A test for direct and indirect effect interaction under treatment interference

by

Alex Becker

B.S. Kansas State University 2023

A REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2024

Approved by:

Major Professor
Mike Higgins

Copyright

Abstract

Treatment interference occurs when the treatment effect of one unit affects the response of another unit. Under interference, treatment effects can be classified as *direct*—those attributable to treatment given to the unit itself—or *indirect*—those that are due to the treatment status of "neighboring" units. However, definitions for many important quantities of interest can be ambiguous and hard to interpret unless no interaction between direct and indirect effects holds. Intuitively, this assumption a the direct effect of treatment for a unit will be the same regardless of what treatment allocation is given to that unit's neighbors. Such an assumption may also help improve precision of direct and indirect effect estimates by allowing additional units to be used in the estimation of these effects. In this report, we develop a test for interaction between direct and indirect effects. Using sampling theory, we argue for approximate normality of our test statistic under a null hypothesis of no treatment interaction between direct and indirect effects. We then perform a simulation study to verify nominal Type I errors for our test and to assess how differences in sample size, interference model, and magnitudes of the treatment effect interactions impact the power of the test.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Causal inference is a branch of Statistics that involves identifying, estimating, and performing inferences on cause-and-effect relationships between variables. In particular, causal inference aims to determine how a treatment affects a response. While this may seem very straightforward, performing causal inference has several challenges. In particular, it is often difficult to separate the effect of treatment from the effect of other confounding variables—those variables that may also affect the response of a unit. For example, in a study on the effect of smoking on the likelihood of developing heart disease, potential confounding variables may include diet and exercise habits.

Confounding variables can be particularly difficult to identify, let alone quantify. Randomized controlled experiments are considered the "gold standard" for causal inference because they allow us to remove the effect of confounding variables because we can control the ground truth of no variables that we can't quantify.

Imbens and Rubin (2015); Rubin (1980) discuss the value of randomized controlled experiments, namely that differences between treatment and control groups can be chalked up as the treatment effect and not pre-existing differences between the two groups.

However, experiments cannot address the fundamental problem of causal inference—that it is impossible to observe both the treatment value and the control value on the same unit. For example, if an individual is given a certain medicine (treatment) that cured their sickness, they can not also be given a control. It would be wrong to directly conclude that the medicine is exactly what cured their sickness due to the possibility that they could have also recovered if they were not given the medicine. Or in other words, even if a treatment appears to have affected the response on a unit, it is impossible to know whether that change of response would have occurred had the unit received control.

Another issue that arises with experiments is that, under many settings, the response of one unit can be affected by the treatment assigned to a different unit, so the interactions and influence between units should be considered and closely studied. This phenomenon is often referred to as *treatment interference*, *treatment spillover*, *network effects*, or *peer effects* (Sobel, 2006; Rosenbaum, 2007; Hudgens and Halloran, 2008).

A considerable amount of recent work has been devoted towards analyzing experiments exhibiting interference.

When interference is present, treatment effects can be classified as either *direct*—changes in response in a unit due to the treatment given to that unit—or *indirect*—changes in a unit's response due to the treatment given to "neighbors" of that unit. However, defining causal estimands may not be straight forward. For example, the direct effect of a unit may depend on what neighboring units are given treatment. Some approaches for overcoming this ambiguity include defining unique direct effects for each possible exposure given to neighbors, and to quantities of interest as a weighted average of direct effects across all possible exposures (Hudgens and Halloran, 2008; Sävje et al., 2021).

However, under a certain assumption—specifically, *no weak interaction between direct and indirect effects* (also known as *weak additivity of main effects*)—these issues become moot. Intuitively, this assumption requires that the average direct effect will be unchanged regardless of the treatment allocation given to units' neighbors. Additionally, not only can

this assumption reduce ambiguity in the definition of causal quantities, but it may also lead to increased precision of estimates of these quantities; the assumption holding may allow additional units to be used in estimators of these quantities.

In this report, we develop a test of the no weak interaction assumption. We argue asymptotic normality of our test statistic under a null hypothesis of no weak interaction and verify our claim via simulation. We then perform a simulation study to assess the power of our test.

1.2 Potential Outcomes Model for Causal Inference

Suppose we have an experiment with n units, numbered 1 through n. Let Y_i denote the response of unit i and let W_i denote a treatment indicator: $W_i = 1$ if unit i is given treatment; otherwise, $W_i = 0$. A common model for causal inference is the Neyman-Rubin potential outcomes model of response: (Rubin, 1974; Holland, 1986; Splawa-Neyman et al., 1990). This model assumes the following model of response:

$$Y_i = y_{i1}W_i + y_{i0}(1 - W_i). \tag{1.1}$$

The values y_{i1} and y_{i0} are *potential outcomes*—the hypothetical outcomes of response for unit i given treatment and control respectively.

