Causal Inference with Interference and Noncompliance in Two-Stage Randomized Experiments

Kosuke Imai† Zhichao Jiang‡ Anup Malani§

First Draft: April 19, 2018
This Draft: February 18, 2020

Abstract

In many social science experiments, subjects often interact with each other and as a result one unit’s treatment influences the outcome of another unit. Over the last decade, a significant progress has been made towards causal inference in the presence of such interference between units. Researchers have shown that the two-stage randomization of treatment assignment enables the identification of average direct and spillover effects. However, much of the literature has assumed perfect compliance with treatment assignment. In this paper, we establish the nonparametric identification of the complier average direct and spillover effects in two-stage randomized experiments with interference and noncompliance. In particular, we consider the spillover effect of the treatment assignment on the treatment receipt as well as the spillover effect of the treatment receipt on the outcome. We propose consistent estimators and derive their randomization-based variances under the stratified interference assumption. We also prove the exact relationship between the proposed randomization-based estimators and the popular two-stage least squares estimators. Our methodology is motivated by and applied to the randomized evaluation of the India’s National Health Insurance Program (RSBY), where we find some evidence of spillover effects on both treatment receipt and outcome. The proposed methods are implemented via an open-source software package.

Keywords: complier average causal effects, encouragement design, program evaluation, randomization inference, spillover effects, two-stage least squares

*The proposed methodology is implemented via an open-source software package experiment [Imai and Jiang, 2018], which is available at https://cran.r-project.org/package=experiment. We thank Naoki Egami for helpful comments. We thank the Editor, Associate Editor, and three reviewers for careful reading and many constructive comments.

†Professor of Government and of Statistics, Harvard University. 1737 Cambridge Street, Institute for Quantitative Social Science, Cambridge MA 02138. Email: imai@harvard.edu URL: https://imai.fas.harvard.edu

‡Assistant Professor, Department of Biostatistics and Epidemiology, University of Massachusetts, Amherst MA 01003. Email: zhichaojiang@umass.edu

§Lee and Brena Freeman Professor, University of Chicago Law School and Pritzker School of Medicine, Chicago IL 60637, and National Bureau of Economic Research, Cambridge MA 02138.
1 Introduction

Early methodological research on causal inference has assumed no interference between units (e.g., Neyman, 1923; Fisher, 1935; Holland, 1986; Rubin, 1990). That is, spillover effects are assumed to be absent. In many social science experiments, however, subjects often interact with each other and as a result one unit’s treatment influences the outcome of another unit. Over the last decade, a significant progress has been made towards causal inference in the presence of such interference between units (e.g., Sobel, 2006; Rosenbaum, 2007; Hudgens and Halloran, 2008; Tchetgen Tchetgen and VanderWeele, 2010; Aronow, 2012; Vanderweele et al., 2013; Liu and Hudgens, 2014; Hong, 2015; Forastiere et al., 2016; Aronow and Samii, 2017; Athey et al., 2017; Basse and Feller, 2018).

Much of this literature, however, has not addressed another common feature of social science experiments where some control units decide to take the treatment while others in the treatment group refuse to receive one. Such noncompliance often occurs in these experiments because for ethical and logistical reasons, researchers typically cannot force experimental subjects to adhere to experimental protocol. The existing methods either assume perfect compliance with treatment assignment or focus on intention-to-treat (ITT) analyses by ignoring the information about actual receipt of treatment. Unfortunately, the ITT analysis is unable to tell, for example, whether a small causal effect arises due to ineffective treatment or low compliance. While researchers have developed methods to deal with noncompliance (e.g., Angrist et al., 1996), they are based on the assumption of no interference between units. This assumption may be unrealistic since there are multiple ways in which spillover effects could arise: for example, one unit’s treatment assignment may influence another unit’s decision to receive the treatment, whereas it is also possible that one’s treatment receipt affects the outcomes of other units.

In this paper, we show how to analyze two-stage randomized experiments with both interference and noncompliance (Section 3). In an influential paper, Hudgens and Halloran, 2008...
loran (2008) propose two-stage randomized experiments as a general approach to causal inference with interference. We extend their framework so that it is applicable even in the presence of two-sided noncompliance. In particular, we define the complier average direct and spillover effects, propose consistent estimators, and derive their variances under the stratified interference assumption. We follow Hudgens and Halloran (2008) by referring to the effect of one’s own treatment as a direct effect and the effect of another unit’s treatment as a spillover effect. In a closely related working paper, Kang and Imbens (2016) also analyze two-stage randomized experiments with interference and noncompliance. We consider a more general pattern of interference by allowing for the spillover effect of the treatment assignment on the treatment receipt as well as the spillover effect of the treatment receipt on the outcome. The proposed methods are implemented via an open-source software package, experiment (Imai and Jiang, 2018), which is available at https://cran.r-project.org/package=experiment.

Finally, we prove the exact relationships between the proposed randomization-based estimators and the popular two-stage least squares estimators as well as those between their corresponding variance estimators. Our results build upon and extend the work of Basse and Feller (2018) to the case with noncompliance. We also conduct simulation studies to investigate the finite sample performance of the confidence intervals based on the proposed variance estimators (see Appendix D).

The proposed methodology is motivated by our own randomized evaluation of the Indian Health Insurance Scheme (known by the acronym RSBY), a study that employed the two-stage randomized design. In Section 2, we briefly describe the background and experimental design of this study. In Section 4, we apply the proposed methodology to this study. We present some evidence for the existence of positive spillover effects of treatment assignment on the enrollment in the RSBY. In addition, we estimate the complier average direct effect to be positive under the “low” treatment assignment mechanism. This suggests
that people may be more likely to visit a hospital when fewer households in their village are encouraged to enroll in the insurance program. Finally, Section 5 concludes.

2 A Motivating Empirical Application

In this section, we describe the randomized evaluation of the Indian health insurance program, which serves as our motivating empirical application. We provide a brief background of the evaluation and introduce its experimental design. For a more detailed description of the design, see the pre-analysis plan posted on the American Economic Association’s Registry at https://www.socialscienceregistry.org/trials/1793

2.1 Randomized Evaluation of the Indian Health Insurance Program

Each year, 150 million people worldwide face financial catastrophe due to spending on health. According to a 2010 study, more than one third of them live in India (Shahrawat and Rao, 2011). Almost 63 million Indians fall below the poverty line (BPL) due to health spending (Berman et al., 2010). In 2008, the Indian government introduced its first national, public health insurance scheme, Rastriya Swasthya Bima Yojana (RSBY), to address the problem. Its aim was to provide coverage for hospitalization to its BPL population, comprising roughly 250 million persons.

RSBY provides access to an insurance plan that covers inpatient hospital care for up to five members of each household. The plan covers all pre-existing diseases and there is no age limit of the beneficiaries. The rates of most surgical procedures are fixed by the government. Beneficiaries can obtain treatment at any hospital empaneled in the RSBY network. The insurance scheme is cashless, with the plan paying providers directly rather than reimbursing beneficiaries for expenses. The plan also covers INR 100 (or approximately USD 1.53) of transportation costs per hospitalization. The coverage lasts one year starting the month after the first enrollment in a particular district, but is often extended without cost to beneficiaries. The insurance plan is provided by private insurance compa-
nies, but the premium are paid by the government. In Karnataka, the state in which the randomized evaluation was conducted, premiums were roughly INR 200 (USD 3.07) per year during the study. Households only have to pay INR 30 (USD 0.46) per year user fee to obtain an insurance card. There are no deductibles or co-payments and there is an annual cap of INR 30,000 (USD 460) per household.

We conducted a randomized controlled trial to determine whether RSBY increases access to hospitalization, and thus health, and reduced impoverishment due to high medical expenses. The findings are policy-relevant because the Indian government has announced a new scheme called the National Health Protection Scheme (NHPS) that seeks to build on RSBY to provide coverage for nearly 500 million Indians, but has not yet decided its design or how much to fund it.

In this evaluation, spillover effects are of concern because formal insurance may crowd out informal insurance, which is a substitute method of smoothing health care shocks (e.g., Jowett, 2003; Lin et al., 2014). That is, the enrollment in RSBY by one household may depend on the treatment assignment of other households. In addition, we also must address noncompliance because some households in the treatment group decided not to enroll in RSBY while others in the control group managed to join the insurance program.

2.2 Experimental Design

Our evaluation study is based on a total of 11,089 above poverty line (APL) households in two districts of Karnataka State who had no pre-existing health insurance coverage and lived within 25 km of an RSBY empaneled hospital. We selected APL households because they are not otherwise eligible for RSBY, but are candidates for any expansion of RSBY. The two districts were Gulbarga and Mysore, which are economically and culturally representative of central and southern India, respectively. We required proximity to a hospital as hospital insurance has little value if there is no local hospital at which to use the insurance.

As shown in Table [1] we employed a two-stage randomized design to study both di-
rect and spillover effects of RSBY. In the first stage, randomly selected 219 villages were assigned to the “High” treatment assignment mechanism whereas the rest of villages were assigned to the “Low” treatment assignment mechanism. In the second stage, 80% of 5714 households under the “High” assignment mechanism are completely randomly assigned to the treatment condition, while the rest of households were assigned to the control group. In contrast, under the “Low” assignment mechanism, 40% of households within a cluster are completely randomly assigned to the treatment condition. The households in the treatment group are given RSBY essentially for free, whereas some households in the control group were able to buy RSBY at the government price of roughly INR 200.

Households were informed of the assigned treatment conditions and were given the opportunities to enroll in RSBY from April to May, 2015. Approximately 18 months later, we carried out a post-treatment survey and measured a variety of outcomes. Policy makers are interested in the health and financial effects of RSBY. To evaluate the efficacy of RSBY, we must estimate the effects of actual treatment receipt as well as the intention-to-treat effects because some households in the treatment group may not enroll in RSBY while others in the control group may do so.

### 3 The Proposed Methodology

In this section, we first review the intention-to-treat (ITT) analysis of two-stage randomized experiments proposed by Hudgens and Halloran (2008) and others. We then introduce a new causal quantity of interest, the complier average direct effect (CADE), present a non-parametric identification result, and propose a consistent estimator. We further consider the identification and inference of the CADE under the assumption of stratified interference,
and derive the randomization-based variance of the proposed estimator. We also establish the direct connections between these randomization-based estimators and the two-stage least squares estimators. Finally, we present analogous results for another new causal quantity, the complier average spillover effect (CASE), in Appendix A.

3.1 Two-Stage Randomized Experiments

We consider a two-stage randomized experiment (Hudgens and Halloran, 2008) with a total of $N$ units and $J$ clusters where each unit belongs to one of the clusters. If we use $n_j$ to denote the number of units in cluster $j$, we have $N = \sum_{j=1}^{J} n_j$. In a two-stage randomized experiment, we first randomly assign each cluster to one of the treatment assignment mechanisms, which in turn assigns different proportions of units within each cluster to the treatment condition. For the sake of simplicity, we consider two assignment mechanisms indicated by $A_j \in \{0, 1\}$ where $A_j = 1$ ($A_j = 0$) indicates that a high (low) proportion of units are assigned to the treatment within cluster $j$. In our application, $A_j = 1$ corresponds to the treatment assignment probability of 80%, whereas $A_j = 0$ represents 40%. We assume complete randomization, in which a total of $J_a$ clusters are assigned to the assignment mechanism $A_j = a$ for $a = 0, 1$ with $J_0 + J_1 = J$. Finally, $A = (A_1, A_2, \ldots, A_J)$ denotes the vector of treatment assignment mechanisms for all clusters.

The second stage of randomization concerns the treatment assignment for each unit within cluster $j$ based on the assignment mechanism $A_j$. Let $Z_{ij}$ be the binary treatment assignment variable for unit $i$ in cluster $j$ where $Z_{ij} = 1$ ($Z_{ij} = 0$) implies that the unit is assigned to the treatment (control) condition. Let $Z_j = (Z_{1j}, \ldots, Z_{nj_j})$ denote the vector of assigned treatments for the $n_j$ units in cluster $j$ and $Pr(Z_j = z_j \mid A_j = a)$ represent the distribution of the treatment assignment vector when cluster $j$ is assigned to the assignment mechanism $A_j = a$. We assume the complete randomization such that a total of $n_{jz}$ units in cluster $j$ are assigned to the treatment condition $z$ for $z = 0, 1$, where $n_{j1} + n_{j0} = n_j$.

Assumption 1 (Two-Stage Randomization)
1. Complete randomization of treatment assignment mechanism at the cluster level:

\[
\Pr(A = a) = \frac{1}{(J_j)}
\]

for all \(a\) such that \(1_J^T a = J_1\) where \(1_J\) is the \(J\) dimensional vector of ones.

2. Complete randomization of treatment assignment within each cluster:

\[
\Pr(Z_j = z \mid A_j = a) = \frac{1}{(n_j)}
\]

for all \(z\) such that \(1_{n_j}^T z = n_{j1}\).

Following the literature, we adopt the finite population framework, in which potential outcomes are treated as constants and randomness comes from treatment assignment alone. We consider two-stage randomized experiments with noncompliance, in which the actual receipt of treatment may differ from the treatment assignment. Let \(D_{ij}\) be the treatment receipt for unit \(i\) in cluster \(j\) and \(D_j = (D_{1j}, \ldots, D_{nj})\) be the vector of treatment receipts for the \(n_j\) units in the cluster. The outcome variable \(Y_{ij}\) is observed for each unit and \(Y_j = (Y_{1j}, \ldots, Y_{nj})\) denotes the vector of observed outcomes for the \(n_j\) units in cluster \(j\).

We use the potential outcomes framework of causal inference (e.g., [Neyman, 1923; Holland, 1986; Rubin, 1990]). For unit \(i\) in cluster \(j\), let \(D_{ij}(z)\) represent the potential value of treatment receipt, when the treatment assignment vector for all \(N\) units in the experiment equals \(z\). In addition, we use \(Y_{ij}(z;d)\) to denote the potential value of outcome, when the treatment assignment vector equals \(z\) and treatment receipt vector equals \(d\). Lastly, let \(Y_{ij}(z)\) represent the potential value of outcome when the treatment assignment vector equals \(z\), i.e., \(Y_{ij}(z) = Y_{ij}(z;D_{ij}(z))\). The observed values of treatment receipt and outcome are given by \(D_{ij} = D_{ij}(Z)\) and \(Y_{ij} = Y_{ij}(Z)\) where \(Z\) is the \(N\) dimensional vector of treatment assignment for all units. If there were no restriction on the pattern of interference, each unit has \(2^N\) potential values of treatment receipt and outcome, making identification infeasible. Hence, following the literature (e.g., [Hong and Raudenbush, 2006; Sobel, 2006; Hudgens and Halloran, 2008]), we only allow interference within each cluster.
Assumption 2 (Partial Interference)

\[ Y_{ij}(z) = Y_{ij}(z') \quad \text{and} \quad D_{ij}(z) = D_{ij}(z') \]

for all \( z \) and \( z' \) with \( z_j = z'_j \).

Assumption 2 implies that although the treatment receipt and outcome of a unit can be influenced by the treatment assignment of another unit within the same cluster, they cannot be affected by units in other clusters. This assumption substantially reduces the number of potential values of treatment receipt and outcome for each unit in cluster \( j \) from \( 2^N \) to \( 2^n_j \).

3.2 Intention-to-Treat Effects: A Review

We next review the previous results about the ITT analysis of two-stage randomized experiments under the partial interference assumption (Hudgens and Halloran, 2008). Our analysis differs from the existing ones in that we weight each unit equally instead of giving an equal weight to each cluster as done in the literature.

3.2.1 Causal Quantities of Interest

We begin by defining preliminary average quantities. First, we define the average potential value of treatment receipt for unit \( i \) in cluster \( j \) when the unit is assigned to the treatment condition \( z \) under the treatment assignment mechanism \( a \). We do so by averaging over the distribution of treatment assignments for the other units within the same cluster.

\[
\overline{D}_{ij}(z, a) = \sum_{Z_{i,j} \in Z_{-i,j}} D_{ij}(Z_{ij} = z, Z_{-i,j} = z_{-i,j}) \Pr(Z_{-i,j} = z_{-i,j} | Z_{ij} = z, A_j = a),
\]

where \( Z_{-i,j} = (Z_{1j}, \ldots, Z_{i-1,j}, Z_{i+1,j}, \ldots, Z_{nj,j}) \) represents the \( (n_j - 1) \) dimensional sub-vector of \( Z_j \) with the entry for unit \( i \) removed and \( Z_{-i,j} = \{(z_{1j}, \ldots, z_{i-1,j}, z_{i+1,j}, \ldots, z_{nj,j}) | z_{ip,j} \in \{0, 1\} \text{ for } i' = 1, \ldots, i-1, i+1, \ldots, n_j\} \) is the set of all possible values of the assignment vector \( Z_{-i,j} \). Similarly, we define the average potential outcome for unit \( i \) in cluster \( j \) as,

\[
\overline{Y}_{ij}(z, a) = \sum_{Z_{-i,j} \in Z_{-i,j}} Y_{ij}(Z_{ij} = z, Z_{-i,j} = z_{-i,j}) \Pr(Z_{-i,j} = z_{-i,j} | Z_{ij} = z, A_j = a).
\]
Given these unit-level average potential outcomes, we consider the cluster-level and population-level average potential values of the treatment receipt and outcome,

\[
\overline{D}_j(z, a) = \frac{1}{n_j} \sum_{i=1}^{n_j} D_{ij}(z, a), \quad \overline{D}(z, a) = \frac{1}{N} \sum_{j=1}^{J} n_j \overline{D}_j(z, a),
\]

\[
\overline{Y}_j(z, a) = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij}(z, a), \quad \overline{Y}(z, a) = \frac{1}{N} \sum_{j=1}^{J} n_j \overline{Y}_j(z, a).
\]

We define the ITT effects, starting with the average direct effect of treatment assignment on the treatment receipt and outcome under the treatment assignment mechanism \(a\), as

\[
\text{DED}_{ij}(a) = \overline{D}_{ij}(1, a) - \overline{D}_{ij}(0, a), \quad \text{DEY}_{ij}(a) = \overline{Y}_{ij}(1, a) - \overline{Y}_{ij}(0, a).
\]

where \(\text{DED}\) and \(\text{DEY}\) stand for the average direct effect on \(D\) and \(Y\), respectively. These parameters quantify how the treatment assignment of a unit may affect its treatment receipt and outcome by averaging the treatment assignments of other units within the same cluster under a specific assignment mechanism. Finally, averaging these unit-level quantities gives the following average direct effects of treatment assignment for each cluster and for the entire (finite) population,

\[
\text{DED}_j(a) = \frac{1}{n_j} \sum_{i=1}^{n_j} \text{DED}_{ij}(a), \quad \text{DED}(a) = \frac{1}{N} \sum_{j=1}^{J} n_j \text{DED}_j(a),
\]

\[
\text{DEY}_j(a) = \frac{1}{n_j} \sum_{i=1}^{n_j} \text{DEY}_{ij}(a), \quad \text{DEY}(a) = \frac{1}{N} \sum_{j=1}^{J} n_j \text{DEY}_j(a).
\]

Another quantity of interest is the spillover effect, which quantifies how one unit’s treatment receipt or outcome is affected by other units’ treatment assignments. Following Halloran and Struchiner (1995), we define the unit-level spillover effects on the treatment receipt and outcome as

\[
\text{SED}_{ij}(z) = \overline{D}_{ij}(z, 1) - \overline{D}_{ij}(z, 0), \quad \text{SEY}_{ij}(z) = \overline{Y}_{ij}(z, 1) - \overline{Y}_{ij}(z, 0),
\]

which compare the average potential values under two different assignment mechanisms, i.e., \(a = 1\) and \(a = 0\), while holding one’s treatment assignment at \(z\). We then define the
spillover effects on the treatment receipt and outcome at the cluster and population levels,

\[
\text{SED}_j(z) = \frac{1}{n_j} \sum_{i=1}^{n_j} \text{SED}_{ij}(z), \quad \text{SED}(z) = \frac{1}{N} \sum_{j=1}^{J} n_j \text{SED}_j(z).
\]

\[
\text{SEY}_j(z) = \frac{1}{n_j} \sum_{i=1}^{n_j} \text{SEY}_{ij}(z), \quad \text{SEY}(z) = \frac{1}{N} \sum_{j=1}^{J} n_j \text{SEY}_j(z).
\]

The quantities defined above differ from those introduced in the literature in that we equally weight each unit (see Basse and Feller, 2018). In contrast, Hudgens and Halloran (2008) give an equal weight to each cluster regardless of its size. While our analysis focuses on the individual-weighted estimands rather than cluster-weighted estimands, our method can be generalized to any weighting scheme, and as such the proofs in the supplementary appendix are based on general weights.

Finally, in actual policy implementations, the treatment assignment is typically based on a deterministic criteria rather than randomization, suggesting that the causal quantities discussed above may not be of direct interest to policy makers. Even in this situation, however, these causal quantities can provide some policy implications by telling us whether or not spillover effects exist at all. We discuss this issue in the context of our application (see Section 4) and consider a model-based approach to further address this point (see Appendix E).

### 3.2.2 Nonparametric Identification

Hudgens and Halloran (2008) establish the nonparametric identification of the ITT effects, which equally weight each cluster regardless of its size. Here, we present analogous results by weighting each unit equally as done above. Define the following quantities,

\[
\hat{D}(z, a) = \frac{1}{N} \sum_{j=1}^{J} n_j \hat{D}_j(z, a) I(A_j = a), \quad \hat{Y}(z, a) = \frac{1}{N} \sum_{j=1}^{J} n_j \hat{Y}_j(z, a) I(A_j = a),
\]

where

\[
\hat{D}_j(z, a) = \frac{\sum_{i=1}^{n_j} D_{ij} I(Z_{ij} = z)}{\sum_{i=1}^{n_j} I(Z_{ij} = z)}, \quad \hat{Y}_j(z, a) = \frac{\sum_{i=1}^{n_j} Y_{ij} I(Z_{ij} = z)}{\sum_{i=1}^{n_j} I(Z_{ij} = z)}.
\]

Then, we can obtain the unbiased estimators of the direct effects and the spillover effects.
THEOREM 1 (UNBIASED ESTIMATION OF THE ITT EFFECTS) Define the following estimators,

\[ \hat{\text{DED}}(a) = \hat{D}(1, a) - \hat{D}(0, a), \quad \hat{\text{SED}}(z) = \hat{D}(z, 1) - \hat{D}(z, 0), \]
\[ \hat{\text{DEY}}(a) = \hat{Y}(1, a) - \hat{Y}(0, a), \quad \hat{\text{SEY}}(z) = \hat{Y}(z, 1) - \hat{Y}(z, 0). \]

Under Assumptions 1 and 2, these estimators are unbiased for the ITT effects,

\[ \mathbb{E}\{\hat{\text{DED}}(a)\} = \text{DED}(a), \quad \mathbb{E}\{\hat{\text{SED}}(z)\} = \text{SED}(z), \]
\[ \mathbb{E}\{\hat{\text{DEY}}(a)\} = \text{DEY}(a), \quad \mathbb{E}\{\hat{\text{SEY}}(z)\} = \text{SEY}(z). \]

Proof is straightforward and hence omitted.

