1.571	1.428	11.1	0.8	1.000
3	Staggered	500	0.450	0.390	13.3	2.0	1.000
3	Staggered	1000	0.451	0.394	13.0	1.3	1.000
3	Staggered	2000	0.451	0.396	11.9	0.8	1.000
5	SC sparse	500	1.532	1.275	17.9	2.8	1.000
5	SC sparse	1000	1.531	1.320	14.3	1.6	0.999
5	SC sparse	2000	1.563	1.391	11.7	1.0	0.999
5	Staggered	500	0.446	0.377	15.8	2.8	1.000
5	Staggered	1000	0.450	0.385	15.0	1.6	1.000
5	Staggered	2000	0.452	0.386	14.5	1.0	1.000
10	SC sparse	500	1.521	1.206	21.6	4.0	0.999
10	SC sparse	1000	1.561	1.297	16.3	2.4	0.999
10	SC sparse	2000	1.559	1.361	13.8	1.5	1.000
10	Staggered	500	0.448	0.360	19.5	4.0	1.000
10	Staggered	1000	0.452	0.376	16.6	2.4	1.000
10	Staggered	2000	0.451	0.381	15.5	1.5	1.000

Note:
Config F only (heterogeneous spillovers, $\tau_x = 2.0$). K: number of covariate strata requested. L_2 width, L_3 width: median identified set width at Levels 2 and 3. Actual (

S8.7 Table S5: Nesting Verification (Configs A–K)

Verification that the estimated L_2 bounds are contained within the L_1 bounds in every replication, and that the true contamination falls within the oracle L_2 region. Rates at 1.000 confirm correct implementation of the bound hierarchy.

Table 57: Nesting Verification: Config A: Baseline

Weight	N	$L_2 \subseteq L_1$	$B^{\text{true}} \in L_2^{\text{oracle}}$
DiD	200	1	1
DiD	500	1	1
DiD	1000	1	1
DiD	2000	1	1
DiD	5000	1	1
SC sparse	200	1	1
SC sparse	500	1	1
SC sparse	1000	1	1
SC sparse	2000	1	1
SC sparse	5000	1	1
SC dense	200	1	1
SC dense	500	1	1
SC dense	1000	1	1
SC dense	2000	1	1
SC dense	5000	1	1
Staggered	200	1	1
Staggered	500	1	1
Staggered	1000	1	1
Staggered	2000	1	1
Staggered	5000	1	1

Note:
$L_2 \subseteq L_1$: fraction of replications where the L_2 (sorting) identified set is contained within the L_1 (support) identified set, validating Theorem 4.4(i). $B^{\text{true}} \in L_2^{\text{oracle}}$: fraction where the true bias falls inside the oracle L_2 bounds (known nuisance parameters). Rates near 1.000 confirm correct implementation of the bound hierarchy. 5,000 replications per cell.

Table 58: Nesting Verification: Config B: No spillover

Weight	N	$L_2 \subseteq L_1$	$B^{\text{true}} \in L_2^{\text{oracle}}$
DiD	500	1	1
DiD	2000	1	1
SC sparse	500	1	1
SC sparse	2000	1	1
SC dense	500	1	1
SC dense	2000	1	1
Staggered	500	1	1
Staggered	2000	1	1

Note:
$L_2 \subseteq L_1$: fraction of replications where the L_2 (sorting) identified set is contained within the L_1 (support) identified set, validating Theorem 4.4(i). $B^{\text{true}} \in L_2^{\text{oracle}}$: fraction where the true bias falls inside the oracle L_2 bounds (known nuisance parameters). Rates near 1.000 confirm correct implementation of the bound hierarchy. 5,000 replications per cell.

Table 59: Nesting Verification: Config C: Uniform spillover

Weight	N	$L_2 \subseteq L_1$	$B^{\text{true}} \in L_2^{\text{oracle}}$
DiD	500	1	1
DiD	2000	1	1
SC sparse	500	1	1
SC sparse	2000	1	1
SC dense	500	1	1
SC dense	2000	1	1
Staggered	500	1	1
Staggered	2000	1	1

Note:
$L_2 \subseteq L_1$: fraction of replications where the L_2 (sorting) identified set is contained within the L_1 (support) identified set, validating Theorem 4.4(i). $B^{\text{true}} \in L_2^{\text{oracle}}$: fraction where the true bias falls inside the oracle L_2 bounds (known nuisance parameters). Rates near 1.000 confirm correct implementation of the bound hierarchy. 5,000 replications per cell.

Table 60: Nesting Verification: Config D: Strong spillover

Weight	N	$L_2 \subseteq L_1$	$B^{\text{true}} \in L_2^{\text{oracle}}$
DiD	200	1	1
DiD	500	1	1
DiD	1000	1	1
DiD	2000	1	1
DiD	5000	1	1
SC sparse	200	1	1
SC sparse	500	1	1
SC sparse	1000	1	1
SC sparse	2000	1	1
SC sparse	5000	1	1
SC dense	200	1	1
SC dense	500	1	1
SC dense	1000	1	1
SC dense	2000	1	1
SC dense	5000	1	1
Staggered	200	1	1
Staggered	500	1	1
Staggered	1000	1	1
Staggered	2000	1	1
Staggered	5000	1	1

Note:
$L_2 \subseteq L_1$: fraction of replications where the L_2 (sorting) identified set is contained within the L_1 (support) identified set, validating Theorem 4.4(i). $B^{\text{true}} \in L_2^{\text{oracle}}$: fraction where the true bias falls inside the oracle L_2 bounds (known nuisance parameters). Rates near 1.000 confirm correct implementation of the bound hierarchy. 5,000 replications per cell.

Table 61: Nesting Verification: Config E: Spatial decay

Weight	N	$L_2 \subseteq L_1$	$B^{\text{true}} \in L_2^{\text{oracle}}$
DiD	200	1	1
DiD	500	1	1
DiD	1000	1	1
DiD	2000	1	1
DiD	5000	1	1
SC sparse	200	1	1
SC sparse	500	1	1
SC sparse	1000	1	1
SC sparse	2000	1	1
SC sparse	5000	1	1
SC dense	200	1	1
SC dense	500	1	1
SC dense	1000	1	1
SC dense	2000	1	1
SC dense	5000	1	1
Staggered	200	1	1
Staggered	500	1	1
Staggered	1000	1	1
Staggered	2000	1	1
Staggered	5000	1	1

Note:
$L_2 \subseteq L_1$: fraction of replications where the L_2 (sorting) identified set is contained within the L_1 (support) identified set, validating Theorem 4.4(i). $B^{\text{true}} \in L_2^{\text{oracle}}$: fraction where the true bias falls inside the oracle L_2 bounds (known nuisance parameters). Rates near 1.000 confirm correct implementation of the bound hierarchy. 5,000 replications per cell.

Table 62: Nesting Verification: Config F: Heterogeneous

Weight	N	$L_2 \subseteq L_1$	$B^{\text{true}} \in L_2^{\text{oracle}}$
DiD	200	1	1
DiD	500	1	1
DiD	1000	1	1
DiD	2000	1	1
DiD	5000	1	1
SC sparse	200	1	1
SC sparse	500	1	1
SC sparse	1000	1	1
SC sparse	2000	1	1
SC sparse	5000	1	1
SC dense	200	1	1
SC dense	500	1	1
SC dense	1000	1	1
SC dense	2000	1	1
SC dense	5000	1	1
Staggered	200	1	1
Staggered	500	1	1
Staggered	1000	1	1
Staggered	2000	1	1
Staggered	5000	1	1

Note:
$L_2 \subseteq L_1$: fraction of replications where the L_2 (sorting) identified set is contained within the L_1 (support) identified set, validating Theorem 4.4(i). $B^{\text{true}} \in L_2^{\text{oracle}}$: fraction where the true bias falls inside the oracle L_2 bounds (known nuisance parameters). Rates near 1.000 confirm correct implementation of the bound hierarchy. 5,000 replications per cell.

Table 63: Nesting Verification: Config G: Wrong μ_0

Weight	N	$L_2 \subseteq L_1$	$B^{\text{true}} \in L_2^{\text{oracle}}$
DiD	200	1	1
DiD	500	1	1
DiD	1000	1	1
DiD	2000	1	1
DiD	5000	1	1
SC sparse	200	1	1
SC sparse	500	1	1
SC sparse	1000	1	1
SC sparse	2000	1	1
SC sparse	5000	1	1
SC dense	200	1	1
SC dense	500	1	1
SC dense	1000	1	1
SC dense	2000	1	1
SC dense	5000	1	1
Staggered	200	1	1
Staggered	500	1	1
Staggered	1000	1	1
Staggered	2000	1	1
Staggered	5000	1	1

Note:
$L_2 \subseteq L_1$: fraction of replications where the L_2 (sorting) identified set is contained within the L_1 (support) identified set, validating Theorem 4.4(i). $B^{\text{true}} \in L_2^{\text{oracle}}$: fraction where the true bias falls inside the oracle L_2 bounds (known nuisance parameters). Rates near 1.000 confirm correct implementation of the bound hierarchy. 5,000 replications per cell.

Table 64: Nesting Verification: Config H: Wrong π

Weight	N	$L_2 \subseteq L_1$	$B^{\text{true}} \in L_2^{\text{oracle}}$
DiD	200	1	1
DiD	500	1	1
DiD	1000	1	1
DiD	2000	1	1
DiD	5000	1	1
SC sparse	200	1	1
SC sparse	500	1	1
SC sparse	1000	1	1
SC sparse	2000	1	1
SC sparse	5000	1	1
SC dense	200	1	1
SC dense	500	1	1
SC dense	1000	1	1
SC dense	2000	1	1
SC dense	5000	1	1
Staggered	200	1	1
Staggered	500	1	1
Staggered	1000	1	1
Staggered	2000	1	1
Staggered	5000	1	1

Note:
$L_2 \subseteq L_1$: fraction of replications where the L_2 (sorting) identified set is contained within the L_1 (support) identified set, validating Theorem 4.4(i). $B^{\text{true}} \in L_2^{\text{oracle}}$: fraction where the true bias falls inside the oracle L_2 bounds (known nuisance parameters). Rates near 1.000 confirm correct implementation of the bound hierarchy. 5,000 replications per cell.

