Causal inference under the K-nearest neighbors interference model

by

Samirah Alzubaidi

M.S., Kansas State University, 2017

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Abstract

In causal inference, an experiment exhibits treatment interference when the treatment status of one unit affects the response of other units. While traditional causal inference methods often assume no interference between units, there has been a recent abundance of work on the design and analysis of experiments under treatment interference—for example, those conducted on social networks. Failure to account for interference may lead to biased estimates of treatment effects and wrong conclusions.

In this dissertation, we propose the K-nearest neighbors interference model (KNNIM)—a model of treatment interference where the response of a unit depends only on its treatment status and the statuses of units within its K-neighborhood. Current methods for detecting interference include carefully designed randomized experiments and conditional randomization tests on a set of focal units. We give guidance on how to choose focal units under KNNIM. We then conduct a simulation study to evaluate the efficacy of existing methods for detecting arbitrary network interference under KNNIM with this choice of focal units. We show that this choice of focal units leads to powerful tests of treatment interference which outperform experimental methods.

Then, we extend the potential outcomes approach and the K-neighborhood interference framework to define causal estimands for direct and K-nearest neighbors indirect effects where interference is allowed within K-neighborhoods of individuals. Under completely randomized and Bernoulli-randomized designs, we provide a closed-form solution to compute the marginal and joint probabilities of units being exposed to treatment exposures of interest. We then propose Horvitz-Thompson unbiased estimators for the defined estimands under K-neighborhood interference assumption. We derive properties of the proposed estimators and provide conservative variance estimators. We then demonstrate how an assumption of no interaction between direct and indirect effects can improve estimates. To demonstrate the

proposed causal methods, we perform a simulation study and apply our proposed methods on an anti-conflict study from a randomized experiment among middle schools students in New Jersey.

Finally, we develop additional estimators of the defined estimands under an assumption of no interaction between the indirect effects. This may enhance the estimation standard errors by increasing the number of units under this assumption. Properties of the developed estimators are derived as well as conservative variance estimators of the defined estimands.

Causal inference under the K-nearest neighbors interference model

by

Samirah Alzubaidi

M.S., Kansas State University, 2017

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2022

Approved by:

Major Professor
Michael Higgins

Copyright

Abstract

In causal inference, an experiment exhibits treatment interference when the treatment status of one unit affects the response of other units. While traditional causal inference methods often assume no interference between units, there has been a recent abundance of work on the design and analysis of experiments under treatment interference—for example, those conducted on social networks. Failure to account for interference may lead to biased estimates of treatment effects and wrong conclusions.

In this dissertation, we propose the K-nearest neighbors interference model (KNNIM)—a model of treatment interference where the response of a unit depends only on its treatment status and the statuses of units within its K-neighborhood. Current methods for detecting interference include carefully designed randomized experiments and conditional randomization tests on a set of focal units. We give guidance on how to choose focal units under KNNIM. We then conduct a simulation study to evaluate the efficacy of existing methods for detecting arbitrary network interference under KNNIM with this choice of focal units. We show that this choice of focal units leads to powerful tests of treatment interference which outperform experimental methods.

Then, we extend the potential outcomes approach and the K-neighborhood interference framework to define causal estimands for direct and K-nearest neighbors indirect effects where interference is allowed within K-neighborhoods of individuals. Under completely randomized and Bernoulli-randomized designs, we provide a closed-form solution to compute the marginal and joint probabilities of units being exposed to treatment exposures of interest. We then propose Horvitz-Thompson unbiased estimators for the defined estimands under K-neighborhood interference assumption. We derive properties of the proposed estimators and provide conservative variance estimators. We then demonstrate how an assumption of no interaction between direct and indirect effects can improve estimates. To demonstrate the

proposed causal methods, we perform a simulation study and apply our proposed methods on an anti-conflict study from a randomized experiment among middle schools students in New Jersey.

Finally, we develop additional estimators of the defined estimands under an assumption of no interaction between the indirect effects. This may enhance the estimation standard errors by increasing the number of units under this assumption. Properties of the developed estimators are derived as well as conservative variance estimators of the defined estimands.

Table of Contents

List of Figures

List of Tables

Acknowledgments

All praise and thanks are to Almighty God for giving me the strength, patience, and the courage to complete this dissertation. This would not have been possible if it had not been for Allah by my side.

I would also like to acknowledge and express my sincere thanks to Umm Al-Qura University and my country, Saudi Arabia, for generously funding my education through a scholarship.

My sincere appreciation and deepest gratitude go to my advisor, Dr. Michael Higgins, for his support, encouragement, guidance, and positivity. I appreciate his insightful suggestions, constant inspiration, and time to carry out this dissertation to its successful completion. It has been an honor and a privilege to work with him.

Special thanks go to Dr. Christopher Vahl for his support, encouragement, and open-door policy during my academic journey at K-State. I would also like to extend my deepest gratitude to the rest of my academic advisory committee members, specifically Dr. Juan Du and Dr. Pietro Poggi-Corradini, for their insightful and valuable guidance.

I would also like to extend my appreciation to the K-State Statistics Department faculty, staff, and graduate students for their support and kindness.

Many thanks to all my lovely family members. My deepest gratitude goes to my beloved father, Hamid, for his continued support and endless love. To my sisters, Samia, Hanan, Amal, Haneen, Raneem, and Sadeem, my brothers, Mosa, Abdullah, and Mohammed, my nephews and nieces, thank you for your support, encouragement, prayers and belief in me.

I am also extremely grateful to the person who supports me more than anyone else, my husband, Sameer. He is always there for me. I cannot thank him enough for his unconditional support, sacrifice, care, encouragement, and patience through this journey. I also extend my thanks to my beloved children, Remas, Khalid, and Raseel, who have been my motivation, inspiration, and drive. I am thankful they took this journey with me. Words are not capable of describing how grateful I am to them.

Dedication

To my father, Hamid, for raising me to pursue my dreams and to always reach for more. You showed me that through hard work and patience, success will follow. You gave me the ability to believe in myself and to follow my dreams.

To the memory of my mother, Fatima. You were the first to believe in me and the first to encourage me. None of this would be possible if it were not for your love, knowledge, and support. You taught me the language of hope, success, and inspiration. I would not be where I am today without you.

Chapter 1

Introduction

1.1 Introduction

Randomized experiments have long been viewed as the most reliable method in causal inference for evaluating the causality of an intervention. There is a rising number of studies on social networks in which the social influence takes place. Technology companies such as Google, Amazon, Facebook, LinkedIn, Netflix, Twitter, and others run online randomized controlled experiments to evaluate the effect of a new feature or product on user engagement. In epidemiology, researchers may want to study the effect of vaccines on a target population to protect individuals who are at risk for an infectious disease.

However, these settings involve interaction between units under study; for example, a user assigned a new feature may interact with a user who is not assigned the feature, thereby impacting the response of the latter user. This interaction complicates the estimation and inference of treatment effects under classical causal inference methodologies.

In particular, a fundamental assumption in the traditional causal inference framework is that there is only a single version of each treatment status and the response of a unit is unaffected by the treatment status of any other unit. This is known as the *stable unit treatment value assumption* (SUTVA) (Rubin, 1980). SUTVA is violated under settings in which there is *treatment interference*—when a treatment assigned to a unit affects the response of other units. Effects on response due to treatment interference are also known as spillover, indirect effects, peer influence, social interaction, or network effects.

The dependence of a unit's outcome on other units' exposures or treatments poses statistical challenges because the potential outcome of a unit—the hypothetical outcome of a unit given a realized treatment assignment—is not only affected by its own treatment status (a *direct effect* of treatment) but also by the treatment conditions received by other units (an *indirect effect*). In traditional causal inference, interference has been considered as a nuisance and researchers may design experiments that control interference and reduce the bias of estimating the primary effect. Although these designs may minimize the effect of interference, such designs are not always possible.

In other settings, there has been a growing interest in estimating the causal effect in the presence of interference. Examples of this include studies on the efficacy of vaccines in which vaccinated and non-vaccinated members of a population interact with one another and researchers are interested in overall infection rates (Ross, 1916; Halloran and Struchiner, 1995; Moulton et al., 2001; King Jr et al., 2006; Hudgens and Halloran, 2008). In behavioral sciences, applications include experiments conducted to study the change of the community social norms and behaviors by interventions applied to a group of the community (Paluck and Shepherd, 2012; Schaefer et al., 2012; Paluck et al., 2016; Basse and Feller, 2018).

Even though causal inference under interference is still an open area of research, considerable work has been devoted to the development of reasonable models of interference to ensure identification of both the direct effect of treatment and the effect of treatment spillover on the response (Toulis and Kao, 2013; Aronow and Samii, 2017; Basse and Feller, 2018; Forastiere et al., 2020; Sussman and Airoldi, 2017).

In this thesis, we extend the potential outcomes approach and introduce a new framework of causal inference under interference called the K-Nearest Neighbors Interference Model (KNNIM). Under KNNIM, the response of a unit is affected only by the treatment given to that unit and the treatment statuses of its K-nearest neighbors (KNN). Such models of interference may be reasonable, for example, under social network settings, where only a few of the observable potential interactions (e.g. accounts that a Twitter user follows) may be influential on a unit's response. Under this setting, the strength of an interaction between two users may be quantified, for example, by assessing the amount of engagement (e.g. likes, comments, retweets) between the two users. We evaluate the performance of existing methods for detecting treatment interference under data generated under a KNNIM model as well as a new developed randomization-based test. We define causal estimands under K-neighborhood interference assumption and propose estimators for these estimands deriving properties and conservative variance estimators of the defined estimands. We also consider improvement of the estimation precision and propose different estimators under different assumptions.

1.2 Causal Inference under Neyman-Rubin Framework

Association does not imply causation. Associational inference focuses on the relationship between two or more variables and how they change together. Causal inference aims to infer the causal effect of an intervention on units and how the intervention affects the response variable.

Literature contains some approaches to causality with different perspectives. The Neyman-Rubin Causal Model (NRCM) framework, or simply the Rubin Causal Model as called in Holland (1986), where the potential outcome is a primary concept in this approach, is a popular model for causal inference in many fields. The concept of potential outcomes defining causal effects was first introduced formally by Neyman (1923) and only in the context of an urn model for assigning varieties to plots where this model is stochastically identical to the completely randomized experiment. Fisher (1925) took this framework a step further, and proposed the physical randomization of units and developed the analysis of randomized experiments. Rubin (1974) extended Neyman's work to a more general framework for causation that applies to both experimental and observational studies. Rubin (1975, 1978) also discussed the importance of randomization and formulated the assignment mechanism in terms of potential outcomes.

Suppose we have a finite population of N units indexed $i = 1, 2, \dots, N$ and each unit

i is assigned to a treatment condition W_i where $W_i \in \{0, 1\}$ is a binary random variable such that $W_i = 1$ if the i^{th} unit is assigned to the treatment condition and $W_i = 0$ if the i^{th} unit is assigned to the control condition. Let $W = (W_1, W_2, \ldots, W_N)$ denote the treatment assignment vector of all N units such that $W \in \mathbb{W}$ where $\mathbb{W} \in \{0, 1\}^N$ is the set of all possible treatments assignments.

The distribution of the treatment assignment vector $W = (W_1, W_2, \ldots, W_N)$ is crucial for the causal inference. In randomized experiments, the assignment mechanism does not depend on the characteristics of the units in the study and the distribution of W is known and the researcher has control over the assignments. In contrast, in observational studies, the assignment mechanism depends on the observed and unobserved characteristics of the units and the distribution of W is unknown.

Let $y_i(1)$ denote the potential outcome for unit i if unit i receives treatment, and let $y_i(0)$ denote the potential outcome for unit i if unit i is exposed to control.

The Neyman-Rubin Causal Model of the observed outcome of the i^{th} unit is

$$Y_i = W_i y_i(1) + (1 - W_i) y_i(0). \tag{1.1}$$

where $Y_i = y_i(1)$ if $W_i = 1$ and $Y_i = y_i(0)$ if $W_i = 0$ because for each unit i, only one potential outcome is observed, namely the potential outcome that corresponds to the realized level of the treatment, and the other potential outcome is unobserved.

The unit-level causal effect of treatment is defined as the difference between the two potential outcomes on unit i by $\delta_i = y_i(1)$ - $y_i(0)$.

However, the difficulty of inferring causality arises from the fact that we can only observe one of the two potential outcomes. This is called the fundamental problem in causal inference by (Holland, 1986) such that causal inference is considered a missing data problem. Even though the advantage of potential outcomes framework is to define causal estimands in terms of individual-level potential outcomes, only typical causal estimands defined in terms of the average of potential outcomes are estimable. Imbens and Rubin (2015) stated: "Although the definition of causal effects does not require more than one unit, learning about causal effects typically requires multiple units. Because with a single unit we can at most observe a single potential outcome, we must observe multiple units, some exposed to the active treatment, some exposed to the alternative (control) treatment." Hence, instead of estimating the causal effect for an individual unit, we estimate the average treatment effect (ATE).

In a finite population of N units, the potential outcomes, $y_i(1)$ and $y_i(0)$, are fixed quantities, nonrandom (there is no super-population beyond the N observed units), and independent of the treatment assignment W. In contrast, the observed outcomes depend on the treatment assignment W and hence are random variables where the only source of randomness in the observed outcomes is induced by the random selection of the realization of the treatment assignment W. In this setting, the average treatment effect becomes

$$\delta = \frac{1}{N}\sum_{i=1}^{N}[y_i(1) - y_i(0)]. \tag{1.2}$$

A fundamental assumption of NRCM is the stable-unit treatment value assumption (SUTVA) (Rubin, 1980). This assumption has two components. First, the no-interference component in which every unit's outcome is affected only by its own treatment and not by any treatment of other units (Cox, 1958; Rubin, 1980). And second, where there is only a single version of each treatment level that defines a unique outcome on each unit. For example, if we consider a teaching strategy to be a treatment and if each teacher applies the strategy in a significantly different way than other teachers, then we must consider this as a different treatment depending on the teacher. In some settings, the first part of the SUTVA assumption is not possible. The following section provides an overview on some of the work that has been developed in settings where interference is present.

1.3 An Overview of Causal Inference under Interference

During the past decade, there exists a considerable body of literature on causal inference under interference. A series of contributions have been made ranging between the no-interference assumption for those who consider interference as a nuisance (i.e., no exposure to other units' treatments), and structured interference assumptions.

Starting from the no-interference assumption and considering interference as a nuisance parameter, some work in the literature focuses on reducing bias in standard estimates of causal effects in the presence of interference using designs that isolate units that might have some connections. Ugander et al. (2013) developed exposure models and proposed a graph cluster randomization scheme for computing exposure probabilities and unbiased estimator of average treatment effects. Gui et al. (2015) extended this work by studying the problem of network A/B testing in real networks proposing a network sampling algorithm and a new estimation method. Eckles et al. (2016) considered methods for bias reduction in the estimates of the average treatment effect through experimental designs and analysis.

Sävje et al. (2021) investigated the behavior of standard estimators under a weak form of interference. Karrer et al. (2020) introduced a framework accounting for interference through cluster-randomized experiments where they ran side-by-side unit-randomized trials and cluster-randomized trials and they introduced a cluster-based regression adjustment estimator to improve the precision of estimating treatment effects.

Rather than considering interference as a nuisance, some researchers tend to relax SUTVA and allow for interference in different ways considering interference effect as of primary interest. In this regard, one line of research focuses on a setting where the population of individuals can be partitioned into mutually exclusive groups such as households, schools, villages, hospitals, etc., where interference is allowed within groups but not across groups. This is referred to as *partial interference* assumption (Sobel, 2006), (i.e., SUTVA is assumed between groups). This can be justified if the groups are divided based on time or location (Hudgens and Halloran, 2008; Tchetgen and VanderWeele, 2012; Rosenbaum, 2007; Sobel, 2006; Basse and Feller, 2018; Offer-Westort and Dimmery, 2021).

Under a partial interference assumption, Sobel (2006) proposed causal estimands to assess house voucher effects by averaging over all possible treatment assignments. Rosenbaum (2007) developed nonparametric tests and confidence intervals to assess treatment effect under partial interference. Hudgens and Halloran (2008) considered a population of groups of individuals where interference is allowed within groups. Following the approach in Sobel (2006) approach, Hudgens and Halloran (2008) proposed estimands for direct, indirect, total and overall causal treatment effects under this setting of interference. They defined the direct causal effect of a treatment on a unit in a group as the difference between potential outcomes when only changing the unit's treatment and holding other units' treatment fixed. In contrast to direct effect, they defined the indirect effect on a unit as the effect of changing the treatments of other units in the same group on that unit holding its own treatment fixed. The total effect combines both direct and indirect effects by changing the treatments of both the particular unit and other units in the same group. Finally, the overall effect on a unit in a group is defined as the difference between potential outcomes changing the treatment for the group of that unit.

Assume we have n groups of units and n_i denotes the number of units in group i for $i = 1, \ldots, n$. Let ψ and ϕ be two treatment assignment strategies. Hudgens and Halloran (2008) defined the individual indirect causal effect estimand as the difference between the potential outcomes with treatment program $\mathbf{W_i}$ compared with $\mathbf{W_i'}$ on the individual j in group i by

$$CE^{I}_{ij}(\mathbf{W_{i(j)}}, \mathbf{W'_{i(j)}}) \equiv \mathbf{y_i}(\mathbf{W_{i(j)}}, \mathbf{W_{ij}} = \mathbf{0})) - \mathbf{y_i}(\mathbf{W'_{i(j)}}, \mathbf{W'_{ij}} = \mathbf{0}) \tag{1.3}$$

where $\mathbf{W_{i(j)}}$ is the subvector of possible values of treatments $\mathbf{W_i}$ for group i with the j^{th} entry deleted and W_{ij} is a possible value of treatment of the individual j in group i. The group and the population average indirect causal effect estimands respectively are defined as

$$\overline{\mathrm{CE}_{\mathrm{i}}}^{I}(\psi, \phi) \equiv \bar{y}_i(0, \psi) - \bar{y}_i(0, \phi) \tag{1.4}$$

and

$$\overline{\mathrm{CE}}^{I}(\psi, \phi) \equiv \bar{y}(0, \psi) - \bar{y}(0, \phi) \tag{1.5}$$

On the design side, Hudgens and Halloran (2008) considered a hierarchical design, also known as two-stage randomization design, in which two randomization procedures are considered: either groups are randomly assigned to treatment assignment strategies (i.e. different proportions of treated units) and then units within groups are randomly assigned to treatments conditional upon the group's strategy; or groups can be randomly assigned to treatment and control and then within each treated group, units are randomly assigned to treatment and control. In the former design, interference effects can be assessed by comparing units across groups with different proportions assigned to treatment (Hudgens and Halloran, 2008); the later design allows assessing interference effects by comparing units assigned to control in treated groups versus units assigned to control in control groups (Basse and Feller, 2018).

Under the two-stage randomization design, Hudgens and Halloran (2008) presented estimators of their proposed estimands that were shown to be unbiased. The unbiased estimator for the population average indirect causal effect is given by

$$\widehat{CE}^{I}(\psi, \phi) \equiv \bar{Y}(0, \psi) - \bar{Y}(0, \phi) \tag{1.6}$$

where $\bar{Y}(0, \psi)$ and $\bar{Y}(0, \phi)$ is the population average of the observed potential outcomes under strategies ψ and ϕ respectively.

Basse and Feller (2018) extended this work and considered the complexity that arises when groups sizes vary. They proposed individual-weighted and household-weighted estimands of the indirect and the total effects providing unbiased estimators for the two-stage weighted estimands.

Moreover, Tchetgen and VanderWeele (2012) expanded upon Hudgens and Halloran (2008) work and developed a finite sample framework for causal inference under interference. Assuming a binary outcome, they constructed a finite sample confidence interval for the four-population average causal effects of interest. They also provided extensions to inverse probability weighting approach to causal inference under interference in observational studies.

Offer-Westort and Dimmery (2021) added to the discussion of estimation and experimental design under partial interference where the main target is to estimate effects of multiple treatment conditions.

In some settings, the partial interference assumption may not be valid (Aronow and Samii, 2017; Toulis and Kao, 2013; Sussman and Airoldi, 2017). Aronow and Samii (2017) extended this work to go beyond partial interference to settings with arbitrary forms of interference generalizing estimation and inference theory. Their estimation framework consisted of experimental design, an exposure mapping and a set of causal estimands. They limited interference extent to a finite set S of exposures ($\mathbf{W_1}$, $\mathbf{W_2}$, ..., $\mathbf{W_S}$) in which case each unit i has a vector of probabilities of exposures $(\pi(\mathbf{W_1}), \pi(\mathbf{W_2}), \ldots, \pi(\mathbf{W_S}))' = \pi_i$ which is the probability of the unit i being subject to each of the S possible exposures.

They defined the average of the units' potential outcomes under exposure $\mathbf{W_s}$ as $\mu(\mathbf{W_s}) = (\mathbf{1/N})\mathbf{y^T}(\mathbf{W_s})$ where $y^T(\mathbf{W_s})$ is the total number of potential outcomes under $\mathbf{W_s}$. Hence, they defined the average unit-level causal effect $\tau(\mathbf{W_s}, \mathbf{W_l}) = \mu(\mathbf{W_s}) - \mu(\mathbf{W_l})$ as the difference between the average of units' potential outcomes under exposure $\mathbf{W_s}$ versus the average under another exposure $\mathbf{W_l}$. In addition, they provided an inverse probability weighted unbiased estimator by Horvitz and Thompson (1952) for the total number of the potential outcomes under $\mathbf{W_s}$ as

$$\widehat{y^T_{HT}}(\mathbf{W_s}) = \sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{I}(\mathbf{W'} = \mathbf{W_s}) \frac{\mathbf{Y_i}}{\pi(\mathbf{W_s})}, \tag{1.7}$$

where $\mathbf{W'}$ is the exposure that unit i receives and therefore, $\widehat{\mu_{HT}}(\mathbf{W_s}) = (\mathbf{1/N})\widehat{\mathbf{y^T_{HT}}}(\mathbf{W_s})$.

An unbiased estimator of the average unit-level causal effect $\tau(\mathbf{W_s}, \mathbf{W_l})$ is

$$\widehat{\tau_{HT}}(\mathbf{W_s}, \mathbf{W_l}) = \widehat{\mu_{\mathbf{HT}}}(\mathbf{W_s}) - \widehat{\mu_{\mathbf{HT}}}(\mathbf{W_l}). \tag{1.8}$$

Toulis and Kao (2013) extended the potential outcomes framework to allow for interference in social networks by defining a k-level estimand of the interference effect. Sussman and Airoldi (2017) developed elements of estimation theory for causal effects assuming that the potential outcomes of an individual depend only on individual's treatment and neighbors' treatment such that changing the treatment of other units in the network does not impact the outcome of unit i.

Aronow et al. (2021) reviewed methods of interference for both general network settings and partial interference within hierarchical structure in the context of randomized experiments.

Moreover, Basse and Airoldi (2018) considered the problem of designing a randomized experiment to minimize estimation error for correlated outcomes where the correlation among outcomes is informed by an available pre-intervention network.

Another research direction focuses on testing for interference and has been developed through a randomization-based approach (Aronow, 2012; Athey et al., 2018) or through an experimental design approach (Saveski et al., 2017; Pouget-Abadie et al., 2019). Aronow (2012) employed the randomization inference approach for testing non-sharp null hypotheses under interference between units where he provided a conditional randomization test of the analyst's choice that allows for the calculation of the exact significance level of the causal dependence of outcomes on the treatment status of other units. Athey et al. (2018) expanded upon this work and developed tests for a large class of hypotheses under interference. Basse et al. (2019) built on this work and considered the validity of the test by conditioning on observed treatment assignment of the subset of units who received an exposure of interest. Saveski et al. (2017) and Pouget-Abadie et al. (2019) presented an experimental design for testing whether SUTVA holds and provided theoretical bounds on the type I error rate.

1.4 K Nearest Neighbors Interference Model

We view units under study as a mathematical graph. Let $\mathbf{G} = (V, E)$ be a directed graph of $\|V\| = N$ vertices; each vertex $i \in V$ corresponds to a unit under study. An edge $ij \in E$ denotes potential interaction between units i and j. These interactions between units can be defined by any type of relationships—membership to the same group, friendship in social media, geographic proximity, etc. (Forastiere et al., 2020). Throughout the thesis, the terms vertex, unit, and individual will be used interchangeably.

Let A denote the $\mathbb{N} \times \mathbb{N}$ adjacency matrix of $\mathbf{G}$. That is, $A_{ij} = 1$ if $ij \in E$ where there is an edge from unit j to i and $A_{ij} = 0$ otherwise. Because $\mathbf{G}$ has no self-loops, the diagonal elements of the adjacency matrix, $A_{ii} = 0$.

Let $d(i, j)$ denote an interaction or dissimilarity measure between units i and j. In the context of treatment interference, smaller values of $d(i, j)$ indicate stronger interactions between units i and j. We assume, for now, that $d(i, j)$ is only computed for units i, j with $A_{ij} = 1$. Let $d(i, (j))$ denote the jth smallest value of $\{d(i, j), j \neq i\}$; that is, $d(i, (1)) < d(i, (2)) < \cdots$. For ease of exposition, we assume that all values of $d(i, j)$ are unique (in practice, ties may be broken arbitrarily). The *K-neighborhood* of unit i, denoted $\mathcal{N}_{iK}$, is the set of the K "closest" units to unit i:

$$\mathcal{N}_{iK} = \{j : d(i, (j)) \leq d(i, (K)), j = 1, 2, \ldots, K\}. \tag{1.9}$$

Define $\mathcal{N}_{-iK} = V \setminus (i \cup N_{iK})$ as all units in V that are outside of i's K-neighborhood. Note that the sets $\{i, \mathcal{N}_{ik}, \mathcal{N}_{-ik}\}$ form a partition of V.

Recall that W_i is a treatment indicator for unit i, and let $W = (W_1, W_2, \ldots, W_N) = \{W_i, W_{\mathcal{N}_{ik}}, W_{\mathcal{N}_{-ik}}\}$ denote treatment assignment vector for all units N. Also recall that Y_i is the outcome measured on the unit i. Each unit's potential outcome $y_i(W)$ is defined as a function of the entire assignment vector of units to the two treatment conditions $W \in \{0, 1\}^N$.

Recall that, under SUTVA, the outcome of unit i depends only on the treatment assigned to unit i. That is, for two randomizations W, W', SUTVA implies that $y_i(W) = y_i(W')$ if

$W_i = W_i'$. However, when treatment effects interfere across units, units' responses are not only affected by their own treatments but also by the treatments assigned to other units in the network.

Treatment interference models range between assuming no exposure to other units' treatments—which is the case for traditional randomized experiments with Neyman-Rubin causal model under SUTVA—and structured interference models. In completely arbitrary interference exposure, each unit will have a unique type of exposure depending on the treatment assignment for all N individuals. This results in distinct 2^N potential outcomes for each unit and $N2^N$ potential outcomes for the experimental population where we only observe N of these potential outcomes. Under the latter exposure, there would be no meaningful way to analyze the experiment so that researchers focus on structured or limited interference.

To make progress on treatment interference problems, researchers make assumptions between those of SUTVA and arbitrary interference models that restrict the extent of interference allowed (Toulis and Kao, 2013; Aronow and Samii, 2017; Ugander et al., 2013; Sussman and Airoldi, 2017).

Remark 1.1. *Aronow and Samii (2017) define exposure mapping as a function that maps an assignment vector and unit i specific traits to an exposure value where there is a finite set of exposure values. Toulis and Kao (2013) introduce k-level interference where unit i is exposed to interference effect if at least one neighbor is treated or unit i is k-exposed if exactly k neighbors are treated for $i \in V_k$ where V_k is the set of units that have at least k neighbors.*

Remark 1.2. *Sussman and Airoldi (2017) consider neighborhood interference assumption (NIA) where the potential outcome of an individual depends only on an individual's treatment and neighbors' treatment such that unit j is a neighbor of unit i if $A_{ij} = 1$. The neighborhood of unit i is the set of all vertices with edges directed toward unit i denoted as $\mathcal{N}_i$ where $\mathcal{N}_i = \{j : A_{ij} = 1\}$. NIA along with other assumptions lead to various models for the potential outcomes where someone can define estimands for the direct and indirect treatment effects in terms of the model parameters. Similarly, Forastiere et al. (2020) consider neighborhood interference assumption excluding the dependence of the potential outcome of unit i from* *treatments assigned outside the neighborhood $\mathcal{N}_i$. Forastiere et al. (2020) assumption differs from Sussman and Airoldi (2017) assumption in that they assume that the dependence of the potential outcomes is defined through a specific function $g_i(.)$. Applying this function to the neighborhood treatment vector results in a variable denoted by $U_i = g_i(W_{\mathcal{N}_i})$, that can explain the type of the dependence, for example, the number of treated neighbors, (i.e., $U_i = \sum_{j \in \mathcal{N}_i} W_j$). Ugander et al. (2013) develop different exposure models of interference. One exposure model is the full neighborhood exposure to a treatment where unit i and all its neighbors receive that treatment condition. An absolute k-neighborhood exposure to treatment is that for vertex i with degree $d \geq K$, vertex i and $\geq K$ neighbors of i receive that treatment condition. In addition, fractional q-neighborhood exposure to a treatment that is a vertex i and $\geq qd$ neighbors of i receive that treatment condition. They also introduce stricter versions of the above exposures using core exposures.*

However, most models in previous work only specify that the units' outcomes are affected by the number or fraction of treated neighbors, but they do not specify which neighbors affect unit outcome and how they affect the outcome. We extend the literature on causal inference under interference and propose an interference model that differs from the above exposure models in that we restrict the interference of treatment on a unit i to its the K nearest neighbors. This allows different neighbors to contribute different effects depending on the proximity of the relationship—neighbors who are close to unit i are more likely to have their treatment statuses affect the response of unit i. In other words, we relax SUTVA to allow treatment effects to interfere between units, but we limit the extent of interference only to their K-nearest neighbors. Thus, we restrict the number of possible treatment assignment vectors and hence, the number of potential outcomes to be 2^{K+1} for each unit. Now we define the K-nearest neighbors interference model under the following assumption:

Assumption 1.1. *(K-Neighborhood Interference Assumption (K-NIA)). For each unit i in a network $\mathbf{G}$ and for all treatment assignments $W_{\mathcal{N}_{-iK}}$, $W'_{\mathcal{N}_{-iK}}$, the potential outcomes* *satisfy K-Neighborhood Interference Assumption if*

$$y_i(W_i, W_{\mathcal{N}_{iK}}, W_{\mathcal{N}_{-iK}}) = y_i(W_i, W_{\mathcal{N}_{iK}}, W'_{\mathcal{N}_{-iK}}) \tag{1.10}$$

Assumption 1.1 states that the potential outcome of unit i is only affected by its treatment and by the treatments assigned to its K-nearest neighbors. Changing treatments for other units outside the K-neighborhood will not affect the potential outcome of unit i. This is a special case of the NIA described in Sussman and Airoldi (2017). In its most general form, the K-nearest neighbors interference model (KNNIM) assumes only that the treatment interference structure satisfies assumption 1.1.

In this thesis, we use the nearest neighbor effect, indirect effect and interference terms interchangeably and we refer to the unit's response to treatment as a direct effect and the unit's response to interference as an indirect effect as in (Hudgens and Halloran, 2008).

1.5 Organization of the Dissertation

In Chapter 2, we evaluate the performance of existing methods for detecting arbitrary interference under the K-nearest neighbors interference model. In Chapter 3, we define estimands for the direct, indirect, total ,and ℓ_{th} nearest neighbor effects under the K-neighborhood interference assumption and provide estimators of the defined effects. In Chapter 4, we propose estimators under the no-interaction between indirect effects. We conclude in Chapter 5.

Chapter 2

Detecting Interference under K-Nearest Neighbors Interference Model

2.1 Introduction

Randomized experiments have long been viewed as the gold standard for causal inference. In epidemiology, researchers may want to study the effect of vaccines on a target population to protect individuals who are at risk of an infectious disease. Technology companies such as Google, Amazon, Facebook, LinkedIn, Netflix, Twitter, and others run online randomized controlled experiments to evaluate the effect of a new feature or product on user engagement. However, in such settings, units under study may interact with one another; for example, a user assigned a new feature may interact with one not assigned the feature, thereby impacting the response of the latter user. This interaction poses challenges in estimating and inferring treatment effects under traditional causal inference methodologies.

In particular, a fundamental assumption in the traditional causal inference framework is that there is only a single version of each treatment status and the response of a unit is unaffected by the treatment status of any other unit (see Imbens and Rubin (2015) for a review). This is known as the *stable unit treatment value assumption* (SUTVA) (Rubin, 1980). SUTVA is violated under settings in which there is *treatment interference*—that is, when a treatment assigned to a unit affects the response of other units. Effects on response due to treatment interference are also known as spillover, peer influence, social interaction, or network effects.

The dependence of a unit's outcome on other units' exposures or treatments poses statistical challenges because the *potential outcome of a unit*—the hypothetical outcome of a unit given a realized treatment assignment—is not only affected by its own treatment status but also by the treatment conditions received by other units. In some settings, interference can be considered as a nuisance and researchers may control for it to reduce the bias by designing experiments in such a way that treatment effects do not interfere. Although these designs may minimize the effect of interference, such designs are not always possible. On the other hand, in other settings, estimating the causal effect in the presence of interference is of interest itself. Examples of this include studies on the efficacy of vaccines in which vaccinated and non-vaccinated members of a population interact with each other and researchers are interested in overall infection rates. Under these latter settings, considerable work has been devoted to the development of reasonable models of interference in order to ensure identification of both the direct effect of treatment and the effect of treatment spillover on the response (Aronow and Samii, 2017; Forastiere et al., 2020; Manski, 2013; Sussman and Airoldi, 2017; Toulis and Kao, 2013).

In this chapter, we introduce a model of treatment interference called the *K-nearest neighbors interference model* (KNNIM). Under KNNIM, the response of a unit is affected only by the treatment given to that unit and the treatment statuses of its K nearest neighbors (KNN). Such models of interference may be reasonable, for example, under social network settings, where only a few of the observable potential interactions (e.g. accounts that a Twitter user follows) may be influential on a unit's response, and the strength of interaction may be measured by the amount of engagement between users.

We then perform a simulation study to determine how existing methods, and one newly developed method, for detecting treatment interference perform under data generated under a KNNIM model. While these methods were originally developed to detect arbitrary interference (Aronow, 2012; Athey et al., 2018; Pouget-Abadie et al., 2019; Saveski et al., 2017), it is reasonable to assume that the efficacy of these methods may vary depending on the structure of interference. However, little work has been done to assess how these methods perform under various interference models. We repeatedly simulate data under a KNNIM model and apply these methods to the simulated data. We then assess the power of these methods to successfully detect treatment interference when it is present and their likelihood of concluding insignificant interference when it is omitted. Results suggest that methods which incorporate structured selection of focal units (Aronow, 2012; Athey et al., 2018) tend to perform reasonably well on this type of data.

The rest of this chapter is organized as follows. An overview on causal inference under interference is presented in Section 2.2. The K−nearest neighbors interference model is introduced in Section 2.3. We discuss the application of conditional randomization inference on non-sharp hypothesis in Section 2.4. An algorithm on the selection of the focal units is provided in Section 2.5. Section 2.6 gives a summary of current methods of detecting interference. Our proposed test statistic is given in Section 2.7. Section 2.8 evaluates current methods as well as our test under KNNIM model through a simulation. Results are discussed in Section 2.9. We conclude in Section 2.10.

2.1.1 Motivating Example

We motivate our approach using data from a randomized field experiment aimed to reduce conflict among middle school students in 56 schools in New Jersey(Paluck et al., 2016). This study assesses the impact of an anti-conflict program on individual students and determines whether benefits of the program are transmitted through social interactions between students.

The anti-conflict program was implemented via a two-stage experimental design. In the first stage, of the 56 schools in the study, 28 schools were randomly assigned to participate in the anti-conflict program. Then, within each school, between 40 and 64 students were determined to be eligible to be "seed" students—students that actively participate and advocate for the anti-conflict program. In the second stage of the randomization, of all the eligible seed students, half were assigned to be seeds. Seed students were encouraged to publicly reflect their opposition to conflict in their school—for example, identifying a common conflict in their school and creating a hashtag about it—and were also asked to distribute orange wristbands with the intervention logo to students that demonstrate anti-conflict attitudes. Analysis was performed only on students that were eligible to be seeds ($N = 2{,}451$).

Of particular note, to assess potential pathways for treatment interference, students were asked to identify, in order, the 10 other students that they spent the most time with during the previous few weeks. This yields a unique dataset in which the strength of the interaction between two individuals under study is explicitly recorded. Hence, statistical analyses may benefit from an interference model, such as KNNIM, that allows for direct incorporation of the relative strengths of the interactions. For this dataset, KNNIM models with K up to 10 may be applicable.

2.2 Background and Related Work

The Neyman-Rubin Causal Model (NRCM) is a popular model of response in causal inference (Holland, 1986; Imbens and Rubin, 2015; Rubin, 1980; Splawa-Neyman et al., 1990). Consider a simple experiment on N units, numbered $1, \ldots, N$, where all units are given either a treatment or a control condition. The NRCM assumes that the response of unit i, denoted Y_i follows the model

$$Y_i = y_i(1)W_i + y_i(0)(1 - W_i).$$

Here, $y_i(W_i)$ is the potential outcome under treatment status $W_i \in \{0, 1\}$—the hypothetical response of unit i had that unit received treatment status W_i—and W_i is a treatment indicator: $W_i = 1$ if unit i receives treatment and $W_i = 0$ if unit i receives control. Inherent in this model is the no interference assumption or SUTVA. This assumption states that there is only a single version of each treatment status and that a unit's outcome is only affected by its own treatment status and is not affected by the treatment status of any other unit (Cox, 1958; Rubin, 1980).

In many settings, SUTVA is not plausible, and considerable work has been performed on analyzing causal effects when SUTVA is violated. Sobel (2006) showed that violating SUTVA can lead to wrong conclusions about the effectiveness of the treatment of interest. Forastiere et al. (2020) derive bias formulas for the treatment effect when SUTVA is wrongly assumed and show that the bias that is due to the presence of interference is proportional to the level of interference and the relationship between the individual and the neighborhood treatments.

When interference is present, the effect of a treatment on a unit may occur through direct application of the treatment to that unit, indirectly through application of treatment to units that interact with the original unit, or both (Hudgens and Halloran, 2008). We can extend the potential outcomes framework to account for both direct and indirect treatment components. Let $y_i(\mathbf{W}) = y_i(W_i, \mathbf{W}_{-i})$ denote the potential outcome of unit i under treatment allocation $\mathbf{W} \in \{0,1\}^N$, where unit i is given treatment W_i, and the remaining treatment statuses are allocated according to $\mathbf{W}_{-i}$. Responses Y_i satisfy

$$Y_i = \sum_{\mathbf{W} \in \{0,1\}^N} y_i(\mathbf{W}) \mathbf{1}(\mathbf{W}^* = \mathbf{W}),$$

where $\mathbf{1}(\mathbf{W}^* = \mathbf{W})$ is an indicator variable that is equal to 1 if and only if the observed treatment status $\mathbf{W}^* = \mathbf{W}$. The *average direct effect* τ_{dir} is the average difference in a unit's potential outcomes when changing that unit's treatment status and holding all other units' treatment status fixed. It may be defined as

$$\tau_{dir} = \frac{1}{N} \sum_{i=1}^{N} (y_i(1, \mathbf{1}) - y_i(0, \mathbf{1})),$$

where $\mathbf{1}$ denotes a vector of all 1's. In contrast to direct effect, the *average indirect effect* τ_{ind} is defined as the average difference in a unit's potential outcome when changing all other treatment statuses from control to treated, holding its own treatment fixed. It can be defined as

$$\tau_{ind} = \frac{1}{N}\sum_{i=1}^{N}(y_i(0, \mathbf{1}) - y_i(0, \mathbf{0})),$$

where $\mathbf{0}$ denotes a vector of all 0's. The *average total effect* τ_{tot} measures the average difference in potential outcomes between all units receiving treatment and all units receiving control:

$$\tau_{tot} = \frac{1}{N}\sum_{i=1}^{N}(y_i(1, \mathbf{1}) - y_i(0, \mathbf{0})).$$

Note that these quantities are defined to satisfy $\tau_{tot} = \tau_{dir} + \tau_{ind}$. Addionally, when SUTVA holds, $\tau_{tot} = \tau_{dir}$ and $\tau_{ind} = 0$.

There are a variety of strategies for designing and analyzing experiments under treatment interference. One approach is to view interference as a nuisance parameter and to reduce the effect of treatment interference on causal estimates through effective experimental design. This line of work aims to use available information on potential interaction of units to design an experiment that mitigates the effect of this interaction. Often, this is done through forming clusters with high within-cluster interaction and randomizing treatment across clusters rather than individual units (Eckles et al., 2016; Gui et al., 2015; Ugander et al., 2013). However, knowledge of the interaction network may not necessary to make progress on this problem—Sävje et al. (2021) investigate methods for consistent estimation of treatment effects when the structure of interference is unknown. This approach may not be ideal when indirect effects are of interest to the researcher.

