Estimating within-cluster and between-cluster spillover effects in randomized saturation designs

Sizhu Lu*[1], Lei Shi*[2], and Peng Ding[1]

[1]Department of Statistics, University of California, Berkeley

[2]Division of Biostatistics, University of California, Berkeley

November 16, 2025

Abstract

Randomized saturation designs are two-stage experiments: they first randomly assign treatment probabilities over the clusters and then randomly assign the treatment to the units within the clusters. The existing literature on randomized saturation designs focuses on estimating within-cluster spillover effects by assuming away between-cluster spillover effects. However, the units may interact across clusters in many practical randomized saturation designs. A leading example is that some units are geographically close to each other, so spillover effects arise across clusters. Based on the potential outcomes framework, we formulate the causal inference problem of estimating within-cluster and between-cluster spillover effects in randomized saturation designs. We clarify the causal estimands and establish the statistical theory for estimation and inference. We also apply our method to analyze a recent randomized saturation design of cash transfer on household expenditure in Kenya.

Key Words: direct effect; exposure mapping; indirect effect; interference; two-stage experiment

1 Introduction

Randomized saturation designs have become increasingly common across disciplines for studying spillovers and interference, with applications spanning economics (e.g., Crépon et al. 2013; Baird et al. 2018; Egger et al. 2022), public health (e.g., Melis et al. 2005; Benjamin-Chung et al. 2018), and political science (e.g., Sinclair et al. (2012)). Most existing studies consider only within-cluster interference and assume there is no interference between clusters (Hudgens and Halloran, 2008; Basse and Feller, 2018; Jiang et al., 2023).

However, in some real-world settings, this assumption may not hold. A leading motivating example is the study by Egger et al. (2022), which implemented a randomized saturation design to evaluate the economic impacts of a large-scale cash transfer program in rural Kenya between 2014 and 2017. The study area consists

*The first two authors contributed equally to this work.

of 653 villages nested within 155 sublocations across two counties. Sublocations are administrative units, and villages within the same sublocation often share common markets, social ties, and economic connections.

The randomization proceeded in two stages following a randomized saturation design. In the first stage, sublocations were randomly assigned to high- or low-saturation groups. In the second stage, within each high-saturation sublocation, two-thirds of villages were randomly assigned to treatment, while in low-saturation sublocations, one-third of villages were treated. All eligible households in treated villages received transfers.

In this design, there are two possible types of interference: (i) within-sublocation interference, where the outcome of a village may be affected by the treatment status of other villages in the same sublocation, and (ii) between-sublocation interference, where the outcome of a village may also be affected by the treatment of villages from different sublocations that are geographically close.

To make the ideas precise, we start by introducing some notation. Villages are grouped into administrative sublocations, and let $k(i)$ denote the sublocation to which village i belongs. Some villages are close to one another, and we use $\mathcal{G}_i$ to represent the set of villages that are geographically close (for example, within a certain distance) to village i. Let $A_i \in \{0,1\}$ denote the binary treatment of each village, with $A_i = 1$ if the village i was assigned a cash transfer, and $A_i = 0$ otherwise.

When there is interference, a village's outcome can depend not only on its own treatment but also on the treatments received by other villages. To describe this dependence, we use the concept of an exposure mapping (Aronow and Samii, 2017). In our setting, an exposure mapping summarizes how the treatments received by other villages, together with village i's own treatment, combine to determine how much exposure village i experiences that affects its outcome. Specifically, we define an exposure mapping $d_i(\boldsymbol{A}) = (A_i, S_i, H_i)$, where A_i is the village's own treatment, S_i summarizes the treatment status of other villages in the same sublocation as village i, capturing within-sublocation exposure, and H_i summarizes the treatment status of nearby villages located in different sub locations, capturing between-sublocation exposure. For each village i, we can write S_i as $S_i = f_i\{A_j : k(j) = k(i)\}$, which is a function of the treatment assignments of other villages within the same sublocation. The mapping f_i is allowed to vary across units. Similarly, we can write H_i as $H_i = g_i\{A_j : k(j) \neq k(i),\ j \in \mathcal{G}_i\}$, which is a function of the treatment assignments of nearby villages outside i's sublocation. The mapping g_i may also vary across units.

As an example, consider the following exposure mapping:

$$
\begin{aligned}
S_i &= \mathbb{1}\left\{\frac{\sum_{j\neq i}\mathbb{1}\{k(j)=k(i)\}A_j}{\sum_{j\neq i}\mathbb{1}\{k(j)=k(i)\}} > \frac{1}{2}\right\}, \\
H_i &= \mathbb{1}\left\{\frac{\sum_{j=1}^{n}\mathbb{1}\{k(j)\neq k(i), j\in\mathcal{G}_i\}A_j}{\sum_{j=1}^{n}\mathbb{1}\{k(j)\neq k(i), j\in\mathcal{G}_i\}} > \frac{1}{2}\right\}.
\end{aligned}
$$

Here, S_i indicates whether more than half of the other villages in the same sublocation as village i are treated, and H_i indicates whether more than half of the nearby villages in different sublocations are treated.

Related work. Most methodological studies on randomized saturation designs focus on within-cluster interference. The only recent work that explicitly allows for interference between clusters is Leung (2025), who studies cluster-randomized trials with cross-cluster interference. Leung (2025) primarily focuses on

improving estimators for direct effects and indirect effects related to treatment saturation to reduce bias when interference extends beyond cluster boundaries. In contrast, our study directly models and estimates the between-cluster spillover effects themselves, providing both identification results and estimators that explicitly account for cross-cluster dependence.

Another related line of work is spatial interference, where researchers study interference that arises through geographic proximity or distance-based exposure (Papadogeorgou et al., 2022; Giffin et al., 2023; Wang et al., 2025). In contrast, we also consider administrative-level interference, where within-sublocation interference may occur even when two villages are not geographically close but belong to the same administrative unit.

Organization of the paper The paper proceeds as follows. Section 2 introduces a set of causal estimands that capture both the direct effects of the treatment and the indirect effects arising from administrative and geographic interference. Section 3 presents point estimators for these causal estimands. Section 4 provides their theoretical properties, including consistency, asymptotic normality, and expressions for the asymptotic variance. Section 5 provides variance estimators based on these theoretical results and shows that they lead to asymptotically valid confidence intervals. Section 6 implements the proposed estimators and inference methods in the cash transfer study of Egger et al. (2022). Section 7 concludes with a discussion of future research directions.

2 Causal estimands of interest

2.1 Conditional causal effects

Conditional direct effects. We first consider the direct effect of the treatment, holding (S, H) at a fixed level (s, h). Define

$$\text{DE}(s,h) \quad = \quad n^{-1}\sum_{i=1}^{n} Y_i(1,s,h) - n^{-1}\sum_{i=1}^{n} Y_i(0,s,h).$$

The quantity $\text{DE}(s,h)$ represents the *conditional direct effect* of treatment A_i on the outcome, while holding the exposure variables (S_i, H_i) fixed at values (s,h).

Conditional indirect effects. We consider two types of indirect effects corresponding to the two sources of interference: within-cluster and between-cluster spillover effects.

For within-cluster indirect effects, we define the *within-cluster conditional indirect effect* of S_i on the outcome as

$$\text{WIE}(s,s',h) \quad = \quad n^{-1}\sum_{i=1}^{n} Y_i(0,s,h) - n^{-1}\sum_{i=1}^{n} Y_i(0,s',h).$$

The quantity $\text{WIE}(s,s',h)$ captures the effect of changing the proportion of treated villages in the same

sublocation (from s to s') on the outcome of village i, while holding its own treatment fixed at zero and the between-cluster exposure H_i at level h.

For between-cluster indirect effects, similarly, we define the *between-cluster conditional indirect effect* of H_i on the outcome as

$$\text{BIE}(s,h,h') \quad = \quad n^{-1}\sum_{i=1}^{n} Y_i(0,s,h) - n^{-1}\sum_{i=1}^{n} Y_i(0,s,h').$$

The quantity $\text{BIE}(s,h,h')$ captures the effect of changing the level of between-cluster exposure from h to h', with the own treatment fixed at zero and the within-cluster exposure S_i fixed at level s.

We can view this setup as a 2^3 factorial experiment defined by the three factors (A_i, S_i, H_i). Although many other causal contrasts can be defined in this framework, we focus on these three effects, which are most relevant to our empirical motivation. See Zhao and Ding (2022b) for a more general discussion on factorial experiments.

2.2 In-policy causal effects

In practice, we can report the aggregated version of the causal estimands by marginalizing over the implemented policy distributions.

Marginal direct effect. We first consider the direct effect of the treatment, averaging over the conditional distribution of (S_i, H_i). Define

$$\begin{aligned}
\text{DE} \quad &= \quad n^{-1}\sum_{i=1}^{n} E_{\boldsymbol{A}_{(-i)}|A_i=1}\left\{Y_i(1,S_i,H_i)\right\} - n^{-1}\sum_{i=1}^{n} E_{\boldsymbol{A}_{(-i)}|A_i=0}\left\{Y_i(0,S_i,H_i)\right\} \qquad (1)\\
&= \quad n^{-1}\sum_{i=1}^{n}\sum_{s=0,1}\sum_{h=0,1} \text{pr}(S_i=s, H_i=h \mid A_i=1) Y_i(1,s,h)\\
&\quad\quad -n^{-1}\sum_{i=1}^{n}\sum_{s=0,1}\sum_{h=0,1} \text{pr}(S_i=s, H_i=h \mid A_i=0) Y_i(0,s,h),
\end{aligned}$$

where $E_{\boldsymbol{A}_{(-i)}|A_i=a}(\cdot)$ denotes the expectation with respect to the conditional distribution of $\boldsymbol{A}_{(-i)}$, the vector of treatment assignments for all villages other than i, given $A_i = a$ for $a = 0, 1$.

DE represents the *marginal direct effect* of treatment A_i, marginalizing over the distribution of treatments for all other villages. Therefore, by definition, DE depends on the treatment assignment mechanism.

Marginal indirect effects. We again consider the two types of indirect effects corresponding to the two sources of interference: within-cluster and between-cluster spillovers.