Table 65: Nesting Verification: Config I: Both wrong

Weight	N	$L_2 \subseteq L_1$	$B^{\text{true}} \in L_2^{\text{oracle}}$
DiD	200	1.000	1.000
DiD	500	1.000	1.000
DiD	1000	1.000	1.000
DiD	2000	1.000	1.000
DiD	5000	1.000	1.000
SC sparse	200	0.989	0.960
SC sparse	500	0.996	0.984
SC sparse	1000	0.999	0.983
SC sparse	2000	1.000	0.981
SC sparse	5000	1.000	0.981
SC dense	200	1.000	0.998
SC dense	500	1.000	1.000
SC dense	1000	1.000	1.000
SC dense	2000	1.000	1.000
SC dense	5000	1.000	1.000
Staggered	200	1.000	0.998
Staggered	500	1.000	1.000
Staggered	1000	1.000	1.000
Staggered	2000	1.000	1.000
Staggered	5000	1.000	1.000

Note:
$L_2 \subseteq L_1$: fraction of replications where the L_2 (sorting) identified set is contained within the L_1 (support) identified set, validating Theorem 4.4(i). $B^{\text{true}} \in L_2^{\text{oracle}}$: fraction where the true bias falls inside the oracle L_2 bounds (known nuisance parameters). Rates near 1.000 confirm correct implementation of the bound hierarchy. 5,000 replications per cell.

Table 66: Nesting Verification: Config J: Small n_1/n_0

Weight	N	$L_2 \subseteq L_1$	$B^{\text{true}} \in L_2^{\text{oracle}}$
DiD	200	1	1
DiD	500	1	1
DiD	1000	1	1
DiD	2000	1	1
DiD	5000	1	1
SC sparse	200	1	1
SC sparse	500	1	1
SC sparse	1000	1	1
SC sparse	2000	1	1
SC sparse	5000	1	1
SC dense	200	1	1
SC dense	500	1	1
SC dense	1000	1	1
SC dense	2000	1	1
SC dense	5000	1	1
Staggered	200	1	1
Staggered	500	1	1
Staggered	1000	1	1
Staggered	2000	1	1
Staggered	5000	1	1

Note:
$L_2 \subseteq L_1$: fraction of replications where the L_2 (sorting) identified set is contained within the L_1 (support) identified set, validating Theorem 4.4(i). $B^{\text{true}} \in L_2^{\text{oracle}}$: fraction where the true bias falls inside the oracle L_2 bounds (known nuisance parameters). Rates near 1.000 confirm correct implementation of the bound hierarchy. 5,000 replications per cell.

Table 67: Nesting Verification: Config K: Large n_1/n_0

Weight	N	$L_2 \subseteq L_1$	$B^{\text{true}} \in L_2^{\text{oracle}}$
DiD	200	1.000	1.000
DiD	500	1.000	1.000
DiD	1000	1.000	1.000
DiD	2000	1.000	1.000
DiD	5000	1.000	1.000
SC sparse	200	0.997	0.989
SC sparse	500	0.999	0.994
SC sparse	1000	1.000	0.994
SC sparse	2000	1.000	0.995
SC sparse	5000	1.000	0.996
SC dense	200	1.000	1.000
SC dense	500	1.000	1.000
SC dense	1000	1.000	1.000
SC dense	2000	1.000	1.000
SC dense	5000	1.000	1.000
Staggered	200	1.000	1.000
Staggered	500	1.000	1.000
Staggered	1000	1.000	1.000
Staggered	2000	1.000	1.000
Staggered	5000	1.000	1.000

Note:
$L_2 \subseteq L_1$: fraction of replications where the L_2 (sorting) identified set is contained within the L_1 (support) identified set, validating Theorem 4.4(i). $B^{\text{true}} \in L_2^{\text{oracle}}$: fraction where the true bias falls inside the oracle L_2 bounds (known nuisance parameters). Rates near 1.000 confirm correct implementation of the bound hierarchy. 5,000 replications per cell.

S8.8 Table S6: Computation Times

Wall-clock computation time per replication for the full pipeline (DGP generation, nuisance estimation, three-level bounds, and inference) across sample sizes.

Table 68: Computational Timing by Sample Size

N	Cells	Mean (s)	Median (s)	Max (s)	Total (h)
200	36	0.01	0.01	0.01	0.5
500	44	0.02	0.02	0.03	1.1
1000	36	0.05	0.05	0.07	2.5
2000	44	0.18	0.15	0.27	11.0
5000	36	1.39	1.26	1.84	69.3

Note:
Mean, Median, Max: per-replication wall-clock time in seconds, averaged across cells at each N. Total (h): cumulative computation time in hours across all DGP$\times$weight cells at each N (mean per-rep time $\times$ number of replications, summed across cells). Times reflect the full pipeline: DGP generation, nuisance estimation, three-level bounds, and inference.

S8.9 Table S7: Propensity Score Clipping Rates

Fraction of replications where $\hat{\psi}_{\mathrm{DR}}$ required clipping to the feasibility range $[0, \bar{d}]$, and the mean magnitude of clipping when it occurs.

Table 69: Clipping Rates for $\hat{\psi}$

DGP	N	Clip rate	Mean clip	DGP	N	Clip rate	Mean clip
A	200	0.002	3e-04	G	200	0.004	5e-04
A	500	0	0	G	500	0	0
A	1000	0	0	G	1000	0	0
A	2000	0	0	G	2000	0	0
A	5000	0	0	G	5000	0	0
B	500	0.047	0.0113	H	200	0.005	0.001
B	2000	0.001	1e-04	H	500	0	1e-04
C	500	0	0	H	1000	0	0
C	2000	0	0	H	2000	0	0
D	200	0.156	0.0222	H	5000	0	0
D	500	0.053	0.0042	I	200	0.229	0.1098
D	1000	0.014	8e-04	I	500	0.132	0.0336
D	2000	0.002	1e-04	I	1000	0.066	0.0119
D	5000	0	0	I	2000	0.022	0.0023
E	200	0	0	I	5000	0.001	1e-04
E	500	0	0	J	200	0.498	0.0946
E	1000	0	0	J	500	0.507	0.0617
E	2000	0	0	J	1000	0.514	0.0468
E	5000	0	0	J	2000	0.494	0.0322
F	200	0	1e-04	J	5000	0.493	0.0201
F	500	0	0	K	200	0.137	0.051
F	1000	0	0	K	500	0.045	0.0088
F	2000	0	0	K	1000	0.009	0.0011
F	5000	0	0	K	2000	0	0
				K	5000	0	0

Note:
Clip rate: fraction of replications where $\hat{\psi}$ was clipped to $[0, \bar{d}]$. Mean clip: mean magnitude of clipping when it occurs ($|\hat{\psi} - \hat{\psi}_{\text{clipped}}|$). Clipping enforces the feasibility constraint $\psi \in [0, \bar{d}]$. High clipping rates (Configs B, J) indicate the estimator frequently falls outside the feasible range, which can occur when the true spillover is near the boundary. Clipping does not depend on the weight type, so DiD is shown as representative.

S8.10 Table S8: Sensitivity to Support Bound a

Bound widths, CI lengths, and coverage across a grid of support lower-bound values $a \in \{-3, -2, -1, -0.5, 0\}$ for Config A. The table documents the bias-coverage tradeoff: a conservative a widens the bounds but maintains coverage, while an aggressive a tightens them but risks support violations.

Table 70: Sensitivity to Support Bound a (Config A)

a	Weight	N	L_2 width	CI length	Cov L_2	$\bar{n}_{d<0}$	Bias($\hat{\psi}$)
-3.0	SC sparse	1000	2.334	2.334	1	0.0	0.0014
-3.0	SC sparse	2000	2.341	2.341	1	0.0	-0.0011
-2.0	SC sparse	1000	1.902	1.902	1	0.0	0.0050
-2.0	SC sparse	2000	1.881	1.881	1	0.0	-0.0002
-1.0	SC sparse	1000	1.456	1.456	1	0.5	-0.0026
-1.0	SC sparse	2000	1.468	1.468	1	1.0	-0.0003
-0.5	SC sparse	1000	1.225	1.225	1	1.6	0.0006
-0.5	SC sparse	2000	1.238	1.238	1	3.3	-0.0004
0.0	SC sparse	1000	0.999	0.999	1	4.8	-0.0007
0.0	SC sparse	2000	1.046	1.046	1	9.7	-0.0020
-3.0	Staggered	1000	0.603	0.733	1	0.0	0.0014
-3.0	Staggered	2000	0.602	0.693	1	0.0	-0.0011
-2.0	Staggered	1000	0.495	0.620	1	0.0	0.0050
-2.0	Staggered	2000	0.494	0.581	1	0.0	-0.0002
-1.0	Staggered	1000	0.376	0.499	1	0.5	-0.0026
-1.0	Staggered	2000	0.378	0.463	1	1.0	-0.0003
-0.5	Staggered	1000	0.319	0.440	1	1.6	0.0006
-0.5	Staggered	2000	0.319	0.404	1	3.3	-0.0004
0.0	Staggered	1000	0.256	0.381	1	4.8	-0.0007
0.0	Staggered	2000	0.259	0.345	1	9.7	-0.0020

Note:
a: analyst-chosen lower bound on the support of potential outcomes (true value $a = -1$). L_2 width: median identified set width at Level 2. CI length: median Imbens–Manski CI length. Cov L_2: empirical coverage of the 95

S8.11 Table S9: Cross-Fitting vs. Standard DR Estimator

Comparison of the standard DR estimator with its cross-fitted variant under Config A. Cross-fitting removes the Donsker condition but inflates variance when the nuisance models are parametric.

Table 71: Cross-Fitting vs. Standard DR Estimator (Config A)

N	B	Bias(std)	Bias(CF)	Δ bias	SD(std)	SD(CF)	RMSE(std)	RMSE(CF)
500	1000	-0.0020	-0.0012	-0.0007	0.1778	0.4149	0.0020	0.0012
1000	1000	-0.0048	-0.0070	0.0022	0.1334	0.3164	0.0048	0.0070
2000	1000	-0.0052	-0.0123	0.0071	0.0937	0.2329	0.0052	0.0123

Note:
Comparison of the standard DR estimator ($\hat{\psi}$) with its cross-fitted variant ($\hat{\psi}_{CF}$). Cross-fitting (sample splitting for nuisance estimation) eliminates the Donsker condition required for $\sqrt{N}$-consistency, at the cost of increased variance from sample splitting. Bias(std), Bias(CF): mean bias of standard and cross-fitted estimators. Δ bias: difference in bias (std $-$ CF); values near zero indicate that cross-fitting provides no bias reduction for this DGP. SD: Monte Carlo standard deviation. RMSE: root mean squared error. Similar performance across both estimators justifies using the simpler (non-cross-fitted) estimator when the Donsker condition is plausible. Config A, 5,000 replications per N.

S8.12 Table S10: Shapiro–Wilk Normality Tests

Shapiro–Wilk W statistics for the standardized DR estimates and bound endpoints, supporting the asymptotic normality result.