Rather than considering interference as a nuisance, some researchers tend to relax SUTVA and allow for different models of interference, considering interference effect as of primary interest. One significant example of this involves experiments in the efficacy of vaccines where the likelihood of a person contracting an infectious disease depends on others in the same population who are vaccinated (Halloran and Struchiner, 1995; Hudgens and Halloran, 2008; Ross, 1916). Under this setting, interference is allowed within groups but not across groups—this is referred to as a partial interference assumption (Sobel, 2006), i.e., SUTVA is assumed between groups (Basse and Feller, 2018; Hudgens and Halloran, 2008; Offer-Westort and Dimmery, 2021; Rosenbaum, 2007; Sobel, 2006; Tchetgen and VanderWeele, 2012).

A similar approach to partial interference assumes that treatment interference on a unit can only occur within a small closed neighborhood of that unit (Sussman and Airoldi, 2017)—the K-nearest-neighbors interference model (KNNIM) introduced in this thesis is a variant of this setting. Another common approach is to assume that the treatment condition can only "spill over" and affect the response of a control unit if a certain number or fraction of potential interactors of that unit receive treatment (Gui et al., 2015; Toulis and Kao, 2013). Finally, in its least restrictive form, Aronow and Samii (2017) consider the use of Horvitz-Thompson estimators for estimating treatment effects under arbitrary forms of interference.

Another research direction focuses on the development of hypothesis tests to detect the presence of treatment interference in an experiment. Aronow (2012) introduces a framework for conditional randomization tests for detecting treatment interference. Athey et al. (2018) extend this approach to develop tests for more general forms of treatment interference. Basse et al. (2019) build on this work and consider the validity of the test by conditioning on observed treatment assignment of the subset of units who received an exposure of interest. Saveski et al. (2017) and Pouget-Abadie et al. (2019) develop an experimental framework to simultaneously estimate treatment effects and test whether treatment interference is present within an experiment.

2.3 K-Nearest Neighbors Interference Model

Treatment interference models range between assuming no exposure to other units' treatments—which is the case for traditional randomized experiments with the NRCM under SUTVA—and structured interference models. For models allowing arbitrary interference, each unit will have a unique type of exposure depending on the treatment assignment for all N individuals. This results in distinct 2^N potential outcomes for each unit and $N2^N$ potential outcomes for the experimental population where we only observe N of these potential outcomes. Under the latter exposure, there would be no meaningful way to analyze the experiment so that researchers focus on structured or limited interference.

To make progress on treatment interference problems, researchers make assumptions between those of SUTVA and arbitrary interference models that restrict the extent of interference allowed (Aronow and Samii, 2017; Sussman and Airoldi, 2017; Toulis and Kao, 2013; Ugander et al., 2013). However, most models in previous work only specify that the units' outcomes are affected by the number or fraction of treated neighbors, but do not specify which neighbors impact unit response and how they affect the response.

We now propose an interference model where we restrict the interference of treatment on a unit i to its K–nearest neighbors. This allows different neighbors to contribute different effects depending on the proximity of the relationship—neighbors that are close to unit i are more likely to have their treatment status affect the response of unit i. This model restrict the number of potential outcomes to be 2^{K+1} for each unit.

We view units under study as a mathematical graph. Let $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ be a directed graph of $\|\mathbf{V}\| = N$ vertices; each vertex $i \in \mathbf{V}$ corresponds to a unit under study. An edge $ij \in \mathbf{E}$ denotes potential interaction between units i and j. These interactions between units can be defined by any type of relationship, for example, membership to the same group, friendship in social media, geographic proximity, etc. (Forastiere et al., 2020). For ease of exposition, we assume that the degree of each vertex i is at least K.

Let $\mathbf{A}$ denote the $\mathbb{N} \times \mathbb{N}$ adjacency matrix of $\mathbf{G}$. That is, $A_{ij} = 1$ if $ij \in \mathbf{E}$ where there is an edge from unit j to i and $A_{ij} = 0$ otherwise. Since $\mathbf{G}$ has no self-loops, the diagonal elements of the adjacency matrix, $A_{ii} = 0$.

Let $d(i,j)$ denote an interaction or dissimilarity measure between units i and j. In the context of treatment interference, smaller values of $d(i,j)$ indicate stronger interactions between units i and j. We assume, for now, that $d(i,j)$ is only computed for units i,j with $A_{ij} = 1$. Let $d(i,(j))$ denote the jth smallest value of $\{d(i,j), j \neq i\}$; that is, $d(i,(1)) < d(i,(2)) < \cdots$. For ease of exposition, we assume that all values of $d(i,j)$ are unique (in practice, ties may be broken arbitrarily). The K*-neighborhood* of unit i, denoted $\mathcal{N}_{iK}$, is the set of the K "closest" units to unit i:

$$\mathcal{N}_{iK} = \{j : d(i,(j)) \leq d(i,(K)), j = 1, 2, \ldots, K\}.$$

Define $\mathcal{N}_{-iK} = \mathbf{V} \setminus (i \cup \mathcal{N}_{iK})$ as all units in $\mathbf{V}$ that are outside of i's K-neighborhood. Note that the sets $\{i, \mathcal{N}_{iK}, \mathcal{N}_{-iK}\}$ form a partition of $\mathbf{V}$.

Recall that W_i is a treatment indicator for unit i, and let $\mathbf{W} = (W_1, W_2, \ldots, W_N) = \{W_i, \mathbf{W}_{\mathcal{N}_{iK}}, \mathbf{W}_{\mathcal{N}_{-iK}}\}$ denote treatment assignment vector for all units N. Also recall that Y_i is the outcome measured on the unit i. Each unit's potential outcome $y_i(\mathbf{W})$ is defined as a function of the entire assignment vector of units to the two treatment conditions $\mathbf{W} \in \{0,1\}^N$.

Now we give the following assumption that defines the K-nearest neighbors interference model:

Assumption 2.1. *(K-Neighborhood Interference Assumption (K-NIA)). For each unit i in a network G and for all treatment assignments $\mathbf{W}_{\mathcal{N}_{-iK}}$, $\mathbf{W}'_{\mathcal{N}_{-iK}}$, the potential outcomes satisfy K-Neighborhood Interference Assumption if*

$$y_i(W_i, \mathbf{W}_{\mathcal{N}_{iK}}, \mathbf{W}_{\mathcal{N}_{-iK}}) = y_i(W_i, \mathbf{W}_{\mathcal{N}_{iK}}, \mathbf{W}'_{\mathcal{N}_{-iK}}).$$

Assumption 2.1 states that the potential outcome of unit i is only affected by its treatment and by the treatments assigned to its K-nearest neighbors. Changing treatments for other units outside the K-neighborhood will not affect the potential outcome of unit i. This is a special case of the neighborhood interference assumption (NIA) described in Sussman and Airoldi (2017). In its most general form, the K-nearest neighbors interference model (KNNIM) assumes only that the treatment interference structure satisfies Assumption 2.1. For convenience, we will suppress the treatment statuses in $\mathbf{W}_{\mathcal{N}_{-iK}}$ when referring to the potential outcomes y_i.

2.3.1 Choosing the neighborhood size K

In applications, the experimenter can choose the number of the K-nearest neighbors based on the field and the purpose of the study. The experimenter also can decide on the size of K using prior knowledge from previous studies that serves the objective of the study. This should be done in early phases of the study, and hence, the network is fixed and known in advance.

However, another factor that should be addressed when choosing the size of K is the sample size in order to be able to quantify, estimate and draw inference on the K-nearest neighbors indirect effects. As mentioned above, number of exposure to treatments is restricted to 2^{K+1} exposures. Hence, to ensure sufficient power, many methods that incorporate KNNIM will require a sufficient number of units exposed to each of these exposure levels. From our experience, a good heuristic is to require roughly 30 observations for each considered treatment exposure.

In the school conflict motivating example described in Section 2.1.1, the experimenters measured $K = 10$ nearest neighbors for each student. However, suppose that analysis is isolated to eligible students in treated schools who have at least $K = 2$ seed-eligible nearest neighbors. This sample contains $N = 348$ units, and there are eight treatment exposures possible for each student. In Table 2.1, we see that each possible exposure has at least 34 students assigned to that exposure.

However, suppose we restrict our analysis further to only eligible students in treated schools who have at least $K = 3$ seed-eligible nearest neighbors ($N = 100$). In Table 2.2, we see there is only one unit given the exposure where the individual and all its three seed-eligible nearest neighbors are all treated. Increasing the size of K when sample sizes are small may lead to hypothetical exposures with few (if any) units assigned to that exposure, which may complicate analyses and reduce power substantially.

Table 2.1: *Number of units in each exposure of Anti-Conflict Program Experiment with K =2 and N = 348*

	Indirect			
Direct	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
Treated	38	42	39	34
Control	40	59	46	50

Table 2.2: *Number of units in each exposure of Anti-Conflict Program Experiment with K =3 and N = 100*

	Indirect							
Direct	(000)	(001)	(010)	(100)	(011)	(101)	(110)	(111)
Treated	5	6	3	6	8	7	11	1
Control	6	8	3	4	11	4	10	7

2.4 Randomization Inference for Detecting Interference

We now describe the framework for randomization inference for testing the presence of treatment interference under KNNIM. Recall that $\mathbf{W}$ is the treatment assignment vector and $y_i(\mathbf{W})$ is the potential outcome of unit i under treatment $\mathbf{W}$. Let $T = T(\mathbf{W}, y(\mathbf{W}))$ denote a test statistic— a random variable where the source of randomness follows from the dependence on the random treatment assignment $\mathbf{W}$. Let $\mathbf{W}^{obs}$ and $\mathbf{Y}^{obs} = \mathbf{Y}(\mathbf{W}^{obs})$ denote the observed treatment assignment vector and the observed outcome vector, respectively. Let $T^{obs} = T(\mathbf{W}^{obs}, \mathbf{Y}^{obs})$ denote the observed value of the test statistic. We aim to test the null hypothesis of no treatment interference for each unit

$$H_0 : y_i(W_i, \mathbf{W}_{\mathcal{N}_{iK}}) = y_i(W_i, \mathbf{W}'_{\mathcal{N}_{iK}}). \tag{2.1}$$

Typically, randomization tests under the potential outcome framework assume a sharp null hypothesis of no unit-level treatment effects, and potential outcomes are able to be inferred under this sharp null across randomizations (Fisher, 1925). However, since the hypothesis (2.1) does not make assumptions about direct effect of treatment on each unit, the potential outcome $y_i(W_i, \mathbf{W}_{\mathcal{N}_{iK}})$ may not be imputable for randomizations under which $W_i \neq W_i^{obs}$. Progress can be made by conditioning on a set of randomizations $\mathbf{\Omega}$ and choosing a test statistic T such that T is *imputable* under randomizations in $\mathbf{\Omega}$ (Basse et al., 2019). Afterward, a conditional p-value is obtained by computing, for example, the fraction of randomizations $\mathbf{W}' \in \mathbf{\Omega}$ such that

$$|T(\mathbf{W}', y(\mathbf{W}'))| \geq |T(\mathbf{W}^{obs}, \mathbf{Y}^{obs})|.$$

Following Aronow (2012) and Athey et al. (2018), this conditional randomization inference can be performed by first selecting a subset of units under study called *focal units* and to only consider randomizations of treatment $\mathbf{W}$ that do not affect the treatment status of the focal units. Only *variant units*—those that are not focal units—can have differing treatment statuses across randomizations. In other words, we simulate draws from the random treatment assignment vectors conditional on the fixed treatment of the focal units. Thus, the null hypothesis of no interference is sharp on the focal units since only treatment statuses of variant units—only those units that can impose indirect effects—are randomized. The test statistic T is only computed on the outcomes of the focal units and hence, the test statistic is imputable under alternative treatment assignment vectors.

2.5 Selection of the Focal Units

Although the choice of the focal units does not affect the validity of the test, it plays a key role for the power of the test (Athey et al., 2018). Aronow (2012) provides a restriction on the size of the focal set when the selection on the focal units is random with a binary treatment conditions. Athey et al. (2018) provide different methods of choosing the focal units and found that systematic ways perform better than random selection of the focal units. Basse et al. (2019) build on this work and provide a conditioning framework that consider the validity of the test by conditioning on observed treatment assignment of the focal units who received an exposure that contributes more to the hypothesis of interest.

Under KNNIM, we suggest choosing focal units in a way such that the K-neighborhoods of the focal units do not overlap. This will enable us to remove dependencies between outcomes of focal units induced by indirect effects. Additionally, a substantial fraction of focal units may still be selected under this condition, increasing the power of the the randomization inference.

Such a selection can be performed as follows. Define the K-nearest neighbors adjacency graph $\mathbf{G}_{KNN} = (\mathbf{V}, \mathbf{E}_{KNN})$ to be the undirected graph with an edge $ij \in \mathbf{E}_{KNN}$ if and only if $j \in \mathcal{N}_{iK}$ or $i \in \mathcal{N}_{jK}$, and second power of $\mathbf{G}_{KNN}$, denoted $\mathbf{G^2}_{KNN} = (\mathbf{V}, \mathbf{E^2}_{KNN})$, as the graph that contains an edge $ik \in \mathbf{E^2}_{KNN}$ if and only if there is a path of two edges or fewer in $\mathbf{E}_{KNN}$ joining i and k. Focal units can then be selected by choosing a maximal independent set of units $\mathbf{F}$ within $\mathbf{G^2}_{KNN}$; that is, all units in $\mathbf{F}$ are independent of each other in $\mathbf{G^2}_{KNN}$, and the addition of any other unit $i \in \mathbf{V} \setminus \mathbf{F}$ into $\mathbf{F}$ will add dependencies (Higgins et al., 2016).

Selection of a maximal independent set of focal units in $\mathbf{G^2}_{KNN}$ may be performed as follows.

Algorithm. *Given a K-nearest neighbors adjacency graph $\mathbf{G}_{KNN} = (\mathbf{V}, \mathbf{E}_{KNN})$, the following algorithm will select a maximal independent set of focal units within $\mathbf{G^2}_{KNN}$.*

1. **Step 1:** (Initialize) Let $\mathbf{U} = \mathbf{V}$. Initialize the set of focal units $\mathbf{F} = \emptyset$. Initialize the set of variant units $\mathbf{I} = \emptyset$.

2. **Step 2:** (Select focal unit) While $|\mathbf{U}| > 0$, choose one vertex $i \in \mathbf{U}$ at random. Set i as a focal unit: $i \in \mathbf{F}$.

3. **Step 3:** (Find nearest neighbors) Set $\mathbf{I}$ equal to all units j such that $ij \in \mathbf{E}_{KNN}$.

4. **Step 4:** (Find neighbors of neighbors) Find all units $k \in \mathbf{V} \setminus \mathbf{I}$ such that, for some unit $j \in \mathbf{I}$, $jk \in \mathbf{G^2}_{KNN}$. Set these units $k \in \mathbf{I}$.

5. **Step 5:** (Remove units) Remove all vertices in $\mathbf{F}$ and $\mathbf{I}$ from $\mathbf{U}$.

6. **Step 6:** (Repeat or terminate) If $|\mathbf{U}| = 0$, stop. The set of focal units $\mathbf{F}$ is a maximally independent set of units within $\mathbf{G^2}_{KNN}$. Otherwise, set $\mathbf{I} = \emptyset$ and return to Step 2.

2.6 Current Methods for Detecting Interference

Current methods for detecting interference include conditional randomization tests (Aronow, 2012; Athey et al., 2018) (as outlined in Section 2.4) and carefully designed experiments performed with the intention to detect interference (Pouget-Abadie et al., 2019; Saveski et al., 2017). We now provide a summary of these methods for testing for interference. For randomization tests, we focus on the choice of test statistic used. For experimental design methods, we describe both experimental setup and the test statistic.

2.6.1 Test Statistics for Randomization Tests

Aronow (2012) introduced the randomization inference approach for testing for interference between units, where units are affected by their own treatment and by the treatment assigned to their immediate neighbors. In this test, the treatment status for a subset of focal units remains fixed; the rest of the units are the variant subset. The randomization inference is conditional on the observed treatment status of the fixed subset. That is, this test is on indirect effects resulting from the variation of the treatment status for the variant subset of units. A variety of test statistics may be used under this framework. Differences in the statistic across randomizations must be resulting from the variation of the treatment status of the variant units.

The Pearson correlation coefficient ρ between the outcomes of the fixed units ($\mathbf{Y_F}$) and the "distance" to the nearest unit of a particular treatment status in the variant subset ($\mathbf{D}_{nearest}$) may be used as the test statistic:

$$\rho = cor(\mathbf{Y_F}, \mathbf{D_{nearest}}). \tag{2.2}$$

A common choice of distance is the Euclidean distance between pretreatment covariates.

This distance can be incorporated into the KNNIM framework through the dissimilarity measure d. Aronow (2012) advocates for computing Pearson correlation coefficient on the ranks of these quantities; however, preliminary simulations suggest that the statistic ρ tends to be more powerful for the models considered in Section 2.8.

Athey et al. (2018) extend this work and develop tests for more general realizations of interference (e.g. no higher-order interference). As part of this work, they suggest additional test statistics for detecting interference. The edge-level contrast statistic T_{elc}—a modification of a test statistic proposed by Bond et al. (2012)—is the difference between the average outcomes of the focal units with treated neighbors and the focal units with control neighbors. Here, T_{elc} averages over edges ij where i is a focal unit and j is not a focal unit:

$$T_{elc} = \frac{\sum_{i,j\neq i} F_i A_{ij}(1-F_j)W_j Y_i^{obs}}{\sum_{i,j\neq i} F_i A_{ij}(1-F_j)W_j} - \frac{\sum_{i,j\neq i} F_i A_{ij}(1-F_j)(1-W_j)Y_i^{obs}}{\sum_{i,j\neq i} F_i A_{ij}(1-F_j)(1-W_j)},$$

where F_i is an indicator variable satisfying $F_i = 1$ if and only if $i \in \mathbf{F}$.

A second test statistic is the score test statistic T_{score} (Athey et al., 2018). This statistic is motivated by a model of treatment interference in which the indirect effect is proportional to the fraction of treated neighbors (Manski, 1993, 2013). The score test begins by computing

$$r_i = Y_i^{obs} - \bar{Y}_{F,0}^{obs} - (\bar{Y}_{F,1}^{obs} - \bar{Y}_{F,0}^{obs})W_i,$$

for each focal unit $i \in \mathbf{F}$, where $\bar{Y}_{F,1}^{obs}$ and $\bar{Y}_{F,0}^{obs}$ are the average outcome for the treated and control focal units respectively. Then, T_{score} is the covariance between these r_i terms and $\sum_{j=1}^{N} \overline{A_{ij}} W_j$—the fraction of treated neighbors for unit i. This statistic is computed across only focal units that have at least one treated neighbor:

$$T_{score} = cov\left(r_i, \sum_{j=1}^{N} \overline{A_{ij}} W_j \middle| F_i = 1, \sum_{j=1}^{N} A_{ij} > 0 \right).$$

Finally, Athey et al. (2018) consider the has-treated-neighbor test statistic T_{htn}, a modification of Pearson correlation coefficient (2.2). Instead of using the distance to the nearest treated neighbor, this statistic uses a indicator variable E_i for whether any of a unit's neighbors in the variant subset are treated: that is, $E_i = 1$ if and only if $\sum_j A_{ij} W_j (1 - F_j) > 0$. Then T_{htn} is the correlation between this indicator and the outcomes for the focal units $\mathbf{F}$.

$$T_{htn} = \frac{1}{S_{Y_F^{obs}}.S_E} \frac{1}{|\mathbf{F}|} \sum_{i \in \mathbf{F}} \left(Y_i^{obs} - \bar{Y}_F^{obs} \right) E_i,$$

where $\bar{Y}_F^{obs}$ and $S_{Y_F^{obs}}$ are the sample mean and standard deviation of the outcomes for focal units respectively and S_E is the sample standard deviation of the E_i variables.

2.6.2 Experimental Design Approach

Saveski et al. (2017) and Pouget-Abadie et al. (2019) present a two-stage experimental design to test for the presence of interference. In this design, the units under study are divided into two groups and two experiments are performed simultaneously: for one group, treatment is assigned completely at random, and for another group, units are clustered and treatment is assigned across clusters rather than units. Then, estimates of the average direct effect are computed under the assumption of no interference for both the completely randomized and cluster randomized designs. Finally, a standardized difference T_{exp} is computed between these estimates:

$$T_{exp} = \frac{|\hat{\tau}_{cr} - \hat{\tau}_{cbr}|}{\hat{\sigma}_p}, \tag{2.3}$$

where $\hat{\tau}_{cr}$ and $\hat{\tau}_{cbr}$ are the estimates of the direct effect under the completely randomized and cluster randomized designs respectively and $\hat{\sigma}_p$ is a pooled standard deviation of responses from both the completely randomized and cluster randomized designs (Saveski et al., 2017). Large values of T_{exp} imply the presence of indirect effects.

A conservative test of the null hypothesis of no treatment interference can be performed at the α significance level by rejecting the null hypothesis if and only if $T_{exp} \geq \alpha^{-1/2}$. Additionally, as the number of units $n \to \infty$, it can be shown that T_{exp} converges to a standard normal distribution (provided that cluster sizes remain fixed). Thus, an approximate size

α test can be conducted by rejecting the null hypothesis of no interference if $T_{exp} \geq z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution.

2.7 K-Nearest Neighbors Indirect Effect Test Statistic

We now propose an additional test statistic designed to detect K-nearest neighbors indirect effects. Let $\overline{Y}^{obs}(W_i, \mathbf{W}_{\ell=W_j})$ denote the average response of observed units that are assigned to treatment status W_i and have their ℓth nearest neighbor assigned to treatment status W_j. The *K-nearest neighbors indirect effect test statistic* T_{knn} is obtained by computing differences in potential outcomes between focal units that receive the same treatment status but differ on the status of their ℓth nearest neighbor, and summing these differences across each of the K nearest neighbors.

That is, for $W_i \in \{0, 1\}$ and $\ell \in \{1, \ldots, K\}$, define

$$T_{knn,\ell}(W_i) = \bar{Y}^{obs}(W_i, \mathbf{W}_{\ell=1}) - \bar{Y}^{obs}(W_i, \mathbf{W}_{\ell=0}),$$

and define $T_{knn,\ell}$ as a weighted average of these terms:

$$T_{knn,\ell} = \frac{N_{Ft}}{|\mathbf{F}|} T_{knn,\ell}(1) + \frac{N_{Fc}}{|\mathbf{F}|} T_{knn,\ell}(0),$$

where N_{Ft} and N_{Fc} are the number of treated focal units and control focal units respectively. We then can define T_{knn} as a sum of these $T_{knn,\ell}$ statistics:

$$T_{knn} = \sum_{\ell=1}^{K} T_{knn,\ell}.$$

Note that, under the null hypothesis of no treatment interference, each of the $T_{knn,\ell}(W_i)$ terms should be close to 0. Thus, since T_{knn} is a linear combination of these terms, values of T_{knn} that are relatively large in magnitude provide evidence against this null hypothesis, and so, $|T_{knn}|$ may be effective as a test statistic. Additionally, note that the statistic $T_{knn,\ell}$

may be used directly for a test of interference stemming from treatments assigned to the ℓth-nearest neighbor.

2.8 Simulation

In this Section, we conduct a comparison and evaluate the performance of the methods covered in Section 2.6 and 2.7 for testing the null hypothesis of no interference under the K-nearest neighbors interference model.

2.8.1 Data Generation Procedure

We generate the responses under the following model which satisfies KNNIM:

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

In this model, we assume that the closest three neighbors affect the response Y_i (i.e. $K = 3$); we use $W_{i\ell}$ to denote the treatment status of the ℓth nearest neighbor of unit i. The covariates X_j, $j = 1, 2, 3$, are independent and identically distributed $Normal(0, 1)$ random variables. We use the Euclidean distance between the covariates $\mathbf{X}_i$ and $\mathbf{X}_j$ as the dissimilarity measure $d(i, j)$—units with more similar values of covariates are more likely to interact with each other. In the initial generation of data, treatment is completely randomized across all N units, with half of the units receiving treatment and the other half receiving control. Different models are generated through varying the $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_d)$ coefficients and the sample size N. We consider sample sizes of $N = 256$ and $N = 1024$.

For each choice of sample size, we consider thirteen different models of interference. We describe these models in Table 2.3 in terms of the coefficients vector $\boldsymbol{\beta}$. The first 3 elements of $\boldsymbol{\beta}$ represent the indirect effect contributed by first, second, and third-nearest-neighbor respectively. The last element β_d is the unit's direct effect. In all models considered, the closer the relationship to unit i, the greater the indirect effect: $|\beta_1| \geq |\beta_2 \geq |\beta_3|$. The indirect effects in every set of three models represent the degree of interference starting from no interference in the first 3 models, followed by weak interference in the second three models, moderate interference in the next three models, and finally strong interference in the last four models.

For datasets with $N = 256$ observations, 1,000 realizations of potential outcomes following each model are generated. Tests of indirect effects are then applied to each of the 1,000 realizations. Results are given in Section 2.9. Due to computational limitations, only 100 realizations are generated for models containing $N = 1024$ units.

2.8.2 Simulation for Randomization Tests

We compare the performance of both conditional randomization tests and experimental design approaches for detecting interference. For the conditional randomization tests, for each set of generated potential outcomes, treatment is initially assigned completely at random to units, with half of the units receiving treatment and the other half receiving control. Then,focal units are selected according to Algorithm 2.5. We then proceed with randomization tests as described in Sections 2.4 and 2.6.1. We evaluate the performance of the following test statistics: the Pearson correlation coefficient (Pearson) (Aronow, 2012), the edge level contrast statistic (ELC), the score statistic (Score), the has-treated-neighbor statistic (HTN) (Athey et al., 2018), and the K-nearest neighbors indirect effect test statistic (KNN).

Test statistics are computed across 1,000 randomizations for each realization of the potential outcomes; for each randomization, treatment statuses are fixed for focal units and are completely randomized across variant units. For each generated dataset and each choice of test statistic, we obtain a p-value for the null hypothesis of no treatment interference. Thus, for $N = 256$, we obtain a distribution of 1,000 p-values for each test statistic under each model. The power of the tests can also be estimated by computing the fraction of p-values that fall beneath a pre-specified significance level α.

2.8.3 Simulation for Experimental Design Approach

In addition, we follow the experimental design in Saveski et al. (2017) (described in Section 2.6.2) to determine its efficacy for testing whether SUTVA holds under KNNIM. For each set of generated potential outcomes, we divide the units into clusters of four units using a heuristic algorithm for the clique partitioning problem with minimum clique size requirement from Ji (2004) (Algorithm 4). This clustering is performed once per dataset.

We then randomly select half of the clusters to be cluster randomized; for this group, treatment is assigned at the cluster level, with half of the clusters receiving treatment and the other half receiving control. For units belonging to the remaining clusters, each unit's cluster assignment is ignored, and treatment is completely randomized across all remaining units. Again, half of these units receive treatment and the other half receive control. For each dataset, the random selection of clusters and the treatment randomization is performed 1,000 times.

For each randomization, the statistic T_{exp} in (2.3) is computed. We then perform a test of the null hypothesis of no treatment interaction at the $\alpha = 0.05$ significance level. A conservative test rejects this null hypothesis if $T_{exp} \geq \alpha^{-1/2}$ and an asymptotic test rejects the null if $T_{exp} \geq z_{1-\alpha/2}$. Thus, for $N = 256$, we perform a total of 1,000,000 tests: that is, 1,000 tests for each of the 1,000 generated potential outcomes. By computing the fraction of rejected null hypotheses, we are able to assess the Type I Error (Models 1–3) and the power (Models 4–13) of the experimental design approach.

2.9 Discussion

Figure 2.1 provides a visual comparison of the distribution of p-values for randomization tests to detect interference under KNNIM. Table 2.4 provides the estimated Type I Error and power of these tests (conducted at significance level $\alpha = 0.05$) across the 13 considered models. As is expected by design (Higgins, 2004), the p-values of all randomization tests under models without treatment interference (Models 1–3) are approximately distributed uniformly between 0 and 1. Under weak interference (Models 4–6), the ELC, Score, and KNN tests seem to outperform the Pearson and the HTN tests; the p-values are smaller overall for these three tests. Similar trends hold under moderate interference (Models 7–9) and strong interference (Models 10–13). In particular, under strong interference, Score, KNN, and ELC tests have near 100% power to detect treatment interference.

However, the ELC, Pearson, and HTN tests seem to have some difficulty with detecting indirect effects when direct effects become large. For example, the p-values for these three tests under Models 6 and 9—models that have comparatively larger direct effects—are substantially larger than under Models 4 and 5 and Models 7 and 8 respectively. The Score and KNN tests do not suffer from this loss of power as direct effects increase. For example, for Model 6, the Score and KNN tests have an estimated power of 0.916 and 0.849 respectively where the ELC, Pearson and HTN tests have an estimated power of 0.693, 0.407, and 0.367 respectively. Thus, for the considered tests, the Score and KNN tests (in that order) seem to have the best combination of power in detecting treatment effects and isolating indirect effects in the presence of direct effects. Similar comparisons between the methods hold for datasets with $N = 1024$ and/or when focal units are selected from only one treatment condition (Table 2.5, Figure 2.3, Figure 2.4, Figure 2.5 and Figure 2.8).

Figure 2.2 gives box plots of the estimated rejection rate across all 1,000 generated potential outcomes for both the conservative and asymptotic tests using the experimental design method (Pouget-Abadie et al., 2019; Saveski et al., 2017) with $N = 256$ and significance level $\alpha = 0.05$. This plot also shows the estimated power of the considered randomization tests under these 13 models. Table 2.4 includes the median values of the rejection rates across the 1,000 generated potential outcomes for these tests. The conservative experimental approach appears to lead to a very conservative test; the true Type I Error is much smaller than $\alpha = 0.05$, and the test appears to have weak power under weak and moderate interference. Even under Models 10–13, which exhibit strong interference, the conservative test only has a median power of approximately 0.696.

The asymptotic test yields much more desirable results for our simulated data. Overall, the Type I Error seems quite close to the nominal $\alpha = 0.05$. The asymptotic test outperforms the Pearson and HTN randomization tests for almost all models of interference, and has a power close to 1 of detecting interference under Models 10–13. However, the power of the asymptotic test still is behind that of the Score, KNN, ELC tests across all models.

When we increase the sample size to $N = 1024$, the conservative approach seems to be powerful for moderate and strong interference while the asymptotic approach is powerful for all interference models. However, both approaches remain comparatively less powerful than the Score, KNN, and ELC randomization tests (Table 2.5 and Figure 2.8).

2.10 Conclusion

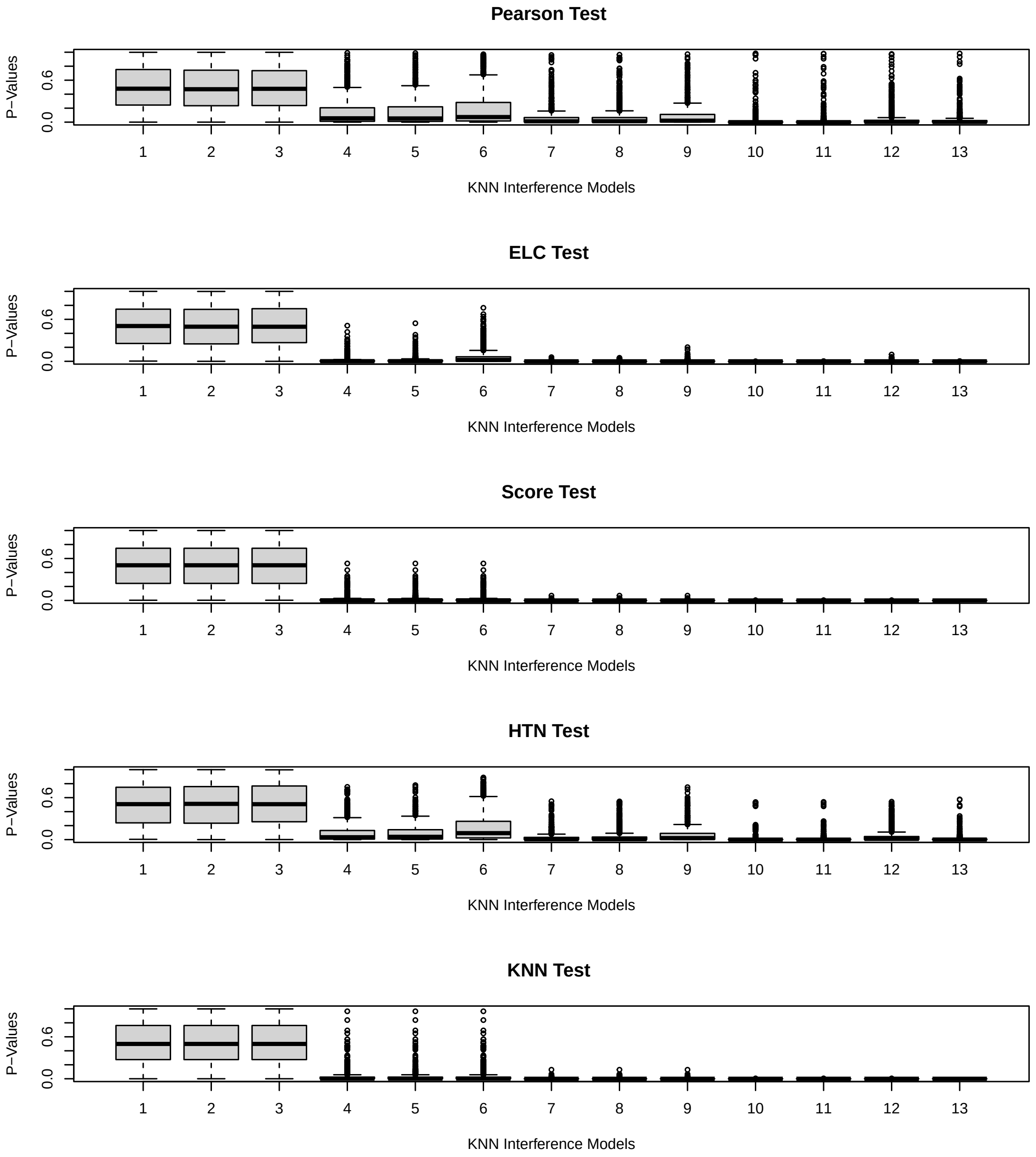
Traditional causal inference methodologies may fail to make reliable causal statements on treatment effects in the presence of interference. A substantial amount of recent work has been devoted to causal inference under interference, including methods for detecting treatment interference (Aronow, 2012; Aronow and Samii, 2017; Athey et al., 2018; Basse et al., 2019; Forastiere et al., 2020; Manski, 2013; Pouget-Abadie et al., 2019; Saveski et al., 2017; Sussman and Airoldi, 2017; Toulis and Kao, 2013).

We consider a new model of treatment interference—the K-nearest-neighbors interference model (KNNIM)—in which the treatment status of a unit i affects the response of a unit j only if i is one of j's K closest neighbors. We give advice for selecting focal units for conditional randomization tests for detecting interference under KNNIM, and suggest a new test-statistic—the *K-nearest neighbors indirect effect test statistic* (KNN)—for these randomization tests. We then perform a simulation study to compare the efficacy of both the randomization tests and experimental design approach for detecting interference under KNNIM.

Results suggest that randomization tests that incorporate our proposed selection of focal units tend to perform reasonably well on data satisfying KNNIM. Additionally, randomization tests using the score and KNN test statistics tended to be most powerful for detecting interference, especially when direct effects are permitted to grow large relative to the indirect effects.

Table 2.3: *Interference Models*

Models	(β_1, β_2, β_3, β_d)
Model 1	(0,0,0,0)
Model 2	(0,0,0,1)
Model 3	(0,0,0,4)
Model 4	(2,1,0.5,0)
Model 5	(2,1,0.5,1)
Model 6	(2,1,0.5,4)
Model 7	(3,2,1,0)
Model 8	(3,2,1,1)
Model 9	(3,2,1,4)
Model 10	(30,20,10,0)
Model 11	(30,20,10,10)
Model 12	(30,20,10,40)
Model 13	(30,30,30,30)


Pearson Test

ELC Test

Score Test

HTN Test

KNN Test

Figure 2.1: *Boxplots of p-values for the Pearson test (Pearson), edge level contrast test (ELC), score test (Score), has treated neighbor test (HTN) and K-nearest neighbors indirect effect test (KNN) under various KNNIM models. We use $N = 256$ units and $K = 3$ nearest neighbors. The p-values are estimated using 1,000 randomizations for each of the 1,000 generated potential outcome realizations.*

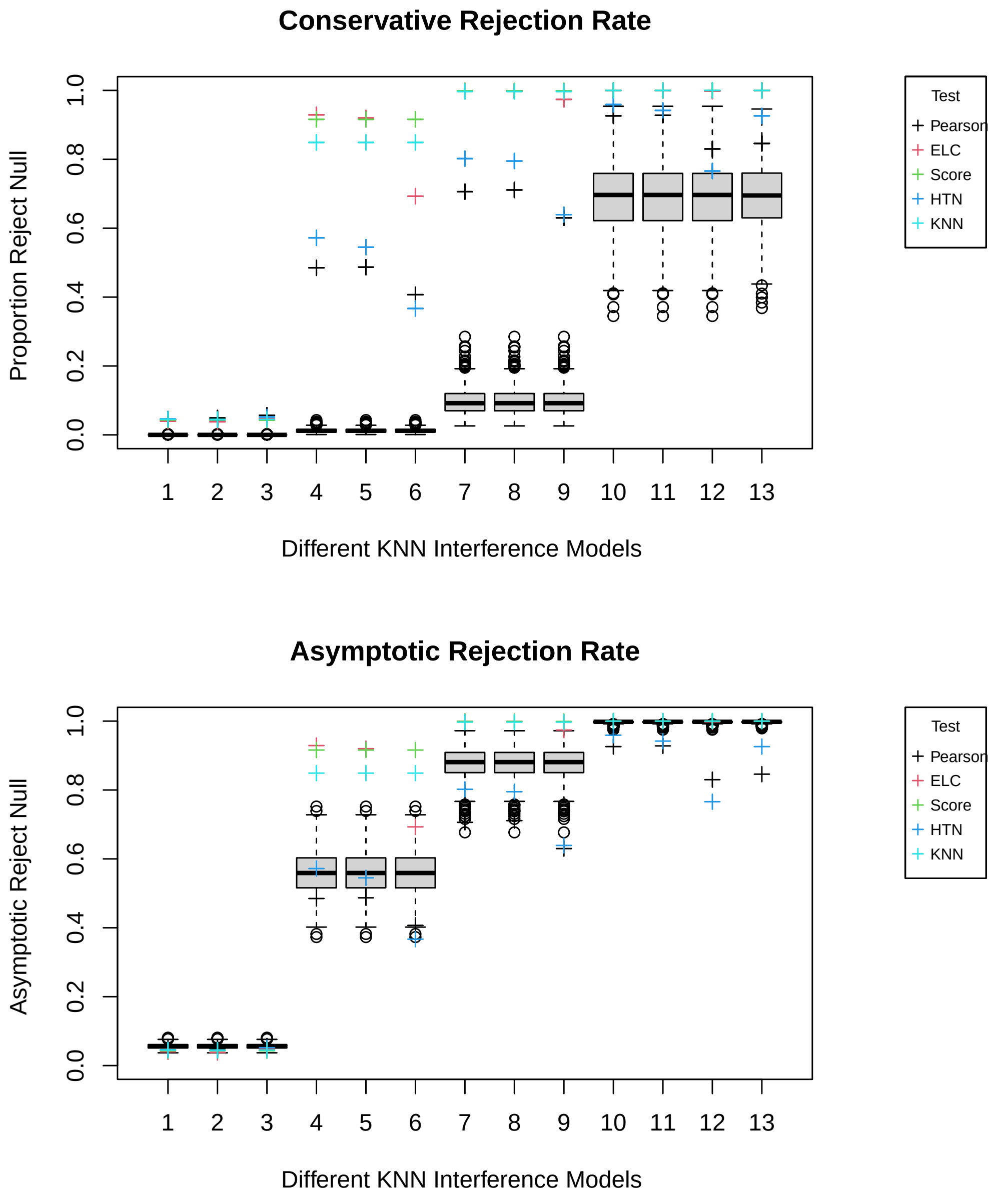


Figure 2.2: *Boxplots of the estimated rejection rates under the experimental design approach for both the conservative and asymptotic tests of the null hypothesis of no treatment interference under various KNNIM models. Plots also contain the estimated Type I Error (Models 1–3) and power (Models 4–13) for the Pearson test (Pearson), edge level contrast test (ELC), score test (Score), has treated neighbor test (HTN) and K-nearest neighbors indirect effect tests (KNN). We use $N = 256$ units and $K = 3$ nearest neighbors. The rejection rates are estimated using 1,000 treatment assignments for each of the 1,000 generated potential outcomes. Tests are performed at significance level $\alpha = 0.05$.*

Table 2.4: *Estimated Type I Errors and power for tests of treatment interference for sample size* $N = 256$

Models	Score	KNN	ELC	HTN	Pearson	Cons	Asymp
Model 1	0.043	0.045	0.040	0.046	0.045	0.000	0.056
Model 2	0.043	0.045	0.038	0.044	0.049	0.000	0.056
Model 3	0.043	0.045	0.047	0.052	0.057	0.000	0.056
Model 4	0.916	0.849	0.929	0.572	0.485	0.012	0.559
Model 5	0.916	0.849	0.920	0.545	0.487	0.012	0.559
Model 6	0.916	0.849	0.693	0.367	0.407	0.012	0.559
Model 7	0.999	0.999	0.999	0.802	0.706	0.092	0.881
Model 8	0.999	0.999	0.999	0.795	0.711	0.092	0.881
Model 9	0.999	0.999	0.974	0.639	0.630	0.092	0.881
Model 10	1.000	1.000	1.000	0.959	0.926	0.696	0.998
Model 11	1.000	1.000	1.000	0.942	0.928	0.696	0.998
Model 12	1.000	1.000	0.998	0.766	0.830	0.696	0.998
Model 13	1.000	1.000	1.000	0.926	0.846	0.695	0.998

Estimated Type I Errors (Models 1–3) and estimated power (Models 4–13) for simulated data under KNNIM. Results are provided for the score test (Score), K-nearest neighbors indirect effect test (KNN), edge level contrast test (ELC), has treated neighbor test (HTN) and the Pearson test (Pearson). Estimates of the median rejection rates under the experimental design approach for both the conservative (Cons) and asymptotic (Asymp) tests are also provided. We use $N = 256$ units and $K = 3$ nearest neighbors. These values are estimated using 1,000 generated potential outcomes with 1,000 treatment assignments performed on each set of potential outcomes. Tests are performed at significance level $\alpha = 0.05$.