For within-cluster indirect effects, we define the *marginal within-cluster indirect effect* of S_i on the outcome as

$$\text{WIE}(s,s') \quad = \quad n^{-1}\sum_{i=1}^{n} E_{\boldsymbol{A}_{(-i)}|(A_i,S_i)=(0,s)}\left\{Y_i(0,s,H_i)\right\} - n^{-1}\sum_{i=1}^{n} E_{\boldsymbol{A}_{(-i)}|(A_i,S_i)=(0,s')}\left\{Y_i(0,s',H_i)\right\}$$

$$
\begin{aligned}
&= n^{-1}\sum_{i=1}^{n}\sum_{h=0,1}\text{pr}(H_i = h \mid A_i = 0, S_i = s)Y_i(0,s,h) \\
&\quad -n^{-1}\sum_{i=1}^{n}\sum_{h=0,1}\text{pr}(H_i = h \mid A_i = 0, S_i = s')Y_i(0,s',h),
\end{aligned}
$$

where $E_{\boldsymbol{A}_{(-i)}|(A_i,S_i)=(0,s)}(\cdot)$ denotes expectation with respect to the distribution of $\boldsymbol{A}_{(-i)}$ conditional on $A_i = 0$ and $S_i = s$. The quantity $\text{WIE}(s,s')$ is a marginal version of the contrasts, marginalized over the conditional distribution of other villages' treatments given $(A_i, S_i) = (0, s)$ and $(A_i, S_i) = (0, s')$, respectively. As with the marginal direct effect, this definition depends on the treatment assignment mechanism.

For between-cluster indirect effects, similarly, we define the *marginal between-cluster indirect effect* of H_i on the outcome as

$$
\begin{aligned}
\text{BIE}(h,h') &= n^{-1}\sum_{i=1}^{n} E_{\boldsymbol{A}_{(-i)}|(A_i,H_i)=(0,h)}\left\{Y_i(0,S_i,h)\right\} - n^{-1}\sum_{i=1}^{n} E_{\boldsymbol{A}_{(-i)}|(A_i,H_i)=(0,h')}\left\{Y_i(0,S_i,h')\right\} \\
&= n^{-1}\sum_{i=1}^{n}\sum_{s=0,1}\text{pr}(S_i = s \mid A_i = 0, H_i = h)Y_i(0,s,h) \\
&\quad -n^{-1}\sum_{i=1}^{n}\sum_{s=0,1}\text{pr}(S_i = s \mid A_i = 0, H_i = h')Y_i(0,s,h'),
\end{aligned}
$$

where $E_{\boldsymbol{A}_{(-i)}|(A_i,H_i)=(0,h)}(\cdot)$ denotes expectation conditional on $A_i = 0$ and $H_i = h$. The quantity $\text{BIE}(h,h')$ is the corresponding marginal effect, integrating over the distribution of other villages' treatments given $(A_i, H_i) = (0, h)$ and $(A_i, H_i) = (0, h')$.

Throughout, we define these indirect effects holding the treatment status $A_i = 0$ following Hudgens and Halloran (2008), though we can similarly define analogous indirect effects holding at $A_i = 1$.

If the treatment assignments A_i are independent across units, such as under Bernoulli randomization within each sublocation, then the conditional expectations $E_{\boldsymbol{A}_{(-i)}|A_i=a}(\cdot)$, $E_{\boldsymbol{A}_{(-i)}|(A_i,S_i)=(0,s)}(\cdot)$, and $E_{\boldsymbol{A}_{(-i)}|(A_i,H_i)=(0,h)}(\cdot)$ reduce to expectations over the joint distribution of (S_i, H_i), the marginal distribution of H_i, and the marginal distribution of S_i, respectively. In this case, the marginal causal effects simplify to

$$
\begin{aligned}
\text{DE} &= n^{-1}\sum_{i=1}^{n}\sum_{s=0,1}\sum_{h=0,1}\text{pr}(S_i = s, H_i = h)Y_i(1,s,h) - n^{-1}\sum_{i=1}^{n}\sum_{s=0,1}\sum_{h=0,1}\text{pr}(S_i = s, H_i = h)Y_i(0,s,h), \\
\text{WIE}(s,s') &= n^{-1}\sum_{i=1}^{n}\sum_{h=0,1}\text{pr}(H_i = h)Y_i(0,s,h) - n^{-1}\sum_{i=1}^{n}\sum_{h=0,1}\text{pr}(H_i = h)Y_i(0,s',h), \\
\text{BIE}(h,h') &= n^{-1}\sum_{i=1}^{n}\sum_{s=0,1}\text{pr}(S_i = s)Y_i(0,s,h) - n^{-1}\sum_{i=1}^{n}\sum_{s=0,1}\text{pr}(S_i = s)Y_i(0,s,h').
\end{aligned}
$$

All quantities are defined conditional on the realization of the first-stage randomization, which determines the distribution of $\boldsymbol{A}$.

2.3 Policy-specific causal estimands

In this section, we further define estimands that compare two policies, similar to Hudgens and Halloran (2008). For a specific treatment assignment policy ψ, define the policy-specific direct effect DE_ψ, within-cluster indirect effect WIE_ψ, and between-cluster indirect effect BIE_ψ as follows

$$\begin{aligned}
\text{DE}_\psi &= n^{-1}\sum_{i=1}^{n} E_\psi\{Y_i(1,S_i,H_i)\} - n^{-1}\sum_{i=1}^{n} E_\psi\{Y_i(0,S_i,H_i)\},\\
\text{WIE}_\psi &= n^{-1}\sum_{i=1}^{n} E_\psi\left\{Y_i(0,1,H_i)\right\} - n^{-1}\sum_{i=1}^{n} E_\psi\left\{Y_i(0,0,H_i)\right\},\\
\text{BIE}_\psi &= n^{-1}\sum_{i=1}^{n} E_\psi\left\{Y_i(0,S_i,1)\right\} - n^{-1}\sum_{i=1}^{n} E_\psi\left\{Y_i(0,S_i,0)\right\},
\end{aligned}$$

where the subscript ψ denotes marginalization of (S_i, H_i) conditional on $A_i = a$ under the distribution induced by policy ψ for $a = 0, 1$ in the definition of DE_ψ, marginalization of H_i conditional on $(A_i, S_i) = (0, s)$ under the distribution induced by policy ψ for $s = 0, 1$ in the definition of WIE_ψ, and marginalization of S_i conditional on $(A_i, H_i) = (0, h)$ under the distribution induced by policy ψ for $h = 0, 1$ in the definition of BIE_ψ. In particular, if we take ψ as the treatment policy implemented in the study, the above effects recover the in-policy causal estimands we defined in the previous section.

For two policies ψ_1, ψ_2, we can therefore compare their direct effect and indirect effects by the contrasts $\text{DE}_{\psi_1} - \text{DE}_{\psi_2}$, $\text{WIE}_{\psi_1} - \text{WIE}_{\psi_2}$, and $\text{BIE}_{\psi_1} - \text{BIE}_{\psi_2}$, respectively.

3 Estimation by inverse propensity score weighting

3.1 Averages of the potential outcomes and conditional effects

We first construct an estimator for the average potential outcome,

$$\bar{Y}(a,s,h) = n^{-1}\sum_{i=1}^{n} Y_i(a,s,h),$$

for given (a, s, h). Consider the Horvitz–Thompson estimator

$$\hat{Y}^{\text{ht}}(a,s,h) = n^{-1}\sum_{i=1}^{n} \frac{\mathbb{I}_i(a,s,h)}{\pi_i(a,s,h)} Y_i,$$

and the Hájek estimator

$$\hat{Y}^{\text{haj}}(a,s,h) = n^{-1}\sum_{i=1}^{n} \frac{\mathbb{I}_i(a,s,h)}{\pi_i(a,s,h)} Y_i \Big/ n^{-1}\sum_{i=1}^{n} \frac{\mathbb{I}_i(a,s,h)}{\pi_i(a,s,h)},$$

where $\mathbb{I}_i(a,s,h) = \mathbb{1}\{A_i = a, S_i = s, H_i = h\}$ and $\pi_i(a,s,h) = \text{pr}(A_i = a, S_i = s, H_i = h)$ is the corresponding propensity score. These propensity scores are determined by both the experimental design and the

network structure. In principle, they depend deterministically on the treatment assignment mechanism and definition of the exposure mapping, but directly calculating them can become infeasible as the treatment space grows. A practical alternative is to approximate these probabilities using Monte Carlo simulations.

Under randomization, the estimator $\hat{Y}^{\text{ht}}(a,s,h)$ is unbiased to $\bar{Y}(a,s,h)$. The Hájek estimator $\hat{Y}^{\text{haj}}(a,s,h)$ is not unbiased in finite sample but is consistent to $\hat{Y}^{\text{haj}}(a,s,h)$ and generally has more stable finite sample performance. Therefore, the corresponding Horvitz–Thompson estimators are unbiased, and the Hájek estimators are consistent.

We then propose the following estimators for the conditional direct and indirect effects defined in Section 2. For $* \in \{\text{ht}, \text{haj}\}$, construct

$$\begin{aligned}
\widehat{\text{DE}}^{*}(s,h) &= \hat{Y}^{*}(1,s,h) - \hat{Y}^{*}(0,s,h), \\
\widehat{\text{WIE}}^{*}(s,s',h) &= \hat{Y}^{*}(0,s,h) - \hat{Y}^{*}(0,s',h), \\
\widehat{\text{BIE}}^{*}(s,h,h') &= \hat{Y}^{*}(0,s,h) - \hat{Y}^{*}(0,s,h').
\end{aligned}$$

3.2 In-policy and policy-specific causal effects

We propose to use a policy-specific re-weighting to estimate the policy-specific effects. Under different policies, the joint distribution of (A_i, S_i, H_i) will, in general, differ. For a given policy of interest, the policy-specific causal effects are defined as marginal expectations of the potential outcomes, where the marginalization is taken with respect to the policy-induced distribution of (A_i, S_i, H_i), as defined in Sections 2.2 and 2.3. Each estimand involves a reweighted average of potential outcomes, where the weights depend on the distribution of (S_i, H_i) under a specific policy. Accordingly, to estimate these effects, we construct a class of reweighting estimators that account for the policy-induced exposure distribution. More concretely, for a given set of weights $\gamma_i(a,s,h)$, define the following class of Horvitz–Thompson estimators indexed by $\Gamma = \{\gamma_i(a,s,h)\}$:

$$\hat{Y}^{\text{ht}}(a,s,h;\Gamma) = n^{-1}\sum_{i=1}^{n} \frac{\mathbb{I}_i(a,s,h)\gamma_i(a,s,h)}{\pi_i(a,s,h)} Y_i,$$

and the corresponding class of Hájek estimators:

$$\hat{Y}^{\text{haj}}(a,s,h;\Gamma) = n^{-1}\sum_{i=1}^{n}\gamma_i(a,s,h) \cdot n^{-1}\sum_{i=1}^{n} \frac{\mathbb{I}_i(a,s,h)\gamma_i(a,s,h)}{\pi_i(a,s,h)} Y_i \Big/ n^{-1}\sum_{i=1}^{n} \frac{\mathbb{I}_i(a,s,h)\gamma_i(a,s,h)}{\pi_i(a,s,h)}.$$

These estimators form a general framework for constructing estimators of the policy-specific direct and indirect effects, as introduced in Section 2.