Table 72: Shapiro–Wilk Normality Tests (Config A, DiD, $N \in \{500, 1000, 2000\}$)

	N	B	W(psi)	W(UB)	W(LB)
W	500	5000	0.9997	0.9997	0.9997
W1	1000	5000	0.9992	0.9992	0.9992
W2	2000	5000	0.9997	0.9997	0.9997

Note:
W: Shapiro–Wilk test statistic for normality. Values close to 1.0 indicate close approximation to the Gaussian distribution, supporting the asymptotic normality claimed in Theorem 5.2. $W(\psi)$: applied to standardized $\hat{\psi}$. W(UB), W(LB): applied to standardized L_2 upper and lower bound plugin estimators. Formal p-values are omitted because the Shapiro–Wilk test at $n = 5{,}000$ has sufficient power to reject exact normality for any finite-sample distribution; the W statistic itself is the informative quantity. Config A, DiD weights.

S8.13 Table S11: Hölder Bound Validation (Configs A, D, E, F)

Verification that the estimated L_2 width falls below the theoretical Hölder upper bound in each replication. Valid rates near 1.000 confirm the theoretical guarantee. These four configurations span the range of weight dispersion and spillover heterogeneity, the two dimensions most directly relevant to width validation.

Table 73: Hölder Bound Validation: Config A: Baseline

Weight	N	L_2 width (med)	Hölder bound (med)	Valid rate
DiD	500	0.000	0.000	1.000
DiD	1000	0.000	0.000	1.000
DiD	2000	0.000	0.000	1.000
DiD	5000	0.000	0.000	1.000
SC sparse	500	1.441	21.542	1.000
SC sparse	1000	1.462	44.251	1.000
SC sparse	2000	1.447	87.331	1.000
SC sparse	5000	1.446	222.380	1.000
Staggered	500	0.375	0.678	0.999
Staggered	1000	0.377	0.678	1.000
Staggered	2000	0.378	0.675	1.000
Staggered	5000	0.378	0.677	1.000

Note:
Validates Theorem 4.4(iii): the width of the L_2 identified set is bounded above by the Hölder bound $2n_1 \max|\tilde{w}_j| \min(\psi, \bar{d}-\psi)$. This is an upper bound on the *width* $B_U - B_L$, not on the endpoints themselves. L_2 width (med): median width of the estimated L_2 identified set. Hölder bound (med): median of the theoretical upper bound on the width. Valid rate: fraction of replications where L_2 width $\leq$ Hölder bound. Rates near 1.000 confirm the theoretical guarantee. Deviations below 1.000 can occur due to finite-sample estimation noise in the plugin bound estimators. 5,000 replications per cell.

Table 74: Hölder Bound Validation: Config D: Strong spillover

Weight	N	L_2 width (med)	Hölder bound (med)	Valid rate
DiD	500	0.000	0.000	1.000
DiD	1000	0.000	0.000	1.000
DiD	2000	0.000	0.000	1.000
DiD	5000	0.000	0.000	1.000
SC sparse	500	1.141	6.522	1.000
SC sparse	1000	1.214	13.212	1.000
SC sparse	2000	1.270	26.870	1.000
SC sparse	5000	1.260	67.590	1.000
Staggered	500	0.137	0.203	0.799
Staggered	1000	0.133	0.203	0.880
Staggered	2000	0.134	0.205	0.950
Staggered	5000	0.136	0.205	0.994

Note:
Validates Theorem 4.4(iii): the width of the L_2 identified set is bounded above by the Hölder bound $2n_1 \max|\tilde{w}_j| \min(\psi, \bar{d}-\psi)$. This is an upper bound on the *width* $B_U - B_L$, not on the endpoints themselves. L_2 width (med): median width of the estimated L_2 identified set. Hölder bound (med): median of the theoretical upper bound on the width. Valid rate: fraction of replications where L_2 width $\leq$ Hölder bound. Rates near 1.000 confirm the theoretical guarantee. Deviations below 1.000 can occur due to finite-sample estimation noise in the plugin bound estimators. 5,000 replications per cell.

Table 75: Hölder Bound Validation: Config E: Spatial decay

Weight	N	L_2 width (med)	Hölder bound (med)	Valid rate
DiD	500	0.000	0.000	1
DiD	1000	0.000	0.000	1
DiD	2000	0.000	0.000	1
DiD	5000	0.000	0.000	1
SC sparse	500	1.733	41.899	1
SC sparse	1000	1.727	86.213	1
SC sparse	2000	1.744	172.949	1
SC sparse	5000	1.743	435.416	1
Staggered	500	0.575	1.333	1
Staggered	1000	0.577	1.334	1
Staggered	2000	0.573	1.323	1
Staggered	5000	0.574	1.329	1

Note:
Validates Theorem 4.4(iii): the width of the L_2 identified set is bounded above by the Hölder bound $2n_1 \max|\tilde{w}_j| \min(\psi, \bar{d} - \psi)$. This is an upper bound on the *width* $B_U - B_L$, not on the endpoints themselves. L_2 width (med): median width of the estimated L_2 identified set. Hölder bound (med): median of the theoretical upper bound on the width. Valid rate: fraction of replications where L_2 width $\leq$ Hölder bound. Rates near 1.000 confirm the theoretical guarantee. Deviations below 1.000 can occur due to finite-sample estimation noise in the plugin bound estimators. 5,000 replications per cell.

Table 76: Hölder Bound Validation: Config F: Heterogeneous

Weight	N	L_2 width (med)	Hölder bound (med)	Valid rate
DiD	500	0.000	0.000	1
DiD	1000	0.000	0.000	1
DiD	2000	0.000	0.000	1
DiD	5000	0.000	0.000	1
SC sparse	500	1.528	29.251	1
SC sparse	1000	1.538	59.229	1
SC sparse	2000	1.511	117.338	1
SC sparse	5000	1.540	299.672	1
Staggered	500	0.450	0.910	1
Staggered	1000	0.449	0.912	1
Staggered	2000	0.451	0.915	1
Staggered	5000	0.449	0.915	1

Note:
Validates Theorem 4.4(iii): the width of the L_2 identified set is bounded above by the Hölder bound $2n_1 \max|\tilde{w}_j| \min(\psi, \bar{d} - \psi)$. This is an upper bound on the *width* $B_U - B_L$, not on the endpoints themselves. L_2 width (med): median width of the estimated L_2 identified set. Hölder bound (med): median of the theoretical upper bound on the width. Valid rate: fraction of replications where L_2 width $\leq$ Hölder bound. Rates near 1.000 confirm the theoretical guarantee. Deviations below 1.000 can occur due to finite-sample estimation noise in the plugin bound estimators. 5,000 replications per cell.

S8.14 Table S12: Sensitivity to Exposure Radius r (Configs A, E)

Bound widths, CI lengths, and coverage across a grid of exposure radii $r \in [0.5, 3.0]$ for Configs A and E. The table documents the tension between identification-region width, which is stable, and estimation precision, which deteriorates as r grows and unexposed controls vanish.

Table 77: Sensitivity to Exposure Radius r: Config A: Baseline

Weight	r	N	L_2 width	CI length	Cov L_2	Bias($\hat{\psi}$)
SC sparse	0.50	1000	1.475	1.475	1	0.0002
SC sparse	0.75	1000	1.454	1.454	1	-0.0019
SC sparse	1.00	1000	1.452	1.452	1	-0.0014
SC sparse	1.25	1000	1.462	1.462	1	-0.0018
SC sparse	1.50	1000	1.461	1.461	1	-0.0072
SC sparse	1.75	1000	1.456	1.456	1	0.0015
SC sparse	2.00	1000	1.452	1.462	1	0.0018
SC sparse	2.25	1000	1.410	1.478	1	-0.0014
SC sparse	2.50	1000	1.292	1.635	1	0.0305
SC sparse	2.75	1000	1.051	1.997	1	-0.0470
SC sparse	3.00	1000	0.000	114.077	1	-0.0319
Staggered	0.50	1000	0.378	0.462	1	0.0002
Staggered	0.75	1000	0.376	0.462	1	-0.0019
Staggered	1.00	1000	0.379	0.470	1	-0.0014
Staggered	1.25	1000	0.380	0.480	1	-0.0018
Staggered	1.50	1000	0.378	0.500	1	-0.0072
Staggered	1.75	1000	0.377	0.543	1	0.0015
Staggered	2.00	1000	0.376	0.640	1	0.0018
Staggered	2.25	1000	0.373	0.886	1	-0.0014
Staggered	2.50	1000	0.358	1.510	1	0.0305
Staggered	2.75	1000	0.269	3.108	1	-0.0470
Staggered	3.00	1000	0.000	6.157	1	-0.0319

Note:
r: exposure radius defining the spatial neighborhood. Units within distance r of a treated unit are classified as exposed controls. Larger r increases the exposed set, potentially changing bound widths and coverage. L_2 width: median identified set width at Level 2. CI length: median Imbens–Manski CI length. Cov L_2: empirical coverage. Bias($\hat{\psi}$): mean DR estimator bias. DiD weights are omitted: point identification holds regardless of r (see Table S4 note). Configs A and E, $N = 1000$, 2,000 replications.

Table 78: Sensitivity to Exposure Radius r: Config E: Spatial decay

Weight	r	N	L_2 width	CI length	Cov L_2	Bias($\hat{\psi}$)
SC sparse	0.50	1000	1.745	1.745	1	-0.0050
SC sparse	0.75	1000	1.737	1.737	1	-0.0018
SC sparse	1.00	1000	1.737	1.737	1	-0.0033
SC sparse	1.25	1000	1.725	1.725	1	-0.0034
SC sparse	1.50	1000	1.725	1.725	1	0.0005
SC sparse	1.75	1000	1.745	1.745	1	-0.0045
SC sparse	2.00	1000	1.741	1.741	1	-0.0065
SC sparse	2.25	1000	1.739	1.750	1	-0.0193
SC sparse	2.50	1000	1.694	1.821	1	0.0094
SC sparse	2.75	1000	1.465	2.038	1	-0.0027
SC sparse	3.00	1000	1.068	2.686	1	-0.0467
Staggered	0.50	1000	0.574	0.683	1	-0.0050
Staggered	0.75	1000	0.576	0.687	1	-0.0018
Staggered	1.00	1000	0.577	0.690	1	-0.0033
Staggered	1.25	1000	0.576	0.698	1	-0.0034
Staggered	1.50	1000	0.577	0.715	1	0.0005
Staggered	1.75	1000	0.574	0.748	1	-0.0045
Staggered	2.00	1000	0.573	0.825	1	-0.0065
Staggered	2.25	1000	0.565	0.993	1	-0.0193
Staggered	2.50	1000	0.545	1.455	1	0.0094
Staggered	2.75	1000	0.463	2.917	1	-0.0027
Staggered	3.00	1000	0.212	5.691	1	-0.0467

Note:
r: exposure radius defining the spatial neighborhood. Units within distance r of a treated unit are classified as exposed controls. Larger r increases the exposed set, potentially changing bound widths and coverage. L_2 width: median identified set width at Level 2. CI length: median Imbens–Manski CI length. Cov L_2: empirical coverage. Bias($\hat{\psi}$): mean DR estimator bias. DiD weights are omitted: point identification holds regardless of r (see Table S4 note). Configs A and E, $N = 1000$, 2,000 replications.

