

Estimating Total Effects in Bipartite Experiments with Spillovers and Partial Eligibility

Albert Tan^{*} Mohsen Bayati^{*,†} James Nordlund^{*} Roman Istomin^{*}

Abstract

We study randomized experiments in bipartite systems where only a subset of treatment-side units are eligible for assignment while all units continue to interact, generating interference. We formalize *eligibility-constrained bipartite experiments* and define estimands aligned with full deployment: the *Primary Total Treatment Effect* (PTTE) on eligible units and the *Secondary Total Treatment Effect* (STTE) on ineligible units. Under randomization within the eligible set, we give identification conditions and develop interference-aware ensemble estimators that combine exposure mappings, generalized propensity scores, and flexible machine learning. We further introduce a projection that links treatment- and outcome-level estimands; this mapping is exact under a *Linear Additive Edges* condition and enables estimation on the (typically much smaller) treatment side with deterministic aggregation to outcomes. In simulations with known ground truth across realistic exposure regimes, the proposed estimators recover PTTE and STTE with low bias and variance and reduce the bias that could arise when interference is ignored. Two field experiments illustrate practical relevance: our method corrects the direction of expected interference bias for a pre-specified metric in both studies and reverses the sign and significance of the primary decision metric in one case.

1 Introduction

Traditional A/B testing assumes the Stable Unit Treatment Value Assumption (SUTVA), under which each unit's outcome depends only on its own treatment (Cox, 1958). Interference or spillover, the violation of this assumption, is ubiquitous in networked settings. Seminal contributions clarified identification and inference under interference (Halloran and Struchiner, 1995; Sobel, 2006; Hudgens and Halloran, 2008); for a concise review of design and analysis choices when SUTVA fails, see (Athey and Imbens, 2017).

In many applications there is a natural bipartite (two-sided) structure underlying the interference. For example, in ride-sharing services the *treatment units* may be drivers (e.g., a new routing or dispatch policy) while outcomes are measured on riders (wait times, booking rates, satisfaction). A rider may interact with multiple drivers, and a driver's treatment can affect not only their riders but also other nearby drivers and their riders, creating interference patterns that invalidate standard A/B analyses. Importantly, only a *subset* of drivers may be eligible for treatment (e.g., those in specific cities, app versions, or fleet programs). Even when assignment is restricted to these eligible drivers, ineligible drivers remain in the system and continue to affect matching, congestion, and learning dynamics.

We study this problem by extending the bipartite experimental design framework of Doudchenko et al. (2020) to *eligibility-constrained bipartite experiments*, where randomization is feasible only among a subset of treatment-side units but interference propagates through all units. Our design formalizes (i) eligibility sets and exposure mappings that allow ineligible units to affect outcomes, and (ii) identification under randomization within the eligible set. Methodologically, we incorporate rich interaction data between treatment and outcome units, build on generalized propensity score ideas of Hirano and Imbens (2004), and leverage modern machine learning to flexibly model exposure-response relationships.

^{*} Amazon [†] Stanford University

1.1 Intuition Behind Primary and Secondary Treatment Effects

Consider a hypothetical ride-sharing service that wants to test a new app design feature: enhanced visual prominence and preferential placement for economy vehicles in the rider interface. The service operates with multiple vehicle categories: economy cars (compact sedans), premium vehicles (luxury sedans), and XL vehicles (SUVs). The treatment involves displaying economy vehicles at the top of the selection screen with highlighted badges indicating “Quick Pickup” or “Eco-Friendly Choice,” enhanced driver profiles showing ratings and estimated arrival times more prominently, and subtle visual animations that draw attention to these options. Due to technical constraints or strategic considerations, this enhanced presentation can only be applied to economy vehicles, while premium and XL vehicles maintain their standard display format.

In this setting, economy vehicles represent our primary units; they are eligible for the enhanced app treatment and can receive the preferential placement. Premium and XL vehicles constitute our secondary units; while they cannot receive the enhanced presentation directly, they remain active in the system and may experience indirect effects. When economy vehicles become more visually prominent and appear first in the selection interface, rider behavior shifts in many ways. Some riders who might have scrolled down to select premium vehicles now choose the prominently displayed economy option. Others, particularly those looking for luxury or larger vehicles, may actually become more likely to select premium or XL options as a conscious reaction against the service’s apparent steering toward economy vehicles.

We introduce the notion of Primary Total Treatment Effect (PTTE) that captures what happens to economy vehicles when we compare two scenarios: one where all economy vehicles receive the enhanced app treatment versus one where none do. This effect encompasses not only the direct impact on economy vehicle utilization rates but also the competitive dynamics that emerge. For instance, when all economy vehicles have enhanced placement, riders cannot simply scroll past to find a “normal” economy option; they must either select from the prominently featured economy vehicles or deliberately choose a different vehicle category. The PTTE thus measures the true impact on the primary units when the feature is fully deployed.

Similarly, we introduce the notion of Secondary Total Treatment Effect (STTE) that quantifies the impact on premium and XL vehicles. Even though these vehicles never receive the enhanced app treatment, their outcomes may change substantially. Premium vehicles might see decreased demand as riders are steered toward the more prominent economy options. Conversely, they might experience increased utilization from riders who interpret the economy vehicle promotion as a signal of lower quality or longer wait times, leading them to “upgrade” their choice. XL vehicles might capture riders traveling in groups who previously would have booked multiple economy vehicles but now, seeing the service’s emphasis on individual economy rides, opt for a single larger vehicle instead. The magnitude and direction of these effects depend on complex behavioral patterns, user interface design principles, and rider preferences that vary across time and geography.

The main insight is that traditional experimentation, which might randomly assign the enhanced app treatment to half of economy vehicles, creates an artificial environment. Riders see an inconsistent interface where some economy vehicles appear with enhanced prominence while others appear normally, potentially causing confusion and unnatural selection patterns. This leads to misestimating the treatment effect; at full deployment, the consistent interface creates different choice architecture and behavioral responses. The methodology studied in this paper aims to correct for these biases by explicitly modeling the bipartite structure of rider-driver interactions and the eligibility constraints that define which units can receive the app treatment.

1.2 Related Work

The foundations of causal inference under interference draw on early epidemiologic formulations for vaccine trials (Halloran and Struchiner, 1995) and potential-outcomes frameworks that formalized estimands under partial interference and group randomization (Sobel, 2006; Hudgens and Halloran, 2008), as well as neighborhood-based interference (Sussman and Airola, 2017). Subsequent work developed randomization-based tests and design principles (Rosenbaum, 2007), generalized estimands and inverse-probability estimators for arbitrary interference (Tchetgen Tchetgen and VanderWeele, 2012; Aronow and Samii, 2017), and clarified what standard estimators target when interference is unknown (Sävje et al., 2021). These threads motivate network-aware designs such as graph-cluster randomization and bias-reducing analyses (Ugander et al., 2013; Eckles et al., 2017), which our bipartite and eligibility-constraint setting builds upon.

More recent work considers settings where the interference network is only partially observed or entirely unknown, developing designs and estimators that remain valid in such cases (Agarwal et al., 2023; Yu et al., 2022; Cortez et al., 2022; Shirani and Bayati, 2024; Shirani et al., 2025). Within the class of experiments where a bipartite interaction structure between treatment and outcome units is observed, a growing literature studies design and analysis of such experiments. Pouget-Abadie et al. (2019) propose correlation-clustering designs that reduce variance under linear exposure mappings, and Harshaw et al. (2023) introduce the Exposure Reweighted Linear estimator and associated clustering strategies for unbiased and asymptotically normal total effect estimation. Brennan et al. (2022) study cluster-randomized designs in one-sided bipartite settings, and Shi et al. (2024) develop scalable randomization-based inference and covariate-adjusted estimators for large-scale bipartite experiments. Our paper builds directly on Doudchenko et al. (2020), who introduced the bipartite framework and generalized propensity score ideas for two-sided experiments, and is complementary to these design- and analysis-focused contributions.

Beyond randomized experiments, there is emerging work on bipartite interference in quasi-experimental and longitudinal settings. Chen et al. (2024) develop a difference-in-differences framework under bipartite network interference with staggered adoption, reconfiguring panel data so that analysis is carried out at the intervention-unit level while preserving interpretation at the outcome level. Conceptually, their mapping between levels is related to our projection between treatment- and outcome-level estimands, but our focus is on randomized experiments with partial eligibility and explicit primary and secondary total effects.

A related line of work examines experimental design and equilibrium responses in large two-sided systems. Bajari et al. (2023) survey multiple randomization designs and interference-aware decision metrics, while Holtz et al. (2020) analyze interference bias in pricing experiments. Johari et al. (2022) and Munro et al. (2021) study equilibrium adjustments to interventions on one side of a service and the implications for experimentation. Our framework differs by imposing explicit eligibility constraints on treatment-side units and defining primary and secondary total treatment effects together with a projection mapping between treatment- and outcome-level estimands.

There is also a broad methodological literature on interference and spillovers spanning statistics, econometrics, and related fields. Representative contributions include design-based inference and randomization tests under general interference (Athey et al., 2018; Basse et al., 2019; Puelz et al., 2022), designs for networked and cluster experiments (Jagadeesan et al., 2020; Baird et al., 2018), two-stage settings and noncompliance (Basse and Feller, 2018; Imai et al., 2021), identification under partial or unknown network exposure (Leung, 2022; Egami, 2021; Forastiere et al., 2021), and estimation of direct and indirect effects (Hu et al., 2022; Choi, 2017). Additional perspectives include auto- g -computation on networks (Tchetgen Tchetgen et al., 2021), peer encouragement designs (Kang and Imbens, 2016), and frameworks for bipartite interference (Zigler and Papadogeorgou, 2021). For broader syntheses of interference and network-aware experimentation, see Ogburn and VanderWeele (2014); VanderWeele and An (2013); Benjamin-Chung et al. (2018); Keele and Kang (2022); Athey and Imbens (2017).