To operationalize the reweighting idea, we specify weight functions for each type of causal effect. Let Γ_ψ^{DE}, Γ_ψ^{BIE}, and Γ_ψ^{WIE} denote the corresponding classes of weighting functions for the direct, within-sublocation indirect, and between-sublocation indirect effects, respectively. The elements of these classes are defined as

$$\gamma_{i,\psi}^{\text{DE}}(a,s,h) = \text{pr}_\psi(S_i = s, H_i = h \mid A_i = a),$$

$$
\begin{aligned}
\gamma^{\mathrm{WIE}}_{i,\psi}(a,s,h) &= \mathrm{pr}_\psi(H_i = h \mid A_i = a, S_i = s),\\
\gamma^{\mathrm{BIE}}_{i,\psi}(a,s,h) &= \mathrm{pr}_\psi(S_i = s \mid A_i = a, H_i = h).
\end{aligned}
$$

These probabilities describe the distribution of (S_i, H_i) under a specific policy ψ, which serve as reweighting terms in our estimators. For either the Horvitz–Thompson or Hájek estimator ($* \in \{\mathrm{ht}, \mathrm{haj}\}$), we then define the corresponding estimators for the policy-specific causal effects as:

$$
\begin{aligned}
\hat{\mathrm{DE}}^*_\psi &= \sum_{s,h=0,1} \{\hat{Y}^*(1,s,h;\Gamma^{\mathrm{DE}}_\psi) - \hat{Y}^*(0,s,h;\Gamma^{\mathrm{DE}}_\psi)\},\\
\hat{\mathrm{WIE}}^*_\psi &= \sum_{h=0,1} \{\hat{Y}^*(0,1,h;\Gamma^{\mathrm{WIE}}_\psi) - \hat{Y}^*(0,0,h;\Gamma^{\mathrm{WIE}}_\psi)\},\\
\hat{\mathrm{BIE}}^*_\psi &= \sum_{s=0,1} \{\hat{Y}^*(0,s,1;\Gamma^{\mathrm{BIE}}_\psi) - \hat{Y}^*(0,s,0;\Gamma^{\mathrm{BIE}}_\psi)\}.
\end{aligned}
$$

As a special case, if we take $\psi = \phi$ as the treatment policy actually implemented in the real study, these expressions yield the in-policy estimators for the marginal marginal causal effects:

$$
\begin{aligned}
\hat{\mathrm{DE}}^*_\phi &= \sum_{s,h=0,1} \{\hat{Y}^*(1,s,h;\Gamma^{\mathrm{DE}}_\phi) - \hat{Y}^*(0,s,h;\Gamma^{\mathrm{DE}}_\phi)\},\\
\hat{\mathrm{WIE}}^*_\phi &= \sum_{h=0,1} \{\hat{Y}^*(0,1,h;\Gamma^{\mathrm{WIE}}_\phi) - \hat{Y}^*(0,0,h;\Gamma^{\mathrm{WIE}}_\phi)\},\\
\hat{\mathrm{BIE}}^*_\phi &= \sum_{s=0,1} \{\hat{Y}^*(0,s,1;\Gamma^{\mathrm{BIE}}_\phi) - \hat{Y}^*(0,s,0;\Gamma^{\mathrm{BIE}}_\phi)\}.
\end{aligned}
$$

Remark 3.1 (Covariate adjustment estimator)**.** *Let X_i denote the vector of centered pre-treatment covariates for unit i, augmented with a constant term in the first position. For each exposure configuration $(a,s,h) \in \{0,1\}^3$, we define a covariate-adjusted estimator of the average potential outcome $\bar{Y}(a,s,h)$ as*

$$
\hat{Y}^{\mathrm{ca}}(a,s,h) = n^{-1}\sum_{i=1}^{n}\left\{\frac{\mathbb{1}_i(a,s,h)}{\pi_i(a,s,h)}(Y_i - \hat{\beta}^{\mathrm{T}}_{a,s,h}X_i) + \hat{\beta}^{\mathrm{T}}_{a,s,h}X_i\right\},
$$

where $\hat{\beta}_{a,s,h}$ is the ordinary least squares coefficient from regressing Y_i on X_i within the subsample $(A_i,S_i,H_i) = (a,s,h)$. The estimator $\hat{Y}^{\mathrm{ca}}(a,s,h)$ combines regression adjustment with inverse probability weighting. It is a standard strategy in the literature and we omit the development of its theoretical properties here.

For the conditional causal effects, we take differences between the corresponding $\hat{Y}^{\mathrm{ca}}(a,s,h)$ estimators. For the marginal causal effects, we form weighted averages of $\hat{Y}^{\mathrm{ca}}(a,s,h)$'s and take contrasts, using the corresponding policy-induced probabilities as weights.

4 Theoretical properties

4.1 Asymptotic variance

We first derive the asymptotic variance of the reweighted estimators, which serves as the building block for our asymptotic analysis. Let $\mathbf{\Gamma}(a,s,h) = \text{diag}\{\gamma_i(a,s,h)\}$ denote the $n \times n$ diagonal matrix with diagonal terms equal to the weights $\gamma_i(a,s,h)$ for $i = 1,\ldots,n$. To characterize dependence induced by the design, we further define two $n \times n$ matrices, $\mathbf{\Lambda}(a,s,h)$ and $\mathbf{\Lambda}(a,s,h;a',s',h')$, that will be crucial in defining the asymptotic variance:

$$\begin{aligned}
\mathbf{\Lambda}_{(i,j)}(a,s,h) &= \mathbb{1}\{j=i\}\frac{1-\pi_i(a,s,h)}{\pi_i(a,s,h)} + \mathbb{1}\{j\neq i\}\frac{\pi_{ij}(a,s,h;a,s,h)-\pi_i(a,s,h)\pi_j(a,s,h)}{\pi_i(a,s,h)\pi_j(a,s,h)},\\
\mathbf{\Lambda}_{(i,j)}(a,s,h;a',s',h') &= -\mathbb{1}\{j=i\} + \mathbb{1}\{j\neq i\}\frac{\pi_{ij}(a,s,h;a',s',h')-\pi_i(a,s,h)\pi_j(a',s',h')}{\pi_i(a,s,h)\pi_j(a',s',h')}.
\end{aligned}$$

In practice, these second-order inclusion probabilities $\pi_{ij}(a,s,h;a',s',h')$ can be simulated given a known randomization policy and network structure. See Aronow and Samii (2017) for further discussion.

In addition, the asymptotic variance depends on whether the potential outcomes are centered, and if so, around what quantity. The Horvitz–Thompson estimator uses the raw potential outcomes and therefore has a variance expression directly in terms of $Y_i(a,s,h)$, so we define $Y_i^{\text{ht}}(a,s,h) = Y_i(a,s,h)$. In contrast, the Hájek estimator normalizes the weights by an estimated denominator, which induces additional dependence across units. It is therefore convenient to rewrite the Hájek estimator as an inverse probability weighting estimator applied to centered potential outcomes. Accordingly, for the Hájek form we define

$$Y_i^{\text{haj}}(a,s,h) = Y_i(a,s,h) - \frac{n^{-1}\sum_{i=1}^n \gamma_i(a,s,h)Y_i(a,s,h)}{n^{-1}\sum_{i=1}^n \gamma_i(a,s,h)}.$$

Let $\boldsymbol{Y}^*(a,s,h) = (Y_1^*(a,s,h),\ldots,Y_n^*(a,s,h))^{\mathrm{T}}$ denote the vector of potential outcomes for all units.

Theorem 4.1. *For a reweighting regime Γ and $* \in \{\text{ht},\text{haj}\}$, the asymptotic variance of $\hat{Y}^*(a,s,h;\Gamma)$ at a fixed treatment and exposure mapping level (a,s,h) is*

$$\text{avar}\{\hat{Y}^*(a,s,h;\Gamma)\} = n^{-2}\boldsymbol{Y}^*(a,s,h)^{\mathrm{T}}\mathbf{\Gamma}(a,s,h)\mathbf{\Lambda}(a,s,h)\mathbf{\Gamma}(a,s,h)\boldsymbol{Y}^*(a,s,h),$$

and the asymptotic covariance for a given pair (a,s,h) and (a',s',h') is

$$\text{acov}\{\hat{Y}^*(a,s,h;\Gamma),\hat{Y}^*(a',s',h';\Gamma)\} = n^{-2}\boldsymbol{Y}^*(a,s,h)^{\mathrm{T}}\mathbf{\Gamma}(a,s,h)\mathbf{\Lambda}(a,s,h;a',s',h')\mathbf{\Gamma}(a',s',h')\boldsymbol{Y}^*(a',s',h'),$$

where $\text{avar}(\cdot)$ *and* $\text{acov}(\cdot,\cdot)$ *denote the asymptotic variance and asymptotic covariance, respectively.*

4.2 Consistency and asymptotic normality

We next establish the asymptotic properties of the proposed estimators. Following the framework of Aronow and Samii (2017) for inverse probability weighting estimators under network interference and Chen and

Shao (2004) for central limit theorems under network dependence, we establish consistency and asymptotic normality under the following four regularity conditions.

Assumption 1 (Bounded potential outcomes)**.** *There exists a constant $C_Y < \infty$ such that $|Y_i(a, s, h)| \leq C_Y$ for all $i = 1, \ldots, n$ and all $(a, s, h) \in \{0, 1\}^3$.*

Assumption 1 is standard in the literature and aligns with many practical metrics (Aronow and Samii, 2017; Gao and Ding, 2025; Lu et al., 2025). Although it can be relaxed, we keep the boundedness assumption for clarity and to simplify the presentation.

Assumption 2 (Positivity of exposure probabilities)**.** *There exist constants $0 < \underline{c}_\pi < \bar{c}_\pi < 1$ such that $\underline{c}_\pi \leq \pi_i(a, s, h) \leq \bar{c}_\pi$ for every unit i and all exposure configurations (a, s, h).*

Assumption 2 extends the classical positivity assumption to the interference setting. It requires that every exposure combination has a non-negligible probability of occurring. This ensures that the invrese probability weights remain well-behaved and do not explode, which is essential both for identification of the causal estimands and for controlling estimator variance.

Assumption 3 (Bounded network degree)**.** *There exists a finite constant $\Delta < \infty$ such that each unit has at most Δ neighbors:* $\max(|\{j : k(j) = k(i)\} \cup \mathcal{G}_i|) \leq \Delta$.