Rubin (1974) is credited as viewing causal inference as a missing data problem. The *fundamental problem of causal inference* is that only one potential outcome can be observed for any given unit—all other outcomes are missing. The observed potential outcome becomes the response, and the unobserved potential outcomes are called *counterfactuals*.

Inherent in this model is the stable unit treatment value assumption (SUTVA) (Rubin, 1980). This assumption states that there is only one version of each treatment, and the treatment status of one unit does not impact the response of any other unit. Note, SUTVA does not hold when treatment interference is present.

1.3 Treatment interference

Treatment interference occurs if the response of a unit is affected by the treatment status of another unit. A good example of this idea is the dependent happenings of infectious diseases and subsequent vaccination. For example, if a subject is not vaccinated for some disease, they may be more likely to transmit the disease to another subject and thus have an effect on their response (i.e. getting sick as a result of no treatment from another unit).

While an experiment that exhibits interference violates SUTVA, we can still use the potential outcomes framework to define response. Let $y_i(W_i, \mathbf{W}_{-i})$ denote the potential outcome of unit i had unit i received the treatment condition W_i and the "neighbors" of i receive treatment condition $\mathbf{W}_{-i}$—where we define a *neighbor* of i as any unit j such that the response of i may be affected by the treatment status of j. While potential outcomes could be different for every one of the 2^n possible treatment allocations, in practice, the set of neighbors for a unit i tends to be small compared to the number of units. Note, to make this concept completely rigorous, we require the introduction of exposure mappings; we rely on more of an intuitive definition instead (Aronow, 2012).

We can then define the response of unit i as

$$Y_i = \sum_{(W_i, \mathbf{W}_{-i})} y_i(W_i, \mathbf{W}_{-i}) \mathbf{1}(W_i, \mathbf{W}_{-i}) \tag{1.2}$$

where $\mathbf{1}(W_i, \mathbf{W}_{-i})$ is a treatment indicator for the exposure $(W_i, \mathbf{W}_{-i})$—that is, it equals 1 if the exposure $(W_i, \mathbf{W}_{-i})$ is observed and is 0 otherwise—and this sum is taken across all possible treatment exposures given to a unit and its neighbors.

The *direct effect* for a unit i is the difference in potential outcomes when the unit is given treatment and when the unit is given control. It may be different given different neighborhood exposure $\mathbf{W}_{-i}$. Thus, it may be defined as

$$\tau_i(\mathbf{W}_{-i}) = y_i(1, \mathbf{W}_{-i}) - y_i(0, \mathbf{W}_{-i}). \tag{1.3}$$

Due to the dependence on the exposures given to neighbors of i, there is no consensus on how to defining average direct effects—the average effect of treatment when treatment is given to that unit. Some approaches consider defining average direct effects for each exposure—which, while interpretable, creates many different competing forms of the direct effects (Hudgens and Halloran, 2008). Others consider taking a weighted average of direct effects across all possible exposures—which will lead to a single estimable quantity, but the interpretation of what that quantity means may not be straight forward (Sävje et al., 2021).

However, as we will see, under stronger assumptions, these concerns about how to define average direct effects can be mitigated.

1.3.1 No interaction between direct and indirect effects

While strong, an assumption of *no strong interaction between direct and indirect effects* can be quite useful for simplifying definitions of average direct and indirect effects. There is no strong interaction between direct and indirect effects if and only if potential outcomes can be decomposed into direct and indirect effects. That is, there is no strong interaction between direct and indirect effects if we can write potential outcomes of a unit i as:

$$y_i(W_i, \mathbf{W_{-i}}) = \alpha_i + \beta_i W_i + \sum_{\mathbf{W}_{-i}} \beta_{i,\mathbf{W}_{-i}} \mathbf{1}(\mathbf{W}_{-i}). \tag{1.4}$$

where the sum is taken over all possible treatment exposures given to i's neighbors, and $1(\mathbf{W}_{-i})$ is an indicator variable denoting that exposure $\mathbf{W}_{-i}$ is observed. This assumption is also known as additivity of main effects (Sussman and Airoldi, 2017).

Note that, if there is no strong interaction, then for any two exposures $\mathbf{W}_{-i}, \mathbf{W}'_{-i}$ given to neighbors, we have

$$\tau_i(\mathbf{W}_{-i}) = \tau_i(\mathbf{W}'_{-i}) = \beta_i. \tag{1.5}$$

That is, for each unit, the unit-level direct effect is always the same regardless of the exposure given to i's neighbors, thereby mitigating the issues of how to define direct effects.