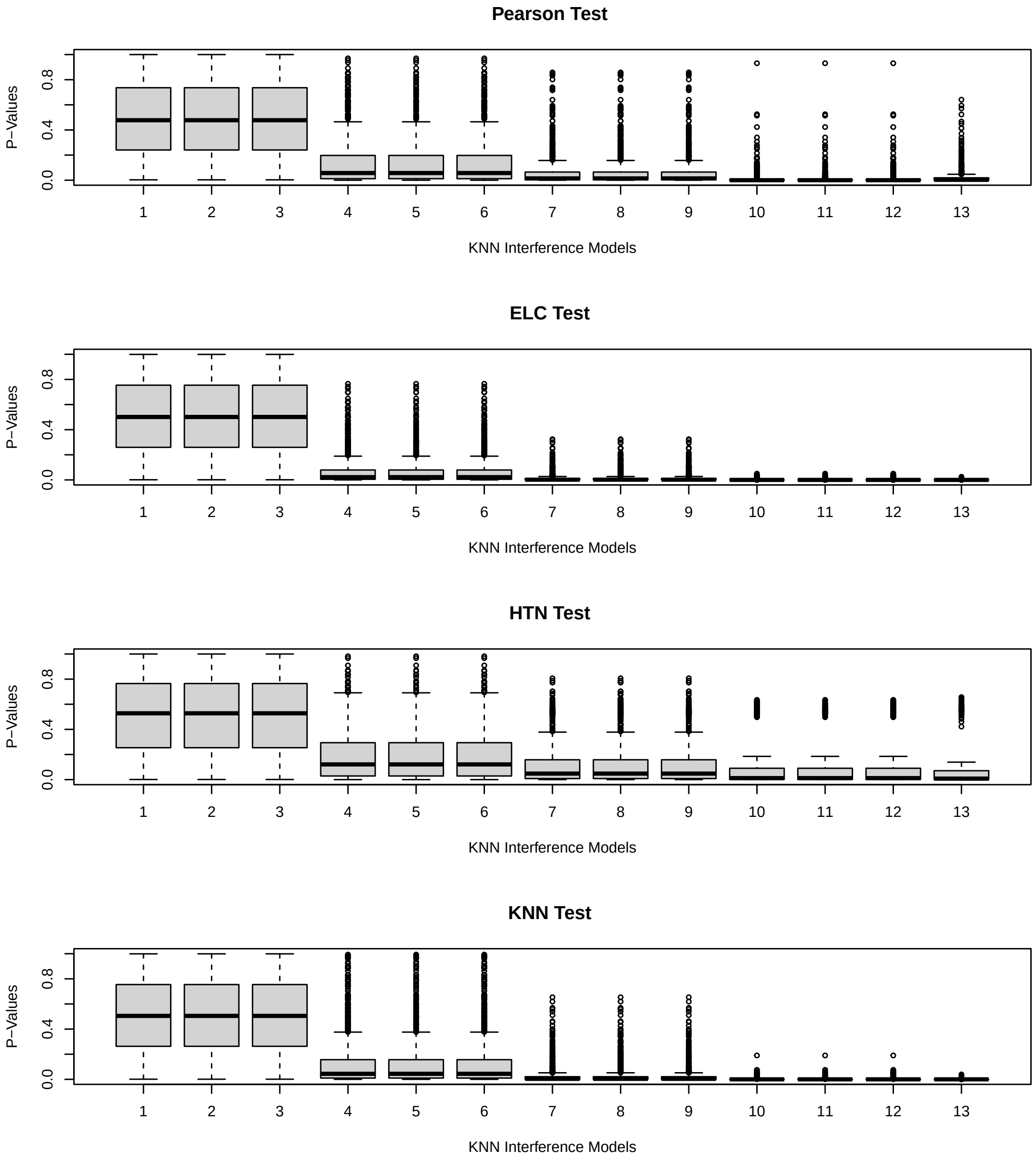


Figure 2.3: *Boxplots of p-values for the Pearson test (Pearson), edge level contrast test (ELC), has treated neighbor test (HTN) and K-nearest neighbors indirect effect test (KNN) under various KNNIM models using only control focal units. We use* $N = 256$ *units and* $K = 3$ *nearest neighbors. The p-values are estimated using 1,000 randomizations for each of the 1,000 generated potential outcome realizations.*

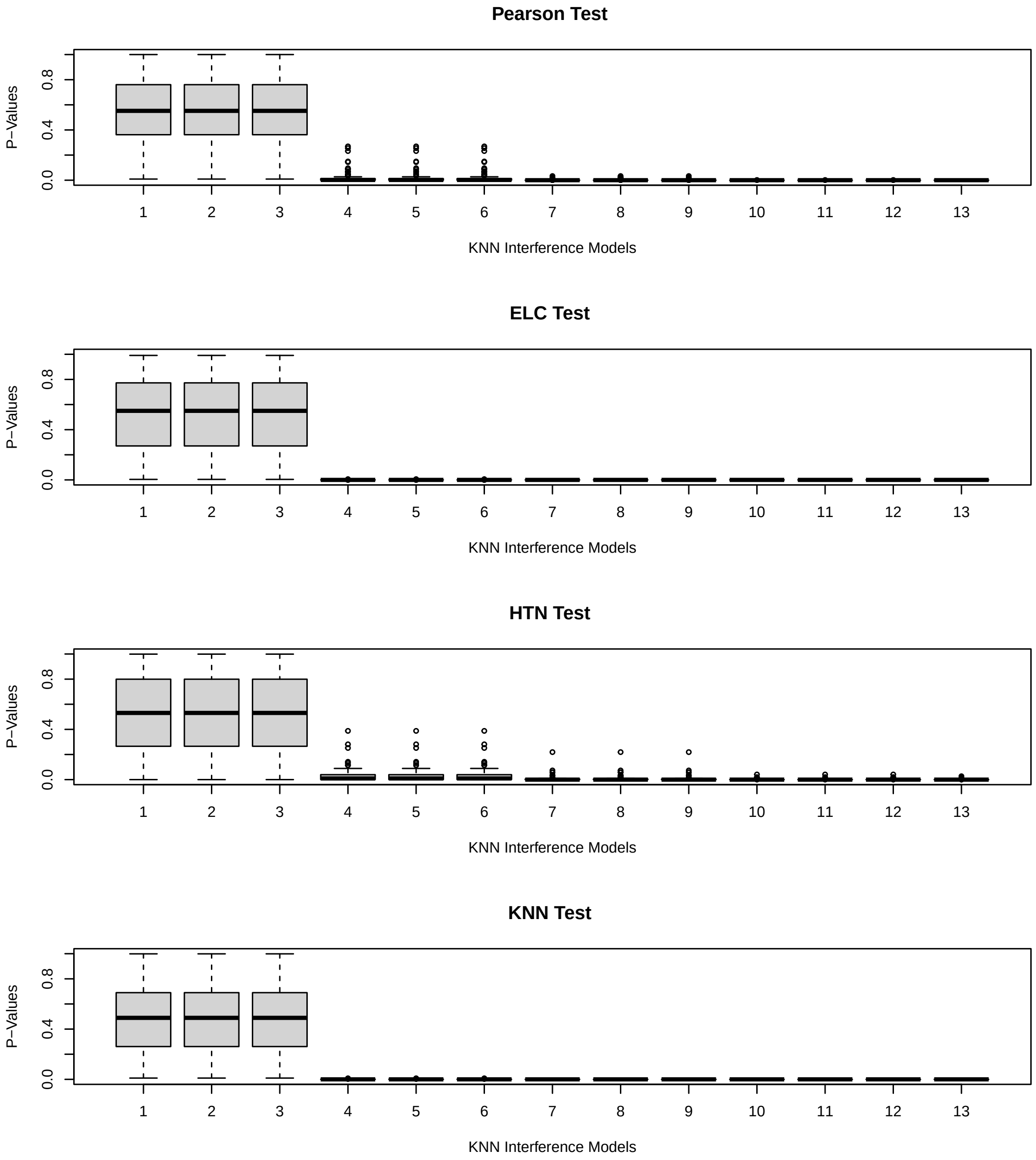


Figure 2.4: *Boxplots of p-values for the Pearson test (Pearson), edge level contrast test (ELC), has treated neighbor test (HTN) and K-nearest neighbors indirect effect test (KNN) under various KNNIM models using only control focal units. We use $N = 1024$ units and $K = 3$ nearest neighbors. The p-values are estimated using 1,000 randomizations for each of the 100 generated potential outcome realizations. .*

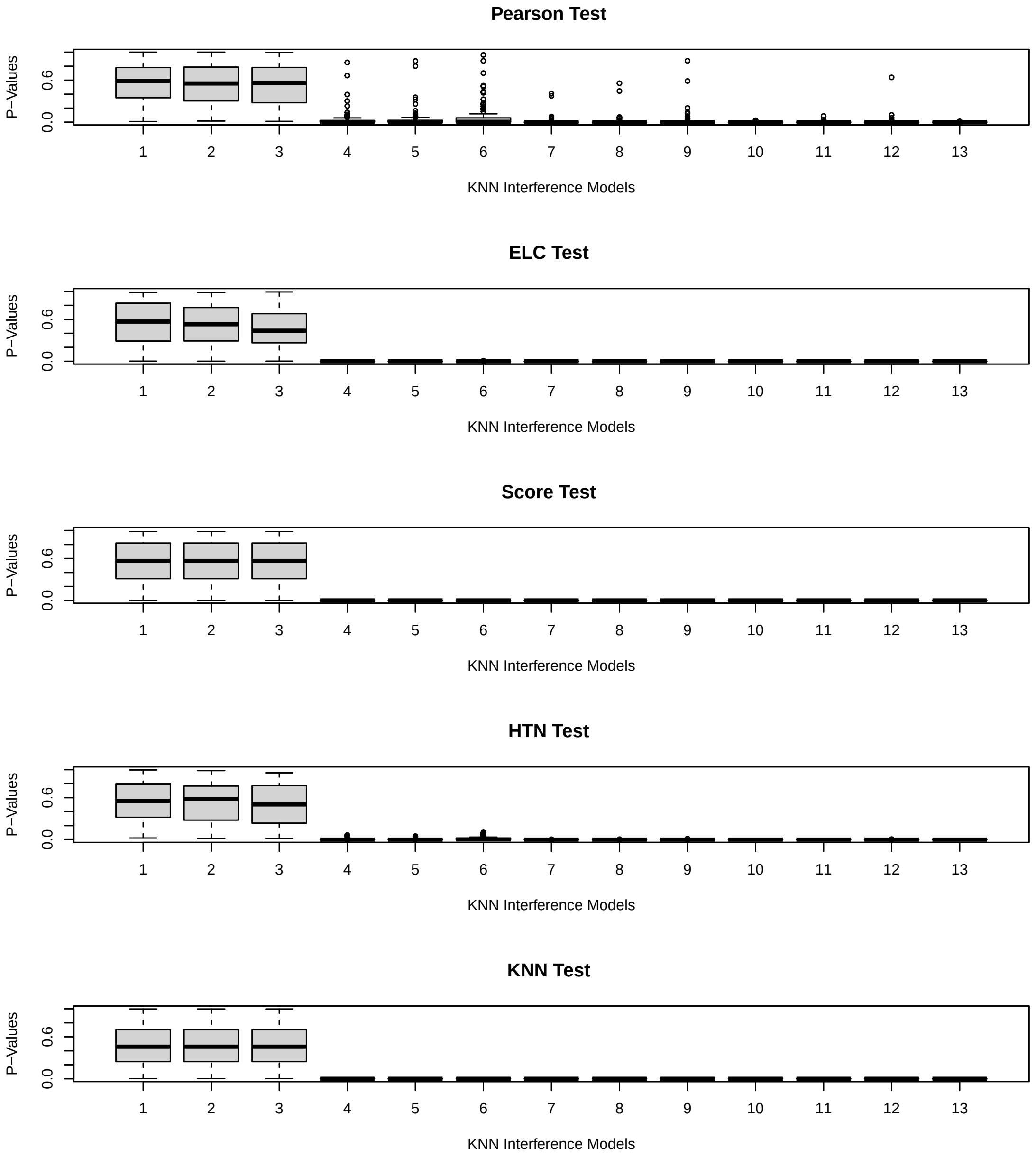


Figure 2.5: *Boxplots of p-values for the Pearson test (Pearson), edge level contrast test (ELC), score test (Score), has treated neighbor test (HTN) and K-nearest neighbors indirect effect test (KNN) under various KNNIM models. We use $N = 1024$ units and $K = 3$ nearest neighbors. The p-values are estimated using 1,000 randomizations for each of the 100 generated potential outcome realizations.*

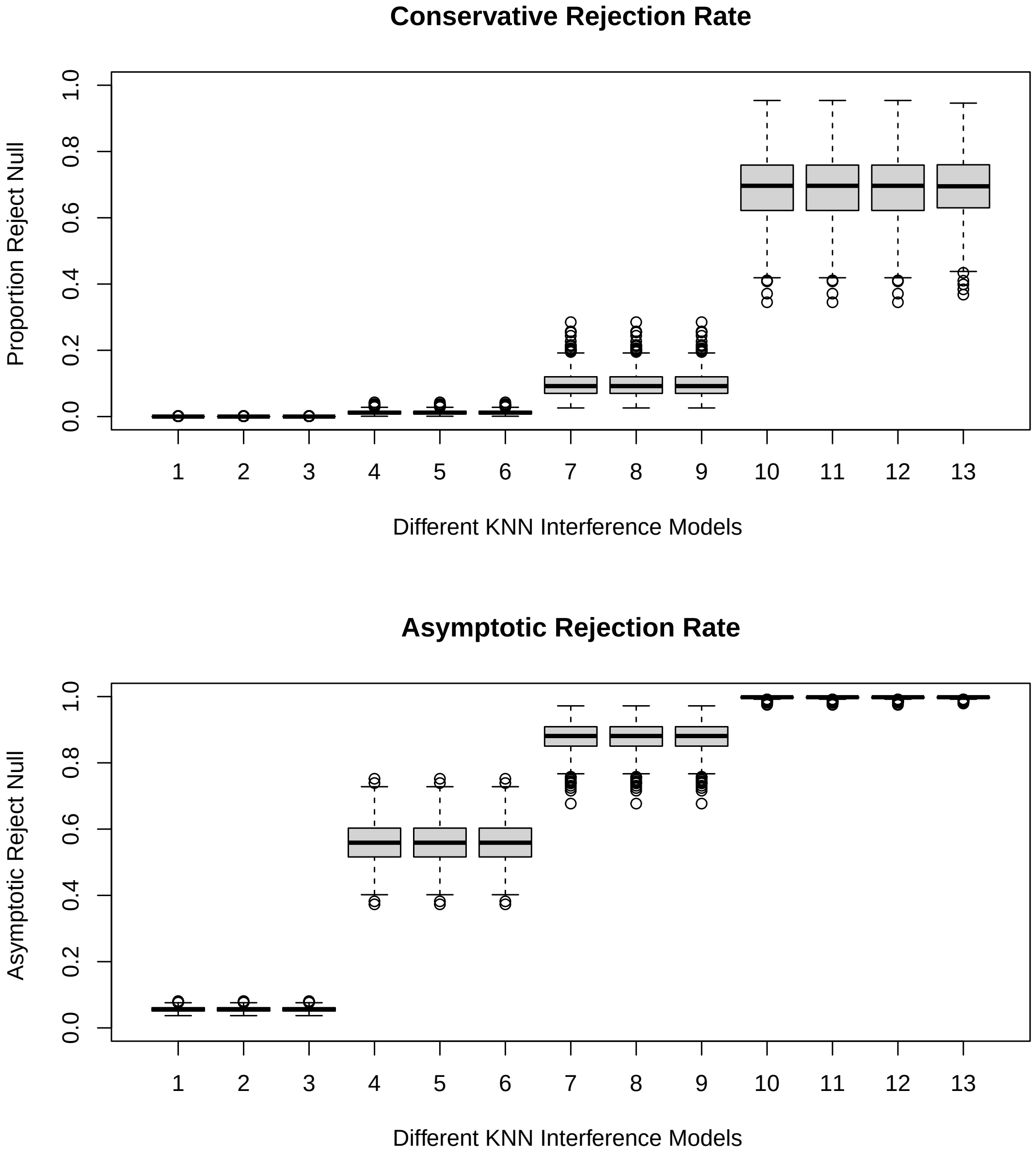


Figure 2.6: *Boxplots of the estimated rejection rates under the experimental design approach for both the conservative and asymptotic tests of the null hypothesis of no treatment interference under various KNNIM models. We use $N = 256$ units and $K = 3$ nearest neighbors. The rejection rates are estimated using 1,000 treatment assignments for each of the 1,000 generated potential outcomes. Tests are performed at significance level $\alpha = 0.05$.*

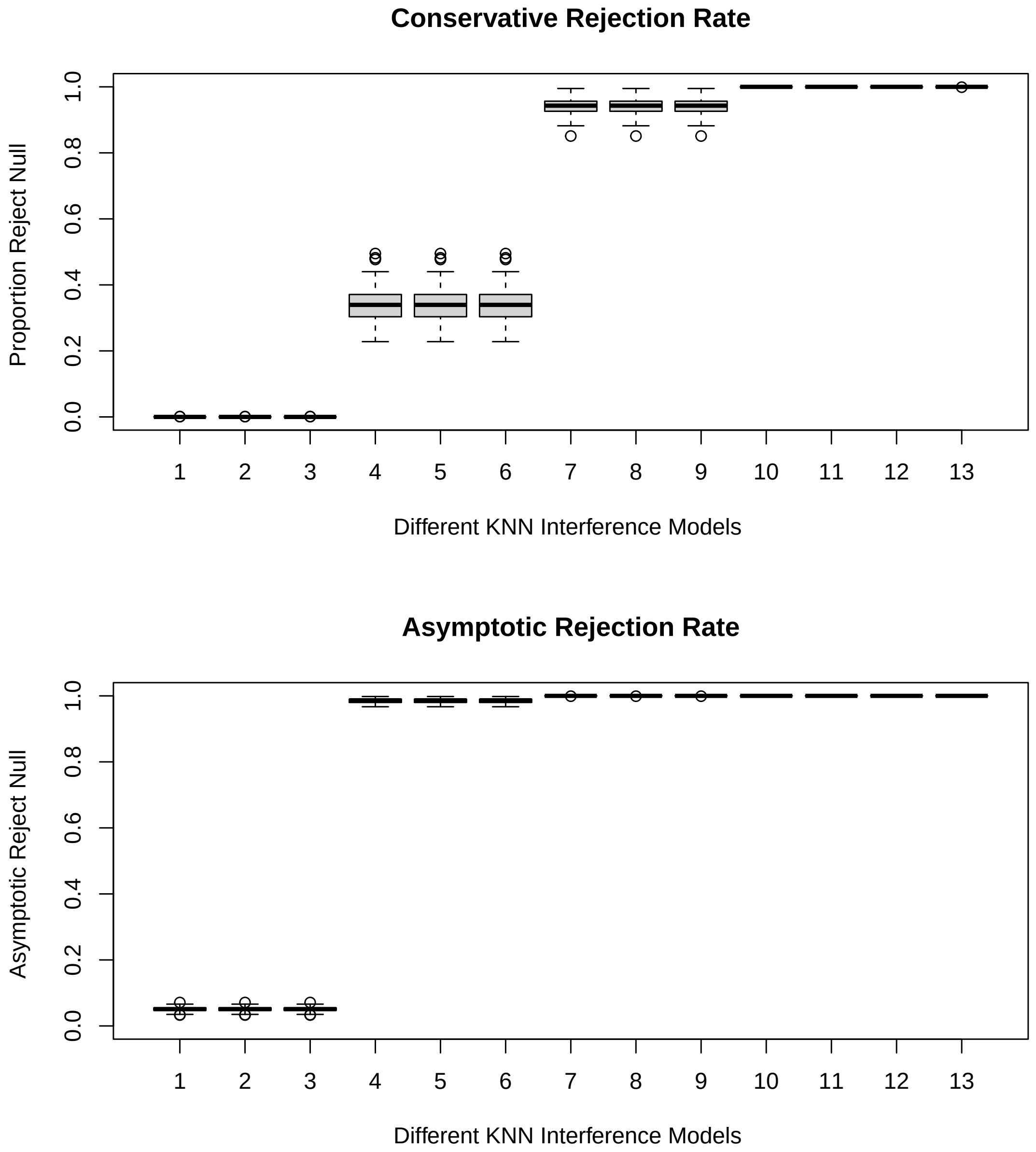


Figure 2.7: *Boxplots of the estimated rejection rates under the experimental design approach for both the conservative and asymptotic tests of the null hypothesis of no treatment interference under various KNNIM models. We use* $N = 1024$ *units and* $K = 3$ *nearest neighbors. The rejection rates are estimated using 1,000 treatment assignments for each of the 100 generated potential outcomes. Tests are performed at significance level* $\alpha = 0.05$.

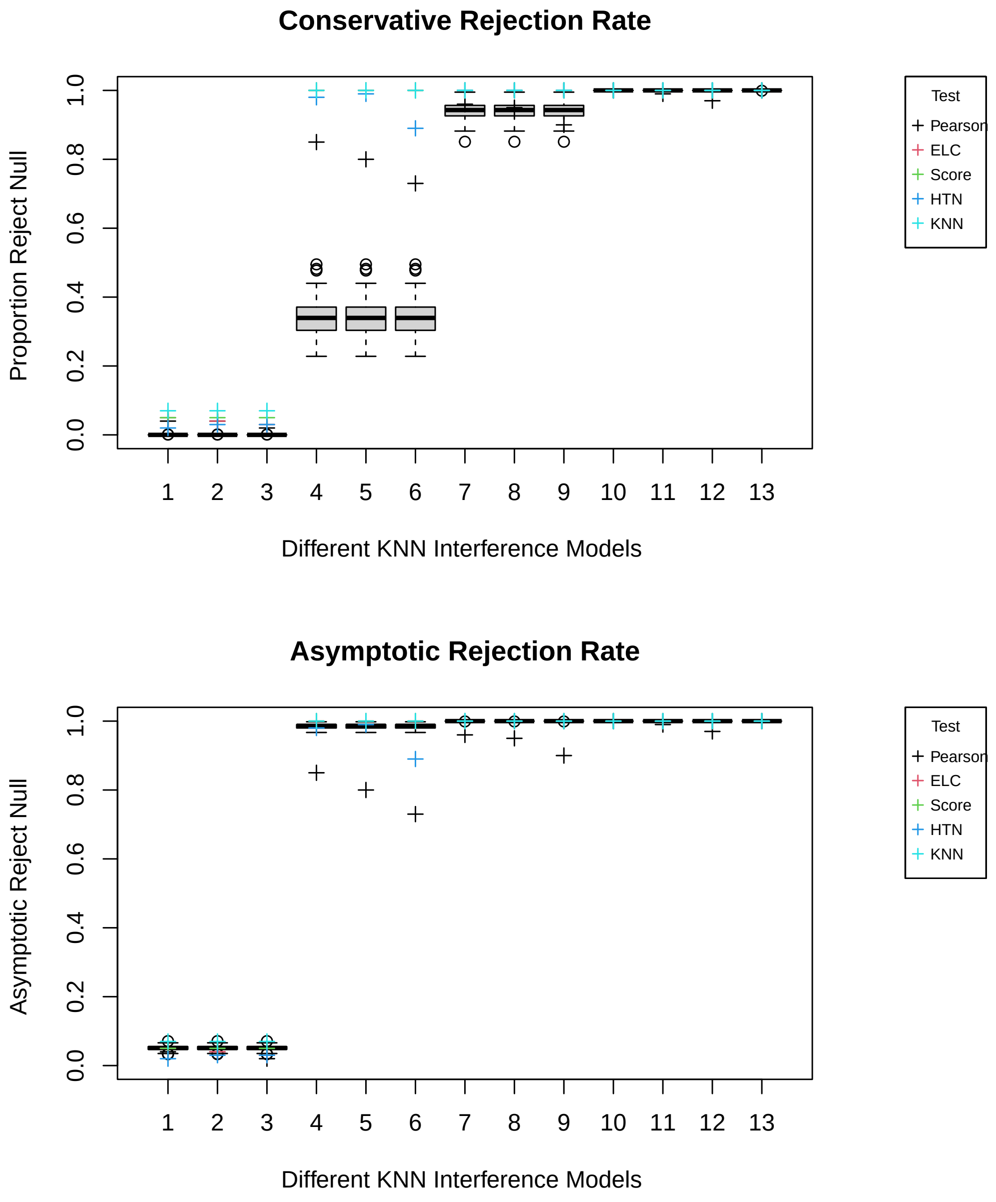


Figure 2.8: *Boxplots of the estimated rejection rates under the experimental design approach for both the conservative and asymptotic tests of the null hypothesis of no treatment interference under various KNNIM models. Plots also contain the estimated Type I Error (Models 1–3) and power (Models 4–13) for the Pearson test (Pearson), edge level contrast test (ELC), score test (Score), has treated neighbor test (HTN) and K-nearest neighbors indirect effect tests (KNN). We use $N = 1024$ units and $K = 3$ nearest neighbors. The rejection rates are estimated using 1,000 treatment assignments for each of the 100 generated potential outcomes. Tests are performed at significance level $\alpha = 0.05$.*

Table 2.5: *Estimated Type I Errors and power for tests of treatment interference for sample size* $N = 1024$

Models	Score	KNN	ELC	HTN	Pearson	Cons	Asymp
Model 1	0.05	0.07	0.05	0.02	0.04	0.00	0.051
Model 2	0.05	0.07	0.04	0.03	0.04	0.00	0.051
Model 3	0.05	0.07	0.03	0.03	0.02	0.00	0.051
Model 4	1.00	1.00	1.00	0.98	0.85	0.33	0.98
Model 5	1.00	1.00	1.00	0.99	0.80	0.33	0.98
Model 6	1.00	1.00	1.00	0.89	0.73	0.33	0.98
Model 7	1.00	1.00	1.00	1.00	0.96	0.94	1.00
Model 8	1.00	1.00	1.00	1.00	0.95	0.94	1.00
Model 9	1.00	1.00	1.00	1.00	0.90	0.94	1.00
Model 10	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Model 11	1.00	1.00	1.00	1.00	0.99	1.00	1.00
Model 12	1.00	1.00	1.00	1.00	0.97	1.00	1.00
Model 13	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Estimated Type I Errors (Models 1–3) and estimated power (Models 4–13) for simulated data under KNNIM. Results are provided for the score test (Score), K-nearest neighbors indirect effect test (KNN), edge level contrast test (ELC), has treated neighbor test (HTN) and the Pearson test (Pearson). Estimates of the median rejection rates under the experimental design approach for both the conservative (Cons) and asymptotic (Asymp) tests are also provided. We use $N = 1024$ units and $K = 3$ nearest neighbors. These values are estimated using 100 generated potential outcomes with 1,000 treatment assignments performed on each set of potential outcomes. Tests are performed at significance level $\alpha = 0.05$.

Chapter 3

Estimation of Causal Effects under K-Nearest Neighbors Interference

3.1 Introduction

In randomized experiments, assessing causal effects requires special care when the treatment condition assigned to one unit is allowed to affect the response of other units. Under this setting, a unit's outcome is not only influenced by its own treatment status—a *direct effect* of treatment—but may also be influenced by other units' treatments—an *indirect effect* (Sobel, 2006; Rosenbaum, 2007; Hudgens and Halloran, 2008).

In causal inference terminology, these experiments exhibit *treatment interference*, *treatment spillover*, *network effects*, or *peer effects*. This interference is especially common in settings with a social factor where units are allowed to interact with each other, for example, in studies on social media networks.

Experiments exhibiting treatment interference violate the *stable unit treatment value assumption* (SUTVA)—a foundational assumption of traditional causal inference methods. In particular, SUTVA requires that the treatment assigned to one unit affects only the outcome of that unit and does not affect the outcomes of other units (Rubin, 1980). The presence of interference may complicate statistical analysis and lead to inaccurate inference if not carefully taken into account (Sobel, 2006).

Traditionally, the effect of social influence and interaction between units has been viewed as a nuisance prohibiting accurate estimation of the direct effect of treatment. Considerable work has focused on designing experiments to mitigate the effect of treatment interference, for example, through clustering units that are likely to interact with each other and assigning treatment to these clusters instead of individual units (Ugander et al., 2013; Gui et al., 2015; Eckles et al., 2016). However, recent applications—for example, studies conducted on social media platforms and those evaluating the efficacy of vaccination strategies—have giving rise to studies in which quantifying and estimating interference effects is of primary interest (Hudgens and Halloran, 2008; Aronow and Samii, 2017; Forastiere et al., 2020; Sussman and Airoldi, 2017; Toulis and Kao, 2013; Alzubaidi and Higgins, 2022).

Methods for estimating indirect effects often begin by classifying the interaction through defining an exposure mapping on the units under study—a network where nodes represent units under study and edges between vertex indicate that the corresponding units may interact with each other. There may additionally be an interaction measure computed between each pair of units in the exposure mapping indicating the strength of that interaction. One example of the exposure mapping is to allow for treatment interference within groups of units but not across groups (Sobel, 2006; Rosenbaum, 2007; Hudgens and Halloran, 2008; Tchetgen and VanderWeele, 2012; Basse and Feller, 2018). Another approach assumes that treatment interference is restricted to a small neighborhood (Sussman and Airoldi, 2017). Aronow and Samii (2017) develop general estimation methods of treatment effects under arbitrary but known forms of interference. Indirect effects are then estimable through making assumptions on this exposure mapping and interaction measure, for example, by allowing interaction if the interaction measure is sufficiently large,... etc.

We build on this literature by analyzing the estimation of direct and indirect treatment effects under the K-nearest neighbor interaction model (KNNIM) (Alzubaidi and Higgins, 2022). In this model, the treatment given to one unit may interfere with the response of another unit if the first unit is one of the K individuals “closest” to the second unit with respect to the interaction measure. Quantifying K-nearest neighbors effects may help researchers tease out peer effects induced by, for example, interactions between best friends, spouses, siblings,or close colleagues. Additionally, this model has several appealing properties. First, this model allows for users with stronger interactions to produce larger indirect effects than users with weaker interactions, and will ignore potential indirect effects due to dilapidated, but technically present, connections (e.g. Facebook friends that no longer interact with each other). Second, the marginal and joint probabilities for possible treatment exposures have closed-form expressions under common experimental settings, allowing for unbiased estimation of treatment effects and precise estimation of standard errors. Finally, KNNIM may be effective in estimating indirect effects in the presence of non-transitive *mogul* effects—effects induced by users that have influence over a large number of individuals, but may themselves only directly interact with a handful of individuals.

In this chapter, using a potential outcomes approach, we define causal estimands for direct and K-nearest-neighbor indirect effects. We then derive Horvitz-Thompson estimators (Horvitz and Thompson, 1952) for these estimands that are unbiased given exact marginal and joint probabilities for possible treatment exposures. We provide a closed-form solution to compute these marginal and joint probabilities under completely-randomized and Bernoulli-randomized experimental designs. We derive conservative standard errors for these estimators. We then demonstrate how these estimators may have significantly stronger precision when an assumption of no interaction between direct and indirect effects. We conclude by showcasing the effectiveness of these methods via simulation and application to a field experiment conducted to study an anti-conflict behaviors among middile school students in New Jersey .

The chapter is organized as follows. Section 3.2 sets up the notation and preliminaries. The K-nearest neighbors interference model is provided in Section 3.3. Section 3.4 defines causal effects under KNNIM. Proposed unbiased estimators under the K-neighborhood assumption with the derived properties are given in Section 3.5. Estimators under the no-interaction between direct and indirect effects assumption with the derived properties are provided in Section 3.6. Section 3.7 presents variance estimation. Simulation studies, discussion, and real data analysis are provided in Sections 3.8, 3.9, and 3.10. We conclude in

Section 3.11.

3.2 Notation and Preliminaries

Consider an experiment on N units where each unit is assigned either a treatment status or a control status. The Neyman-Rubin Causal Model (NRCM) (Splawa-Neyman et al., 1990; Rubin, 1974; Holland, 1986) is a commonly-assumed model of response for making causal inferences. Under this model, the observed response of a unit is determined by the treatment status given to that unit and the potential outcomes for that unit—the hypothetical responses of that unit under the possible treatment statuses. A fundamental assumption of this model is the stable-unit treatment value assumption (SUTVA), which requires that there is only a single version of each treatment status and the response of a unit is unaffected by the treatment status of any other unit (Rubin, 1978, 1980; Imbens and Rubin, 2015). Of note, experiments in which the outcome of a unit is affected by others treatments—a phenomenon known as *interference* (Cox, 1958; Rubin, 1980)—violate SUTVA. Failing to account for violations of SUTVA can lead to inaccurate treatment effect estimates (Sobel, 2006).

Under interference, the effect of a treatment on a unit may occur through direct application of the treatment to that unit, indirectly through application of treatment to units that affect the response of the original unit, or both (Hudgens and Halloran, 2008). When interference is allowed to take completely arbitrary forms, treatment effect estimates are often estimated with very low power, or may be unidentifiable (Aronow and Samii, 2017).

Thus, to make progress on this problem, researchers often make assumptions that restrict which units are allowed to interfere with each other (Toulis and Kao, 2013; Aronow and Samii, 2017; Ugander et al., 2013; Sussman and Airoldi, 2017).

We can extend the potential outcomes framework to account for both direct and indirect treatment components. Let $y_i(\mathbf{W}) = y_i(W_i, \mathbf{W}_{-i})$ denote the potential outcome of unit i under treatment allocation $\mathbf{W} = (W_1, W_2, \ldots, W_N) \in \{0, 1\}^N$, where unit i is given treatment W_i, and the remaining treatment statuses are allocated according to $\mathbf{W}_{-i}$. Responses

Y_i satisfy

$$Y_i = \sum_{\mathbf{W} \in \{0,1\}^N} y_i(\mathbf{W}) \mathbf{1}(\mathbf{W}' = \mathbf{W}),$$

where $\mathbf{1}(\mathbf{W}' = \mathbf{W})$ is an indicator variable that is equal to 1 if and only if the observed treatment status $\mathbf{W}' = \mathbf{W}$. That is, the response of unit i only depends on the potential outcomes of unit i and the treatment assignment given to that unit.

Without making assumptions on the amount of interference allowed in a study, it may be impossible to estimate common causal quantities of interest in any practical way—for example, each unit may have up to 2^N potential outcomes when interference is unconstrained. Thus, to make progress on treatment interference problems, researchers often place strong restrictions on the extent of interference allowed (Toulis and Kao, 2013; Aronow and Samii, 2017; Ugander et al., 2013; Sussman and Airoldi, 2017). This often begins by constructing an *exposure mapping* $G = (V, E)$—a directed graph where each vertex $i \in V$ represents a unit under study and each edge $\vec{ij} \in E$ denotes that the treatment status of unit i may potentially interfere with the response of unit j. Each edge $\vec{ij} \in E$ may also have a weight $d(i,j)$ denoting the strength of the the potential interference which may be observed through studying interactions between i and j—stronger interactions between i and j correspond to smaller values of $d(i,j)$. Under an assumed exposure mapping G, for two treatment allocations $\mathbf{W}$, $\mathbf{W}'$, we have that $y_i(\mathbf{W}) = y_i(\mathbf{W}')$ if $W_j = W'_j$ for all $j \in V$ such that $\vec{ji} \in E$. If $\vec{ki} \notin E$, then treatment statuses W_k, W'_k may differ without affecting equality of the potential outcomes.

Once the exposure mapping is specified, models can then further restrict the nature of interference allowable. For example, a common assumption is that interference can only occur if a certain number or fraction of neighbors within the exposure mapping are given the treatment condition. However, few existing models specify exactly which neighbors in the exposure mapping are allowed to interfere with a unit's response, or allow for indirect effects to differ across neighbors. As an alternative, we propose a model where interference of treatment on a unit i is restricted to its K–nearest neighbors (Alzubaidi and Higgins, 2022). This model allows neighbors with stronger interactions to contribute larger indirect effects, and limits the ability of weakly-interacting units to affect response.

3.3 K-Nearest Neighbors Interference Model

The K-nearest neighbors interference model (KNNIM) is a recently-proposed model in which a unit j is only allowed to interfere with the response of unit i if j is within i's K-neighborhood (Alzubaidi and Higgins, 2022). The *K-neighborhood* of unit i, denoted $\mathcal{N}_{iK}$, is the set of the K "closest" units to unit i:

$$\mathcal{N}_{iK} = \{j : d(i,(j)) \geq d(i,(K)), j = 1, 2, \ldots, K\}. \tag{3.1}$$

Define $\mathcal{N}_{-iK} = V \setminus (i \cup N_{iK})$ as all units in V that are outside of i's K-neighborhood. Note that the sets $\{i, \mathcal{N}_{ik}, \mathcal{N}_{-ik}\}$ form a partition of V.

Let $\mathbf{W} = (W_i, \mathbf{W}_{\mathcal{N}_{iK}}, \mathbf{W}_{\mathcal{N}_{-iK}})$ denote treatment assignment vector for all units N, partitioned into the treatment given to unit i, the treatments given to i's K-nearest neighbors, and the treatments given to all other units. Treatment statuses in $\mathbf{W}_{\mathcal{N}_{iK}}$ and $\mathbf{W}_{\mathcal{N}_{-iK}}$ are given in descending order with respect to $d(i,j)$ (*e.g.* the first entry of $\mathbf{W}_{\mathcal{N}_{iK}}$ is the treatment assignment given to the nearest neighbors of unit i). The defining assumption of KNNIM is as follows:

Assumption 3.1. *(K-Neighborhood Interference Assumption (K-NIA)). The potential outcomes $y_i(\mathbf{W})$ for all units i satisfy*

$$y_i(W_i, \mathbf{W}_{\mathcal{N}_{iK}}, \mathbf{W}_{\mathcal{N}_{-iK}}) = y_i(W_i, \mathbf{W}_{\mathcal{N}_{iK}}, \mathbf{W}'_{\mathcal{N}_{-iK}}). \tag{3.2}$$

In words, K-NIA ensures that the potential outcome of unit i is only affected by its treatment and by the treatments assigned to its K-nearest neighbors. Changing treatments for other units outside the K-neighborhood will not affect the potential outcome of unit i. Note that this model restricts the number of potential outcomes to be 2^{K+1} for each unit. The choice of K is ultimately left to the researcher, though large values of K may not be sufficiently restrictive to allow for reliable estimates and inferences. For brevity, we will suppress the treatment assignment outside of the K-nearest neighbors when denoting potential outcomes under KNNIM: $y_i(W_i, \mathbf{W}_{\mathcal{N}_{iK}}) = y_i(W_i, \mathbf{W}_{\mathcal{N}_{iK}}, \mathbf{W}_{\mathcal{N}_{-iK}})$.