3.3 Complier Average Direct Effects

We now address the issue of noncompliance in the presence of interference between units. In a seminal paper, Angrist et al. (1996) show how to identify the complier average causal effect (CACE) in standard randomized experiments under the assumption of no interference. The CACE represents the average effect of treatment receipt among the compliers who would receive the treatment only when assigned to the treatment condition. Below, we introduce the complier average direct effect, which is a generalization of the CACE to settings with interference, and show how to nonparametrically identify and consistently estimate it using the data from two-stage randomized experiments.

3.3.1 Causal Quantity of Interest

We first generalize the definition of compliers to settings with interference between units. Under the assumption of no interference, compliers are those who receive the treatment only when assigned to the treatment condition. However, in the presence of partial interference, the treatment receipt is also affected by the treatment assignment of other units in the same cluster. Thus, the compliance status of a unit is a function of the treatment assignment of other units in the same cluster,

\[ C_{ij}(z_{-i,j}) = I\{D_{ij}(1, z_{-i,j}) = 1, D_{ij}(0, z_{-i,j}) = 0\}. \]

We consider a measure of compliance behavior for each unit by averaging over the distribution of treatment assignments of the other units within the same cluster under the
treatment assignment mechanism \( a \). This general measure of compliance behavior ranges from 0 to 1 and is defined as,

\[
\sum_{z_{-i,j} \in Z_{-i,j}} C_{ij}(z_{-i,j}) \Pr(Z_{-i,j} = z_{-i,j} \mid A_j = a).
\] (2)

for \( a = 0, 1 \). Given this compliance measure, we now define the complier average direct effect (CADE) as the average direct effect of treatment assignment among compliers,

\[
\text{CADE}(a) = \frac{\sum_{j=1}^{J} \sum_{i=1}^{n_j} \sum_{z_{-i,j} \in Z_{-i,j}} \{Y_{ij}(1, z_{-i,j}) - Y_{ij}(0, z_{-i,j})\} C_{ij}(z_{-i,j}) \Pr(Z_{-i,j} = z_{-i,j} \mid A_j = a)}{\sum_{j=1}^{J} \sum_{i=1}^{n_j} \sum_{z_{-i,j} \in Z_{-i,j}} C_{ij}(z_{-i,j}) \Pr(Z_{-i,j} = z_{-i,j} \mid A_j = a)}.
\]

The definition requires that there exists at least one complier in the population. If units do not influence each other, we have \( Y_{ij}(z_{ij}, z_{-i,j}) = Y_{ij}(z_{ij}) \) and \( D_{ij}(z_{ij}, z_{-i,j}) = D_{ij}(z_{ij}) \). Hence, the compliance status for each unit in equations (1) and (2) no longer depends on the treatment assignment of the other units. As a result, under this setting, the CADE equals the finite sample version of the complier average causal effect defined in Angrist et al. (1996). Finally, in the absence of noncompliance, i.e., \( C_{ij}(z_{-i,j}) = 1 \) for all \( z_{-i,j} \) and \( i,j \), then \( \text{CADE}(a) \) asymptotically equals \( \text{DEY}(a) \) as the cluster size grows.

The CADE combines two causal pathways: a unit’s treatment assignment \( Z_{ij} \) can affect its outcome \( Y_{ij} \) either through its own treatment receipt \( D_{ij} \) or that of the other units \( D_{-i,j} = (D_{1j}, \ldots, D_{i-1,j}, D_{i+1,j}, \ldots, D_{n_j}) \). If there is either no spillover effect of encouragement on treatment receipt or no spillover effect of treatment receipt on outcome, then the second causal pathway no longer exists. Under this scenario, the CADE corresponds to the average direct effect of one’s own treatment receipt among compliers because the treatment assignment is the same as the treatment receipt. In contrast, when both types of spillover effects exist, the CADE includes the indirect effect of one’s own encouragement on the outcome through the treatment receipt of other units in the same village as well as the direct effect of one’s own treatment receipt on the outcome. Unfortunately, without additional assumptions, the CADE is not identifiable. We therefore propose a set of assumptions for nonparametric identification. In addition, Appendix E.2 considers a
model-based approach to the identification and estimation for further distinguishing the two causal pathways.

### 3.3.2 Nonparametric Identification

To establish the nonparametric identification of the CADE, we begin by generalizing the exclusion restriction of \[\text{Angrist et al. (1996)},\] which assumes no interference between units.

**Assumption 3 (Exclusion Restriction with Interference between Units)**

\[
Y_{ij}(z_j; d_j) = Y_{ij}(z_j'; d_j) \quad \text{for any } z_j, z_j' \text{ and } d_j.
\]

Assumption 3 states that the outcome of a unit does not depend on the treatment assignment of any unit within the same cluster (including itself) so long as the treatment receipt for all the units of the cluster remains identical. In other words, the outcome of a unit depends only on the treatment receipt vector of all units within its own cluster. The assumption is violated if the outcome of one unit is influenced by its own treatment assignment or that of another unit within the same cluster even when the treatment receipts of all the units in the cluster including itself are held constant. In our application, the assumption is plausible since the encouragement to enroll in the RSBY is unlikely to affect the hospital expenditure other than through the actual enrollment itself.

Under Assumption 3, we can write the potential outcome as the function of treatment receipt alone, \(Y_{ij}(d_j)\). Thus, the observed outcome is written as \(Y_{ij}(D_j)\) where \(D_j = D_j(Z_j)\). We maintain Assumption 3 for the remainder of the paper. To avoid confusion, we will explicitly write out the treatment receipt as the argument of potential outcome. For example, \(Y_{ij}(D_j = 1_{n_j})\) represents the potential outcome when \(D_{ij} = 1\) for \(j = 1, \ldots, n_j\), while \(Y_{ij}(1_{n_j})\) denotes the potential outcome when \(Z_{ij} = 1\) for \(j = 1, \ldots, n_j\).

We next generalize the monotonicity assumption of \[\text{Angrist et al. (1996)}.\]

**Assumption 4 (Monotonicity with Interference between Units)**

\[
D_{ij}(1, z_{-i,j}) \geq D_{ij}(0, z_{-i,j}) \quad \text{for all } z_{-i,j} \in Z_{-i,j}.
\]
The assumption states that being assigned to the treatment condition never negatively affects the treatment receipt of a unit, regardless of how the other units within the same cluster are assigned to the treatment/control conditions. Assumption 4 is plausible in our application because the encouragement is expected to increase the enrollment in the RSBY.

In the absence of interference between units, exclusion restriction and monotonicity are sufficient for the nonparametric identification of the complier average causal effect. However, when interference exists, an additional restriction on the interference structure is necessary. The reason is that there are two types of possible spillover effects: the spillover effect of treatment assignment on the treatment receipt and the spillover effect of treatment receipt on the outcome. As a result, even under exclusion restriction, the treatment assignment of a noncomplier can still affect its outcome through the treatment receipts of other units within in the same cluster.

To address this problem, we propose the following identification assumption.

**Assumption 5 (Restricted Interference under Noncompliance)** For any unit \(i\) in cluster \(j\), if \(D_{ij}(1, z_{-i,j}) = D_{ij}(0, z_{-i,j})\) for some \(z_{-i,j} \in Z_{-i,j}\), then \(Y_{ij}(D_j(1, z_{-i,j})) = Y_{ij}(D_j(0, z_{-i,j}))\) holds.

The assumption states that if the treatment receipt of a unit is not affected by its own treatment assignment (i.e., the unit is a noncomplier), then its outcome should also not be affected by its own treatment assignment through the treatment receipts of other units in the same cluster. Although Assumption 5 appears to be concerned only with the spillover effects of treatment receipt on the outcome, its plausibility also depends on the spillover effects of treatment assignment on the treatment receipt.

To facilitate the understanding of this assumption, we consider the following three scenarios under which Assumption 5 is satisfied. First, assume no spillover effect of treatment receipt on the outcome (Scenario I of Figure 1(a)),

\[
Y_{ij}(d_{ij}, d_{-i,j}) = Y_{ij}(d_{ij}, d'_{-i,j}) \text{ for } d_{ij} = 0, 1, \text{ and any } d_{-i,j}, d'_{-i,j}. \tag{3}
\]

Testable conditions for this scenario are given in Appendix B.1.
Second, suppose that the treatment assignment has no spillover effect on the treatment receipt (Scenario II of Figure 1(b)),

$$D_{ij}(z_{ij}, z'_{-i,j}) = D_{ij}(z_{ij}, z_{-i,j}) \text{ for } z_{ij} = 0, 1, \text{ and any } z_{-i,j}, z'_{-i,j}. \quad (4)$$

Such an assumption is made by Kang and Imbens (2016) in the context of online experiments, in which the assignment of treatment (e.g., social media messaging) can be individualized but units may interact with each other once they receive the treatment. We can test this scenario by estimating $\text{SED}(1)$ and $\text{SED}(0)$.

Third, we can weaken the condition in equation (4) by considering an alternative condition that if a unit’s treatment receipt is not affected by its own treatment assignment (i.e., the unit is a noncomplier), then the treatment assignment of this unit has no effect on the treatment receipts of the other units in the same cluster (the absence of dotted edges in Scenario III of Figure 1(c)),

$$\text{if } D_{ij}(1, z_{-i,j}) = D_{ij}(0, z_{-i,j}), \text{ then } D_{-i,j}(1, z_{-i,j}) = D_{-i,j}(0, z_{-i,j}).$$

In our application, this scenario is violated, for example, if a household that already has insurance and is not going to be affected by the encouragement influences the enrollment decision of another household by recommending the RSBY to it. To increase the plausi-
bility of this scenario in our application, we excluded all the households with pre-existing insurance from the experiment. As a result, this scenario is plausible because one’s encouragement is expected to have a much greater influence on his/her own enrollment than the enrollment of another unit.

Although all three scenarios above satisfy Assumption 5, the interpretation of the CADE is different. In particular, under Scenarios I and II, we can interpret the CADE as the average direct effect of one’s own treatment receipt on the outcome among compliers. In contrast, under Scenario III, the CADE also includes the average direct effect of one’s own encouragement on the outcome through the treatment receipts of other units. Nevertheless, this combined direct effect of encouragement may be of interest to policy makers because most government programs including the RSBY are based on the encouragement design. In Appendix E.2, we address this issue using a model-based approach.

The next theorem establishes the nonparametric identification of the CADE as the cluster size tends to infinity. Under Assumptions 1–5, we show that in the limit, the CADE equals the ratio of the average direct effects of treatment assignment on the outcome and on the treatment receipt while holding the treatment assignment mechanism fixed. Although the unbiased estimation of DEY(a) and DED(a) is readily available (Hudgens and Hallo-
ran, 2008), for the consistent estimation of the CADE, we need an additional restriction on the structure of interference. We follow Sävje et al. (2017)’s result on the consistency of average causal effect in finite population framework, and assume that the average amount of interference per unit does not grow proportionally to the cluster size (see Appendix B.2 for a proof of the theorem and the details).

**Theorem 2 (Nonparametric Identification and Consistent Estimation of the Complier Average Direct Effect)**

1. Under Assumptions 1–5 we have

   $$\lim_{n_j \to \infty} \frac{\text{DEY}(a)}{\text{DED}(a)} = \lim_{n_j \to \infty} \text{CADE}(a).$$

2. Suppose that the outcome is bounded and the restriction on interference in Sävje et al. (2017) holds for both the treatment receipt and the outcome. Then, as both
the cluster size \( n_j \) and the number of clusters \( J \) go to infinity, we can consistently estimate the CADE,

\[
\operatorname{plim}_{n_j \to \infty, J \to \infty} \frac{\hat{DEY}(a)}{\hat{DED}(a)} = \lim_{n_j \to \infty, J \to \infty} \text{CADE}(a)
\]

for each \( a = 0, 1 \).

The CADE is nonparametrically identifiable as the cluster size and the number of clusters tend to infinity, and can be consistently estimated by the ratio of two estimated ITT effects. The asymptotic properties are derived within the finite population framework, approximating the sampling distribution of an estimator by embedding it in an asymptotically stable sequence of finite populations (Hájek, 1960; Lehmann, 2004).

### 3.4 Stratified Interference

Unfortunately, as pointed out by Hudgens and Halloran (2008), a valid estimator of the variances of these ITT effect estimators is unavailable without an additional assumption. Hudgens and Halloran (2008) rely upon the stratified interference assumption that the outcome of one unit depends on the treatment assignment of other units only through the number of those who are assigned to the treatment condition within the same cluster. In other words, what matters is the number of units rather than which units are assigned to the treatment condition.

We assume that stratified interference applies to both the outcome and treatment receipt.

**Assumption 6 (Stratified Interference)**

\[ D_{ij}(z_j) = D_{ij}(z_j') \quad \text{and} \quad Y_{ij}(z_j) = Y_{ij}(z_j') \quad \text{if} \quad z_{ij} = z_{ij}' \quad \text{and} \quad \sum_{i=1}^{n_j} z_{ij} = \sum_{i=1}^{n_j} z_{ij}'. \]

In our application, stratified interference for the treatment receipt requires that the enrollment decisions of households depend only on their own encouragement and the number of encouraged households in their village. Under the assumption of no spillover effect of treatment receipt on the outcome, stratified interference for the outcome holds so long as it is applicable to the treatment receipt. However, for more general scenarios, Assumption 6 may not be satisfied for the outcome even if it holds for the treatment receipt.
3.4.1 Nonparametric Identification

Under Assumption 6, we can simplify the CADE because the number of the units assigned to the treatment condition in each cluster is fixed given treatment assignment mechanism. This implies that we can write $D_{ij}(z_j)$ and $Y_{ij}(z_j)$ as $D_{ij}(z, a)$ and $Y_{ij}(z, a)$, respectively, and as a result $CADE(a)$ equals,

$$CADE(a) = \frac{\sum_{J} \sum_{i=1}^{n_j} \{Y_{ij}(1, a) - Y_{ij}(0, a)\} I\{D_{ij}(1, a) - D_{ij}(0, a) = 1\}}{\sum_{J} \sum_{i=1}^{n_j} I\{D_{ij}(1, a) - D_{ij}(0, a) = 1\}},$$

where the complier status can also be simplified as a function of assignment mechanism alone, i.e., $C_{ij}(a) = I\{D_{ij}(1, a) = 1, D_{ij}(0, a) = 0\}$.

We now present the results on nonparametric identification and consistent estimation under stratified interference.

**Theorem 3 (Nonparametric Identification and Consistent Estimation of the Complier Average Direct Effect under Stratified Interference)** Suppose that the outcome is bounded. Then, under Assumptions 1–6 we have

$$\lim_{n_j \to \infty, J \to \infty} CADE(a) = \text{plim}_{n_j \to \infty, J \to \infty} \frac{DEY(a)}{DED(a)},$$

for $a = 0, 1$.

Proof is in Appendix B.4. Under the stratified interference assumption, the consistent estimation of CADE no longer requires the restrictions on interference in Sävje et al. (2017).

3.4.2 Effect Decomposition

Under stratified interference, we can decompose the average direct effect of treatment assignment as the sum of the average direct effects for compliers and noncompliers,

$$DEY(a) = CADE(a) \cdot \pi_c(a) + NADE(a) \cdot \{1 - \pi_c(a)\},$$

where $NADE(a)$ is the noncomplier average direct effect and is defined as,

$$NADE(a) = \frac{\sum_{J} \sum_{i=1}^{n_j} \{Y_{ij}(1, a) - Y_{ij}(0, a)\} I\{D_{ij}(1, a) = D_{ij}(0, a)\}}{\sum_{J} \sum_{i=1}^{n_j} I\{D_{ij}(1, a) = D_{ij}(0, a)\}},$$

for $a = 0, 1$. 

and the proportion of compliers is given by,

$$\pi_c(a) = \frac{1}{N} \sum_{j=1}^{J} \sum_{i=1}^{n_j} I\{D_{ij}(1, a) = 1, D_{ij}(0, a) = 0\}.$$

According to the exclusion restriction given in Assumption 3 for compliers with $D_{ij}(1, a) = 1$ and $D_{ij}(0, a) = 0$, we can write the unit-level direct effect on the outcome as the sum of the direct effect through its own treatment receipt and the indirect effect through the treatment receipts of other units within the same cluster,

$$Y_{ij}(Z_{ij} = 1, a) - Y_{ij}(Z_{ij} = 0, a) = \{Y_{ij}(D_{ij} = 1, D_{-i,j}(Z_{ij} = 1, a)) - Y_{ij}(D_{ij} = 0, D_{-i,j}(Z_{ij} = 1, a))\}$$

$$+ \{Y_{ij}(D_{ij} = 0, D_{-i,j}(Z_{ij} = 1, a)) - Y_{ij}(D_{ij} = 0, D_{-i,j}(Z_{ij} = 0, a))\}.$$  \hspace{1cm} (6)

Thus, the treatment assignment can affect its outcome either directly through its own treatment or indirectly through the treatment receipts of the other units in the same cluster.

For noncompliers ($D_{ij}(1, a) = D_{ij}(0, a) = d$), the exclusion restriction implies,

$$Y_{ij}(Z_{ij} = 1, a) - Y_{ij}(Z_{ij} = 0, a) = Y_{ij}(D_{ij} = d, D_{-i,j}(Z_{ij} = 1, a)) - Y_{ij}(D_{ij} = d, D_{-i,j}(Z_{ij} = 0, a)).$$  \hspace{1cm} (7)

The treatment assignment affects its own outcome only through the treatment receipt of the other units in the same cluster. Furthermore, Assumption 5 implies $Y_{ij}(D_{ij} = d, D_{-i,j}(Z_{ij} = 1, a)) = Y_{ij}(D_{ij} = d, D_{-i,j}(Z_{ij} = 0, a))$. Under this assumption, equation (7) equals zero, implying $\text{NADE}(a) = 0$ and the identification of $\text{CADE}(a)$.

3.4.3 Randomization-based Variances

We derive the randomization-based variances of the proposed estimators within the finite population framework, in which the uncertainty is solely based on the two-stage randomization. As shown by Hudgens and Halloran (2008) in the context of ITT analysis, stratified interference enables the estimation of variance. Here, we first derive the variances of the
ITT effect estimators and then derive the variance of the proposed CADE estimator. We begin by defining the following quantities,

\[
\sigma^2_j(z, a) = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} \left\{ Y_{ij}(z, a) - \bar{Y}_j(z, a) \right\}^2, \quad \sigma^2_{DE}(a) = \frac{1}{J - 1} \sum_{j=1}^{J} \left\{ \frac{n_j J}{N} \text{DEY}_j(a) - \text{DEY}(a) \right\}^2, \]

\[
\omega^2_j(a) = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} \left\{ \left[ Y_{ij}(1, a) - Y_{ij}(0, a) \right] - \left[ \bar{Y}_j(1, a) - \bar{Y}_j(0, a) \right] \right\}^2,
\]

where \( \sigma^2_j(z, a) \) is the within-cluster variance of potential outcomes, \( \sigma^2_{DE}(a) \) is the between-cluster variance of \( \text{DEY}_{ij}(a) \), and \( \omega^2_j(a) \) is the within-cluster variance of \( \text{DEY}_{ij}(a) \). Using this notation, we give the results for the ITT effects of treatment assignment on the outcome. The results for the ITT effects of treatment assignment on the treatment receipt can be obtained in the same way.

**Theorem 4 (Randomization-based Variances of the ITT Effect Estimators)**

Under Assumptions 1, 2 and 6, we have

\[
\text{var}\left\{ \hat{\text{DEY}}(a) \right\} = \left( 1 - \frac{J_a}{J} \right) \frac{\hat{\sigma}_{DE}^2(a)}{J_a} + \frac{1}{J_a J} \sum_{j=1}^{J} \text{var}\left\{ \frac{n_j J}{N} \hat{\text{DEY}}_j(a) \mid A_j = a \right\},
\]

where

\[
\text{var}\left\{ \hat{\text{DEY}}_j(a) \mid A_j = a \right\} = \frac{\hat{\sigma}_j^2(1, a)}{n_{j1}} + \frac{\hat{\sigma}_j^2(0, a)}{n_{j0}} - \frac{\omega_j^2(a)}{n_j}.
\]

Proof is given in Appendix B.5. Because we cannot observe \( Y_{ij}(1, a) \) and \( Y_{ij}(0, a) \) simultaneously, no unbiased estimator exists for \( \omega_j^2(a) \), implying that no unbiased estimation of the variances is possible. Thus, following Hudgens and Halloran (2008), we propose a conservative estimator,

\[
\hat{\text{var}}\left\{ \hat{\text{DEY}}(a) \right\} = \left( 1 - \frac{J_a}{J} \right) \frac{\hat{\sigma}_{DE}^2(a)}{J_a} + \frac{1}{J_a J} \sum_{j=1}^{J} n_j^2 J^2 \left( \frac{\hat{\sigma}_j^2(1, a)}{n_{j1}} + \frac{\hat{\sigma}_j^2(0, a)}{n_j - n_{j1}} \right) I(A_j = a),
\]

(8)

where

\[
\hat{\sigma}_j^2(z, a) = \frac{\sum_{i=1}^{n_j} \left\{ Y_{ij} - \hat{Y}_j(z, a) \right\}^2 I(Z_{ij} = z)}{n_{jz} - 1},
\]

\[
\hat{\sigma}_{DE}^2(a) = \frac{\sum_{j=1}^{J} \left\{ \frac{n_j J}{N} \hat{\text{DEY}}_j(a) - \hat{\text{DEY}}(a) \right\}^2 I(A_j = a)}{J_a - 1}.
\]
In equation (8), \( \hat{\sigma}^2_{DE}(a) \) represents the between-cluster sample variance, and \( \hat{\sigma}^2_j(z, a) \) is the within-cluster variance in cluster \( j \). Thus, the variance of the ITT direct effect estimator is a weighted average of the between-cluster sample variance and the within-cluster sample variance.

It can be shown that this variance estimator is on average no less than the true variance,

\[
E \left[ \hat{\text{var}} \left\{ \hat{DEY}(a) \right\} \right] \geq \text{var} \left\{ \hat{DEY}(a) \right\},
\]

where the inequality becomes equality when the unit-level direct effect, i.e., \( Y_{ij}(1, a) - Y_{ij}(0, a) \), is constant within each cluster (see Appendix B.7 for a proof). In Appendix B.3 we provide the asymptotic normality result of the ITT effect estimators under additional regularity conditions based on the finite population central limit theorems in Hájek (1960), Ohlsson (1989) and Li and Ding (2017). These conditions are satisfied for a bounded outcome as the cluster size and the number of clusters go to infinity. See Chin (2018) for more refined results on the asymptotic normality of the ITT effect estimators without stratified interference.

We next derive the asymptotic randomization-based variance of the proposed estimator.