However, if we are only interested in ensuring well-defined average direct effects, it suffices to consider a weaker version of this assumption. There is *no weak interaction between direct and indirect effects* if *averages* of potential outcomes can be expressed as

$$\bar{y}(W_i, \mathbf{W}_{-i}) = \alpha + \beta W_i + \sum_{\mathbf{W}_{-i}} \beta_{\mathbf{W}_{-i}} \mathbf{1}(\mathbf{W}_{-i}). \tag{1.6}$$

Note, we have a similar result about average direct effects as we did before with unit-level direct effects. Define

$$\tau(\mathbf{W}_{-i}) = \bar{y}(1, \mathbf{W}_{-i}) - \bar{y}(0, \mathbf{W}_{-i}) \tag{1.7}$$

as the average direct effect when units have a neighborhood exposure of $\mathbf{W}_{-i}$. Note, under no weak interaction between direct and indirect effects, we have for any two neighborhood exposures $\mathbf{W}_{-i}, \mathbf{W}'_{-i}$, we have

$$\tau(\mathbf{W}_{-i}) = \tau(\mathbf{W}'_{-i}). \tag{1.8}$$

This also implies that estimates of $\tau(\mathbf{W}_{-i})$—that would typically need to use only responses that have a neighborhood exposure of $\mathbf{W}_{-i}$—now provide an estimate of the average direct effect. Hence, we may be able to get a precise estimate of the direct effect by taking a weighted average of these estimates.

Finally, a common way to define the *average total effect* is to take the difference average of potential outcomes where everyone is treated and where everyone is control:

$$\tau_{TOT} = \bar{y}(1, \mathbf{1}) - \bar{y}(0, \mathbf{0}) \tag{1.9}$$

where $\mathbf{1}$ and $\mathbf{0}$ denote the exposures where all neighbors receive treatment and control respectively. We can consider two decompositions of the average total effect into direct and indirect effects:

$$\tau_{TOT} = \tau(\mathbf{1}) - (\bar{y}(0, \mathbf{1}) - \bar{y}(0, \mathbf{0})) \tag{1.10}$$

and

$$\tau_{TOT} = \tau(\mathbf{0}) - (\bar{y}(1, \mathbf{1}) - \bar{y}(1, \mathbf{0})). \tag{1.11}$$

An assumption of no weak interaction between direct and indirect effects ensures that the direct effects in these expressions (and hence, indirect effects) are the same across both decompositions. That is, if a hypothesis test determines that $\tau(\mathbf{1}) \neq \tau(\mathbf{0})$, then we are able to show a violation of the no weak interaction between direct and indirect effects assumptions. In this report, we develop a test for this last claim.

1.3.2 Literature Review

Recent work by Alzubaidi and Higgins (2022) seeks to detect treatment interference using this same anti-conflict program highlighted above as their motivating example. In their paper, Alzubaidi and Higgins suggest using a K-Nearest Neighbors Interference Model (KNNIM) to detect treatment interference. In the context of the anti-conflict program, this model only assumes that treatment interference can only occur within the nearest neighbors that a seeded student listed, and the closer of a neighbor somebody is to the treatment may lead to a larger influence on the response.

Sussman and Airoldi (2017) mention many assumptions about interference structure that are important to the larger scope of this work. The first assumption is the Additivity of Main Effects, which states that there is not interaction between a unit's treatment and the treatments of its neighbors. Thus, the potential outcomes are reduced to a summation of the outcome under no treatment $Y_i(0, 0)$, the direct effect of the treatment on the unit $Y_i(z_i, 0)$–$Y_i(0, 0)$, and the interference effect $Yi(0, Z_{ni})$–$Y_i(0, 0)$. The second assumption about the interference structure is the Additivity of Interference Effects, which deals with how interference effects accumulate among those units effected by the treatment of their neighbors. This idea is given by the summation of the outcome when no neighbors are treated and the additional effects of the neighbor's treatment on a unit. Both assump-

tions about interference structure will serve to be important in our case, because we wish to consider various assumptions on interaction between direct and indirect effects and develop hypothesis tests for these assumptions in 2.

Athey et al. (2018) propose calculation of p-values for network interference for non-sharp null hypotheses about treatment effects. Many of their considerations have useful logic in the context of the KNNIM where some hypotheses consider limiting the effect of treatment status from one unit to another according to the distance between units, which is similar to the idea that the effect of treatment status does not matter for units not in the same K-neighborhood. The scope of this paper focuses on scenarios where treatments, outcomes, and network connections are observed (remember that this is the same case for the New Jersey Middle School example where students identified their own network connections). Due to the interconnected nature of social networks, Athey et al. note that traditional large sample approximations are rendered difficult. Because of this, Athey et al. propose an approach based on randomization to derive p-values for interference without depending on network structure or sample size. The selection of focal units is again at the center of the methodology to reach empirical results when testing interference structure, because the statistical power of the tests is impacted with focal unit selection (see Alzubaidi and Higgins (2022)) and allows for derivation of exact p-values through artificial experiment (Athey et al., 2018).