3.4 Causal Estimands under KNNIM

Using the potential outcomes framework and following Hudgens and Halloran (2008), we now define causal estimands under KNNIM. We start with general definitions of direct and indirect effects and conclude with KNNIM-specific nearest neighbors effects.

3.4.1 Direct, Indirect, and Total Effects

The *average direct effect* (ADE) δ_{dir} is the average difference in a unit's potential outcomes when changing that unit's treatment status and holding all other units' treatment status fixed. It may be defined as

$$\delta_{dir} = \frac{1}{N}\sum_{i=1}^{N}(y_i(1,\mathbf{1}) - y_i(0,\mathbf{1})), \tag{3.3}$$

where $\mathbf{1}$ denotes a vector of 1's of length K. In contrast to direct effect, the *average indirect effect* (AIE) δ_{ind} is defined as the average difference in a unit's potential outcome when changing all other treatment statuses from control to treated, holding its own treatment fixed. It may be defined as

$$\delta_{ind} = \frac{1}{N}\sum_{i=1}^{N}(y_i(0,\mathbf{1}) - y_i(0,\mathbf{0})), \tag{3.4}$$

where $\mathbf{0}$ denotes a vector of 0's of length K. The *average total effect* (ATOT) δ_{tot} measures the average difference in potential outcomes between all units receiving treatment and all units receiving control:

$$\delta_{tot} = \frac{1}{N}\sum_{i=1}^{N}(y_i(1,\mathbf{1}) - y_i(0,\mathbf{0})). \tag{3.5}$$

Note that these quantities are defined to satisfy $\delta_{tot} = \delta_{dir} + \delta_{ind}$. Addionally, when SUTVA holds, $\delta_{tot} = \delta_{dir}$ and $\delta_{ind} = 0$.

3.4.2 The ℓ^{th}–Nearest Neighbor Indirect Effect

Let $\mathbf{W}^*_\ell = (W^*_{\ell,1}, W^*_{\ell,2}, \dots, W^*_{\ell,K}) \in \{0,1\}^K$ denote the treatment vector assignment of length K where the first ℓ nearest neighbors are given treatment and the rest are control:

$$W^*_{\ell,j} = \begin{cases} 1, & j \leq \ell, \\ 0, & \text{otherwise.} \end{cases} \tag{3.6}$$

Note that $\mathbf{W}^*_K = \mathbf{1}$, and define $\mathbf{W}^*_0 = \mathbf{0}$. The *average ℓ^{th}–nearest neighbor indirect effect* (AℓNNIE) is defined as

$$\delta_\ell = \frac{1}{N} \sum_{i=1}^{N} (y_i(0, \mathbf{W}^*_\ell) - y_i(0, \mathbf{W}^*_{\ell-1})). \tag{3.7}$$

Note that $\mathbf{W}^*_\ell$ and $\mathbf{W}^*_{\ell-1}$ are identical except that $W^*_{\ell,\ell} = 1$ and $W^*_{\ell-1,\ell} = 0$. Hence, δ_ℓ may be interpreted as the average difference in response due to the treatment status of the ℓ^{th}–nearest-neighbor. Additionally, under KNNIM, the AIE is the sum of the AℓNNIEs.

Lemma 3.1.

$$\delta_{ind} = \sum_{\ell=1}^{K} \delta_\ell, \tag{3.8}$$

$$\delta_{tot} = \delta_{dir} + \delta_{ind} \tag{3.9}$$

Proofs are provided in Appendix B

3.5 Horvitz–Thompson Estimators

We now derive Horvitz–Thompson (HT) estimators for the estimands described in Section 3.4. Our approach closely follows that in Aronow and Samii (2017). Of particular note, these HT estimators require computing the marginal and joint probabilities of observing various treatment allocations. Thankfully, for many common designs, these probabilities can be computed exactly under KNNIM; we give closed-form solutions for these probabilities under completely-randomized and Bernoulli-randomized designs. When these probabilities cannot be computed exactly, they may still be estimated, for example, using the approach in Aronow and Samii (2017).

Let $Y_i^{obs} = Y_i^{obs}(W_i, \mathbf{W}_{\mathcal{N}_{ik}})$ denote the observed potential outcome of unit i. Let $\pi_i(W, \mathbf{W}_{\mathcal{N}_K})$ denote the marginal probability that unit i is given exposure $(W, \mathbf{W}_{\mathcal{N}_K})$—that is, the overall treatment allocation assigns treatment W to unit i and assigns treatment conditions $\mathbf{W}_K$ to i's K-neighborhood. Define $\pi_{ij}(W, \mathbf{W}_{\mathcal{N}_K})$ as the joint probability that units i and j are both given exposure $(W, \mathbf{W}_{\mathcal{N}_K})$ and define $\pi_{ij}((W, \mathbf{W}_{\mathcal{N}_K}), (W', \mathbf{W}'_{\mathcal{N}_K}))$ as the joint probability that unit i receives exposure $(W, \mathbf{W}_{\mathcal{N}_K})$ and unit j receives exposure $(W', \mathbf{W}'_{\mathcal{N}_K})$. Additionally, define indicator variables $I_i(W, \mathbf{W}_{\mathcal{N}_K})$ that are equal to 1 if unit i is given exposure $(W, \mathbf{W}_{\mathcal{N}_K})$, and is 0 otherwise.

Define

$$\bar{y}(W, \mathbf{W}_{\mathcal{N}_K}) = \frac{1}{N}\sum_{i=1}^{N} y_i(W, \mathbf{W}_{\mathcal{N}_K}) \tag{3.10}$$

as the average response across all N units under treatment allocation $(W, \mathbf{W}_{\mathcal{N}_K})$. The Horvitz-Thompson (HT) estimator (Horvitz and Thompson, 1952) for $\bar{y}(W, \mathbf{W}_{\mathcal{N}_K})$ is

$$\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{\mathcal{N}_K}) = \frac{1}{N}\sum_{i=1}^{N} I_i(W, \mathbf{W}_{\mathcal{N}_K}) \frac{Y_i^{obs}(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})}. \tag{3.11}$$

This estimator is unbiased for $\bar{y}(W, \mathbf{W}_{\mathcal{N}_K})$,

$$\mathbf{E}(\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{\mathcal{N}_K})) = \bar{y}(W, \mathbf{W}_{\mathcal{N}_K}), \tag{3.12}$$

and has variance

$$\mathbf{Var}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K})) = \frac{1}{N^2}\sum_{i=1}^{N} \pi_i(W, \mathbf{W}_{\mathcal{N}_K})[1 - \pi_i(W, \mathbf{W}_{\mathcal{N}_K})] \left[\frac{y_i(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})}\right]^2$$
$$+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j \neq i}[\pi_{ij}(W, \mathbf{W}_{\mathcal{N}_K}) - \pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_j(W, \mathbf{W}_{\mathcal{N}_K})] \frac{y_i(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})}\frac{y_j(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_j(W, \mathbf{W}_{\mathcal{N}_K})}. \quad (3.13)$$

The covariance of the HT estimators under any two exposures $(W, \mathbf{W}_{\mathcal{N}_K})$ and $(W', \mathbf{W}'_{\mathcal{N}_K})$ is

$$\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{\mathcal{N}_K})) =$$
$$\frac{1}{N^2}\left(\sum_{i=1}^{N}\sum_{j \neq i}\left[\pi_{ij}((W, \mathbf{W}_{\mathcal{N}_K}), (W', \mathbf{W}'_{\mathcal{N}_K})) - \pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_j(W', \mathbf{W}'_{\mathcal{N}_K})\right]\right.$$
$$\left. \times \frac{y_i(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})}\frac{y_j(W', \mathbf{W}'_{\mathcal{N}_K})}{\pi_j(W', \mathbf{W}'_{\mathcal{N}_K})}\right) - \frac{1}{N^2}\sum_{i=1}^{N} y_i(W, \mathbf{W}_{\mathcal{N}_K})y_i(W', \mathbf{W}'_{\mathcal{N}_K}). \quad (3.14)$$

From (3.12), (3.13), and (3.14), the expectation and variance for the difference in HT estimators for the average response under any two unique exposures $(W, \mathbf{W}_{\mathcal{N}_K})$, $(W', \mathbf{W}'_{\mathcal{N}_K})$ can be computed as follows:

Theorem 3.1.

$$\mathbf{E}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K}) - \bar{Y}^{obs}_{HT}(W', \mathbf{W}'{}_{\mathcal{N}_K})) = \bar{y}(W, \mathbf{W}_{\mathcal{N}_K}) - \bar{y}(W', \mathbf{W}'{}_{\mathcal{N}_K}) \quad (3.15)$$

$$\mathbf{Var}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K}) - \bar{Y}^{obs}_{HT}(W', \mathbf{W}'{}_{\mathcal{N}_K})) =$$
$$\mathbf{Var}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K})) + \mathbf{Var}(\bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{\mathcal{N}_K})) - 2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{\mathcal{N}_K})). \quad (3.16)$$

We can then obtain unbiased estimators for the ADE, AIE, ATOT, and the AℓNNIE as follows:

$$\widehat{\delta}_{HT,dir} = \bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{1}), \tag{3.17}$$

$$\widehat{\delta}_{HT,ind} = \bar{Y}^{obs}_{HT}(0,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0}), \tag{3.18}$$

$$\widehat{\delta}_{HT,tot} = \bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0}), \tag{3.19}$$

$$\widehat{\delta}_{HT,\ell} = \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1}). \tag{3.20}$$

Variances for these estimators are derived as in (3.16). As with the estimands, we have the following relationships between these estimators.

Lemma 3.2.

$$\widehat{\delta}_{HT,tot} = \widehat{\delta}_{HT,dir} + \widehat{\delta}_{HT,ind}, \tag{3.21}$$

$$\widehat{\delta}_{HT,ind} = \sum_{\ell=1}^{K} \widehat{\delta}_{HT,\ell}. \tag{3.22}$$

3.5.1 Marginal and Joint Exposure Probabilities

One significant benefit of KNNIM is that this model allows for closed-form expressions of the marginal and joint exposure probabilities for many common experimental designs. We now provide these exposure probabilities under completely-randomized and Bernoulli-randomized designs.

3.5.1.1 Exposure Probabilities Under Complete Randomization

In a completely-randomized design, the number of treated units N_t in the study is selected prior to randomization. Each possible treatment has the same $\binom{N}{N_t}^{-1}$ probability of occurring (Kuehl, 2000).

Consider $i \cup \mathcal{N}_{iK}$, the closed K-neighborhood for unit i. Suppose that the exposure $(W, \mathbf{W}_{N_K})$ has a total of N_{itK} treatment conditions and N_{icK} control conditions—the specific units given these conditions do not factor into the probability computations. The marginal probability that $i \cup \mathcal{N}_{iK}$ receives exposure $(W, \mathbf{W}_{N_K})$ is

$$\pi_i(W, \mathbf{W}_{N_K}) = \frac{\binom{N-K-1}{N_t - N_{itK}}}{\binom{N}{N_t}}. \tag{3.23}$$

Two treatment exposures $(W, \mathbf{W}_{N_K})$, $(W', \mathbf{W}'_{N_K})$ for units i and j respectively are called *compatible* if they can co-occur within a given treatment assignment. For example, if j is unit i's nearest neighbor, the exposures $(1, \mathbf{1})$ for unit i and $(0, \mathbf{0})$ for unit j are *incompatible* since unit j is given treatment in the first exposure and control in the second exposure. The joint probability of observing two incompatible treatment assignments is 0.

For two closed K-neighborhoods $i \cup \mathcal{N}_{iK}$, $j \cup \mathcal{N}_{jK}$, let b_{ij} denote the number of units in the overlap of the two neighborhoods: $b_{ij} = |(i \cup \mathcal{N}_{iK}) \cap (j \cup \mathcal{N}_{jK})|$. For two exposures $(W, \mathbf{W}_{N_K})$, $(W', \mathbf{W}'_{N_K})$ for units i and j respectively, let N_{jitK} and N_{jicK} denote the number of treated and control units respectively in $(W', \mathbf{W}'_{N_K})$ not already belonging to i's closed K-neighborhood:

$$N_{jitk} = \sum_{\substack{j' \in \{j \cup \mathcal{N}_{jK}\} \\ j' \notin \{i \cup \mathcal{N}_{iK}\}}} W_{j'}, \ N_{jick} = \sum_{\substack{j' \in \{j \cup \mathcal{N}_{jK}\} \\ j' \notin \{i \cup \mathcal{N}_{iK}\}}} 1 - W_{j'}. \tag{3.24}$$

Then, the joint probability of units i and j being exposed to $(W, \mathbf{W}_{N_K})$ and $(W', \mathbf{W}'_{N_K})$ respectively is

$$\begin{aligned} &\pi_{ij}((W, \mathbf{W}_{N_K}), (W', \mathbf{W}'_{N_K})) \\ &= \begin{cases} \frac{\binom{N-K-1}{N_t - N_{itK}}}{\binom{N}{N_t}} \frac{\binom{N-2K-2+b_{ij}}{N_t - N_{itK} - N_{jitK}}}{\binom{N-K-1}{N_t - N_{itK}}}, & (W, \mathbf{W}_{N_K}), (W', \mathbf{W}'_{N_K}) \text{ are compatible for } i \text{ and } j, \\ 0, & \text{otherwise.} \end{cases}, \end{aligned} \tag{3.25}$$

3.5.1.2 Exposure Probabilities Under Bernoulli Randomization

In a Bernoulli-randomized design, each unit has a pre-specified probability p of being assigned treatment, and treatments are assigned independently across units (*e.g.* treatment assignment is determined for each unit by flipping a coin that has probability p of landing heads). Under Bernoulli-randomization, the marginal probability that $i \cup \mathcal{N}_{iK}$ receives exposure $(W, \mathbf{W}_{N_K})$ is

$$\pi_i(W, \mathbf{W}_{N_K}) = p^{N_{itK}}(1-p)^{N_{icK}}, \tag{3.26}$$

and joint probability of units i and j being exposed to $(W, \mathbf{W}_{N_K})$ and $(W', \mathbf{W}'_{N_K})$ respectively is

$$\begin{aligned} &\pi_{ij}((W, \mathbf{W}_{N_K}), (W', \mathbf{W}'_{N_K})) \\ &= \begin{cases} p^{Ni_{tK}}(1-p)^{Ni_{cK}} p^{N_{jitK}}(1-p)^{N_{jicK}}, & (W, \mathbf{W}_{N_K}), (W', \mathbf{W}'_{N_K}) \text{ are compatible for } i \text{ and } j, \\ 0, & \text{otherwise.} \end{cases}, \end{aligned} \tag{3.27}$$

3.6 Estimation Under No Interaction Between Direct and Indirect Effects

Next, we strengthen Assumption 3.1 in order to improve the power of our HT estimates.

Assumption 3.2. *(No Interaction Between Direct and Indirect Effects) For each unit i in a network G , there is no interaction between direct and indirect effects if* $[y_i(1, W_{\mathcal{N}_{ik}}) - y_i(1, W'_{\mathcal{N}_{ik}})] - [y_i(0, W_{\mathcal{N}_{ik}}) - y_i(0, W'_{\mathcal{N}_{ik}})] = 0$.

Under Assumption 3.2, if we assume that there is no interaction between the direct and indirect effects, the unbiased Horvitz–Thompson estimator of ATOT is provided as follows.

$$\widehat{\delta^*}_{HT,tot} = \widehat{\delta^*}_{HT,dir} + \widehat{\delta^*}_{HT,ind}. \tag{3.28}$$

Lemma 3.3. *For $C_1 = C_2 = \frac{1}{2}$,*

$$\widehat{\delta^*}_{HT,tot} = \widehat{\delta}_{HT,tot}. \tag{3.29}$$

Theorem 3.2. *Under the no-interaction between direct and indirect effects assumption, and for $C_1 = C_2 = \frac{1}{2}$,*

$$\mathbf{E}(\widehat{\delta^*}_{HT,tot}) = \delta_{tot}. \tag{3.30}$$

$$\mathbf{Var}(\widehat{\delta^*}_{HT,tot}) = \mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) + \mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0}). \tag{3.31}$$

Under Assumption 3.2, the unbiased Horvitz–Thompson estimator of ADE is as follows.

$$\widehat{\delta^*}_{HT,dir} = \frac{1}{2}[\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{1})] + \frac{1}{2}[\bar{Y}^{obs}_{HT}(1,\mathbf{0}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})]. \tag{3.32}$$

Theorem 3.3. *Under the no-interaction between direct and indirect effects assumption,*

$$\mathbf{E}(\widehat{\delta^*}_{HT,dir}) = \delta_{dir}. \tag{3.33}$$

$$\begin{aligned}\mathbf{Var}(\widehat{\delta^*}_{HT,dir}) = {} & \frac{1}{4}\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) + \frac{1}{4}\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{1})) \\ & + \frac{1}{4}\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{0})) + \frac{1}{4}\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) \\ & - 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{1})) + 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(1,\mathbf{0})) \\ & - 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}), \bar{Y}^{obs}_{HT}(1,\mathbf{0})) \\ & + 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{0}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})).\end{aligned} \tag{3.34}$$

Under Assumption 3.2, the unbiased Horvitz–Thompson estimator of AIE is as follows.

$$\widehat{\delta^*}_{HT,ind} = \frac{1}{2}[\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(1,\mathbf{0})] + \frac{1}{2}[\bar{Y}^{obs}_{HT}(0,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})]. \tag{3.35}$$

Theorem 3.4. *Under the no-interaction between direct and indirect effects assumption,*

$$\mathbf{E}(\widehat{\delta^*}_{HT,ind}) = \delta_{ind}. \tag{3.36}$$

$$\begin{aligned}\mathbf{Var}(\widehat{\delta^*}_{HT,ind}) = &\frac{1}{4}\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) + \frac{1}{4}\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{0})) \\ &+ \frac{1}{4}\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{1})) + \frac{1}{4}\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) \\ &- 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}),\bar{Y}^{obs}_{HT}(1,\mathbf{0})) + 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}),\bar{Y}^{obs}_{HT}(0,\mathbf{1})) \\ &- 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}),\bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{0}),\bar{Y}^{obs}_{HT}(0,\mathbf{1})) \\ &+ 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{0}),\bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}),\bar{Y}^{obs}_{HT}(0,\mathbf{0})).\end{aligned} \tag{3.37}$$

Under Assumption 3.2, if we assume that there is no interaction between the direct and indirect effects, the unbiased Horvitz–Thompson estimator of AℓNNIE is as follows,

$$\widehat{\delta^*}_{HT,\ell} = \frac{1}{2}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})] + \frac{1}{2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})]. \tag{3.38}$$

Lemma 3.4.

$$\widehat{\delta^*}_{HT,ind} = \sum_{\ell=1}^{K}\widehat{\delta^*}_{HT,\ell}. \tag{3.39}$$

Unbiasedness, and theoretical variance of HT-AℓNNIE estimator under Assumption 3.2 are provided in the following theorem.

Theorem 3.5. *Under the no-interaction between direct and indirect effects assumption,*

$$\mathbf{E}(\widehat{\delta^*}_{HT,\ell}) = \delta_\ell. \tag{3.40}$$

$$
\begin{aligned}
\mathbf{Var}(\widehat{\delta^*}_{HT,\ell}) = \frac{1}{4}\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell})) + \frac{1}{4}\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})) \\
+ \frac{1}{4}\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell})) + \frac{1}{4}\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) \\
- 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}), \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})) + 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}), \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell})) \\
- 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}), \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) - 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1}), \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell})) \\
+ 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1}), \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) - 2(\frac{1}{4})\mathbf{Cov}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell}), \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})). \quad (3.41)
\end{aligned}
$$

Proofs of the lemmas and theorems are given in Appendix B.

Note that in $\widehat{\delta^*}_{HT,\ell}$ under Assumption 3.2, we chose the weights of treated and control units to be $C_{\ell 1} = C_{\ell 2} = \frac{1}{2}$. However, let $S^2_{11} = \mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}))$, $S^2_{12} = \mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1}))$, $S^2_{21} = \mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell}))$ and $S^2_{22} = \mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1}))$ and assuming that the covariance components in 3.41 are equal to zero. If the experimenter has prior knowledge from previous studies on S^2_{11}, S^2_{12}, S^2_{21} and S^2_{22}, then $C_{\ell 1}$ and $C_{\ell 2}$ can be chosen such that $C_{\ell 1} = \frac{S^2_{21}+S^2_{22}}{S^2_{11}+S^2_{12}+S^2_{21}+S^2_{22}}$ and $C_{\ell 2} = \frac{S^2_{11}+S^2_{12}}{S^2_{11}+S^2_{12}+S^2_{21}+S^2_{22}}$ which give the minimum variance of HT-AℓNNIEE under Assumption 3.2. This applies likewise to other estimators under Assumption 3.2. The proof is given in Appendix B.

Restricting the neighborhood interference assumption (NIA) in (Sussman and Airoldi, 2017) to K nearest neighbors, Assumption 3.1 states that the potential outcome of unit i is only affected by its treatment and by the treatments assigned to its K nearest neighbors such that changing treatments for other units outside the K-neighborhood will not affect the potential outcome of unit i (i.e., for each unit i and for all $j \in \mathcal{N}_{-ik}$, $\delta_j = 0$ on unit i where δ_j is the j^{th} nearest neighbor indirect effect on unit i). On the other hand, Assumption 3.2 states that the indirect effect will be the same across all units' treatment groups. Similarly, the direct effect will be the same across all K-nearest neighbors treatment groups.

3.7 Variance Estimators

We extend the work provided in (Aronow and Samii, 2013, 2017) and (Lohr, 2019) to the K-nearest neighbors interference and provide an estimator for the variance of all estimators in the previous section. In order to estimate the variance of the provided estimators, we estimate all variance and covariance components such that the Horvitz–Thompson estimated variance of the population average potential outcomes under exposure $(W, \mathbf{W}_{N_K})$ as follows.

$$\begin{aligned}\widehat{\mathbf{Var}}_{HT}(\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{N_K})) &= \frac{1}{N^2}\sum_{i\in U} I_i(W, \mathbf{W}_{N_K})[1-\pi_i(W, \mathbf{W}_{N_K})]\left[\frac{Y_i^{obs}(W, \mathbf{W}_{N_K})}{\pi_i(W, \mathbf{W}_{N_K})}\right]^2 \\ &+ \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i} I_i(W, \mathbf{W}_{N_K})I_j(W, \mathbf{W}_{N_K})\frac{[\pi_{ij}(W, \mathbf{W}_{N_K}) - \pi_i(W, \mathbf{W}_{N_K})\pi_j(W, \mathbf{W}_{N_K})]}{\pi_{ij}(W, \mathbf{W}_{N_K})} \\ &\times \frac{Y_i^{obs}(W, \mathbf{W}_{N_K})}{\pi_i(W, \mathbf{W}_{N_K})}\frac{Y_j^{obs}(W, \mathbf{W}_{N_K})}{\pi_j(W, \mathbf{W}_{N_K})}. \qquad (3.42)\end{aligned}$$

If the joint probabilities $\pi_{ij}(W, \mathbf{W}_{N_K}) > 0$ for all i and j, then this estimated variance is unbiased. However, if $\pi_{ij}(W, \mathbf{W}_{N_K}) = 0$ for some i and j, then this estimate of variance will be biased such that $\mathbf{E}(\widehat{\mathbf{Var}}_{HT}(\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{N_K}))) = \mathbf{Var}_{HT}(\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{N_K})) + A$ where $A = \sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}(W, \mathbf{W}_{N_K})=0} y_i(W, \mathbf{W}_{N_K})y_j(W, \mathbf{W}_{N_K})$. Using Young's inequality as derived in Aronow and Samii (2013, 2017), we have the following variance bias correction,

$$\widehat{\mathbf{Var}}_A(\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{N_K})) = \widehat{\mathbf{Var}}_{HT}(\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{N_K})) + \widehat{A}^*(W, \mathbf{W}_{N_K}), \qquad (3.43)$$

where

$$\begin{aligned}\widehat{A}^*(W, \mathbf{W}_{N_K}) = \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}(W, \mathbf{W}_{N_K})=0}&\left[\frac{I_i(W, \mathbf{W}_{N_K})Y_i^{2^{obs}}(W, \mathbf{W}_{N_K})}{2\pi_i(W, \mathbf{W}_{N_K})}\right. \\ &\left. + \frac{I_j(W, \mathbf{W}_{N_K})Y_j^{2^{obs}}(W, \mathbf{W}_{N_K})}{2\pi_j(W, \mathbf{W}_{N_K})}\right]. \qquad (3.44)\end{aligned}$$

Then, $\widehat{\mathbf{Var}}_A(\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{N_K}))$ is a conservative estimator for the variance of the Horvitz–Thompson estimator of the average potential outcomes under exposure $(W, \mathbf{W}_{N_K})$(the proof in Aronow and Samii (2013, 2017) is reproduced in Appendix A).

Moreover, last term in equation 3.14 is unidentified because each unit receives only one exposure and can only be observed under this exposure. Hence, there is no unbiased estimator for the variance of the proposed estimators. However, if the joint probabilities $\pi_{ij}((W, \mathbf{W}_{N_K}), (W', \mathbf{W}'_{N_K})) > 0$ for two different exposures $(W, \mathbf{W}_{N_K})$ and $(W', \mathbf{W}'_{N_K})$ for all i and j, an estimator for the covariance in 3.14 can be as follows.

$$\begin{aligned}
&\widehat{\mathbf{Cov}}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{N_K})) = \\
&\quad \frac{1}{N^2} \sum_{i \in U} \sum_{j \in U, j \neq i} \left[\frac{I_i(W, \mathbf{W}_{N_K}) I_j(W', \mathbf{W}'_{N_K})}{\pi_{ij}((W, \mathbf{W}_{N_K}), (W', \mathbf{W}'_{N_K}))} \frac{Y_i(W, \mathbf{W}_{N_K})}{\pi_i(W, \mathbf{W}_{N_K})} \frac{Y_j(W', \mathbf{W}'_{N_K})}{\pi_j(W', \mathbf{W}'_{N_K})} \right. \\
&\qquad \left. \times [\pi_{ij}((W, \mathbf{W}_{N_K}), (W', \mathbf{W}'_{N_K})) - \pi_i(W, \mathbf{W}_{N_K}) \pi_j(W', \mathbf{W}'_{N_K})] \right] \\
&\quad - \frac{1}{N^2} \sum_{i \in U} \left[\frac{I_i(W, \mathbf{W}_{N_K}) Y_i^{2^{obs}}(W, \mathbf{W}_{N_K})}{2\pi_i(W, \mathbf{W}_{N_K})} + \frac{I_i(W', \mathbf{W}'_{N_K}) Y_i^{2^{obs}}(W', \mathbf{W}'_{N_K})}{2\pi_i(W', \mathbf{W}'_{N_K})} \right], \qquad (3.45)
\end{aligned}$$

where the expected value of the estimated covariance is less than or equal to the true covariance (Aronow and Samii, 2013, 2017).

For the case where the joint probabilities $\pi_{ij}((W, \mathbf{W}_{N_K}), (W', \mathbf{W}'_{N_K})) = 0$ for some i and j, the covariance in 3.14 can be refined as follows:

$$\begin{aligned}
&\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{N_K})) = \\
&\frac{1}{N^2} \sum_{i \in U} \sum_{j \in U, j \neq i : \pi_{ij}((W, \mathbf{W}_{N_K}), (W', \mathbf{W}'_{N_K})) > 0} [\pi_{ij}((W, \mathbf{W}_{N_K}), (W', \mathbf{W}'_{N_K})) - \pi_i(W, \mathbf{W}_{N_K}) \pi_j(W', \mathbf{W}'_{N_K})] \\
&\qquad \times \frac{y_i(W, \mathbf{W}_{N_K})}{\pi_i(W, \mathbf{W}_{N_K})} \frac{y_j(W', \mathbf{W}'_{N_K})}{\pi_j(W', \mathbf{W}'_{N_K})} \\
&\qquad - \frac{1}{N^2} \sum_{i \in U} \sum_{j \in U : \pi_{ij}((W, \mathbf{W}_{N_K}), (W', \mathbf{W}'_{N_K})) = 0} y_i(W, \mathbf{W}_{N_K}) y_j(W', \mathbf{W}'_{N_K}). \qquad (3.46)
\end{aligned}$$

Consequently, the more general covariance estimator is

$$\widehat{\mathbf{Cov}}_A(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{N_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W}'_{N_K}))=$$

$$\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U,j\neq i:\pi_{ij}((W,\mathbf{W}_{N_K}),(W',\mathbf{W}'_{N_K}))>0}\left[\frac{I_i(W,\mathbf{W}_{N_K})I_j(W',\mathbf{W}'_{N_K})}{\pi_{ij}((W,\mathbf{W}_{N_K}),(W',\mathbf{W}'_{N_K}))}\frac{Y_i^{obs}(W,\mathbf{W}_{N_K})}{\pi_i(W,\mathbf{W}_{N_K})}\frac{Y_j^{obs}(W',\mathbf{W}'_{N_K})}{\pi_j(W',\mathbf{W}'_{N_K})}\right.$$
$$\left.\times[\pi_{ij}((W,\mathbf{W}_{N_K}),(W',\mathbf{W}'_{N_K}))-\pi_i(W,\mathbf{W}_{N_K})\pi_j(W',\mathbf{W}'_{N_K})]\right]$$
$$-\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U:\pi_{ij}((W,\mathbf{W}_{N_K}),(W',\mathbf{W}'_{N_K}))=0}\left[\frac{I_i(W,\mathbf{W}_{N_K})Y_i^2}{2\pi_i(W,\mathbf{W}_{N_K})}+\frac{I_j(W',\mathbf{W}'_{N_K})Y_j^2}{2\pi_j(W',\mathbf{W}'_{N_K})}\right]. \quad (3.47)$$

Proposition 3.1.

$$\mathbf{E}(\widehat{\mathbf{Cov}}_A(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{N_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W}'_{N_K}))\leq\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{N_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W}'_{N_K})). \quad (3.48)$$

Since $\widehat{\mathbf{Var}}_A(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{N_K})$ is a conservative variance estimator, the covariance estimator in A.19, provides a conservative variance estimator of any estimator of the the form $\widehat{\delta}=X-Y$ such that $\mathbf{Var}(X-Y)=\mathbf{Var}(X)+\mathbf{Var}(Y)$ -2$\mathbf{Cov}(X,Y)$ which apply to all estimators under Assumption 3.1. However, under Assumption 3.2, we have estimators of the form $\widehat{\delta}=(X-Y)+(W-Z)$ such that $\mathbf{Var}((X-Y)+(W-Z))=\mathbf{Var}(X)+\mathbf{Var}(Y)+\mathbf{Var}(W)+\mathbf{Var}(Z)$ -2$\mathbf{Cov}(X,Y)$ + 2$\mathbf{Cov}(X,W)$ -2$\mathbf{Cov}(X,Z)$ -2$\mathbf{Cov}(Y,W)$ +2$\mathbf{Cov}(Y,Z)$ -2$\mathbf{Cov}(W,Z)$. To get conservative variance estimator of any estimator of the second form, $\widehat{\mathbf{Cov}}_A(X,Y)$ can be used as a lower bound estimator of the covariance (i.e., for covariance components between two averages with negative and positive coefficients) while for positive covariance components (i.e., the covariance between two averages both with positive coefficients or negative coefficients), we need an upper bound covariance estimator that is guaranteed to have expectation greater than or equal to the covariance in 3.14.

We provide the following covariance estimator for the second case,

$$\widehat{\mathbf{Cov}}_B(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{N_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W'}_{N_K}))=$$

$$\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U,j\neq i:\pi_{ij}((W,\mathbf{W}_{N_K}),(W',\mathbf{W'}_{N_K}))>0}\left[\frac{I_i(W,\mathbf{W}_{N_K})I_j(W',\mathbf{W'}_{N_K})}{\pi_{ij}((W,\mathbf{W}_{N_K}),(W',\mathbf{W'}_{N_K}))}\frac{Y_i^{obs}(W,\mathbf{W}_{N_K})}{\pi_i(W,\mathbf{W}_{N_K})}\frac{Y_j^{obs}(W',\mathbf{W'}_{N_K})}{\pi_j(W',\mathbf{W'}_{N_K})}\right.$$

$$\left.\times[\pi_{ij}((W,\mathbf{W}_{N_K}),(W',\mathbf{W'}_{N_K}))-\pi_i(W,\mathbf{W}_{N_K})\pi_j(W',\mathbf{W'}_{N_K})]\right]$$

$$+\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U:\pi_{ij}((W,\mathbf{W}_{N_K}),(W',\mathbf{W'}_{N_K}))=0}\left[\frac{I_i(W,\mathbf{W}_{N_K})Y_i^2}{2\pi_i(W,\mathbf{W}_{N_K})}+\frac{I_j(W',\mathbf{W'}_{N_K})Y_j^2}{2\pi_j(W',\mathbf{W'}_{N_K})}\right]. \quad (3.49)$$

Proposition 3.2.

$$\mathbf{E}(\widehat{\mathbf{Cov}}_B(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{N_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W'}_{N_K}))\geq\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{N_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W'}_{N_K})). \quad (3.50)$$

The proof follows by Young's inequality in equations A.22 and A.23.

Thereby, the conservative variance estimators of all estimators in the previous section under Assumptions 3.1 and 3.2 respectively are as follows,

$$\widehat{\mathbf{Var}}(\hat{\delta}_{HT,tot})=\widehat{\mathbf{Var}}_A(\bar{Y}_{HT}^{obs}(1,\mathbf{1}))+\widehat{\mathbf{Var}}_A(\bar{Y}_{HT}^{obs}(0,\mathbf{0}))$$

$$-2\widehat{\mathbf{Cov}}_A(\bar{Y}_{HT}^{obs}(1,\mathbf{1}),\bar{Y}_{HT}^{obs}(0,\mathbf{0})). \quad (3.51)$$

$$\widehat{\mathbf{Var}}(\hat{\delta}_{HT,dir})=\widehat{\mathbf{Var}}_A(\bar{Y}_{HT}^{obs}(1,\mathbf{1}))+\widehat{\mathbf{Var}}_A(\bar{Y}_{HT}^{obs}(0,\mathbf{1}))$$

$$-2\widehat{\mathbf{Cov}}_A(\bar{Y}_{HT}^{obs}(1,\mathbf{1}),\bar{Y}_{HT}^{obs}(0,\mathbf{1})). \quad (3.52)$$

$$\begin{aligned}\widehat{\mathbf{Var}}(\widehat{\delta}_{HT,ind}) = \widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{1})) + \widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{0}))& \\ - 2\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})).&\end{aligned} \tag{3.53}$$

$$\begin{aligned}\widehat{\mathbf{Var}}(\widehat{\delta}_{HT,\ell}) = \widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell})) + \widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1}))& \\ - 2\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell}), \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1}).&\end{aligned} \tag{3.54}$$

$$\widehat{\mathbf{Var}}(\widehat{\delta^*}_{HT,tot}) = \widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) + \widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})). \tag{3.55}$$

$$\begin{aligned}\widehat{\mathbf{Var}}(\widehat{\delta^*}_{HT,dir}) = \frac{1}{4}\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) + \frac{1}{4}\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{1}))& \\ + \frac{1}{4}\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{0})) + \frac{1}{4}\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{0}))& \\ - 2(\frac{1}{4})\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{1})) + 2(\frac{1}{4})\widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(1,\mathbf{0}))& \\ - 2(\frac{1}{4})\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2(\frac{1}{4})\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{1}), \bar{Y}^{obs}_{HT}(1,\mathbf{0}))& \\ + 2(\frac{1}{4})\widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(0,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2(\frac{1}{4})\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{0}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})).&\end{aligned} \tag{3.56}$$

$$
\begin{aligned}
\widehat{\mathbf{Var}}(\widehat{\delta^*}_{HT,ind}) = &\frac{1}{4}\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) + \frac{1}{4}\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{0})) \\
&+ \frac{1}{4}\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{1})) + \frac{1}{4}\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) \\
&- 2(\frac{1}{4})\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{1}),\bar{Y}^{obs}_{HT}(1,\mathbf{0})) + 2(\frac{1}{4})\widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(1,\mathbf{1}),\bar{Y}^{obs}_{HT}(0,\mathbf{1})) \\
&- 2(\frac{1}{4})\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{1}),\bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2(\frac{1}{4})\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{0}),\bar{Y}^{obs}_{HT}(0,\mathbf{1})) \\
&+ 2(\frac{1}{4})\widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(1,\mathbf{0}),\bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2(\frac{1}{4})\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{1}),\bar{Y}^{obs}_{HT}(0,\mathbf{0})).
\end{aligned}
\tag{3.57}
$$

$$
\begin{aligned}
\widehat{\mathbf{Var}}(\widehat{\delta^*}_{HT,\ell}) = &\frac{1}{4}\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell)) + \frac{1}{4}\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})) \\
&+ \frac{1}{4}\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)) + \frac{1}{4}\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) \\
&- 2(\frac{1}{4})\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell),\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})) + 2(\frac{1}{4})\widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell),\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)) \\
&- 2(\frac{1}{4})\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell),\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) - 2(\frac{1}{4})\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1}),\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)) \\
&+ 2(\frac{1}{4})\widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1}),\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) - 2(\frac{1}{4})\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell),\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})).
\end{aligned}
\tag{3.58}
$$

By linearity of expectation, Propositions A.1 and A.2 and Equation A.15 in Appendix A, the covariance estimators provided here are conservative. The conservative estimator of $\mathbf{Var}(\bar{Y}^{obs}_{HT}(W_i, W_{\mathcal{N}_{ik}}))$ and proofs are provided in Appendix A.

3.8 Simulation

In this section, we assess the performance of our proposed estimators through a simulation study. We consider three different scenarios of the indirect effects where the indirect effects in each scenario represent the degree of interference starting from no interference in the first three models scenario, followed by weak interference in the second three models scenario and moderate interference in the last three models scenario. In each scenario, the direct effect takes also three different scenarios such that δ_i takes the values in (0,1,4).

We generate responses under the following KNNIM model:

$$Y_i = X_1 + X_2 + X_3 + \delta_1 W_1 + \delta_2 W_2 + \delta_3 W_3 + \delta_{dir} W_i \tag{3.59}$$

where the covariates $X_j \sim N(0,1)$, $j = 1,2,3$, and we consider the case where we only have $K = 3$ nearest neighbors. We assess our estimators under two randomization designs: completely randomized design (CRD) where half of the N units are assigned to treatment, completely at random and Bernoulli randomization design (BR) with probability p = 0.5.

The Nine interference models are shown in the second column of Table 3.1. The first three elements of the vector $(\delta_1, \delta_2, \delta_3, \delta_{dir})$ in Table 3.1 represent the first, second, and third nearest neighbor's indirect effect where the last element is the unit's direct effect.

In all models considered, the closer the distance to unit i, the greater the indirect effect: $|\delta_1| \geq |\delta_2 \geq |\delta_3|$.

In each model, we evaluate the performance of the total, direct, indirect and ℓ_{th} nearest neighbor estimators with the estimated variance under Assumption 3.1 and Assumption 3.2. The experiment is replicated repeatedly 1000 times with sample size N = 256. The marginal and joint probabilities are computed as in equations 3.23, 3.25, 3.26 and 3.27. The empirical expected value of the estimates (Emp.Estimates), empirical variance (Emp.Var) and standard deviation (Emp.S.D.), and the of the estimated variance (Var Estimate) are computed. The results are illustrated for CRD in table 3.2 for model 1 to table 3.10 for model 9 and for BR in table 3.11 for model 1 to table 3.19 for model 9. The comparison of CRD and BR results under Assumptions 3.1 and 3.2 are illustrated in Figures 3.1 to 3.9.