S8.15 Table S13: Contamination Magnitude Relative to ATT

Ratio of the true contamination to the ATT, documenting the practical severity of interference. The fraction of replications where contamination exceeds 10% and 25% of the ATT gives practical guidance on when interference is substantively consequential.

Table 79: Substantive Contamination: B/τ Relative to ATT ($N = 2000$)

DGP	Weight	Median B/τ	$B/\tau > 0.10$	$B/\tau > 0.25$
A	DiD	0.254	1.000	0.677
A	SC sparse	0.273	0.988	0.585
A	Staggered	0.254	1.000	0.626
D	DiD	0.077	0.000	0.000
D	SC sparse	0.067	0.341	0.052
D	Staggered	0.076	0.003	0.000
E	DiD	0.498	1.000	1.000
E	SC sparse	0.561	1.000	0.977
E	Staggered	0.497	1.000	1.000
F	DiD	0.342	1.000	1.000
F	SC sparse	0.348	0.913	0.662
F	Staggered	0.342	1.000	1.000

Note:
Median B/τ: median ratio of true contamination bias to the ATT across replications. Values above 0.10 indicate contamination exceeding 10

S8.16 Application Detection-and-Bounds Scorecards

The two scorecards below collect, for each empirical study, the exposure specification, the doubly robust detection estimate, and the full nested-region ladder that the main text reports only in summary. Table 80 covers the point-identified difference-in-differences case (police pacification) and Table 81 the partially identified synthetic-control case (a municipal basic income), carried through two outcomes at two exposure radii.

Table 80: Police Pacification (UPP): Detection and Bounds Scorecard

Quantity	Value
Estimator	Staggered-adoption difference-in-differences
Panel	156 bairros (55 treated, 101 control)
Naive ATT $\hat{\tau}(w)$	+0.0397 (Callaway–Sant'Anna / Sun–Abraham: +0.0377)
Exposure map	First-order spatial contiguity to a treated bairro
Exposed / unexposed controls (n_1, n_0)	40 / 60
Exposure fraction ρ	0.40
Detection $\hat{\psi}$ (SE)	+0.0133 (0.0050), $t = 2.66$
Contamination $\mathcal{B}(w) = \rho\hat{\psi}$	+0.0053
Imbens–Manski CI for $\mathcal{B}(w)$	$[+0.0014,\ +0.0092]$
Identification regime	Point (symmetric weights)
Manski baseline	$[-0.0295,\ +0.4269]$
Level 1 region $\mathcal{B}_1(w)$	$[0,\ +0.4269]$
Level 2 region $\mathcal{B}_2(w)$	$\{+0.0053\}$
De-contaminated target $\tau^{\text{SUTVA}}(w)$	+0.0450 (simplex), +0.0434 (estimator-faithful)
Spillover share of naive effect	$\approx 13\%$

Note: Uniform difference-in-differences weights place the study in the point-identified regime (symmetric weights, Proposition 4.6 of the main text), so the mean constraint collapses the Level 2 region to the single point $\rho\hat{\psi}$. The contamination is positive, so the naive estimator is biased toward zero and de-contaminating raises the effect. The simplex headline +0.0450 projects the exposed-control weight in full onto the differenced outcome. The estimator-faithful headline +0.0434 uses the control mass 0.69 that the Callaway–Sant'Anna estimator places on the differenced path, giving $0.69\,\mathcal{B}_2(w) = 0.0037$. Sign, order of magnitude, and substantive reading hold under either convention.

S9 Exposure-Mapping Robustness

The framework detects and bounds the contamination of a linear ATT estimator under interference, and it reads the exposure mapping in a single place: the control observation equation of Section 2, whose consequence is that an unexposed control directly reveals its clean counterfactual $\Delta Y_j(0,0)$. That single fact identifies the spillover mean ψ, and ψ is the constraint that turns the support-only region into the informative Level 2 region. The decomposition, the sorting linear program, and the estimator never touch the mapping. In the applications the mapping is radial, $E_j = \mathbb{1}\{\min_p d_{jp} < r\}$ for a reach radius r, and r is an analyst choice: the true contamination process is not known, so the exposed/clean partition is observed only *given* a mapping, not in an absolute sense. This section isolates

Table 81: Maricá Basic Income: Detection and Bounds Scorecard by Outcome and Exposure Radius

Quantity	Incumbent-party share	Vote concentration
Estimator	Synthetic control	Synthetic control
Naive ATT $\hat{\tau}(w)$	+0.59	+0.268
Exposure radius	$\approx$ 75 km	100 km
Exposed / unexposed donors (n_1, n_0)	21 / 65	40 / 46
Exposure fraction ρ	0.24	0.47
Detection $\hat{\psi}$ (SE)	+0.089 (0.042), $t = 2.12$	+0.059 (0.020), $t = 3.0$
Imbens–Manski detection interval	[0, +0.568]	[0, +0.440]
Identification regime	Partial (asymmetric weights)	Partial (asymmetric weights)
Manski baseline	[0, +0.733]	[0, +0.615]
Level 1 region $\mathcal{B}_1(w)$	[0, +0.458]	[0, +0.440]
Level 2 region $\mathcal{B}_2(w)$	[0, +0.387]	[0, +0.440]
De-contaminated target $\tau^{\text{SUTVA}}(w)$	[+0.59, +0.98]	[+0.268, +0.708]

Note: Concentrated synthetic-control weights place both outcomes in the partially identified regime (asymmetric weights, Proposition 4.7 of the main text), so the Level 2 region stays a nondegenerate interval rather than the singleton of the difference-in-differences case. For incumbent-party share the mean constraint tightens the region from Level 1 to Level 2, while for vote concentration it adds nothing to the monotonicity restriction, so $\mathcal{B}_1(w) = \mathcal{B}_2(w)$. Both corrections are signed upward, so the true effect is at least the naive estimate. A replication for the broader left bloc reproduces the vote-concentration pattern at $t = 4.2$.

the residual dependence on that choice into one weakened condition (No-Leakage), names the safe direction of error in a proposition (Directional-Validity), supplies a data-side test that certifies the condition (the $\mathcal{C}_0$ placebo), and gives a selection rule that turns a pool of candidate mappings into one reported answer. The core of the paper is unchanged. This section adds an honest account of the mapping's role and a procedure that is robust to getting the mapping wrong in the common, benign direction while detecting and refusing the dangerous one.

S9.1 The exposure mapping as a cut on a reach axis

Order the control units along a *reach axis* s_j, the distance to the nearest treated unit (or a graph distance, or a directional coordinate, all three of which I use in Section S9.3). The true spillover δ_j is an unknown, weakly decreasing function $\delta(s_j)$: units near treatment are contaminated ($\delta_j > 0$), far units are clean ($\delta_j = 0$), and there is a true boundary $s = r^*$ beyond which $\delta = 0$. Write $\mathcal{C}_1^* = \{j : \delta_j > 0\}$ for the true exposed set and $\mathcal{C}_0^* = \{j : \delta_j = 0\}$ for the true clean set. The contamination target $\mathcal{B}(w) = \sum_{j \in \mathcal{C}_1^*} w_j \delta_j$ is fixed by the truth. It does not move when the analyst changes the mapping.

A radial mapping m with radius r declares $\{s_j < r\}$ exposed, giving $\mathcal{C}_1(m)$, and $\{s_j \geq r\}$ clean, giving $\mathcal{C}_0(m)$. That vertical cut is the mapping's complete content. Its one job is to nominate a set $\mathcal{C}_0(m)$ it claims is clean and can therefore serve as the reference that identifies $\psi(m)$. Figure 6 shows the object and the two ways a cut can miss the true boundary r^*.

When $r > r^*$ the cut sweeps genuinely-clean units ($s \in [r^*, r]$) into $\mathcal{C}_1(m)$. These carry $\delta_j = 0$, contribute nothing to $\mathcal{B}(w)$, and leave $\mathcal{C}_0(m)$ entirely clean (over-coverage, the green band). When $r < r^*$ the cut leaves truly-exposed units ($s \in [r, r^*]$) inside the set called clean, so $\mathcal{C}_0(m)$ is contaminated (under-coverage, the red band). The whole robustness argument rests on the fact that these two directions are not symmetric.

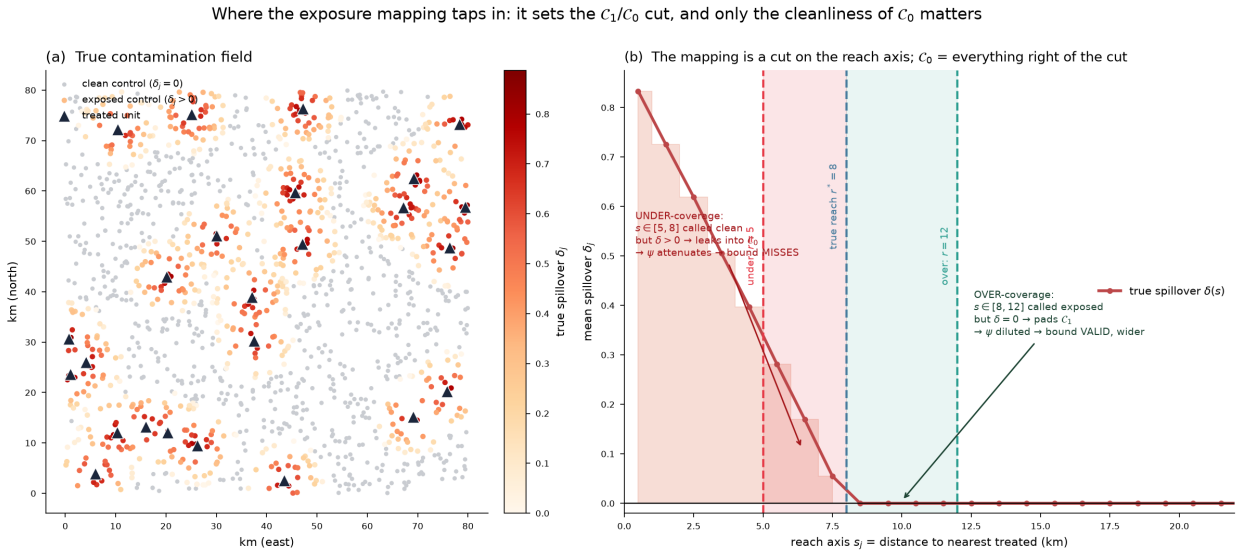


Figure 6: The exposure mapping sets the $\mathcal{C}_1/\mathcal{C}_0$ cut, and only the cleanliness of $\mathcal{C}_0$ matters. Panel (a) shows one draw of the true contamination field: treated units, clean controls ($\delta_j = 0$), and exposed controls shaded by δ_j. Panel (b) orders controls by the reach axis s_j and plots the true $\delta(s)$: an over-covering cut at $r = 12$ pads $\mathcal{C}_1$ with $\delta = 0$ units and dilutes ψ into a valid but wider bound, while an under-covering cut at $r = 5$ leaves $\delta > 0$ units in $\mathcal{C}_0$, attenuates ψ, and makes the bound miss. True reach $r^* = 8$.