We contribute along two dimensions. First, we formalize eligibility-constrained bipartite designs in which only a subset of treatment-side units can be randomized while ineligible units still participate in interference; we define estimands (PTTE and STTE) and give identification under randomization within the eligible set. Second, we develop and analyze a projection mapping between treatment- and outcome-level total effects that holds under a *Linear Additive Edges* assumption. This mapping preserves the estimand, enables estimation on the typically much smaller set of treatment units with projection-based inference to the outcome level, and delivers substantial computational gains without sacrificing interpretability.

2 Methodology

This section develops our framework for eligibility-constrained bipartite experiments. For ease of reference, all notation is summarized in Table 4 in the Appendix. We begin by formalizing the experimental setup and introducing the distinction between primary (eligible) and secondary (ineligible) treatment units (Section 2.1). We then develop ensemble estimators at two levels: first for outcome units like riders (Section 2.2) and then for treatment units like drivers (Section 2.3), defining the Primary Total Treatment Effect (PTTE) at each level. Section 2.4 presents the formal assumptions required for identification. Our projection method (Section 2.5) links treatment effects across levels under a linear additive edges assumption and enables com-

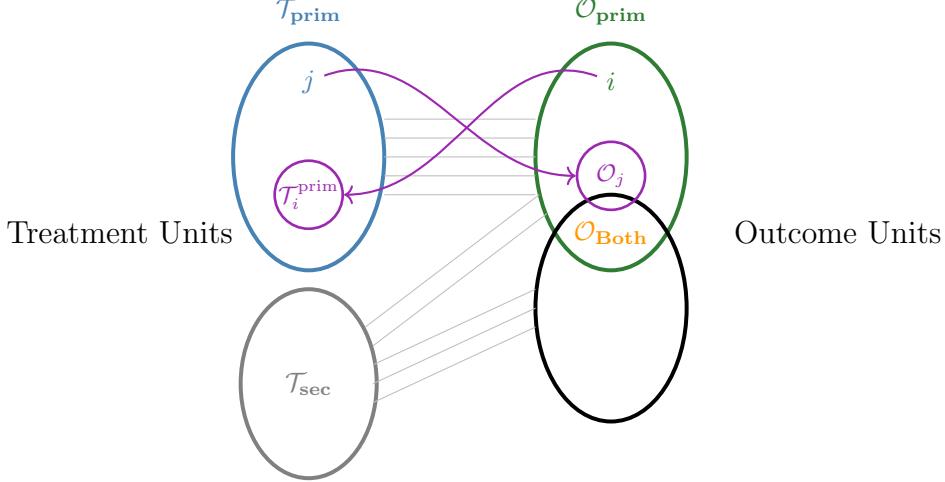


Figure 1. Bipartite structure in eligibility-constrained experiments. Primary units $\mathcal{T}_{\text{prim}}$ (top left) are eligible for treatment; secondary units \mathcal{T}_{sec} (bottom left) are ineligible but connected to outcomes. Outcome units $\mathcal{O}_{\text{prim}}$ (top right) connect to primaries; $\mathcal{O}_{\text{Both}}$ (overlap) connect to both.

putationally efficient estimation. Section 2.6 details our machine learning-based estimation procedures that leverage generalized propensity scores and network structure. Finally, Section 2.7 extends the framework to quantify Secondary Total Treatment Effects (STTE) on ineligible units, capturing the full ecosystem impact of the intervention.

2.1 Problem setup

Consider a bipartite experiment with two types of units: treatment units (e.g., drivers in a ride-sharing service) and outcome units (e.g., riders), where outcomes are observed for both types. In our ride-sharing example, enhanced app placement may be deployed to certain vehicle types, affecting both driver utilization and rider selection patterns. We distinguish between two categories of treatment units, as illustrated in Figure 1. Primary units ($\mathcal{T}_{\text{prim}}$) are those eligible for treatment. Secondary units (\mathcal{T}_{sec}), while not eligible for direct treatment, may be affected through spillover effects due to shared connections with outcome units. The right side of Figure 1 represents the outcome units, with connections between treatment and outcome units shown as edges in the graph. We denote $\mathcal{T}_i^{\text{prim}}$ as the set of primary units connected to outcome unit i , and conversely, \mathcal{O}_j represents the set of outcome units connected to treatment unit j , and $\mathcal{O}_{\text{prim}}$ is the set of outcomes connected to at least one primary treatment unit. The overlap region $\mathcal{O}_{\text{Both}}$ captures outcome units exposed to both primary and secondary treatment units. The matrix \mathbf{w} captures the relationship structure, where w_{ij} represents the connection strength between outcome unit i and treatment unit j .

Edge-level representation and additive metrics. In many applications of interest, including our motivating ride-sharing example, it is natural to think of outcomes as aggregating over interactions along edges of the bipartite graph: each ride corresponds to a rider–driver edge, each impression to an advertiser–viewer edge, and so on. We formalize this by introducing edge-level potential outcomes $Y_{ij}(\mathbf{Z})$ for every existing edge (i, j) , with $Y_{ij}(\mathbf{Z}) = 0$ when (i, j) is absent. We then define

$$Y_i(\mathbf{Z}) = \sum_j Y_{ij}(\mathbf{Z}), \quad Y_j(\mathbf{Z}) = \sum_i Y_{ij}(\mathbf{Z}),$$

so that each unit’s outcome is the sum over its incident edges. Many business metrics (counts of rides, orders, impressions, revenue, and their additively weighted variants) are naturally of this form.

Primary and secondary outcome components. Let $\mathcal{T}_{\text{prim}}$ and \mathcal{T}_{sec} denote the sets of primary (eligible) and secondary (ineligible) treatment units, respectively. For any outcome unit i we decompose its outcome

as

$$Y_i(\mathbf{Z}) = \sum_{j \in \mathcal{T}_{\text{prim}}} Y_{ij}(\mathbf{Z}) + \sum_{j \in \mathcal{T}_{\text{sec}}} Y_{ij}(\mathbf{Z}) =: Y_{i,\text{prim}}(\mathbf{Z}) + Y_{i,\text{sec}}(\mathbf{Z}).$$

Here $Y_{i,\text{prim}}(\mathbf{Z})$ collects the contribution coming through primary treatment units, while $Y_{i,\text{sec}}(\mathbf{Z})$ collects the contribution coming through secondary units. In the ride-sharing example, $Y_{i,\text{prim}}$ would be the part of rider i 's metric attributable to trips taken with primary (eligible) vehicles, and $Y_{i,\text{sec}}$ the part attributable to secondary vehicle categories.

2.2 Outcome-Unit Ensemble

For each outcome unit i , we define exposure as:

$$E_i(\mathbf{Z}) = \sum_{j \in \mathcal{T}_i^{\text{prim}}} w_{ij} Z_j, \quad (1)$$

where $\mathcal{T}_i^{\text{prim}}$ is the set of primary units connected to i , Z_j indicates treatment assignment of unit j , and \mathbf{Z} is a vector of size $|\mathcal{T}_{\text{prim}}| + |\mathcal{T}_{\text{sec}}|$, containing treatment assignment of all units. Note that treatment assignment of secondary units is always equal zero because they are not eligible for treatment, i.e., the last $|\mathcal{T}_{\text{sec}}|$ coordinates of \mathbf{Z} are always equal to zero. The outcome specification becomes:

$$Y_{i,\text{prim}}(\mathbf{w}, E_i, X_i) = \Phi\left(n_i^{\text{prim}}, E_i, r(E_i, \mathbf{w}, X_i)\right) + \epsilon_i \quad (2)$$

where n_i^{prim} is the number of primary units connected to i , $r(\cdot)$ is the generalized propensity score, X_i are covariates, ϵ_i is the error term, and Φ is the outcome function. In the implementations, we utilize binomial expression for the propensity score which effectively specializes to the case $w_{ij} = 1/n_i^{\text{prim}}$ which means $E_i(\mathbf{Z})$ would be the fraction of treated primary neighbors.

Let $\mathbf{Z}^{(1)}$ denote the assignment where all primary units are treated (and secondary units are ineligible), and $\mathbf{Z}^{(0)} \equiv \mathbf{0}$ denote the assignment where no units are treated. It is straightforward to see that under $\mathbf{Z}^{(1)}$ all exposures in $\mathcal{O}_{\text{prim}}$ are equal to 1 and under $\mathbf{Z}^{(0)}$ all such exposures are equal to 0. Hence, at the outcome level, we define the Primary Total Treatment Effect (PTTE) as

$$\begin{aligned} \text{PTTE}_{\text{outcome}} &= \frac{1}{|\mathcal{O}_{\text{prim}}|} \sum_{i \in \mathcal{O}_{\text{prim}}} \left(\mathbb{E}[Y_{i,\text{prim}}(\mathbf{Z}^{(1)})] - \mathbb{E}[Y_{i,\text{prim}}(\mathbf{Z}^{(0)})] \right) \\ &= \frac{1}{|\mathcal{O}_{\text{prim}}|} \sum_{i \in \mathcal{O}_{\text{prim}}} \left(\mathbb{E}[Y_{i,\text{prim}}(E_i = 1)] - \mathbb{E}[Y_{i,\text{prim}}(E_i = 0)] \right). \end{aligned} \quad (3)$$

That is, $\text{PTTE}_{\text{outcome}}$ captures the change in the primary-edge component of outcome-unit metrics when we move from no primary units treated to all primary units treated. Note that, expectation (with notation \mathbb{E}) is taken with respect to all randomness. We use $Y_{i,\text{prim}}(e)$ to denote the potential outcome of unit i under the exposure level e , and model $Y_{i,\text{prim}}(e)$ via (2) with \mathbf{w} and X_i being fixed throughout. This specification corresponds to an *exposure mapping assumption*: for any two assignments \mathbf{Z} and \mathbf{Z}' with $E_i(\mathbf{Z}) = E_i(\mathbf{Z}')$, we have $Y_{i,\text{prim}}(\mathbf{Z}) = Y_{i,\text{prim}}(\mathbf{Z}')$