**Theorem 5 (Randomization-based Variance of the CADE Estimator)** Under Assumptions 1–6, the asymptotic variance of \( \hat{CADE}(a) \) is

\[
\frac{1}{\text{DED}(a)^2} \left[ \text{var} \left\{ \hat{DEY}(a) \right\} - 2 \frac{\text{DEY}(a)}{\text{DED}(a)} \text{cov} \left\{ \hat{DEY}(a), \hat{\text{DED}}(a) \right\} + \frac{\text{DEY}(a)^2}{\text{DED}(a)^2} \text{var} \left\{ \hat{\text{DED}}(a) \right\} \right],
\]

Proof of Theorem 5 is a direct application of the Delta method based on the asymptotic normality of the ITT effect estimators shown in Appendix B.3. Due to the space limitation, we give the expression of \( \text{cov} \left\{ \hat{DEY}(a), \hat{\text{DED}}(a) \right\} \) in Appendix B.6. Because the proposed CADE estimator is a ratio estimator, its variance blows up when \( \text{DED} \) is close to zero. This is similar to the weak instrument problem in the standard instrumental variable settings.

We obtain the following variance estimator by replacing each term in the brackets with their conservative estimators,

\[
\hat{\text{var}} \left\{ \hat{CADE}(a) \right\}
\]
\[
\text{var} \left\{ \hat{\text{CADE}}(a) \right\} = \frac{1}{\text{DED}(a)^2} \left[ \text{var} \left\{ \hat{\text{DEY}}(a) \right\} - 2 \frac{\text{DEY}(a)}{\text{DED}(a)} \text{cov} \left\{ \hat{\text{DEY}}(a), \hat{\text{DED}}(a) \right\} + \frac{\text{DEY}(a)^2}{\text{DED}(a)^2} \text{var} \left\{ \hat{\text{DED}}(a) \right\} \right],
\]

(9)

where \( \text{var} \left\{ \hat{\text{DED}}(a) \right\} \) and \( \text{cov} \left\{ \hat{\text{DEY}}(a), \hat{\text{DED}}(a) \right\} \) are obtained by replacing \( Y \) with \( D \) and the sample variances with the sample covariances in equation (8), respectively. Similar to the ITT analysis, each of the three terms in the brackets of \( \text{var} \left\{ \hat{\text{CADE}}(a) \right\} \) is a weighted average of between-cluster and within-cluster sample variances.

Because the expectation of product is generally not equal to the product of expectations, unlike the ITT analysis, \( \text{var} \left\{ \hat{\text{CADE}}(a) \right\} \) is not a conservative variance estimator in finite samples. In Appendix B.8, however, we show that it is asymptotically conservative. Finally, to evaluate the robustness of the variance estimator based on Assumption 6, we conduct simulation studies and find that the proposed variance estimator works well so long as the number of clusters is relatively large (see Appendix D).

### 3.5 Connections to Two-stage Least Squares Regression

In this section, we establish direct connections between the proposed estimator of the CADE and the two-stage least squares estimator, which is popular among applied researchers. Basse and Feller (2018) study the relationships between the ordinary least squares and randomization-based estimators for the ITT analysis under a particular two-stage randomized experiment design. Here, we further extend these previous results.

#### 3.5.1 Point Estimates

We begin with the ITT analysis. To account for different cluster sizes, we transform the treatment and outcome variables so that each unit, rather than each cluster, is equally weighted. Specifically, we multiply them by the weights proportional to the cluster size, i.e., \( D_{ij}^* = n_j JD_{ij} / N \) and \( Y_{ij}^* = n_j JY_{ij} / N \) (see Appendix C for the results with general
weights). We consider the following linear models for the treatment receipt and outcome,

\[
D_{ij}^* = \sum_{a=0,1} \gamma_a I(A_j = a) + \sum_{a=1} \gamma_{1a} Z_{ij} I(A_j = a) + \xi_{ij}, \quad (10)
\]

\[
Y_{ij}^* = \sum_{a=0,1} \alpha_a I(A_j = a) + \sum_{a=1} \alpha_{1a} Z_{ij} I(A_j = a) + \epsilon_{ij}, \quad (11)
\]

where \( \xi_{ij} \) and \( \epsilon_{ij} \) are error terms.

Unlike the two-step procedure in Basse and Feller (2018), we fit the weighted least squares regression with the following inverse probability weights,

\[
w_{ij} = \frac{1}{J_{A_j} \cdot n_{jZ_{ij}}}. \quad (12)
\]

The next theorem shows that the resulting weighted least squares estimators are equivalent to the randomization-based ITT effect estimators. Proof is given in Appendix C.1.

**THEOREM 6 (WEIGHTED LEAST SQUARES REGRESSION ESTIMATORS FOR THE ITT ANALYSIS)**

Let \( \hat{\gamma}_{wls} \) and \( \hat{\alpha}_{wls} \) be the weighted least squares estimators of the coefficients in the models given in equations (10) and (11), respectively. The regression weights are given in equation (12). Then,

\[
\hat{\gamma}_{1a} = \widehat{DE}(a), \quad \hat{\gamma}_a = \widehat{D}(0, a), \quad \hat{\alpha}_{1a} = \widehat{DY}(a), \quad \hat{\alpha}_a = \widehat{Y}(0, a).
\]

For the CADE, we consider the weighted two-stage least squares regression where the weights are the same as before and given in equation (12). In our setting, the first-stage regression model is given by equation (10) while the second-stage regression is given by

\[
Y_{ij}^* = \sum_{a=0,1} \beta_a I(A_j = a) + \sum_{a=0,1} \beta_{1a} D_{ij}^* I(A_j = a) + \eta_{ij}, \quad (13)
\]

where \( \eta_{ij} \) is an error term. The weighted two-stage least squares estimators of the coefficients for the model in equation (13) can be obtained by first fitting the model in equation (10) with weighted least squares and then fitting the model in equation (13) again via weighted least squares, in which \( D_{ij}^* \) is replaced by its predicted values based on the first stage regression model. The following theorem establishes the equivalence between the resulting weighted two-stage least squares regression and randomization-based estimators.

Proof is given in Appendix C.2.
**Theorem 7 (Weighted Two-stage Least Squares Regression Estimator for the CADE)**

Let $\beta_{w2sls}^{a}$ and $\beta_{w2sls}^{a1}$ be the weighted two-stage least squares estimators of the coefficients for the model given in equation (13). The first stage regression model is given in equation (10), and the regression weights are given in equation (12). Then,

$$\hat{\beta}_{w2sls}^{a} = \text{CADE}(a), \quad \hat{\beta}_{w2sls}^{a1} = \hat{Y}(0, a) - \text{CADE}(a) \cdot \hat{D}(0, a).$$

### 3.5.2 Variances

Basse and Feller (2018) show that the cluster-robust HC2 variance (Bell and McCaffrey, 2002) is equal to the randomization-based variance of the average spillover effect estimator under the assumption of equal cluster size. We first generalize this equivalence result to the case where the cluster size varies and then propose a regression-based variance estimator for the CADE estimator that is equivalent to the randomization-based variance estimator.

We begin by introducing additional notation. Let $X_{j} = (X_{1j}, \ldots, X_{nj})^{\top}$ be the design matrix of cluster $j$ for the model given in equations (10) and (11) with $X_{ij} = (I(A_{j} = 1), I(A_{j} = 0), Z_{ij}I(A_{j} = 1), Z_{ij}I(A_{j} = 0))^{\top}$. Let $X = (X_{1}^{\top}, \ldots, X_{J}^{\top})^{\top}$ be the entire design matrix, and $W_{j} = \text{diag}(w_{1j}, \ldots, w_{nj})$ be the weight matrix in cluster $j$, $W = \text{diag}(W_{1}, \ldots, W_{J})$ be the entire weight matrix. We use $\hat{\epsilon}_{j} = (\hat{\epsilon}_{1j}, \ldots, \hat{\epsilon}_{nj})$ to denote the residual vector in cluster $j$ obtained from the weighted least squares fit of the model given in equation (11), and $\hat{\epsilon} = (\hat{\epsilon}_{1}, \ldots, \hat{\epsilon}_{j})$ to represent the residual vector for the entire sample.

Using the weights, the cluster-robust generalization of HC2 variance is given by

$$\hat{\text{var}}_{\text{hc2}}^{\text{cluster}}(\alpha_{wls}) = (X^{\top}WX)^{-1}\left\{ \sum_{j} X_{j}^{\top}W_{j}(I_{nj} - P_{j})^{-1/2}\hat{\epsilon}_{j}^{\top}(I_{nj} - P_{j})^{-1/2}W_{j}X_{j} \right\} (X^{\top}WX)^{-1},$$

where $I_{nj}$ is the $n_{j} \times n_{j}$ identity matrix and $P_{j}$ is the following cluster leverage matrix,

$$P_{j} = W_{j}^{1/2}X_{j}(X^{\top}WX)^{-1}X_{j}^{\top}W_{j}^{1/2}.$$

It can be shown that $\hat{\text{var}}_{\text{hc2}}^{\text{cluster}}(\alpha_{wls}) = \hat{\sigma}_{DE}(a)^{2}/J_{a}$, representing the between-cluster sample variance. However, as shown in Theorem 4, $\hat{\text{var}}\left\{ \hat{D}EY(a) \right\}$ is a weighted average of between-cluster and within-cluster sample variances. Thus, unlike the results in Basse and Feller (2018), the cluster-robust HC2 variance no longer equals the randomization-based variance estimator, because it only takes into account the between-cluster variance.
To address this problem, we introduce the following individual-robust HC2 variance,

$$\hat{\text{var}}_{\text{hc2}}^{\text{ind}}(\hat{\alpha}_{wls}) = (X^\top WX)^{-1} \left\{ \sum_{j=1}^{J} \sum_{i=1}^{n_j} w_{ij}^2 \hat{\epsilon}_{ij}^2 (1 - P_{ij})^{-1} X_{ij} X_{ij}^\top \right\} (X^\top WX)^{-1},$$

where $P_{ij} = w_{ij} X_{ij} (X_j^\top W_j X_j)^{-1} X_{ij}$ is the individual leverage and $\hat{\epsilon}_{ij}^* = \hat{\epsilon}_{ij} - \sum_{i'=1}^{n_j} \hat{\epsilon}_{i'j} I(Z_{i'j} = z)/n_{jz}$ is the adjusted residuals for $Z_{ij} = z$ so that we have $X_j^\top \hat{\epsilon}_{ij}^* = 0$. The next theorem establishes that the weighted average of the cluster-robust and individual-robust HC2 variance estimators is numerically equivalent to the randomization-based variance estimator.

**Theorem 8 (Regression-based Variance Estimators for the ITT Effects)** The randomization-based variance estimator of the direct effect is a weighted average of the cluster-robust and individual-robust HC2 variances,

$$\hat{\text{var}} \left\{ \hat{\text{DED}}(a) \right\} = \left( 1 - \frac{J_a}{J} \right) \hat{\text{var}}_{\text{hc2}}^{\text{cluster}}(\hat{\gamma}_{1a}) + \frac{J_a}{J} \hat{\text{var}}_{\text{hc2}}^{\text{ind}}(\hat{\gamma}_{1a}),$$

$$\hat{\text{var}} \left\{ \hat{\text{DEY}}(a) \right\} = \left( 1 - \frac{J_a}{J} \right) \hat{\text{var}}_{\text{hc2}}^{\text{cluster}}(\hat{\alpha}_{1a}) + \frac{J_a}{J} \hat{\text{var}}_{\text{hc2}}^{\text{ind}}(\hat{\alpha}_{1a}).$$

Proof is given in Appendix C.3.

To gain some intuition about the weighted average of two robust variances, consider the following model commonly used for split-plot designs,

$$Y_{ij}^* = \sum_{a=0,1} \alpha_a I(A_j = a) + \sum_{a=0,1} \alpha_{1a} Z_{ij} I(A_j = a) + \epsilon_{Bj} + \epsilon_{Wij},$$

where $\epsilon_{Bj}$ represents the random effects for whole plots (or clusters), and $\epsilon_{Wij}$ is the random effects for split-plots (or individuals). The cluster-robust HC2 variance is related to $\epsilon_{Bj}$ and the individual-robust HC2 variance is related to $\epsilon_{Wij}$ In Appendix C.4 we discuss the connection between the random effects model and the randomization-based inference and explain why the adjustment for $\hat{\epsilon}_{ij}^*$ is necessary.

Finally, we consider the weighted two-stage least squares regression given in equations (10) and (13). Let $M_j = (M_{1j}, \ldots, M_{nj})^\top$ be the design matrix for cluster $j$ in the second-stage regression with $M_{ij}^\top = (I(A_j = 1), I(A_j = 0), \hat{D}_{ij}^* I(A_j = 1), \hat{D}_{ij}^* I(A_j = 0))$ where $\hat{D}_{ij}^*$ represents the fitted value given in equation (10). Let $M = (M_1^\top, \ldots, M_J^\top)^\top$
be the entire design matrix. We define the cluster-robust HC2 variance as,

\[
\hat{\text{var}}_{\text{cluster}}(\hat{\beta}_{\text{w2sls}}) = (M^\top WM)^{-1} \left\{ \sum_{j=1}^{J} M_j^\top W_j (I_{n_j} - Q_j)^{-1/2} \hat{\eta}_j^\top (I_{n_j} - Q_j)^{-1/2} W_j M_j \right\} (M^\top WM)^{-1},
\]

where \( Q_j \) is the cluster leverage matrix,

\[
Q_j = W_j^{1/2} M_j (M^\top WM)^{-1} M_j^\top W_j^{1/2}.
\]

The individual-robust HC2 variance is given by,

\[
\hat{\text{var}}_{\text{ind}}(\hat{\beta}_{\text{w2sls}}) = (M^\top WM)^{-1} \left\{ \sum_{j=1}^{J} \sum_{i=1}^{n_j} w_{ij}^2 \hat{\eta}_{ij}^2 (1 - Q_{ij})^{-1} M_{ij} M_{ij}^\top \right\} (M^\top WM)^{-1},
\]

where \( Q_{ij} = w_{ij} M_{ij}^\top (M^\top W_j M_j)^{-1} M_{ij} \) is the individual leverage and \( \hat{\eta}_{ij} = \hat{\eta}_{ij} - \sum_{i'=1}^{n_j} \hat{\eta}_{i'j} I(Z_{ij} = z) / n_{ij} \) for \( Z_{ij} = z \) is the adjusted residual with \( X_j^\top \hat{\eta}_j = 0 \). As in the case of ITT analysis, we can show that the weighted average of cluster-robust and individual-robust variance estimators is numerically equivalent to the randomization-based variance estimator.

**Theorem 9 (Regression-based Variance Estimator for the CADE)** The randomization-based variance estimator of the average complier direct effect is a weighted average of the cluster-robust and individual-robust HC2 variances,

\[
\text{var} \{ \hat{\text{CADE}}(\alpha) \} = \left( 1 - \frac{J_a}{J} \right) \hat{\text{var}}_{\text{cluster}}(\hat{\beta}_{1a}^{\text{w2sls}}) + \frac{J_a}{J} \hat{\text{var}}_{\text{ind}}(\hat{\beta}_{1a}^{\text{w2sls}}).
\]

Proof is given in Appendix C.5

### 4 Empirical Analysis

In this section, we analyze the data introduced in Section 2 by applying the proposed methodology. We focus on the annual household hospital expenditure, which ranges from 0 to INR 500,000 with the median value of 1,000. The outcome is missing for 926 households, which is less than 10% of the sample. For simplicity, we discard the observations with missing data from the current analysis and leave the development of a method for analyzing two-stage randomized experiments with missing data to future research.
As expected, the enrollment rate in the villages assigned to the “High” assignment mechanism is 67.0%, whereas the enrollment rate in the villages under the “Low” assignment mechanism is just 46.2%. Because the encouragement proportion is 80% under the “High” assignment mechanism and 40% under the “Low” assignment mechanism, this implies the existence of two-sided noncompliance, in which some households in the treatment group did not receive the treatment and others in the control group managed to receive it.

Table 2 presents the estimates of ITT effects and complier average direct and spillover effects. We show the results for both the individual/household-weighted and cluster/village-weighted estimands. In the top row, we show the estimated average direct effect on enrollment in RSBY under the “High” treatment mechanism (DED(1)) and under the “Low” treatment mechanism (DED(0)), the average spillover effects under the treatment (SED(1)) and control (SED(0)) conditions. The middle row presents the same set of ITT estimates for hospital expenditure. Finally, the bottom row presents the complier average direct effect under the “High” (CADE(1)) and “Low” (CADE(0)) treatment assignment mechanisms as well as the complier average spillover effect under the treatment (CASE(1)) and control (CASE(0)) conditions.

Table 2: Estimated Intention-to-Treat (ITT) and Complier Average Direct and Spillover Effects. For the “household-weighted” estimates, we equally weight households, whereas each village is equally weighted for the “village-weighted” estimates. The top row presents the average direct effects on enrollment in RSBY under the high treatment mechanism (DED(1)) and under the low treatment mechanism (DED(0)), the average spillover effects under the treatment (SED(1)) and control (SED(0)) conditions. The middle row presents the same set of ITT estimates for hospital expenditure. Finally, the bottom row presents the complier average direct effect under the “High” (CADE(1)) and “Low” (CADE(0)) treatment assignment mechanisms as well as the complier average spillover effect under the treatment (CASE(1)) and control (CASE(0)) conditions.

<table>
<thead>
<tr>
<th>Enrollment in RSBY</th>
<th>DED(1)</th>
<th>DED(0)</th>
<th>SED(1)</th>
<th>SED(0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household-weighted</td>
<td>0.482</td>
<td>0.441</td>
<td>0.086</td>
<td>0.045</td>
</tr>
<tr>
<td>Village-weighted</td>
<td>0.457</td>
<td>0.445</td>
<td>0.044</td>
<td>0.031</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hospital expenditure</th>
<th>DEY(1)</th>
<th>DEY(0)</th>
<th>SEY(1)</th>
<th>SEY(0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household-weighted</td>
<td>-795</td>
<td>875</td>
<td>-1374</td>
<td>297</td>
</tr>
<tr>
<td>Village-weighted</td>
<td>-222</td>
<td>1666</td>
<td>-1677</td>
<td>211</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hospital expenditure</th>
<th>CADE(1)</th>
<th>CADE(0)</th>
<th>CASE(1)</th>
<th>CASE(0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household-weighted</td>
<td>-1649</td>
<td>1984</td>
<td>-15900</td>
<td>6568</td>
</tr>
<tr>
<td>Village-weighted</td>
<td>-485</td>
<td>3752</td>
<td>-38341</td>
<td>6846</td>
</tr>
</tbody>
</table>
sometimes, statistically significant. In particular, the village-weighted average spillover effect on enrollment is 4.4 percentage points with the standard error of 1.8 under the treatment condition. The finding implies that assigning a greater proportion of households to the treatment condition makes another household of the same village more likely to enroll in RSBY especially if the latter is also encouraged to enroll.

The middle row of Table 2 presents the estimated ITT effects on the outcome. The estimated average direct effect under the “Low” assignment mechanism \( \text{DEY}(0) \) tends to be positive where the village-weighted estimate is statistically significant. In contrast, the estimated average direct effect under the “High” assignment mechanism \( \text{DEY}(1) \) is negative although not statistically significant.

One possible explanation for this difference is that the assignment to the treatment condition makes people visit hospitals more often and spend more on healthcare so long as fewer households within the same village are assigned to the treatment. When a large number of households within the same village are assigned to the treatment condition, the overcrowding of hospitals may reduce hospital visits of each treated household.

We examine the plausibility of this explanation by estimating the direct effect of the treatment assignment on the number of hospital visits. The estimated direct effect under the “High” treatment assignment mechanism is \(-0.157\), whereas that under the “Low” treatment assignment mechanism is \(0.132\). Although these estimates are not statistically significant, they provide suggestive evidence consistent with the overcrowding hypothesis.

The bottom row of Table 2 presents the estimates of the complier average direct and spillover effects. The village-weighted complier average direct effect under the “Low” assignment mechanism \( \text{CADE}(0) \) is positive and statistically significant, implying that enrollment in RSBY directly increases the household hospital expenditure when many households are assigned to the treatment condition. In contrast, the complier average direct effect under the “High” assignment mechanism \( \text{CADE}(1) \) is negative. This difference is
consistent with the overcrowding hypothesis discussed above.

In addition, we also estimate the complier average spillover effects. Unfortunately, they are imprecisely estimated, making it difficult to draw a definite conclusion about whether or not the proportion of treated households in a village directly affects one’s outcome among those who enroll in RSBY only when a greater proportion of households is encouraged to sign up for the insurance program.

Because most of the estimates are not statistically significant, it is difficult to draw a definitive conclusion. However, our analysis provides some suggestive policy recommendations. First, the estimated positive spillover effect of encouragement on enrollment suggests that the government could increase the enrollment rate by leveraging existing social networks among households within each village. Second, the estimated negative CADE under the “High” treatment assignment mechanism condition suggests that there might be overcrowding of local hospitals when many households newly enroll in the RSBY. The government can address this issue by increasing the capacity of local hospitals.

In addition to the quantities in our analysis above, we may also be interested in other quantities, e.g., the average spillover effect of the treatment assignment when all households are assigned to the treatment condition versus households are assigned to the treatment condition, and the direct effect of one’s own treatment receipt when the treatment receipts of other households are fixed at some constant levels. Unfortunately, without modeling assumptions, we are unable to identify these quantities. In Appendix E, we propose a model-based approach and estimate these quantities using our application data.

5 Concluding Remarks

In this paper, we consider two-stage randomized experiments with noncompliance and interference. We merge two strands of the causal inference literature, one on experiments with noncompliance and the other on experiments with interference. We introduce new causal quantities of interest, propose nonparametric identification results and consistent es-
timators, and derive their variances. We connect these randomization-based estimators to two-stage least squares regressions that are commonly used by applied researchers. Our motivating empirical application illustrates the usefulness of our proposed methodology in practice. We believe that the proposed methodology can help applied researchers make best use of this effective experimental design for studying interference problems.

References


Supplementary Appendix

The supplementary material contains the following five sections:

- Section A provides the results for the complier average spillover effects under stratified interference.
- Section B gives the proofs for the randomization-based inference approach.
- Section C gives the proofs for the regression-based approach.
- Section D presents the simulation studies.
- Section E proposes a model-based approach to overcome the limitations of the nonparametric approach in the main text.

A Complier Average Spillover Effects under Stratified Interference

Stratified interference, i.e., Assumption 6, allows us to define the complier average spillover effect (CASE), representing the average causal effect of treatment assignment mechanism among compliers while holding their own treatment assignment at a fixed value,

\[
\text{CASE}(z) = \frac{\sum_{j=1}^{J} \sum_{i=1}^{n_j} \{Y_{ij}(z, 1) - Y_{ij}(z, 0)\} I\{D_{ij}(z, 1) = 1, D_{ij}(z, 0) = 0\}}{\sum_{j=1}^{J} \sum_{i=1}^{n_j} I\{D_{ij}(z, 1) = 1, D_{ij}(z, 0) = 0\}}.
\]

We emphasize that this estimand is defined only when the spillover effect of treatment assignment on the treatment receipt is present (otherwise, the denominator is zero). Note that the compliers here are defined differently than those for the CADE. Specifically, the compliers for the CASE are those who receive the treatment only when the assignment mechanism \(A_j\) is equal to 1, i.e., \(I\{D_{ij}(z, 1) = 1, D_{ij}(z, 0) = 0\}\). Thus, the CASE represents the average causal effect of the assignment mechanism on the outcome among the compliers while holding their treatment assignment status constant.