1.4 Contribution

In this report, we develop a test of the no weak interaction between direct and indirect effects assumption. We argue that our test statistic—a form of a t-statistic—is approximately normal when interaction between effects is absent. We then perform a simulation study to investigate properties of our test. Through our simulation, we verify approximate normality of our test statistic and Type I errors close to the given nominal significance level α. We then investigate how the power of the test changes across different specifications of interaction

between direct and indirect effects.

Chapter 2

Hypothesis test for no interaction between direct and indirect effects

We now detail our test of no interaction between direct and indirect effects.

2.1 Derivation of the Hypothesis Test

2.1.1 Hypotheses

Under no interaction, we expect the direct effect to be same regardless of the exposure given to the neighboring units (see Chapter 1). This yields the following null hypothesis:

$$H_0 : \bar{y}(1, \mathbf{1}) - \bar{y}(1, \mathbf{0}) = \bar{y}(0, \mathbf{1}) - \bar{y}(0, \mathbf{0}) \tag{2.1}$$

Or equivalently

$$H_0 : \tau_{int} = \bar{y}(1, \mathbf{1}) - \bar{y}(1, \mathbf{0}) - \bar{y}(0, \mathbf{1}) + \bar{y}(0, \mathbf{0}) = 0 \tag{2.2}$$

2.1.2 Test statistics

To estimate the left hand side of (2.2), we can take the sample means of the 4 defined exposures:

$$\hat{\tau}_{int} = \overline{Y}_{obs}(1, \mathbf{1}) - \overline{Y}_{obs}(1, \mathbf{0}) - \overline{Y}_{obs}(0, \mathbf{1}) + \overline{Y}_{obs}(0, \mathbf{0}) \tag{2.3}$$

where

$$\overline{Y}_{obs}(W_i, \mathbf{W}_{-i}) = \frac{1}{N(W_i, \mathbf{W}_{-i})} \sum_{i=1}^{n} Y_i^{obs}(W_i, \mathbf{W}_{-i}) \mathbf{1}(W_i, \mathbf{W}_{-i}). \tag{2.4}$$

Here, $N(W_i, \mathbf{W}_{-i})$ is the number of units assigned exposure $(W_i, \mathbf{W}_{-i})$, $Y_i^{obs}(W_i, \mathbf{W}_{-i})$ is the observed response of unit i when i is given exposure $\mathbf{1}(W_i, \mathbf{W}_{-i})$, and $\mathbf{1}(W_i, \mathbf{W}_{-i})$ is an indicator variable denoting that unit i received exposure $(W_i, \mathbf{W}_{-i})$—this variable equals 1 when this exposure is observed, and is 0 otherwise. Note, depending on the randomization scheme—including complete and Bernoulli randomization—the quantity $N(W_i, \mathbf{W}_{-i})$ is a random variable.

Note, the quantity $\overline{Y}_{obs}(W_i, \mathbf{W}_{-i})$ is the Hájek estimator for $\bar{y}(W_i, \mathbf{W}_{-i})$. Hájek estimators tend to be biased but consistent, and the variance of these estimators tend to be quite small. Under certain conditions (which we will talk about in Section 2.1.3), Hájek estimators can be unbiased for their respective estimands (Lohr, 2019).

Given unbiased estimation of the Hájek estimators, through linearity of expectations, we have that

$$E(\hat{\tau}_{int}) = \tau_{int}. \tag{2.5}$$

Additionally, given independence of responses Y_i, we will have

$$Var(\hat{\tau}_{int}) = \frac{\sigma^2(1, \mathbf{1})}{N(1, \mathbf{1})} + \frac{\sigma^2(1, \mathbf{0})}{N(1, \mathbf{0})} + \frac{\sigma^2(0, \mathbf{1})}{N(0, \mathbf{1})} + \frac{\sigma^2(0, \mathbf{0})}{N(0, \mathbf{0})} \tag{2.6}$$

where $\sigma^2(W_i, \mathbf{W}_{-i})$ denotes the variance in the potential outcomes under exposure $(W_i, \mathbf{W}_{-i})$:

$$\sigma^2(W_i, \mathbf{W}_{-i}) = \frac{1}{n-1} \sum_{i=1}^{n} y_i(\sigma^2(W_i, \mathbf{W}_{-i}) - \bar{y}(W_i, \mathbf{W}_{-i}))^2. \tag{2.7}$$

We estimate each of the variance terms using the standard sample variance:

$$\hat{\sigma}^2(W_i, \mathbf{W}_{-i}) = \frac{1}{N(W_i, \mathbf{W}_{-i}) - 1} \sum_{i=1}^{N} (Y_i^{obs}(W_i, \mathbf{W}_{-i}) - \overline{Y}_{obs}(W_i, \mathbf{W}_{-i}))^2 \mathbf{1}(W_i, \mathbf{W}_{-i}). \tag{2.8}$$