2.3 Treatment-Unit Ensemble

Denoting indicator function of an event A by $\mathbb{I}(A)$, at the treatment-unit level, we define direct exposure by

$$E_j^{\text{Dir}}(\mathbf{Z}) = Z_j \sum_{i \in \mathcal{O}_{\text{prim}}} \mathbb{I}(j \in \mathcal{T}_i^{\text{prim}}), \quad (4)$$

and indirect exposure by

$$E_j^{\text{Ind}}(\mathbf{Z}) = \sum_{i \in \mathcal{O}_{\text{prim}}} \left[\mathbb{I}(j \in \mathcal{T}_i^{\text{prim}}) \sum_{k \neq j} Z_k \cdot \mathbb{I}(k \in \mathcal{T}_i^{\text{prim}}) \right]. \quad (5)$$

At the treatment-unit level, utilizing $Y_j(\mathbf{Z}) = \sum_i Y_{ij}(\mathbf{Z})$, we define

$$\text{PTTE}_{\text{treatment}} = \frac{1}{|\mathcal{T}_{\text{prim}}|} \sum_{j \in \mathcal{T}_{\text{prim}}} \left(\mathbb{E}[Y_j(\mathbf{Z}^{(1)})] - \mathbb{E}[Y_j(\mathbf{Z}^{(0)})] \right), \quad (6)$$

i.e., the average total effect of full deployment on primary treatment units. Under our parametric representation $Y_j(E_j^{\text{Dir}}(\mathbf{Z}), E_j^{\text{Ind}}(\mathbf{Z}), X_j)$, (6) can be written as,

$$\text{PTTE}_{\text{treatment}} = \frac{1}{|\mathcal{T}_{\text{prim}}|} \sum_{j \in \mathcal{T}_{\text{prim}}} \mathbb{E} \left[\Psi \left(E_j^{\text{Dir}}(\mathbf{Z}^{(1)}), E_j^{\text{Ind}}(\mathbf{Z}^{(1)}), X_j \right) - \Psi \left(E_j^{\text{Dir}}(\mathbf{Z}^{(0)}), E_j^{\text{Ind}}(\mathbf{Z}^{(0)}), X_j \right) \right], \quad (7)$$

where it is easy to see that $E_j^{\text{Dir}}(\mathbf{Z}^{(0)})$ and $E_j^{\text{Ind}}(\mathbf{Z}^{(0)})$ are both equal to 0.

2.4 Formal Assumptions

Our framework relies on three key assumptions adapted from Doudchenko et al. (2020).

Assumption 2.1 (Exogenous Network). \mathbf{w} is not affected by the treatment assignment \mathbf{Z} .

This assumption requires that the connections between outcome and treatment units remain stable throughout the experiment. In our application, this means treatment does not alter which outcome units interact with which treated units.

Assumption 2.2 (Weak Unconfoundedness). For all exposure levels e : $\mathbb{I}[E_i = e] \perp Y_i(e)|\mathbf{w}_i$ where \mathbf{w}_i represents outcome unit i 's connection weights to all treatment-side units.

Under randomization of treated units, this assumption holds by design, as exposure is determined by the random treatment assignment and the fixed network structure.

Assumption 2.3 (Overlap). For all exposure levels e and weight vectors \mathbf{w} : $0 < \mathbb{P}(E_i = e|\mathbf{w}) < 1$.

This positivity condition ensures sufficient variation in exposure levels for identification. In practice, it requires that the treatment probability and network structure create adequate overlap in the exposure distribution.

The final assumption, already discussed at the beginning of this section, formalizes additivity of outcomes along the edges of the bipartite graph.

Assumption 2.4 (Linear additive edges). For all assignments \mathbf{Z} , unit-level outcomes can be written as

$$Y_i(\mathbf{Z}) = \sum_j Y_{ij}(\mathbf{Z}), \quad Y_j(\mathbf{Z}) = \sum_i Y_{ij}(\mathbf{Z}),$$

where $Y_{ij}(\mathbf{Z})$ is the potential outcome associated with edge (i, j) and $Y_{ij}(\mathbf{Z}) = 0$ if (i, j) is absent.

Assumption 2.4 holds exactly for additive metrics such as total rides, completed orders, or revenue, and for any linear reweighting of these (e.g., weighted revenue). It is not appropriate for inherently non-additive metrics (medians, quantiles, capped scores).

2.5 Projection Between Treatment and Outcome Effects

We formalize the link between the treatment- and outcome-level estimands when edge outcomes add linearly.

Recall that $\text{PTTE}_{\text{outcome}}$ is defined in (3) via the exposure E_i , which aggregates *only* primary-unit assignments, and $\text{PTTE}_{\text{treatment}}$ is given in (7).

Theorem 2.5 (Exact projection of PTTE under edge additivity). Suppose Assumption 2.4 holds. Then

$$\text{PTTE}_{\text{outcome}} = \frac{|\mathcal{T}_{\text{prim}}|}{|\mathcal{O}_{\text{prim}}|} \text{PTTE}_{\text{treatment}}, \quad (8)$$

where $\text{PTTE}_{\text{outcome}}$ is defined in (3) in terms of $Y_{i,\text{prim}}(\mathbf{Z})$ and $\text{PTTE}_{\text{treatment}}$ is defined in (7) in terms of $Y_j(\mathbf{Z})$.

Proof. By Assumption 2.4, for any assignment \mathbf{Z} and any $i \in \mathcal{O}_{\text{prim}}$,

$$Y_{i,\text{prim}}(\mathbf{Z}) = \sum_{j \in \mathcal{T}_{\text{prim}}} Y_{ij}(\mathbf{Z}), \quad Y_j(\mathbf{Z}) = \sum_i Y_{ij}(\mathbf{Z}).$$

Using the definition in (3) and linearity of expectation,

$$\begin{aligned} \text{PTTE}_{\text{outcome}} &= \frac{1}{|\mathcal{O}_{\text{prim}}|} \sum_{i \in \mathcal{O}_{\text{prim}}} \mathbb{E}[Y_{i,\text{prim}}(\mathbf{Z}^{(1)}) - Y_{i,\text{prim}}(\mathbf{Z}^{(0)})] \\ &= \frac{1}{|\mathcal{O}_{\text{prim}}|} \sum_{i \in \mathcal{O}_{\text{prim}}} \mathbb{E}\left[\sum_{j \in \mathcal{T}_{\text{prim}}} (Y_{ij}(\mathbf{Z}^{(1)}) - Y_{ij}(\mathbf{Z}^{(0)}))\right] \\ &= \frac{1}{|\mathcal{O}_{\text{prim}}|} \sum_{j \in \mathcal{T}_{\text{prim}}} \mathbb{E}\left[\sum_{i \in \mathcal{O}_{\text{prim}}} (Y_{ij}(\mathbf{Z}^{(1)}) - Y_{ij}(\mathbf{Z}^{(0)}))\right], \end{aligned}$$

where we interchange the (finite) sums over i and j .

Every edge (i, j) with $j \in \mathcal{T}_{\text{prim}}$ necessarily has $i \in \mathcal{O}_{\text{prim}}$, so

$$\sum_{i \in \mathcal{O}_{\text{prim}}} Y_{ij}(\mathbf{Z}) = \sum_i Y_{ij}(\mathbf{Z}) = Y_j(\mathbf{Z}).$$

Thus the inner sum equals $Y_j(\mathbf{Z}^{(1)}) - Y_j(\mathbf{Z}^{(0)})$, and

$$\text{PTTE}_{\text{outcome}} = \frac{1}{|\mathcal{O}_{\text{prim}}|} \sum_{j \in \mathcal{T}_{\text{prim}}} \mathbb{E}[Y_j(\mathbf{Z}^{(1)}) - Y_j(\mathbf{Z}^{(0)})].$$

By the definition of $\text{PTTE}_{\text{treatment}}$ in (7),

$$\text{PTTE}_{\text{treatment}} = \frac{1}{|\mathcal{T}_{\text{prim}}|} \sum_{j \in \mathcal{T}_{\text{prim}}} \mathbb{E}[Y_j(\mathbf{Z}^{(1)}) - Y_j(\mathbf{Z}^{(0)})],$$

so

$$\text{PTTE}_{\text{outcome}} = \frac{|\mathcal{T}_{\text{prim}}|}{|\mathcal{O}_{\text{prim}}|} \text{PTTE}_{\text{treatment}},$$

which is (8). \square

Remark 2.6. In applications where one suspects displacement onto ineligible units, Section 2.7 provides STTE estimands; comparing the projected quantity in (8) to outcome-level estimates is an empirical check.

Remark 2.7. If a decision metric applies known edge weights $\omega_{ij} \geq 0$, the same proof holds with Y_{ij} replaced by $\omega_{ij} Y_{ij}$ and normalizations based on total ω -mass.

Remark 2.8. For ratios or quantiles, a first-order linearization around the observed edge measure provides an approximate projection with a curvature remainder; see the discussion in Section 4.

2.6 Estimation Methods

Our estimation procedure utilizes the observed bipartite network structure to construct features (including exposure levels, number of exposed treatment-units, and propensity scores) which serve as inputs to machine learning models (linear, polynomial, non-parametric). These models estimate the outcome function Φ or Ψ from which we derive corresponding PTTE estimates by comparing predicted outcomes at full versus zero exposure levels.