A.1 Nonparametric Identification

To establish the nonparametric identification result for the CASE, we need two assumptions similar to Assumptions 4 and 5 for the CADE.

**Assumption A1 (Monotonicity with Respect to the Assignment Mechanism)**

\[D_{ij}(z, 1) \geq D_{ij}(z, 0) \quad \text{for all } z = 0, 1.\]

The assumption states that a unit is no less likely to receive the treatment under the treatment assignment mechanism \(A_j = 1\) than under the treatment assignment mechanism...
$A_j = 0$, holding its own treatment assignment at a constant. In our application, Assumption \[ A_1 \] implies that a household is no less likely to enroll in the RSBY when a greater number of households are encouraged to do so.

Next, we introduce the assumption of restricted interference similar to Assumption \[ 5 \].

**Assumption A2 (Restricted Interference under Noncompliance for the Assignment Mechanism)** Under Assumption \[ 6 \] for a given unit \( i \) in cluster \( j \), if \( D_{ij}(z, 1) = D_{ij}(z, 0) \) for some \( z \), then \( Y_{ij}(D_j(z, 1)) = Y_{ij}(D_j(z, 0)) \).

The assumption states that if the treatment receipt of a unit is not affected by the assignment mechanism of the cluster, its outcome should also not be affected by the assignment mechanism. Similar to Assumption \[ 5 \], this assumption holds in case of no spillover effect of treatment receipt on the outcome (equation (3)). As noted above, however, in case of no spillover effect on the treatment receipt (equation (4)), the CASE is not well defined. Furthermore, when both spillover effects are present, Assumption \[ A2 \] is likely to be violated.

We provide the nonparametric identification and consistent estimation results for the CASE that are analogous to those presented in Theorem \[ 3 \] for the CADE.

**Theorem A1 (Nonparametric Identification and Consistent Estimation of the Complier Average Spillover Effect under Stratified Interference)**

Suppose that the outcome is bounded. Then, under Assumptions \[ 1-3, 6 \] and \[ A1-A2 \] we have

$$\lim_{n_j \to \infty, J \to \infty} \text{CASE}(z) = \lim_{n_j \to \infty, J \to \infty} \frac{\text{SEY}(z)}{\text{SED}(z)},$$

for \( z = 0, 1 \).

Proof is in Appendix B.4.

### A.2 Effect Decomposition

We consider the following decomposition of the CASE analogous to that of the CADE,

$$\text{SEY}(z) = \text{CASE}(z) \cdot \lambda_c(z) + \text{NASE}(z) \cdot \{1 - \lambda_c(z)\},$$

(A1)

where the average noncomplier spillover effect is defined as,

$$\text{NASE}(z) = \frac{\sum_{j=1}^{J} \sum_{i=1}^{n_j} \{Y_{ij}(z, 1) - Y_{ij}(z, 0)\} I\{D_{ij}(z, 1) = D_{ij}(z, 0)\}}{\sum_{j=1}^{J} \sum_{i=1}^{n_j} I\{D_{ij}(1, a) = D_{ij}(0, a)\}},$$

and the proportion of compliers with respect to the treatment assignment is given by,

$$\lambda_c(z) = \frac{1}{N} \sum_{j=1}^{J} \sum_{i=1}^{n_j} I\{D_{ij}(z, 1) = 1, D_{ij}(z, 0) = 0\}.$$

For compliers with \( D_{ij}(z_{ij}, 1) = 1 \) and \( D_{ij}(z_{ij}, 0) = 0 \), the exclusion restriction, i.e., Assumption \[ 3 \] implies the following decomposition,

$$Y_{ij}(z_{ij}, 1) - Y_{ij}(z_{ij}, 0) = \{Y_{ij}(D_{ij} = 1, D_{-i,j}(z_{ij}, 1)) - Y_{ij}(D_{ij} = 0, D_{-i,j}(z_{ij}, 0))\}.$$
which shows that the effect of treatment assignment mechanism on the outcome for a unit equals the sum of the direct effect through its own treatment receipt and the indirect effect through the treatment receipts of the other units in the same cluster. For noncompliers with \( D_{ij}(z_{ij}, 1) = D_{ij}(z_{ij}, 0) = d \) where \( d = 0, 1 \), we can write the total effect of treatment assignment mechanism as,

\[
Y_{ij}(z_{ij}, 1) - Y_{ij}(z_{ij}, 0) = Y_{ij}(D_{ij} = d, D_{-i,j}(Z_{ij} = z_{ij}, 1)) - Y_{ij}(D_{ij} = d, D_{-i,j}(Z_{ij} = z_{ij}, 0)),
\]

which characterizes the spillover effect of the treatment assignments of the other units on the outcome through their treatment receipts. Assumption A2 guarantees this effect is zero for noncompliers, implying \( \text{NASE}(z) = 0 \) and the identification of \( \text{CASE}(z) \).

### A.3 Randomization-based Variances

We can also derive the randomization-based variances of the proposed spillover effect estimators. We begin by defining the following quantities,

\[
\sigma^2_b(z, a) = \frac{1}{J - 1} \sum_{j=1}^{J} \left\{ \frac{n_j J}{N} \bar{Y}_j(z, a) - \bar{Y}(z, a) \right\}^2,
\]

\[
\sigma^2_{SE}(z) = \frac{1}{J - 1} \sum_{j=1}^{J} \left\{ \frac{n_j J}{N} \text{SEY}_j(z) - \text{SEY}(z) \right\}^2,
\]

where \( \sigma^2_b(z, a) \) is the between-cluster variance of \( Y_{ij}(z, a) \), and \( \sigma^2_{SE}(z) \) is the between-cluster variance of \( \text{SEY}_{ij}(a) \). The next theorem presents the randomization-based variance.

**Theorem A2 (Variances of the ITT Spillover Effect Estimators)** Under Assumptions 7 and 8, we have

\[
\text{var}\left\{ \widehat{\text{SEY}}(z) \right\} = \frac{\sigma^2_b(z, 1)}{J_1} + \frac{\sigma^2_b(z, 0)}{J_0} - \frac{\sigma^2_{SE}(z)}{J} + \frac{1}{J_a} \sum_{a=0}^{1} \sum_{j=1}^{J} \text{var}\left\{ \widehat{Y}_j(z, a) \mid A_j = a \right\}
\]

where

\[
\text{var}\left\{ \widehat{Y}_j(z, a) \mid A_j = a \right\} = \frac{1}{n_{jz}} \left( 1 - \frac{n_{jz}}{n_j} \right) \sigma^2_j(z, a),
\]

In contrast to the case of the direct effect estimators, the variances of the spillover effect estimators are based on cluster-robust variances alone. However, no unbiased estimation of the variance of \( \widehat{\text{SEY}}(z) \) is available because \( \sigma^2_{SE}(z) \) is not identifiable. Hence, we propose a conservative estimator of the variance,

\[
\text{var}\left\{ \widehat{\text{SEY}}(z) \right\} = \frac{\widehat{\sigma}^2_b(z, 1)}{J_1} + \frac{\widehat{\sigma}^2_b(z, 0)}{J_0},
\]

36
which is no less than the true variance in expectation, i.e., $\mathbb{E} \left[ \text{var} \left\{ \text{SEY}(z) \right\} \right] \geq \text{var} \left\{ \text{SEY}(z) \right\}$. The inequality becomes equality when the cluster-level spillover effect, i.e., $n_j \sum_j \{ \overline{Y}_j(z, 1) - \overline{Y}_j(z, 0) \}/N$, is constant (see Appendix B.7 for a proof).

Finally, the variance of the CASE estimator can be derived by applying the Delta method as done in Theorem 5. The resulting variance involves the covariance between $\text{SEY}$ and $\text{SED}$ whose expression is shown in Appendix B.6.

## B Proofs for the Randomization Inference Approach

### B.1 Testable Conditions for No Spillover Effect of Treatment Receipt on the Outcome

When there is no spillover effect of treatment receipt on the outcome, we define

$$
\overline{(Y = y)D_{ij}(z, a)} = \sum_{z_{-i,j} \in Z_{-i,j}} I\{Y_{ij}(Z_{ij} = z, Z_{-i,j} = z_{-i,j}) = y\} D_{ij}(Z_{ij} = z, Z_{-i,j} = z_{-i,j})
$$

for any $y$. Because

$$
\lim_{n_j \to \infty} \{ \Pr(Z_{-i,j} = z_{-i,j} | Z_{ij} = 1, A_j = a) - \Pr(Z_{-i,j} = z_{-i,j} | Z_{ij} = 0, A_j = a) \} = 0
$$

we can obtain $\lim_{n_j \to \infty} \overline{(Y = y)D_{ij}(1, a)} \geq \lim_{n_j \to \infty} \overline{(Y = y)D_{ij}(0, a)}$ under Assumption 4. As a result, we have $\overline{(Y = y)D(1, a)} \geq \overline{(Y = y)D(0, a)}$, where

$$
\overline{(Y = y)D(z)} = \frac{1}{N} \sum_{j=1}^{J} \sum_{i=1}^{n_j} \overline{(Y = y)D_{ij}(z)}.
$$

Similarly, we can obtain $\lim_{n_j \to \infty} \overline{(Y = y)(1 - D)(1, a)} \geq \lim_{n_j \to \infty} \overline{(Y = y)(1 - D)(0, a)}$. Therefore, we obtain the following testable conditions for no spillover effect of the treatment receipt on the outcome,

$$
\lim_{n_j \to \infty} \overline{(Y = y)D_{ij}(1, a)} \geq \lim_{n_j \to \infty} \overline{(Y = y)D_{ij}(0, a)},
$$

$$
\lim_{n_j \to \infty} \overline{(Y = y)(1 - D)(1, a)} \geq \lim_{n_j \to \infty} \overline{(Y = y)(1 - D)(0, a)}.
$$

Similar to the unbiased estimation of the ITT effects, we can use

$$(Y = y)D(z, a) = \frac{1}{N} \sum_{j=1}^{J} n_j \overline{(Y = y)D_j(z, a)I(A_j = a)} = \frac{1}{J} \sum_{j=1}^{J} I(A_j = a),$$

$$(Y = y)(1 - D)(z, a) = \frac{1}{N} \sum_{j=1}^{J} n_j \overline{(Y = y)(1 - D)_j(z, a)I(A_j = a)} = \frac{1}{J} \sum_{j=1}^{J} I(A_j = a),$$

37
where

\[
(Y = y)D_j(z, a) = \sum_{i=1}^{n_j} I(Y_{ij} = y)D_{ij} I(Z_{ij} = z) / \sum_{i=1}^{n_j} I(Z_{ij} = z),
\]

\[
(Y = y)(1 - D)_j(z, a) = \sum_{i=1}^{n_j} I(Y_{ij} = y)(1 - D_{ij}) I(Z_{ij} = z) / \sum_{i=1}^{n_j} I(Z_{ij} = z),
\]

to unbiasedly estimate \((Y = y)D(z, a)\) and \((Y = y)(1 - D)(z, a)\), respectively. As a result, we can use the observed data to test whether there is a spillover effect of treatment receipt on the outcome.

B.2 Proof of Theorem 2

We first prove the nonparametric identification as the cluster size \(n_j\) goes to infinity for each \(j\). Under this scenario, the treatment assignment of one unit becomes asymptotically independent of another unit’s treatment assignment given its treatment assignment mechanism within the same cluster. This yields

\[
\lim_{n_j \to \infty} \{ \Pr(Z_{-i,j} = z_{-i,j} \mid Z_{ij} = z, A_j = a) - \Pr(Z_{-i,j} = z_{-i,j} \mid A_j = a) \} = 0
\]

for \(z = 0, 1\). Therefore, we have,

\[
\lim_{n_j \to \infty} \text{DED}(a) = \lim_{n_j \to \infty} J \sum_{j=1}^{J} n_j \sum_{i=1}^{n_j} \sum_{z_{-i,j} \in Z_{-i,j}} \{D_{ij}(1, z_{-i,j}) - D_{ij}(0, z_{-i,j})\} \Pr(Z_{-i,j} = z_{-i,j} \mid A_j = a)
\]

\[
= \lim_{n_j \to \infty} J \sum_{j=1}^{J} n_j \sum_{i=1}^{n_j} \sum_{z_{-i,j} \in Z_{-i,j}} C_{ij}(z_{-i,j}) \Pr(Z_{-i,j} = z_{-i,j} \mid A_j = a), \quad (A2)
\]

where the last equality follows from Assumption 4. Next, we show that the numerator of \(CADE(a)\) is equal to,

\[
\lim_{n_j \to \infty} \text{DEY}(a) = \lim_{n_j \to \infty} J \sum_{j=1}^{J} n_j \sum_{i=1}^{n_j} \sum_{z_{-i,j} \in Z_{-i,j}} \{Y_{ij}(1, z_{-i,j}) - Y_{ij}(0, z_{-i,j})\} \Pr(Z_{-i,j} = z_{-i,j} \mid A_j = a)
\]

\[
= \lim_{n_j \to \infty} J \sum_{j=1}^{J} n_j \sum_{i=1}^{n_j} \sum_{z_{-i,j} \in Z_{-i,j}} \{Y_{ij}(1, z_{-i,j}) - Y_{ij}(0, z_{-i,j})\} C_{ij}(z_{-i,j}) \Pr(Z_{-i,j} = z_{-i,j} \mid A_j = a), \quad (A3)
\]

where the last equality follows from Assumption 5. Thus, we obtain the desired result,

\[
\lim_{n_j \to \infty} \text{CADE}(a) = \lim_{n_j \to \infty} \text{DEY}(a) / \text{DED}(a).
\]

Next, we establish the consistent estimation. We assume the following restriction on interference in (Sävje et al., 2017) hold, which still allows the total amount of interference
within each cluster to grow with the cluster size,
\begin{align*}
\frac{1}{n_j} \sum_{i=1}^{n_j} \sum_{i'=1}^{n_j} t_{ii'} &= o(n_j) \quad \text{where} \quad t_{ii'} = \begin{cases} 
1 & \text{if } I_{ii}I_{ii'} = 1 \text{ for some } \ell = 1, 2, \ldots, n_j \\
0 & \text{otherwise}
\end{cases} \\
\sum_{i=1}^{n_j} \sum_{i'=1}^{n_j} \kappa_{ii'} &= o(n_j) \quad \text{where} \quad \kappa_{ii'} = \begin{cases} 
1 & \text{if } K_{ii}K_{ii'} \text{ for some } \ell = 1, 2, \ldots, n_j \\
0 & \text{otherwise}
\end{cases}
\end{align*}

and
\begin{align*}
I_{ii} &= \begin{cases} 
1 & \text{if } D_{ij}(z_j) \neq D_{ij}(z'_j) \quad \text{for } z_{ij} \neq z'_{ij} \text{ and } z_{-\ell,j} = z'_{-\ell,j} \\
1 & \text{if } i = \ell \\
0 & \text{otherwise}
\end{cases} \\
K_{ii} &= \begin{cases} 
1 & \text{if } Y_{ij}(z_j) \neq Y_{ij}(z'_j) \quad \text{for } z_{ij} \neq z'_{ij} \text{ and } z_{i,j} = z'_{i,j} \\
1 & \text{if } i = \ell \\
0 & \text{otherwise}
\end{cases}
\end{align*}

According to Sävje et al. (2017), under the proposed conditions, we can consistently estimate the DED and the DEY within each cluster,
\begin{align*}
\text{plim}_{n_j \to \infty} \hat{\text{DED}}_j(a) &= \text{DED}_j(a), \\
\text{plim}_{n_j \to \infty} \hat{\text{DEY}}_j(a) &= \text{DEY}_j(a) \quad \text{(A4)}
\end{align*}

for all \( j \). Furthermore, because \( A_j \) is the sampling indicator of a simple random sampling from \((n_1J \bar{D}_1(z,a)/N, \ldots, n_JJ \bar{D}_J(z,a)/N)\), as the number of clusters also tends to infinity, we have,
\begin{align*}
\text{plim}_{n_j \to \infty, J \to \infty} \hat{\text{DED}}(a) &= \text{DED}(a).
\end{align*}

Similarly, we can obtain
\begin{align*}
\text{plim}_{n_j \to \infty, J \to \infty} \hat{\text{DEY}}(a) &= \text{DEY}(a).
\end{align*}

Putting all together establishes the consistent estimation of the CADE. \( \square \)

**B.3 Asymptotic Normality of the ITT Effect Estimators Under Stratified Interference**

We provide the conditions for the asymptotic normality of the estimator of the ITT effect on the outcome. The conditions for the estimator of the ITT effect on the treatment receipt can be obtained in a similar fashion. Ohlsson (1989) establishes the asymptotic normality for two-stage sampling from a finite population. In our setting, we generalize his result to the two-stage randomized experiments by verifying the conditions required by his result in our context. We first state the following finite population central limit theorem.

**Theorem A3 (Finite Population Central Limit Theorem)** Let \( \bar{v}_S \) be the average of a simple random sample of size \( n \) from a finite population \( \{v_1, \ldots, v_N\} \). As \( N \to \infty \), if
\begin{align*}
\frac{1}{\min(n, N - n)} \cdot \max_{1 \leq i \leq N} (v_i - \bar{v}_N)^2 \rightarrow 0,
\end{align*}

where \( \bar{v}_N \) is the average of the population, then
\begin{align*}
(\bar{v}_S - \bar{v}_N)/\sqrt{\text{var}(\bar{v}_S)} & \xrightarrow{d} N(0, 1).
\end{align*}

39
Equation (A5) holds if \( v_i \)'s are bounded and \( n \) and \( N \) go to infinity.

We next introduce the central limit theorem under two-stage sampling.

**THEOREM A4 (FINITE POPULATION CENTRAL LIMIT THEOREM UNDER TWO-STAGE SAMPLING (OHLSSON, 1989))** Let \( v_{ij} \) be the outcome of interest of unit \( i \) in cluster \( j \), where \( i = 1, \ldots, n_j \) and \( J = 1, \ldots, J \). Define \( v_j = \sum_{i=1}^{n_j} v_{ij} \). Clusters are sampled from the population in the first stage and units are sampled in the second stage within the sampled clusters from the first stage. Let \( W_j \) be the sample indicator of the first stage and \( I_{ij} \) be the sample indicator of the second stage. Define \( T_j = \sum_{i=1}^{J} I_{ij} v_{ij}, T = \sum_{j=1}^{J} W_j T_j, Q = \sum_{j=1}^{J} W_j v_j, \sigma_j^2 = \text{var}(T_i \mid W_i, \ldots, W_j) \) and \( \mu_j^{(4)} = \mathbb{E}\{(T_j - v_j)^4 \mid W_i, \ldots, W_j\}. \) Then,

\[
\frac{T - \mathbb{E}(T)}{\sqrt{\text{var}(T)}} \xrightarrow{d} N(0,1)
\]

if the following three conditions hold

\[
\frac{Q - \mathbb{E}(Q)}{\sqrt{\text{var}(Q)}} \xrightarrow{d} N(0,1), \quad (A6)
\]

\[
\sum_{j=1}^{J} \mu_j^{(4)} \mathbb{E}(W_j^4) \rightarrow 0, \quad (A7)
\]

\[
\text{cov}(W_j^2, W_j^2) \leq 0, \text{ for } j \neq j'. \quad (A8)
\]

To apply Theorem A4, we decompose \( \hat{DEY}(a) \) as,

\[
\hat{DEY}(a) = \frac{1}{J} \sum_{j=1}^{J} I(A_j = a) \cdot \frac{n_j J}{N} \hat{DEY}_j(a) + \frac{1}{J} \sum_{j=1}^{J} I(A_j = a) \cdot \frac{n_j J}{N} \{\hat{DEY}_j(a) - \hat{DEY}_j(a)\}, \quad (A9)
\]

and \( \hat{SEY}(z) \) as,

\[
\hat{SEY}(z) = \sum_{j=1}^{J} A_j \cdot \frac{n_j J}{N} \left\{ \frac{\bar{Y}_j(z, 1)}{J_1} + \frac{\bar{Y}_j(z, 0)}{J_0} \right\} - \sum_{j=1}^{J} A_j \frac{n_j J}{N} \left\{ \frac{\bar{Y}_j(z, 1)}{J_0} - \frac{\bar{Y}_j(z, 0)}{J_0} \right\} + \frac{1}{J_1} \sum_{j=1}^{J} A_j \cdot \frac{n_j J}{N} \{\bar{Y}_j(z, 1) - \bar{Y}_j(z, 1)\} - \frac{1}{J_0} \sum_{j=1}^{J} (1 - A_j) \cdot \frac{n_j J}{N} \{\bar{Y}_j(z, 0) - \bar{Y}_j(z, 0)\}. \quad (A10)
\]

Denote the first part of each equation above as,

\[
\hat{DEY}_{\text{cluster}}(a) = \frac{1}{J} \sum_{j=1}^{J} I(A_j = a) \cdot \frac{n_j J}{N} \hat{DEY}_j(a),
\]

\[
\hat{SEY}_{\text{cluster}}(z) = \sum_{j=1}^{J} A_j \cdot \frac{n_j J}{N} \left\{ \frac{\bar{Y}_j(z, 1)}{J_1} + \frac{\bar{Y}_j(z, 0)}{J_0} \right\} - \sum_{j=1}^{J} A_j \frac{n_j J}{N} \left\{ \frac{\bar{Y}_j(z, 1)}{J_0} - \frac{\bar{Y}_j(z, 0)}{J_0} \right\}.
\]

We can treat \( \hat{DEY}_{\text{cluster}}(a) \) and \( \hat{SEY}_{\text{cluster}}(z) \) as \( Q \) in Theorem A4. For \( \hat{DEY}(a) \), we treat \( I(A_j = a) \) as \( W_j \) and \( \frac{n_j J}{N} \hat{DEY}_j(a) \) as \( v_{ij} \) in Theorem A4. For \( \hat{SEY}(z) \), we treat \( A_j \) as \( W_j \).
and $n_j J \frac{Y_{j}(z,1) + Y_{j}(z,0)}{J}$ as $v_{ij}$ in Theorem \ref{thm:asymptotic_normality}. We first give the regularity conditions for the asymptotic normality of $\widehat{\text{DEY}}_{\text{cluster}}^{(a)}$ and $\widehat{\text{SEY}}_{\text{cluster}}^{(z)}$:

(a) Equation (A5) holds for $n = J_1$, $N = J$ and $v_i = \frac{n_j J}{N} Y_{j}(z,a)$ for $z = 0, 1$ and $a = 0, 1$.

(b) Equation (A5) holds for $n = J_1$, $N = J$ and $v_i = \frac{n_j J}{N} Y_{j}(z,1)/J_1 + \frac{n_j J}{N} Y_{j}(z,0)/J_0$ for $z = 0, 1$.