Hence, we obtain an estimated variance of the estimator $\hat{\tau}_{int}$ as

$$\widehat{Var}(\hat{\tau}_{int}) = \frac{\hat{\sigma}^2(1, \mathbf{1})}{N(1, \mathbf{1})} + \frac{\hat{\sigma}^2(1, \mathbf{0})}{N(1, \mathbf{0})} + \frac{\hat{\sigma}^2(0, \mathbf{1})}{N(0, \mathbf{1})} + \frac{\hat{\sigma}^2(0, \mathbf{0})}{N(0, \mathbf{0})} \tag{2.9}$$

And derive our t-statistic for our hypothesis test

$$t_{stat} = \frac{\hat{\tau}_{int}}{\sqrt{\widehat{Var}(\hat{\tau}_{int})}} \tag{2.10}$$

Again, under independence of responses can be ensured through careful selection of "focal" units, central limit theorems from sampling theory will show approximate normality of t_{stat} (Lohr, 2019). If sample sizes are small, provided no outliers in the potential potential outcomes, we can estimate the distribution of t_{stat} "conservatively" as a t-distribution where the degrees of freedom are given as

$$df = \min_{W_i, \mathbf{W}_{-i}} N(W_i, \mathbf{W}_{-i}) - 1 \tag{2.11}$$

where this minimum is taken over the four considered exposures.

2.1.3 Selecting a subset of units for analysis

A difficulty with estimation of these quantities is dependence between responses—the neighborhood exposures from one unit may affect the treatment status (and hence, response) of a neighboring unit. One way to subvert this issue is to perform analysis on a subset of "focal units" which do not have overlapping neighborhoods. If this is the case, we can independence of responses of focal units if, for example, we assign treatment through Bernoulli randomization.

One way to do this is to select a set of focal units that create a 2-net of the original dataset (Alzubaidi and Higgins, 2022; Athey et al., 2018).

Intuitively, if we know which units may interfere with each other, we draw a set of units such that no neighborhoods of units overlap and such that, for any non focal unit, it either belongs to the neighborhood of a focal unit or belongs the neighborhood of a unit in the neighborhood of a focal unit.

In particular, Athey et al. (2018) propose a symmetric adjacency matrix to measure the network between units, where $G_{ij} = 1$ if i and j are neighbors. Athey et al. (2018) also discuss a vector of attributes X_i which are pre-treatment variables not affected by the treatment. In our case, we will call these covariates which will be used to determine nearest neighbors (determined by smallest euclidean distances) and subsequent focal unit selection.

2.2 Simulation under no interaction

The simulation begins by specifying the number of units (rows in the dataset) that are generated. In our case as shown in 2.3, this number was either set to 1024 or 2048. Next, the size of the K-neighborhood was chosen to be 4 (1 unit and its 3 nearest neighbors). Also shown in 2.3, the coefficient vector varied across the series of simulations to include models for direct effects only, indirect effects only, and both direct and indirect effects.

For each run of the simulation, we repeated generating the units 1000 times (i.e. gen-

erating a 1024 row or 2048 row dataset 1000 times). Covariates are then generated from i.i.d. $N(0,1)$, and a (Euclidean) distance matrix is created. We then find the three nearest neighbors for all unit by taking the three smallest distances. Next, we created a function to get the response Y_i for every unit, which is shown in (2.12) for response models under no interference and (2.13) for response models with interference. We assume the model of response

$$Y_i = X_1 + X_2 + X_3 + \beta_1 W_{i1} + \beta_2 W_{i2} + \beta_3 W_{i3} + \beta_d W_i \tag{2.12}$$

where X_1, X_2, X_3 are the covariates, $\beta_1, \beta_2, \beta_3$ are treatment effects for the 1st, 2nd, and 3rd nearest neighbor, W_{i1}, W_{i2}, W_{i3} are the treatment statuses of the 1st, 2nd, and 3rd nearest neighbor, β_d is the direct effect.

Next, a function is created to get the potential outcomes for every unit. Since this is an artificial experiment, we can simulate the response for each unit for if they would have received treatment and if they would have received control. See 1.2. To determine which units receive control and which units receive treatment, we use simple Bernoulli randomization ($Bernoulli(0.5)$), with 0 being a unit receiving control and 1 being a unit receiving treatment.

The dataset is then generated with the columns being the response value for each unit, the 3 randomly generated covariates, the treatment status of the unit, and the treatment status of the 3 nearest neighbors to the unit. We explore 3 different models of response by varying the β coefficients; these 3 different models do not exhibit interaction between direct and indirect effects.