Table 3.1: *Interference Models*

Models	$(\delta_{1^{st}},\, \delta_{2^{nd}},\, \delta_{3^{rd}},\, \delta_{dir})$
Model 1	(0,0,0,0)
Model 2	(0,0,0,1)
Model 3	(0,0,0,4)
Model 4	(2,1,0.5,0)
Model 5	(2,1,0.5,1)
Model 6	(2,1,0.5,4)
Model 7	(3,2,1,0)
Model 8	(3,2,1,1)
Model 9	(3,2,1,4)

Table 3.2: *Estimates Under Completely Randomized Design Model 1*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	0	0.0779	1.1668	1.0802	1.1759
$\widehat{\delta^*}_{HT,tot}$	0	0.0779	1.1668	1.0802	1.1759
$\widehat{\delta}_{HT,dir}$	0	0.0337	0.6078	0.7796	0.6683
$\widehat{\delta^*}_{HT,dir}$	0	0.0303	0.2894	0.5380	0.5036
$\widehat{\delta}_{HT,ind}$	0	0.0442	0.9404	0.9697	0.9393
$\widehat{\delta^*}_{HT,ind}$	0	0.0476	0.5615	0.7493	0.6619
$\widehat{\delta}_{HT,1^{st}}$	0	0.0346	0.6273	0.7920	0.6821
$\widehat{\delta^*}_{HT,1^{st}}$	0	0.0191	0.2647	0.5145	0.3950
$\widehat{\delta}_{HT,2^{nd}}$	0	-0.0111	0.4642	0.6813	0.6023
$\widehat{\delta^*}_{HT,2^{nd}}$	0	-0.0021	0.2694	0.5190	0.4625
$\widehat{\delta}_{HT,3^{rd}}$	0	0.0207	0.4541	0.6738	0.6076
$\widehat{\delta^*}_{HT,3^{rd}}$	0	0.0306	0.3014	0.5490	0.4514

Table 3.3: *Estimates Under Completely Randomized Design Model 2*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	1	1.0707	1.306	1.1430	1.3314
$\widehat{\delta^*}_{HT,tot}$	1	1.0707	1.306	1.1430	1.3314
$\widehat{\delta}_{HT,dir}$	1	1.0265	0.7273	0.8528	0.8050
$\widehat{\delta^*}_{HT,dir}$	1	1.0300	0.3377	0.5812	0.6214
$\widehat{\delta}_{HT,ind}$	0	0.0442	0.9404	0.9697	0.9393
$\widehat{\delta^*}_{HT,ind}$	0	0.0406	0.6081	0.7798	0.7381
$\widehat{\delta}_{HT,1^{st}}$	0	0.0346	0.6273	0.7920	0.6821
$\widehat{\delta^*}_{HT,1^{st}}$	0	0.01758	0.3009	0.5486	0.4749
$\widehat{\delta}_{HT,2^{nd}}$	0	-0.0111	0.4642	0.6813	0.6023
$\widehat{\delta^*}_{HT,2^{nd}}$	0	-0.0052	0.3328	0.5769	0.5601
$\widehat{\delta}_{HT,3^{rd}}$	0	0.0207	0.4541	0.6738	0.6076
$\widehat{\delta^*}_{HT,3^{rd}}$	0	0.0282	0.3525	0.5937	0.5359

Table 3.4: *Estimates Under Completely Randomized Design Model 3*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	4	4.0489	2.9225	1.7095	3.3552
$\widehat{\delta^*}_{HT,tot}$	4	4.0489	2.9225	1.7095	3.3552
$\widehat{\delta}_{HT,dir}$	4	4.0047	2.2827	1.5108	2.5258
$\widehat{\delta^*}_{HT,dir}$	4	4.0292	0.9267	0.9626	2.1155
$\widehat{\delta}_{HT,ind}$	4	0.0442	0.9404	0.9697	0.9393
$\widehat{\delta^*}_{HT,ind}$	0	0.01978	1.1883	1.0901	1.7146
$\widehat{\delta}_{HT,1^{st}}$	0	0.0346	0.6273	0.7920	0.6821
$\widehat{\delta^*}_{HT,1^{st}}$	0	0.01286	0.8246	0.9081	1.4569
$\widehat{\delta}_{HT,2^{nd}}$	0	-0.0111	0.4642	0.6813	0.6023
$\widehat{\delta^*}_{HT,2^{nd}}$	0	-0.01445	1.0133	1.0066	1.7389
$\widehat{\delta}_{HT,3^{rd}}$	0	0.0207	0.4541	0.6738	0.6076
$\widehat{\delta^*}_{HT,3^{rd}}$	0	0.0213	1.0711	1.0349	1.6299

Table 3.5: *Estimates Under Completely Randomized Design Model 4*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	3.5	3.5526	2.5285	1.5901	2.8557
$\widehat{\delta^*}_{HT,tot}$	3.5	3.5526	2.5285	1.5901	2.8557
$\widehat{\delta}_{HT,dir}$	0	-0.0123	2.9637	1.7215	3.2323
$\widehat{\delta^*}_{HT,dir}$	0	0.0073	0.8913	0.9441	1.5485
$\widehat{\delta}_{HT,ind}$	3.5	3.5649	1.5750	1.2550	2.3108
$\widehat{\delta^*}_{HT,ind}$	3.5	3.5453	0.9523	0.9758	1.6828
$\widehat{\delta}_{HT,1^{st}}$	2	2.0414	0.8733	0.9345	1.1102
$\widehat{\delta^*}_{HT,1^{st}}$	2	2.0261	0.3901	0.6245	0.7582
$\widehat{\delta}_{HT,2^{nd}}$	1	1.0003	1.4039	1.1848	1.9204
$\widehat{\delta^*}_{HT,2^{nd}}$	1	0.9961	0.7423	0.8615	1.4553
$\widehat{\delta}_{HT,3^{rd}}$	0.5	0.5230	1.8611	1.3642	2.5928
$\widehat{\delta^*}_{HT,3^{rd}}$	0.5	0.5230	1.1157	1.0563	1.8726

Table 3.6: *Estimates Under Completely Randomized Design Model 5*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	4.5	4.5453	3.3664	1.8347	3.9196
$\widehat{\delta^*}_{HT,tot}$	4.5	4.5453	3.3664	1.8347	3.9196
$\widehat{\delta}_{HT,dir}$	1	0.9804	3.9013	1.9751	4.1313
$\widehat{\delta^*}_{HT,dir}$	1	1.0070	1.1503	1.0725	1.9781
$\widehat{\delta}_{HT,ind}$	3.5	3.5649	1.5750	1.2550	2.3108
$\widehat{\delta^*}_{HT,ind}$	3.5	3.5383	1.1485	1.0717	1.9574
$\widehat{\delta}_{HT,1^{st}}$	2	2.0414	0.8733	0.9345	1.1102
$\widehat{\delta^*}_{HT,1^{st}}$	2	2.0245	0.4873	0.6981	0.9667
$\widehat{\delta}_{HT,2^{nd}}$	1	1.0003	1.4039	1.1848	1.9204
$\widehat{\delta^*}_{HT,2^{nd}}$	1	0.9930	0.9747	0.9873	1.8982
$\widehat{\delta}_{HT,3^{rd}}$	0.5	0.5230	1.8611	1.3642	2.5928
$\widehat{\delta^*}_{HT,3^{rd}}$	0.5	0.5207	1.4523	1.2051	2.3927

Table 3.7: *Estimates Under Completely Randomized Design Model 6*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	7.5	7.5236	7.0767	2.6602	8.6687
$\widehat{\delta^*}_{HT,tot}$	7.5	7.5236	7.0767	2.6602	8.6687
$\widehat{\delta}_{HT,dir}$	4	3.9586	7.9109	2.8126	8.1389
$\widehat{\delta^*}_{HT,dir}$	4	4.0061	2.3713	1.5399	4.4075
$\widehat{\delta}_{HT,ind}$	3.5	3.5649	1.5750	1.2550	2.3108
$\widehat{\delta^*}_{HT,ind}$	3.5	3.5174	2.1777	1.4757	3.5289
$\widehat{\delta}_{HT,1^{st}}$	2	2.0414	0.8733	0.9345	1.1102
$\widehat{\delta^*}_{HT,1^{st}}$	2	2.0198	1.1942	1.0928	2.3348
$\widehat{\delta}_{HT,2^{nd}}$	1	1.0003	1.4039	1.1848	1.9204
$\widehat{\delta^*}_{HT,2^{nd}}$	1	0.9838	2.1622	1.4704	4.1131
$\widehat{\delta}_{HT,3^{rd}}$	0.5	0.5230	1.8611	1.3642	2.5928
$\widehat{\delta^*}_{HT,3^{rd}}$	0.5	0.5138	3.0275	1.7399	4.7936

Table 3.8: *Estimates Under Completely Randomized Design Model 7*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	6	6.0344	4.9972	2.2354	6.0021
$\widehat{\delta^*}_{HT,tot}$	6	6.0344	4.9972	2.2354	6.0021
$\widehat{\delta}_{HT,dir}$	0	-0.0452	7.4190	2.7237	7.9636
$\widehat{\delta^*}_{HT,dir}$	0	-0.0091	2.0143	1.4192	3.4718
$\widehat{\delta}_{HT,ind}$	6	6.0797	2.7909	1.6705	4.850
$\widehat{\delta^*}_{HT,ind}$	6	6.0436	1.6675	1.2913	3.5938
$\widehat{\delta}_{HT,1^{st}}$	3	3.0449	1.1582	1.0762	1.6049
$\widehat{\delta^*}_{HT,1^{st}}$	3	3.0295	0.5412	0.7357	1.1811
$\widehat{\delta}_{HT,2^{nd}}$	2	2.0092	2.8500	1.6882	3.9616
$\widehat{\delta^*}_{HT,2^{nd}}$	2	1.9961	1.4097	1.1873	2.9671
$\widehat{\delta}_{HT,3^{rd}}$	1	1.0256	4.5078	2.1231	6.1172
$\widehat{\delta^*}_{HT,3^{rd}}$	1	1.0179	2.6139	1.6167	4.4043

Table 3.9: *Estimates Under Completely Randomized Design Model 8*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	7	7.0272	6.3337	2.5166	7.7149
$\widehat{\delta}^*{}_{HT,tot}$	7	7.0272	6.3337	2.5166	7.7149
$\widehat{\delta}_{HT,dir}$	1	0.9474	8.9409	2.9901	9.4071
$\widehat{\delta}^*{}_{HT,dir}$	1	0.9905	2.4239	1.5568	4.1241
$\widehat{\delta}_{HT,ind}$	6	6.0797	2.7909	1.6705	4.8503
$\widehat{\delta}^*{}_{HT,ind}$	6	6.0366	1.9707	1.4038	4.0101
$\widehat{\delta}_{HT,1^{st}}$	3	3.0449	1.1582	1.0762	1.6049
$\widehat{\delta}^*{}_{HT,1^{st}}$	3	3.0280	0.6690	0.8179	1.4541
$\widehat{\delta}_{HT,2^{nd}}$	2	2.0092	2.8500	1.6882	3.9616
$\widehat{\delta}^*{}_{HT,2^{nd}}$	2	1.9930	1.7408	1.3194	3.6167
$\widehat{\delta}_{HT,3^{rd}}$	1	1.0256	4.5078	2.1231	6.1172
$\widehat{\delta}^*{}_{HT,3^{rd}}$	1	1.0156	3.1489	1.7745	5.2252

Table 3.10: *Estimates Under Completely Randomized Design Model 9*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	10	10.0055	11.54004	3.3970	14.4106
$\widehat{\delta}^*{}_{HT,tot}$	10	10.0055	11.54004	3.3970	14.4106
$\widehat{\delta}_{HT,dir}$	4	3.9257	14.7035	3.8345	15.0482
$\widehat{\delta}^*{}_{HT,dir}$	4	3.9896	4.0963	2.0239	7.2216
$\widehat{\delta}_{HT,ind}$	6	6.0797	2.7909	1.6705	4.8503
$\widehat{\delta}^*{}_{HT,ind}$	6	6.0158	3.3206	1.8222	6.0066
$\widehat{\delta}_{HT,1^{st}}$	3	3.0449	1.1582	1.0762	1.6049
$\widehat{\delta}^*{}_{HT,1^{st}}$	3	3.0232	1.4675	1.2114	3.0152
$\widehat{\delta}_{HT,2^{nd}}$	2	2.0092	2.8500	1.6882	3.9616
$\widehat{\delta}^*{}_{HT,2^{nd}}$	2	1.9838	3.2241	1.7955	6.4517
$\widehat{\delta}_{HT,3^{rd}}$	1	1.0256	4.5078	2.1231	6.1172
$\widehat{\delta}^*{}_{HT,3^{rd}}$	1	1.0086	5.3193	2.3063	8.5285

Table 3.11: *Estimates Under Bernoulli Randomization Model 1*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	0	0.0059	1.2480	1.1171	1.2037
$\widehat{\delta^*}_{HT,tot}$	0	0.0059	1.2480	1.1171	1.2037
$\widehat{\delta}_{HT,dir}$	0	-0.01509	0.6850	0.8277	0.7016
$\widehat{\delta^*}_{HT,dir}$	0	0.0103	0.3116	0.5582	0.5153
$\widehat{\delta}_{HT,ind}$	0	0.0210	0.8232	0.9073	0.9258
$\widehat{\delta^*}_{HT,ind}$	0	-0.0044	0.5622	0.7498	0.6696
$\widehat{\delta}_{HT,1^{st}}$	0	0.0502	0.6502	0.8063	0.6804
$\widehat{\delta^*}_{HT,1^{st}}$	0	0.01735	0.2460	0.4960	0.3922
$\widehat{\delta}_{HT,2^{nd}}$	0	-0.0181	0.4430	0.6656	0.5985
$\widehat{\delta^*}_{HT,2^{nd}}$	0	-0.0085	0.2702	0.5198	0.4599
$\widehat{\delta}_{HT,3^{rd}}$	0	-0.0110	0.4489	0.6700	0.6048
$\widehat{\delta^*}_{HT,3^{rd}}$	0	-0.0132	0.3134	0.5599	0.4629

Table 3.12: *Estimates Under Bernoulli Randomization Model 2*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	1	1.0326	1.4420	1.2008	1.4406
$\widehat{\delta^*}_{HT,tot}$	1	1.0326	1.4420	1.2008	1.4406
$\widehat{\delta}_{HT,dir}$	1	1.0116	0.8786	0.9373	0.8960
$\widehat{\delta^*}_{HT,dir}$	1	1.0277	0.3573	0.5978	0.6357
$\widehat{\delta}_{HT,ind}$	0	0.0210	0.8232	0.9073	0.9258
$\widehat{\delta^*}_{HT,ind}$	0	0.0049	0.6462	0.8038	0.7994
$\widehat{\delta}_{HT,1^{st}}$	0	0.0502	0.6502	0.8063	0.6804
$\widehat{\delta^*}_{HT,1^{st}}$	0	0.0142	0.2858	0.5346	0.4783
$\widehat{\delta}_{HT,2^{nd}}$	0	-0.0181	0.4430	0.6656	0.5985
$\widehat{\delta^*}_{HT,2^{nd}}$	0	-0.0177	0.3250	0.5700	0.5625
$\widehat{\delta}_{HT,3^{rd}}$	0	-0.0110	0.4489	0.6700	0.6048
$\widehat{\delta^*}_{HT,3^{rd}}$	0	0.0084	0.3846	0.6202	0.5497

Table 3.13: *Estimates Under Bernoulli Randomization Model 3*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	4	4.1129	4.0302	2.0075	4.5346
$\widehat{\delta^*}_{HT,tot}$	4	4.1129	4.0302	2.0075	4.5346
$\widehat{\delta}_{HT,dir}$	4	4.0919	3.4654	1.8615	3.6069
$\widehat{\delta^*}_{HT,dir}$	4	4.0797	0.9582	0.9789	2.2095
$\widehat{\delta}_{HT,ind}$	0	0.0210	0.8232	0.9073	0.9258
$\widehat{\delta^*}_{HT,ind}$	0	0.0331	1.8081	1.3446	2.3816
$\widehat{\delta}_{HT,1^{st}}$	0	0.0502	0.6502	0.8063	0.6804
$\widehat{\delta^*}_{HT,1^{st}}$	0	0.0051	0.8466	0.9201	1.5260
$\widehat{\delta}_{HT,2^{nd}}$	0	-0.0181	0.4430	0.6656	0.5985
$\widehat{\delta^*}_{HT,2^{nd}}$	0	-0.0456	1.0226	1.0112	1.7932
$\widehat{\delta}_{HT,3^{rd}}$	0	-0.0110	0.4489	0.6700	0.6048
$\widehat{\delta^*}_{HT,3^{rd}}$	0	0.0737	1.2557	1.1205	1.7099

Table 3.14: *Estimates Under Bernoulli Randomization Model 4*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	3.5	3.5995	3.3899	1.8411	3.7707
$\widehat{\delta^*}_{HT,tot}$	3.5	3.5995	3.3899	1.8411	3.7707
$\widehat{\delta}_{HT,dir}$	0	0.0745	3.1656	1.7792	3.4406
$\widehat{\delta^*}_{HT,dir}$	0	0.0552	0.9018	0.9496	1.6046
$\widehat{\delta}_{HT,ind}$	3.5	3.5249	1.8276	1.3519	2.5207
$\widehat{\delta^*}_{HT,ind}$	3.5	3.5443	1.5038	1.2262	2.1947
$\widehat{\delta}_{HT,1^{st}}$	2	2.0532	0.9882	0.9940	1.1739
$\widehat{\delta^*}_{HT,1^{st}}$	2	2.0206	0.3871	0.6222	0.7689
$\widehat{\delta}_{HT,2^{nd}}$	1	0.9878	1.4253	1.1938	2.0036
$\widehat{\delta^*}_{HT,2^{nd}}$	1	0.9675	0.8350	0.9137	1.5577
$\widehat{\delta}_{HT,3^{rd}}$	0.5	0.4838	2.2543	1.5014	2.8237
$\widehat{\delta^*}_{HT,3^{rd}}$	0.5	0.5561	1.5671	1.2518	2.1355

Table 3.15: *Estimates Under Bernoulli Randomization Model 5*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	4.5	4.6263	4.7542	2.1804	5.3978
$\widehat{\delta^*}_{HT,tot}$	4.5	4.6263	4.7542	2.1804	5.3978
$\widehat{\delta}_{HT,dir}$	1	1.1013	4.3757	2.0918	4.6502
$\widehat{\delta^*}_{HT,dir}$	1	1.0725	1.1720	1.0826	2.0765
$\widehat{\delta}_{HT,ind}$	3.5	3.5249	1.8276	1.3519	2.5207
$\widehat{\delta^*}_{HT,ind}$	3.5	3.5537	2.0192	1.4210	2.7832
$\widehat{\delta}_{HT,1^{st}}$	2	2.0532	0.9882	0.9940	1.1739
$\widehat{\delta^*}_{HT,1^{st}}$	2	2.0176	0.4648	0.6818	0.9666
$\widehat{\delta}_{HT,2^{nd}}$	1	0.9878	1.4253	1.1938	2.0036
$\widehat{\delta^*}_{HT,2^{nd}}$	1	0.9582	1.1082	1.0527	2.0400
$\widehat{\delta}_{HT,3^{rd}}$	0.5	0.4838	2.2543	1.5014	2.8237
$\widehat{\delta^*}_{HT,3^{rd}}$	0.5	0.5778	2.0343	1.4263	2.7262

Table 3.16: *Estimates Under Bernoulli Randomization Model 6*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	7.5	7.7065	10.8531	3.2944	12.6623
$\widehat{\delta^*}_{HT,tot}$	7.5	7.7065	10.8531	3.2944	12.6623
$\widehat{\delta}_{HT,dir}$	4	4.1815	10.0122	3.1642	10.4070
$\widehat{\delta^*}_{HT,dir}$	4	4.1245	2.4463	1.5640	4.7047
$\widehat{\delta}_{HT,ind}$	3.5	3.5249	1.8276	1.3519	2.5207
$\widehat{\delta^*}_{HT,ind}$	3.5	3.5819	4.4755	2.1155	5.7415
$\widehat{\delta}_{HT,1^{st}}$	2	2.0532	0.9882	0.9940	1.1739
$\widehat{\delta^*}_{HT,1^{st}}$	2	2.0084	1.1391	1.0672	2.3490
$\widehat{\delta}_{HT,2^{nd}}$	1	0.9878	1.4253	1.1938	2.0036
$\widehat{\delta^*}_{HT,2^{nd}}$	1	0.9304	2.4611	1.5687	4.4098
$\widehat{\delta}_{HT,3^{rd}}$	0.5	0.4838	2.2543	1.5014	2.8237
$\widehat{\delta^*}_{HT,3^{rd}}$	0.5	0.6431	4.0935	2.0232	5.3983

Table 3.17: *Estimates Under Bernoulli Randomization Model 7*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	6	6.1664	7.4275	2.7253	8.5832
$\widehat{\delta^*}_{HT,tot}$	6	6.1664	7.4275	2.7253	8.5832
$\widehat{\delta}_{HT,dir}$	0	0.1386	7.8388	2.7997	8.5881
$\widehat{\delta^*}_{HT,dir}$	0	0.0872	2.0487	1.4313	3.6387
$\widehat{\delta}_{HT,ind}$	6	6.0277	3.5973	1.8966	5.4461
$\widehat{\delta^*}_{HT,ind}$	6	6.0791	3.2309	1.7974	5.0241
$\widehat{\delta}_{HT,1^{st}}$	3	3.0547	1.3709	1.1708	1.7503
$\widehat{\delta^*}_{HT,1^{st}}$	3	3.0223	0.5510	0.7423	1.2101
$\widehat{\delta}_{HT,2^{nd}}$	2	1.9923	2.9090	1.7056	4.1353
$\widehat{\delta^*}_{HT,2^{nd}}$	2	1.9521	1.6716	1.2929	3.2181
$\widehat{\delta}_{HT,3^{rd}}$	1	0.9806	5.3764	2.3187	6.7556
$\widehat{\delta^*}_{HT,3^{rd}}$	1	1.1047	3.8051	1.9506	5.1576

Table 3.18: *Estimates Under Bernoulli Randomization Model 8*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	7	7.1931	9.6277	3.1028	11.2033
$\widehat{\delta^*}_{HT,tot}$	7	7.1931	9.6277	3.1028	11.2033
$\widehat{\delta}_{HT,dir}$	1	1.1654	9.7750	3.1265	10.5230
$\widehat{\delta^*}_{HT,dir}$	1	1.1045	2.4792	1.5745	4.3616
$\widehat{\delta}_{HT,ind}$	6	6.0277	3.5973	1.8966	5.4461
$\widehat{\delta^*}_{HT,ind}$	6	6.0885	4.0546	2.0136	5.9402
$\widehat{\delta}_{HT,1^{st}}$	3	3.0547	1.3709	1.1708	1.7503
$\widehat{\delta^*}_{HT,1^{st}}$	3	3.0192	0.6476	0.8047	1.4636
$\widehat{\delta}_{HT,2^{nd}}$	2	1.9923	2.9090	1.7056	4.1353
$\widehat{\delta^*}_{HT,2^{nd}}$	2	1.9428	2.0746	1.4403	3.9275
$\widehat{\delta}_{HT,3^{rd}}$	1	0.9806	5.3764	2.3187	6.7556
$\widehat{\delta^*}_{HT,3^{rd}}$	1	1.1264	4.5503	2.1331	6.0993

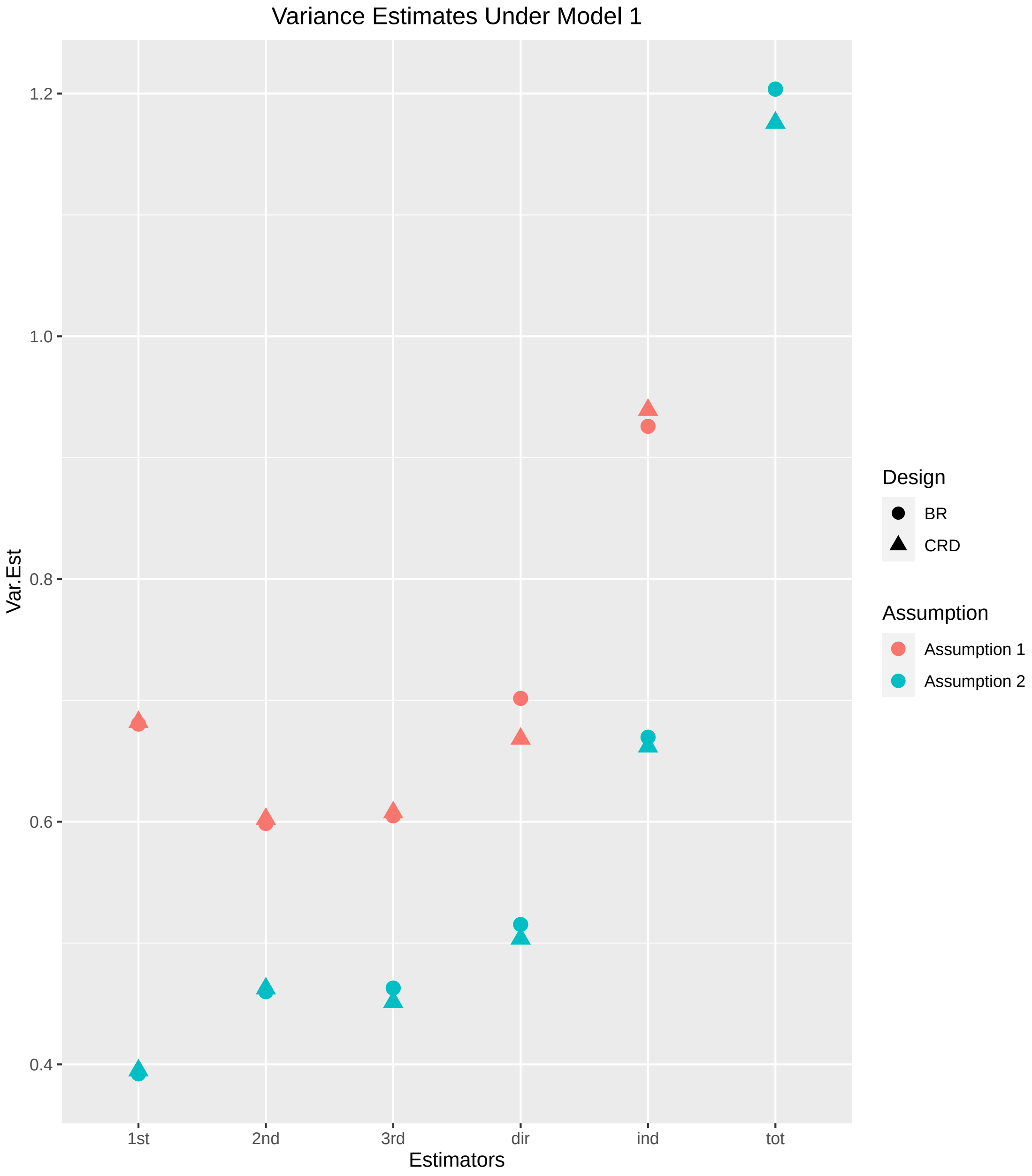


Figure 3.1: *Variance estimates for model 1 of all estimators under the K-nearest neighbors interference assumption and no-interaction between direct and indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 1,000 randomizations.*

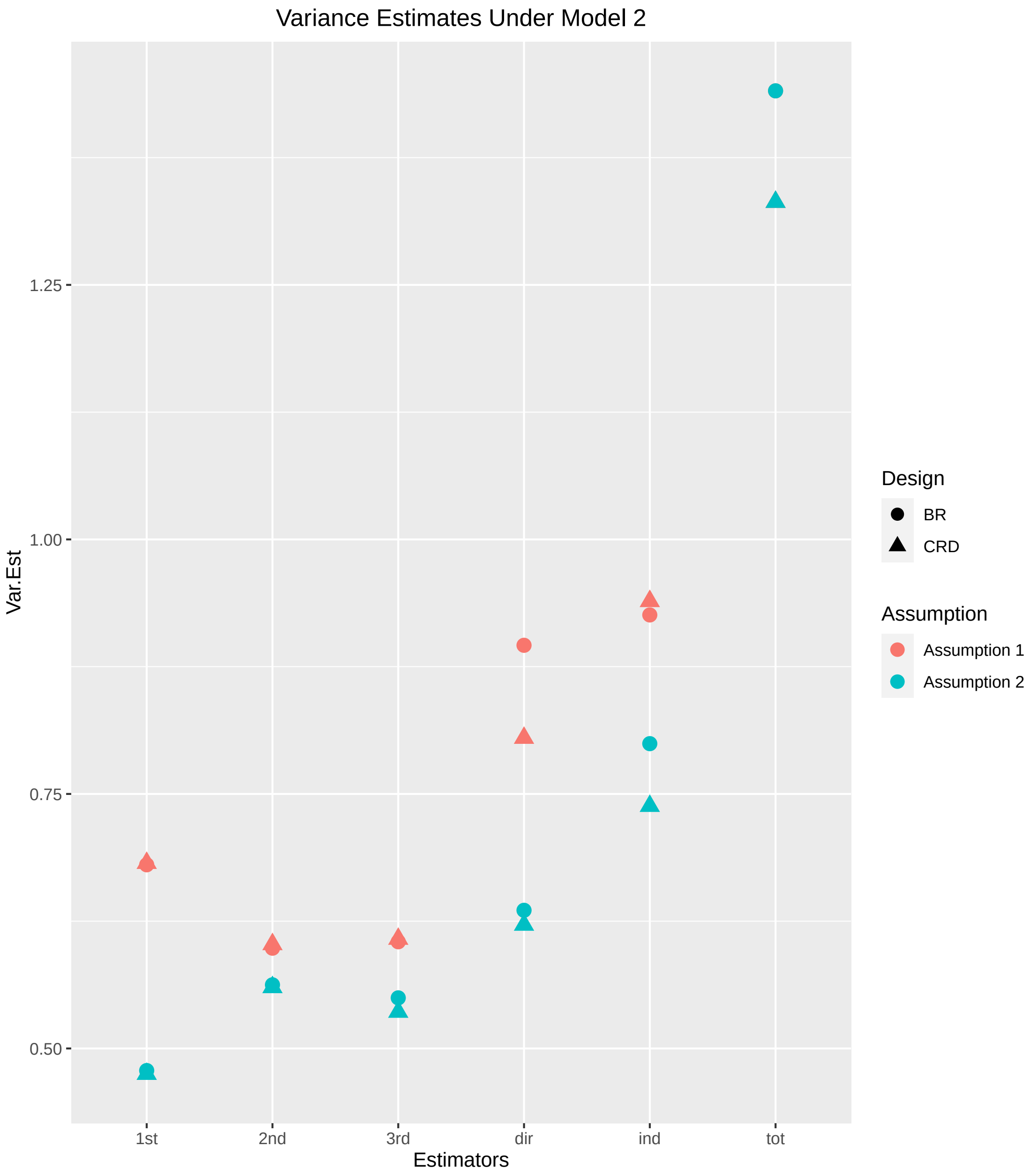


Figure 3.2: *Variance estimates for model 2 of all estimators under the* K*-nearest neighbors interference assumption and no-interaction between direct and indirect effects assumption. We use* $N = 256$ *units and* $K = 3$ *nearest neighbors. The effects are estimated using 1,000 randomizations.*

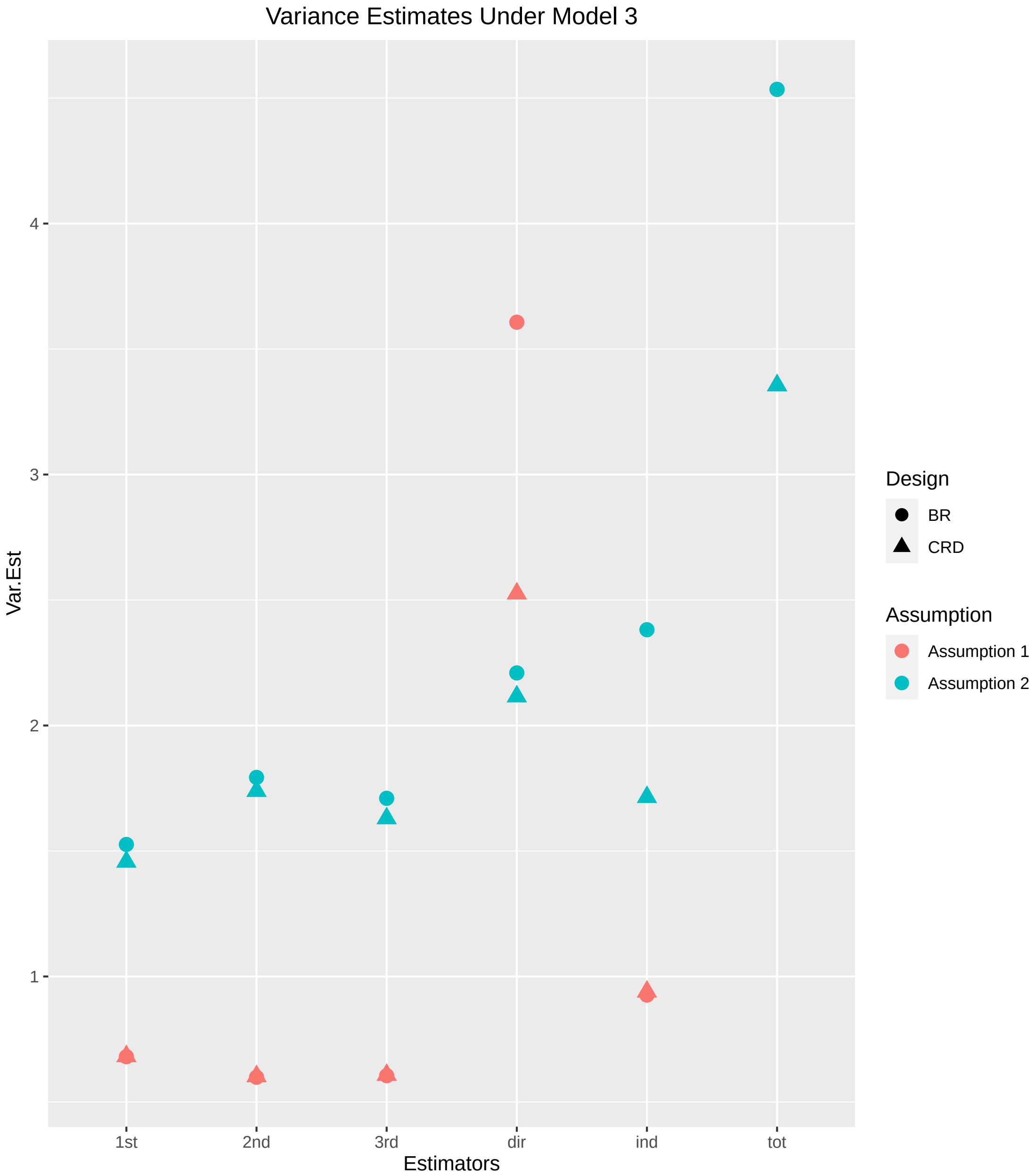


Figure 3.3: *Variance estimates for model 3 of all estimators under the K-nearest neighbors interference assumption and no-interaction between direct and indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 1,000 randomizations.*

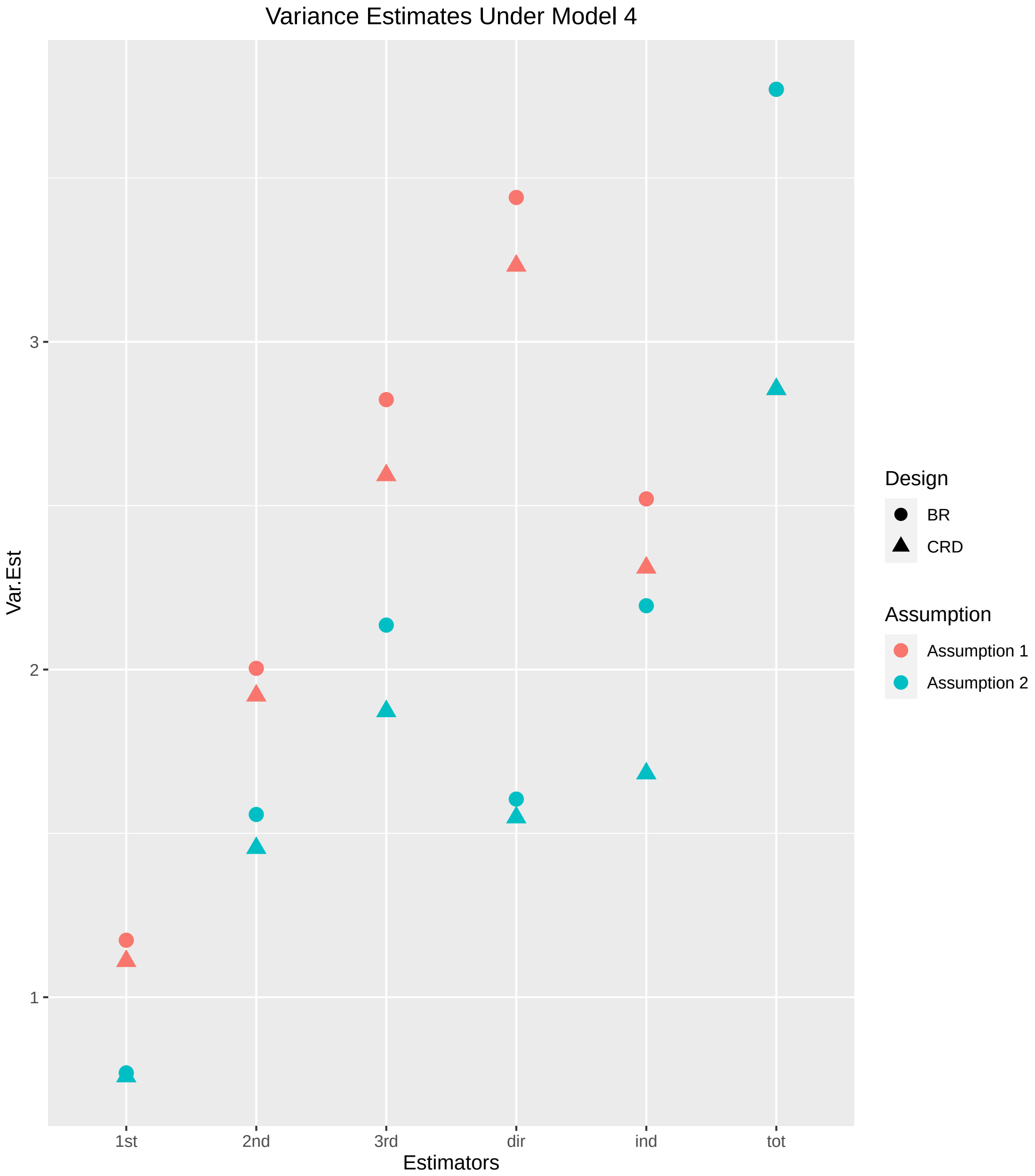


Figure 3.4: *Variance estimates for model 4 of all estimators under the K-nearest neighbors interference assumption and no-interaction between direct and indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 1,000 randomizations.*

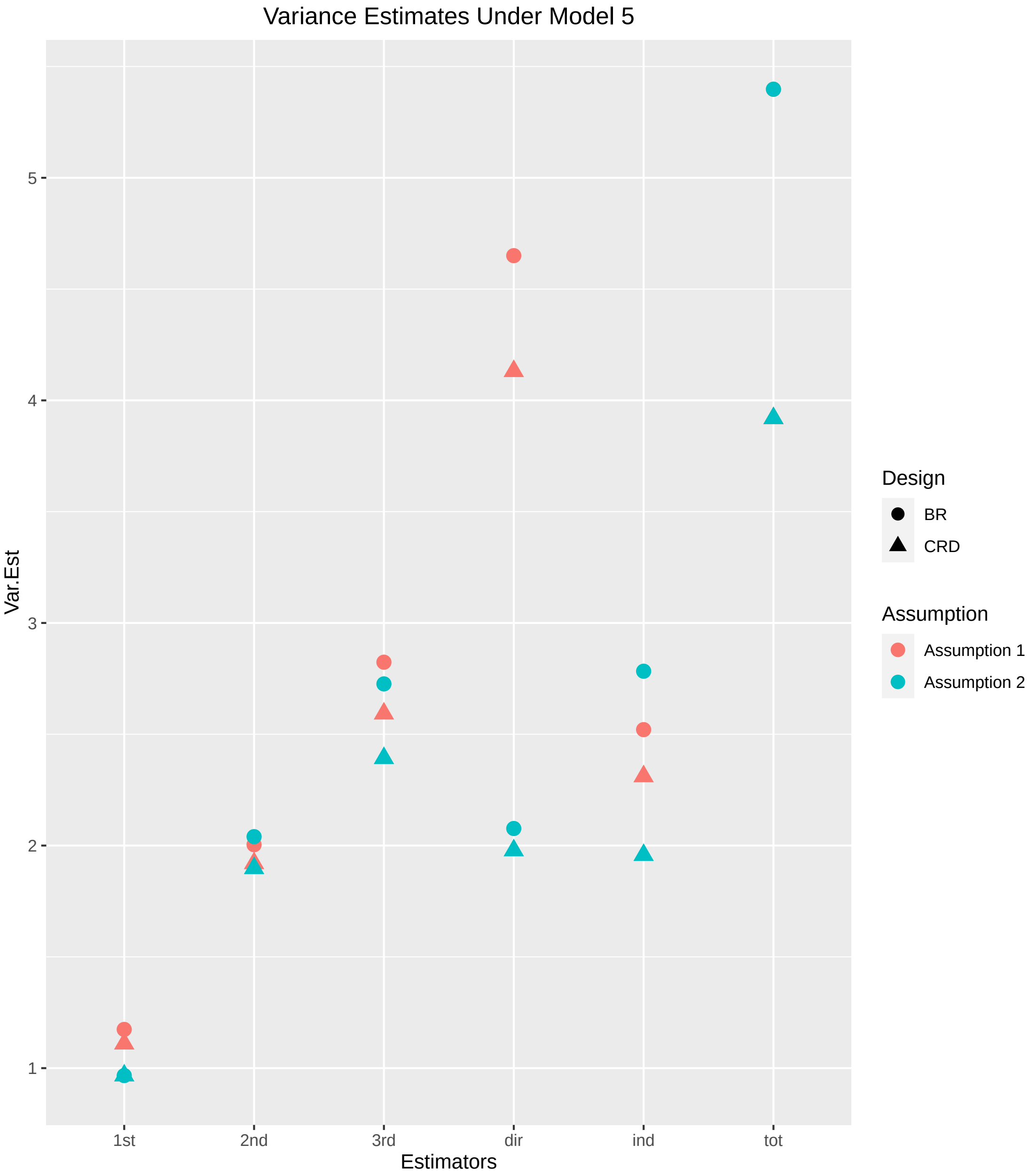


Figure 3.5: *Variance estimates for model 5 of all estimators under the K-nearest neighbors interference assumption and no-interaction between direct and indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 1,000 randomizations.*