For a bounded outcome, these two conditions are satisfied as the number of clusters goes to infinity. According to Theorem \ref{thm:asymptotic_normality}, under Condition (a), as $J \to \infty$, we have

$$\frac{\widehat{\text{DEY}}_{\text{cluster}}^{(a)} - \text{DEY}(a)}{\sqrt{\text{var}\{\widehat{\text{DEY}}_{\text{cluster}}^{(a)}\}}} \overset{d}{\to} N(0, 1).$$

(A11)

and under Condition (b), as $J \to \infty$, we have,

$$\frac{\widehat{\text{SEY}}_{\text{cluster}}^{(z)} - \text{SEY}(z)}{\sqrt{\text{var}\{\widehat{\text{SEY}}_{\text{cluster}}^{(z)}\}}} \overset{d}{\to} N(0, 1).$$

(A12)

Second, we require the following conditions,

(c) As $J \to \infty$

$$\frac{\sum_{j=1}^{J} \mathbb{E}[\{\widehat{\text{DEY}}_{j}(a) - \text{DEY}_{j}(a)\}^4 \mid A_j = a] \text{pr}(A_j = a)}{\left(\sum_{j=1}^{J} \mathbb{E}[\{\widehat{\text{DEY}}_{j}(a) - \text{DEY}_{j}(a)\}^2 \mid A_j = a] \text{pr}(A_j = a)\right)^2} \to 0$$

for $a = 0, 1$.

(d) As $J \to \infty$

$$\frac{\sum_{j=1}^{J} \mathbb{E}[\{\widehat{\text{SEY}}_{j}(z,a) - \text{SEY}_{j}(z,a)\}^4 \mid A_j = a] \text{pr}(A_j = a)}{\left(\sum_{j=1}^{J} \mathbb{E}[\{\widehat{\text{SEY}}_{j}(z,a) - \text{SEY}_{j}(z,a)\}^2 \mid A_j = a] \text{pr}(A_j = a)\right)^2} \to 0$$

for $z = 0, 1$ and $a = 0, 1$.

To give some intuition on when these two conditions hold, we show that Conditions (c) and (d) hold if the outcome is bounded and $\mathbb{E}[\{\widehat{\text{DEY}}_{j}(a) - \text{DEY}_{j}(a)\}^2 \mid A_j = a]$ and $\mathbb{E}[\{\widehat{\text{SEY}}_{j}(z,a) - \text{SEY}_{j}(z,a)\}^2 \mid A_j = a]$ are equal across different clusters. In this case, the term on the left hand side of Condition (c) is of the same order as

$$\frac{1}{J} \frac{\mathbb{E}[\{\widehat{\text{DEY}}_{j}(a) - \text{DEY}_{j}(a)\}^4 \mid A_j = a]}{\mathbb{E}[\{\widehat{\text{DEY}}_{j}(a) - \text{DEY}_{j}(a)\}^2 \mid A_j = a]^2 \text{pr}(A_j = a)},$$

which converges to zero if the outcome is bounded and $J$ goes to infinity. Therefore, Condition (c) holds. Similar argument applies to Condition (d).
We now verify the conditions in Theorem A4. For $\widehat{DEY}(a)$, Condition (A6) follows from equation (A11). Condition (A7) follows from Condition (c) above, and Condition (A8) follows from $\text{cov}(A_j, A_{j'}) \leq 0$ for $j \neq j'$. Therefore, as $J \to \infty$, we have,

$$
\frac{\text{DEY}(a) - \text{DEY}(a)}{\text{var} \left\{ \text{DEY}(a) \right\}} \xrightarrow{d} N(0, 1).
$$

Similarly, for $\text{SEY}(z)$, Condition (A6) in follows from equation (A12), Condition (A7) follows from Condition (d) above, and Condition (A8) follows from $\text{cov}(A_j, A_{j'}) \leq 0$ for $j \neq j'$. Therefore, as $J \to \infty$, we have,

$$
\frac{\text{SEY}(z) - \text{SEY}(z)}{\text{var} \left\{ \text{SEY}(z) \right\}} \xrightarrow{d} N(0, 1).
$$

Analogous conditions can be provided for the the asymptotic normality of $\widehat{DED}(a)$ and $\widehat{SED}(z)$. In a similar way, we can show the asymptotic normality for any linear combinations of $\{\widehat{DED}(a), \widehat{DEY}(a)\}$ and $\{\widehat{SED}(z), \text{SEY}(z)\}$. The conditions are satisfied if $Y$ is bounded. As a result, we can further show that

$$
\begin{pmatrix}
\text{var} \left\{ \widehat{DED}(a) \right\} & \text{cov} \left\{ \widehat{DED}(a), \widehat{DEY}(a) \right\} \\
\text{cov} \left\{ \widehat{DED}(a), \widehat{DEY}(a) \right\} & \text{var} \left\{ \widehat{DEY}(a) \right\}
\end{pmatrix}
^{-1/2}
\begin{pmatrix}
\widehat{DED}(a) - \text{DED}(a) \\
\widehat{DEY}(a) - \text{DEY}(a)
\end{pmatrix}
\xrightarrow{d} N_2(0_2, I_2),
$$

$$
\begin{pmatrix}
\text{var} \left\{ \widehat{SED}(z) \right\} & \text{cov} \left\{ \widehat{SED}(z), \text{SEY}(z) \right\} \\
\text{cov} \left\{ \widehat{SED}(z), \text{SEY}(z) \right\} & \text{var} \left\{ \text{SEY}(z) \right\}
\end{pmatrix}
^{-1/2}
\begin{pmatrix}
\widehat{SED}(z) - \text{SED}(z) \\
\text{SEY}(z) - \text{SEY}(z)
\end{pmatrix}
\xrightarrow{d} N_2(0_2, I_2).
$$

In general, the asymptotic normality of the ITT effects only requires some mild conditions as long as the outcome is bounded and $J$ goes to infinity. We leave the development of more refined CLTs under the two-stage randomized experiments to future work.

Finally, because $\text{CADE}(a)$ equals the ratio of $\widehat{DEY}(a)$ and $\widehat{DED}(a)$, and $\text{CASE}(z)$ equals the ratio of $\text{SEY}(a)$ and $\widehat{SED}(a)$, we can obtain the CLT for the CADE and CASE by applying the Delta method.

**B.4 Proof of Theorems 3 and A1**

According to the asymptotic normality results shown in Appendix B.3, we have

$$
\lim_{n_j \to \infty, J \to \infty} \widehat{DEY}(a) = \text{DEY}(a), \quad \lim_{n_j \to \infty, J \to \infty} \widehat{DED}(a) = \text{DED}(a).
$$

As a result,

$$
\lim_{n_j \to \infty, J \to \infty} \text{CADE}(a) = \lim_{n_j \to \infty, J \to \infty} \frac{\widehat{DEY}(a)}{\widehat{DED}(a)}.
$$

For the CASE, under Assumption A1, we have

$$
\text{SED}(z) = \sum_{j=1}^{J} \sum_{i=1}^{n_j} \{D_{ij}(z, 1) - D_{ij}(z, 0)\} = \sum_{j=1}^{J} \sum_{i=1}^{n_j} C_{ij}(z).
$$
We can then obtain

$$\text{SEY}(z) = \sum_{j=1}^{J} \sum_{i=1}^{n_j} \{Y_{ij}(z, 1) - Y_{ij}(z, 0)\} = \sum_{j=1}^{J} \sum_{i=1}^{n_j} \{Y_{ij}(z, 1) - Y_{ij}(z, 1)\} \{D_{ij}(z, 1) - D_{ij}(z, 0)\}$$

$$= \sum_{j=1}^{J} \sum_{i=1}^{n_j} \{Y_{ij}(z, 1) - Y_{ij}(z, 1)\} I\{D_{ij}(z, 1) = 1, D_{ij}(z, 0) = 0\},$$

where the second equality follows from Assumption A2. Thus, we have the following equality,

$$\text{CASE}(z) = \frac{\text{SEY}(z)}{\text{SED}(z)}.$$

According to the asymptotic normality results shown in Appendix B.3, we have

$$\text{plim} \hat{\text{SEY}}(z) = \text{SEY}(z), \quad \text{plim} \hat{\text{SED}}(a) = \text{SED}(a).$$

This implies the nonparametric identification and consistent estimation of the CASE. □

B.5 Proof of Theorem 4

We prove a general version of Theorem 4 using the general weight $w_j^*$ whereas in the main text, we consider the special case with $w_j^* = n_j J / N$. Using this general weight, we can rewrite the causal quantities as follows,

$$\text{DED}(a) = \frac{1}{J} \sum_{j=1}^{J} \frac{w_j^*}{n_j} \sum_{i=1}^{n_j} D_{ij}(1, a) - D_{ij}(0, a), \quad \text{SED}(z) = \frac{1}{J} \sum_{j=1}^{J} \frac{w_j^*}{n_j} \sum_{i=1}^{n_j} D_{ij}(z, 1) - D_{ij}(z, 0),$$

$$\text{DEY}(a) = \frac{1}{J} \sum_{j=1}^{J} \frac{w_j^*}{n_j} \sum_{i=1}^{n_j} Y_{ij}(1, a) - Y_{ij}(0, a), \quad \text{SEY}(z) = \frac{1}{J} \sum_{j=1}^{J} \frac{w_j^*}{n_j} \sum_{i=1}^{n_j} Y_{ij}(z, 1) - Y_{ij}(z, 0),$$

$$\text{CADE}(a) = \frac{\sum_{j=1}^{J} \frac{w_j^*}{n_j} \sum_{i=1}^{n_j} \{Y_{ij}(1, a) - Y_{ij}(0, a)\} I\{D_{ij}(1, a) = 1, D_{ij}(0, a) = 0\}}{\sum_{j=1}^{J} \frac{w_j^*}{n_j} \sum_{i=1}^{n_j} I\{D_{ij}(1, a) = 1, D_{ij}(0, a) = 0\}},$$

$$\text{CASE}(z) = \frac{\sum_{j=1}^{J} \frac{w_j^*}{n_j} \sum_{i=1}^{n_j} \{Y_{ij}(z, 1) - Y_{ij}(z, 0)\} I\{D_{ij}(z, 1) = 1, D_{ij}(z, 0) = 0\}}{\sum_{j=1}^{J} \frac{w_j^*}{n_j} \sum_{i=1}^{n_j} I\{D_{ij}(z, 1) = 1, D_{ij}(z, 0) = 0\}}.$$

The corresponding estimators for the ITT effects can be written as,

$$\hat{\text{DED}}(a) = \frac{1}{J_a} \sum_{j=1}^{J} w_j^* \hat{D}_j(1, a) I(A_j = a) - \frac{1}{J_a} \sum_{j=1}^{J} w_j^* \hat{D}_j(0, a) I(A_j = a),$$

$$\hat{\text{SED}}(z) = \frac{1}{J_1} \sum_{j=1}^{J} w_j^* \hat{D}_j(z, 1) I(A_j = 1) - \frac{1}{J_0} \sum_{j=1}^{J} w_j^* \hat{D}_j(z, 0) I(A_j = 0),$$

$$\hat{\text{DEY}}(a) = \frac{1}{J_a} \sum_{j=1}^{J} w_j^* \hat{Y}_j(1, a) I(A_j = a) - \frac{1}{J_a} \sum_{j=1}^{J} w_j^* \hat{Y}_j(0, a) I(A_j = a),$$

$$\hat{\text{CADE}}(a) = \frac{\sum_{j=1}^{J} \frac{w_j^*}{n_j} \sum_{i=1}^{n_j} \{Y_{ij}(1, a) - Y_{ij}(0, a)\} I\{D_{ij}(1, a) = 1, D_{ij}(0, a) = 0\}}{\sum_{j=1}^{J} \frac{w_j^*}{n_j} \sum_{i=1}^{n_j} I\{D_{ij}(1, a) = 1, D_{ij}(0, a) = 0\}},$$

$$\hat{\text{CASE}}(z) = \frac{\sum_{j=1}^{J} \frac{w_j^*}{n_j} \sum_{i=1}^{n_j} \{Y_{ij}(z, 1) - Y_{ij}(z, 0)\} I\{D_{ij}(z, 1) = 1, D_{ij}(z, 0) = 0\}}{\sum_{j=1}^{J} \frac{w_j^*}{n_j} \sum_{i=1}^{n_j} I\{D_{ij}(z, 1) = 1, D_{ij}(z, 0) = 0\}}.$$
\[ \hat{SEY}(z) = \frac{1}{J_1} \sum_{j=1}^{J} w_j^s \hat{Y}_j(z,1) I(A_j = 1) - \frac{1}{J_0} \sum_{j=1}^{J} w_j^s \hat{Y}_j(z,0) I(A_j = 0), \]

where
\[
\hat{D}_j(z, a) = \frac{\sum_{i=1}^{n_j} D_{ij} I(Z_{ij} = z)}{\sum_{i=1}^{n_j} I(Z_{ij} = z)}, \quad \hat{Y}_j(z, a) = \frac{\sum_{i=1}^{n_j} Y_{ij} I(Z_{ij} = z)}{\sum_{i=1}^{n_j} I(Z_{ij} = z)}.
\]

Theory of simple random sampling implies,
\[
\mathbb{E}\{\hat{D}_j(z, a) | A_j = a\} = \frac{1}{n_j} \sum_{i=1}^{n_j} D_{ij}(z, a), \quad \mathbb{E}\{\hat{Y}_j(z, a) | A_j = a\} = \frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij}(z, a).
\]

Thus, it is straightforward to show that the estimators for the ITT direct and spillover effects are unbiased. Below, without loss of generality, we prove the theorem only for the case with \( w_j^s = 1 \) since the general results can be obtained by simply transforming the outcome \( Y_{ij}^* = w_j^* Y_{ij} \) and treatment receipt \( D_{ij}^* = w_j^* D_{ij} \).

First, note that
\[
\text{cov}(Z_{ij}, Z_{i'j}) = \begin{cases} 
\frac{n_{i1}}{n_j} \left(1 - \frac{n_{i1}}{n_j}\right) & \text{if } i = i', \\
-\frac{n_{i1}}{n_j(n_j-1)} \left(1 - \frac{n_{i1}}{n_j}\right) & \text{if } i \neq i'.
\end{cases}
\]

Then, from Assumption 1 we have
\[
\text{var}\{\hat{Y}_j(z, a) | A_j = a\} = \frac{1}{n_{jz}^2} \text{var}\left\{\sum_{i=1}^{n_j} Y_{ij} I(Z_{ij} = z) | A_j = a\right\}
\]
\[
= \frac{1}{n_{jz}^2} \sum_{i=1}^{n_j} Y_{ij}^2(z, a) \text{cov}\{I(Z_{ij} = z), I(Z_{ij} = z)\}
\]
\[
+ \frac{1}{n_{jz}^2} \sum_{i \neq i'} Y_{ij}(z, a) Y_{i'j}(z, a) \text{cov}\{I(Z_{ij} = z), I(Z_{i'j} = z)\}
\]
\[
= \frac{1}{n_j n_{jz}} \left(1 - \frac{n_{jz}}{n_j}\right) \sum_{i=1}^{n_j} Y_{ij}^2(z, a) - \frac{1}{n_j n_{jz} (n_j - 1)} \sum_{i \neq i'} Y_{ij}(z, a) Y_{i'j}(z, a)
\]
\[
= \frac{1}{n_j n_{jz}} \left(1 - \frac{n_{jz}}{n_j}\right) \sum_{i=1}^{n_j} Y_{ij}^2(z, a) - \frac{1}{n_j n_{jz} (n_j - 1)} \left[\sum_{i=1}^{n_j} Y_{ij}(z, a)^2 - \sum_{i=1}^{n_j} Y_{ij}^2(z, a)\right]
\]
\[
= \frac{1}{(n_j - 1) n_{jz}} \left(1 - \frac{n_{jz}}{n_j}\right) \sum_{i=1}^{n_j} Y_{ij}^2(z, a) - \frac{n_j}{n_j (n_j - 1)} \left(1 - \frac{n_{jz}}{n_j}\right) \bar{Y}_j^2(z, a)
\]
\[
= \frac{1}{n_{jz}} \left(1 - \frac{n_{jz}}{n_j}\right) \sigma_j^2(z, a),
\]

and
\[
\text{cov}\{\hat{Y}_j(1, a), \hat{Y}_j(0, a) | A_j = a\}
\]
Therefore, we obtain,

\[ \begin{align*}
& = \frac{1}{n_1n_0} \sum_{i=1}^{n_j} Y_{ij}(1, a) Y_{ij}(0, a) \text{cov}\{I(Z_{ij} = 1), I(Z_{ij} = 0)\} \\
& \quad + \frac{1}{n_1n_0} \sum_{i \neq i'} Y_{ij}(1, a) Y_{i'j}(0, a) \text{cov}\{I(Z_{ij} = 1), I(Z_{i'j} = 0)\} \\
& = -\frac{1}{n_j} \sum_{i=1}^{n_j} Y_{ij}(1, a) Y_{ij}(0, a) + \frac{1}{n_j^2(n_j - 1)} \sum_{i \neq i'} Y_{ij}(1, a) Y_{i'j}(0, a) \\
& = -\frac{1}{(n_j - 1)n_j} \sum_{i=1}^{n_j} Y_{ij}(1, a) Y_{ij}(0, a) + \frac{1}{n_j - 1} \sum_{i=1}^{n_j} \{Y_{ij}(1, a) - \bar{Y}_j(1, a)\} \{Y_{ij}(0, a) - \bar{Y}_j(0, a)\}. 
\end{align*} \]

Next, we compute the variance of SEY. From Assumption I, we have,

\[ \begin{align*}
& = \text{var}\{\text{DEY}(a)\} \\
& = \mathbb{E} \left[ \frac{1}{J_a} \sum_{j=1}^{J} \text{var}\{\text{DEY}(a) \mid A_j = a\} I(A_j = a) \right] + \text{var} \left\{ \frac{1}{J_a} \sum_{j=1}^{J} \text{DEY}(a) I(A_j = a) \right\} \\
& = \left( 1 - \frac{J_a}{J} \right) \frac{\sigma^2_{DE}(a)}{J_a} + \frac{1}{J_a J} \sum_{j=1}^{J} \text{var} \left\{ \text{DEY}(a) \mid A_j = a \right\}. 
\end{align*} \]

Next, we compute the variance of SEY(z). From Assumption I, we have,

\[ \begin{align*}
& \quad \text{cov}\{\hat{Y}(z, 1), \hat{Y}(z, 0)\} \\
& = \frac{1}{J_1 J_0} \text{cov} \left\{ \sum_{j=1}^{J} \hat{Y}_j(z, 1) I(A_j = 1), \sum_{j=1}^{J} \hat{Y}_j(z, 0) I(A_j = 0) \right\} \\
& = \frac{1}{J_1 J_0} \mathbb{E} \left[ \text{cov} \left\{ \sum_{j=1}^{J} \hat{Y}_j(z, 1) I(A_j = 1), \sum_{j=1}^{J} \hat{Y}_j(z, 0) I(A_j = 0) \mid A_1, \ldots, A_J \right\} \right] \\
& \quad + \frac{1}{J_1 J_0} \text{cov} \left\{ \sum_{j=1}^{J} \bar{Y}_j(z, 1) I(A_j = 1), \sum_{j=1}^{J} \bar{Y}_j(z, 0) I(A_j = 0) \right\} \\
& = \frac{1}{J_1 J_0} \text{cov} \left\{ \sum_{j=1}^{J} \bar{Y}_j(z, 1) I(A_j = 1), \sum_{j=1}^{J} \bar{Y}_j(z, 0) I(A_j = 0) \right\} \\
& = \frac{1}{J_1 J_0} \sum_{j=1}^{J} \bar{Y}_j(z, 1) \bar{Y}_j(z, 0) \text{cov}\{I(A_j = 1), I(A_j = 0)\}. 
\end{align*} \]
follows from the conditional independence where the second equality follows from the law of total variance and the third equality follows from the conditional independence 

\[ \mathbf{Z}_j \perp \perp \mathbf{Z}_{j'} \mid (A_1, \ldots, A_J) \text{ for } j \neq j'. \]

Therefore, we have

\[
\text{var}\{\widehat{\text{SE}}\mathbf{Y}(z)\} = \text{var}\{\widehat{\mathbf{Y}}(z, 1)\} + \text{var}\{\widehat{\mathbf{Y}}(z, 0)\} - 2\text{cov}\{\widehat{\mathbf{Y}}(z, 1), \widehat{\mathbf{Y}}(z, 0)\}
\]

\[
= \left(1 - \frac{J_1}{J}\right) \frac{\sigma_b^2(z, 1)}{J_1} + \left(1 - \frac{J_0}{J}\right) \frac{\sigma_b^2(z, 0)}{J_0} - \frac{1}{J} \{\sigma_{\text{SE}}^2(z) - \sigma_b^2(z, 1) - \sigma_b^2(z, 0)\}
\]

\[
+ \frac{1}{J_1 J} \sum_{j=1}^J \text{var}\{\widehat{Y}_j(z, 1) \mid A_j = 1\} + \frac{1}{J_0 J} \sum_{j=1}^J \text{var}\{\widehat{Y}_j(z, 0) \mid A_j = 0\}
\]

\[
= \frac{\sigma_b^2(z, 1)}{J_1} + \frac{\sigma_b^2(z, 0)}{J_0} - \frac{\sigma_{\text{SE}}^2(z)}{J} + \frac{1}{J_1 J} \sum_{j=1}^J \frac{1}{n_j} \left(1 - \frac{n_{jz}}{n_j}\right) \sigma_j^2(z, 1)
\]

\[
+ \frac{1}{J_0 J} \sum_{j=1}^J \frac{1}{n_j} \left(1 - \frac{n_{jz}}{n_j}\right) \sigma_j^2(z, 0).
\]

\[\square\]

### B.6 Covariances

#### B.6.1 \( \text{cov}(\text{DEY}(a), \text{DED}(a)) \)

We introduce some notation. Define

\[
\zeta_{jzz'}(a) = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} \{Y_{ij}(z, a) - \overline{Y}_j(z, a)\} \{D_{ij}(z', a) - \overline{D}_j(z', a)\},
\]

\[
\zeta_{j(1-0)}(a) = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} \{\text{DEY}_{ij}(a) - \overline{\text{DEY}}_j(a)\} \{\text{DED}_{ij}(a) - \overline{\text{DED}}_j(a)\}.
\]
Similar to the calculation of $\text{var}\{\hat{Y}_j(z, a) \mid A_j = a\}$ and $\text{cov}\{\hat{Y}_j(1, a), \hat{Y}_j(0, a) \mid A_j = a\}$ in the proof of Theorem 4, it is easy to show that

$$
\begin{align*}
\text{cov}\{\hat{Y}_j(1, a), \hat{D}_j(0, a) \mid A_j = a\} &= -\frac{1}{n_j} \zeta_{j01}(a), \\
\text{cov}\{\hat{Y}_j(0, a), \hat{D}_j(1, a) \mid A_j = a\} &= -\frac{1}{n_j} \zeta_{j01}(a), \\
\text{cov}\{\hat{Y}_j(z, a), \hat{D}_j(z, a) \mid A_j = a\} &= \frac{1}{n_j} \left(1 - \frac{n_j}{n_j} \right) \zeta_{jzz}(a).
\end{align*}
$$