S9.2 No-Leakage and Directional-Validity

The asymmetry reframes the dependence from the impossible question "did you get the exposed set exactly right?" to the tractable one "is everything to the right of your cut actually clean?" I state the tractable condition as an assumption on the mapping.

Assumption S1 (No-Leakage)**.** *A candidate mapping m satisfies No-Leakage if every unit it assigns to the clean set is truly clean: $\delta_j = 0$ for all $j \in \mathcal{C}_0(m)$. Equivalently, the mapping's exposed set contains the entire true exposed set, $\mathcal{C}_1^* \subseteq \mathcal{C}_1(m)$, so the mapping may over-cover but does not leave a truly-exposed unit in the reference.*

No-Leakage is weaker than correct specification. Correct specification requires $\mathcal{C}_1(m) = \mathcal{C}_1^*$ exactly. No-Leakage requires only the one-sided containment $\mathcal{C}_1^* \subseteq \mathcal{C}_1(m)$, which every over-covering mapping satisfies. The next proposition shows this weaker condition is enough for validity at all three levels of the identification region.

Proposition S21 (Directional-Validity)**.** *Fix a weight vector w and a candidate mapping m that satisfies No-Leakage (Assumption S1). Then the true contamination $\mathcal{B}(w)$ lies inside the Level 1, Level 2, and Level 3 identification regions built at m, and the Imbens–Manski interval built at m covers $\mathcal{B}(w)$. Over-coverage preserves validity at every level and only widens the region, with the width premium governed by the weight dispersion of Theorem 4.10 (Width of the Level 2 Identification Region) of the main text. If instead No-Leakage fails, so that some $j \in \mathcal{C}_0(m)$ has $\delta_j > 0$, then the identified mean $\psi(m)$ is attenuated toward zero and the region can exclude $\mathcal{B}(w)$.*

Proof. The argument reuses the Level 2 linear-program feasibility argument of the main text. The Level 2 region is the set of values $\sum_j w_j \delta_j$ attained by profiles $\{\delta_j\}$ in the box $\delta_j \in [0, d_j]$ with capacity $d_j = \Delta Y_j^{\text{obs}} - a$ (Assumption 6, Feasibility), subject to the identified mean constraint $\sum_{j \in \mathcal{C}_1(m)} w_j \delta_j = \psi(m)$ (the Level 2 linear programs, equations (35) and (36) of the main text). Consider the true profile $\delta^* = \{\delta_j^*\}$.

Under No-Leakage the true profile is feasible for this program. First, every truly-exposed unit lies in $\mathcal{C}_1(m)$, because $\mathcal{C}_1^* \subseteq \mathcal{C}_1(m)$. The padded units in $\mathcal{C}_1(m) \setminus \mathcal{C}_1^*$ carry $\delta_j^* = 0$ and satisfy the box lower bound, and each unit satisfies $0 \leq \delta_j^* \leq d_j$ by the capacity definition together with monotone nonnegative spillovers, so δ^* is box-feasible. Second, because $\mathcal{C}_0(m)$ is clean, the reference mean identifies $\psi(m)$ with no contamination, and the true profile meets the mean constraint, $\sum_{j \in \mathcal{C}_1(m)} w_j \delta_j^* = \psi(m)$, since the padded zeros add nothing to the sum. Feasibility of δ^* places its objective value $\mathcal{B}(w) = \sum_j w_j \delta_j^*$ between the program's minimum and maximum, so $\mathcal{B}(w)$ lies in the Level 2 region. Level 1 (Definition 4.1 of the main text) drops the mean constraint and is a relaxation, so by Lemma 4.4 (Nesting of Feasible Sets) $\mathcal{B}(w)$ lies in the Level 1 region as well. The additional Level 3 restrictions (Definition 4.3 of the main text) are shape conditions the true spillover profile satisfies by construction, so δ^* remains feasible and $\mathcal{B}(w)$ lies in the Level 3 region. Validity therefore holds at all three levels, and the Imbens–Manski interval, which contains the Level 2 region, covers $\mathcal{B}(w)$. Over-coverage enlarges $\mathcal{C}_1(m)$ by zero-spillover coordinates, which enlarges the feasible set and weakly widens the region without moving the truth out of it, and the size of that widening is the weight-dispersion premium of Theorem 4.10.

Now suppose No-Leakage fails, so a truly-exposed unit j_0 with $\delta_{j_0} > 0$ sits in $\mathcal{C}_0(m)$. The reference group is contaminated, so the estimated clean mean $\mathbb{E}[\Delta Y(0,0) \mid \mathcal{C}_0(m)]$ is inflated by the leaked spillover and the identified gap $\psi(m) = \mathbb{E}[\Delta Y(0,1) - \Delta Y(0,0) \mid E_j = 1, A_j = 0]$ is attenuated toward zero. A too-small $\psi(m)$ shifts the mean constraint downward, and the region built from it can lie entirely below $\mathcal{B}(w)$, excluding the truth. The failure is self-serving, because it makes contamination look smaller than it is. □

Validity is one-sided for every estimator, but the price of over-coverage is estimator-dependent, because the width premium of Theorem 4.10 scales with weight dispersion. For symmetric weights (difference-in-differences, staggered) the padded units receive near-zero marginal weight, so over-coverage is nearly free. For concentrated weights (sparse synthetic control) a padded unit can attract large weight, so over-coverage is expensive. One subtlety survives this account and matters for the selection rule below: for point-identified weights, severe over-coverage does not merely widen the region, it can bias the point, so point-

identified coverage is a two-sided window rather than a one-sided free lunch. I measure all of this in Section S9.5.

S9.3 The clean-set placebo

Directional-Validity makes robustness possible, but it is not usable on its own, because the analyst cannot see which direction the cut erred in. The bridge is a data-side test of No-Leakage, and it is the paper's own doubly robust detector pointed one ring inward (Figure 7). The detector compares $\mathcal{C}_1$ against $\mathcal{C}_0$ and reads a non-zero contrast as interference. The placebo runs the same AIPW contrast, but inside the declared-clean set: the inner boundary ring of $\mathcal{C}_0(m)$ plays the exposed role and the outer core plays the reference role. If $\mathcal{C}_0(m)$ is truly clean, both sub-groups are clean and the contrast is zero. If truly-exposed units leaked into the ring, the contrast is non-zero, which is $\psi \neq 0$ on that inner comparison.

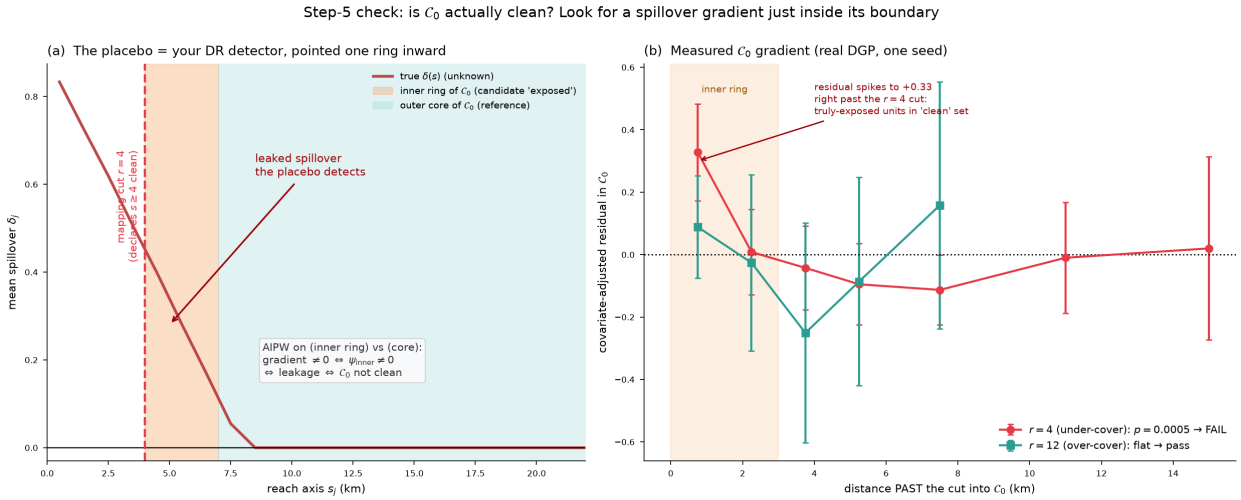


Figure 7: The placebo is the doubly robust detector pointed one ring inward. Panel (a): inside a declared-clean $\mathcal{C}_0$, the inner ring just past the cut is the candidate exposed group and the outer core is the reference. The placebo detects a residual spillover gradient across the ring. Panel (b): the measured covariate-adjusted residual on one draw, showing an under-covering cut at $r = 4$ spiking to $+0.33$ and rejecting ($p = 0.0005$, FAIL) while an over-covering cut at $r = 12$ stays flat and passes.

Given a candidate mapping m with cut at r, the placebo has three steps. First, split $\mathcal{C}_0(m)$ along the reach axis into an inner ring $s_j \in [r, r + h)$ of a few kilometers and an outer core $s_j \geq r + 2h$, the gap avoiding an abutting boundary. The ring is where residual contamination would sit if the cut was early, and the core is the part I am most willing to treat as clean. Second, fit $\hat{\mu}_0(X)$ on the core only and form covariate-adjusted residuals $\hat{r}_j = \Delta Y_j^{\text{obs}} - \hat{\mu}_0(X_j)$ for both groups, so that a surviving ring-versus-core gap is not explained by covariate composition. Third, run the one-sided test $H_0 : \bar{r}_{\text{ring}} \leq \bar{r}_{\text{core}}$. A significant positive gap means leftover spillover in the ring, so the cut is inside the contaminated region and the mapping FAILs. A flat comparison passes, certifying $\mathcal{C}_0(m)$ clean. The test is one-sided because leakage only raises the ring ($\delta \geq 0$), and guarding the other direction would

waste power. This is one call to the existing estimator with the roles reassigned, adding no structural assumption beyond ignorability restricted to $\mathcal{C}_0$ and one new input, the reach axis s.