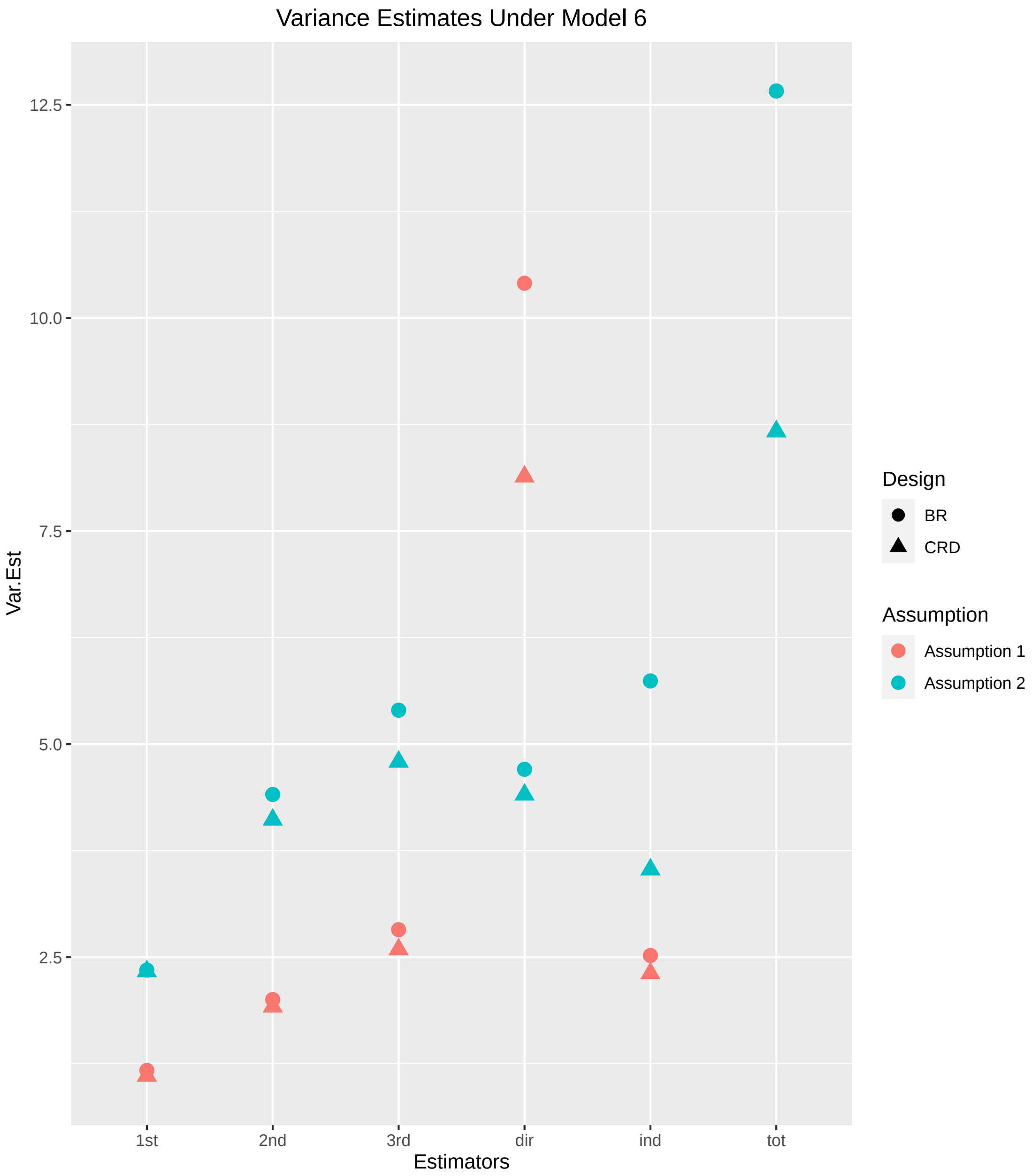


Figure 3.6: *Variance estimates for model 6 of all estimators under the K-nearest neighbors interference assumption and no-interaction between direct and indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 1,000 randomizations.*

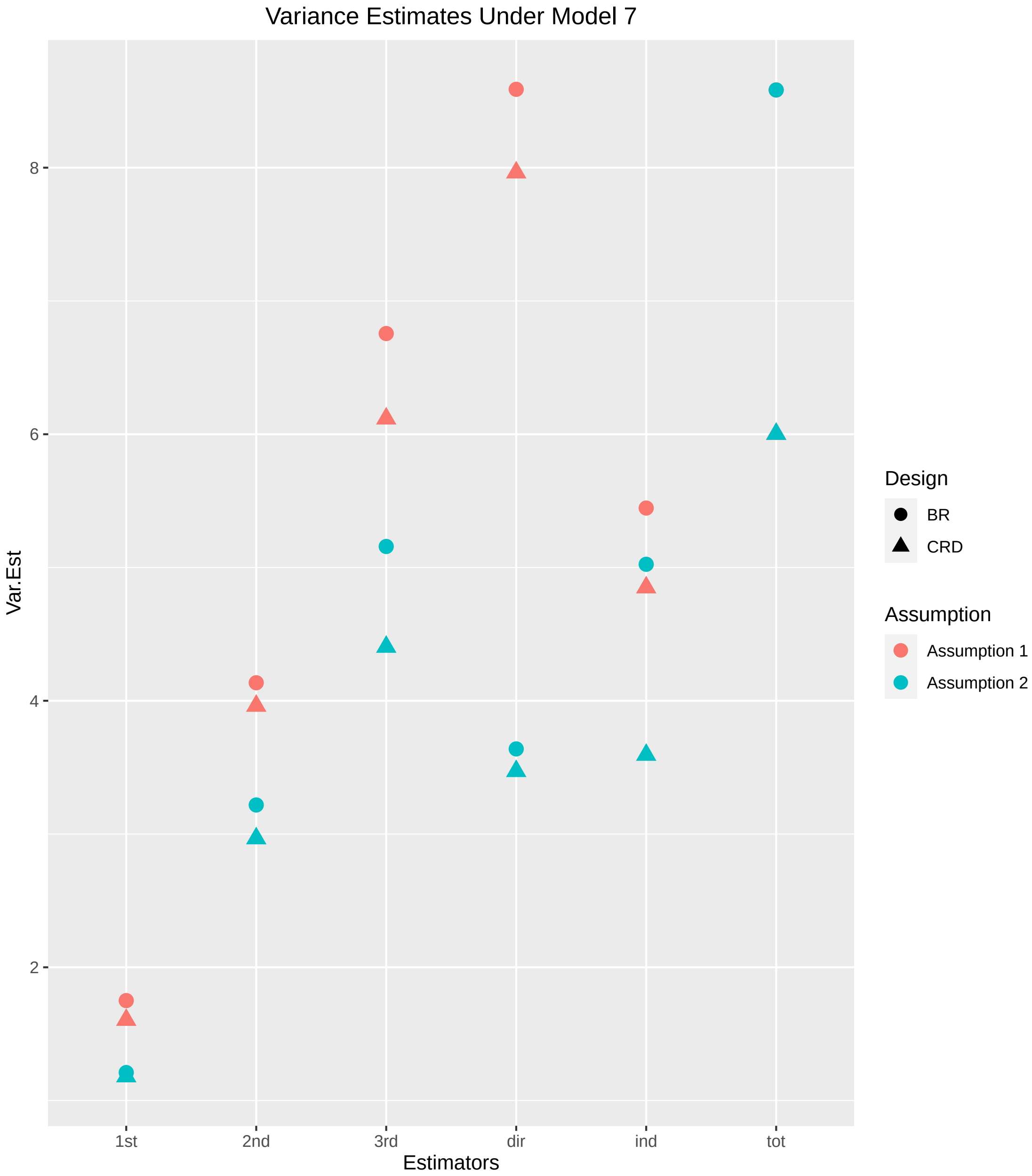


Figure 3.7: *Variance estimates for model 7 of all estimators under the K-nearest neighbors interference assumption and no-interaction between direct and indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 1,000 randomizations.*

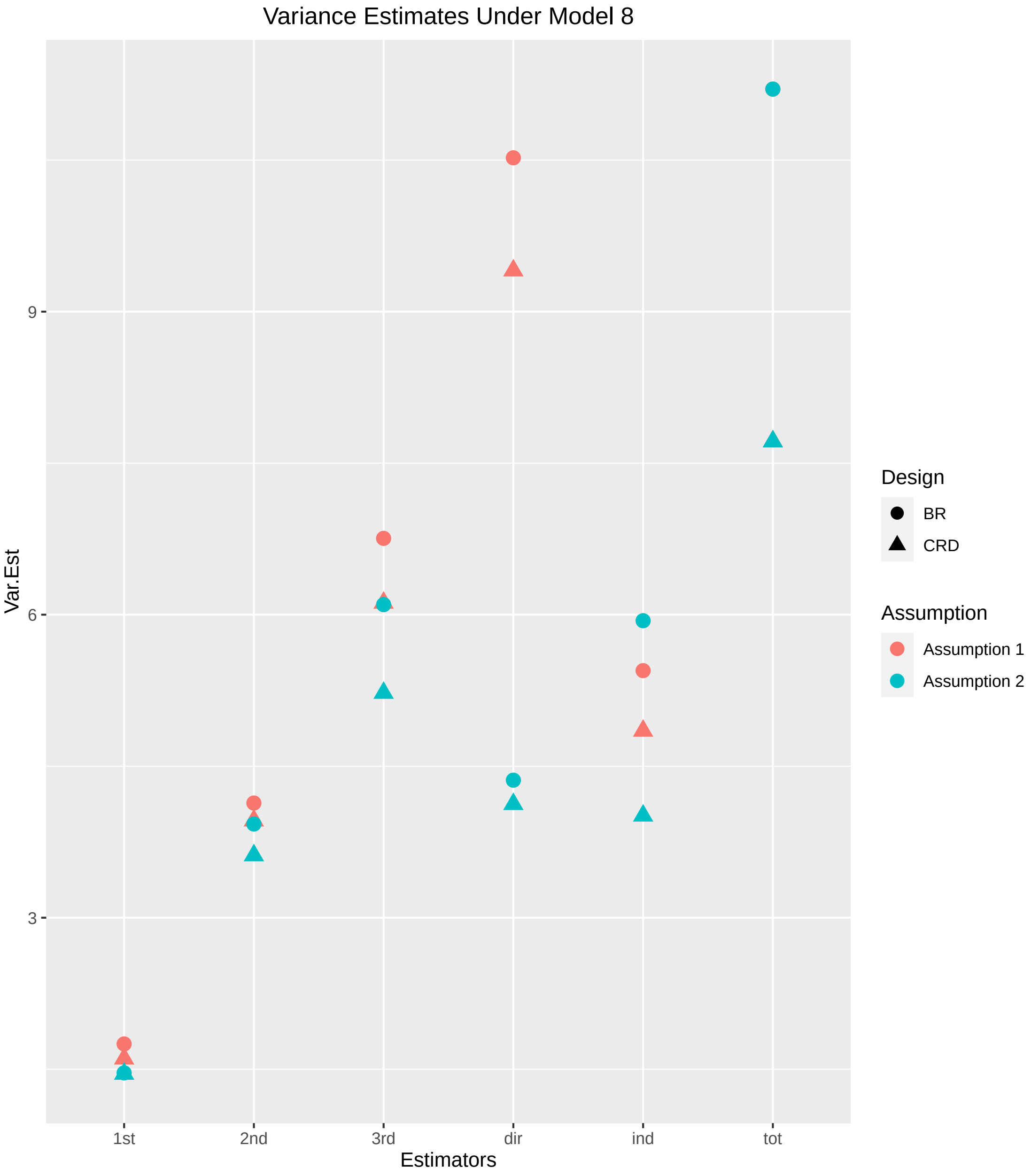


Figure 3.8: *Variance estimates for model 8 of all estimators under the K-nearest neighbors interference assumption and no-interaction between direct and indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 1,000 randomizations.*

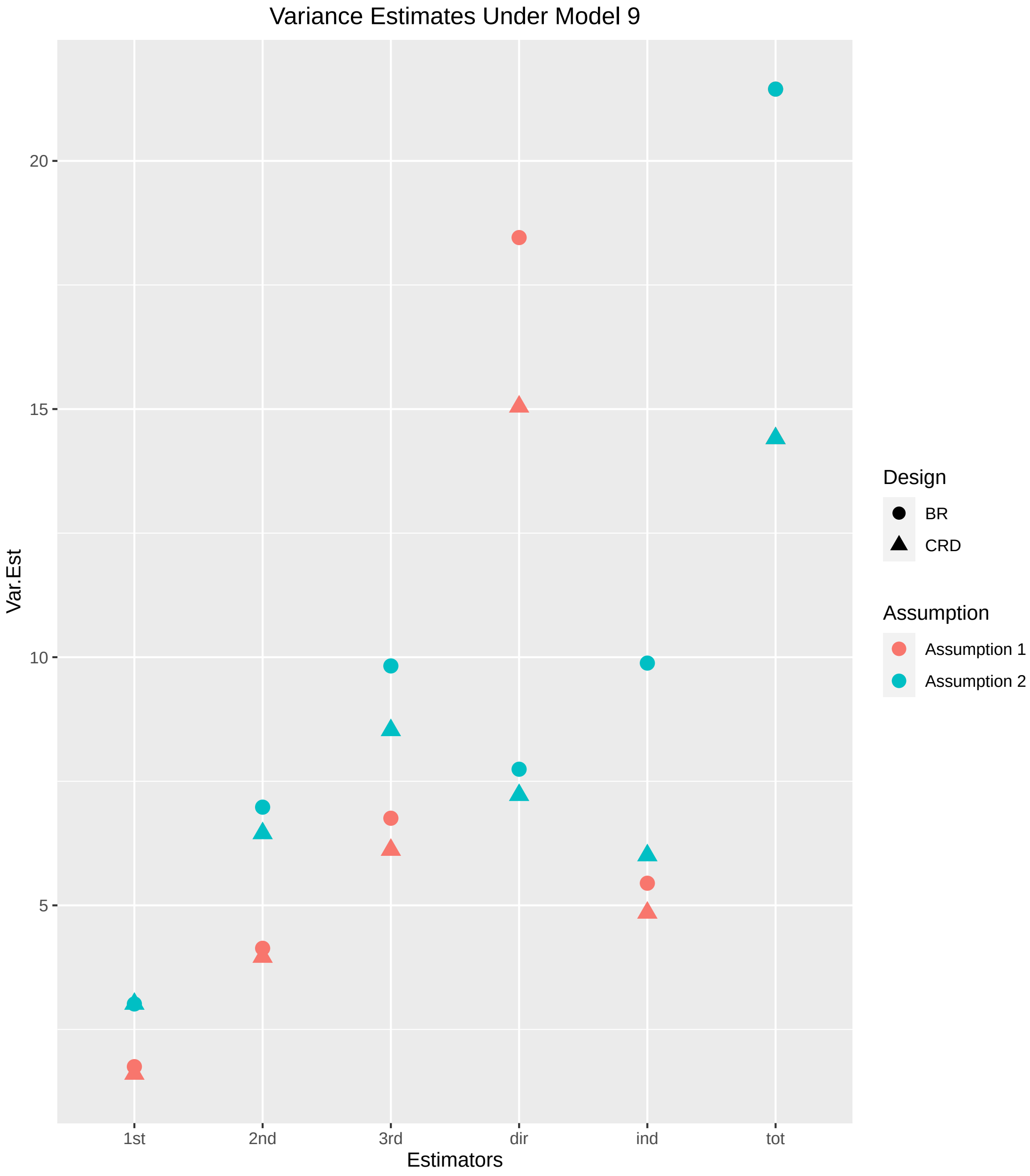


Figure 3.9: *Variance estimates for model 9 of all estimators under the K-nearest neighbors interference assumption and no-interaction between direct and indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 1,000 randomizations.*

Table 3.19: *Estimates Under Bernoulli Randomization Model 9*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate
$\widehat{\delta}_{HT,tot}$	10	10.2734	18.2343	4.2701	21.4468
$\widehat{\delta^*}_{HT,tot}$	10	10.2734	18.2343	4.2701	21.4468
$\widehat{\delta}_{HT,dir}$	4	4.2456	17.5899	4.1940	18.4553
$\widehat{\delta^*}_{HT,dir}$	4	4.1566	4.2346	2.0578	7.7429
$\widehat{\delta}_{HT,ind}$	6	6.0277	3.5973	1.8966	5.4461
$\widehat{\delta^*}_{HT,ind}$	6	6.1168	7.4354	2.7267	9.8815
$\widehat{\delta}_{HT,1^{st}}$	3	3.0547	1.3709	1.1708	1.7503
$\widehat{\delta^*}_{HT,1^{st}}$	3	3.0100	1.3786	1.1741	3.0134
$\widehat{\delta}_{HT,2^{nd}}$	2	1.9923	2.9090	1.7056	4.1353
$\widehat{\delta^*}_{HT,2^{nd}}$	2	1.9150	3.8171	1.9537	6.9785
$\widehat{\delta}_{HT,3^{rd}}$	1	0.9806	5.3764	2.3187	6.7556
$\widehat{\delta^*}_{HT,3^{rd}}$	1	1.1917	7.4434	2.7282	9.8241

3.9 Anti-Conflict Program: An Analysis of Social Network Experiment

In this section, we analyze a field experiment conducted on 56 schools for changing climates of conflict among middle schools students in New Jersey (Paluck et al., 2016). In this multilevel experiment, 28 out of 56 schools were randomly assigned to receive an anti-conflict program that aims to reduce conflicts among adolescents and to understand how the effects on participants transmit to their social peers. In each treated school, 20-32 students were nonrandomly selected as eligible students in which 50% of those eligible students were randomly assigned to participate in the anti-conflict intervention blocked by gender and grade. Following three weeks of the start of school, students in each school were asked to nominate up to 10 students with whom they chose to spent time with in the last few weeks either in school, out of school or online where number 1 is the first person a student spent most time with, number 2 is the second person a student spent most time with and so on. These nominations are used to measure the social connections between students providing the adjacency graph. Every two weeks over the course of the year, trained research assistants in each

Table 3.20: *Number of units in each exposure of Anti-Conflict Program Experiment with K =2 and N= 348*

	Indirect			
Direct	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
Treated	38	42	39	34
Control	40	59	46	50

Table 3.21: *Estimates of Anti-Conflict Program with K = 2 for only Treated School N = 348*

Estimator	Estimates	S.E.
$\widehat{\delta}_{HT,Total}$	0.1899	0.0985
$\widehat{\delta^*}_{HT,Total}$	0.1899	0.0985
$\widehat{\delta}_{HT,direct}$	-0.0254	0.1332
$\widehat{\delta^*}_{HT,direct}$	0.0559	0.0863
$\widehat{\delta}_{HT,indirect}$	0.2154	0.0927
$\widehat{\delta^*}_{HT,indirect}$	0.1340	0.0781
$\widehat{\delta}_{HT,1^{st}}$	0.1788	0.0822
$\widehat{\delta^*}_{HT,1^{st}}$	0.1019	0.0683
$\widehat{\delta}_{HT,2^{nd}}$	0.0365	0.1148
$\widehat{\delta^*}_{HT,2^{nd}}$	0.0320	0.0934

treated school had held meetings with participants to identify conflict types in their school and discuss strategies that encourage participants to reduce conflict behaviour among other students. At the end of the school year, a survey was conducted to measure conflict norms that measures multiple outcomes. In this analysis, we focus on one particular response in which students self reported if they had worn an orange wristband issued and distributed to students reflecting their attitudes of anti-conflict norms. We restrict our analysis to eligible students in treated school who nominated at least two eligible friends ($N = 348$ with $K = 2$) and we assume that the carried randomization design was completely at random with 153 treated students. There are $2^{K+1} = 8$ exposures and number of students for each of the eight exposures in this dataset are illustrated in table 3.20. The marginal and the joint inclusion probabilities are computed using equations 3.23and 3.25. Results are presented in table 3.21.

3.10 Discussion

Under completely randomized design, results of the first and second models in the first scenario where there is no indirect effects show that all estimators under Assumption 3.1 and Assumption 3.2 are unbiased. All variance estimates are greater than the empirical variance for all estimators under both Assumption 3.1 and Assumption 3.2. The variance estimates are smaller for all estimators under Assumption 3.2 than under Assumption 3.1 (Tables 3.2, 3.3). However, when we increase the direct effect, the variance estimates for the indirect effects estimators under Assumption 1 are about 50% smaller than under Assumption 3.2 (Table 3.4).

Similarly, under weak and moderate interference scenarios, all estimators are unbiased and all variance estimates are greater than the empirical variance for all estimators under Assumption 3.1 and Assumption 3.2. In addition, the variance estimates are smaller for all estimators under Assumption 3.2 than under Assumption 3.1 except model 3, model 6 and model 9 when the direct effect is relatively large and close to the overall indirect effects (Tables 3.5, 3.6, 3.7, 3.8, 3.9 and 3.10).

Moreover, Assumption 3.2 has improved the variance estimates of the direct effect estimator for all indirect effect scenarios specially when the direct effect increases, the variance estimates get smaller.

All previous results of completely randomized apply to Bernoulli randomization except that Assumption 3.2 didn't improve standard errors of the indirect effects in models 5 and model 8. However, under Bernoulli randomization, all variance estimates for all estimators in all scenarios seem to be larger than those of the completely randomized design (Tables 3.11 to table 3.19 and Figures 3.1 to 3.9). Hence, completely randomized design is preferable over Bernoulli randomization for this type of data.

For anti-conflict study, results of Assumption 3.1 suggest that the estimate of indirect effect is about 21% increase in the probability of wearing a wrist band. The estimate of the total effect (direct + indirect) of the anti-conflict program on eligible students with at least two eligible nearest neighbors is 18% increase in the probability of wearing a wrist band while the direct effect has smaller estimate on the probability of wearing a wrist band. On the other hand, under Assumption 3.2, the estimate of the direct effect is about 5% increase in the probability of wearing a wrist band while the indirect effect has an estimate about 13% increase in the probability of wearing a wrist band. The difference between the estimates of the indirect effects under Assumptions 3.1 and 3.2 indicates that Assumption 3.2 might be violated in this data. The first nominated eligible friend with whom a student chose to spend most time with has an effect estimated about 10 to 18% increase in the probability of wearing a wrist band where it is 3% for the second nominated eligible friend under both Assumptions 3.1 and Assumption 3.2. All estimates of standard errors of all estimators are reduced under Assumption 3.2.

3.11 Conclusion

Causal inference in the presence of interference has been a trend in the past decade. We extended the potential outcomes approach and the K-nearest neighbors interference framework and we defined causal effects under the K-neighborhood assumption. We provided unbiased estimators of the defined estimands and derived properties of the proposed estimators under both K-neighborhood and the no-interaction between direct and indirect effects assumptions. We uncover indirect effects of the K-nearest neighbors which has not been studied in previous work. Assumption 3.2 achieved better results with respect to estimation precision than Assumption 3.1 specifically, for smaller direct effects. An extension to this work could be an improvement of the estimation standard errors, specially for large direct effects. Another future work is to obtain an optimal design for the K-nearest neighbors interference.

Chapter 4

Improving Estimation of Causal Effects under K-Nearest Neighbors Interference

4.1 Introduction

Classical causal inference approaches assume that treatment assigned to one unit only affects the outcome of this unit (direct effect) and doesn't affect other units outcomes (indirect effect) (Cox, 1958; Rubin, 1980). The common term used in causal inference approaches to describe this setting is *interference*. Failure to account for interference when it is present may result in wrong conclusions about the treatment effectiveness; Sobel (2006) demonstrated this risk.

Interference is common in settings where a social factor is present. For example, in infectious disease studies, the risk of catching an infectious disease depends on the vaccination status of others (Halloran and Struchiner, 1995). In educational studies, the behavior and attitude of all students in schools can be affected by the behavior of a few students; the effects of communication and educational interventions may extend to other students through social connections (Paluck et al., 2016). Moreover, in voting behavior studies, an individual's decision on whether to vote might be impacted by the attitudes of others in the same network (Bond et al., 2012).

Traditionally, treatment interference is considered a nuisance and mitigated through designs that combine highly connected individuals into the same group and the analysis is performed on the group level (Ugander et al., 2013; Gui et al., 2015; Eckles et al., 2016).

However, estimating the indirect effects of treatment interference has become the focus of many researchers in the past decade. Typically, the estimation of the indirect effects is done through assumptions that restrict the extent of interference to a limited form. One example is *partial interference* (Sobel, 2006), where interference is allowed within disjointed groups but not across groups. However, the partial interference assumption does not always hold. Another interference structure assumes that interference is allowed within a limited number of individuals in a neighborhood.

Previously, we presented the *K-nearest neighbors interference* framework to extend causal inference methodologies and limit interference between units to their K-neighborhood. In this model, we account for the strength of the relationship within the neighborhood in which each individual is affected by its direct treatment and by treatments assigned to its K nearest neighbors.

In Chapter 3, we define direct, indirect and total effects under this framework and we provide estimators of the defined estimands. This chapter extends the estimation of treatment effects under the K-nearest neighbors interference to improve estimation standard errors under the assumption of no interaction between indirect effects. This allows for the inclusion of more units, which increases the amount of information and may improve estimation precision. We propose estimators under this assumption and derive properties of the estimators. We also provide conservative variance estimates of the proposed estimators. The provided methods are demonstrated through a simulation study to evaluate and compare their performance to the methods referenced in the previous chapter.

The rest of this chapter is organized as follows. Section 4.2 provides estimators with the derived properties of causal effects defined in the previous chapter. Conservative variance estimators are given in Section 4.3. Simulation and discussion on the performance of the proposed methods are given in Sections 4.4 and 4.5. We conclude in Section 4.6.

4.2 Estimation of Causal Effects

In this section, we extend Assumption 3.1 and 3.2 and provide new estimators under the following assumption:

Assumption 4.1. *(No Interaction between Indirect Effects Assumption). For each unit i in a network G and for $j \in \mathcal{N}_{ik} \cup i$, the potential outcome of unit i is a linear combination of the unit's treatment and the neighborhood treatments,*

$$y_i(W_i, W_{\mathcal{N}_{ik}}) = y_i(0, \mathbf{0}) + \sum_{j \in \mathcal{N}_{ik} \cup i} \beta_{ij} W_j. \tag{4.1}$$

Note that Assumption 4.1 inherently assumes that the contribution of the treatment effect of the ℓth nearest neighbor towards the potential outcome for unit i only depends on the treatment status of this neighbor and is unaffected by neither the treatment status of unit i nor any other neighbor of unit i. This is a strictly stronger assumption than Assumption 3.2. The baseline $y_i(0, \mathbf{0})$ is the potential outcome of unit i when unit i and its K nearest neighbors receive the control treatment condition, which includes the covariates and the noise.

Each unit i under the KNNIM, has 2^{K+1} possible exposures to treatment, hence 2^{K+1} potential outcomes. Following Aronow and Samii (2017), the unbiased Horvitz–Thompson estimator of the average potential outcomes of units under any exposure $(W, \mathbf{W}_{N_K})$ is

$$\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}) = \frac{1}{N} \sum_{i=1}^{N} I_i(W, \mathbf{W}_{N_K}) \frac{Y_i^{obs}(W, \mathbf{W}_{N_K})}{\pi_i(W, \mathbf{W}_{N_K})}, \tag{4.2}$$

The $\pi_i(W, \mathbf{W}_{N_K})$ and $\pi_{ij}((W, \mathbf{W}_{N_K}), (W', \mathbf{W}'_{N_K}))$ terms as defined in 3.5.1 are the marginal and joint probabilities of units i and j under exposure $(W, \mathbf{W}_{N_K})$ and $(W', \mathbf{W}'_{N_K})$ for completely randomized design (CRD).

Under Assumption 4.1, for $\sum_{e=1}^{2^K} C_e = \sum_{e=1}^{2^K} \frac{1}{2^K} = 1$, the unbiased Horvitz–Thompson estimator of ADE is as follows.

$$\widehat{\delta^{**}}_{HT,dir} = \sum_{e=1}^{2^K} C_e[\bar{Y}_{HT}^{obs}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}) - \bar{Y}_{HT}^{obs}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}})] \tag{4.3}$$

Theorem 4.1. *Under the no-interaction between the indirect effects assumption,*

$$\mathbf{E}(\widehat{\delta^{**}}_{HT,dir}) = \delta_{dir}. \tag{4.4}$$

$$\begin{aligned}\mathbf{Var}(\widehat{\delta^{**}}_{HT,dir}) = & \sum_{e,e',W_i=1} C_e C_{e'} \mathbf{Cov}(\bar{Y}_{HT}^{obs}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}_{HT}^{obs}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\ & + \sum_{e,e',W_i=0} C_e C_{e'} \mathbf{Cov}(\bar{Y}_{HT}^{obs}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}_{HT}^{obs}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\ & - 2 \sum_{e,e',W_i=1,W_i=0} C_e C_{e'} \mathbf{Cov}(\bar{Y}_{HT}^{obs}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}_{HT}^{obs}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})).\end{aligned} \tag{4.5}$$

Let $Y_i^{obs} = Y_i^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}})$ be the observed outcome of unit i that receives treatment W_i with its ℓth nearest neighbor being treated (i.e., $W_\ell = 1$) and the rest of its KNN's receive treatment $\mathbf{W_{e,K-1}}$. Similarly, let $Y_{ei}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})$ be the observed outcome of unit i that receives treatment W_i with its ℓth nearest neighbor being control (i.e., $W_\ell = 0$) and the rest of its KNN's receive treatment $\mathbf{W_{e,K-1}}$.

Under Assumption 4.1, an unbiased Horvitz–Thompson estimator of AℓNNIE estimand becomes as follows.

$$\widehat{\delta^{**}}_{HT,\ell} = \sum_{e=1}^{2^K} \frac{1}{2^K}[\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})]. \tag{4.6}$$

Note that $\mathbf{W_{e,K-1}}$ is fixed in the two averages in $[\bar{Y}_{e,HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}_{e,HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})]$ where we have 2^{K-1} different exposures for the K-1 nearest neighbors of unit i and 2^K different exposures if we account for the direct treatment on unit i. Under Assumption 4.1, we weighted each difference of the averages with the corresponding $\mathbf{W_{e,K-1}}$ by $\frac{1}{2^K}$ such that $\sum_{e=1}^{2^K} C_{\ell e} = \sum_{e=1}^{2^K} \frac{1}{2^K} = 1$. For example, if we have only two nearest neighbors i.e., $K = 2$, then we will have two differences for $W_i = 0$ and two differences for $W_i = 1$ such that

$$\begin{aligned}\widehat{\delta^{**}}_{HT,\ell} = {} & \frac{1}{4}[\bar{Y}_{HT}^{obs}(0, W_\ell = 1, \mathbf{W_{e,K-1}} = 1) - \bar{Y}_{HT}^{obs}(0, W_\ell = 0, \mathbf{W_{e,K-1}} = 1)] \\ & + \frac{1}{4}[\bar{Y}_{HT}^{obs}(0, W_\ell = 1, \mathbf{W_{e,K-1}} = 0) - \bar{Y}_{HT}^{obs}(0, W_\ell = 0, \mathbf{W_{e,K-1}} = 0)] \\ & + \frac{1}{4}[\bar{Y}_{HT}^{obs}(1, W_\ell = 1, \mathbf{W_{e,K-1}} = 1) - \bar{Y}_{HT}^{obs}(1, W_\ell = 0, \mathbf{W_{e,K-1}} = 1)] \\ & + \frac{1}{4}[\bar{Y}_{HT}^{obs}(1, W_\ell = 1, \mathbf{W_{e,K-1}} = 0) - \bar{Y}_{HT}^{obs}(1, W_\ell = 0, \mathbf{W_{e,K-1}} = 0)]. \quad (4.7)\end{aligned}$$

Unbiasedness, and the theoretical variance of HT-AℓNNIEE estimator in Equation 4.6, are then provided in the following theorem.

Theorem 4.2. *Under the no-interaction between the indirect effects assumption,*

$$\mathbf{E}(\widehat{\delta^{**}}_{HT,\ell}) = \delta_\ell. \quad (4.8)$$

$$\begin{aligned}\mathbf{Var}(\widehat{\delta^{**}}_{HT,\ell}) = {} & \sum_{e,e',W_\ell=1} C_{\ell e} C_{\ell e'} \mathbf{Cov}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e',K-1}})) \\ & + \sum_{e,e',W_\ell=0} C_{\ell e} C_{\ell e'} \mathbf{Cov}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}}), \bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e',K-1}})) \\ & - 2 \sum_{e,e',W_\ell=1,W_\ell=0} C_{\ell e} C_{\ell e'} \mathbf{Cov}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e',K-1}})). \quad (4.9)\end{aligned}$$

For $\sum_{e=1}^{2^K} C_{\ell e} = \sum_{e=1}^{2^K} \frac{1}{2^K} = 1$, the unbiased Horvitz–Thompson estimator of AIE under the no-interaction between the indirect effects assumption is provided in the following definition.

Definition 4.1.

$$\widehat{\delta^{**}}_{HT,ind} = \sum_{\ell=1}^{K} \widehat{\delta^{**}}_{HT,\ell} = \sum_{\ell=1}^{K}\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})] \tag{4.10}$$

Theorem 4.3. *Under the no-interaction between the indirect effects assumption,*

$$\mathbf{E}(\widehat{\delta^{**}}_{HT,ind}) = \delta_{ind}. \tag{4.11}$$

$$\begin{aligned}
\mathbf{Var}(\widehat{\delta^{**}}_{HT,ind}) = & \sum_{e,e',W_\ell = W_{\ell'}=1} \sum_{\ell,\ell'=1}^{K} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 1, \mathbf{W_{e',K-1}})) \\
+ & \sum_{e,e',W_\ell = W_{\ell'}=0} \sum_{\ell,\ell'=1}^{K} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e',K-1}})) \\
-2 & \sum_{e,e',W_\ell = 1, W_{\ell'}=0} \sum_{\ell,\ell'=1}^{K} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e',K-1}})).
\end{aligned} \tag{4.12}$$

Under Assumption 4.1, for $\sum_{e=1}^{2^K} C_e = \sum_{e=1}^{2^K} C_{\ell e} = \sum_{e=1}^{2^K} \frac{1}{2^K} = 1$, the unbiased Horvitz–Thompson estimator of ATOT is provided in the following definition.

Definition 4.2.

$$\widehat{\delta^{**}}_{HT,tot} = \widehat{\delta^{**}}_{HT,dir} + \widehat{\delta^{**}}_{HT,ind}. \tag{4.13}$$

Theorem 4.4. *Under the no-interaction between the indirect effects assumption,*

$$\mathbf{E}(\widehat{\delta^{**}}_{HT,tot}) = \delta_{tot}. \tag{4.14}$$

$$\begin{aligned}\mathbf{Var}(\widehat{\delta^{**}}_{HT,tot}) = &\sum_{e,e'} C_e C_{e'} \mathbf{Cov}(\bar{Y}^{obs}_{HT}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}^{obs}_{HT}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\
&+ \sum_{e,e'} C_e C_{e'} \mathbf{Cov}(\bar{Y}^{obs}_{HT}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}^{obs}_{HT}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\
&- 2 \sum_{e,e', W_i=1, W_i=0} C_e C_{e'} \mathbf{Cov}(\bar{Y}^{obs}_{HT}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}^{obs}_{HT}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\
+ &\sum_{e,e', W_\ell = W_{\ell'} = 1} \sum_{\ell, \ell' = 1}^{K} C_{\ell e} C_{\ell e'} \mathbf{Cov}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 1, \mathbf{W_{e',K-1}})) \\
+ &\sum_{e,e', W_\ell = W_{\ell'} = 0} \sum_{\ell, \ell' = 1}^{K} C_{\ell e} C_{\ell e'} \mathbf{Cov}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e',K-1}})) \\
- 2 &\sum_{e,e', W_\ell = 1, W_{\ell'} = 0} \sum_{\ell, \ell' = 1}^{K} C_{\ell e} C_{\ell e'} \mathbf{Cov}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e',K-1}})) \\
+ 2 &I(C\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), C'\bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{N_K})) \sum_{e,e'} CC' \mathbf{Cov}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{N_K})) \\
-2 &I(-C\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), C'\bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{N_K})) \sum_{e,e'} CC' \mathbf{Cov}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{N_K})).
\end{aligned} \tag{4.15}$$

Proofs of the theorems are given in Appendix C. Variance and covariance components proofs are provided in Appendix A.

4.3 Variance Estimators

In this section, we follow Aronow and Samii (2013, 2017) and Lohr (2019) and provide variance estimates of the presented estimators in the previous section.

Under the no-interaction between the indirect effects assumption, the estimated conservative variance of the provided estimators in Section 4.2 is as follows.

$$\begin{aligned}\widehat{\mathbf{Var}}(\widehat{\delta^{**}}_{HT,dir}) = &\sum_{e,e',W_i=1} C_e C_{e'} \widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(1,\mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}),\bar{Y}^{obs}_{HT}(1,\mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}}))\\ &+\sum_{e,e',W_i=0} C_e C_{e'} \widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(0,\mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}),\bar{Y}^{obs}_{HT}(0,\mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}}))\\ &-2\sum_{e,e',W_i=1,W_i=0} C_e C_{e'} \widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(1,\mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}),\bar{Y}^{obs}_{HT}(0,\mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})).\end{aligned}\tag{4.16}$$

$$\begin{aligned}\widehat{\mathbf{Var}}(\widehat{\delta^{**}}_{HT,\ell}) = &\sum_{e,e',W_\ell=1} C_{\ell e} C_{\ell e'} \widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(W_i,W_\ell=1,\mathbf{W_{e,K-1}}),\bar{Y}^{obs}_{HT}(W_i,W_\ell=1,\mathbf{W_{e',K-1}}))\\ &+\sum_{e,e',W_\ell=0} C_{\ell e} C_{\ell e'} \widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(W_i,W_\ell=0,\mathbf{W_{e,K-1}}),\bar{Y}^{obs}_{HT}(W_i,W_\ell=0,\mathbf{W_{e',K-1}}))\\ &-2\sum_{e,e',W_\ell=1,W_\ell=0} C_{\ell e} C_{\ell e'} \widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(W_i,W_\ell=1,\mathbf{W_{e,K-1}}),\bar{Y}^{obs}_{HT}(W_i,W_\ell=0,\mathbf{W_{e',K-1}})).\end{aligned}\tag{4.17}$$

$$\begin{aligned}\widehat{\mathbf{Var}}(\widehat{\delta^{**}}_{HT,ind}) = &\sum_{e,e',W_\ell=W_{\ell'}=1}\sum_{\ell,\ell'=1}^{K} C_{\ell e} C_{\ell e'} \widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(W_i,W_\ell=1,\mathbf{W_{e,K-1}}),\bar{Y}^{obs}_{HT}(W_i,W_{\ell'}=1,\mathbf{W_{e',K-1}}))\\ &+\sum_{e,e',W_\ell=W_{\ell'}=0}\sum_{\ell,\ell'=1}^{K} C_{\ell e} C_{\ell e'} \widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(W_i,W_\ell=0,\mathbf{W_{e,K-1}}),\bar{Y}^{obs}_{HT}(W_i,W_{\ell'}=0,\mathbf{W_{e',K-1}}))\\ &-2\sum_{e,e',W_\ell=1,W_{\ell'}=0}\sum_{\ell,\ell'=1}^{K} C_{\ell e} C_{\ell e'} \widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(W_i,W_\ell=1,\mathbf{W_{e,K-1}}),\bar{Y}^{obs}_{HT}(W_i,W_{\ell'}=0,\mathbf{W_{e',K-1}})).\end{aligned}\tag{4.18}$$

$$
\begin{aligned}
\widehat{\mathbf{Var}}(\widehat{\delta^{**}}_{HT,tot}) = & \sum_{e,e'} C_e C_{e'} \widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}^{obs}_{HT}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\
& + \sum_{e,e'} C_e C_{e'} \widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}^{obs}_{HT}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\
& - 2 \sum_{e,e',W_i=1,W_i=0} C_e C_{e'} \widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}^{obs}_{HT}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\
& + \sum_{e,e',W_\ell=W_{\ell'}=1} \sum_{\ell,\ell'=1}^{K} C_{\ell e} C_{\ell e'} \widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W}_{\mathbf{e},\mathbf{K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 1, \mathbf{W}_{\mathbf{e'},\mathbf{K-1}})) \\
& + \sum_{e,e',W_\ell=W_{\ell'}=0} \sum_{\ell,\ell'=1}^{K} C_{\ell e} C_{\ell e'} \widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W}_{\mathbf{e},\mathbf{K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W}_{\mathbf{e'},\mathbf{K-1}})) \\
& - 2 \sum_{e,e',W_\ell=1,W_{\ell'}=0} \sum_{\ell,\ell'=1}^{K} C_{\ell e} C_{\ell e'} \widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W}_{\mathbf{e},\mathbf{K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W}_{\mathbf{e'},\mathbf{K-1}})) \\
& + 2I(C\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), C'\bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{N_K})) \sum_{e,e'} CC' \widehat{\mathbf{Cov}}_B(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{N_K})) \\
& -2I(-C\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), C'\bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{N_K})) \sum_{e,e'} CC' \widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{N_K})).
\end{aligned}
\tag{4.19}
$$

where $\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W}_{\mathbf{e},\mathbf{K-1}})) = \widehat{\mathbf{Var}}_{HT}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W}_{\mathbf{e},\mathbf{K-1}})) + \widehat{A}^*(W_i, W_\ell = 1, \mathbf{W}_{\mathbf{e},\mathbf{K-1}})$.