Thus, we can obtain

$$
\begin{align*}
\text{cov}\left\{\text{DEY}_j(a), \text{DED}_j(a) \mid A_j = a\right\} &= \sum_{z=0,1} \frac{1}{n_j} \left(1 - \frac{n_j}{n_j} \right) \zeta_j(z, a) + \frac{1}{n_j} \zeta_{j10}(a) + \frac{1}{n_j} \zeta_{j01}(a) \\
&= \frac{\zeta_{j11}(a)}{n_j} + \frac{\zeta_{j00}(a)}{n_j} - \frac{\zeta_{j(1-0)}(a)}{n_j}.
\end{align*}
$$

As a result,

$$
\begin{align*}
\text{cov}\left\{\text{DEY}(a), \text{DED}(a)\right\} &= \mathbb{E} \left[ \frac{1}{J^2 a} \sum_{j=1}^J \text{cov}\{\text{DEY}_j(a), \text{DED}_j(a) \mid A_j = a\} I(A_j = a) \right] \\
&\quad + \text{cov} \left[ \frac{1}{J a} \sum_{j=1}^J \text{DEY}_j(a) I(A_j = a) \right] \left( \frac{1}{J a} \sum_{j=1}^J \text{DED}_j(a) I(A_j = a) \right) \\
&= \left(1 - \frac{J a}{J}\right) \frac{\zeta_{DE}(a)}{J a} + \frac{1}{J a J} \sum_{j=1}^J \text{cov}\left\{\text{DEY}_j(a), \text{DED}_j(a) \mid A_j = a\right\}.
\end{align*}
$$

B.6.2 $\text{cov}(\text{SEY}(a), \text{SED}(a))$

We introduce some notation. Define

$$
\begin{align*}
\zeta_6^2(z, a) &= \frac{1}{J-1} \sum_{j=1}^J \{\hat{Y}_j(z, a) - \overline{Y}(z, a)\}\{\overline{D}_j(z, a) - \overline{D}(z, a)\}, \\
\zeta_{6E}^2(z) &= \frac{1}{J-1} \sum_{j=1}^J \{|\text{SEY}_j(z) - \text{SEY}(z, a)\} |\text{SED}_j(z) - \text{SEY}(z, a)\}.
\end{align*}
$$

We can decompose $\text{cov}\{\text{SEY}(z), \text{SED}(z)\}$ as,

$$
\text{cov}\{\text{SEY}(z), \text{SED}(z)\} = \text{cov}\{\hat{Y}(z, 1), \hat{D}(z, 1)\} + \text{cov}\{\hat{Y}(z, 0), \hat{D}(z, 0)\} - \text{cov}\{\hat{Y}(z, 1), \hat{D}(z, 0)\} - \text{cov}\{\hat{Y}(z, 0), \hat{Y}(z, 0)\}.
$$

We then calculate each component,

$$
\begin{align*}
\text{cov}\{\hat{Y}(z, a), \hat{D}(z, a)\} &= \mathbb{E} \left[ \frac{1}{J a} \sum_{j=1}^J \text{cov}\{\hat{Y}_j(z, a), \hat{D}_j(z, a) \mid A_j = a\} I(A_j = a) \right] \\
&\quad + \text{cov} \left\{ \frac{1}{J a} \sum_{j=1}^J \hat{Y}_j(z, a) I(A_j = a) \right\} \left( \frac{1}{J a} \sum_{j=1}^J \hat{D}_j(z, a) I(A_j = a) \right).
\end{align*}
$$

47
Thus, we first show that the variance estimator is conservative for \( \text{var} \). Therefore, we have

\[
\text{cov}\{\hat{Y}(z, 1), \hat{D}(z, 0)\} = \text{cov}\left\{ \frac{1}{J_1} \sum_{j=1}^{J} Y_j(z, 1) I(A_j = 1), \frac{1}{J_0} \sum_{j=1}^{J} D_j(z, 0) I(A_j = 0) \right\}
\]

\[
= -\frac{1}{J(J - 1)} \sum_{j=1}^{J} \{Y_j(z, 1) - \bar{Y}(z, 1)\} \{D_j(z, 0) - \bar{D}(z, 0)\},
\]

\[
\text{cov}\{\hat{Y}(z, 0), \hat{D}(z, 1)\} = \text{cov}\left\{ \frac{1}{J_0} \sum_{j=1}^{J} Y_j(z, 0) I(A_j = 0), \frac{1}{J_1} \sum_{j=1}^{J} D_j(z, 1) I(A_j = 1) \right\}
\]

\[
= -\frac{1}{J(J - 1)} \sum_{j=1}^{J} \{Y_j(z, 0) - \bar{Y}(z, 0)\} \{D_j(z, 1) - \bar{D}(z, 1)\}.
\]

Therefore,

\[
\text{cov}\{\hat{Y}(z, 1), \hat{D}(z, 0)\} + \text{cov}\{\hat{Y}(z, 0), \hat{D}(z, 1)\}
\]

\[
= -\frac{1}{J(J - 1)} \sum_{j=1}^{J} \{Y_j(z, 0) - \bar{Y}(z, 0)\} \{D_j(z, 1) - \bar{D}(z, 1)\}
\]

\[
-\frac{1}{J(J - 1)} \sum_{j=1}^{J} \{Y_j(z, 1) - \bar{Y}(z, 1)\} \{D_j(z, 0) - \bar{D}(z, 0)\}
\]

\[
= \frac{1}{J}\{\zeta_{SE}^2 - \zeta_b^2(z, 1) - \zeta_b^2(z, 0)\}.
\]

Thus,

\[
\text{cov}\{SE\hat{Y}(z), SE\hat{D}(z)\}
\]

\[
= \left( 1 - \frac{J_1}{J} \right) \frac{\zeta_b^2(z, 1)}{J_1} + \frac{1}{J_1 J} \sum_{j=1}^{J} \text{cov}\{\hat{Y}_j(z, 1), \hat{D}_j(z, 1) \mid A_j = 1\} + \left( 1 - \frac{J_0}{J} \right) \frac{\zeta_b^2(z, 0)}{J_0}
\]

\[
+ \frac{1}{J_0 J} \sum_{j=1}^{J} \text{cov}\{\hat{Y}_j(z, 0), \hat{D}_j(z, 0) \mid A_j = 0\} - \frac{1}{J}\{\zeta_{SE}^2 - \zeta_b^2(z, 1) - \zeta_b^2(z, 0)\}
\]

\[
= \frac{\zeta_b^2(z, 1)}{J_1} + \frac{\zeta_b^2(z, 0)}{J_0} - \frac{\zeta_{SE}^2(z)}{J} + \frac{1}{J_1 J} \sum_{j=1}^{J} \text{cov}\{\hat{Y}_j(z, 1), \hat{D}_j(z, 1) \mid A_j = 1\}
\]

\[
+ \frac{1}{J_0 J} \sum_{j=1}^{J} \text{cov}\{\hat{Y}_j(z, 0), \hat{D}_j(z, 0) \mid A_j = 0\}.
\]

**B.7 Variance Estimators for the ITT Effects**

We first show that the variance estimator is conservative for \( \text{var}\{\hat{D}E\hat{Y}(a)\} \). From the classical theory of simple random sampling, we know \( \mathbb{E}\{\hat{\sigma}_j^2(z, a) \mid A_j = a\} = \sigma_j^2(z, a) \). In addition, we have

\[
\mathbb{E}\{\hat{\sigma}_{DE}^2(a)\}
\]
and

\[
\mathbb{E}\{\hat{\sigma}_b^2(z, a)\} = \frac{1}{J_a - 1} \mathbb{E}\left\{\sum_{j=1}^{J} \hat{Y}_j^2(z, a) I(A_j = a) - J_a \hat{Y}(z, a)^2\right\} \\
= \frac{1}{J_a - 1} \mathbb{E}\left(\sum_{j=1}^{J} \var{\hat{Y}_j(z, a)} | A_j = a\right) + \frac{J_a(J - 1)}{(J_a - 1)J} \sigma_{DE}^2(a) - \frac{J_a}{J_a - 1} \var{\hat{Y}(z, a)} \\
= \frac{J_a}{J_a - 1} \sum_{j=1}^{J} \var{\hat{Y}_j(z, a)} | A_j = a\right) + \frac{J_a}{J_a - 1} \frac{J_a(J - 1)}{(J_a - 1)J} \sigma_{DE}^2(a) - \frac{J_a}{J_a - 1} \var{\hat{Y}(z, a)} \\
= \frac{J_a}{J_a - 1} \sum_{j=1}^{J} \var{\hat{Y}_j(z, a)} | A_j = a\right) + \frac{J_a}{J_a - 1} \frac{J_a(J - 1)}{(J_a - 1)J} \sigma_{DE}^2(a) - \frac{J_a}{J_a - 1} \var{\hat{Y}(z, a)} \\
= \frac{J_a}{J_a - 1} \left[\left(1 - \frac{J_a}{J}\right) \frac{\sigma_b^2(z, a)}{J_a} + \frac{1}{J_aJ} \sum_{j=1}^{J} \var{\hat{Y}_j(z, 1) | A_j = a}\right] \\
= \sigma_b^2(z, a) + \frac{1}{J} \sum_{j=1}^{J} \var{\hat{Y}_j(z, a) | A_j = a}. \\
\]

Therefore, we have,

\[
\mathbb{E}\{\var{\hat{DEY}(a)}\} \\
= \left(1 - \frac{J_a}{J}\right) \frac{\sigma_{DE}^2(a)}{J_a} + \left(1 - \frac{J_a}{J}\right) \frac{1}{J_aJ} \sum_{j=1}^{J} \var{\hat{DEY}_j(a) | A_j = a} \\
+ \frac{1}{J^2} \sum_{j=1}^{J} \left\{\frac{\sigma_b^2(1, a)}{n_{j1}} + \frac{\sigma_b^2(0, a)}{n_{j0}}\right\}
\]
\[
\begin{align*}
\text{var}\{\hat{\text{DEY}}(a)\} + \frac{1}{J_2} \sum_{j=1}^{J} \frac{\omega_j^2(a)}{n_{j1}} \\
\geq\ var\{\hat{\text{DEY}}(a)\}.
\end{align*}
\]

The variance estimator for \(SEY(z)\) is,
\[
\text{var}\left\{\hat{SEY}(z)\right\} = \frac{\hat{\sigma}_k^2(z,1)}{J_1} + \frac{\hat{\sigma}_k^2(z,0)}{J_0}.
\]

Therefore, we have
\[
\begin{align*}
\mathbb{E}[\text{var}\{\hat{SEY}(z)\}] &= \frac{\sigma_k^2(z,1)}{J_1} + \frac{\sigma_k^2(z,0)}{J_0} + \sum_{a=0}^{1} \frac{1}{J_a} \sum_{j=1}^{J} \text{var}\{\hat{Y}_j(z,a) | A_j = a\} \\
&= \text{var}\{\hat{SEY}(z)\} + \frac{\sigma_{SE}^2(z)}{J} \\
&\geq\ \text{var}\{\hat{SEY}(z)\}.
\end{align*}
\]

Next, we consider the estimator of \(\text{cov}\{\hat{\text{DEY}}(a), \hat{\text{DED}}(a)\}\). Similarly, we have \(\mathbb{E}\{\hat{\zeta}_j(z,a) | A_j = a\} = \zeta_j(z,a)\) and hence,
\[
\begin{align*}
\mathbb{E}\{\hat{\zeta}_{DE}^2(a)\} &= \frac{1}{J_a - 1} \mathbb{E} \left\{ \sum_{j=1}^{J} \hat{\text{DEY}}_j(a) \hat{\text{DED}}_j(a) I(A_j = a) - J_a \hat{\text{DEY}}(a) \hat{\text{DED}}(a) \right\} \\
&= \frac{1}{J_a - 1} \mathbb{E} \left( \sum_{j=1}^{J} \left[ \text{cov}\{\hat{\text{DEY}}_j(a), \hat{\text{DED}}_j(a) | A_j = a\} + \hat{\text{DEY}}_j(a) \hat{\text{DED}}_j(a) \right] I(A_j = a) \right) \\
&\quad - \frac{J_a}{J_a - 1} \left[ \text{cov}\{\hat{\text{DEY}}(a), \hat{\text{DED}}(a)\} + \hat{\text{DEY}}(a) \hat{\text{DED}}(a) \right] \\
&= \frac{J_a}{J(J_a - 1)} \sum_{j=1}^{J} \left[ \text{cov}\{\hat{\text{DEY}}_j(a), \hat{\text{DED}}_j(a) | A_j = a\} + \frac{J_a(J - 1)}{(J_a - 1)J} \zeta_{DE}(a) \right. \\
&\quad \left. - \frac{J_a}{J_a - 1} \text{cov}\{\hat{\text{DEY}}(a), \hat{\text{DED}}(a)\} \right] \\
&= \zeta_{DE}^2(a) + \frac{1}{J} \sum_{j=1}^{J} \left[ \text{cov}\{\hat{\text{DEY}}_j(a), \hat{\text{DED}}_j(a) | A_j = a\} \right].
\end{align*}
\]

Therefore, we have,
\[
\begin{align*}
\mathbb{E}[\text{cov}\{\hat{\text{DEY}}(a), \hat{\text{DED}}(a)\}] &= \left(1 - \frac{J_a}{J}\right) \frac{\zeta_{DE}^2(a)}{J_a} + \left(1 - \frac{J_a}{J}\right) \frac{1}{J_a} \sum_{j=1}^{J} \text{cov}\{\hat{\text{DEY}}_j(a), \hat{\text{DED}}_j(a) | A_j = a\} \\
&\quad + \frac{1}{J^2} \sum_{j=1}^{J} \left\{ \frac{\zeta_j(1,a)}{n_{j1}} + \frac{\zeta_j(0,a)}{n_{j0}} \right\}
\end{align*}
\]
\[
\begin{align*}
&= \text{cov}\{\hat{\text{DEY}}(a), \hat{\text{DED}}(a)\} + \frac{1}{J^2} \sum_{j=1}^{J} \frac{\zeta_{j(1-0)}(a)}{n_{j1}} \\
&\geq \text{cov}\{\hat{\text{DEY}}(a), \hat{\text{DED}}(a)\}.
\end{align*}
\]

### B.8 Asymptotically Conservative Variance Estimator for the CADE

Although \(\text{var}\{\hat{\text{CADE}}(a)\}\) in equation (9) is not a conservative variance estimator in finite samples, we show that it is asymptotically conservative. First, the asymptotic variance of \(\hat{\text{CADE}}(a)\) can be rewritten as,

\[
\text{var}\left[\frac{1}{\text{DED}(a)} \{\hat{\text{DEY}}(a) - \text{CADE}(a) \cdot \hat{\text{DED}}(a)\}\right],
\]

which is the variance of a linear combination of \(\hat{\text{DED}}(a)\) and \(\hat{\text{DEY}}(a)\). Similar to the proof in Section B.7, we can show,

\[
\mathbb{E}\left\{\frac{1}{\text{DED}(a)^2} \left[\text{var}\{\hat{\text{DEY}}(a)\} - 2 \frac{\text{DEY}(a)}{\text{DED}(a)} \text{cov}\{\hat{\text{DEY}}(a), \hat{\text{DED}}(a)\} + \frac{\text{DEY}(a)^2}{\text{DED}(a)^2} \text{var}\{\hat{\text{DED}}(a)\}\right]\right\}
\geq \text{var}\left[\frac{1}{\text{DED}(a)} \{\hat{\text{DEY}}(a) - \text{CADE}(a) \cdot \hat{\text{DED}}(a)\}\right].
\]

Under the restriction on interference in [Sävje et al., 2017], \(\hat{\text{DED}}(a)\) converges to \(\text{DED}(a)\) and \(\hat{\text{DEY}}(a)\) converges to \(\text{DEY}(a)\). Therefore, we obtain the desired result for a bounded outcome,

\[
\mathbb{E}\left[\text{var}\{\hat{\text{CADE}}(a)\}\right] \geq \text{avar}\{\hat{\text{CADE}}(a)\}. 
\]

### C Proofs for the Regression-Based Approach

As in case of the randomization inference approach, it is suffice to prove the case with \(w_{j+}^* = 1\) since the results are applicable directly with any other weight by multiplying \(D_{ij}\) and \(Y_{ij}\) with appropriate constants. Because the columns in the design matrix of the regression models corresponding to different treatment assignment mechanisms are orthogonal to each other, we can prove the results separately for each treatment assignment mechanism. Therefore, we prove the theorems for a given \(a\) and with abuse of notation use the same notation for the sub-matrix that consists of the columns corresponding to the treatment assignment mechanism \(a\) in a full matrix. For example, the proof of Theorem 6 uses \(X_{ij}\) to denote \((I(A_j = a), Z_{ij} I(A_j = a))\) while in the main text we use \(X_{ij}\) to represent \((I(A_j = 1), I(A_j = 0), Z_{ij} I(A_j = 1), Z_{ij} I(A_j = 0))\).

#### C.1 Proof of Theorem 6

Define,

\[
N_{za} = \sum_{j=1}^{J} \sum_{i=1}^{n_j} I(Z_{ij} = z, A_j = a)w_{ij} = \sum_{j=1}^{J} \sum_{i=1}^{n_j} I(Z_{ij} = z, A_j = a) \frac{1}{J a n_{jz}} = 1
\]

51
and $N_{+a} = N_{0a} + N_{1a}$. Then, the OLS estimate can be written as,

$$(\hat{\alpha}_a, \hat{\alpha}_{1a})^\top = (X^\top WX)^{-1}X^\top WY,$$

where

$$(X^\top WX)^{-1} = \begin{pmatrix} N_{+a} & N_{1a} \\ N_{1a} & N_{1a} \end{pmatrix}^{-1} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix},$$

$$X^\top WY = \begin{pmatrix} \hat{Y}(1, a) + \hat{Y}(0, a) \\ \hat{Y}(1, a) \end{pmatrix}.$$ 

Therefore, we have,

$$\hat{\alpha}_a = \hat{Y}(0, a), \quad \hat{\alpha}_{1a} = \text{DEY}(a).$$

\[\square\]

C.2 Proof of Theorem 7

We first write the two-stage least squares regression as,

$$Y_{ij} = \sum_{a=0}^{1} \beta_a I(A_j = a) + \sum_{a=0}^{1} \beta_{1a}\hat{D}_{ij}I(A_j = a) + \epsilon_{ij},$$

$$\hat{D}_{ij} = \sum_{a=0}^{1} \hat{\gamma}_{a}^{wls}I(A_j = a) + \sum_{a=0}^{1} \hat{\gamma}_{1a}^{wls}Z_{ij}I(A_j = a).$$

where $\hat{\gamma}_{a}^{wls}$ and $\hat{\gamma}_{1a}^{wls}$ are the weighted least squares estimate of the corresponding coefficients from the model given in equation (10). Then, we obtain,

$$Y_{ij} = \sum_{a=0}^{1} \beta_a I(A_j = a) + \sum_{a=0}^{1} \beta_{1a} \left\{ \sum_{a=0}^{1} \hat{\gamma}_{a}^{wls}I(A_j = a) + \sum_{a=0}^{1} \hat{\gamma}_{1a}^{wls}Z_{ij}I(A_j = a) \right\} I(A_j = a) + \epsilon_{ij}$$

$$= \sum_{a=0}^{1} \beta_a I(A_j = a) + \sum_{a=0}^{1} \beta_{1a} \left\{ \hat{\gamma}_{a}^{wls}I(A_j = a) + \hat{\gamma}_{1a}^{wls}Z_{ij}I(A_j = a) \right\} + \epsilon_{ij}$$

$$= \sum_{a=0}^{1} (\beta_a + \beta_{1a} \hat{\gamma}_{a}^{wls})I(A_j = a) + \sum_{a=0}^{1} \beta_{1a} \hat{\gamma}_{1a}^{wls}Z_{ij}I(A_j = a) + \epsilon_{ij}. $$

Comparison of this with the weighted regression model of $Y_{ij}$ on $Z_{ij}$ given in equation (11) implies,

$$\hat{\alpha}_{a}^{wls} = \hat{\beta}_{a}^{w2lds} + \hat{\beta}_{1a}^{w2lds} \hat{\gamma}_{a}^{wls}, \quad \hat{\alpha}_{1a}^{wls} = \hat{\beta}_{1a}^{w2lds} \hat{\gamma}_{1a}^{wls}.$$

Thus, using Theorem 6 we have,

$$\hat{\beta}_{1a}^{w2lds} = \text{CADE}(a), \quad \hat{\beta}_{a}^{w2lds} = \hat{Y}(0, a) - \text{CADE}(a) \cdot \hat{D}(0, a).$$
C.3 Proof of Theorem 8

We prove the results only for the direct effects on the outcome. The results for the direct effects on the treatment receipt are similar.

\[
P_j = W_j^{1/2} X_j (X^T W X)^{-1} X_j^T W_j^{1/2}
\]

\[
= \begin{pmatrix}
\frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j1}} & \frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j2}} \\
\frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}} & \frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}}
\end{pmatrix} \begin{pmatrix} 1 & 1 \\
-1 & 2 \end{pmatrix} \begin{pmatrix}
\frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j1}} & \frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j2}} \\
\frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{1}_{n_{j0}} & \frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}}
\end{pmatrix}^T
\]

\[
= \begin{pmatrix}
\frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j1}} & \frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j2}} \\
\frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}} & \frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}}
\end{pmatrix} \begin{pmatrix}
\frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j1}} & \frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j2}} \\
\frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{1}_{n_{j0}} & \frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}}
\end{pmatrix}^T
\]

where \( \mathbf{1}_m \) (\( \mathbf{0}_m \)) is an \( m \)-dimensional vector of ones (zeros) and \( \mathbf{1}_{m_1 \times m_2} \) (\( \mathbf{0}_{m_1 \times m_2} \)) is an \( m_1 \times m_2 \) dimensional matrix of ones (zeros).