The placebo's value rests entirely on whether it can reject a cut placed just inside the true boundary, because small under-coverage is the danger, not gross under-coverage. I therefore measure the power directly rather than assume it. Figure 8 plots the rejection curve and Table 82 gives the numbers, over 150 seeds on the radial DGP with true $r^* = 8$.

Table 82: Placebo rejection rate $\mathbb{P}$(FAIL) by analyst radius, 150 seeds, radial DGP, true $r^* = 8$. Size is nominal for $r \geq r^*$, and power decays into a dead zone at $r = 6$–7, where under-coverage by 1–2 km sits below the placebo's resolving power.

analyst r	4	5	6	7	8	9	10	12	14
single-axis	1.00	0.71	0.21	0.05	0.03	0.06	0.07	0.08	0.03
multi-axis	0.97	0.56	0.13	0.03	0.03	0.03	0.04	0.08	0.07

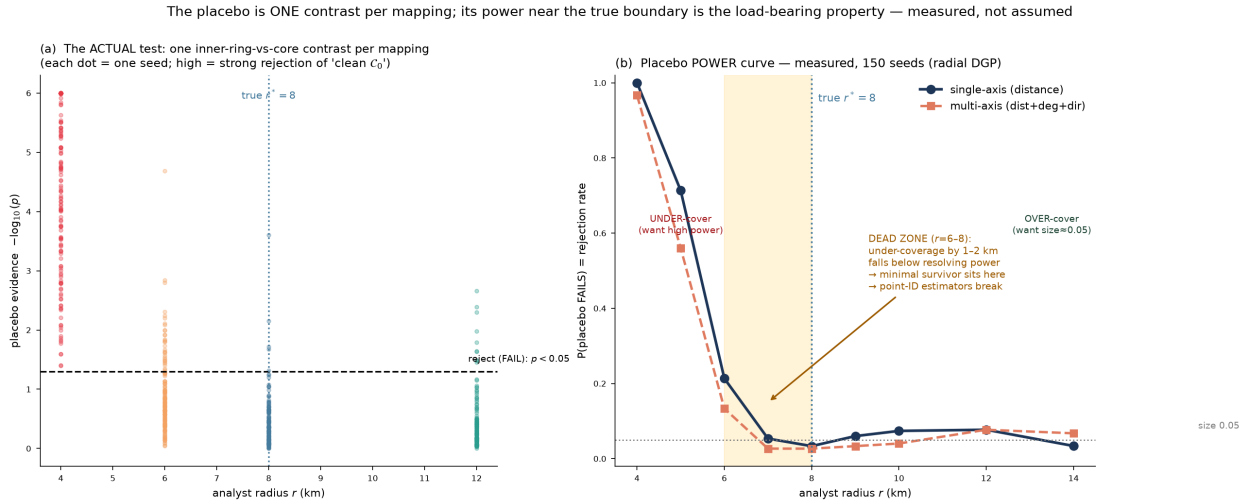


Figure 8: The placebo is one inner-ring-versus-core contrast per mapping, and its power near the true boundary is measured. Panel (a): each dot is one seed's single test statistic at a given analyst radius, not a per-distance-bin profile. Panel (b): the rejection rate for single-axis and multi-axis placebos across analyst radii. Size is nominal (3–8%) for $r \geq r^*$, and power collapses through the dead zone at $r = 6$–7.

Two facts follow. The size is nominal for $r \geq r^*$, at 3 to 8 percent, so the placebo correctly passes clean and over-covering mappings. And there is a dead zone at $r = 6$–7, where power falls from 100 percent at $r = 4$ to 21 percent at $r = 6$ to 5 percent at $r = 7$. A mapping that under-covers by 1 to 2 kilometers sits below the placebo's resolving power. This is the honest finite-sample limit, and it has two consequences the procedure must handle. The minimal placebo-passing mapping sits on this power boundary, where residual undetected leakage is most likely, so the frontier must step off it. And point-identified estimators, which carry no slack, break in the dead zone.

The distance-axis placebo catches leakage organized by distance, but a wrong-family mapping can leave a $\mathcal{C}_0$ that is dirty along a different axis. A radial cut used when contamination is driven by network degree leaves high-degree units in $\mathcal{C}_0$ even though they are far in kilometers, and the distance placebo is blind to that. I run the ring-versus-core contrast on each candidate axis (distance, network degree, directional offset) and combine them with a Bonferroni guard, failing the mapping if any axis rejects at level α/K. The guard is conservative by construction, so it controls false FAILs in the safe direction, and the price is slightly lower power on the axis-aligned case. The recommended default is multi-axis, with the axis set drawn from the substantive candidate mechanisms.

S9.4 The placebo-filtered frontier

The procedure takes a researcher-assembled pool of candidate mappings, filters it with the placebo, selects one mapping to report, and displays the sensitivity of the answer across the pool. Figure 9 shows the flow, and Algorithm 1 states it.

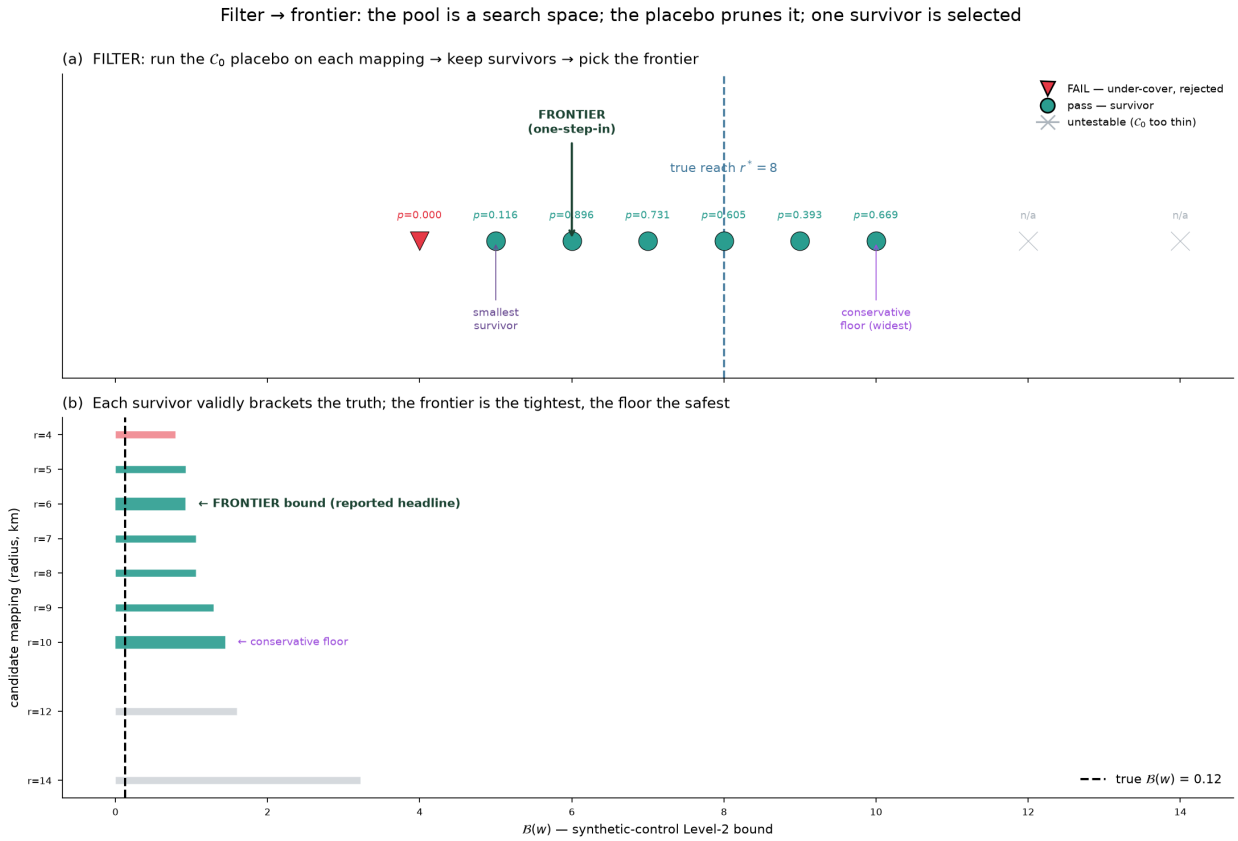


Figure 9: Filter then frontier. Panel (a): the placebo runs on each mapping in the pool. Under-covering mappings FAIL and are discarded, over-covering and exact mappings pass, and very wide mappings whose $\mathcal{C}_0$ is too thin to form a ring and core are untestable. The frontier is the minimal survivor stepped one candidate inward, off the placebo power boundary. Panel (b): every survivor validly brackets the truth, with the frontier the tightest and the conservative floor the widest.

Algorithm 1 Placebo-filtered frontier selection

Require: pool of candidate mappings $\{m_1, \ldots, m_n\}$, reach axes $\{s^{(1)}, \ldots, s^{(K)}\}$, ring width h, level α, weight vector w.

Step 1. Filter (validity).

for each mapping m in the pool **do**
  for each axis $k = 1, \ldots, K$ **do**
    split $\mathcal{C}_0(m)$ into inner ring $\{s_j^{(k)} \in [r, r+h)\}$ and outer core $\{s_j^{(k)} \geq r + 2h\}$
    fit $\hat{\mu}_0(X)$ on the core, then form residuals $\hat{r}_j = \Delta Y_j^{\text{obs}} - \hat{\mu}_0(X_j)$
    run the one-sided AIPW ring-versus-core contrast at level α/K
  end for
  mark m as FAIL if any axis rejects, PASS if all axes are flat, UNTESTABLE if $\mathcal{C}_0(m)$ is too thin to form a ring and a core
end for
survivors $\leftarrow$ the PASS mappings

Step 2. Frontier (selection).

order survivors by exposed-set size $n_1(m)$
frontier $\leftarrow$ the minimal survivor stepped one candidate inward (one notch wider, off the placebo power boundary)

Step 3. Run once.

at the frontier mapping, detect $\hat{\psi}$ and its CI, and if interference is detected, build the Level 1 and Level 2 regions and the Imbens–Manski interval

Step 4. Report.

headline $\leftarrow$ the frontier bound
floor $\leftarrow$ the widest survivor's bound
profile $\leftarrow$ every survivor's bound as a function of reach
return headline, floor, profile

Selection uses validity and reach ordering only, both computed without reference to the sign or magnitude of ψ, so the selected mapping is ancillary to the detection contrast and choosing it cannot bias $\hat{\psi}$ toward detection. There is no winner's curse and no detection multiplicity in the headline. Which survivor to report as the headline depends on the identification type, and this is the one design subtlety. Point-identified weights (difference-in-differences, the point-identified case of the paper's UPP application) must take the frontier, because over-coverage dilutes the point and the conservative floor over-covers maximally. Wide partial-identified weights (sparse synthetic control, the partially-identified case of the paper's Maricá application) benefit from the floor, whose over-coverage is a genuine free lunch at a width cost. Narrow partial-identified weights (staggered, dense synthetic control) fall in between, where the frontier and the floor agree within a few points and the frontier is preferred for width. The recommended default reports the frontier as the headline for all estimators, since it is valid for all and optimal for point-identification, and shows the floor and the profile as the robustness display.