More formally, our estimation proceeds in three steps:

Step 1: Feature Construction. For each unit, construct:

- *Exposure variables:* $E_i, E_j^{\text{Dir}}, E_j^{\text{Ind}}$ as defined above
- *Network features:* n_i^{prim} (number of connections), graph-based covariates
- *Propensity scores:* When neighbors are unweighted ($w_{ij} = 1/n_i^{\text{prim}}$), E_i takes values in $\{0, 1/n_i^{\text{prim}}, \dots, 1\}$. Writing $e = k/n_i^{\text{prim}}$ with $k \in \{0, \dots, n_i^{\text{prim}}\}$, the generalized propensity score has the binomial form

$$r(e, n_i^{\text{prim}}, p) = \binom{n_i^{\text{prim}}}{k} p^k (1-p)^{n_i^{\text{prim}}-k}. \quad (9)$$

For treatment-side units, we do not explicitly calculate the propensity score. Instead, we assume that the propensity score is captured through the non-parametric function of ψ .

Step 2: Outcome Function Estimation. Using the constructed features, estimate Φ or Ψ via machine learning:

- *Linear polynomial (LP):* Second-order polynomial in $(E_i, r, E_i \times r, E_i^2, r^2)$ for outcome-level, similar for treated-unit level (we only use LP in synthetic settings where there are no covariates)
- *Kernel Ridge Regression (KRR):* Non-parametric regression with Gaussian kernel, bandwidth selected via 5-fold cross-validation
- *Other methods:* Random forests, gradient boosting (details in online supplement)

Step 3: Counterfactual Prediction and Aggregation. For each unit, predict outcomes under full treatment ($E = 1$ or $E_j^{\text{Dir}}(\mathbf{Z}^{(1)})$, $E_j^{\text{Ind}}(\mathbf{Z}^{(1)})$) and no treatment ($E = 0$ or $E_j^{\text{Dir}}(\mathbf{Z}^{(0)})$, $E_j^{\text{Ind}}(\mathbf{Z}^{(0)})$). Average the differences to obtain PTTE estimates via Eq. (3) or (7).

Variance Estimation. We construct confidence intervals via bootstrap: sample treatment-side units with replacement, re-estimate on each bootstrap sample, and compute quantiles of the resulting PTTE distribution. For the projection approach, we project each bootstrap replicate and compute quantiles at the target level.

2.7 Extension to Secondary Effects

While primary units receive direct treatment, secondary units (\mathcal{T}_{sec}) experience spillover effects through shared connections with outcome units. These secondary total treatment effects (STTE) capture the indirect impact on ineligible units, an important component for understanding the full ecosystem effects of eligibility-constrained experiments.

2.7.1 Outcome-Unit Level STTE

At the outcome-unit level, we extend our framework to account for connections to both primary and secondary units. Recall that $\mathcal{O}_{\text{Both}}$ denotes outcome units connected to both primary and secondary treatment units. For $i \in \mathcal{O}_{\text{Both}}$, the secondary-edge component $Y_{i,\text{sec}}(\mathbf{Z})$ can change when we treat the primary units. We define the Secondary Total Treatment Effect (STTE) at the outcome level as

$$\text{STTE}_{\text{outcome}} = \frac{1}{|\mathcal{O}_{\text{Both}}|} \sum_{i \in \mathcal{O}_{\text{Both}}} \left(\mathbb{E}[Y_{i,\text{sec}}(\mathbf{Z}^{(1)})] - \mathbb{E}[Y_{i,\text{sec}}(\mathbf{Z}^{(0)})] \right). \quad (10)$$

Outcome units connected only to secondary treatment units are unaffected by the randomization among primary units and therefore do not contribute to $\text{STTE}_{\text{outcome}}$.

In estimation we model $Y_{i,\text{sec}}$ via a function Φ_{sec} as follows. For outcome units in $\mathcal{O}_{\text{Both}}$, let n_i^{sec} be the number of secondary treatment units connected to outcome unit i . We then assume:

$$Y_{i,\text{sec}}(\tilde{\mathbf{w}}, \mathbf{Z}, X_i) = \Phi_{\text{sec}}(n_i^{\text{prim}}, n_i^{\text{sec}}, E_i, r(E_i, \tilde{\mathbf{w}}, X_i)) + \epsilon_i \quad (11)$$

where $\tilde{\mathbf{w}}$ denotes the enriched graph structure including treatment unit classifications, i.e., primary and secondary. Under the same exposure mapping assumption as in Section 2.2, we can equivalently write (10) as

$$\text{STTE}_{\text{outcome}} = \frac{1}{|\mathcal{O}_{\text{Both}}|} \sum_{i \in \mathcal{O}_{\text{Both}}} \mathbb{E}[Y_{i,\text{sec}}(e_{i,\text{max}})] - \mathbb{E}[Y_{i,\text{sec}}(0)], \quad (12)$$

where $e_{i,\text{max}}$ denotes the exposure of unit i when all primary units are treated.

2.7.2 Treatment-Unit Level STTE

For a secondary treatment unit $j \in \mathcal{T}_{\text{sec}}$, we focus on the part of its outcome that flows through outcome units exposed to primary units. Define

$$Y_{j,\text{sec}}(\mathbf{Z}) := \sum_{i \in \mathcal{O}_{\text{Both}}} Y_{ij}(\mathbf{Z}).$$

Edges connecting j to outcome units that are never connected to any primary unit cannot be affected by the primary treatment assignment and are thus excluded from the STTE. We define the treatment-level STTE as

$$\text{STTE}_{\text{treatment}} = \frac{1}{|\mathcal{T}_{\text{sec}}|} \sum_{j \in \mathcal{T}_{\text{sec}}} (\mathbb{E}[Y_{j,\text{sec}}(\mathbf{Z}^{(1)})] - \mathbb{E}[Y_{j,\text{sec}}(\mathbf{Z}^{(0)})]). \quad (13)$$

In our implementation, $Y_{j,\text{sec}}(\mathbf{Z})$ is modeled as a function exposure and covariates. Specifically, under the representation,

$$Y_{j,\text{sec}} = \Psi_{\text{sec}}(E_j^{\text{Ind}}, X_j) + \epsilon_j, \quad (14)$$

where E_j^{Ind} follows a variant of Eq. (5) for secondary units, Eq. (15) can be written as:

$$\text{STTE}_{\text{treatment}} = \frac{1}{|\mathcal{T}_{\text{sec}}|} \sum_{j \in \mathcal{T}_{\text{sec}}} \mathbb{E} [\Psi_{\text{sec}}(E_j^{\text{Ind}}(\mathbf{Z}^{(1)}), X_j) - \Psi_{\text{sec}}(E_j^{\text{Ind}}(\mathbf{Z}^{(0)}), X_j)]. \quad (15)$$

2.7.3 Projection and Estimation

Under Assumption 2.4, the same counting argument as in Theorem 2.5 yields

$$\text{STTE}_{\text{outcome}} = \frac{|\mathcal{T}_{\text{sec}}|}{|\mathcal{O}_{\text{Both}}|} \text{STTE}_{\text{treatment}}. \quad (16)$$

This enables efficient computation at the treatment-unit level with projection to outcome-level effects, maintaining the computational advantages discussed for PTTE.

The ability to quantify STTE is important for bipartite experiments where ineligible units constitute a substantial portion of the ecosystem. In our ride-sharing example, understanding how visual prominence of economy vehicles affects premium and XL vehicles informs both immediate design decisions and long-term fleet composition strategies.

3 Validation on Simulated and Real Experimental Data

We validate our methodology through two complementary approaches. First, we conduct extensive simulations where ground truth effects are known, allowing us to assess both primary (PTTE) and secondary (STTE) treatment effects while evaluating the accuracy of our estimators and projection methods. Second, we apply our framework to two real experiments, focusing on PTTE estimation where business logic provides directional expectations for bias validation.

3.1 Simulation Results

3.1.1 Simulation Design

We simulate a bipartite system where outcome units (riders) interact with both primary and secondary treatment units (different vehicle types) through an app interface. The outcome function for rider i selecting vehicle type j follows:

$$f_{ij} = \alpha_j + \gamma_i \sum_{k \in \mathcal{E}_i} \beta_{jk} \log(v_k) + \epsilon_{ij} \quad (17)$$

where f_{ij} represents the selection outcome (e.g., booking probability or ride frequency), α_j is the baseline selection rate for vehicle type j , γ_i captures rider-specific interface responsiveness, β_{jk} represents the *attention spillover coefficient* between vehicle types j and k , v_k is the *visibility score* for vehicle type k (with baseline $v_k = 1.0$ for standard display), and \mathcal{E}_i is the set of vehicle types (primary and secondary) available to rider i .

The attention spillover coefficients β_{jk} encode how visual prominence affects selection patterns through cognitive attention mechanisms. The key patterns are:

- $\beta_{jj} < 0$: *Prominence saturation effect*, excessive visual emphasis on a vehicle type can trigger adverse user reactions, where riders perceive over-promotion as pushy or interpret it as a negative quality signal
- $\beta_{jk} > 0$ for substitutes: *Complementary positioning*, when economy vehicles are prominently featured, some premium vehicles benefit from appearing as the deliberate “upgrade” choice in contrast
- $\beta_{jk} < 0$ for competitors: *Competitive attention*, direct competition for limited user attention, where prominence of one vehicle type draws cognitive resources away from others

The logarithmic transformation captures *diminishing attention returns*: initial gains in visual prominence have larger effects than additional prominence, reflecting cognitive limits in processing visual stimuli and the natural saturation of attention capture.

Treatment consists of enhanced app placement for primary units (economy vehicles), implemented as a 10% increase in visibility score ($v_k = 1.1$). This could represent various interface changes: moving from position 3 to position 1 in the selection list, increasing badge size and animation intensity, or expanding the screen real estate allocated to these vehicles. Only primary units are eligible for this enhanced display treatment due to technical constraints or strategic considerations, while secondary units (premium and XL vehicles) maintain standard display formatting.