Since \( (\mathbf{1}_{n_{j1}}^T, \mathbf{0}_{n_{j0}}^T)^T \) and \( (\mathbf{0}_{n_{j1}}^T, \mathbf{1}_{n_{j0}}^T)^T \) are two eigenvectors of \( I_{n_j} - P_j \) whose eigenvalue is \( (J_a - 1) / J_a \), we have,

\[
(I_{n_j} - P_j)^{-1/2} (\mathbf{1}_{n_{j1}}^T, 0_{n_{j0}}^T)^T = \sqrt{\frac{J_a}{J_a - 1}} \begin{pmatrix}
\frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j1}} & \frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j2}} \\
\frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}} & \frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}}
\end{pmatrix}^T,
\]

\[
(I_{n_j} - P_j)^{-1/2} (0_{n_{j1}}^T, 1_{n_{j0}}^T)^T = \sqrt{\frac{J_a}{J_a - 1}} \begin{pmatrix}
\frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{0}_{n_{j1}} & \frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j2}} \\
\frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{1}_{n_{j0}} & \frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}}
\end{pmatrix}^T.
\]

Thus,

\[
(I_{n_j} - P_j)^{-1/2} W_j X_j = \sqrt{\frac{J_a}{J_a - 1}} \begin{pmatrix}
\frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j1}} & \frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j2}} \\
\frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}} & \frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}}
\end{pmatrix}
\]

For a unit with \( (A_j = a, Z_{ij} = 1) \), we have \( \hat{\epsilon}_{ij} = Y_{ij} - \hat{\epsilon}_a = \hat{\epsilon}_{1a} = Y_{ij} - \hat{Y}(1, a) \), and for a unit with \( (A_j = a, Z_{ij} = 0) \), we have \( \hat{\epsilon}_{ij} = Y_{ij} - \hat{\epsilon}_a = Y_{ij} - \hat{Y}(0, a) \). As a result,

\[
\hat{\epsilon}_j^T (I_{n_j} - P_j)^{-1/2} W_j X_j
\]

\[
= \sqrt{\frac{J_a}{J_a - 1}} (Y_{ij} - \hat{Y}(1, a), \ldots, Y_{n_{j1}} - \hat{Y}(0, a)) \begin{pmatrix}
\frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j1}} & \frac{1}{\sqrt{J_{a} n_{j1}}} \mathbf{1}_{n_{j2}} \\
\frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}} & \frac{1}{\sqrt{J_{a} n_{j0}}} \mathbf{0}_{n_{j0}}
\end{pmatrix}^T
\]

\[
= \sqrt{\frac{J_a}{J_a - 1}} \begin{pmatrix}
\frac{1}{J_{a} n_{j1}} \left\{ \sum_{i=1}^{n_{j1}} Y_{ij} Z_{ij} - n_{j1} \hat{Y}(1, 1) \right\} + \frac{1}{J_{a} n_{j0}} \left\{ \sum_{i=1}^{n_{j0}} Y_{ij} (1 - Z_{ij}) - n_{j0} \hat{Y}(0, 1) \right\}
\end{pmatrix}^T
\]

\[
\frac{1}{J_{a} n_{j1}} \left\{ \sum_{i=1}^{n_{j1}} Y_{ij} Z_{ij} - n_{j1} \hat{Y}(1, 1) \right\}
\]

\[
\frac{1}{J_{a} (J_a - 1)} \left( \hat{Y}_j(1, 1) - \hat{Y}(1, 1) + \hat{Y}_j(0, 1) - \hat{Y}(0, 1) \right)
\]

Let \( V_j = X_j^T W_j (I_{n_j} - P_j)^{-1/2} \hat{\epsilon}_j^T (I_{n_j} - P_j)^{-1/2} W_j X_j \) and \( v_{jk_1 k_2} \) be the \( k_1 k_2 \)-th entry of \( V_j \). Then,

\[
v_{j11} = \frac{1}{J_{a} (J_a - 1)} \left\{ \hat{Y}_j(1, 1) - \hat{Y}(1, 1) + \hat{Y}_j(0, 1) - \hat{Y}(0, 1) \right\}^2,
\]
\[ v_{j12} = v_{j22} + \frac{1}{J_a(J_a - 1)} \left\{ \hat{Y}_j(1, a) - \hat{Y}(1, a) \right\} \left\{ \hat{Y}_j(0, a) - \hat{Y}(0, a) \right\}, \]
\[ v_{j22} = \frac{1}{J_a(J_a - 1)} \left\{ \hat{Y}_j(1, a) - \hat{Y}(1, a) \right\}^2. \]

The definition of the cluster-robust HC2 variance implies,

\[ \hat{\text{var}}_{hc2}^{\text{cluster}}((\hat{\alpha}_a^\text{wls}, \hat{\alpha}_{1a}^\text{wls})^\top) = (X^\top WX)^{-1} \left\{ \sum_{j=1}^{J} V_j I(A_j = a) \right\} (X^\top WX)^{-1}, \]

which yields

\[ \hat{\text{var}}_{hc2}^{\text{cluster}}(\hat{\alpha}_{1a}^\text{wls}) = \sum_{j=1}^{J} (4v_{j22} - 2v_{j12} - 2v_{j21} + v_{j11}) I(A_j = a) \]
\[ = \frac{1}{J_a(J_a - 1)} \sum_{j=1}^{J} \left\{ \{ \hat{Y}_j(1, a) - \hat{Y}(1, a) \} - \{ \hat{Y}(1, a) - \hat{Y}(0, a) \} \right\}^2 \]
\[ = \frac{\hat{\sigma}_D^2(a)}{J_a}, \]

\[ \hat{\text{var}}_{hc2}^{\text{cluster}}(\hat{\alpha}_{1a}^\text{wls}) = \sum_{j=1}^{J} (v_{j22} - v_{j12} - v_{j21} + v_{j11}) I(A_j = a) \]
\[ = \frac{1}{J_a(J_a - 1)} \sum_{j=1}^{J} \left\{ \hat{Y}_j(0, a) - \hat{Y}(0, a) \right\}^2 \]
\[ = \frac{\hat{\sigma}_b^2(0, a)}{J_a}, \]

and

\[ \hat{\text{cov}}_{hc2}^{\text{cluster}}(\hat{\alpha}_a^\text{wls}, \hat{\alpha}_{1a}^\text{wls}) = \frac{1}{J^2_a} \sum_{j} (-2v_{j22} + 2v_{j12} + v_{j21} - v_{j11}) I(A_j = a) \]
\[ = \frac{1}{J_a(J_a - 1)} \left\{ \hat{Y}_j(1, a) - \hat{Y}(1, a) \right\} \left\{ \hat{Y}_j(0, a) - \hat{Y}(0, a) \right\} \]
\[ - \frac{1}{J_a(J_a - 1)} \sum_{j=1}^{J} \left\{ \hat{Y}_j(0, a) - \hat{Y}(0, a) \right\}^2. \]

Thus, we obtain,

\[ \hat{\text{var}}_{hc2}^{\text{cluster}}(\hat{\alpha}_a^\text{wls} + \hat{\alpha}_{1a}^\text{wls}) = \hat{\text{var}}_{hc2}^{\text{cluster}}(\hat{\alpha}_a^\text{wls}) + \hat{\text{var}}_{hc2}^{\text{cluster}}(\hat{\alpha}_{1a}^\text{wls}) + 2\hat{\text{cov}}_{hc2}^{\text{cluster}}(\hat{\alpha}_a^\text{wls}, \hat{\alpha}_{1a}^\text{wls}) = \frac{\hat{\sigma}_b^2(1, a)}{J_a}. \]

Next, we calculate the individual-robust HC2 variance. Similarly, using the orthogonality among the covariates, we have

\[ \hat{\text{var}}_{hc2}^{\text{ind}}((\hat{\alpha}_a^\text{wls}, \hat{\alpha}_{1a}^\text{wls})^\top) = (X^\top WX)^{-1} \left\{ \sum_{j=1}^{J} \sum_{i=1}^{n_j} w_{ij}^2 \epsilon_{ij}^2 I(A_j = a)(1 - P_{ij})^{-1} X_{ij} X_{ij}^\top \right\} (X^\top WX)^{-1}, \]

54
For units with \( Z_{ij} = z \), we obtain
\[
P_{ij} = w_{ij}^{a}X_{ij}^{T}\begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}X_{ij} = \frac{1}{n_{jz}},
\]
\[
\hat{\epsilon}_{ij} = Y_{ij} - \hat{Y}_{j}(z, a).
\]

Therefore, we have
\[
\begin{align*}
\sum_{j=1}^{J} \sum_{i=1}^{n_{j}} w_{ij} \hat{\epsilon}_{ij}^{2}I(A_{j} = a)(1 - P_{ij})^{-1}X_{ij}X_{ij}^{T} \\
&= \frac{1}{j^{2}a} \sum_{j=1}^{J} I(A_{j} = a) \sum_{i=1}^{n_{j}} \left \{ \frac{(Y_{ij} - \hat{Y}_{j}(1, a))^{2}I(Z_{ij} = 1)}{n_{j1}(n_{j1} - 1)} + \frac{(Y_{ij} - \hat{Y}_{j}(0, a))^{2}I(Z_{ij} = 0)}{n_{j0}(n_{j0} - 1)} \right \}X_{ij}X_{ij}^{T} \\
&= \frac{1}{j^{2}a} \left( \sum_{j=1}^{J} I(A_{j} = a) \sum_{i=1}^{n_{j}} \frac{\hat{\sigma}_{j}^{2}(1, a)}{n_{j1}} \right) \left( I(A_{j} = a) \sum_{j=1}^{J} \frac{\hat{\sigma}_{j}^{2}(0, a)}{n_{j1}} \right) \\
&= \frac{1}{j^{2}a} \left( \sum_{j=1}^{J} \frac{\hat{\sigma}_{j}^{2}(1, a)}{n_{j1}} \right) \left( \sum_{j=1}^{J} \frac{\hat{\sigma}_{j}^{2}(0, a)}{n_{j1}} \right). \tag{A13}
\end{align*}
\]

As a result,
\[
\hat{\text{var}}_{\text{hc}2}(\hat{\alpha}_{11})_{\text{ws}}^{\text{ind}} = \frac{1}{j^{2}a} \sum_{j=1}^{J} \left( \frac{\hat{\sigma}_{j}^{2}(1, a)}{n_{j1}} + \frac{\hat{\sigma}_{j}^{2}(0, a)}{n_{j1} - n_{j1}} \right) I(A_{j} = a)
\]

and hence,
\[
\hat{\text{var}} \left\{ \hat{\text{DE}}Y(a) \right\} = \left( 1 - \frac{J}{J} \right) \hat{\text{var}}_{\text{hc}2}(\hat{\alpha}_{1a}) + \frac{J}{J} \hat{\text{var}}_{\text{hc}2}(\hat{\alpha}_{1a}).
\]

\[\square\]

### C.4 Relations to Random Effects Models for Split-Plot Designs

Because two-stage experiments have a hierarchical structure, we re-express the linear model as a random effects model. First, suppose that \( Y_{ij}(1, a) - Y_{ij}(0, a) = \alpha_{aij} \), then we can write the potential outcomes as,
\[
Y_{ij}(z, a) = I(z = 1)\alpha_{aj} + \alpha_{0aj} + I(z = 1)r_{aij} + r_{0aij},
\]
where \( \alpha_{0aj} = \bar{Y}_{j}(0, a), \alpha_{aj} = Y_{j}(1, a) - Y_{j}(0, a), r_{0aij} = Y_{ij}(0, a) - \bar{Y}_{j}(0, a) \) and \( r_{aij} = \alpha_{aij} - \bar{\alpha}_{aj} \). Then, the realized outcome can be expressed as,
\[
Y_{ij} = \sum_{a=0,1} \alpha_{0aij}I(A_{j} = a) + \sum_{a=0,1} \alpha_{aij}Z_{ij}I(A_{j} = a) + \sum_{a=0,1} I(A_{j} = a)r_{0aij} + \sum_{a=0,1} Z_{ij}I(A_{j} = a)r_{aij},
\]
\[
= \sum_{a=0,1} \alpha_{a}I(A_{j} = a) + \sum_{a=0,1} \bar{\alpha}_{a}Z_{ij}I(A_{j} = a) + \sum_{a=0,1} I(A_{j} = a)r_{0aij} + \sum_{a=0,1} Z_{ij}I(A_{j} = a)r_{aij}
\]
\[
+ \sum_{a=0,1} s_{0aij}I(A_{j} = a) + \sum_{a=0,1} s_{aij}Z_{ij}I(A_{j} = a), \tag{A13}
\]
where \( \bar{\alpha}_{a} = \bar{Y}(1, a) - \bar{Y}(0, a), \bar{\alpha}_{0a} = \bar{Y}(0, a), s_{aj} = \bar{\alpha}_{aj} - \bar{\alpha}_{a} \) and \( s_{0aj} = \bar{\alpha}_{0aj} - \bar{\alpha}_{0a} \).
Given the similarity between equations [(A13)] and [(11)], we can treat the last four terms of equation [(A13)] as the error term $\epsilon_{ij}$ in equation [(11)], and decompose it into two parts, i.e., $\epsilon_{ij} = \epsilon_{Bj} + \epsilon_{Wij}$, where

\[
\begin{align*}
\epsilon_{Bj} &= \sum_{a=0,1} s_{0aj} I(A_j = a) + \sum_{a=0,1} s_{aj} Z_{ij} I(A_j = a) = \sum_{a=0,1} I(A_j = a)(s_{0aj} + s_{aj} Z_{ij}), \\
\epsilon_{Wij} &= \sum_{a=0,1} I(A_j = a) r_{0aij} + \sum_{a=0,1} Z_{ij} I(A_j = a) r_{aij} = \sum_{a=0,1} I(A_j = a)(r_{0aij} + r_{aij} Z_{ij}),
\end{align*}
\]

where $\epsilon_{Bj}$ can be viewed as the between-cluster residual and $\epsilon_{Wij}$ can be viewed as the within-cluster residual. The cluster-robust HC2 variance in our regression-based approach corresponds to $\epsilon_{Bj}$ and the individual-robust HC2 variance corresponds to $\epsilon_{Wij}$. Because $\sum_{i=1}^{n_j} \epsilon_{Wij} = \sum_{i=1}^{n_j} \epsilon_{Wij} I(Z_{ij} = z) = 0$ holds for all $j$ and $z = 0, 1$, the adjustment for $\hat{\epsilon}_{ij}$ is necessary to ensure the residual from our regression also satisfies this condition. The term $w_{ij}^2 \hat{\epsilon}_{ij}^2 X_{ij} X_{ij}^\top$ in the individual-robust variance corresponds to $r_{aij} + Z_{ij} r_{0aij}$, and the term $X_{ij}^\top W_j \hat{\epsilon}_{ij} \hat{\epsilon}_{ij}^\top W_j X_j$ in the cluster-robust variance corresponds to $s_{0aj} + s_{aj} Z_{ij}$.

### C.5 Proof of Theorem 9

Define

\[
\begin{align*}
\tilde{\zeta}_j^2(z, a) &= \frac{\sum_{i=1}^{n_j} \{D_{ij} - \hat{D}_j(z, a)\}^2 I(Z_{ij} = z)}{n_j - 1}, \\
\tilde{\zeta}_{DE}^2(a) &= \frac{\sum_{j=1}^{J} \{\hat{D}_{E}(a) - \hat{D}(a)\}^2 I(A_j = a)}{J_n - 1}. 
\end{align*}
\]

First, for a given $a$, it is straightforward to show,

\[
(M^\top W M)^{-1} = \begin{pmatrix}
\frac{\hat{D}^2(1,a) + \hat{D}^2(0,a)}{\hat{D}_E^2(a)} & -\frac{\hat{D}(1,a) + \hat{D}(0,a)}{\hat{D}_E^2(a)} \\
-\frac{\hat{D}(1,a) + \hat{D}(0,a)}{\hat{D}_E^2(a)} & 2
\end{pmatrix}.
\]

We then compute the projection matrix,

\[
P_j = W_j^{1/2} M_j (M^\top W M)^{-1} M_j^\top W_j^{1/2}
\]

\[
= \begin{pmatrix}
\frac{1}{\sqrt{J_n n_j} 1_{n_j}} & \\ \frac{1}{\sqrt{J_n n_j} 1_{n_j}} & \\ \frac{1}{\sqrt{J_n n_j} 1_{n_j}} & \\ \frac{1}{\sqrt{J_n n_j} 1_{n_j}} & \\ \\
\end{pmatrix} \cdot \begin{pmatrix}
\hat{D}(1,a) 1_{n_j} \\
\hat{D}(0,a) 1_{n_j}
\end{pmatrix}^\top
\]

\[
\times \begin{pmatrix}
\frac{1}{\sqrt{J_n n_j} 1_{n_j}} & \\ \frac{1}{\sqrt{J_n n_j} 1_{n_j}} & \\ \frac{1}{\sqrt{J_n n_j} 1_{n_j}} & \\ \frac{1}{\sqrt{J_n n_j} 1_{n_j}} & \\ \\
\end{pmatrix} \cdot \begin{pmatrix}
\hat{D}(1,a) 1_{n_j} \\
\hat{D}(0,a) 1_{n_j}
\end{pmatrix}
\]

\[
= \begin{pmatrix}
\frac{1}{\sqrt{J_n n_j} 1_{n_j}} & \\ \frac{1}{\sqrt{J_n n_j} 1_{n_j}} & \\ \frac{1}{\sqrt{J_n n_j} 1_{n_j}} & \\ \frac{1}{\sqrt{J_n n_j} 1_{n_j}} & \\ \\
\end{pmatrix} \cdot \begin{pmatrix}
\hat{D}(1,a) 1_{n_j} \\
\hat{D}(0,a) 1_{n_j}
\end{pmatrix}^\top
\]

\[
= \begin{pmatrix}
\frac{1}{J_n n_j} 1_{n_j} \\
\frac{1}{J_n n_j} 1_{n_j} \\
\frac{1}{J_n n_j} 1_{n_j} \\
\frac{1}{J_n n_j} 1_{n_j} \\
\end{pmatrix}
\]
Similar to the proof of Theorem 1, $\left(1_{n_j}^\top, 0_{n_{j_0}}^\top\right)^\top$ and $\left(0_{n_{j_1}}, 1_{n_{j_0}}^\top\right)^\top$ are two eigenvectors of $I_{n_j} - P_b$ whose eigenvalue is $(J_a - 1)/J_a$. Thus, we have,

\[
(I_{n_j} - P_j)^{-1/2}W_jM_j = \sqrt{\frac{J_a}{J_a - 1}} \left( \frac{1}{J_{n_{j_1}}} 1_{n_{j_1}} \cdot \hat{D}(1, a) \frac{1}{J_{n_{j_0}}} 1_{n_{j_0}} \cdot \hat{D}(0, a) 1_{n_{j_0}} \right).
\]

From the regression model given in equation (13), we obtain the following residuals for observations with $Z_{ij} = z$,

\[
\hat{\eta}_{ij} = Y_{ij} - \hat{Y}(z, a) - \text{CADE}(a)\{D_{ij} - \hat{D}(z, a)\}.
\]

This implies,

\[
\hat{\eta}_{ij}^\top (I_{n_j} - Q_j)^{-1/2}WM_j
\]

\[
= \sqrt{\frac{J_a}{J_a - 1}} (Y_{ij} - \hat{Y}(1, a) - \text{CADE}(a)\{D_{ij} - \hat{D}(1, a)\}, \ldots, Y_{n_{j_1}} - Y(0, a) - \text{CADE}(a)\{D_{ij} - \hat{D}(0, a)\})
\]

\[
\times \left( \frac{1}{J_{n_{j_1}}} 1_{n_{j_1}} \cdot \hat{D}(1, a) 1_{n_{j_1}} \right)
\]

\[
\times \left( \frac{1}{J_{n_{j_0}}} 1_{n_{j_0}} \cdot \hat{D}(0, a) 1_{n_{j_0}} \right)
\]

\[
= \sqrt{\frac{1}{J_a(J_a - 1)}} \left( \hat{Y}_{j}(1, 1) - \hat{Y}(1, a) + \text{CADE}(a)\{\hat{D}_{j}(1, a) - \hat{D}(1, a)\}ight)
\]

\[
+ \hat{Y}_{j}(0, a) - \hat{Y}(0, a) + \text{CADE}(a)\{\hat{D}_{j}(0, a) - \hat{D}(0, a)\}
\]

\[\times \left( \hat{D}(1, a) \{\hat{Y}_{j}(1, a) - \hat{Y}(1, a)\} + \hat{D}(1, a) \text{CADE}(a)\{\hat{D}_{j}(1, a) - \hat{D}(1, a)\}ight)
\]

\[+ \hat{D}(0, a) \{\hat{Y}_{j}(0, a) - \hat{Y}(0, a)\} + \hat{D}(0, a) \text{CADE}(a)\{\hat{D}_{j}(0, a) - \hat{D}(0, a)\}\right)^\top
\]

\[= \sqrt{\frac{1}{J_a(J_a - 1)}} \left( \hat{Y}_{j}(1, 1) - \hat{Y}(1, a) + \text{CADE}(a)\{\hat{D}_{j}(1, a) - \hat{D}(1, a)\}ight)
\]

\[+ \hat{Y}_{j}(0, a) - \hat{Y}(0, a) + \text{CADE}(a)\{\hat{D}_{j}(0, a) - \hat{D}(0, a)\}
\]

\[\times \left( \hat{D}(1, a) \{\hat{Y}_{j}(1, a) - \hat{Y}(1, a)\} + \hat{D}(1, a) \text{CADE}(a)\{\hat{D}_{j}(1, a) - \hat{D}(1, a)\}ight)
\]

\[+ \hat{D}(0, a) \{\hat{Y}_{j}(0, a) - \hat{Y}(0, a)\} + \hat{D}(0, a) \text{CADE}(a)\{\hat{D}_{j}(0, a) - \hat{D}(0, a)\}\right)^\top
\]

Let $U_j = M_j^\top W_j (I_{n_j} - P_j)^{-1/2} \eta_{ij}^\top (I_{n_j} - P_j)^{-1/2}W_jM_j$ and $u_{j_1k_2}$ be the $k_1k_2$-th entry of $U_j$,

\[
u_{j_11} = \frac{1}{J_a(J_a - 1)} \left[ \hat{Y}_{j}(1, 1) - \hat{Y}(1, a) + \text{CADE}(a)\{\hat{D}_{j}(1, a) - \hat{D}(1, a)\}ight.\]

\[+ \hat{Y}_{j}(0, a) - \hat{Y}(0, a) + \text{CADE}(a)\{\hat{D}_{j}(0, a) - \hat{D}(0, a)\}\right]^2,
\]

\[
u_{j_12} = \frac{1}{J_a(J_a - 1)} \left[ \hat{Y}_{j}(1, a) - \hat{Y}(1, a) + \text{CADE}(a)\{\hat{D}_{j}(1, a) - \hat{D}(1, a)\}ight.\]

\[+ \hat{Y}_{j}(0, a) - \hat{Y}(0, a) + \text{CADE}(a)\{\hat{D}_{j}(0, a) - \hat{D}(0, a)\}\]

\[\times \left( \hat{D}(1, a) \{\hat{Y}_{j}(1, a) - \hat{Y}(1, a)\} + \hat{D}(1, a) \text{CADE}(a)\{\hat{D}_{j}(1, a) - \hat{D}(1, a)\}ight)
\]

\[+ \hat{D}(0, a) \{\hat{Y}_{j}(0, a) - \hat{Y}(0, a)\} + \hat{D}(0, a) \text{CADE}(a)\{\hat{D}_{j}(0, a) - \hat{D}(0, a)\}\right],
\]

\[
u_{j_22} = \frac{1}{J_a(J_a - 1)} \left[ \hat{D}(1, a) \{\hat{Y}_{j}(1, a) - \hat{Y}(1, a)\} + \hat{D}(1, a) \text{CADE}(a)\{\hat{D}_{j}(1, a) - \hat{D}(1, a)\}ight.
\]

\[+ \hat{D}(0, a) \{\hat{Y}_{j}(0, a) - \hat{Y}(0, a)\} + \hat{D}(0, a) \text{CADE}(a)\{\hat{D}_{j}(0, a) - \hat{D}(0, a)\}\right]^2.
\]