Detection is run once, at the frontier, so there is no detection multiplicity in the headline. The placebo tests are multiple, but they run on the ancillary within-$\mathcal{C}_0$ contrast rather than the detection contrast, and they only prune the pool rather than select a winning p-value,

with the Bonferroni guard controlling their family-wise error in the conservative direction. Bounds are built once, at the frontier, and the profile is a display rather than a family of competing inferences.

S9.5 Simulation evidence

All results below run on the paper's own R engine (the `dr_psi_estimator`, `level2_bounds`, `imbens_manski_ci`, and `weights` routines). The DGP places $N = 1500$ units on an 80×80 km map, 25 treated, true reach $r^* = 8$ km, and four true contamination geometries. The roster is four weight families, difference-in-differences (uniform, point-identified), staggered (cohort blocks), sparse synthetic control (concentrated, HHI ≈ 0.30), and dense synthetic control (spread, HHI ≈ 0.012), with both coverage objects recorded for all four, the Imbens–Manski interval and the Level 2 identified set. The grid is 300 seeds $\times$ 4 geometries $\times$ 10 radii $\times$ 4 families $\times$ two placebos, giving 47,472 sweep rows and 28,800 pool rows.

S9.5.1 Coverage by family, and the point-identified window

Table 83 reports coverage of $\mathcal{B}(w)$ on the radial geometry by analyst radius, with difference-in-differences read through the Imbens–Manski interval and the other families through the Level 2 set. Figure 10 shows the same across all four geometries.

Table 83: Coverage of $\mathcal{B}(w)$ by weight family, radial geometry, 300 seeds, true $r^* = 8$. Difference-in-differences is a two-sided window. Sparse synthetic control is the one-sided free lunch. Staggered and dense synthetic control interpolate by weight dispersion.

r	4	5	6	7	8	9	10	12	14	18
DiD	.82	.97	1.0	1.0	1.0	.99	.96	.86	.68	.42
Staggered	.45	.86	.98	1.0	1.0	1.0	.99	.98	.91	.67
SC-sparse	.67	.89	.96	1.0	1.0	1.0	1.0	1.0	1.0	.93
SC-dense	.95	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.92

The estimator-dependence is now measured for all four families. Difference-in-differences is a two-sided window, at 1.0 for $r = 6$–8 and decaying both below (0.82 at $r = 4$) and above (0.42 at $r = 18$), so over-coverage is not a free lunch for point-identification. Sparse synthetic control is the one-sided free lunch, flat at 1.0 from $r = 7$ to $r = 14$, where over-coverage costs only width. Staggered and dense synthetic control interpolate by weight dispersion: the staggered narrow Level 2 set is more fragile to under-coverage (0.45 at $r = 4$), and the dense synthetic control set is robust to over-coverage but dips at the extreme (0.92 at $r = 18$). The practical consequence for point-identification is that the safe region is a window, roughly $r \in [6, 10]$ here, so the analyst should report the frontier rather than the conservative floor, which sits at the wide end of the window and can fall out of it.

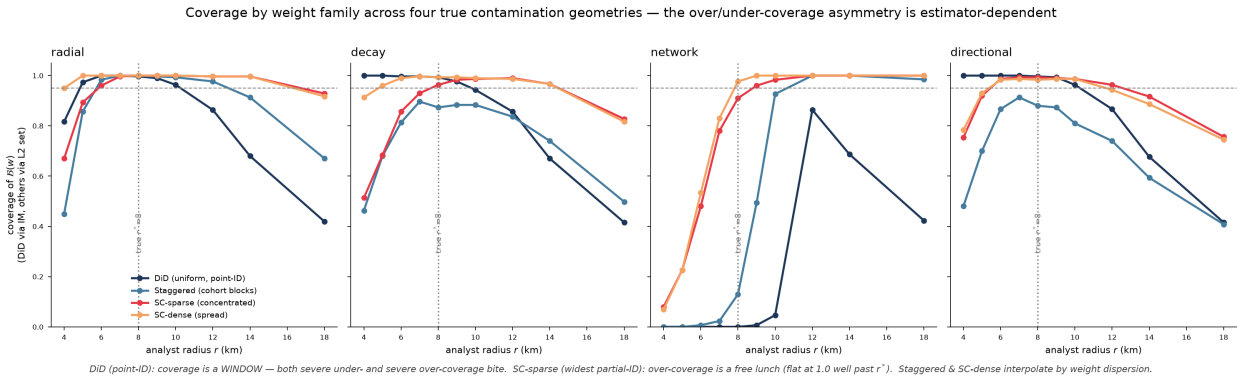


Figure 10: Coverage by weight family across the four true contamination geometries. The over/under-coverage asymmetry is estimator-dependent: difference-in-differences is a window, sparse synthetic control is flat at 1.0 well past r^*, and the network geometry is the wrong-family stress test, where radial mappings fail until the radius is large because a small radial cut leaves degree-driven contamination in $\mathcal{C}_0$.

S9.5.2 The price of over-coverage

Table 84 and Figure 11 report Level 2 bound widths under under-coverage, perfect coverage, and over-coverage, averaged over 300 seeds on the radial geometry.

Table 84: Level 2 bound width by weight family and coverage regime, radial geometry, 300 seeds. The over-coverage premium tracks weight concentration exactly as the width theorem predicts.

family	under ($r = 6$)	perfect ($r = 8$)	over ($r = 12$)	over-coverage cost
DiD	0.00[a]	0.00	0.00	free in width, but point dilutes
Staggered	0.26	0.32	0.32	near free (saturates)
SC-dense	0.67	0.92	1.20	moderate (+30%)
SC-sparse	0.85	1.28	2.57	expensive (+100%)

[a] Difference-in-differences point-identifies, so the Level 2 set is a point of zero width at every radius. The cost of over-coverage shows up in the location of the point rather than in width.

The width premium of over-coverage tracks weight concentration, free for symmetric weights and expensive for concentrated ones, matching Theorem 4.10 of the main text.

S9.5.3 Why bad mappings survive, and the multi-axis fix

Table 85 and Figure 12 decompose an adversarial mixed pool on the radial-true DGP by how each mapping is bad, over 80 seeds.

There are two kinds of bad mapping, needing different remedies. The dirty-$\mathcal{C}_0$ leakers, the network and directional mappings, leave true spillover in $\mathcal{C}_0$ along the degree or direction axis that the distance-only placebo cannot see, and the multi-axis placebo catches them, dropping the network survival rate from 0.99 to 0.58 and the directional rate from 0.96

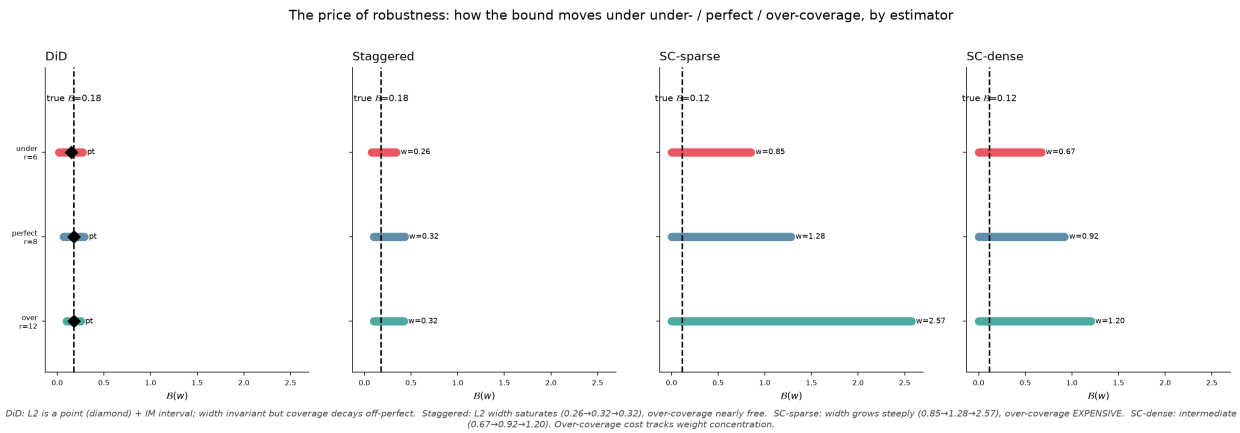


Figure 11: How the bound moves under under, perfect, and over-coverage, by estimator. Difference-in-differences stays a point with width invariant across regimes. Staggered width saturates ($0.26 \to 0.32 \to 0.32$). Dense synthetic control is intermediate ($0.67 \to 0.92 \to 1.20$). Sparse synthetic control grows steeply ($0.85 \to 1.28 \to 2.57$), so over-coverage is expensive there.