2.3 Results of simulation under no interaction

When the null hypothesis holds, the simulation produces test statistics that follow an approximately normal distribution with roughly 5 percent of the observations lying outside of the standard normal distribution's 1.96 (or -1.96) critical value. When the null hypothesis holds, the test statistics will generate a non-significant result (p is greater than .05) from

the Kolmogorov-Smirnov test. Since we are aiming for a more conservative estimate as mentioned in the previous section, can use t-distribution with degrees of freedom based on the minimum exposure minus 1. This leads to higher critical value and thus a lower rejection rate as opposed to the asymptotic distribution. As hypothesized, the test statistics generated as a result appeared to be approximately normal.

For all models of response, roughly 5 percent of the test statistics were above 1.96 or below the -1.96 threshold of the standard normal distribution. For a more conservative estimate, the exposure with the least amount of samples in the 1000 run simulation we will call "min". We then calculate the critical value associated with the $t_{(.025,min-1)}$ distribution. This leads to a lower type 1 error rate and provides a concrete example how this critical value approach will lead to more conservative estimates as shown in 2.1, 2.2, 2.3, 2.4.

Table 2.1: *Type I error rates under the null hypothesis of no interaction*

Coefficient Vector	N	Asymptotic	Conservative
0,0,0,4 (Direct Effects Only)	1024	0.06	0.01
0,0,0,4 (Direct Effects Only)	2048	0.048	0.032
3,2,1,4 (Direct and Indirect Effects)	1024	0.063	0.02
3,2,1,4 (Direct and Indirect Effects)	2048	0.054	0.034
3,2,1,0 (Indirect Effects Only)	1024	0.068	0.035

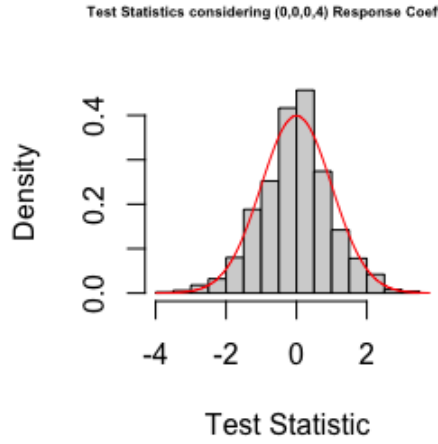
2.4 Simulation under interaction

Next, we will consider a set of simulations where we make sure that there is interaction between units by adding an interaction factor between the treated unit and its nearest neighbor. The model of response is as follows:

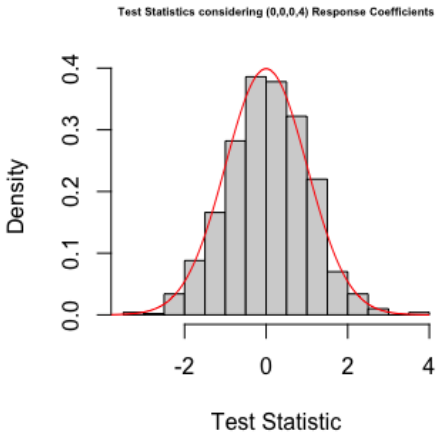
$$Y_i = X_1 + X_2 + X_3 + \beta_1 W_{i1} + \beta_2 W_{i2} + \beta_3 W_{i3} + \gamma(W_{i1} \times W_i) \tag{2.13}$$

for $\gamma = 0.25, 0.5, 1, 2$.

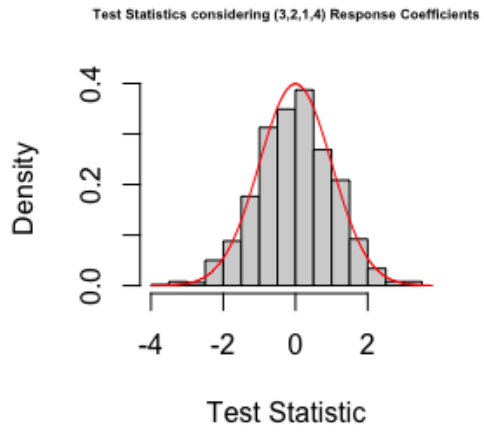
It can be seen that this is very similar to (2.12), except now we are adding γ as an