By linearity of expectation, Propositions A.1 and A.1 and Equation A.15 in Appendix A, the covariance of the estimators provided here is conservative. The conservative estimator of $\mathbf{Var}(\bar{Y}^{obs}_{HT}(W_i, W_{\mathcal{N}_{ik}}))$ and proofs are provided in Appendix A.

4.4 Simulation

Simulated experiments assess the performance of the proposed estimators in the previous sections. We explore three sets of three scenarios of the indirect effects. In these scenarios, the indirect effects represent the degree of interference and vary from no interference to weak interference to moderate interference. The direct effect takes also three scenarios varying δ_i values in (0,1,4) for each of the three interference scenario.

We generate responses under the following KNNIM model:

$$Y_i = X_1 + X_2 + X_3 + \delta_1 W_1 + \delta_2 W_2 + \delta_3 W_3 + \delta_4 W_i \tag{4.20}$$

where the covariates $X_j \sim N(0,1)$, $j = 1,2,3$. We consider the case where we only have $K = 3$ nearest neighbors. Half of the N units are assigned to treatment, completely at random.

The nine interference models are shown in the second column of Table 4.1 where $\delta_1, \delta_2, \delta_3$ represent the first, second, and third nearest neighbor's indirect effect such that $|\delta_1| \geq |\delta_2 \geq |\delta_3|$ and δ_i represents the unit's direct effect.

In each model, we compare total, direct, indirect and ℓ_{th} nearest neighbors estimators performance under the no-interaction between indirect effects assumption to estimators presented in the previous chapter under Assumptions 3.1 and 3.2. We simulate 100 experiments with sample size N = 256. For each realization, the marginal and joint probabilities are computed as in Equations 3.23 and 3.25. The empirical expected value of the estimates (Emp.Estimates), empirical variance (Emp.Var), empirical standard deviation (Emp.S.D.), the estimated variance (Var Estimate), and standard errors estimates of the estimators (SE) are computed. The results are shown in Table 4.2 for Model 1 to Table 4.10 for Model 9 and in Figures 4.1 to 4.9.

4.5 Discussion

For all scenarios, the variance estimates corroborate the theoretical results that the estimates are unbiased and that the variance estimates are conservative.

Smaller values of empirical variance for all estimators in all scenarios indicate estimation precision improvement under the no interaction between indirect effects assumption.

The standard errors for the direct, 1^{st}, 2^{nd} and 3^{rd} estimators are smaller than those under

Table 4.1: *Interference Models*

Models	$(\delta_{1^{st}},\ \delta_{2^{nd}},\ \delta_{3^{rd}},\ \delta_i)$
Model 1	(0,0,0,0)
Model 2	(0,0,0,1)
Model 3	(0,0,0,4)
Model 4	(2,1,0.5,0)
Model 5	(2,1,0.5,1)
Model 6	(2,1,0.5,4)
Model 7	(3,2,1,0)
Model 8	(3,2,1,1)
Model 9	(3,2,1,4)

Table 4.2: *Estimates Under Completely Randomized Design Model 1*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate.	SE
$\widehat{\delta}_{HT,tot}$	0	0.2481	1.0170	1.0085	1.1028	1.0172
$\widehat{\delta^*}_{HT,tot}$	0	0.2481	1.0170	1.0085	1.1028	1.0172
$\widehat{\delta^{**}}_{HT,tot}$	0	0.1906	0.6703	0.8187	4.2099	2.0516
$\widehat{\delta}_{HT,dir}$	0	0.0275	0.5907	0.7685	0.6068	0.7550
$\widehat{\delta^*}_{HT,dir}$	0	0.0959	0.2634	0.5132	0.4764	0.6819
$\widehat{\delta^{**}}_{HT,dir}$	0	0.0327	0.0437	0.2091	0.3233	0.5685
$\widehat{\delta}_{HT,ind}$	0	0.2205	0.6247	0.7904	0.9036	0.9270
$\widehat{\delta^*}_{HT,ind}$	0	0.1521	0.4363	0.6605	0.6329	0.7864
$\widehat{\delta^{**}}_{HT,ind}$	0	0.1578	0.3973	0.6303	2.3052	1.5180
$\widehat{\delta}_{HT,1^{st}}$	0	0.1189	0.5122	0.7157	0.6711	0.7902
$\widehat{\delta^*}_{HT,1^{st}}$	0	0.0097	0.2501	0.5001	0.3912	0.6179
$\widehat{\delta^{**}}_{HT,1^{st}}$	0	0.0507	0.0709	0.2664	0.2443	0.4939
$\widehat{\delta}_{HT,2^{nd}}$	0	0.1360	0.4295	0.6554	0.6155	0.7732
$\widehat{\delta^*}_{HT,2^{nd}}$	0	0.1386	0.2507	0.5007	0.4546	0.6706
$\widehat{\delta^{**}}_{HT,2^{nd}}$	0	0.0647	0.0916	0.3027	0.2573	0.5069
$\widehat{\delta}_{HT,3^{rd}}$	0	-0.0344	0.4234	0.6507	0.6388	0.7922
$\widehat{\delta^*}_{HT,3^{rd}}$	0	0.0036	0.2224	0.4716	0.4431	0.6601
$\widehat{\delta^{**}}_{HT,3^{rd}}$	0	0.0423	0.0807	0.2841	0.2667	0.5161

Table 4.3: *Estimates Under Completely Randomized Design Model 2*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate.	SE
$\widehat{\delta}_{HT,tot}$	1	1.2657	1.1412	1.0683	1.2883	1.0981
$\widehat{\delta^*}_{HT,tot}$	1	1.2657	1.1412	1.0683	1.2883	1.0981
$\widehat{\delta^{**}}_{HT,tot}$	1	1.1929	0.6979	0.8354	4.9994	2.2355
$\widehat{\delta}_{HT,dir}$	1	1.0452	0.7125	0.8441	0.7581	0.8407
$\widehat{\delta^*}_{HT,dir}$	1	1.1023	0.2984	0.5463	0.6013	0.7661
$\widehat{\delta^{**}}_{HT,dir}$	1	1.0328	0.0437	0.2090	0.3691	0.6074
$\widehat{\delta}_{HT,ind}$	0	0.2205	0.6247	0.7904	0.9036	0.9270
$\widehat{\delta^*}_{HT,ind}$	0	0.1634	0.4967	0.7048	0.7133	0.8347
$\widehat{\delta^{**}}_{HT,ind}$	0	0.1601	0.4256	0.6524	2.7357	1.6536
$\widehat{\delta}_{HT,1^{st}}$	0	0.1189	0.5122	0.7157	0.6711	0.7902
$\widehat{\delta^*}_{HT,1^{st}}$	0	0.0253	0.3057	0.5529	0.4637	0.6734
$\widehat{\delta^{**}}_{HT,1^{st}}$	0	0.0476	0.0801	0.2831	0.2931	0.5411
$\widehat{\delta}_{HT,2^{nd}}$	0	0.1360	0.4295	0.6554	0.6155	0.7732
$\widehat{\delta^*}_{HT,2^{nd}}$	0	0.1155	0.3016	0.5492	0.5476	0.7356
$\widehat{\delta^{**}}_{HT,2^{nd}}$	0	0.0616	0.1000	0.3162	0.3061	0.5529
$\widehat{\delta}_{HT,3^{rd}}$	0	-0.0344	0.4234	0.6507	0.6388	0.7922
$\widehat{\delta^*}_{HT,3^{rd}}$	0	0.0224	0.2859	0.5347	0.5322	0.7227
$\widehat{\delta^{**}}_{HT,3^{rd}}$	0	0.0508	0.0813	0.2852	0.3126	0.5587

Table 4.4: *Estimates Under Completely Randomized Design Model 3*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate.	SE
$\widehat{\delta}_{HT,tot}$	4	4.3186	2.6425	1.6255	3.4606	1.8082
$\widehat{\delta^*}_{HT,tot}$	4	4.3186	2.6425	1.6255	3.4606	1.8082
$\widehat{\delta^{**}}_{HT,tot}$	4	4.1998	1.0353	1.0175	14.4232	3.7966
$\widehat{\delta}_{HT,dir}$	4	4.0981	2.2066	1.4854	2.5702	1.5534
$\widehat{\delta^*}_{HT,dir}$	4	4.1213	0.8541	0.9241	2.1387	1.4484
$\widehat{\delta^{**}}_{HT,dir}$	4	4.0329	0.0436	0.2088	0.9059	0.9516
$\widehat{\delta}_{HT,ind}$	0	0.2205	0.6247	0.7904	0.9036	0.9270
$\widehat{\delta^*}_{HT,ind}$	0	0.1973	1.0702	1.0345	1.7123	1.2972
$\widehat{\delta^{**}}_{HT,ind}$	0	0.1668	0.7619	0.8729	7.9620	2.8208
$\widehat{\delta}_{HT,1^{st}}$	0	0.1189	0.5122	0.7157	0.6711	0.7902
$\widehat{\delta^*}_{HT,1^{st}}$	0	0.0722	0.8950	0.9460	1.4281	1.1882
$\widehat{\delta^{**}}_{HT,1^{st}}$	0	0.0384	0.2018	0.4492	0.8726	0.9336
$\widehat{\delta}_{HT,2^{nd}}$	0	0.1360	0.4295	0.6554	0.6155	0.7732
$\widehat{\delta^*}_{HT,2^{nd}}$	0	0.0462	0.9249	0.9617	1.7116	1.3008
$\widehat{\delta^{**}}_{HT,2^{nd}}$	0	0.0522	0.1967	0.4435	0.8914	0.9435
$\widehat{\delta}_{HT,3^{rd}}$	0	-0.0344	0.4234	0.6507	0.6388	0.7922
$\widehat{\delta^*}_{HT,3^{rd}}$	0	0.0788	0.9261	0.9623	1.6405	1.2662
$\widehat{\delta^{**}}_{HT,3^{rd}}$	0	0.0762	0.1582	0.3978	0.8899	0.9428

Table 4.5: *Estimates Under Completely Randomized Design Model 4*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate.	SE
$\widehat{\delta}_{HT,tot}$	3.5	3.8098	2.2747	1.5082	2.9302	1.6621
$\widehat{\delta^*}_{HT,tot}$	3.5	3.8098	2.2747	1.5082	2.9302	1.6621
$\widehat{\delta^{**}}_{HT,tot}$	3.5	3.6956	0.9502	0.9747	10.3000	3.2079
$\widehat{\delta}_{HT,dir}$	0	0.1176	2.6306	1.6219	3.2627	1.7721
$\widehat{\delta^*}_{HT,dir}$	0	0.1409	0.7409	0.8607	1.5610	1.2378
$\widehat{\delta^{**}}_{HT,dir}$	0	0.02400	0.07988	0.2826	0.7601	0.8716
$\widehat{\delta}_{HT,ind}$	3.5	3.6922	1.2860	1.1340	2.3522	1.5166
$\widehat{\delta^*}_{HT,ind}$	3.5	3.6688	0.9084	0.9531	1.7241	1.2986
$\widehat{\delta^{**}}_{HT,ind}$	3.5	3.6716	0.5757	0.7588	5.6947	2.3851
$\widehat{\delta}_{HT,1^{st}}$	2	2.1052	0.8123	0.9013	1.0711	1.0142
$\widehat{\delta^*}_{HT,1^{st}}$	2	2.0292	0.3511	0.5925	0.7336	0.8494
$\widehat{\delta^{**}}_{HT,1^{st}}$	2	2.04894	0.1046	0.3235	0.6608	0.8122
$\widehat{\delta}_{HT,2^{nd}}$	1	1.2768	1.2864	1.1342	1.9647	1.3878
$\widehat{\delta^*}_{HT,2^{nd}}$	1	1.1529	0.6301	0.7938	1.4534	1.2001
$\widehat{\delta^{**}}_{HT,2^{nd}}$	1	1.0891	0.1517	0.3895	0.6253	0.7901
$\widehat{\delta}_{HT,3^{rd}}$	0.5	0.3100	1.8290	1.3524	2.7502	1.6484
$\widehat{\delta^*}_{HT,3^{rd}}$	0.5	0.4867	1.2103	1.1001	1.9331	1.3823
$\widehat{\delta^{**}}_{HT,3^{rd}}$	0.5	0.5335	0.1411	0.3756	0.6120	0.7817

Table 4.6: *Estimates Under Completely Randomized Design Model 5*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate.	SE
$\widehat{\delta}_{HT,tot}$	4.5	4.8275	3.0573	1.7485	4.0582	1.9600
$\widehat{\delta^*}_{HT,tot}$	4.5	4.8275	3.0573	1.7485	4.0582	1.9600
$\widehat{\delta^{**}}_{HT,tot}$	4.5	4.6979	1.1044	1.0509	13.2126	3.6332
$\widehat{\delta}_{HT,dir}$	1	1.1352	3.4722	1.8634	4.1986	2.0074
$\widehat{\delta^*}_{HT,dir}$	1	1.1473	0.9651	0.9824	2.0101	1.4032
$\widehat{\delta^{**}}_{HT,dir}$	1	1.0240	0.0800	0.2829	0.9155	0.9566
$\widehat{\delta}_{HT,ind}$	3.5	3.6922	1.2860	1.1340	2.3522	1.5166
$\widehat{\delta^*}_{HT,ind}$	3.5	3.6801	1.1104	1.0537	2.0145	1.4027
$\widehat{\delta^{**}}_{HT,ind}$	3.5	3.6738	0.6917	0.8317	7.1571	2.6738
$\widehat{\delta}_{HT,1^{st}}$	2	2.1052	0.8123	0.9013	1.0711	1.0142
$\widehat{\delta^*}_{HT,1^{st}}$	2	2.0448	0.4531	0.6731	0.9373	0.9596
$\widehat{\delta^{**}}_{HT,1^{st}}$	2	2.0458	0.1373	0.3705	0.8158	0.9025
$\widehat{\delta}_{HT,2^{nd}}$	1	1.2768	1.2864	1.1342	1.9647	1.3878
$\widehat{\delta^*}_{HT,2^{nd}}$	1	1.1297	0.8375	0.9151	1.8875	1.3679
$\widehat{\delta^{**}}_{HT,2^{nd}}$	1	1.0859	0.1799	0.4241	0.7917	0.8890
$\widehat{\delta}_{HT,3^{rd}}$	0.5	0.3100	1.8290	1.3524	2.7502	1.6484
$\widehat{\delta^*}_{HT,3^{rd}}$	0.5	0.5055	1.5409	1.2413	2.4631	1.5589
$\widehat{\delta^{**}}_{HT,3^{rd}}$	0.5	0.5420	0.1668	0.4085	0.7780	0.8813

Table 4.7: *Estimates Under Completely Randomized Design Model 6*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate.	SE
$\widehat{\delta}_{HT,tot}$	7.5	7.8804	6.5336	2.5560	9.0579	2.9378
$\widehat{\delta^*}_{HT,tot}$	7.5	7.8804	6.5336	2.5560	9.0579	2.9378
$\widehat{\delta^{**}}_{HT,tot}$	7.5	7.7047	1.8217	1.3497	29.0056	5.3838
$\widehat{\delta}_{HT,dir}$	4	4.1881	7.1258	2.6694	8.3648	2.8264
$\widehat{\delta^*}_{HT,dir}$	4	4.1663	2.0884	1.4451	4.5201	2.1028
$\widehat{\delta^{**}}_{HT,dir}$	4	4.0241	0.0806	0.2840	1.7810	1.3343
$\widehat{\delta}_{HT,ind}$	3.5	3.6922	1.2860	1.1340	2.3522	1.5166
$\widehat{\delta^*}_{HT,ind}$	3.5	3.7140	2.1083	1.4520	3.6433	1.8871
$\widehat{\delta^{**}}_{HT,ind}$	3.5	3.6806	1.2912	1.1363	15.4792	3.9327
$\widehat{\delta}_{HT,1^{st}}$	2	2.1052	0.8123	0.9013	1.0711	1.0142
$\widehat{\delta^*}_{HT,1^{st}}$	2	2.0916	1.1817	1.0870	2.2955	1.5040
$\widehat{\delta^{**}}_{HT,1^{st}}$	2	2.0365	0.3292	0.5738	1.7140	1.3084
$\widehat{\delta}_{HT,2^{nd}}$	1	1.2768	1.2864	1.1342	1.9647	1.3878
$\widehat{\delta^*}_{HT,2^{nd}}$	1	1.0604	1.9307	1.3895	4.0752	2.0097
$\widehat{\delta^{**}}_{HT,2^{nd}}$	1	1.0766	0.3360	0.5797	1.7297	1.3142
$\widehat{\delta}_{HT,3^{rd}}$	0.5	0.3100	1.8290	1.3524	2.7502	1.6484
$\widehat{\delta^*}_{HT,3^{rd}}$	0.5	0.5619	2.9822	1.7269	4.8939	2.1931
$\widehat{\delta^{**}}_{HT,3^{rd}}$	0.5	0.5674	0.3191	0.5649	1.7159	1.3090

Table 4.8: *Estimates Under Completely Randomized Design Model 7*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate.	SE
$\widehat{\delta}_{HT,tot}$	6	6.3539	4.5838	2.1409	6.2551	2.4383
$\widehat{\delta^*}_{HT,tot}$	6	6.3539	4.5838	2.1409	6.2551	2.4383
$\widehat{\delta^{**}}_{HT,tot}$	6	6.2060	1.4393	1.1997	20.6590	4.5433
$\widehat{\delta}_{HT,dir}$	0	0.1819	6.5108	2.5516	8.0926	2.8013
$\widehat{\delta^*}_{HT,dir}$	0	0.1731	1.6877	1.2991	3.5277	1.8626
$\widehat{\delta^{**}}_{HT,dir}$	0	0.0181	0.1299	0.3604	1.5061	1.2268
$\widehat{\delta}_{HT,ind}$	6	6.1720	2.5073	1.5834	4.9432	2.2036
$\widehat{\delta^*}_{HT,ind}$	6	6.1808	1.7198	1.3114	3.7121	1.9091
$\widehat{\delta^{**}}_{HT,ind}$	6	6.1879	0.9250	0.9618	11.4523	3.3824
$\widehat{\delta}_{HT,1^{st}}$	3	3.0984	1.1387	1.0671	1.5504	1.2264
$\widehat{\delta^*}_{HT,1^{st}}$	3	3.0389	0.4756	0.6896	1.1477	1.0637
$\widehat{\delta^{**}}_{HT,1^{st}}$	3	3.0472	0.1872	0.4327	1.3216	1.1487
$\widehat{\delta}_{HT,2^{nd}}$	2	2.3685	2.5428	1.5946	4.0576	1.9985
$\widehat{\delta^*}_{HT,2^{nd}}$	2	2.1656	1.2107	1.1003	2.9755	1.7181
$\widehat{\delta^{**}}_{HT,2^{nd}}$	2	2.1053	0.2470	0.4970	1.2755	1.1284
$\widehat{\delta}_{HT,3^{rd}}$	1	0.7050	4.3580	2.08758	6.3803	2.5144
$\widehat{\delta^*}_{HT,3^{rd}}$	1	0.9762	2.8469	1.6873	4.5391	2.1205
$\widehat{\delta^{**}}_{HT,3^{rd}}$	1	1.0353	0.2571	0.5070	1.2202	1.1038

Table 4.9: *Estimates Under Completely Randomized Design Model 8*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate.	SE
$\widehat{\delta}_{HT,tot}$	7	7.3716	5.8366	2.4159	8.0563	2.7697
$\widehat{\delta^*}_{HT,tot}$	7	7.3716	5.8366	2.4159	8.0563	2.7697
$\widehat{\delta^{**}}_{HT,tot}$	7	7.2083	1.6796	1.2960	25.0822	5.0061
$\widehat{\delta}_{HT,dir}$	1	1.1995	7.8666	2.8047	9.5890	3.0456
$\widehat{\delta^*}_{HT,dir}$	1	1.1794	2.0471	1.4307	4.2084	2.0326
$\widehat{\delta^{**}}_{HT,dir}$	1	1.0181	0.1302	0.3609	1.7396	1.3185
$\widehat{\delta}_{HT,ind}$	6	6.1720	2.5073	1.5834	4.9432	2.2036
$\widehat{\delta^*}_{HT,ind}$	6	6.1921	2.0228	1.4222	4.1524	2.0177
$\widehat{\delta^{**}}_{HT,ind}$	6	6.1902	1.1002	1.0489	13.6475	3.6924
$\widehat{\delta}_{HT,1^{st}}$	3	3.0984	1.1387	1.0671	1.5504	1.2264
$\widehat{\delta^*}_{HT,1^{st}}$	3	3.0545	0.6008	0.7751	1.4171	1.1810
$\widehat{\delta^{**}}_{HT,1^{st}}$	3	3.0441	0.2382	0.4880	1.5531	1.2453
$\widehat{\delta}_{HT,2^{nd}}$	2	2.3685	2.5428	1.5946	4.0576	1.9985
$\widehat{\delta^*}_{HT,2^{nd}}$	2	2.1425	1.5115	1.2294	3.6130	1.8933
$\widehat{\delta^{**}}_{HT,2^{nd}}$	2	2.1022	0.2885	0.5371	1.5247	1.2338
$\widehat{\delta}_{HT,3^{rd}}$	1	0.7050	4.3580	2.08758	6.3803	2.5144
$\widehat{\delta^*}_{HT,3^{rd}}$	1	0.9950	3.3635	1.8339	5.3739	2.3057
$\widehat{\delta^{**}}_{HT,3^{rd}}$	1	1.04383	0.2992	0.5470	1.4713	1.2120

Table 4.10: *Estimates Under Completely Randomized Design Model 9*

Estimator	Effects	Emp.Estimates	Emp.Var	Emp.S.D.	Var Estimate.	S.E
$\widehat{\delta}_{HT,tot}$	10	10.4245	10.7237	3.2747	15.0755	3.7942
$\widehat{\delta^*}_{HT,tot}$	10	10.4245	10.7237	3.2747	15.0755	3.7942
$\widehat{\delta^{**}}_{HT,tot}$	10	10.2152	2.6553	1.6295	45.4071	6.7362
$\widehat{\delta}_{HT,dir}$	4	4.2525	13.0627	3.6142	15.4366	3.8532
$\widehat{\delta^*}_{HT,dir}$	4	4.1985	3.5759	1.8910	7.4130	2.6941
$\widehat{\delta^{**}}_{HT,dir}$	4	4.0182	0.1313	0.3624	2.8394	1.6848
$\widehat{\delta}_{HT,ind}$	6	6.1720	2.5073	1.5834	4.9432	2.2036
$\widehat{\delta^*}_{HT,ind}$	6	6.2260	3.3239	1.8231	6.2311	2.4692
$\widehat{\delta^{**}}_{HT,ind}$	6	6.1969	1.8774	1.3701	24.1681	4.9141
$\widehat{\delta}_{HT,1^{st}}$	3	3.0984	1.1387	1.0671	1.5504	1.2264
$\widehat{\delta^*}_{HT,1^{st}}$	3	3.1014	1.3991	1.1828	2.9722	1.7108
$\widehat{\delta^{**}}_{HT,1^{st}}$	3	3.0348	0.4849	0.6963	2.6805	1.6363
$\widehat{\delta}_{HT,2^{nd}}$	2	2.3685	2.5428	1.5946	4.0576	1.9985
$\widehat{\delta^*}_{HT,2^{nd}}$	2	2.0731	2.8845	1.6984	6.4103	2.5215
$\widehat{\delta^{**}}_{HT,2^{nd}}$	2	2.0928	0.4845	0.6961	2.7112	1.6454
$\widehat{\delta}_{HT,3^{rd}}$	1	0.7050	4.3580	2.0875	6.3803	2.5144
$\widehat{\delta^*}_{HT,3^{rd}}$	1	1.0514	5.3629	2.3158	8.7193	2.9314
$\widehat{\delta^{**}}_{HT,3^{rd}}$	1	1.0692	0.5004	0.7074	2.6641	1.6311

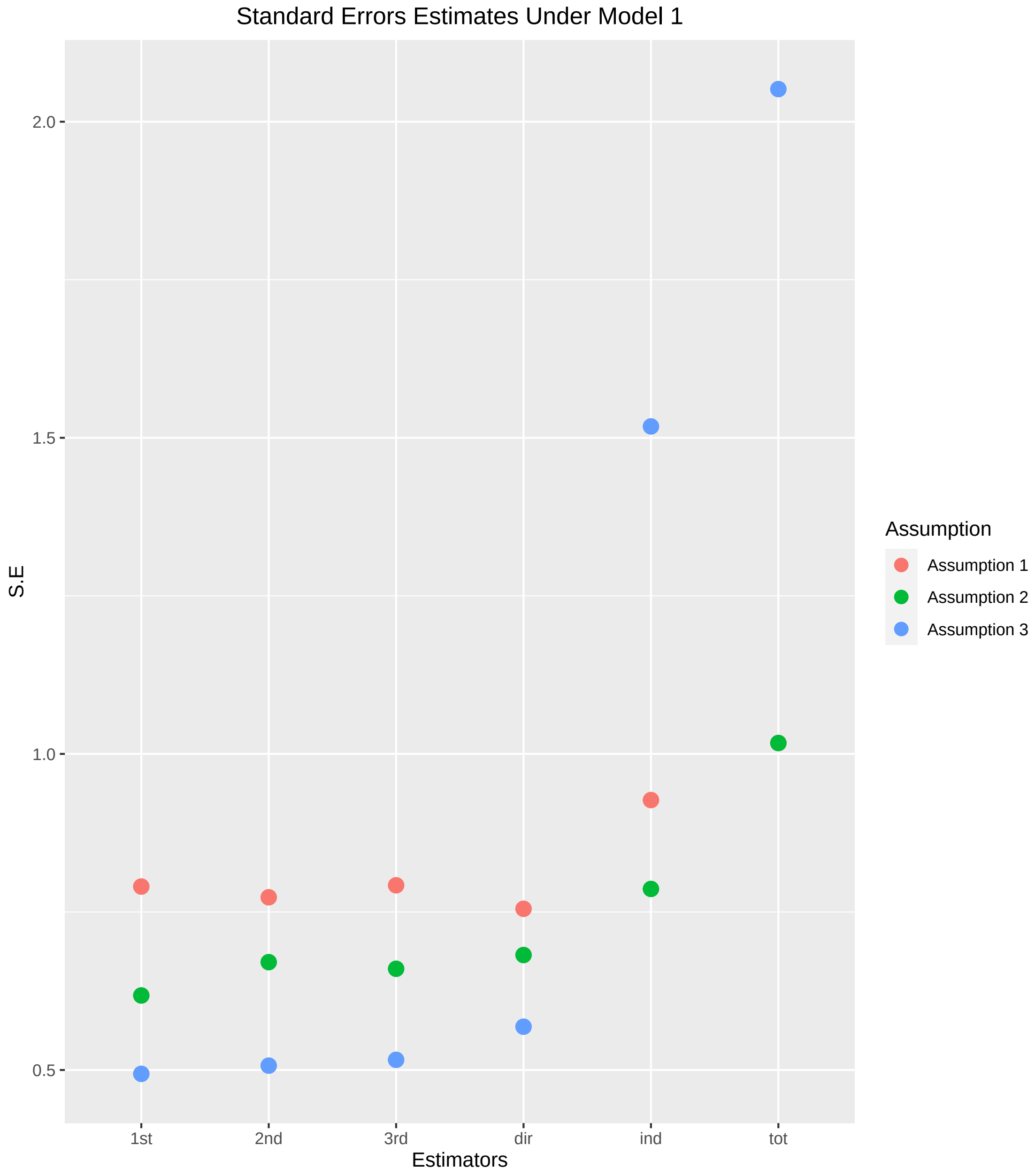


Figure 4.1: *Standard errors estimates of model 1 of all estimators under the K-nearest neighbors interference assumption, no-interaction between direct and indirect effects assumption and no-interaction between indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 100 randomizations.*

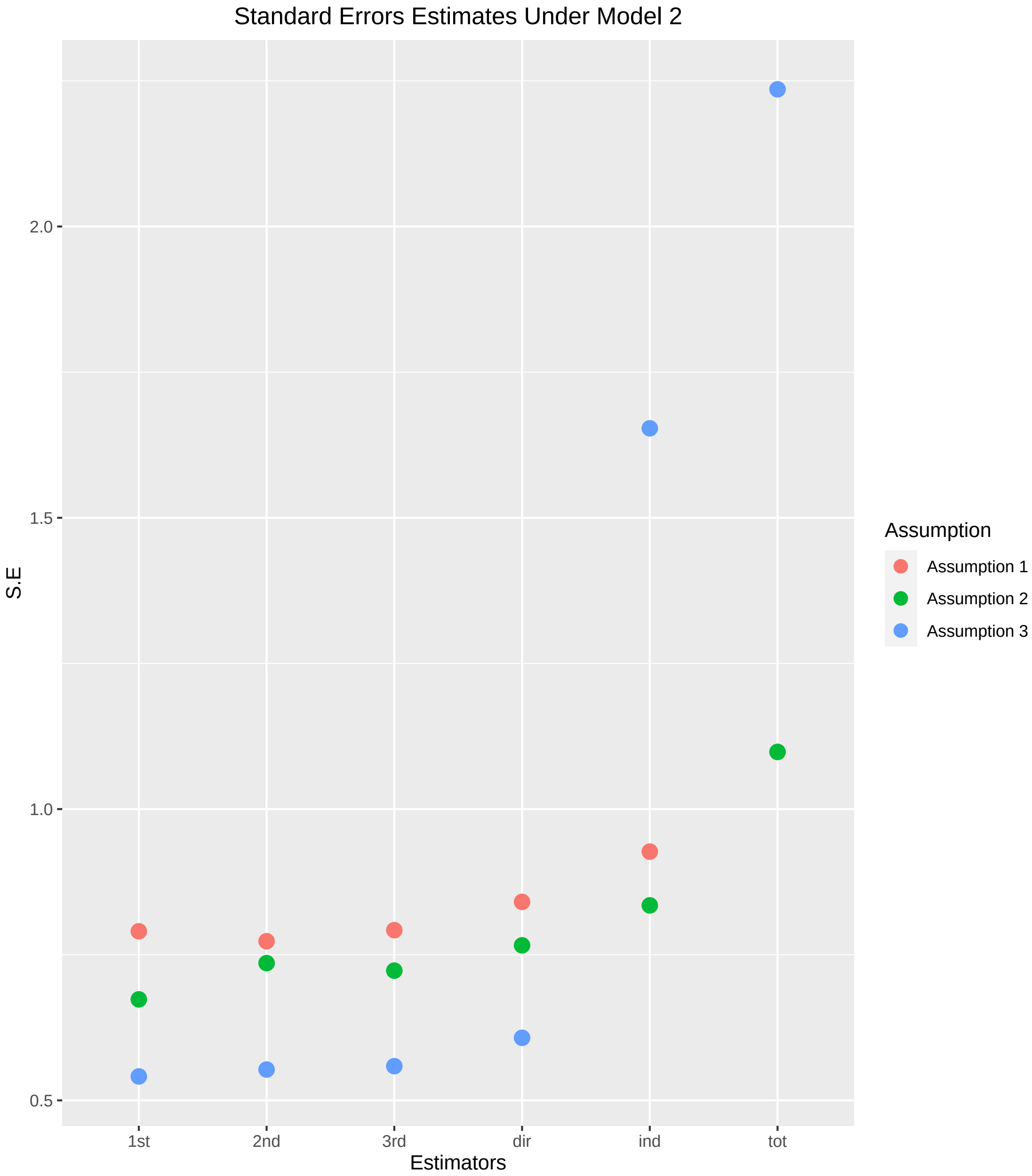


Figure 4.2: *Standard errors estimates of model 2 of all estimators under the* K*-nearest neighbors interference assumption, no-interaction between direct and indirect effects assumption and no-interaction between indirect effects assumption. We use* $N = 256$ *units and* $K = 3$ *nearest neighbors. The effects are estimated using 100 randomizations.*

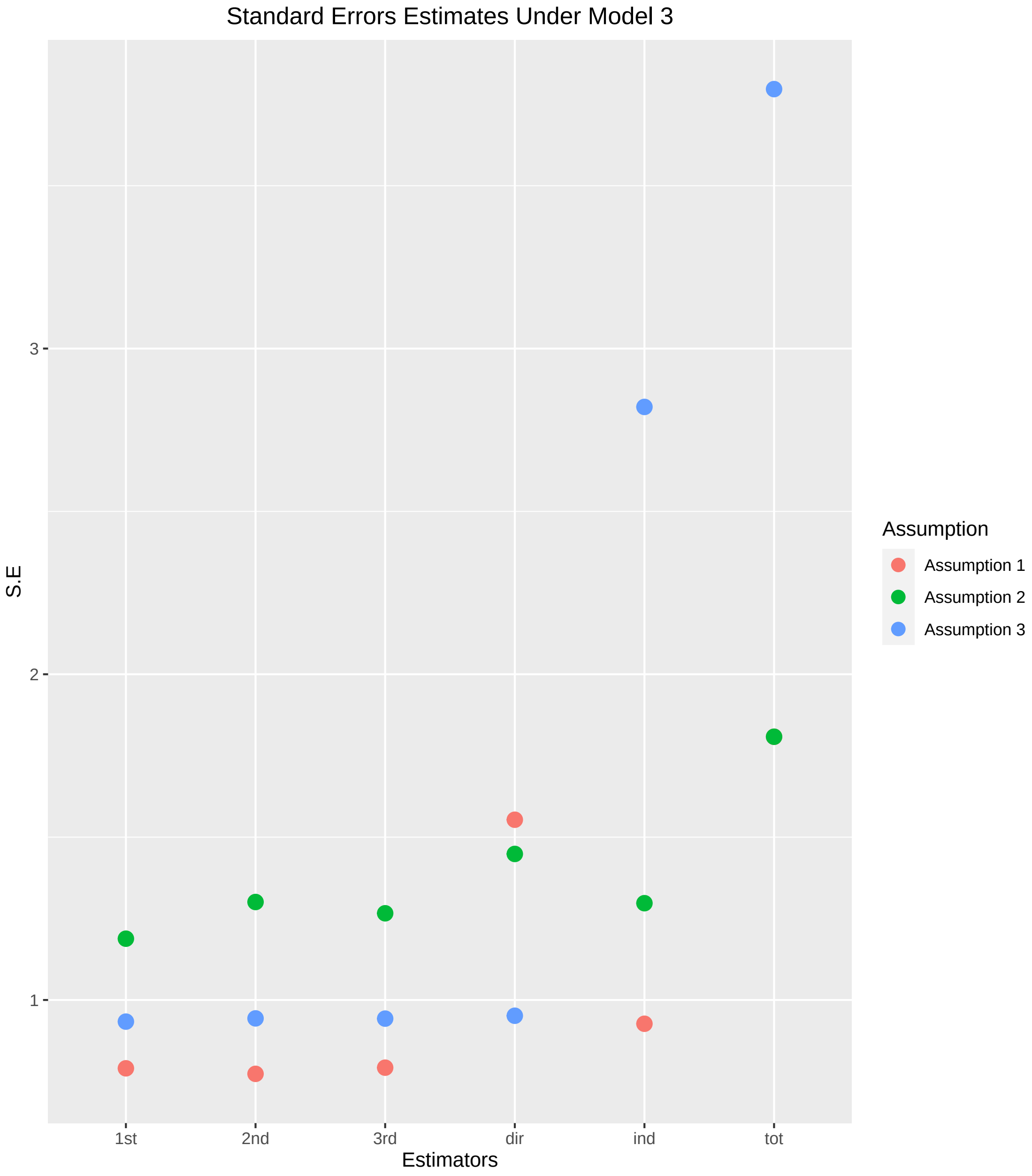


Figure 4.3: *Standard errors estimates of model 3 of all estimators under the K-nearest neighbors interference assumption, no-interaction between direct and indirect effects assumption and no-interaction between indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 100 randomizations.*

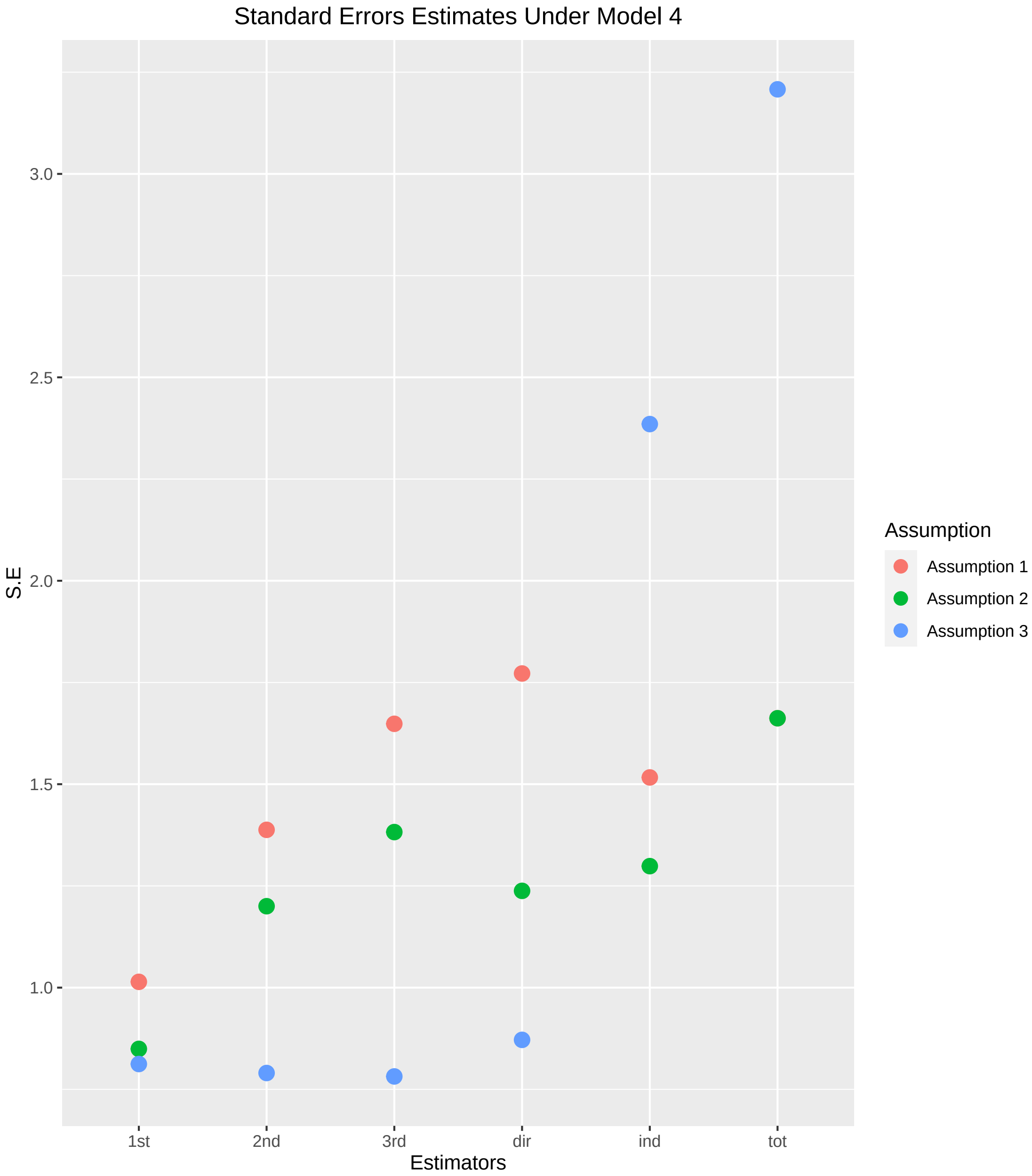


Figure 4.4: *Standard errors estimates of model 4 of all estimators under the K-nearest neighbors interference assumption, no-interaction between direct and indirect effects assumption and no-interaction between indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 100 randomizations.*