We vary the average number of primary units per rider (2.8 to 8.0, representing market density) and treatment probability (40-50%, representing rollout coverage), running 50 replications per configuration to ensure robust estimates. For each replication we compute ground-truth *PTTEs* and *STTEs* by evaluating the data generating process at $\mathbf{Z}^{(1)}$ and $\mathbf{Z}^{(0)}$.

3.1.2 Primary Treatment Effects (PTTE)

Table 1 presents PTTE estimates across five specifications. The Basic approach (a variant of difference-in-means) that ignores interferences consistently underestimates effects by 15-20%, as it fails to account for positive interferences that increase outcomes. Non-parametric methods (KRR) achieve estimates within 0.5% of ground truth across all specifications, while linear polynomial (LP) approaches degrade substantially as exposure complexity increases.

Figure 2 illustrates the distribution of estimates for Specification 1, top row of Table 1, demonstrating that treatment-level estimates projected to outcome level (Proj. KRR) match direct estimation while offering computational advantages, critical when outcome units outnumber treatment units by orders of magnitude. Similar figures for all specifications are provided in Appendix B

3.1.3 Secondary Treatment Effects (STTE)

Table 2 presents STTE estimates for secondary units. These effects, while smaller than PTTE, remain substantial, approximately 17% of primary effects in our simulations. Importantly, STTE would be exactly zero under the no-spillover assumption, highlighting the magnitude of bias when ignoring network effects.

Table 1. Outcome-Level PTTE Simulation Results: Median values across 50 replications

No	$\text{avg}(\mathcal{T}_i^{\text{prim}})$	% Primary Treated	Specification			Outcome Level			Treatment to Outcome Level	
			GT	Basic	LP	KRR	Proj. GT	Proj. KRR		
1	2.8	50%	0.5	0.42	0.51	0.50	0.5	0.5		
2	5.4	50%	0.97	0.81	0.92	0.97	0.97	0.97		
3	8.0	50%	1.46	1.21	0.98	1.46	1.46	1.46		
4	8.0	45%	1.46	1.19	0.79	1.46	1.46	1.46		
5	8.0	40%	1.46	1.16	0.58	1.46	1.46	1.46		

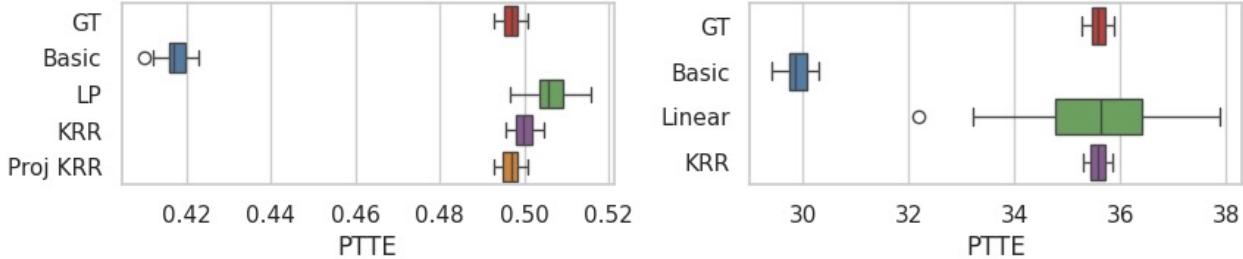


Figure 2. PTTE estimates for outcome units (left) and treatment units (right), showing ground truth (GT), Basic, linear polynomial (LP), Kernel Ridge Regression (KRR), and its projected variant (Proj. KRR, outcome level only).

For secondary effects, gradient boosting methods (XGBoost) perform well at the outcome level, while KRR excels at the treatment level. The projection approach again demonstrates remarkable accuracy, with projected estimates matching ground truth within 0.1%. Similar distributional plots as in Figure 2 are provided in Appendix B. These results validate our framework’s ability to capture the complete spillover ecosystem, quantifying both direct effects on eligible units and indirect effects on ineligible units that continue to operate in the bipartite setting.

3.1.4 Computational Benefits

The projection methodology offers substantial computational advantages. With typical ratios of outcome to treatment units exceeding 1000:1, estimation at the treatment level followed by projection reduces computation time from hours to minutes. For a system with 300,000 outcome units and 300 treatment units, we observe 1000x speedup while maintaining accuracy, needed for practical deployments.

3.2 Real Experiment Results

Focusing on PTTE, we validate the method on two real experiments, denoted by A and B, with three pre-specified metrics, M_1 , M_2 , and M_3 . For both experiments, business logic and economic theory imply that ignoring spillovers should bias M_1 upward (direction known ex ante), whereas no directional prior is available for M_2 or M_3 . Among the three, M_3 is the primary decision metric.

Table 2. STTE Simulation Results: Median values across 50 replications

No	$\text{avg}(\mathcal{T}_i^{\text{prim}})$	% Primary Treated	Specification			Outcome-Level		Treatment-Level	
			GT	XGBoost	Proj. KRR	GT	KRR	GT	KRR
1	2.8	50%	0.10	0.10	0.10	11.4	11.4		
2	5.4	50%	0.17	0.17	0.17	24.0	24.0		
3	8.0	50%	0.25	0.24	0.25	36.6	36.6		
4	8.0	45%	0.25	0.24	0.25	36.6	36.6		
5	8.0	40%	0.25	0.24	0.25	36.6	36.7		

Table 3. Comparison of Basic Method (ignoring spillovers) and Our Approach on Two Real Experiments. For each anonymized metric, we report treatment effect estimates (ATE for Basic, PTTE for our method), statistical significance at 5% level, and observed bias direction. Expected bias direction for M_1 follows from economic theory; M_3 is the primary decision metric. Sample sizes: $|\mathcal{T}_{\text{prim}}| \approx 7,000$ (Experiment A) and 8,000 (Experiment B).

Metric	Basic Method		Our Approach		Bias	
	ATE	Sig.	PTTE	Sig.	Basic – Ours (Observed)	Basic – Ground Truth (Expected)
Experiment A						
M_1	Pos.	Yes	Neg.	No	Pos.	Pos.
M_2	Pos.	No	Neg.	No	Pos.	—
M_3	Neg.	No	Pos.	Yes	Neg.	—
Experiment B						
M_1	Pos.	No	Neg.	No	Pos.	Pos.
M_2	Pos.	Yes	Pos.	No	Pos.	—
M_3	Pos.	No	Neg.	No	Pos.	—

From the results, presented in Table 3, two observations follow. First, in both A and B our approach moves M_1 in the expected direction (the difference, Basic–Ours, is positive), recovering the correct *direction of bias*. Second, for the decision metric, M_3 , ignorance of spillover effects yields an alternate launch conclusion for Experiment A from what is concluded using our method. In Experiment B, both methods yield the same “no effect” conclusion on M_3 (non-significant), leaving the decision unchanged. Overall, it is reassuring that for M_1 our method consistently identifies the correct direction of spillover bias across both experiments.

4 Conclusion

This paper studies eligibility-constrained bipartite experiments in which only a subset of treatment-side units can be randomized while interference propagates through all units. We formalize primary and secondary total treatment effects (PTTE, STTE), give identification under randomization within the eligible set, and propose flexible estimators that leverage generalized propensity scores and machine learning. Under a linear additive edges condition, we derive a projection linking treatment- and outcome-level estimands, enabling estimation on the typically smaller set of treatment units with aggregation to outcomes. In simulations and two case studies, accounting for interference yields effect estimates that differ materially from analyses that ignore spillovers, including a sign difference for one primary decision metric. These findings illustrate that, in settings with interaction across unit types, effect definitions and estimators that target the total impact at rollout can lead to different conclusions than conventional A/B analyses.

Scope and limitations. Our framework is useful when (i) the bipartite structure is substantively meaningful, (ii) randomization is feasible within an eligible subset of treatment units, and (iii) one wishes to estimate the effect of full deployment to the eligible set. Several limitations should be highlighted.

- *Network observability.* Identification and estimation rely on access to the interaction structure (e.g., weights w_{ij} or close proxies). When the network is unavailable, only coarsely measured, or measured with substantial error, the proposed methods would not be applicable or may be biased. Designs and estimators for unknown networks exist (e.g., Yu et al., 2022; Cortez et al., 2022; Shirani and Bayati, 2024; Shirani et al., 2025) but are not developed here.
- *Exogenous network.* Assumption 2.1 requires that the network is unaffected by treatment. If assignment changes matching, congestion, or availability in ways that alter edges, the estimands here are not identified without additional modeling of joint assignment–network dynamics.
- *Unconfoundedness and overlap.* With randomization inside the eligible set, Assumption 2.2 holds by design, but positivity (Assumption 2.3) can fail in sparse or highly skewed graphs. In such cases, models extrapolate to exposure regions with little support, increasing estimator sensitivity.

- *Projection validity.* The projection between treatment and outcome levels requires linear additivity of edge-level outcomes (Assumption 2.4). When the target metric is non-additive (e.g., medians, capped rates, or composite satisfaction scores), the projection is invalid and estimation should remain at the metric’s native level.
- *Finite-sample and algorithmic considerations.* The ensemble estimators depend on tuning choices and on the accuracy of generalized propensity features. Inference is implemented via bootstrap resampling; while practical, its coverage under complex dependence structures may deviate from nominal levels in small samples.
- *External validity.* Empirical validation is limited to two experiments in a particular context. The sign and magnitude of spillovers are design- and environment-specific; conclusions need not transport to other systems, policies, or time periods.

Practical implications. When eligibility constraints coexist with cross-side interactions, analysts seeking the total effect of rollout should (i) collect and audit the interaction data used to form exposure, (ii) assess overlap in exposure distributions prior to modeling, (iii) report estimates at the natural measurement level of each metric and invoke projection only when Assumption 2.4 is defensible, and (iv) accompany point estimates with diagnostic sensitivity analyses to plausible network misspecification. In settings where the network cannot be measured or Assumption 2.1 is doubtful, alternative designs that reduce interference (e.g., graph/cluster randomization) may be preferable (Ugander et al., 2013; Eckles et al., 2017).