Assumption 3 imposes a standard sparsity condition on the interference network: each unit has a relatively sparse number of neighbors. This condition is important for guaranteeing a stable asymptotic distribution for the estimators, as it prevents the dependence structure from becoming too dense shen n grows. Dense networks can violate the classical dependency-graph conditions required for central limit theorems, and in such regimes we would need either additional restrictions on how quickly degrees are allowed to grow or alternative asymptotic framework designed for dense networks. In our empirical application with $n = 653$ villages, this assumption holds with $\Delta \approx 20$, corresponding to the maximum number of villages in any sublocation plus nearby villages within 4 km.

Assumption 4 (Bounded order of dependence)**.** *The exposure mappings (A_i, S_i, H_i) satisfy a bounded dependence condition: for any two units i and j with graph distance greater than m (where graph distance is measured on the the union of sublocation and neighborhood network), the random vectors (A_i, S_i, H_i) and (A_j, S_j, H_j) are independent. The constant $m < \infty$ is the order of dependence.*

Assumption 4 describes how far dependence can propagate in the interference structure. It requires that units sufficiently far apart in the network behave independently, which is another key requirement for applying the dependency graph central limit theorems. Intuitively, the dependence induced by randomization does not extend indefinitely: units outside each other's m-step neighborhoods cannot influence each other's exposure conditions. In our empirical setting, the bounded dependence assumption holds with $m = 2$ because each village's exposure depends only on its own sublocation and immediate neighbors, and any units separated by two or more steps behave independently conditional on the randomization.

Under these assumptions, we establish the following asymptotic results for our policy-specific estimators.

Theorem 4.2 (Consistency and asymptotic normality)**.** *Suppose Assumptions 1–4 hold. Let* $\hat{\boldsymbol{Y}}_\Gamma^*$ *and* $\bar{\boldsymbol{Y}}_\Gamma^*$ *denote the vectors of estimators and average weighted potential outcomes across all exposure combinations under weighting regime* Γ*, for* $* \in \{\text{ht}, \text{haj}\}$*. Then as* $n \to \infty$*:*

1. ***Consistency:*** *For any weight function* Γ *and any exposure* (a, s, h)*, we have for* $* \in \{\text{ht}, \text{haj}\}$*,*

$$\hat{\boldsymbol{Y}}_\Gamma^* \xrightarrow{p} \bar{\boldsymbol{Y}}_\Gamma.$$

   Consequently, all proposed estimators for the causal effects, including the conditional effects, the in-policy marginal effects, and the policy-specific effects, are consistent for their respective population estimands.

2. ***Asymptotic normality:*** *The joint vector of estimators satisfies*

$$\text{acov}(\hat{\boldsymbol{Y}}_\Gamma^*)^{-1/2} \left(\hat{\boldsymbol{Y}}_\Gamma^* - \bar{\boldsymbol{Y}}_\Gamma \right) \xrightarrow{d} N\left(0, I_8\right)$$

   for $* \in \{\text{ht}, \text{haj}\}$*, where* I_8 *is the identity matrix of dimension 8.*

5 Variance estimation and inference

5.1 Variance estimation

The asymptotic variances in Theorem 4.1 require knowledge of the true potential outcomes $Y_i(a, s, h)$, which are not directly observable. In this section, we propose conservative variance estimators that rely only on observed data.

Define the aggregated vector $\hat{\boldsymbol{Y}}^* = (\hat{Y}_1^*, \ldots, \hat{Y}_n^*)^\text{T}$, where for $* \in \{\text{ht}, \text{haj}\}$, $\hat{Y}_i^{\text{ht}} = \mathbb{I}_i(a, s, h) Y_i / \pi_i(a, s, h)$, and $\hat{Y}_i^{\text{haj}} = \mathbb{I}_i(a, s, h) \tilde{Y}_i / \pi_i(a, s, h)$ where

$$\tilde{Y}_i = Y_i - \frac{\hat{Y}^{\text{haj}}(a, s, h; \Gamma)}{n^{-1} \sum_{i=1}^n \gamma_i(a, s, h)}.$$

Next, define the $n \times n$ matrix $\boldsymbol{\Omega}$ with entries:

$$\boldsymbol{\Omega}_{(i,j)} = \frac{\pi_{ij}(a, s, h; a, s, h) - \pi_i(a, s, h)\pi_j(a, s, h)}{\pi_{ij}(a, s, h; a, s, h)},$$

which is a reweighted version of $\boldsymbol{\Lambda}_{(i,j)}$. Then we construct the following variance estimator for $\text{var}\{\hat{Y}^*(a, s, h; \Gamma)\}$ for $* \in \{\text{ht}, \text{haj}\}$:

$$\hat{\text{var}}\{\hat{Y}^*(a, s, h; \Gamma)\} \quad = \quad n^{-2} (\hat{\boldsymbol{Y}}^*)^\text{T} \boldsymbol{\Gamma}(a, s, h) \boldsymbol{\Omega}(a, s, h) \boldsymbol{\Gamma}(a, s, h) \hat{\boldsymbol{Y}}^*.$$

For conditional causal effects such as $\hat{\text{DE}}^*(s, h) = \hat{Y}^*(1, s, h) - \hat{Y}^*(0, s, h)$, the true asymptotic variance includes the covariance term that depends on the joint of different potential outcomes and is not identified.

Decomposing the variance,

$$\text{var}\{\hat{\text{DE}}^*(s,h)\} \quad = \quad \text{var}\{\hat{Y}^*(1,s,h)\} + \text{var}\{\hat{Y}^*(0,s,h)\} - 2\text{cov}\{\hat{Y}^*(1,s,h), \hat{Y}^*(0,s,h)\},$$

reveals that the two variance components are identifiable but the covariance is not. For valid inference, we obtain an upper bound that is identifiable based on the observed data for the covariance term. We use the fact that $\text{cov}^2\{\hat{Y}^*(1,s,h), \hat{Y}^*(0,s,h)\} \leq \text{var}\{\hat{Y}^*(1,s,h)\}\text{var}\{\hat{Y}^*(0,s,h)\}$ guaranteed by the Cauchy–Schwarz inequality, with equality holding when the two estimators $\hat{Y}^*(1,s,h)$ and $\hat{Y}^*(0,s,h)$ are perfectly correlated. This motivates the following conservative variance estimator for $* \in \{\text{ht}, \text{haj}\}$,

$$\hat{\text{var}}\{\hat{\text{DE}}^*(s,h)\} = \left[\hat{\text{se}}\{\hat{Y}^*(1,s,h)\} + \hat{\text{se}}\{\hat{Y}^*(0,s,h)\}\right]^2.$$

We can similarly construct conservative variance estimators for other conditional causal effects.

For policy-specific causal effects, such as the in-policy marginal direct effect $\hat{\text{DE}}^* = \sum_{s,h}\{\hat{Y}^*(1,s,h;\Gamma_\phi^{\text{DE}}) - \hat{Y}^*(0,s,h;\Gamma_\phi^{\text{DE}})\}$, we similarly use $\hat{\text{var}}(\hat{\text{DE}}^*) = [\hat{\text{se}}\{\hat{Y}^*(1,s,h;\Gamma_\phi^{\text{DE}})\} + \hat{\text{se}}\{\hat{Y}^*(0,s,h;\Gamma_\phi^{\text{DE}})\}]^2$ as a variance estimator. We can similarly construct conservative variance estimators for other policy-specific causal effects.

5.2 Inference

In this section, we provide asymptotic results for the variance estimator to support valid inference.

Corollary 5.1 (Asymptotic validity of confidence intervals)**.** *Under Assumptions 1–4,*

1. *The variance estimators* $\hat{\text{var}}\{\hat{Y}^*(a,s,h;\Gamma)\}$ *for* $* \in \{\text{ht}, \text{haj}\}$ *is consistent:*

$$\hat{\text{var}}\{\hat{Y}^*(a,s,h;\Gamma)\} \xrightarrow{p} \text{var}\{\hat{Y}^*(a,s,h;\Gamma)\},$$

and therefore for a single exposure configuration (a,s,h)*, the confidence interval*

$$\hat{Y}^*(a,s,h;\Gamma) \pm z_{\alpha/2} \cdot \widehat{\text{se}}\{\hat{Y}^*(a,s,h;\Gamma)\}$$

achieves asymptotic coverage rate $1-\alpha$ *for* $* \in \{\text{ht}, \text{haj}\}$.

2. *For conditional causal effects and policy-specific causal effects, the conservative variance estimator based on the Cauchy–Schwarz bound provides asymptotically valid confidence intervals with asymptotic coverage rate at least* $1-\alpha$*. In general, the actual coverage may exceed the nominal level due to the conservativeness of the bound.*

Remark 5.1. *If the second-order inclusion probability* $\pi_{ij}(a,s,h;a,s,h) = 0$ *for some pair* (i,j)*, the variance of* $\hat{Y}^*(a,s,h)$ *cannot be consistently estimated either. We can construct conservative estimators for the variance term following the estimators proposed in Aronow and Samii (2017).*

6 Real data analysis

6.1 Analysis paradigm

We apply our methodology to re-analyze the cash transfer experiment studied in Egger et al. (2022). The experiment assigned villages to treatment using a two-stage randomized saturation design across 653 villages in 155 sublocations, resulting in 328 treated and 325 control villages. Our empirical analysis focuses on four village-level enterprise outcomes measured at endline: winsorized average profit, revenue, total cost, and wage bill. We apply our proposed methodology to quantify direct and spillover effects.

A key feature of this setting is that interference may arise through two distinct channels: (i) villages in the same sublocation share administrative and economic ties, and (ii) nearby villages in different sublocations may also influence each other through geographic proximity. To capture these two sources of interference, we construct two binary exposure variables S_i and H_i following Section 2. For S_i, we summarize treatment saturation within i's sublocation. For H_i, we build a geographic network using distance data, identifying for each village i up to three nearest neighbors *outside* its sublocation and within 4 km.

Because both S_i and H_i depend on the treatment assignments of multiple nearby villages, the propensity scores $\pi_i(a, s, h)$ are not available in closed form due to the complex dependency between (A_i, S_i, H_i) induced by the network structure. We therefore estimate the propensity scores $\pi_i(a, s, h)$ and the second-order inclusion probabilities $\pi_{ij}(a, s, h; a', s', h')$ using 100,000 Monte Carlo draws following Aronow and Samii (2017). Table 1 reports summary statistics for the estimated propensity scores across the 653 villages in our sample.