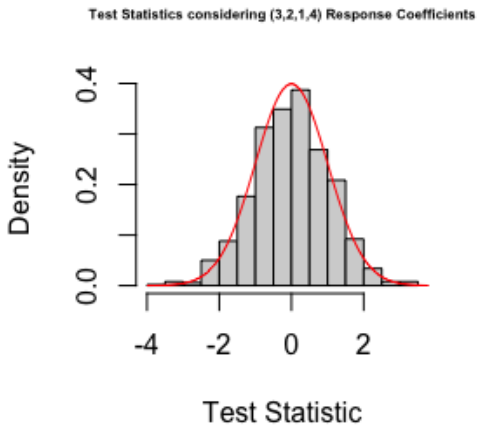
n = 1024



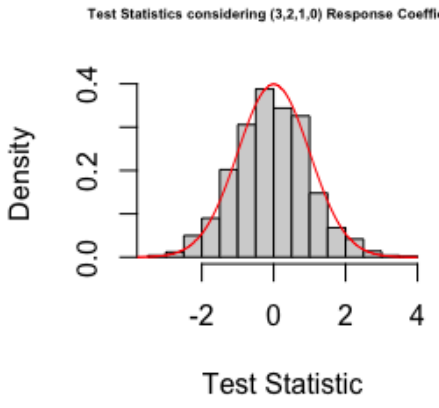
n = 2048



n = 1024



n = 2048



n = 1024

Figure 2.1: *Distribution of test statistics under no interaction*

interaction coefficient between the unit who received treatment and its nearest neighbor. This model of response could be expanded to more than 3 nearest neighbors, and could be changed to assume other sorts of interactions between a treated unit and its neighbors.

2.5 Results of simulation under interaction

When an interaction effect is added to the simulation, we see a shift to the right in the distribution of test statistics. The test statistics are still normally distributed, proven by non-significant p-values given by the Shapiro-Wilks normality test. However, we do find significant p-values resulting from the Kolmogorov-Smirnov test, meaning that we have sufficient evidence to believe that the distribution of test statistics under interaction is different than the distribution of test statistics under no interaction as shown in 2.5. As the interaction effect becomes more extreme, the distribution shifts farther to the right, leading to higher asymptotic and conservative rejection rates.

Table 2.2: *Rejection rates under interaction for direct effects only*

Coefficient Vector	Gamma	N	Asymptotic	Conservative
0,0,0,4	0.25	1024	0.069	0.036
0,0,0,4	0.5	1024	0.077	0.013
0,0,0,4	1	1024	0.142	0.087
0,0,0,4	2	1024	0.415	0.212
0,0,0,4	0.25	2048	0.068	0.049
0,0,0,4	0.5	2048	0.127	0.085
0,0,0,4	1	2048	0.229	0.198
0,0,0,4	2	2048	0.698	0.636