Future directions. Promising extensions include: inference with partially observed or mismeasured networks (including formal sensitivity analyses); experimental designs that jointly learn the interaction structure while estimating PTTE/STTE; dynamic or sequential treatments where networks evolve; heterogeneity of total effects across network positions; and methods for non-additive outcomes. Finally, formal results on efficiency and inference under general dependence structures would help clarify the conditions under which the computational advantages of treatment-level estimation translate to statistical gains.

Overall, the contribution is conceptual and methodological: we articulate estimands aligned with deployment in eligibility-constrained, bipartite settings and provide tools to estimate them under explicit assumptions. Equally important, we discuss the conditions under which these tools should be used with extra caution.

References

- Anish Agarwal, Sarah H. Cen, Devavrat Shah, and Christina Lee Yu. Network synthetic interventions: A causal framework for panel data under network interference, 2023. URL <https://arxiv.org/abs/2210.11355>.
- Peter M. Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, 11(4):1912–1947, 2017. doi: 10.1214/16-AOAS1005.
- S. Athey and G.W. Imbens. The econometrics of randomized experiments. In Abhijit Vinayak Banerjee and Esther Duflo, editors, *Handbook of Field Experiments*, volume 1 of *Handbook of Economic Field Experiments*, pages 73–140. North-Holland, 2017. doi: <https://doi.org/10.1016/bs.hefe.2016.10.003>. URL <https://www.sciencedirect.com/science/article/pii/S2214658X16300174>.
- Susan Athey, Dean Eckles, and Guido W. Imbens. Exact p-values for network interference. *Journal of the American Statistical Association*, 113(521):230–240, 2018. doi: 10.1080/01621459.2016.1241178.
- Sarah Baird, J. Aislinn Bohren, Craig McIntosh, and Berk Özler. Optimal design of experiments in the presence of interference. *Review of Economics and Statistics*, 100(5):844–860, 2018. doi: 10.1162/rest_a_00716.

Patrick Bajari, Brian Burdick, Guido W. Imbens, Lorenzo Masoero, James McQueen, Thomas S. Richardson, and Ido M. Rosen. Experimental Design in Marketplaces. *Statistical Science*, 38(3):458 – 476, 2023. doi: 10.1214/23-STS883. URL <https://doi.org/10.1214/23-STS883>.

Guillaume W. Basse and Avi Feller. Analyzing two-stage experiments in the presence of interference. *Journal of the American Statistical Association*, 113(521):41–55, 2018. doi: 10.1080/01621459.2017.1319839.

Guillaume W. Basse, Avi Feller, and Panos Toulis. Randomization tests of causal effects under interference. *Biometrika*, 106(2):487–494, 2019. doi: 10.1093/biomet/asy072.

Jade Benjamin-Chung, Benjamin F. Arnold, David Berger, Stephen P. Luby, Edward Miguel, Jr. Collford, John M., and Alan E. Hubbard. Spillover effects in epidemiology: parameters, study designs and methodological considerations. *International Journal of Epidemiology*, 47(1):332–347, 2018. doi: 10.1093/ije/dyx201.

Jennifer Brennan, Vahab Mirrokni, and Jean Pouget-Abadie. Cluster randomized designs for one-sided bipartite experiments. In *Advances in Neural Information Processing Systems*, volume 35, pages 37962–37974, 2022.

Kevin L. Chen, Falco J. Bargagli-Stoffi, Raphael C. Kim, Lucas R. F. Henneman, and Rachel C. Nethery. Difference-in-differences under bipartite network interference: A framework for quasi-experimental assessment of the effects of environmental policies on health, 2024. URL <https://arxiv.org/abs/2404.13442>.

David S. Choi. Estimation of monotone treatment effects in network experiments. *Journal of the American Statistical Association*, 112(519):1147–1155, 2017. doi: 10.1080/01621459.2016.1194845.

Mayleen Cortez, Matthew Eichhorn, and Christina Yu. Staggered rollout designs enable causal inference under interference without network knowledge. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 7437–7449. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/3103b25853719847502559bf67eb4037-Paper-Conference.pdf.

D. R. Cox. *Planning of Experiments*. Wiley, New York,, 1958.

Nick Doudchenko, Minzhengxiong Zhang, Evgeni Drynkin, Edoardo Airoldi, Vahab Mirrokni, and Jean Pouget-Abadie. Causal inference with bipartite designs. *arXiv preprint arXiv:2010.02108*, 2020.

Dean Eckles, Brian Karrer, and Johan Ugander. Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, 5(1):20150021, 2017. doi: 10.1515/jci-2015-0021.

Naoki Egami. Spillover effects in the presence of unobserved networks. *Political Analysis*, 29(3):287–316, 2021. doi: 10.1017/pan.2020.28.

Laura Forastiere, Edoardo M. Airoldi, and Fabrizia Mealli. Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, 116(534):901–918, 2021. doi: 10.1080/01621459.2020.1768100.

M. Elizabeth Halloran and Claudio J. Struchiner. Causal inference in infectious diseases. *Epidemiology*, 6(2):142–151, 1995. doi: 10.1097/00001648-199503000-00010.

Christopher Harshaw, Fredrik Sävje, David Eisenstat, Vahab Mirrokni, and Jean Pouget-Abadie. Design and analysis of bipartite experiments under a linear exposure-response model. *Electronic Journal of Statistics*, 17(1):464–518, 2023. doi: 10.1214/23-EJS2111.

Keisuke Hirano and Guido W. Imbens. *The Propensity Score with Continuous Treatments*, chapter 7, pages 73–84. John Wiley & Sons, Ltd, 2004. ISBN 9780470090459. doi: <https://doi.org/10.1002/0470090456.ch7>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/0470090456.ch7>.

D Holtz, R Lobel, I Liskovich, and S Aral. Reducing interference bias in online marketplace pricing experiments. *arXiv preprint arXiv:2004.12489*, 2020.

- Yingjie Hu, Shuangning Li, and Stefan Wager. Average direct and indirect causal effects under interference. *Biometrika*, 109(4):1165–1172, 2022. doi: 10.1093/biomet/asac008.
- Michael G Hudgens and M Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008. doi: 10.1198/016214508000000292.
- Kosuke Imai, Zhichao Jiang, and Anup Malani. Causal inference with interference and noncompliance in two-stage randomized experiments. *Journal of the American Statistical Association*, 116(534):632–644, 2021. doi: 10.1080/01621459.2020.1765787.
- Ramesh Jagadeesan, Natesh S. Pillai, and Alexander Volfovsky. Designs for estimating the treatment effect in networks with interference. *The Annals of Statistics*, 48(2):679–712, 2020. doi: 10.1214/19-AOS1810.
- Ramesh Johari, Hannah Li, Inessa Liskovich, and Gabriel Y. Weintraub. Experimental design in two-sided platforms: An analysis of bias. *Management Science*, 68(10):7069–7089, 2022.
- Hyunseung Kang and Guido W. Imbens. Peer encouragement designs in causal inference with partial interference and identification of local average network effects, 2016.
- Luke Keele and Hyunseung Kang. An introduction to spillover effects in cluster randomized trials with noncompliance. *Clinical Trials*, 19(4):375–379, 2022. doi: 10.1177/17407745221093580.
- Michael P. Leung. Causal inference under approximate neighborhood interference. *Econometrica*, 90(1): 267–293, 2022. doi: 10.3982/ECTA17841.
- Evan Munro, Stefan Wager, and Kuang Xu. Treatment effects in market equilibrium. *arXiv preprint arXiv:2109.11647*, 2021.
- Elizabeth L. Ogburn and Tyler J. VanderWeele. Causal diagrams for interference. *Statistical Science*, 29(4): 559–578, 2014. doi: 10.1214/14-STS501.
- Jean Pouget-Abadie, Kevin Aydin, Warren Schudy, Kay Brodersen, and Vahab S. Mirrokni. Variance reduction in bipartite experiments through correlation clustering. In *Advances in Neural Information Processing Systems*, volume 32, 2019.
- David Puelz, Guillaume W. Basse, Avi Feller, and Panos Toulis. A graph-theoretic approach to randomization tests of causal effects under general interference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84(1):174–204, 2022. doi: 10.1111/rssb.12474.
- Paul R Rosenbaum. Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477):191–200, 2007. doi: 10.1198/016214506000001112. URL <https://doi.org/10.1198/016214506000001112>.
- Fredrik Sävje, Peter M. Aronow, and Michael G. Hudgens. Average treatment effects in the presence of unknown interference. *The Annals of Statistics*, 49(2):673–701, 2021. doi: 10.1214/20-AOS1973.
- Liang Shi, Edvard Bakhitov, Kenneth Hung, Brian Karrer, Charlie Walker, Monica Bhole, and Okke Schrijvers. Scalable analysis of bipartite experiments, 2024. URL <https://arxiv.org/abs/2402.11070>.
- Sadegh Shirani and Mohsen Bayati. Causal message-passing for experiments with unknown and general network interference. *Proceedings of the National Academy of Sciences*, 121(40):e2322232121, 2024.
- Sadegh Shirani, Yuwei Luo, William Overman, Ruoxuan Xiong, and Mohsen Bayati. Can We Validate Counterfactual Estimations in the Presence of General Network Interference? *arXiv e-prints*, art. arXiv:2502.01106, February 2025. doi: 10.48550/arXiv.2502.01106.
- Michael E Sobel. What do randomized studies of housing mobility demonstrate? *Journal of the American Statistical Association*, 101(476):1398–1407, 2006. doi: 10.1198/016214506000000636.
- Daniel L. Sussman and Edoardo M. Airoldi. Elements of estimation theory for causal effects in the presence of network interference, 2017. URL <https://arxiv.org/abs/1702.03578>.