Table 85: Placebo survival and outcome by mapping, radial-true DGP, 80 seeds. The distance-only single-axis placebo lets wrong-family leakers through, and the multi-axis placebo catches them. Clean-$\mathcal{C}_0$ over-coverers pass validly and cannot be rejected.

mapping	passes single	passes multi	leak in $\mathcal{C}_0$	DiD coverage
radial $r = 4$ (under)	0.00	0.01	0.092 (dirty)	0.81
network deg ≥ 3 (wrong family)	0.99	0.58	0.119 (dirty)	0.44
directional (wrong shape)	0.96	0.45	0.121 (dirty)	0.50
radial $r = 8$ (perfect)	0.95	0.96	0.000 (clean)	1.00
decay (good)	0.94	0.96	0.010 (clean)	1.00
radial $r = 14$ (over)	0.98	0.93	0.000 (clean)	0.68

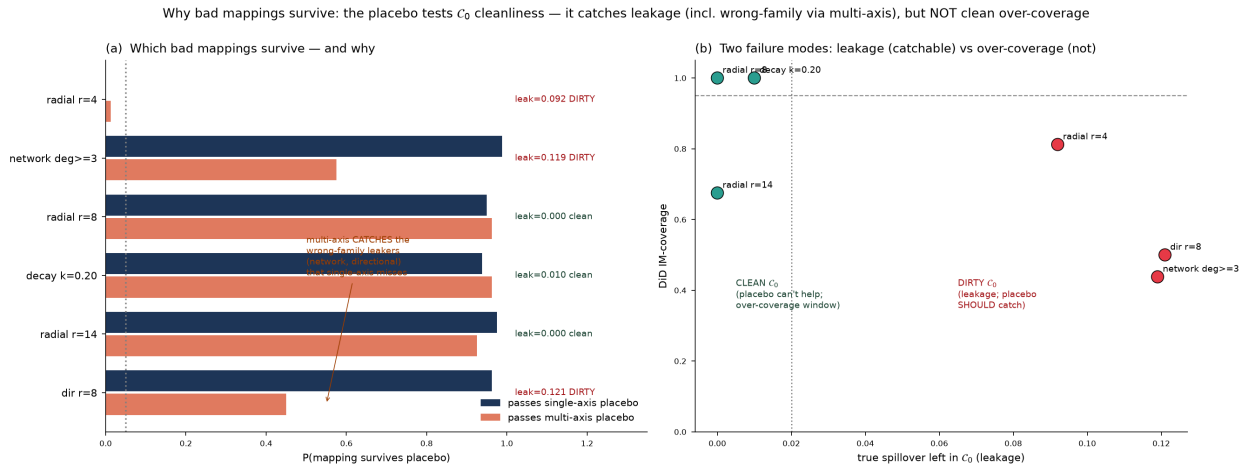


Figure 12: Two failure modes. Panel (a): survival probability under the single-axis and multi-axis placebos, with the leak left in $\mathcal{C}_0$ annotated. Panel (b): DiD coverage against the leakage in $\mathcal{C}_0$, separating clean-$\mathcal{C}_0$ over-coverers (the placebo cannot and should not reject them) from dirty-$\mathcal{C}_0$ leakers (the placebo should catch them).

to 0.45. This directly improves the point-identified validity certificate. The clean-$\mathcal{C}_0$ over-coverers, radial $r = 14$, have a genuinely clean $\mathcal{C}_0$ (leak 0.000), so no placebo can or should reject them. They are valid but wide for partial-identification (synthetic control coverage 1.0) and harmful only for point-identification (0.68), where the remedy is the frontier rule rather than a better test. The mechanism is structural rather than sampling noise, so more seeds would change neither conclusion.

S9.5.4 Frontier versus floor, and the sensitivity plot

Table 86 and Figure 13 report coverage and Level 2 width for the frontier and the floor in the adversarial mixed pool under the multi-axis placebo.

Table 86: Frontier versus floor, mixed pool, multi-axis placebo. Point-identification takes the frontier, and wide partial-identification benefits from the floor at a real width cost. Coverage is of the primary object, and width is Level 2.

family	frontier cov	floor cov	frontier width	floor width
DiD	0.84	0.69	0.00	0.00
Staggered	0.75	0.82	0.35	0.40
SC-sparse	0.88	0.97	1.31	2.26
SC-dense	0.92	0.97	1.00	1.37

Point-identification must use the frontier, 0.84 against 0.69, because the floor over-dilutes it, while wide partial-identification benefits from the floor, 0.97 against 0.88, at a real width cost of 2.26 against 1.31 for sparse synthetic control. The frontier is valid for every estimator, optimal for point-identification, and within a few points of the floor for narrow partial-

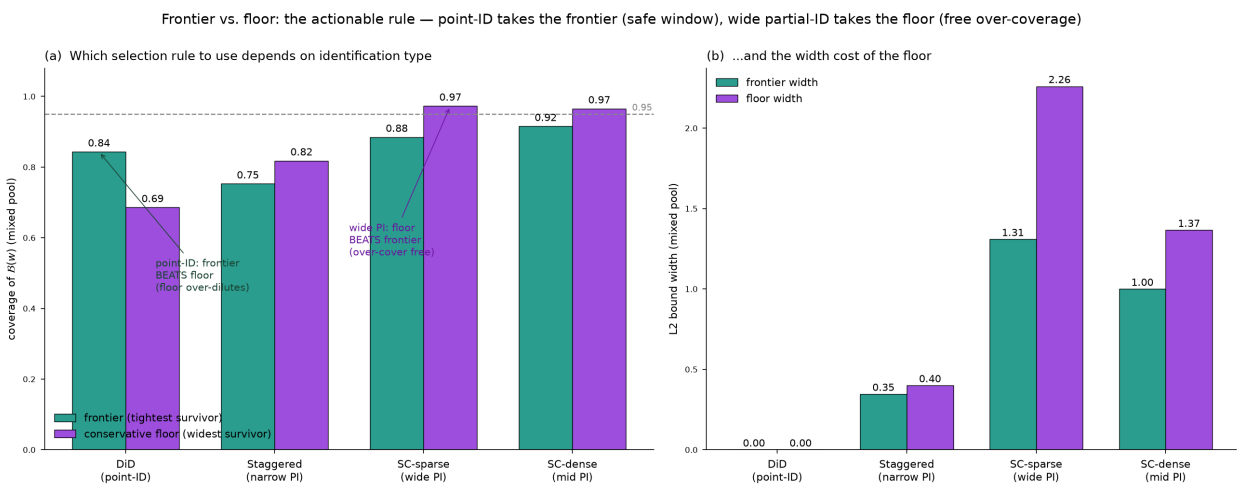


Figure 13: The actionable rule. Panel (a): point-identification takes the frontier (0.84 versus 0.69 for difference-in-differences, where the floor over-dilutes the point), while wide partial-identification benefits from the floor (0.97 versus 0.88 for sparse synthetic control). Panel (b): the width cost of the floor, which is largest for concentrated weights.

identification. Only this adversarial mixed pool separates the rules. The perfect and few-mapping pools reach at least 0.93 frontier coverage for all families.

The object the applied researcher reports is the sensitivity plot, Figure 14: bounds against exposure radius, one panel per estimator, showing the Level 2 identified set as a band, the Imbens–Manski interval as a dashed pair, the placebo verdict at each radius, and the true $\mathcal{B}(w)$. Read from left to right, the bound becomes valid as the radius crosses out of the placebo-rejection region and then widens with over-coverage, and the true value sits inside every valid bound. This is what an applied researcher publishes to answer a referee's misspecification question.

References

Bang, H. and Robins, J. M. (2005), 'Doubly robust estimation in missing data and causal inference models', *Biometrics* **61**(4), 962–973.

Bertsekas, D. P. (1999), *Nonlinear Programming*, 2nd edn, Athena Scientific, Belmont, MA.

Bertsimas, D. and Tsitsiklis, J. N. (1997), *Introduction to Linear Optimization*, Athena Scientific, Belmont, MA.

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.

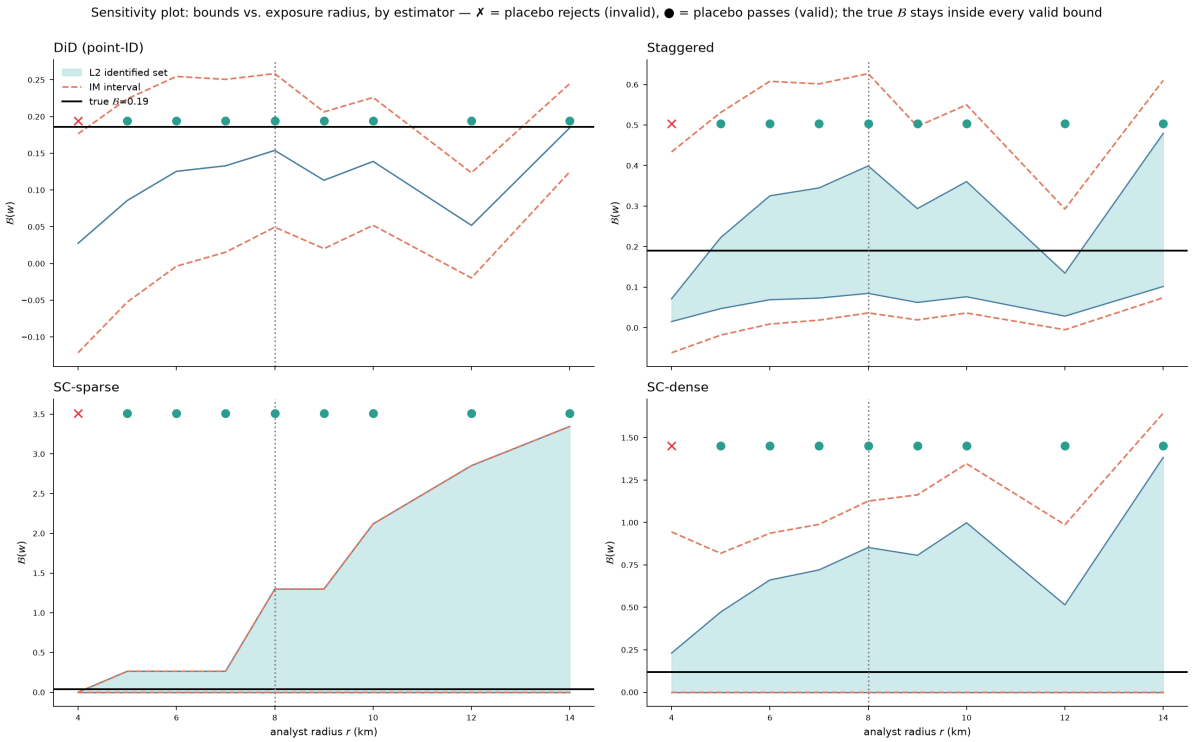


Figure 14: Sensitivity plot: bounds against exposure radius, by estimator. A cross marks a placebo rejection (invalid mapping) and a dot marks a placebo pass (valid mapping). The Level 2 identified set is the band and the Imbens–Manski interval the dashed pair. As the radius grows past the placebo-rejection region the bound becomes valid and widens, the point-identified Imbens–Manski interval stays tight, the partial-identified sets widen with over-coverage, and the true $\mathcal{B}(w)$ stays inside every valid bound.

Fang, Z. and Santos, A. (2019), 'Inference on directionally differentiable functions', *Review of Economic Studies* **86**(1), 377–412.

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.

Marshall, A. W., Olkin, I. and Arnold, B. C. (2011), *Inequalities: Theory of Majorization and Its Applications*, 2nd edn, Springer, New York.

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.

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