From the definition of the cluster-robust HC2 variance, we have,

\[
\var_{\text{hc2}} \left( (\hat{\beta}_a^{n_{2ds}}, \hat{\beta}_{1a}^{n_{2ds}})^\top \right) = (M^\top W M)^{-1} \left\{ \sum_{j=1}^J U_j I(A_j = a) \right\} (M^\top W M)^{-1}.
\]
We can then obtain that
\[
\hat{\text{var}}_{\text{hc2}}(\hat{\beta}_{1a}^{2\text{sls}}) = \frac{1}{\text{DED}(a)} \sum_j \left[ 4u_{j22} - 4\{\hat{D}(1,a) + \hat{D}(0,a)\}u_{j12} + \{\hat{D}(1,a) + \hat{D}(0,a)\}^2u_{j11} \right] I(A_j = a)
\]
\[
= \frac{1}{\text{DED}^2(a)} \left\{ \frac{\hat{\sigma}^2_{DE}(a)}{J_a} - 2\text{CADE}(1) \frac{\hat{\theta}_{DE}(a)}{J_a} + \text{CADE}^2(1) \frac{\hat{\theta}_{DE}(a)}{J_a} \right\}.
\]

We then calculate the individual-robust generalization of HC2 variance,
\[
\hat{\text{var}}_{\text{hc2}}^{\text{ind}}(\hat{\beta}_{1a}^{2\text{sls}}) = (M^T WM)^{-1} \left\{ \sum_{j=0}^{n_j} \sum_{i=1}^{n_j} w_{ij}^2 \hat{\eta}_{ij}^2 (1 - Q_{ij})^{-1} M_{ij} M_{ij}^T I(A_j = a) \right\} (M^T WM)^{-1},
\]
where \(Q_{ij}\) is the individual leverage and \(\hat{\eta}_{ij}^*\) is the adjusted residual to have \(X_j^T \hat{\eta}_{ij}^* = 0_n\),
\[
\hat{\eta}_{ij}^* = Y_{ij} - \hat{Y}_j(z,a) - \text{CADE}(a) \{ D_{ij} - \hat{D}_j(z,a) \}, \text{ if } Z_{ij} = z.
\]
For units with \(Z_{ij} = z\), we have,
\[
Q_{ij} = \frac{w_{ij} J_a}{\text{DED}^2(a)} M_{ij}^T \left( \begin{array}{cc} \hat{D}^2(1) + \hat{D}^2(0) & -\hat{D}(1) - \hat{D}(0) \\ -\hat{D}(1) - \hat{D}(0) & 2 \end{array} \right) M_{ij} = \frac{1}{n_{jz}}.
\]
Thus, we have
\[
\sum_{j=1}^{J} \sum_{i=1}^{n_j} w_{ij}^2 \hat{\eta}_{ij}^2 I(A_j = a)(1 - Q_{ij})^{-1} M_{ij} M_{ij}^T
\]
\[
= \frac{1}{J_a^2} \sum_{j=1}^{J} I(A_j = a) \sum_{i=1}^{n_j} M_{ij} \left\{ \frac{[Y_{ij} - \hat{Y}_j(1,a) - \text{CADE}(a) \{ D_{ij} - \hat{D}_j(1,a) \}]^2 I(Z_{ij} = 1)}{n_{j1} (n_{j1} - 1)} + \frac{[Y_{ij} - \hat{Y}_j(0,a) - \text{CADE}(a) \{ D_{ij} - \hat{D}_j(0,a) \}]^2 I(Z_{ij} = 0)}{n_{j0} (n_{j0} - 1)} \right\} M_{ij}^T
\]
\[
= \frac{1}{J_a^2} \left( S_1 + S_0 \begin{array}{c} \hat{D}(1,a) S_1 + \hat{D}(0,a) S_0 \\ \hat{D}^2(1,a) S_1 + \hat{D}^2(0,a) S_0 \end{array} \right),
\]
where
\[
S_1 = \sum_{j=1}^{J} \left\{ \frac{\hat{\sigma}^2_j(1,a) + \text{CADE}(a) \hat{\theta}_j^2(1,a) + \text{CADE}^2(a) \hat{\theta}_j^2(1,a)}{n_{j1}} \right\} I(A_j = a),
\]
\[
S_0 = \sum_{j=1}^{J} \left\{ \frac{\hat{\sigma}^2_j(0,a) + \text{CADE}(a) \hat{\theta}_j^2(0,a) + \text{CADE}^2(a) \hat{\theta}_j^2(0,a)}{n_{j} - n_{j1}} \right\} I(A_j = a).
\]
Putting all together,
\[
\hat{\text{var}}_{\text{hc2}}^{\text{ind}}(\hat{\beta}_{1a}^{w2\text{sls}})
\]
\[
\frac{1}{J_a \hat{D}_{\hat{E}}D(a)} \sum_{j=1}^{J} \left[ 4 \hat{D}^2(1, a)S_1 + \hat{D}^2(0, a)S_0 - 4\{ \hat{D}(1, a) + \hat{D}(0, a) \} \{ \hat{D}(1, a)S_1 + \hat{D}(0, a)S_0 \} \\
+ \{ \hat{D}(1, a) + \hat{D}(0, a) \}^2 \{ S_1 + S_0 \} \right] I(A_j = a)
\]

\[
= \frac{1}{J_a \hat{D}_{\hat{E}}D(a)} \frac{S_1 + S_0}{J_a^2}.
\]

Therefore, we have,
\[
\tilde{\text{var}} \left\{ \text{CADE}(a) \right\} = \left( 1 - \frac{J_a}{J} \right) \tilde{\text{var}}_{\text{hc2}} (\hat{\beta}_{wa}^{w2sls}) + \frac{J_a}{J} \tilde{\text{var}}_{\text{ind}} (\hat{\beta}_{wa}^{w2sls}).
\]

D Simulation Studies

We examine the performance of different variance estimators for the CADE. In particular, we compare our variance estimator with three other commonly used variance estimators: HC2 variance (MacKinnon and White, 1985), cluster-robust variance (Liang and Zeger, 1986), and cluster-robust HC2 variance (Bell and McCaffrey, 2002). Following the notation used in Section 3.5.2, the HC2 variance is defined as,
\[
(M^\top W M)^{-1} \left\{ \sum_{j=1}^{J} \sum_{i=1}^{n_j} w_{ij}^2 \hat{\eta}_{ij}^2 (1 - Q_{ij})^{-1} M_{ij} M_{ij}^\top \right\} (M^\top W M)^{-1}
\]

whereas the cluster-robust variance is defined as,
\[
(M^\top W M)^{-1} \left\{ \sum_{j=1}^{J} M_j^\top W_j \hat{\eta}_j \hat{\eta}_j^\top W_j M_j \right\} (M^\top W M)^{-1}
\]

Finally, the cluster-robust HC2 variance is defined in equation (14).

Below, we consider three scenarios: no spillover effect of the treatment receipt on the outcome, no spillover effect of the treatment assignment on the treatment receipt, and both spillover effects present. In each scenario, we choose equal size \( n \) in all \( J \) clusters, and generate the data with three different settings regarding the number of clusters and cluster sizes while holding the total number of units \( N = nJ \) constant: \( (n = 10, J = 250) \), \( (n = 250, J = 10) \), and \( (n = 50, J = 50) \). We find that our proposed variance estimators perform well so long as the number of clusters is large. The HC2 variance estimator tends to underestimate the true variance while the cluster-robust HC2 variance estimator tends to overestimate it.

D.1 No Spillover Effect of Treatment Receipt on the Outcome

We first conduct a simulation study under the assumption of no spillover effect of treatment receipt on the outcome. In this scenario, Assumptions 1–6 are satisfied. We begin by defining the complier status variable \( C_{ij}^*(a) \) for a given treatment assignment mechanism \( a = 0, 1 \) as,

\[
C_{ij}^*(a) = \begin{cases} 
0 & \text{if } D_{ij}(1, a) = D_{ij}(0, a) = 0, \\
1 & \text{if } D_{ij}(1, a) = 1, D_{ij}(0, a) = 0, \\
2 & \text{if } D_{ij}(1, a) = D_{ij}(0, a) = 1.
\end{cases}
\]
Figure A1: Coverage rates of 95% confidence intervals when there is no spillover effect of the treatment receipt on the outcome. The confidence intervals based on the proposed variance estimator (solid circle with black line) are compared with those based on the HC2 variance (open circle with dotted line), the cluster-robust variance (grey solid triangle with dotted line), and the cluster-robust HC2 variance (grey open triangle with solid line). The cluster size is indicated by $n$ whereas $J$ is the number of clusters. The horizontal axis represents the intracluster correlation coefficient.

where 0, 1, and 2 represent never-taker, complier, and always-taker, respectively. We sample $C_{ij}^a$ from a categorical distribution with probabilities $(0.1, 0.6, 0.3)$ for $a = 1$ and $(0.3, 0.6, 0.1)$ for $a = 0$. We obtain the potential values $D_{ij}(z, a)$ from the realized value of $C_{ij}^a$ according to equation (A14).

In the absence of the spillover effect of the treatment receipt on the outcome, the potential values of the outcome only depend on one’s treatment receipt. Therefore, we first generate $Y_{ij}(D_{ij} = 0)$ via $Y_{ij}(D_{ij} = 0) \overset{i.i.d.}{\sim} N(0, 1)$ and then generate the $Y_{ij}(D_{ij} = 1)$ as,

$$\theta_j \overset{i.i.d.}{\sim} N(\theta, \sigma_b^2), \quad Y_{ij}(D_{ij} = 1) - Y_{ij}(D_{ij} = 0) \overset{i.i.d.}{\sim} N(\theta_j, \sigma_w^2)$$

where $\sigma_b^2$ represents the between-cluster variance and $\sigma_w^2$ is the within-cluster variance. We generate the treatment assignment mechanism $A_j$ with $\Pr(A_j = a) = 1/2$ for $a = 0, 1$ such that $J_1 = J_0 = J/2$. We then completely randomize the treatment assignment $Z_{ij}$ so that 60% (40%) of the units assigned to treatment if $A_j = 1$ ($A_j = 0$). For population parameters, we fix the average cluster specific effect $\theta = 1$ as well as the total variance $\sigma_b^2 + \sigma_w^2 = 1$. We use four different levels of intraclass correlation, i.e., $(0, 0.2, 0.4, 0.6)$, which is defined as $\rho = \sigma_b^2 / (\sigma_b^2 + \sigma_w^2)$.

Figure [A1] shows the coverage rates of the confidence intervals for the CADEs calculated by averaging over 1,000 Monte Carlo simulations (the top and bottom rows present...
the coverage rates for the $\text{CADE}(1)$ and $\text{CADE}(0)$, respectively). When the number of clusters is relatively large (i.e., $(n = 10, J = 250)$ and $(n = 50, J = 50)$), our variance estimator leads to the coverage rates closest to the nominal rate of 95%. However, when the number of cluster is small but the cluster size is large $(n = 250, J = 10)$, all the four variance estimators have a tendency to undercover the true value especially when the intra-cluster correlation is large. Across all scenarios, the confidence intervals based on the HC2 variance tend to perform poorly.

**D.2 No Spillover Effect of Treatment Assignment on the Treatment Receipt**

We next consider the setting with no spillover effect of treatment assignment on the treatment receipt. In this scenario, Assumptions 1–5 are satisfied and Assumption 6 is violated. Since the potential values of treatment receipt depend only on one’s own treatment assignment, the complier status does not depend on the treatment assignment mechanism, i.e., $C_{ij}^*(1) = C_{ij}^*(0) = C_{ij}^*$.

We sample $C_{ij}^*$ from a categorical distribution with probabilities $(0.2, 0.6, 0.2)$, and compute the potential values of $Y_{ij}$ as,

$$Y_{ij}(D_{ij} = 0, D_{-i,j} = d_{-i,j}) \overset{\text{indep.}}{\sim} N\left(\frac{\beta}{n} \sum_{i=1}^{n} d_{ij}, 1\right), \quad \theta_j \overset{\text{i.i.d.}}{\sim} N(\theta, \sigma^2_b),$$

$$Y_{ij}(D_{ij} = 1, D_{-i,j} = d_{-i,j}) - Y_{ij}(D_{ij} = 0, D_{-i,j} = d_{-i,j}) \overset{\text{indep.}}{\sim} N(\theta_j, \sigma^2_w).$$
Thus, the potential values of the outcome depend on the number of treated units in the cluster. We then generate the treatment assignment and its mechanism in the same way as done in Section D.1 while setting $\theta = 1$ and $\beta = 1$. Figure A2 shows the coverage rates of the confidence intervals for the CADEs in the same way as in Figure A1. The results are similar to those under the setting with no spillover effect of the treatment receipt on the outcome shown in Figure A1.

### D.3 Presence of Two Spillover Effects

Finally, we consider the setting with both types of spillover effects by combining the data generating mechanisms used above. We keep the data generating mechanism of the potential values of the treatment receipt introduced in Section D.1 while generating the outcome according to the data generating mechanism of Section D.2. Thus, this setting permits the presence of two spillover effects: the spillover effect of treatment receipt on the outcome and the spillover effect of treatment assignment on the treatment receipt. In this scenario, Assumptions 1–4 are satisfied and Assumptions 5 and 6 are violated. Figure A3 shows the coverage of the 95% confidence intervals for the CADEs. Again, the results are quite similar to those obtained under the other two scenarios.

### D.4 Zero-Inflated Outcome

Because the outcome variable (annual hospital expenditure) in our application data has many zeros (20.7%) and is skewed, we consider a simulation study with a zero-inflated
outcome variable. In particular, we maintain the core data generating mechanisms used in Sections D.1–D.3 and replace the distribution of the outcome variable with a mixture of the point mass at zero and a truncated normal distribution. For example, for the simulation setup used in Section D.1, we generate the outcome variable as,

$$U_i \sim \text{Bernoulli}(p),$$

$$Y_{ij}(D_{ij} = 1) - Y_{ij}(D_{ij} = 0) \sim (1 - U) \cdot \text{TruncN}(\theta_j, \sigma_w^2, (0, +\infty)), $$

where $\text{TruncN}(\theta_j, \sigma_w^2, (0, +\infty))$ is a normal distribution with mean $\theta_j$ and variance $\sigma_w^2$ truncated from below at zero. We choose $p = 0.2$ to emulate our application data.

Figure A4 shows the coverage rates of the confidence intervals for the CADEs when the treatment receipt has no spillover effect on the outcome. The top and bottom rows present the coverage rates for the CADE(1) and CADE(0), respectively. The result is similar to that of the simulation in Section D.1. When the number of clusters is relatively large (i.e., $(n = 10, J = 250)$ and $(n = 50, J = 50)$), our variance estimator leads to the coverage rates closest to the nominal rate of 95%. However, when the number of cluster is small but the cluster size is large $(n = 250, J = 10)$, all the four variance estimators have a tendency to undercover the true value especially when the intracluster correlation is large.

Figure A5 presents the coverage of the 95% confidence intervals for the CADEs when there is no spillover effect of treatment assignment on the treatment receipt. Figure A6 shows the coverage of the 95% confidence intervals for the CADEs in the presence of
two spillover effects. The results are similar to those with the corresponding cases of the normally distributed outcomes (see Sections D.2 and D.3).

**E Model-Based Analysis**

In the main text, we focus on the nonparametric identification of the ITT causal effects and complier causal effects and establish the connection between the randomization-based estimators and the regression-based estimators. Although the model-free identification results are appealing, they are limited by the experimental design and may not directly estimate the causal quantities that are of interest to applied researchers and policy makers. In this section, we consider a model-based analysis to overcome the limitations of our nonparametric approach. In the following, we assume linear models for the sake of illustration but other modeling assumptions are possible. We note that the model-based analysis is based on the super population framework, which is different from the finite population framework used for our nonparametric analysis.

**E.1 Intention-to-Treat Analysis**

We first consider the model-based ITT analysis. We model the potential outcome as a linear function of one’s own encouragement, the proportion of encouraged households within the
same village, and their interaction.

\[
Y_{ij}(Z_j = z_j) = \alpha_0 + \alpha_1 z_{ij} + \alpha_2 \cdot \frac{\sum_{j=1}^{n_j} z_{ij}}{n_j} + \alpha_3 z_{ij} \cdot \frac{\sum_{j=1}^{n_j} z_{ij}}{n_j} + \iota_{ij}, \quad (A15)
\]

where \(\iota_{ij}\) is the error term.

This model is applicable under two-stage randomized experiments with more than two treatment assignment mechanisms. In addition, the model can be used to extrapolate the average direct and spillover effects under different treatment assignment mechanisms. For example, under the scenario that all other households within the same cluster are assigned to the treatment condition, the average direct effect is given by

\[
E\{Y_{ij}(Z_{ij} = 1, Z_{-i,j} = 1) - Y_{ij}(Z_{ij} = 0, Z_{-i,j} = 1)\} \approx \alpha_1 + \alpha_3.
\]

where the approximation results from an additional term \(\alpha_2/n_j\), which is negligible so long as \(n_j\) is large. If, on the other hand, all other households within the same cluster are assigned to the treatment condition, the average direct effect is equal to,

\[
E\{Y_{ij}(Z_{ij} = 1, Z_{-i,j} = 0) - Y_{ij}(Z_{ij} = 0, Z_{-i,j} = 0)\} \approx \alpha_1.
\]

Similarly, the average spillover effect of assigning all other households in the same cluster to the treatment condition (versus no household assigned to the treatment condition) depends on one’s own encouragement status and is given by,

\[
E\{Y_{ij}(Z_{ij} = 1, Z_{-i,j} = 1) - Y_{ij}(Z_{ij} = 1, Z_{-i,j} = 0)\} \approx \alpha_2 + \alpha_3.
\]
Table A1: Estimated average direct and spillover effects under Model (A15). The first (second) column presents the estimated average direct effect when all other households within the same cluster are assigned to the treatment (control) condition. The third (fourth) column shows the estimated average spillover effect of all other households within the same cluster are assigned to the treatment condition versus no households are assigned to the treatment condition when the household itself is assigned to the treatment (control) condition. The confidence intervals are based on cluster-robust HC2 standard errors.

<table>
<thead>
<tr>
<th>Direct effects</th>
<th>Spillover effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_1 + \alpha_3$</td>
<td>$\alpha_1$</td>
</tr>
<tr>
<td>$\alpha_2 + \alpha_3$</td>
<td>$\alpha_2$</td>
</tr>
<tr>
<td>$-1253$</td>
<td>$1477$</td>
</tr>
<tr>
<td>$-2881$</td>
<td>$-180$</td>
</tr>
<tr>
<td>$(-2646, 139)$</td>
<td>$(-94, 2988)$</td>
</tr>
<tr>
<td>$(-4631, -1131)$</td>
<td>$(-1991, 1630)$</td>
</tr>
</tbody>
</table>

We apply this model-based approach to our application data. Table A1 shows the results. Although not statistically significant, the estimated average direct effect when all other households are assigned to the treatment condition ($\alpha_1 + \alpha_3$) is negative whereas the estimated average direct effect when all other households are assigned to the control condition ($\alpha_1$) is positive. The spillover effects are generally negative especially when a household is encouraged to enroll in the RSBY. These results are similar to those based on the randomization-based approach.

E.2 Complier Average Direct Effect of Treatment Receipt

Next, we consider a model-based approach to the estimation of the complier average direct effect (CADE). In the standard settings without interference between units, the complier average causal effect represents the average causal effect of one’s own treatment receipt on the outcome among compliers (Angrist et al., 1996). This interpretation is still applicable to our settings if there exists no spillover effect of encouragement on treatment receipt or no spillover effect of treatment receipt on outcome (i.e., Scenarios I and II of Figure 1). However, when both types of spillover effects exist (i.e., Scenario III of the figure), the CADE represents the indirect effect of one’s own encouragement on the outcome through the treatment receipt of other units in the same village as well as the direct effect of one’s own treatment receipt on the outcome. Unfortunately, without an additional assumption, we cannot identify the latter in the presence of these two spillover effects.

Here, we address this issue by assuming the following parametric structure for the spillover effects.

\[
Y_{ij}(D_j = d_j) = \beta_0 + \beta_1 d_{ij} + \beta_2 \cdot \frac{\sum_{j=1}^{n_j} d_{ij}}{n_j} + \beta_3 d_{ij} \cdot \frac{\sum_{j=1}^{n_j} d_{ij}}{n_j} + U_{ij},
\]  

(A16) where $U_{ij}$ represents the latent confounders between the treatment receipt and the outcome. Model (A16) posits the potential outcome as a linear function of one’s own treatment receipt and the proportion of households in the same cluster who received the treatment.

Under this model, the average direct effect of one’s own treatment receipt on the outcome when all the other households within the same cluster receive the treatment is given
Table A2: Estimated direct and spillover effects of the treatment receipt under Model (A16). The first (second) column presents the average direct effect of one’s own treatment receipt when all other households within the same cluster receive the treatment (control) condition. The third (fourth) column shows the average spillover effect of all other households within the same cluster receive the treatment condition versus no households receive the treatment condition when the household itself receives the treatment (control) condition. The confidence intervals are based on cluster-robust HC2 standard errors.

<table>
<thead>
<tr>
<th>Direct effects</th>
<th>Spillover effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1 + \beta_3$</td>
<td>$\beta_1$</td>
</tr>
<tr>
<td>$\beta_2 + \beta_3$</td>
<td>$\beta_2$</td>
</tr>
<tr>
<td>$(-6013, 8724)$</td>
<td>$(-11715, 3022)$</td>
</tr>
<tr>
<td>$(-11872, -154)$</td>
<td>$(407, 17041)$</td>
</tr>
<tr>
<td>$(-19445, -3985)$</td>
<td>$(-4927, 10970)$</td>
</tr>
</tbody>
</table>

Theorem A5 Suppose Model (A16) is correctly specified and Assumption 1 holds. Then, the coefficients of the model are identified by solving the following estimating equations

$$\sum_{i=1}^{J} \sum_{j=1}^{n_j} \{ Y_{ij} - (\beta_0 + \beta_1 D_{ij} + \beta_2 \bar{D}_j + \beta_3 D_{ij} \bar{D}_j) \} \mathbf{H}_{ij} = 0,$$

where $\mathbf{H}_{ij} = (1, Z_{ij}, A_j, Z_{ij} A_j)^\top$.

We fit this model to our application data. Table A2 shows the results. The estimated average direct effect of enrollment in the RSBY receipt when all other households are also enrolled $(\beta_1 + \beta_3)$ is negative while the estimated average direct effect when all other households are not enrolled $(\beta_1)$ is positive. This finding is similar to the one obtained under our nonparametric approach.