Table 1: Summary statistics of estimated propensity scores

Exposure	Mean	Std	Median
$(A=0, S=0, H=0)$	0.184	0.175	0.137
$(A=0, S=0, H=1)$	0.143	0.141	0.099
$(A=0, S=1, H=0)$	0.088	0.062	0.071
$(A=0, S=1, H=1)$	0.086	0.068	0.068
$(A=1, S=0, H=0)$	0.118	0.078	0.088
$(A=1, S=0, H=1)$	0.098	0.065	0.078
$(A=1, S=1, H=0)$	0.139	0.143	0.069
$(A=1, S=1, H=1)$	0.144	0.154	0.051
Marginal probabilities			
$\text{pr}(A_i = 1)$	0.499	0.167	0.500
$\text{pr}(S_i = 1)$	0.458	0.342	0.500
$\text{pr}(H_i = 1)$	0.471	0.224	0.500

Notes: Propensity scores are estimated using 100,000 Monte Carlo draws. Marginal probabilities are computed by summing the relevant joint propensities for each village. For each village, the eight joint propensities sum to one.

6.2 Results

We now present the results of our proposed estimators. We organize the results into two parts: Table 2 reports the conditional direct and indirect effects, and Table 3 reports the in-policy marginal direct and indirect effects. For each estimand, we report three estimators: the Horvitz–Thompson estimator, the Hájek estimator, and the covariate-adjusted estimator.

Overall, the Hájek estimator provides substantially more precise estimates than the Horvitz–Thompson estimator, with standard errors typically 50–70% smaller. This is consistent with the well-known finite-sample efficiency advantages of Hájek estimation (Aronow and Samii, 2017; Ding, 2024; Gao and Ding, 2025). Also, the results are relatively stable between the Hájek and the covariate-adjusted estimator in terms of both point estimators and the standard error.

Conditional causal effects. Panel A of Table 2 reports the estimated conditional direct effects $\widehat{\text{DE}}^*(s,h)$, which measure the treatment effect of village i receiving a cash transfer, conditional on its exposure environment $(S_i, H_i) = (s, h)$.

The results reveal several important patterns. Focusing on the more precise Hájek and covariate-adjusted estimates, we find significant positive effects on profits and revenues when villages are in high within-sublocation saturation but low geographic exposure environments ($S = 1, H = 0$). In this setting, treated villages experience increases of 3,539 KES in monthly profit ($p < 0.01$) and 4,159 KES in monthly revenue ($p < 0.01$). This suggests that within-sublocation indirect effects create favorable conditions for treated enterprises to grow.

By contrast, enterprises in villages with low local saturation but high geographic exposure ($S = 0, H = 1$) face higher costs and wages, consistent with increased competition for inputs and labor. Effects are close to zero or negative when both exposure levels are simultaneously high or low, highlighting the importance of accounting for multiple indirect effect channels.

We next examine the conditional indirect effects, which isolate indirect effect channels by holding the village's own treatment status fixed. In addition to the conditional indirect effects introduced in Section 2 for $A_i = 0$, we also report estimation results for analogous estimands holding $A_i = 1$. Panel B of Table 2 reports the within-sublocation indirect effects, which measure the impact of changing within-sublocation treatment saturation from low to high, conditional on the village's own treatment status $A_i = a$ and geographic exposure $H_i = h$.

The results reveal distinct patterns depending on treatment status. Among control villages ($A = 0$), the Horvitz–Thompson estimator suggests large positive effects on profits and revenues from increased sublocation saturation, though with substantial uncertainty. The Hájek estimator, on the contrary, shows negative effects when geographic exposure is high ($H = 1$): control villages in high-saturation sublocations with many nearby treated villages experience profit and revenue decreases, consistent with competitive pressure from treated neighbors.

Among treated villages ($A = 1$), the Hájek estimator finds large positive effects when geographic exposure is low ($H = 0$): treated villages benefit substantially from being in high-saturation sublocations when they

have few treated neighbors outside their sublocation. However, when geographic exposure is also high ($H = 1$), these benefits disappear.

Panel C of Table 2 presents the between-sublocation indirect effects, which measure the impact of increasing geographic exposure to treated villages in other sublocations from low to high, conditional on own treatment $A_i = a$ and within-sublocation saturation $S_i = s$.

These between-sublocation indirect effects are heterogeneous across exposure configurations. Using the Hájek estimator, control villages in low-saturation sublocations ($A = 0, S = 0$) experience significant positive effects from greater geographic exposure to treated villages. This suggests that control villages benefit from proximity to treated villages, possibly through increased economic activity, though they also face higher input costs. However, among treated villages in high-saturation sublocations ($A = 1, S = 1$), the effects are strongly negative.

In-policy marginal effects. Table 3 reports estimates of the in-policy marginal direct and indirect effects, where the weighting scheme reflects the exposure distribution induced by the implemented policy. The marginal direct effect averages over exposure levels using weights $\gamma_i(a, s, h) = \text{pr}(S_i = s, H_i = h \mid A_i = a)$, while the within- and between-sublocation indirect effects average over the distributions using weights $\text{pr}(S_i = s \mid A_i = a, H_i = h)$ and $\text{pr}(H_i = h \mid A_i = a, S_i = s)$, respectively.

In general, the in-policy marginal causal effects have relatively large standard errors, and most estimates are not statistically significant. As before, the Hájek and the covariate-adjusted estimators are considerably more precise, with standard errors roughly 50-60% smaller than the Horvitz–Thompson estimator. The lack of significance is consistent with the substantial heterogeneity documented in Table 2: when the conditional causal effects vary strongly across exposure environments, marginalizing over these environments tends to dilute the signal and reduce statistical power.

7 Discussion

Beyond the nonparametric inverse propensity score weighting estimators developed in Section 3, it is also natural to consider regression-assisted approaches that partially pool information across exposure configurations. Because our estimands involve eight possible (a, s, h) combinations, some cells may be sparsely populated in the real data. A partially saturated regression that restricts the three-way interaction and some two-way interactions among (A, S, H) is of interest as a model-assisted perspective. The idea connects directly to the model-assisted framework of Zhao and Ding (2022a) and to classical analyses of 2^3 factorial designs. In Section A.2 of the supplementary material, we explore the perspective by considering a regression that includes main effects of (A, S, H) and selected two-way interactions, with S and H centered at their population means. We provide causal interpretations of the regression coefficients under independent Bernoulli assignment.

Throughout this paper, we have focused on binary exposure mappings for (S_i, H_i), which leads to a finite number of exposure environments. This setting is conceptually simple and allows for nonparametric identification of all relevant average potential outcomes. The framework can be extended to categorical

Table 2: Conditional direct and indirect effects

Panel A. Conditional direct effects: $\widehat{\text{DE}}^*(s,h)$

	Horvitz–Thompson				Hájek				Covariate-adjusted			
(s,h)	Profit	Revenue	Costs	Wage	Profit	Revenue	Costs	Wage	Profit	Revenue	Costs	Wage
$(0,0)$	2,142	3,739	542	406	−360	−46	200	163	−101	217	172	136
	(2,180)	(3,302)	(352)	(262)	(909)	(1,143)	(135)	(116)	(883)	(1,117)	(137)	(117)
$(0,1)$	3,211	5,841	825*	625*	−558	300	329**	268*	−835	−124	260*	211
	(2,406)	(3,940)	(476)	(361)	(616)	(877)	(168)	(145)	(559)	(696)	(148)	(131)
$(1,0)$	272	−440	−300	−253	3,539***	4,159***	87	23	2,817**	3,043***	−150	−181
	(3,431)	(4,518)	(469)	(372)	(1,263)	(1,247)	(169)	(159)	(1,268)	(1,167)	(157)	(149)
$(1,1)$	−3,233	−4,785	−458	−340	−227	140	115	73	37	481	142	95
	(2,225)	(3,364)	(353)	(260)	(688)	(825)	(158)	(134)	(677)	(776)	(153)	(131)

Panel B. Conditional within-sublocation indirect effects: $\widehat{\text{WIE}}^*(a,h)$

	Horvitz–Thompson				Hájek				Covariate-adjusted			
(a,h)	Profit	Revenue	Costs	Wage	Profit	Revenue	Costs	Wage	Profit	Revenue	Costs	Wage
$(0,0)$	2,881	4,420	584*	457*	−182	−304	132	134	67	363	333**	304***
	(1,895)	(2,813)	(337)	(262)	(837)	(961)	(170)	(148)	(823)	(943)	(159)	(140)
$(0,1)$	3,078	4,848	541	396	−1,136*	−1,579*	−92	−67	−1,333**	−1,728**	−76	−50
	(2,072)	(3,131)	(331)	(248)	(671)	(813)	(138)	(121)	(654)	(750)	(132)	(116)
$(1,0)$	1,011	242	−258	−202	3,718***	3,902***	18	−6	2,985**	3,189**	11	−13
	(3,715)	(5,007)	(483)	(372)	(1,335)	(1,429)	(134)	(127)	(1,327)	(1,342)	(134)	(126)
$(1,1)$	−3,366	−5,778	−741	−569	−805	−1,738**	−306	−262*	−461	−1,124	−194	−166
	(2,558)	(4,173)	(498)	(373)	(633)	(889)	(188)	(158)	(581)	(722)	(169)	(147)

Panel C. Conditional between-sublocation indirect effects: $\widehat{\text{BIE}}^*(a,s)$

	Horvitz–Thompson				Hájek				Covariate-adjusted			
(a,s)	Profit	Revenue	Costs	Wage	Profit	Revenue	Costs	Wage	Profit	Revenue	Costs	Wage
$(0,0)$	−598	−964	−82	−52	1,330**	1,950***	222**	176**	1,442***	1,998***	204**	161**
	(1,458)	(2,190)	(233)	(177)	(554)	(645)	(93)	(83)	(511)	(555)	(86)	(76)
$(0,1)$	−402	−537	−125	−113	375	674	−2	−25	42	−92	−204	−194
	(2,509)	(3,755)	(435)	(333)	(954)	(1,128)	(214)	(186)	(967)	(1,138)	(205)	(180)
$(1,0)$	471	1,138	200	167	1,132	2,296*	350*	282	708	1,658	293*	236
	(3,127)	(5,052)	(594)	(446)	(971)	(1,374)	(209)	(178)	(930)	(1,258)	(199)	(172)
$(1,1)$	−3,906	−4,882	−283	−201	−3,391***	−3,344***	26	25	−2,739***	−2,654***	88	83
	(3,146)	(4,127)	(387)	(299)	(997)	(944)	(112)	(106)	(978)	(805)	(104)	(101)