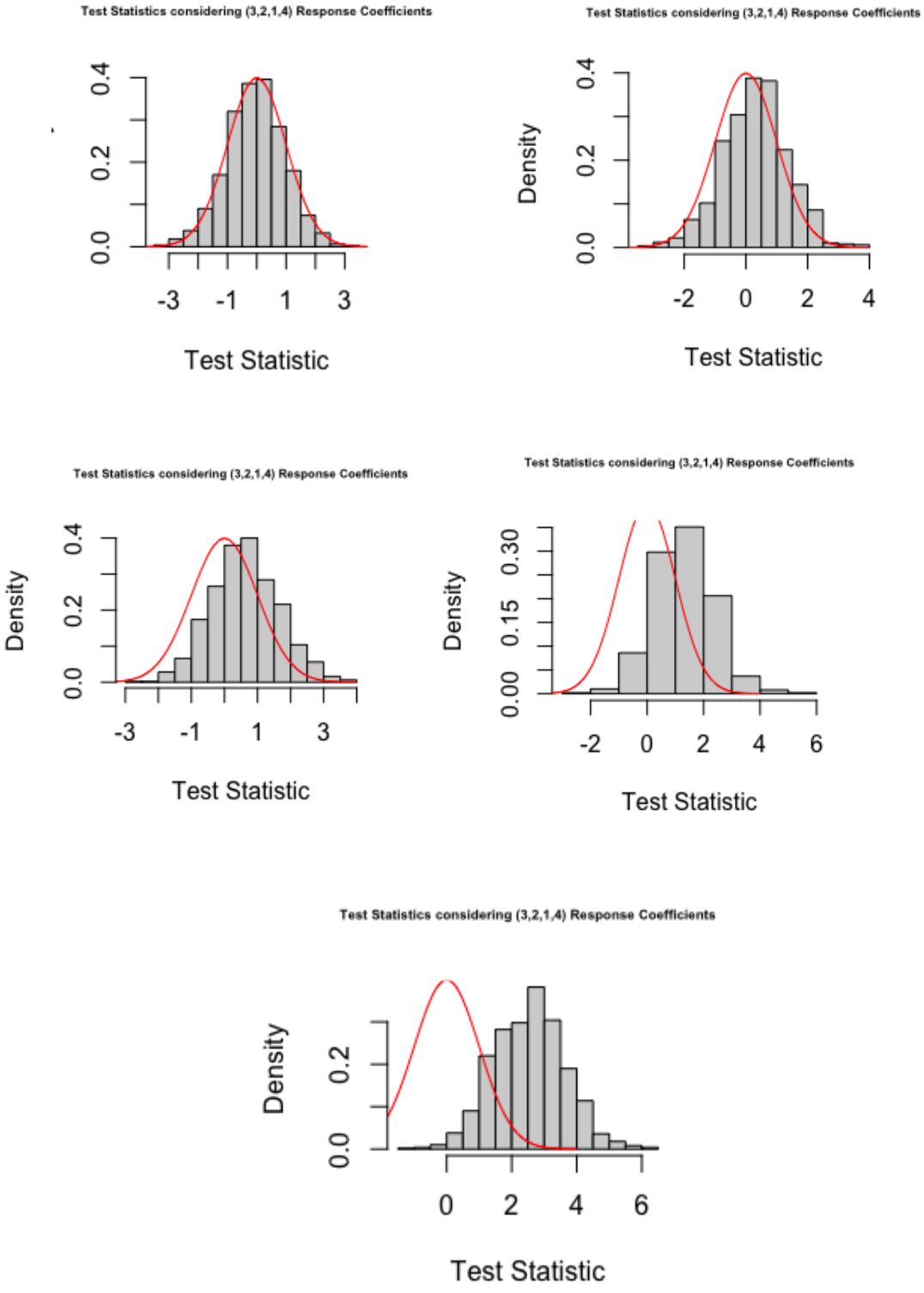


Figure 2.2: *Distribution of test statistics under response model* (2.13) for Coefficient Vector 3,2,1,4 (n=2048) and $\gamma = 0, 0.25, 0.5, 1, 2$

Table 2.3: *Rejection rates under interaction for both direct and indirect effects*

Coefficient Vector	Gamma	N	Asymptotic	Conservative
3,2,1,4	0.25	1024	0.06	0.028
3,2,1,4	0.5	1024	0.09	0.046
3,2,1,4	1	1024	0.156	0.037
3,2,1,4	2	1024	0.411	0.216
3,2,1,4	0.25	2048	0.082	0.05
3,2,1,4	0.5	2048	0.1	0.071
3,2,1,4	1	2048	0.268	0.213
3,2,1,4	2	2048	0.689	0.629

Table 2.4: *Rejection rates under interaction for indirect effects only*

Coefficient Vector	Gamma	N	Asymptotic	Conservative
3,2,1,0	0.25	1024	0.068	0.026
3,2,1,0	0.5	1024	0.08	0.048
3,2,1,0	1	1024	0.147	0.088
3,2,1,0	2	1024	0.388	0.261

Chapter 3

Conclusion

3.1 Conclusion

Experiments exhibit interference when the treatment status of one unit affects the response of another unit. When interference is present, treatment effects can be classified as either direct—effects due to the treatment status given to that unit—or indirect—due the the treatment status of neighboring units. However, unless there is no interaction between direct and indirect effects, the direct effects may vary depending on the exposure given to indirect effects. This makes defining causal effects difficult and estimation of these effects tricky.

We successfully developed a test and ran various simulations to detect weak interaction between direct and indirect effects across units. We performed a simulation study to verify approximate normality of our test statistics, and we perform a thorough simulation study to verify that Type I errors meet the nominal significance levels and to assess the power of our tests.

3.2 Future Work

In future work, simulations could be ran that consider generating larger datasets as this was the biggest time constraint across the entire project. Additionally, we could consider different models of interference since we only considered interference between a unit and its nearest neighbor in (2.13). This could be updated to include interference between a unit and all of its nearest neighbors etc. Another point of interest may be to consider changing the response structure to see how the distribution of test statistics is affected. Finally, we may consider creating tests for stronger forms of interaction between direct and indirect effects

We will also consider applying our test to real world data, like the New Jersey Middle Schools example to determine if there is interaction between direct and indirect effects.

Bibliography

Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

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

Michael E Sobel. What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *Journal of the American Statistical Association*, 101 (476):1398–1407, 2006.

Paul R Rosenbaum. Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477):191–200, 2007.

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

Fredrik Sävje, Peter M Aronow, and Michael G Hudgens. Average treatment effects in the presence of unknown interference. *The Annals of Statistics*, 49(2):673–701, 2021.

Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688, 1974.

Paul W Holland. Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960, 1986.

Jerzy Splawa-Neyman, Dorota M Dabrowska, and TP Speed. On the application of probability theory to agricultural experiments. essay on principles. section 9. *Statistical Science*, pages 465–472, 1990.

Peter M Aronow. A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research*, 41(1):3–16, 2012.

Daniel L Sussman and Edoardo M Airoldi. Elements of estimation theory for causal effects in the presence of network interference. *arXiv preprint arXiv:1702.03578*, 2017.

Samirah H Alzubaidi and Michael J Higgins. Detecting treatment interference under the k-nearest-neighbors interference model. *arXiv preprint arXiv:2203.16710*, 2022.

Susan Athey, Dean Eckles, and Guido W Imbens. Exact p-values for network interference. *Journal of the American Statistical Association*, 113(521):230–240, 2018.

Sharon L Lohr. *Sampling: design and analysis*. Chapman and Hall/CRC, 2019.