Eric J. Tchetgen Tchetgen and Tyler J. VanderWeele. On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21(1):55–75, 2012. doi: 10.1177/0962280210386779.

Eric J. Tchetgen Tchetgen, Irene R. Fulcher, and Ilya Shpitser. Auto-g-computation of causal effects on a network. *Journal of the American Statistical Association*, 116(534):833–844, 2021. doi: 10.1080/01621459.2020.1773839.

Johan Ugander, Brian Karrer, Lars Backstrom, and Jon Kleinberg. Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 329–337. ACM, 2013. doi: 10.1145/2487575.2487695.

Tyler J. VanderWeele and Weihua An. Social networks and causal inference. In Stephen L. Morgan, editor, *Handbook of Causal Analysis for Social Research*, pages 353–374. Springer, Dordrecht, 2013. doi: 10.1007/978-94-007-6094-3_17.

Christina Lee Yu, Edoardo M. Airoldi, Christian Borgs, and Jennifer T. Chayes. Estimating the total treatment effect in randomized experiments with unknown network structure. *Proceedings of the National Academy of Sciences*, 119(44), 2022. ISSN 1091-6490.

Christopher M. Zigler and Georgia Papadogeorgou. Bipartite causal inference with interference. *Statistical Science*, 36(1):109–123, 2021. doi: 10.1214/19-STS749.

A Key Notation

This section provides a summary of the key mathematical notation and variable definitions in Table 4.

Table 4. Key notation for eligibility-constrained bipartite experiments

Symbol	Definition
Units and Sets	
$\mathcal{T}_{\text{prim}}, \mathcal{T}_{\text{sec}}$	Primary (eligible) and secondary (ineligible) treatment units
$\mathcal{O}_{\text{prim}}$	Outcome units connected to ≥ 1 primary unit
$\mathcal{O}_{\text{Both}}$	Outcome units connected to both primary & secondary treatment units
$\mathcal{T}_i^{\text{prim}}$	Set of primary units connected to outcome unit i
\mathcal{O}_j	Set of outcome units connected to treatment unit j
Treatment and Assignment	
\mathbf{Z}	Vector of treatment assignments (size is the total number of treatment-side units)
Z_j	Treatment assignment indicator for treatment-side unit j
p	Treatment probability
Network Structure	
\mathbf{w}	$n_{\text{outcome}} \times n_{\text{treatment}}$ matrix of connection weights
w_{ij}	Connection weight between outcome unit i and treatment unit j
$\tilde{\mathbf{w}}$	Enriched graph structure including unit classifications
Exposure Variables	
E_i	Treatment exposure of outcome unit i given assignment \mathbf{Z}
$E_j^{\text{Dir}}, E_j^{\text{Ind}}$	Direct and indirect exposure of treatment unit j
$e_{i,\max}$	Maximum exposure when all primary units are treated
n_i^{prim}	Number of primary units connected to outcome unit i
n_i^{sec}	Number of secondary units connected to outcome unit i
Outcome Variables	
Y_i	Outcome for outcome unit i
Y_j	Outcome for treatment unit j
Y_{ij}	Edge-level outcome between units i and j
$Y_{i,\text{sec}}$	Outcomes attributable to secondary units
$Y_{j,\text{sec}}$	Secondary treatment unit outcome
Functions and Parameters	
Φ, Ψ	Outcome functions for outcome and treatment units
$r(E_i, \mathbf{w}, X_i)$	Generalized propensity score: $\mathbb{P}(E_i \mathbf{w}, X_i)$
X_i, X_j	Covariates for outcome and treatment units
ϵ_i	Error term
Estimands	
$\text{PTTE}_{\text{outcome}}, \text{STTE}_{\text{outcome}}$	Primary and secondary total treatment effect at outcome-side
$\text{PTTE}_{\text{treatment}}, \text{STTE}_{\text{treatment}}$	Primary and secondary total treatment effect at treatment-side
Mathematical Operators	
$\mathbb{E}[\cdot]$	Expectation operator
$\mathbb{P}(\cdot)$	Probability operator
$\mathbb{I}(\cdot)$	Indicator function (1 if true, 0 otherwise)

B Additional Simulation Results

Figures 3-4 show the boxplot of the ground truth and estimates of PTTE for outcome-level and treatment unit-level, respectively, for specifications 1-5. Similar quantities for STTE are presented in Figures 5-6.