Notes: Each panel reports conditional causal effects estimated using Horvitz–Thompson, Hájek, and Covariate-adjusted estimators. Panel A reports conditional direct effects comparing treated ($A_i = 1$) versus control ($A_i = 0$) villages for each (s,h). Panel B reports conditional within-sublocation indirect effects comparing high ($S_i = 1$) versus low ($S_i = 0$) saturation levels for each (a,h). Panel C reports between-sublocation indirect effects comparing high ($H_i = 1$) versus low ($H_i = 0$) exposure levels for each (a,s). All monetary values are in Kenyan Shillings (KES) per enterprise per month. Point estimates are reported in the first line, with robust standard errors in parentheses below. The covariate-adjusted estimator includes baseline covariates to improve efficiency. Significance levels: $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

exposure mappings by defining additional exposure levels, for example, low/medium/high proportions of treated neighboring villages. However, the number of exposure cells grows quickly with the number of categories, and the corresponding number of average potential outcomes increases combinatorially. A more challenging extension involves allowing (S_i, H_i) to take continuous values, such as the exact proportion of treated neighbors within and outside the sublocation. In this case, the object of interest becomes an

Table 3: In-policy marginal direct and indirect effects

Horvitz–Thompson				Hájek				Covariate-adjusted			
Profit	Revenue	Costs	Wage	Profit	Revenue	Costs	Wage	Profit	Revenue	Costs	Wage
Marginal direct effect											
577	1,180	191	140	335	810	164	121	289	655	99	64
(2,597)	(3,930)	(476)	(370)	(1,051)	(1,242)	(232)	(204)	(994)	(1,119)	(215)	(191)
Within-sublocation indirect effect											
232	360	102	89	−412	−639	15	29	−378	−485	62	67
(1,085)	(1,652)	(200)	(156)	(366)	(434)	(88)	(78)	(377)	(460)	(83)	(73)
Between-sublocation indirect effect											
−374	−565	−49	−33	433	663	74	58	499	682	47	33
(1,385)	(2,068)	(224)	(170)	(548)	(636)	(101)	(89)	(514)	(564)	(95)	(84)

Notes: This table reports in-policy marginal direct and indirect effects estimates. The marginal direct effect uses weights $\gamma_i = \pi_i(\cdot, s, h)$ that marginalize over the treatment distribution. Within-sublocation spillover effects compare $S = 1$ to $S = 0$ for control and treated villages using weights $\gamma_i = \pi_i(\cdot, h, \cdot)$, while between-sublocation spillover effects compare $H = 1$ to $H = 0$ using $\gamma_i = \pi_i(\cdot, \cdot, s)$. All monetary values are reported in Kenyan Shillings (KES) per enterprise per month. Point estimates are shown in the first line, with robust standard errors reported in parentheses beneath. The covariate-adjusted estimator incorporates baseline covariates for improved efficiency. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

exposure-response function $(s, h) \mapsto \bar{Y}(a, s, h)$, and recovering it requires either nonparametric methods or additional model structure. A fully nonparametric approach would require smoothing or kernel methods over a two-dimensional continuous exposure space, which may suffer from the curse of dimensionality and require a large sample size for stable estimation (Kennedy et al., 2017). Alternatively, one may impose parametric or semiparametric assumptions on the exposure–response relationship to obtain more stable inference at the cost of additional modeling assumptions.

In our real data analysis, we implemented an intuitive covariate-adjustment strategy, analogous to Lin's estimator (Lin, 2013) in the no-interference setting, but without a formal justification for variance reduction under interference. Recent studies have begun to explore covariate adjustment under various settings with interference (Gao and Ding, 2025; Lu et al., 2025; Chang, 2025). A natural direction for future research is to develop a rigorous model-assisted covariate adjustment framework for interference settings, together with provable guarantees on efficiency gains. We leave it for future work.

References

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

Baird, S., Bohren, J. A., McIntosh, C., and Özler, B. (2018). Optimal design of experiments in the presence of interference. *Review of Economics and Statistics*, 100(5):844–860.

Basse, G. and Feller, A. (2018). Analyzing two-stage experiments in the presence of interference. *Journal of the American Statistical Association*, 113(521):41–55.

Benjamin-Chung, J., Arnold, B. F., Berger, D., Luby, S. P., Miguel, E., Colford Jr, J. M., and Hubbard, A. E.

(2018). Spillover effects in epidemiology: parameters, study designs and methodological considerations. *International Journal of Epidemiology*, 47(1):332–347.

Chang, H. (2025). Design-based estimation theory for complex experiments.

Chen, L. H. and Shao, Q.-M. (2004). Normal approximation under local dependence. *The Annals of Probability*, 32(3A):1985–2028.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. (2013). Do labor market policies have displacement effects? evidence from a clustered randomized experiment. *Quarterly Journal of Economics*, 128(2):531–580.

Ding, P. (2024). *A First Course in Causal Inference*. London: Chapman and Hall.

Egger, D., Haushofer, J., Miguel, E., Niehaus, P., and Walker, M. (2022). General equilibrium effects of cash transfers: experimental evidence from kenya. *Econometrica*, 90(6):2603–2643.

Gao, M. and Ding, P. (2025). Causal inference in network experiments: regression-based analysis and design-based properties. *Journal of Econometrics*, 252:106119.

Giffin, A., Reich, B., Yang, S., and Rappold, A. (2023). Generalized propensity score approach to causal inference with spatial interference. *Biometrics*, 79(3):2220–2231.

Hudgens, M. G. and Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842.

Jiang, Z., Imai, K., and Malani, A. (2023). Statistical inference and power analysis for direct and spillover effects in two-stage randomized experiments. *Biometrics*, 79(3):2370–2381.

Kennedy, E. H., Ma, Z., McHugh, M. D., and Small, D. S. (2017). Non-parametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(4):1229–1245.

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

Leung, M. P. (2025). Cluster-randomized trials with cross-cluster interference. *Journal of the American Statistical Association*, 0(ja):1–20.

Li, F., Morgan, K. L., and Zaslavsky, A. M. (2018). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521):390–400.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *The Annals of Applied Statistics*, 7(1):295–318.

Lu, S., Shi, L., Fang, Y., Zhang, W., and Ding, P. (2025). Design-based causal inference in bipartite experiments.

Melis, R. J., van Eijken, M. I., Borm, G. F., Wensing, M., Adang, E., van de Lisdonk, E. H., van Achterberg, T., and Olde Rikkert, M. G. (2005). The design of the dutch easycare study: a randomised controlled trial on the effectiveness of a problem-based community intervention model for frail elderly people [nct00105378]. *BMC Health Services Research*, 5(1):65.

Mukerjee, R., Dasgupta, T., and Rubin, D. B. (2018). Using standard tools from finite population sampling to improve causal inference for complex experiments. *Journal of the American Statistical Association*, 113(522):868–881.

Papadogeorgou, G., Imai, K., Lyall, J., and Li, F. (2022). Causal inference with spatio-temporal data: Estimating the effects of airstrikes on insurgent violence in iraq. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84(5):1969–1999.

Sinclair, B., McConnell, M., and Green, D. P. (2012). Detecting spillover effects: Design and analysis of multilevel experiments. *American Journal of Political Science*, 56(4):1055–1069.

Wang, Y., Samii, C., Chang, H., and Aronow, P. (2025). Design-based inference for spatial experiments under unknown interference. *The Annals of Applied Statistics*, 19(1):744–768.

Wu, C. J. and Hamada, M. S. (2011). *Experiments: planning, analysis, and optimization*. John Wiley & Sons.

Zhao, A. and Ding, P. (2022a). Reconciling design-based and model-based causal inferences for split-plot experiments. *The Annals of Statistics*, 50(2):1170–1192.

Zhao, A. and Ding, P. (2022b). Regression-based causal inference with factorial experiments: estimands, model specifications and design-based properties. *Biometrika*, 109(3):799–815.

Supplementary Material

A Additional technical results

A.1 Explicit forms of the variance and covariance

In Theorem 4.1, we give a compact form of asymptotic variance and covariance for the reweighting estimators. In this section, we provide more explicit forms, which are more straightforward and interpretable.

For a reweighting regime Γ and $* \in \{\text{ht}, \text{haj}\}$, for a fixed treatment and exposure mapping level (a, s, h), the asymptotic variance of $\hat{Y}^*(a, s, h; \Gamma)$ is

$$\begin{aligned}
& \text{avar}\{\hat{Y}^*(a,s,h;\Gamma)\} \\
= \; & n^{-2}\sum_{i=1}^{n} \frac{1-\pi_i(a,s,h)}{\pi_i(a,s,h)}\{\gamma_i(a,s,h)Y_i^*(a,s,h)\}^2 \\
& + n^{-2}\sum_{i=1}^{n}\sum_{j\neq i} \frac{\pi_{ij}(a,s,h;a,s,h)-\pi_i(a,s,h)\pi_j(a,s,h)}{\pi_i(a,s,h)\pi_j(a,s,h)}\gamma_i(a,s,h)\gamma_j(a,s,h)Y_i^*(a,s,h)Y_j^*(a,s,h) \\
= \; & n^{-2}\boldsymbol{Y}^*(a,s,h)^{\mathrm{T}}\boldsymbol{\Gamma}(a,s,h)\boldsymbol{\Lambda}(a,s,h)\boldsymbol{\Gamma}(a,s,h)\boldsymbol{Y}^*(a,s,h).
\end{aligned}$$

Similarly, for a pair (a, s, h) and (a', s', h'), the asymptotic covariance is

$$\begin{aligned}
& \text{acov}\{\hat{Y}^*(a,s,h;\Gamma), \hat{Y}^*(a',s',h';\Gamma)\} \\
= \; & -n^{-2}\sum_{i=1}^{n} \gamma_i(a,s,h)\gamma_i(a',s',h')Y_i^*(a,s,h)Y_i^*(a',s',h') \\
& + n^{-2}\sum_{i=1}^{n}\sum_{j\neq i} \frac{\pi_{ij}(a,s,h;a',s',h')-\pi_i(a,s,h)\pi_j(a',s',h')}{\pi_i(a,s,h)\pi_j(a',s',h')}\gamma_i(a,s,h)\gamma_j(a',s',h')Y_i^*(a,s,h)Y_j^*(a',s',h') \\
= \; & n^{-2}\boldsymbol{Y}^*(a,s,h)^{\mathrm{T}}\boldsymbol{\Gamma}(a,s,h)\boldsymbol{\Lambda}(a,s,h;a',s',h')\boldsymbol{\Gamma}(a',s',h')\boldsymbol{Y}^*(a',s',h').
\end{aligned}$$

A.2 Partially saturated regression estimator

The inverse probability weighting estimators introduced earlier represent a nonparametric approach, as they estimate all eight cell-specific means corresponding to $(a, s, h) \in \{0, 1\}^3$. Depending on the sample sizes available within each cell, this nonparametric strategy may be challenging. An alternative is to come up with regression models that restrict certain two-way or three-way interactions among A, S, and H. Under such restrictions, we can pool information across (a, s, h) groups and estimate the resulting estimands via an ordinal least squares regression. Although we do not pursue this modeling approach in our empirical analysis, it provides a useful complementary perspective.