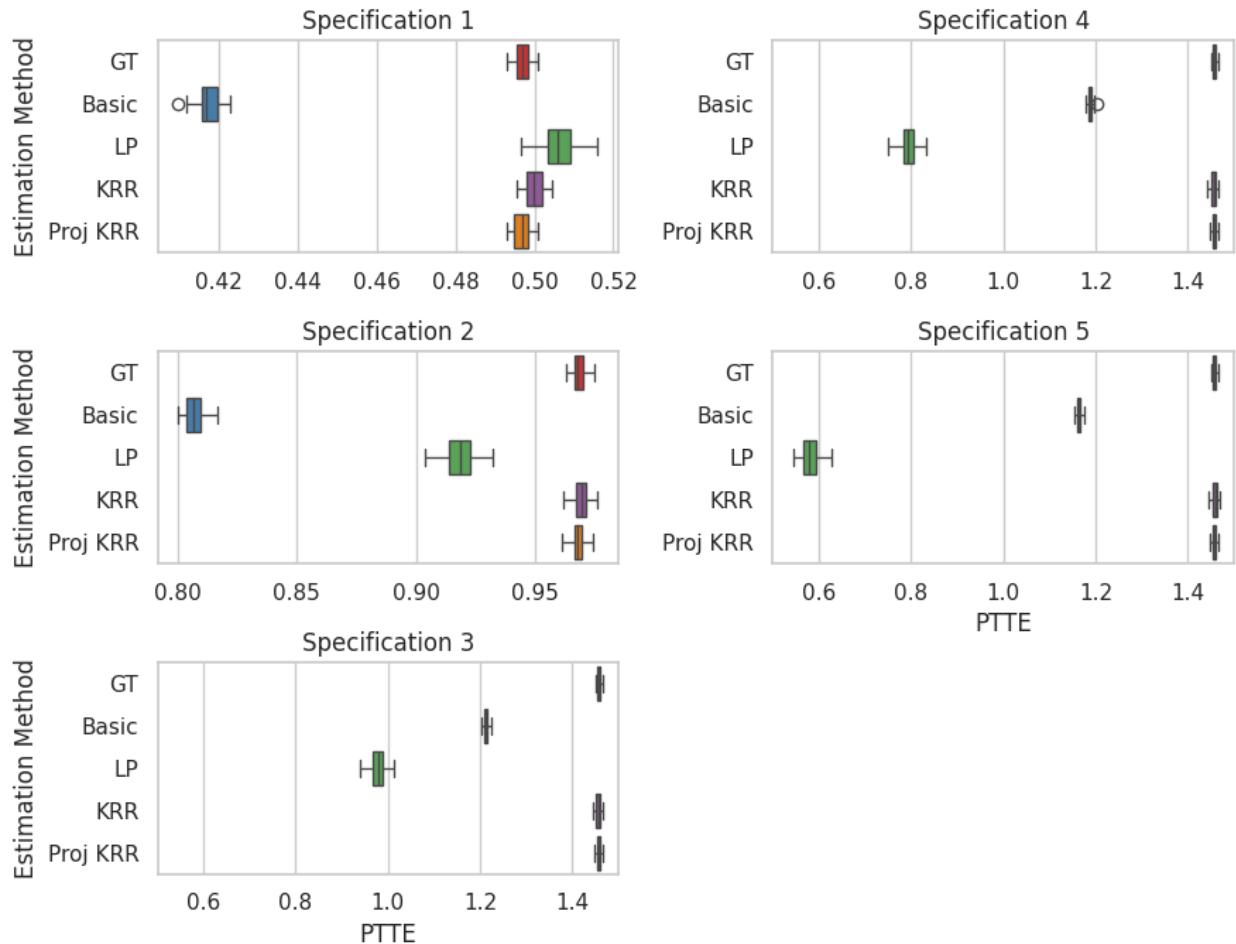


Figure 3. Box plot of PTTE at the outcome unit granularity

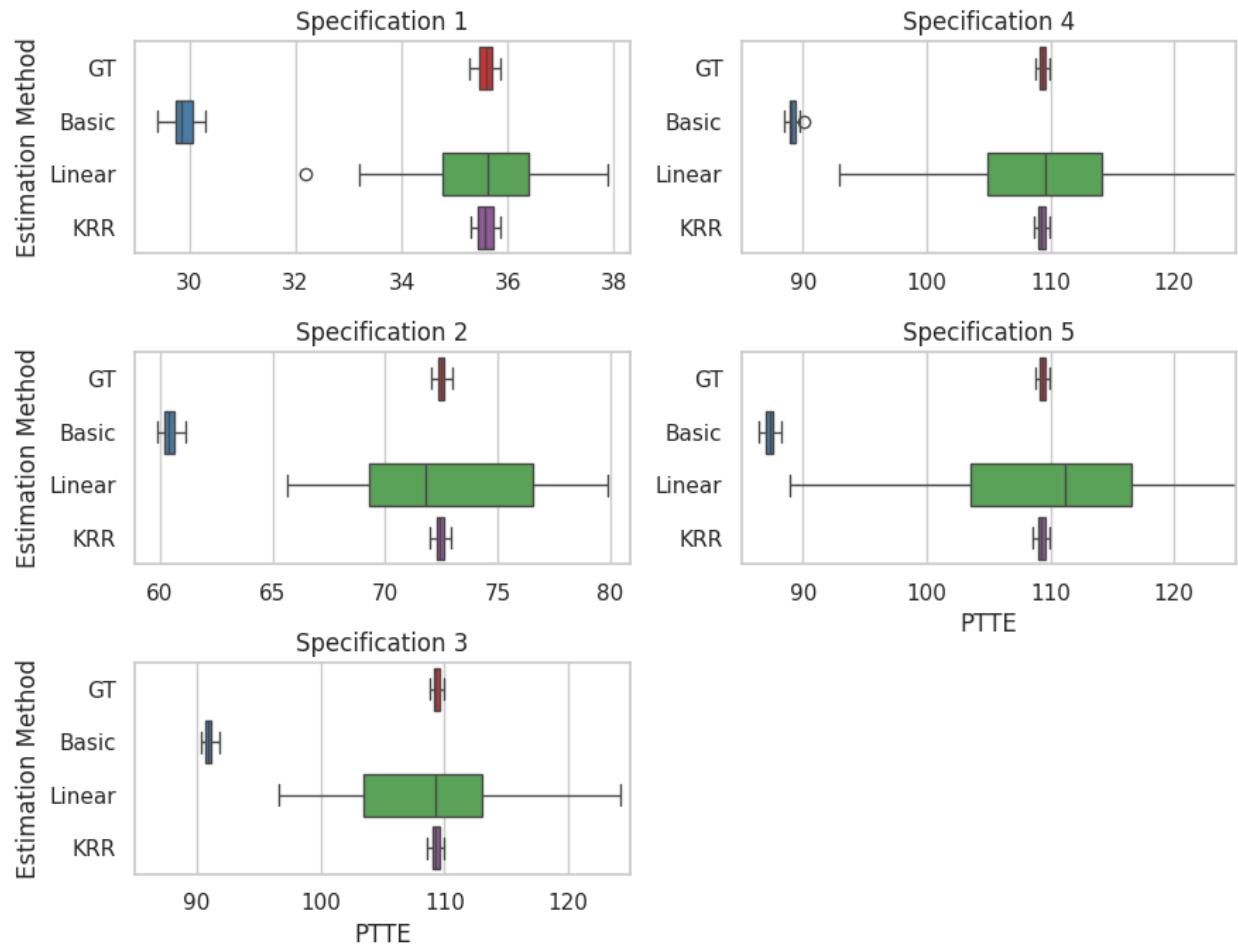


Figure 4. Box plot of PTTE at the treatment unit granularity

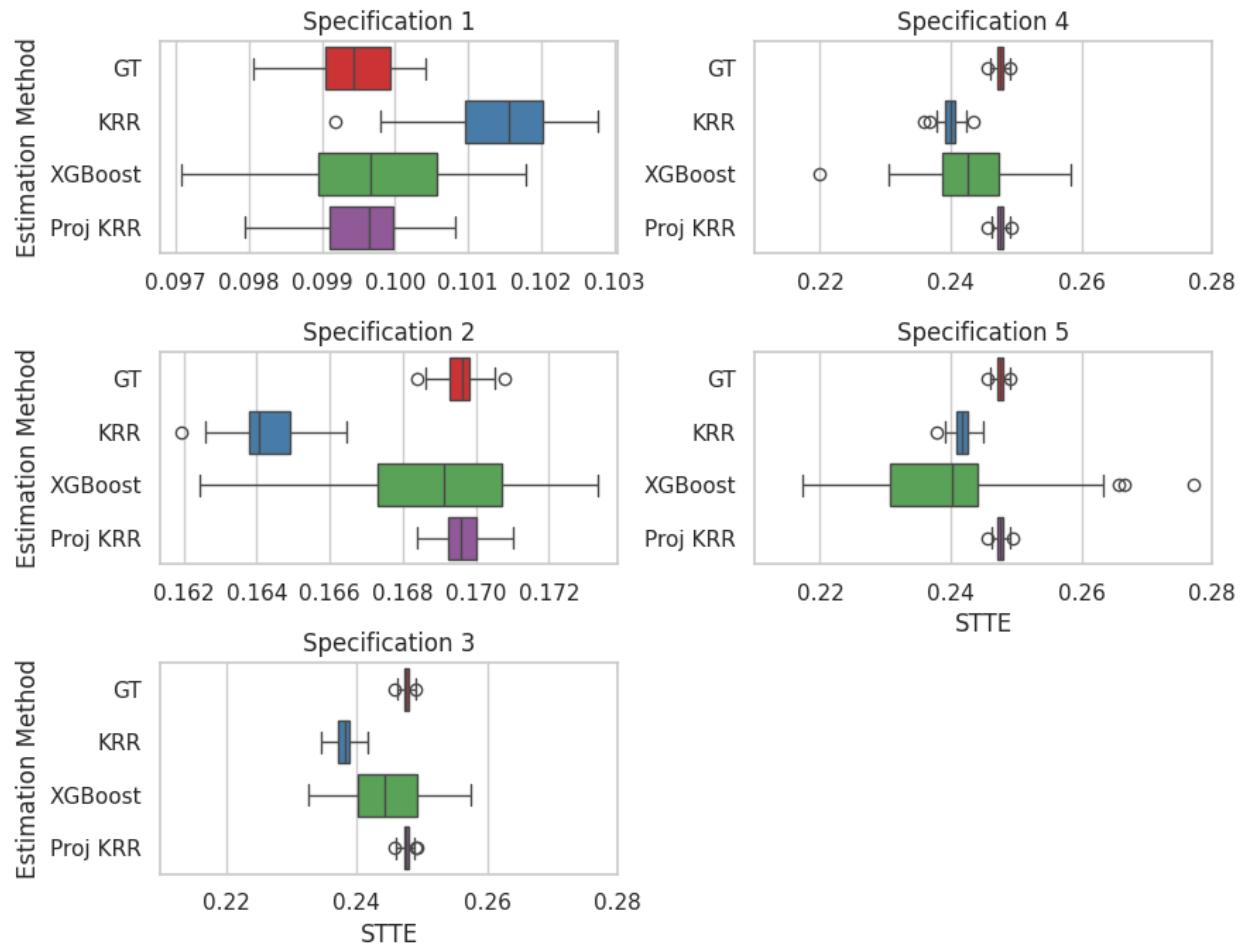


Figure 5. Box plot of STTE at the outcome unit granularity

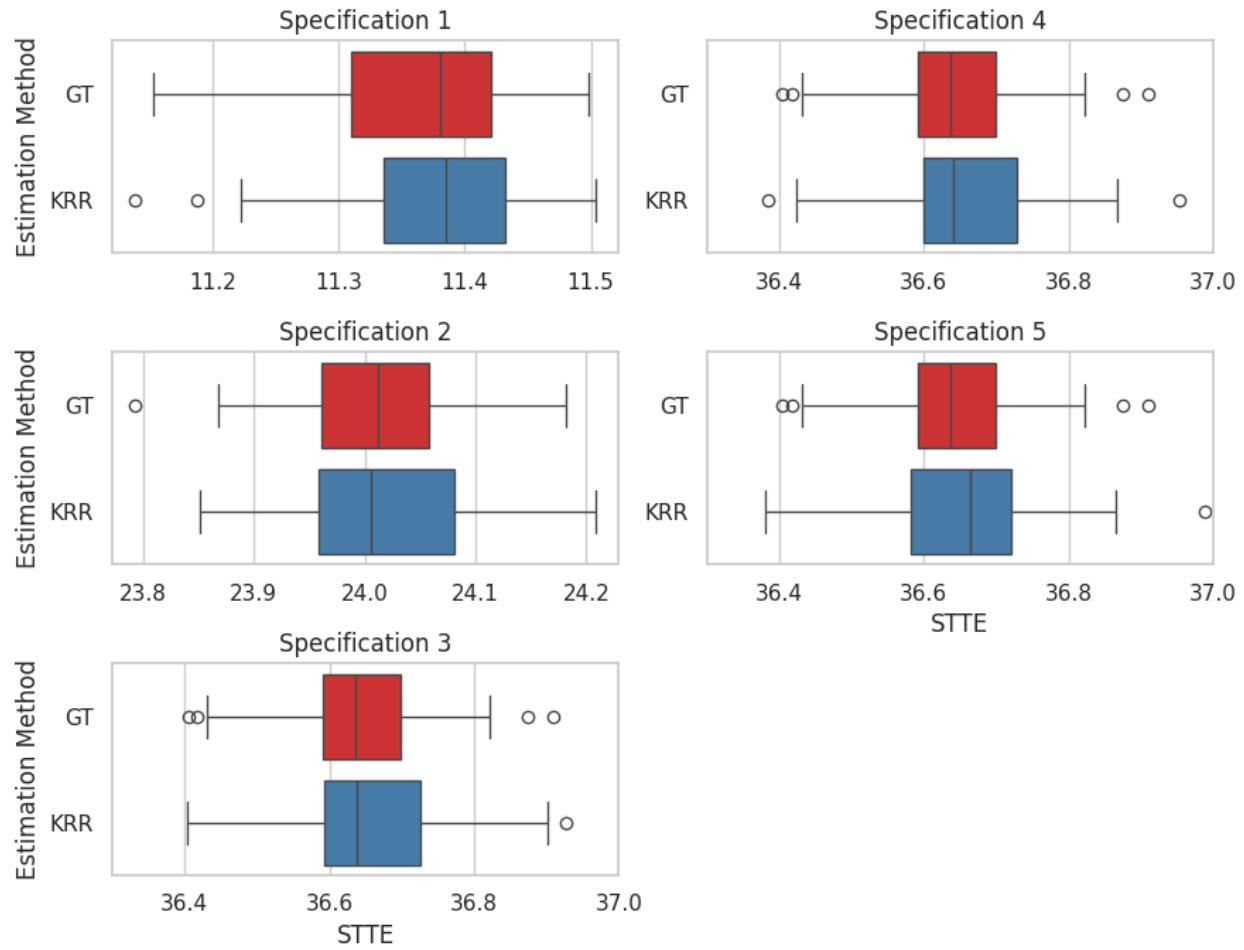


Figure 6. Box plot of STTE at the treatment unit granularity