Relatedly, it is also useful to consider a model-assisted approach to estimation and inference. While our main estimands are defined without imposing outcome models, working with a partially saturated regression can help recover point estimates and variance estimators in a more stable manner, in line with the model-assisted framework of Zhao and Ding (2022a).

Motivated by these considerations, the definition of (A, S, H) is reminiscent of factorial regressions in a 2^3 factorial experiment (Wu and Hamada, 2011; Zhao and Ding, 2022b). Based on this connection, we can study the main factorial effects of A, S, and H, as well as selected interaction effects. We explore this perspective and provide a causal interpretation for these factorial effects.

Concretely, we consider the following partially saturated regression:

$$Y_i \sim 1 + A_i + \tilde{S}_i + \tilde{H}_i + A_i\tilde{S}_i + A_i\tilde{H}_i, \tag{2}$$

where $\tilde{S}_i$ and $\tilde{H}_i$ are the centered versions of S_i and H_i, defined by subtracting their population means $\pi_{S,i}$ and $\pi_{H,i}$:

$$\tilde{S}_i = S_i - \pi_{S,i}, \quad \tilde{H}_i = H_i - \pi_{H,i}.$$

These population means depend on the geographic network and can be computed or simulated. This regression formulation does not include the three-way interaction among $(A, \tilde{S}, \tilde{H})$ nor the two-way interaction between $\tilde{S}$ and $\tilde{H}$, and therefore can be viewed as pooling information across the corresponding (a, s, h) cells.

In the special case where both stages of the randomized saturation design use independent Bernoulli assignments, (A_i, S_i, H_i) are mutually independent and the A_i's are identically distributed. Under this setting, the following identification results hold.

Theorem A.1. *The limit of the coefficients from the partially saturated regression (2) is* $(\beta_0, \beta_A, \beta_S, \beta_H, \beta_{AS}, \beta_{AH})^{\mathrm{T}}$, *where*

$$\begin{aligned}
\beta_A &= n^{-1}\sum_{i=1}^{n} E\{Y_i(1, S_i, H_i) - Y_i(0, S_i, H_i)\} \;=\; \text{DE}, \\
\beta_S &= \frac{n^{-1}\sum_{i=1}^{n} \pi_{S,i}(1-\pi_{S,i}) E\{Y_i(0,1,H_i) - Y_i(0,0,H_i)\}}{n^{-1}\sum_{i=1}^{n} \pi_{S,i}(1-\pi_{S,i})}, \\
\beta_H &= \frac{n^{-1}\sum_{i=1}^{n} \pi_{H,i}(1-\pi_{H,i}) E\{Y_i(0,S_i,1) - Y_i(0,S_i,0)\}}{n^{-1}\sum_{i=1}^{n} \pi_{H,i}(1-\pi_{H,i})}, \\
\beta_{AS} &= \frac{n^{-1}\sum_{i=1}^{n} \pi_{S,i}(1-\pi_{S,i}) E\{Y_i(1,1,H_i) - Y_i(1,0,H_i) - Y_i(0,1,H_i) + Y_i(0,0,H_i)\}}{n^{-1}\sum_{i=1}^{n} \pi_{S,i}(1-\pi_{S,i})}, \\
\beta_{AH} &= \frac{n^{-1}\sum_{i=1}^{n} \pi_{H,i}(1-\pi_{H,i}) E\{Y_i(1,S_i,1) - Y_i(1,S_i,0) - Y_i(0,S_i,1) + Y_i(0,S_i,0)\}}{n^{-1}\sum_{i=1}^{n} \pi_{H,i}(1-\pi_{H,i})},
\end{aligned}$$

where $\pi_A = \text{pr}(A_i = 1)$ *is the marginal probability of being assigned to treatment, given by the weighted average of the high and low saturation probabilities.*

From Theorem A.1, we obtain five effects. The coefficient β_A recovers exactly the direct effect defined in (1). The coefficients β_S and β_H do not recover the effects we defined earlier; isntead, they correspond to a reweighted average of the indirect effects with weights $\pi_{S,i}(1-\pi_{S,i})$ and $\pi_{H,i}(1-\pi_{H,i})$, respectively, known as overlap weights in the literature (Li et al., 2018). These overlap indirect effects target the subpopulation with good overlap in the corresponding exposure level.

The second-order interaction terms yield two additional estimands. They correspond to overlap-weighted versions of the marginal interaction factorial effects. For example, the coefficient β_{AS} is an overlap-weighted version of the interaction effect $E\{Y_i(1,1,H_i) - Y_i(1,0,H_i) - Y_i(0,1,H_i) + Y_i(0,0,H_i)\}$, which measures how the direct effect varies across different levels of S while marginalizing over H.

B Proof

B.1 Proof of Theorem 4.1

The theorem largely follows from existing results in complex randomized experiments, such as Aronow and Samii (2017); Leung (2022); Mukerjee et al. (2018), by defining pseudo potential outcomes using the weighting matrix Γ and the original/centered potential outcomes. For demonstration, we showcase the derivation of the asymptotic variance of the Horvitz–Thompson estimator for the average of potential outcomes at level (a, s, h), which is given by

$$\begin{aligned} & \operatorname{var}\left\{ n^{-1} \sum_{i=1}^{n} \frac{\mathbb{I}_i(a,s,h) \cdot \gamma_i(a,s,h)}{\pi_i(a,s,h)} Y_i \right\} \\ = \; & n^{-2} \sum_{i=1}^{n} \frac{\operatorname{var}\{\mathbb{I}_i(a,s,h)\} \gamma_i(a,s,h)^2 Y_i(a,s,h)^2}{\pi_i(a,s,h)^2} \\ & + n^{-2} \sum_{j \neq i} \frac{\operatorname{cov}\{\mathbb{I}_i(a,s,h), \mathbb{I}_j(a,s,h)\} \cdot \gamma_i(a,s,h) \gamma_j(a,s,h) Y_i(a,s,h) Y_j(a,s,h)}{\pi_i(a,s,h) \pi_j(a,s,h)}. \end{aligned}$$

Using the variance formula for a Bernoulli variable, we have

$$\operatorname{var}\{\mathbb{I}_i(a,s,h)\} = \pi_i(a,s,h)\{1 - \pi_i(a,s,h)\}. \tag{3}$$

Meanwhile, we can compute the covariance term between units i and j based on the definition:

$$\begin{aligned} \operatorname{cov}\{\mathbb{I}_i(a,s,h), \mathbb{I}_j(a,s,h)\} &= E\{\mathbb{I}_i(a,s,h)\mathbb{I}_j(a,s,h)\} - E\{\mathbb{I}_i(a,s,h)\} E\{\mathbb{I}_j(a,s,h)\} \\ &= \pi_{ij}(a,s,h;a,s,h) - \pi_i(a,s,h)\pi_j(a,s,h). \end{aligned} \tag{4}$$

Equations (3) and (4) together lead to the definition of $\Lambda_{(i,j)}(a,s,h)$ and also complete the variance computation for the Horvitz–Thompson estimator.

We now derive the asymptotic variance of the Hájek estimator

$$\hat{Y}^{\text{haj}}(a,s,h;\Gamma) = \frac{n^{-1} \sum_{i=1}^{n} \frac{\mathbb{I}_i(a,s,h)\gamma_i(a,s,h)}{\pi_i(a,s,h)} Y_i}{n^{-1} \sum_{i=1}^{n} \frac{\mathbb{I}_i(a,s,h)\gamma_i(a,s,h)}{\pi_i(a,s,h)} \Big/ n^{-1} \sum_{i=1}^{n} \gamma_i(a,s,h)}.$$

Denote the numerator and denominator as:

$$A_n(a,s,h) = n^{-1} \sum_{i=1}^{n} \frac{\mathbb{I}_i(a,s,h)\,\gamma_i(a,s,h)}{\pi_i(a,s,h)} Y_i(a,s,h),$$

$$B_n(a,s,h) \quad = \quad n^{-1}\sum_{i=1}^{n}\frac{\mathbb{I}_i(a,s,h)\,\gamma_i(a,s,h)}{\pi_i(a,s,h)}\Big/ n^{-1}\sum_{i=1}^{n}\gamma_i(a,s,h).$$

Accordingly, we have $E\{A_n(a,s,h)\} = n^{-1}\sum_{i=1}^{n}\gamma_i(a,s,h)Y_i(a,s,h)$ and $E\{B_n(a,s,h)\} = 1$. Denote the corresponding target estimand under weighting regime Γ as $\bar{Y}(a,s,h;\Gamma) = n^{-1}\sum_{i=1}^{n}\gamma_i(a,s,h)Y_i(a,s,h)$, by Taylor expansion of the ratio A_n/B_n around $\bar{Y}(a,s,h;\Gamma) = E(A_n)/E(B_n)$, we have

$$\hat{Y}^{\text{haj}}(a,s,h;\Gamma) - \bar{Y}(a,s,h;\Gamma) \quad = \quad \frac{1}{E(B_n)}\left[A_n - E(A_n) - \frac{E(A_n)}{E(B_n)}\{B_n - E(B_n)\}\right] + o_p(n^{-1/2}).$$

The main term on the right hand side is equal to

$$\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\frac{\mathbb{I}_i(a,s,h)-\pi_i(a,s,h)}{\pi_i(a,s,h)}\gamma_i(a,s,h)Y_i(a,s,h) \\
& -n^{-1}\sum_{i=1}^{n}\frac{\mathbb{I}_i(a,s,h)-\pi_i(a,s,h)}{\pi_i(a,s,h)}\gamma_i(a,s,h)\cdot\frac{n^{-1}\sum_{i=1}^{n}\gamma_i(a,s,h)Y_i(a,s,h)}{n^{-1}\sum_{i=1}^{n}\gamma_i(a,s,h)} \\
= \quad & n^{-1}\sum_{i=1}^{n}\frac{\mathbb{I}_i(a,s,h)-\pi_i(a,s,h)}{\pi_i(a,s,h)}\gamma_i(a,s,h)Y_i^{\text{haj}}(a,s,h),
\end{aligned}$$

and therefore we have

$$\hat{Y}^{\text{haj}}(a,s,h;\Gamma) - \bar{Y}(a,s,h;\Gamma) \quad = \quad n^{-1}\sum_{i=1}^{n}\frac{\mathbb{I}_i(a,s,h)-\pi_i(a,s,h)}{\pi_i(a,s,h)}\gamma_i(a,s,h)Y_i^{\text{haj}}(a,s,h) + o_p(n^{-1/2}),$$

whose first order behaves like a Horvitz–Thompson estimator applied to the centered potential outcomes $Y_i^{\text{haj}}(a,s,h)$. Using the previous derivation on asymptotic variance of the Horvitz–Thompson estimator, we have the results in Theorem 4.1. □

B.2 Proof of Theorem 4.2 and Corollary 5.1

The consistency result of the point estimators and variance estimators follow from Aronow and Samii (2017), by the law of large numbers for weakly dependent random variables under the bounded degree and dependence assumptions. The asymptotic normality follows from the central limit theorem for m-dependent arrays (Chen and Shao, 2004), where the dependence structure is determined by the network topology. □

B.3 Proof of Theorem A.1

The OLS regression estimators are given by $\hat{\beta} = \Omega_{XX}^{-1}\Omega_{XY}$, where

$$\Omega_{XX} = n^{-1}\begin{pmatrix} n & \sum_{i=1}^n A_i & \sum_{i=1}^n \tilde{S}_i & \sum_{i=1}^n \tilde{H}_i & \sum_{i=1}^n A_i\tilde{S}_i & \sum_{i=1}^n A_i\tilde{H}_i \\ \sum_{i=1}^n A_i & \sum_{i=1}^n A_i^2 & \sum_{i=1}^n A_i\tilde{S}_i & \sum_{i=1}^n A_i\tilde{H}_i & \sum_{i=1}^n A_i^2\tilde{S}_i & \sum_{i=1}^n A_i^2\tilde{H}_i \\ \sum_{i=1}^n \tilde{S}_i & \sum_{i=1}^n A_i\tilde{S}_i & \sum_{i=1}^n \tilde{S}_i^2 & \sum_{i=1}^n \tilde{S}_i\tilde{H}_i & \sum_{i=1}^n A_i\tilde{S}_i^2 & \sum_{i=1}^n A_i\tilde{S}_i\tilde{H}_i \\ \sum_{i=1}^n \tilde{H}_i & \sum_{i=1}^n A_i\tilde{H}_i & \sum_{i=1}^n \tilde{S}_i\tilde{H}_i & \sum_{i=1}^n \tilde{H}_i^2 & \sum_{i=1}^n A_i\tilde{S}_i\tilde{H}_i & \sum_{i=1}^n A_i\tilde{H}_i^2 \\ \sum_{i=1}^n A_i\tilde{S}_i & \sum_{i=1}^n A_i^2\tilde{S}_i & \sum_{i=1}^n A_i\tilde{S}_i^2 & \sum_{i=1}^n A_i\tilde{S}_i\tilde{H}_i & \sum_{i=1}^n A_i^2\tilde{S}_i^2 & \sum_{i=1}^n A_i^2\tilde{S}_i\tilde{H}_i \\ \sum_{i=1}^n A_i\tilde{H}_i & \sum_{i=1}^n A_i^2\tilde{H}_i & \sum_{i=1}^n A_i\tilde{S}_i\tilde{H}_i & \sum_{i=1}^n A_i\tilde{H}_i^2 & \sum_{i=1}^n A_i^2\tilde{S}_i\tilde{H}_i & \sum_{i=1}^n A_i^2\tilde{H}_i^2 \end{pmatrix}$$

and

$$\Omega_{XY} = n^{-1}\left(\sum_{i=1}^n Y_i \quad \sum_{i=1}^n A_iY_i \quad \sum_{i=1}^n \tilde{S}_iY_i \quad \sum_{i=1}^n \tilde{H}_iY_i \quad \sum_{i=1}^n A_i\tilde{S}_iY_i \quad \sum_{i=1}^n A_i\tilde{H}_iY_i\right)^{\mathrm{T}}.$$

Due to the demeaning step and the independence among A_i, $\tilde{S}_i$, and $\tilde{H}_i$, some off-diagonal values of the expectation of the Ω_{XX} matrix are zero. The diagonal values of Ω_{XX} are:

$$\begin{aligned} E\left(n^{-1}\sum_{i=1}^n A_i^2\right) &= \pi_A, \\ E\left(n^{-1}\sum_{i=1}^n \tilde{S}_i^2\right) &= n^{-1}\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i}), \\ E\left(n^{-1}\sum_{i=1}^n \tilde{H}_i^2\right) &= n^{-1}\sum_{i=1}^n \pi_{H,i}(1-\pi_{H,i}), \\ E\left(n^{-1}\sum_{i=1}^n A_i^2\tilde{S}_i^2\right) &= \pi_A n^{-1}\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i}), \\ E\left(n^{-1}\sum_{i=1}^n A_i^2\tilde{H}_i^2\right) &= \pi_A n^{-1}\sum_{i=1}^n \pi_{H,i}(1-\pi_{H,i}). \end{aligned}$$

The non-zero off-diagonal terms are

$$\begin{aligned} E\left(n^{-1}\sum_{i=1}^n A_i\tilde{S}_i^2\right) &= \pi_A\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i}), \\ E\left(n^{-1}\sum_{i=1}^n A_i\tilde{S}_i^2\right) &= \pi_A\sum_{i=1}^n \pi_{H,i}(1-\pi_{H,i}). \end{aligned}$$

Therefore,

$$E(\Omega_{XX}) = n^{-1}\begin{pmatrix} n & n\pi_A & 0 & 0 & 0 & 0 \\ n\pi_A & n\pi_A & 0 & 0 & 0 & 0 \\ 0 & 0 & \sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i}) & 0 & \pi_A \sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i}) & 0 \\ 0 & 0 & 0 & \sum_{i=1}^n \pi_{H,i}(1-\pi_{H,i}) & 0 & \pi_A \sum_{i=1}^n \pi_{H,i}(1-\pi_{H,i}) \\ 0 & 0 & \pi_A \sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i}) & 0 & \pi_A \sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i}) & 0 \\ 0 & 0 & 0 & \pi_A \sum_{i=1}^n \pi_{H,i}(1-\pi_{H,i}) & 0 & \pi_A \sum_{i=1}^n \pi_{H,i}(1-\pi_{H,i}) \end{pmatrix}.$$

Taking the inverse, we have

$$E(\Omega_{XX})^{-1} = n\begin{pmatrix} \frac{(1-\pi_A)^{-1}}{n} & \frac{-(1-\pi_A)^{-1}}{n} & 0 & 0 & 0 & 0 \\ \frac{-(1-\pi_A)^{-1}}{n} & \frac{\pi_A^{-1}(1-\pi_A)^{-1}}{n} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{(1-\pi_A)^{-1}}{\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i})} & 0 & \frac{-(1-\pi_A)^{-1}}{\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i})} & 0 \\ 0 & 0 & 0 & \frac{(1-\pi_A)^{-1}}{\sum_{i=1}^n \pi_{H,i}(1-\pi_{H,i})} & 0 & \frac{-(1-\pi_A)^{-1}}{\sum_{i=1}^n \pi_{H,i}(1-\pi_{H,i})} \\ 0 & 0 & \frac{-(1-\pi_A)^{-1}}{\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i})} & 0 & \frac{\pi_A^{-1}(1-\pi_A)^{-1}}{\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i})} & 0 \\ 0 & 0 & 0 & \frac{-(1-\pi_A)^{-1}}{\sum_{i=1}^n \pi_{H,i}(1-\pi_{H,i})} & 0 & \frac{\pi_A^{-1}(1-\pi_A)^{-1}}{\sum_{i=1}^n \pi_{H,i}(1-\pi_{H,i})} \end{pmatrix}.$$

Now we can compute the population mean of Ω_{XY}. For the intercept coordinate, we have

$$E\left(n^{-1}\sum_{i=1}^n Y_i\right) = n^{-1}\sum_{i=1}^n \left[\pi_A E\{Y_i(1,S_i,H_i)\} + (1-\pi_A)E\{Y_i(0,S_i,H_i)\}\right].$$

For the cross term between Y_i and A_i, we have

$$E\left(n^{-1}\sum_{i=1}^n A_i Y_i\right) = \pi_A n^{-1}\sum_{i=1}^n E\{Y_i(1,S_i,H_i)\}.$$

For the cross term between Y_i and $\tilde{S}_i$, we have

$$\begin{aligned} E\left(n^{-1}\sum_{i=1}^n \tilde{S}_i Y_i\right) &= \pi_A n^{-1}\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i})E\{Y_i(1,1,H_i)\} + (1-\pi_A)n^{-1}\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i})E\{Y_i(0,1,H_i)\} \\ &\quad -\pi_A n^{-1}\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i})E\{Y_i(1,0,H_i)\} - (1-\pi_A)n^{-1}\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i})E\{Y_i(0,0,H_i)\} \\ &= \pi_A n^{-1}\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i})E\{Y_i(1,1,H_i) - Y_i(1,0,H_i)\} \\ &\quad +(1-\pi_A)n^{-1}\sum_{i=1}^n \pi_{S,i}(1-\pi_{S,i})E\{Y_i(0,1,H_i) - Y_i(0,0,H_i)\}. \end{aligned}$$

Similarly,

$$E\left(n^{-1}\sum_{i=1}^n \tilde{H}_i Y_i\right) = \pi_A n^{-1}\sum_{i=1}^n \pi_{H,i}(1-\pi_{H,i})E\{Y_i(1,S_i,1) - Y_i(1,S_i,0)\}$$

$$+(1-\pi_A)n^{-1}\sum_{i=1}^{n}\pi_{H,i}(1-\pi_{H,i})E\{Y_i(0,S_i,1)-Y_i(0,S_i,0)\}.$$

For the rest of the two terms, we have

$$\begin{aligned}
E\left(n^{-1}\sum_{i=1}^{n}A_i\tilde{S}_iY_i\right) &= \pi_A n^{-1}\sum_{i=1}^{n}\pi_{S,i}(1-\pi_{S,i})E\{Y_i(1,1,H_i)-Y_i(1,0,H_i)\},\\
E\left(n^{-1}\sum_{i=1}^{n}A_i\tilde{H}_iY_i\right) &= \pi_A n^{-1}\sum_{i=1}^{n}\pi_{H,i}(1-\pi_{H,i})E\{Y_i(1,S_i,1)-Y_i(1,S_i,0)\}.
\end{aligned}$$

Combining all the computations above, we conclude the results. □