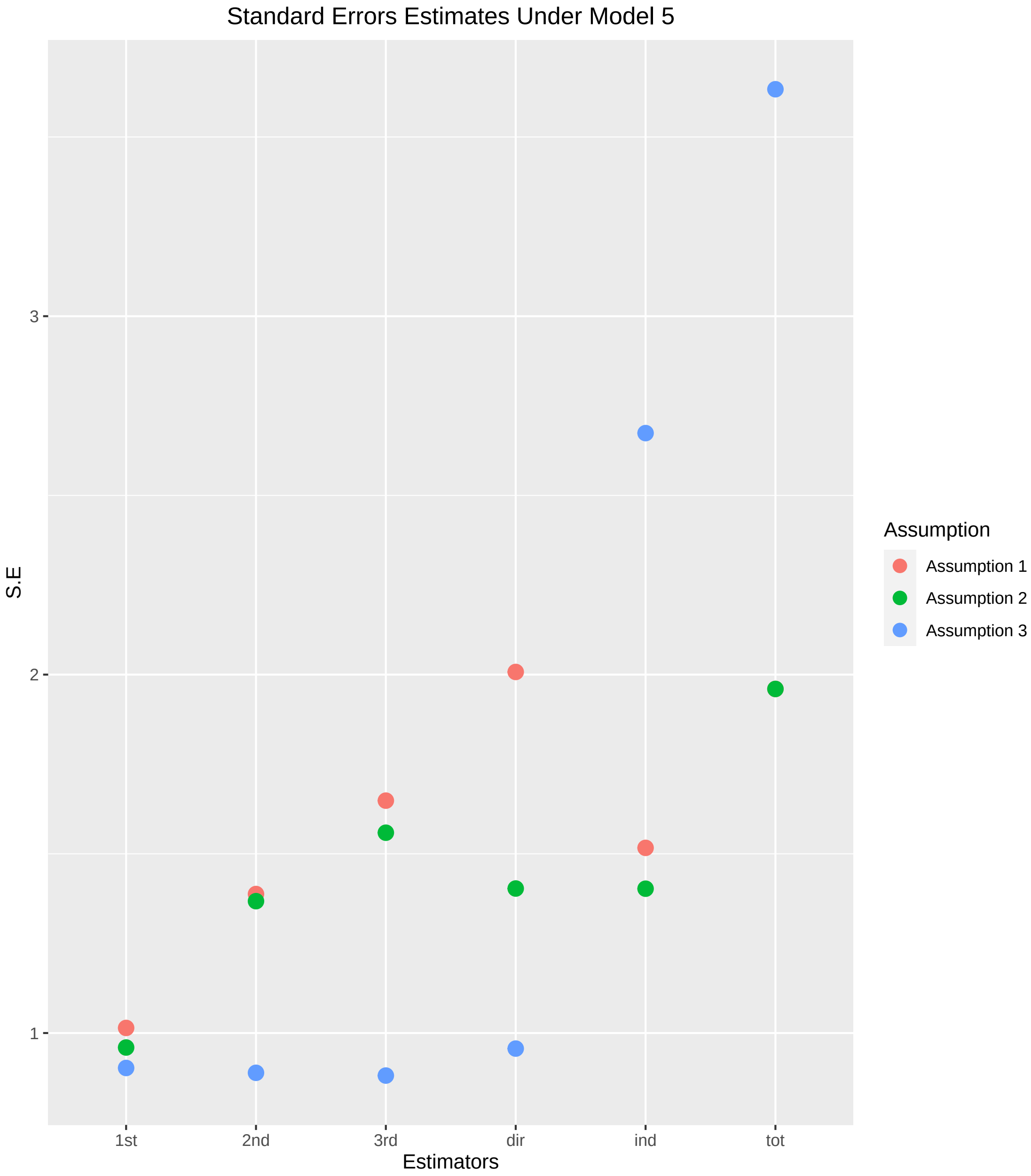


Figure 4.5: *Standard errors estimates of model 5 of all estimators under the K-nearest neighbors interference assumption, no-interaction between direct and indirect effects assumption and no-interaction between indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 100 randomizations.*

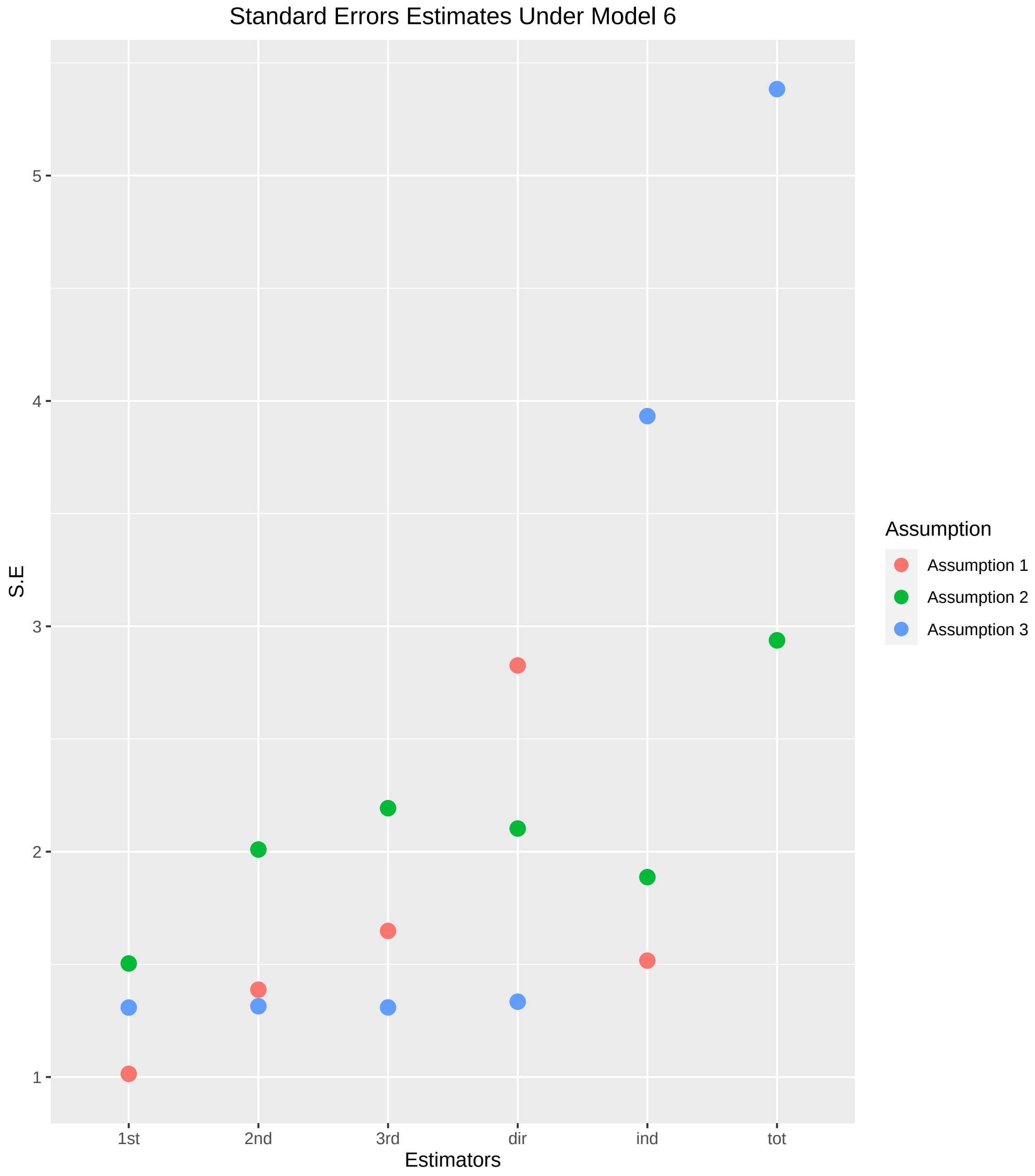


Figure 4.6: *Standard errors estimates of model 6 of all estimators under the K-nearest neighbors interference assumption, no-interaction between direct and indirect effects assumption and no-interaction between indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 100 randomizations.*

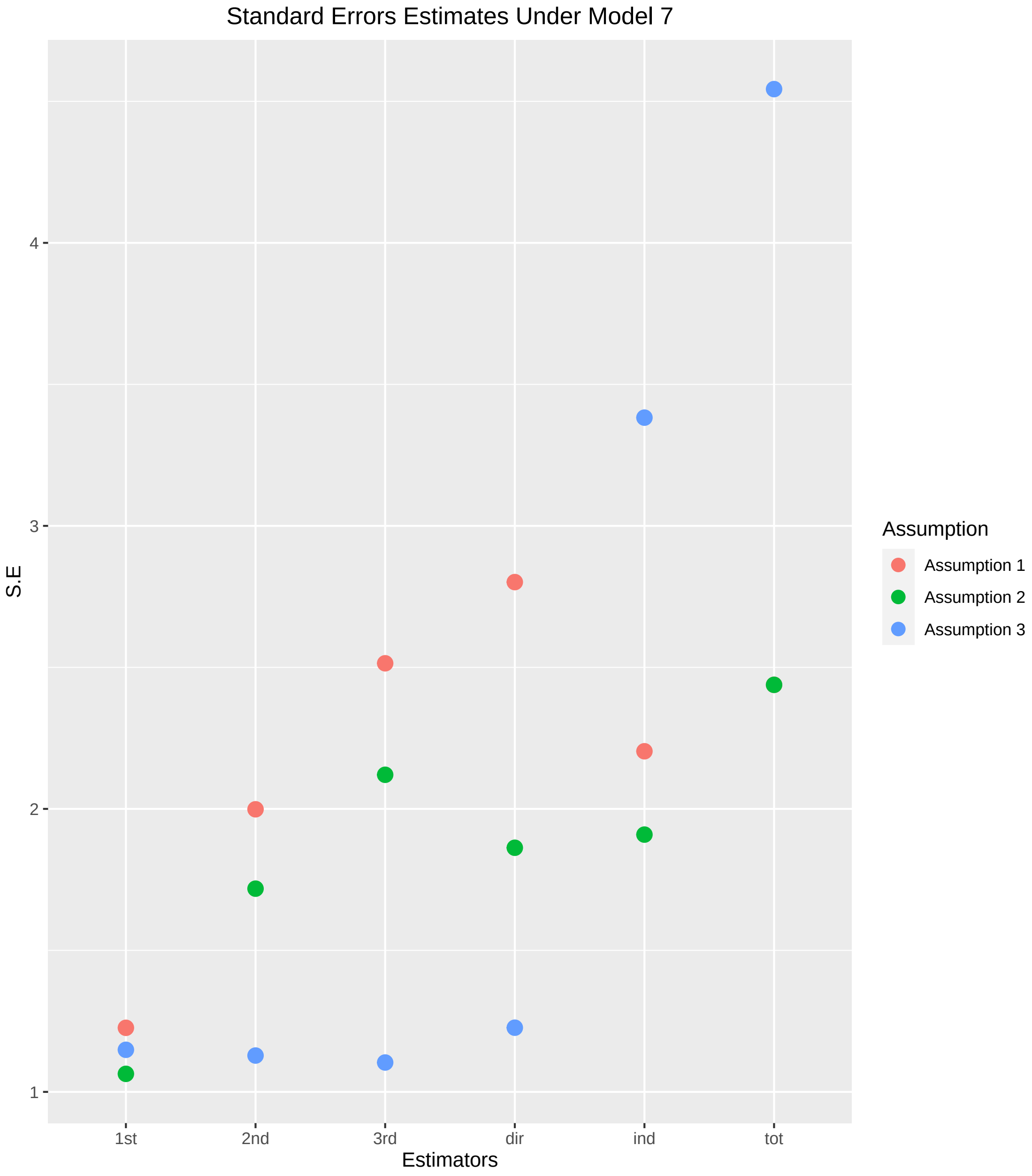


Figure 4.7: *Standard errors estimates of model 7 of all estimators under the K-nearest neighbors interference assumption, no-interaction between direct and indirect effects assumption and no-interaction between indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 100 randomizations.*

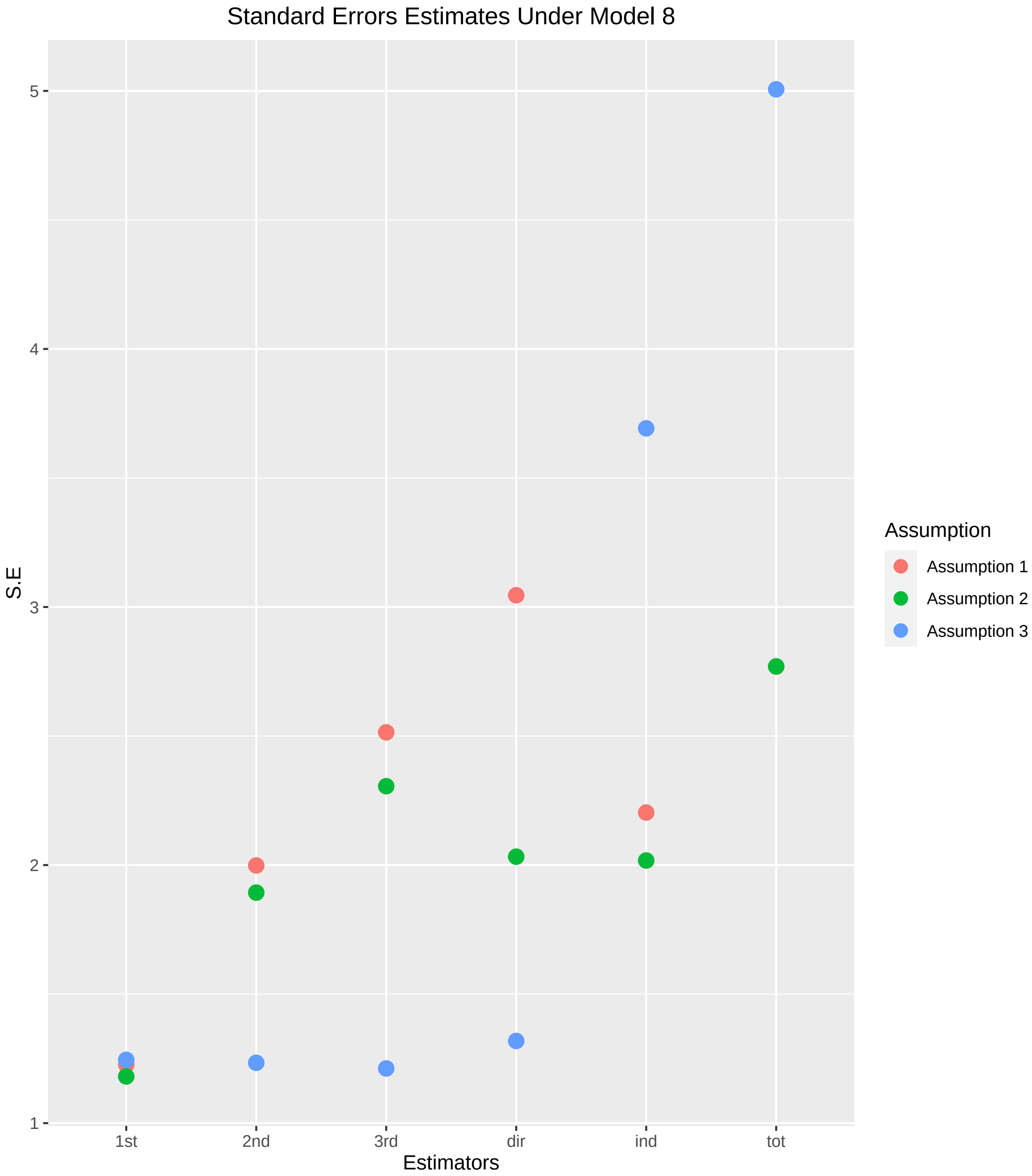


Figure 4.8: *Standard errors estimates of model 8 of all estimators under the K-nearest neighbors interference assumption, no-interaction between direct and indirect effects assumption and no-interaction between indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 100 randomizations.*

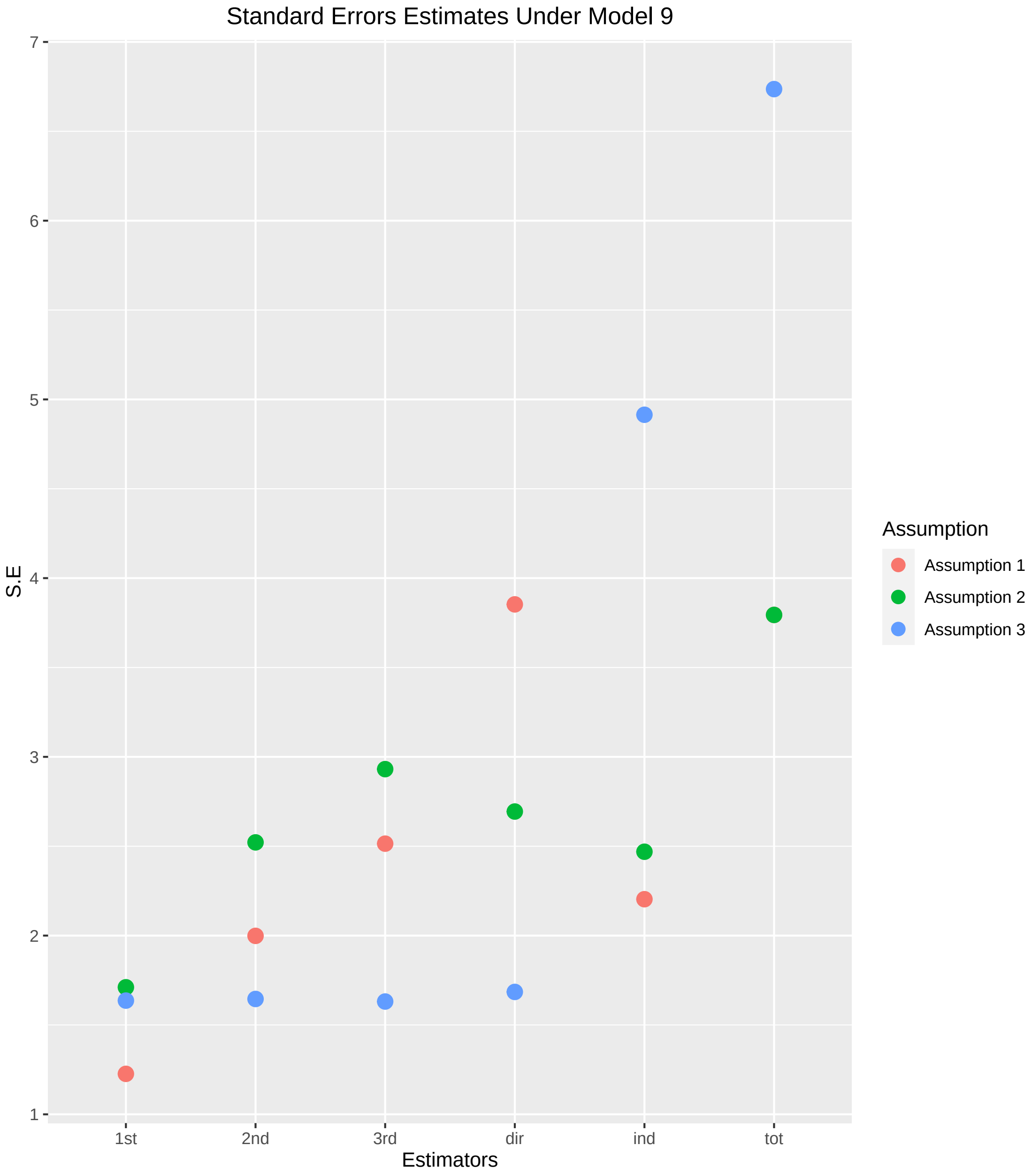


Figure 4.9: *Standard errors estimates of model 9 of all estimators under the K-nearest neighbors interference assumption, no-interaction between direct and indirect effects assumption and no-interaction between indirect effects assumption. We use $N = 256$ units and $K = 3$ nearest neighbors. The effects are estimated using 100 randomizations.*

Assumptions 3.1 and 3.2 for interference structure with smaller direct effects (see Models 1, 2, 4, 5, 7, and 8). However, for interference with slightly large direct effect (see Models 3, 6 and 9), those estimators increase estimation precision only when compared to Assumption 3.2. On the other hand, the total and indirect effect estimators under Assumption 4.1 have larger standard errors for all scenarios than those under Assumptions 3.1 and 3.2.

4.6 Conclusion

Traditional causal inference methodologies are inappropriate in the presence of interference. The K-nearest neighbors interference framework can answer interesting questions in different applications. The no-interaction between indirect effects assumption has improved estimation precision, in particular, when the direct effect is relatively small. More work could be done in order to estimate total and indirect effects when the direct effect is large.

Chapter 5

Conclusion

5.1 Summary

Causal inference under settings with interference has gained the attention of researchers in the past decade, and it remains an active research area. Many experimental studies seeking the causal treatment effect encounter treatment interference between units under study. In such settings, interference complicates the analysis rendering traditional causal inference methodologies inadequate.

This dissertation focuses on developing a new framework of causal inference in the presence of interference as an extension of the Neyman-Rubin causal model and allowing for interference within the K-neighborhood of units. In the K-nearest neighbors interference model, a unit assigned to treatment is allowed to interfere with another unit's outcome if it is one of the K closest units to the second one. (i.e., if the first unit belongs to the K-neighborhood of the second unit using an interaction measure). Typically, methodologies account for the number of treated neighbors regardless of which neighbors are included. Under the K-neighborhood framework, we account for the proximity of the neighbors to the unit. To the best of our knowledge, this interference structure previously has not been considered.

Detecting treatment interference is one way to address the problem. Existing testing methods of arbitrary interference vary between conditional randomization testing and an experimental design approach. We develop a randomization-based test and evaluate the performance of this test as well as existing methods under KNNIM through extensive simulation studies. Given an algorithm of choosing independent focal units, conditional randomization tests achieved better results than experimental design approach.

Estimation of indirect effects has become a primary interest in many applications recently. We define direct effects, indirect effects, total effects, and the ℓ_{th} nearest neighbor indirect effect under the K-neighborhood assumption. We uncover and examine the indirect effects of the K-nearest neighbors, which has not been studied. Using the Horvitz–Thompson estimator, we propose estimators of each of the defined effects under the K-neighborhood assumption, derive properties of the proposed estimators, and provide conservative variance estimators. To achieve better estimation precision, we propose another set of estimators with their properties under the no-interaction between direct and indirect effects. To demonstrate the proposed methods, under completely randomized and Bernoulli randomization designs, we evaluate the performance of estimators under both, the K-neighborhood as well as the no-interaction between direct and indirect effects assumptions through simulation study and a case study.

Finally, we develop new estimators of direct, indirect, total, and the ℓ_{th} nearest neighbor effect under a new assumption of no-interaction between indirect effects. This assumption allows for including more units to improve estimation precision. The proposed estimators, specifically the direct and the ℓ_{th} nearest neighbors, achieve smaller standard errors for almost all interference models.

5.2 Future Research

For detecting treatment interference—even though the newly developed test obtained better results than other methods—further theoretical research is needed to investigate the asymptotic behavior of the test. Comparing estimators under the K-neighborhood and the no-interaction between direct and indirect effects assumptions for both completely randomized and Bernoulli randomization designs suggests that completely randomized designs achieve better results. A research extension is to consider experimental designs—for example, cluster-randomized or two-stage designs—that can best estimate treatment effects under K-nearest neighbors interference.

The K-nearest neighbors interference framework accounts for the closeness between the unit and its K-nearest neighbors with respect to an interaction measure. Next step is to assume monotonic magnitude of the indirect effects where we expect closer neighbors to provide larger indirect effects on the outcome of the unit. Hence, sequential tests could test the significance of the indirect effects of the ℓ_{th} nearest neighbor, such that if the effect of the closest neighbor is insignificant, then the rest of the nearest neighbors effects may be insignificant accordingly. This assumption might be addressed in future research.

Bibliography

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

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

Peter M Aronow and Cyrus Samii. Conservative variance estimation for sampling designs with zero pairwise inclusion probabilities. *Surv. Methodol*, 39:231–241, 2013.

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

Peter M Aronow, Dean Eckles, Cyrus Samii, and Stephanie Zonszein. Spillover effects in experimental data. *Advances in Experimental Political Science*, page 289, 2021.

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.

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

Guillaume W Basse and Edoardo M Airoldi. Model-assisted design of experiments in the presence of network-correlated outcomes. *Biometrika*, 105(4):849–858, 2018.

GW Basse, A Feller, and P Toulis. Randomization tests of causal effects under interference. *Biometrika*, 106(2):487–494, 2019.

Robert M Bond, Christopher J Fariss, Jason J Jones, Adam DI Kramer, Cameron Marlow, Jaime E Settle, and James H Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298, 2012.

David Roxbee Cox. *Planning of experiments.* Wiley, 1958.

Dean Eckles, Brian Karrer, and Johan Ugander. Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, 5(1), 2016.

Ronald A Fisher. Statistical methods for research workers. oliver and boyd. *Edinburgh, Scotland*, 6, 1925.

Laura Forastiere, Edoardo M Airoldi, and Fabrizia Mealli. Identification and estimation of treatment and interference effects in observational studies on networks. *Journal of the American Statistical Association*, pages 1–18, 2020.

Huan Gui, Ya Xu, Anmol Bhasin, and Jiawei Han. Network a/b testing: From sampling to estimation. In *Proceedings of the 24th International Conference on World Wide Web*, pages 399–409, 2015.

M Elizabeth Halloran and Claudio J Struchiner. Causal inference in infectious diseases. *Epidemiology*, pages 142–151, 1995.

James J Higgins. *An introduction to modern nonparametric statistics.* Brooks/Cole Pacific Grove, CA, 2004.

Michael J Higgins, Fredrik Sävje, and Jasjeet S Sekhon. Improving massive experiments with threshold blocking. *Proceedings of the National Academy of Sciences*, 113(27):7369–7376, 2016.

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

Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260): 663–685, 1952.

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

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

Xiaoyun Ji. *GRAPH PARTITION PROBLEMS WITH MINIMUM SIZE CONSTRAINTS*. PhD thesis, Rensselaer Polytechnic Institute, 2004.

Brian Karrer, Liang Shi, Monica Bhole, Matt Goldman, Tyrone Palmer, Charlie Gelman, Mikael Konutgan, and Feng Sun. Network experimentation at scale. *arXiv preprint arXiv:2012.08591*, 2020.

James C King Jr, Jeffrey J Stoddard, Manjusha J Gaglani, Kristine A Moore, Laurence Magder, Elizabeth McClure, Judith D Rubin, Janet A Englund, and Kathleen Neuzil. E11ectiveness of school-based influenza vaccination. *New England Journal of Medicine*, 355(24):2523–2532, 2006.

Robert OO Kuehl. *Designs of experiments: statistical principles of research design and analysis*. Duxbury press, 2000.

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

Charles F Manski. Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542, 1993.

Charles F Manski. Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1):S1–S23, 2013.

Lawrence H Moulton, Katherine L O'Brien, Robert Kohberger, Ih Chang, Raymond Reid, Robert Weatherholtz, Jill G Hackell, George R Siber, and Mathuram Santosham. Design of a group-randomized streptococcus pneumoniae vaccine trial. *Controlled Clinical Trials*, 22(4):438–452, 2001.

J Neyman. On the application of probability theory to agricultural experiments: Principles (translated from polish original). *Roczniki Nauk Rolniczch*, 10(1):21–51, 1923.

Molly Offer-Westort and Drew Dimmery. Experimentation for homogenous policy change. *arXiv preprint arXiv:2101.12318*, 2021.

Elizabeth Levy Paluck and Hana Shepherd. The salience of social referents: a field experiment on collective norms and harassment behavior in a school social network. *Journal of personality and social psychology*, 103(6):899, 2012.

Elizabeth Levy Paluck, Hana Shepherd, and Peter M Aronow. Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences*, 113(3):566–571, 2016.

Jean Pouget-Abadie, Guillaume Saint-Jacques, Martin Saveski, Weitao Duan, S Ghosh, Y Xu, and Edoardo M Airoldi. Testing for arbitrary interference on experimentation platforms. *Biometrika*, 106(4):929–940, 2019.

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

Ronald Ross. An application of the theory of probabilities to the study of a priori pathometry.—part i. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 92(638):204–230, 1916.

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

Donald B Rubin. Bayesian inference for causality: The importance of randomization. In *The Proceedings of the social statistics section of the American Statistical Association*, volume 233, page 239. American Statistical Association Alexandria, VA, 1975.

Donald B Rubin. Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, pages 34–58, 1978.

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

Martin Saveski, Jean Pouget-Abadie, Guillaume Saint-Jacques, Weitao Duan, Souvik Ghosh, Ya Xu, and Edoardo M Airoldi. Detecting network effects: Randomizing over randomized experiments. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1027–1035, 2017.

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.

David R Schaefer, Steven A Haas, and Nicholas J Bishop. A dynamic model of us adolescents' smoking and friendship networks. *American journal of public health*, 102(6):e12–e18, 2012.

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.

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.

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.

Eric J Tchetgen Tchetgen and Tyler J VanderWeele. On causal inference in the presence of interference. *Statistical methods in medical research*, 21(1):55–75, 2012.

Panos Toulis and Edward Kao. Estimation of causal peer influence effects. In *International conference on machine learning*, pages 1489–1497, 2013.

Johan Ugander, Brian Karrer, Lars Backstrom, and Jon Kleinberg. Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 329–337, 2013.

Appendix A

Properties of Horvitz–Thompson Estimator of the Average Potential Outcomes under any Exposure

Here, we derive properties of the Horvitz–Thompson estimator of the average potential outcomes under any exposure $(W, \mathbf{W}_{\mathcal{N}_K})$—that is, the overall treatment allocation assigns treatment W to unit i and assigns treatment conditions $\mathbf{W}_{\mathcal{N}_K}$ to i's K-neighborhood.

First, the Horvitz–Thompson estimator of the total potential outcomes of units under any exposure $(W, \mathbf{W}_{\mathcal{N}_K})$ is

$$\widehat{y^T_{HT}}(W, \mathbf{W}_{\mathcal{N}_K}) = \sum_{i=1}^{N} I_i(W, \mathbf{W}_{\mathcal{N}_K}) \frac{Y_i^{obs}(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})}. \tag{A.1}$$

The unbiased Horvitz–Thompson estimator of the average potential outcomes of units under any exposure $(W, \mathbf{W}_{\mathcal{N}_K})$ is

$$\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K}) = \frac{1}{N} \sum_{i=1}^{N} I_i(W, \mathbf{W}_{\mathcal{N}_K}) \frac{Y_i^{obs}(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})}. \tag{A.2}$$

where $I_i(W, \mathbf{W}_{\mathcal{N}_K})$ is the treatment allocation indicator of unit i which is the only stochastic component of the expression. Hence, $I_i(W, \mathbf{W}_{\mathcal{N}_K})$ is a Bernoulli random variable with $\mathbf{E}[I_i(W, \mathbf{W}_{\mathcal{N}_K})] = \pi_i(W, \mathbf{W}_{\mathcal{N}_K})$, $\mathbf{Var}(I_i(W, \mathbf{W}_{\mathcal{N}_K})) = \mathbf{Cov}(I_i(W, \mathbf{W}_{\mathcal{N}_K}), I_i(W, \mathbf{W}_{\mathcal{N}_K})) = \pi_i(W, \mathbf{W}_{\mathcal{N}_K})(1-\pi_i(W, \mathbf{W}_{\mathcal{N}_K}))$ and $\mathbf{Cov}(I_i(W, \mathbf{W}_{\mathcal{N}_K}), I_j(W, \mathbf{W}_{\mathcal{N}_K})) = (\pi_{ij}(W, \mathbf{W}_{\mathcal{N}_K}) - \pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_j(W, \mathbf{W}_{\mathcal{N}_K}))$ where

$\mathbf{E}[I_i(W, \mathbf{W}_{\mathcal{N}_K}) I_j(W, \mathbf{W}_{\mathcal{N}_K})] = \pi_{ij}(W, \mathbf{W}_{\mathcal{N}_K})$ is the inclusion probability of units i and j.

A.1 The Expected value of Horvitz–Thompson Estimator

The expected value of $\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{\mathcal{N}_K})$ is

$$\begin{aligned}
\mathbf{E}(\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{\mathcal{N}_K})) &= \mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N} I_i(W, \mathbf{W}_{\mathcal{N}_K}) \frac{Y_i^{obs}(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})}\right] \\
&= \frac{1}{N}\sum_{i=1}^{N} \mathbf{E}\left[I_i(W, \mathbf{W}_{\mathcal{N}_K})\right] \frac{y_i(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})} \\
&= \frac{1}{N}\sum_{i=1}^{N} \pi_i(W, \mathbf{W}_{\mathcal{N}_K}) \frac{y_i(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})} \\
&= \frac{1}{N}\sum_{i=1}^{N} y_i(W, \mathbf{W}_{\mathcal{N}_K}) \\
&= \bar{y}(W, \mathbf{W}_{\mathcal{N}_K}).
\end{aligned} \tag{A.3}$$

A.2 The Variance of Horvitz–Thompson Estimator

Recall the property that $\mathbf{Var}(X) = \mathbf{E}(X^2)$ - $(\mathbf{E}(X))^2$ and $(\sum_{i=1}^{N} a_i X_i)^2 = \sum_{i=1}^{N} a_i^2 X_i^2 + \sum_{i=1}^{N}\sum_{j\neq i} a_i a_j X_i X_j$ and note that $\mathbf{E}[I_i^2(W, \mathbf{W}_{\mathcal{N}_K})] = \mathbf{E}[I_i(W, \mathbf{W}_{\mathcal{N}_K})] = \pi_i(W, \mathbf{W}_{\mathcal{N}_K})$ and $\pi_i((W, \mathbf{W}_{\mathcal{N}_K}), (W', \mathbf{W}'_{\mathcal{N}_K})) = 0$ because unit i cannot be exposed to two exposures at the same time.

$$\begin{aligned}
\left(\mathbf{E}(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{\mathcal{N}_K})\right)^2 &= \left(\mathbf{E}\left(\frac{1}{N}\sum_{i=1}^{N} I_i(W,\mathbf{W}_{\mathcal{N}_K})\frac{Y_i^{obs}(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}\right)\right)^2 \\
&= \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\left(I_i(W,\mathbf{W}_{\mathcal{N}_K})\right)\frac{y_i(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}\right)^2 \\
&= \frac{1}{N^2}\left(\sum_{i=1}^{N} y_i(W,\mathbf{W}_{\mathcal{N}_K})\right)^2
\end{aligned}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} y_i^2(W,\mathbf{W}_{\mathcal{N}_K}) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W,\mathbf{W}_{\mathcal{N}_K}) \tag{A.4}$$

$$
\begin{aligned}
\mathbf{E}((\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{\mathcal{N}_K}))^2) &= \mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} I_i(W,\mathbf{W}_{\mathcal{N}_K})\frac{Y_i^{obs}(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}\right)^2\right] \\
&= \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i^2(W,\mathbf{W}_{\mathcal{N}_K})\frac{Y_i^{2^{obs}}(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i^2(W,\mathbf{W}_{\mathcal{N}_K})}\right)\right. \\
&\left.+\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(W,\mathbf{W}_{\mathcal{N}_K})I_j(W,\mathbf{W}_{\mathcal{N}_K})\frac{Y_i^{obs}(W,\mathbf{W}_{\mathcal{N}_K})Y_j^{obs}(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})}\right)\right] \\
&= \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i^2(W,\mathbf{W}_{\mathcal{N}_K})\frac{Y_i^{2^{obs}}(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i^2(W,\mathbf{W}_{\mathcal{N}_K})}\right)\right] \\
&+\mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(W,\mathbf{W}_{\mathcal{N}_K})I_j(W,\mathbf{W}_{\mathcal{N}_K})\frac{Y_i^{obs}(W,\mathbf{W}_{\mathcal{N}_K})Y_j^{obs}(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})}\right)\right] \\
&= \left(\frac{1}{N^2}\sum_{i=1}^{N}\mathbf{E}[I_i^2(W,\mathbf{W}_{\mathcal{N}_K})]\frac{y_i^2(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i^2(W,\mathbf{W}_{\mathcal{N}_K})}\right) \\
&+\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\mathbf{E}[I_i(W,\mathbf{W}_{\mathcal{N}_K})I_j(W,\mathbf{W}_{\mathcal{N}_K})]\frac{y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})}\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\frac{y_i^2(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i^2(W,\mathbf{W}_{\mathcal{N}_K})} \\
&+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})\frac{y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})} \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\frac{y_i^2(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})\frac{y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})} \qquad \text{(A.5)}
\end{aligned}
$$

Hence, the variance of $\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{\mathcal{N}_K})$ is

$$\begin{aligned}
\mathbf{Var}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K})) &= \mathbf{E}((\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K}))^2) - \left(\mathbf{E}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K})\right)^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\frac{y_i^2(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})\frac{y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})} \\
&\quad - \frac{1}{N^2}\sum_{i=1}^{N}y_i^2(W,\mathbf{W}_{\mathcal{N}_K}) - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W,\mathbf{W}_{\mathcal{N}_K})
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{N^2}\sum_{i=1}^{N}\pi_i(W,\mathbf{W}_{\mathcal{N}_K})[1-\pi_i(W,\mathbf{W}_{\mathcal{N}_K})]\left[\frac{y_i(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}\right]^2 \\
&+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K}) - \pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})]\frac{y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})}. \quad \text{(A.6)}
\end{aligned}$$

A.3 The Covariance between two Horvitz–Thompson Estimators of the Averages

The covariance between the averages of potential outcomes under any two exposures to treatments $(W, \mathbf{W}_{\mathcal{N}_K})$ and $(W', \mathbf{W}'_{\mathcal{N}_K}))$ using the property that $\mathbf{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$ is as follows:

Note the following expectations:

$$
\begin{aligned}
&\mathbf{E}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K})\bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{\mathcal{N}_K}))) = \\
&\quad \mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} I_i(W, \mathbf{W}_{\mathcal{N}_K})\frac{Y_i^{obs}(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})}\right)\left(\frac{1}{N}\sum_{i=1}^{N} I_i(W', \mathbf{W}'_{\mathcal{N}_K}))\frac{Y_i^{obs}(W', \mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W', \mathbf{W}'_{\mathcal{N}_K}))}\right)\right] \\
&\qquad = \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i(W, \mathbf{W}_{\mathcal{N}_K})I_i(W', \mathbf{W}'_{\mathcal{N}_K}))\frac{Y_i^{obs}(W, \mathbf{W}_{\mathcal{N}_K})Y_i^{obs}(W', \mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_i(W', \mathbf{W}'_{\mathcal{N}_K}))}\right)\right. \\
&\qquad \left. + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(W, \mathbf{W}_{\mathcal{N}_K})I_j(W', \mathbf{W}'_{\mathcal{N}_K}))\frac{Y_i^{obs}(W, \mathbf{W}_{\mathcal{N}_K})Y_j^{obs}(W', \mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_j(W', \mathbf{W}'_{\mathcal{N}_K}))}\right)\right] \\
&\qquad = \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i(W, \mathbf{W}_{\mathcal{N}_K})I_i(W', \mathbf{W}'_{\mathcal{N}_K}))\frac{Y_i^{obs}(W, \mathbf{W}_{\mathcal{N}_K})Y_i^{obs}(W', \mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_i(W', \mathbf{W}'_{\mathcal{N}_K}))}\right)\right] \\
&\qquad + \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(W, \mathbf{W}_{\mathcal{N}_K})I_j(W', \mathbf{W}'_{\mathcal{N}_K}))\frac{Y_i^{obs}(W, \mathbf{W}_{\mathcal{N}_K})Y_j^{obs}(W', \mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_j(W', \mathbf{W}'_{\mathcal{N}_K}))}\right)\right] \\
&\qquad = \left(\frac{1}{N^2}\sum_{i=1}^{N}\mathbf{E}\left[I_i(W, \mathbf{W}_{\mathcal{N}_K})I_i(W', \mathbf{W}'_{\mathcal{N}_K}))\right]\frac{y_i(W, \mathbf{W}_{\mathcal{N}_K})y_i(W', \mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_i(W', \mathbf{W}'_{\mathcal{N}_K}))}\right) \\
&\qquad + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\mathbf{E}\left[I_i(W, \mathbf{W}_{\mathcal{N}_K})I_j(W', \mathbf{W}'_{\mathcal{N}_K}))\right]\frac{y_i(W, \mathbf{W}_{\mathcal{N}_K})y_i(W', \mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_j(W', \mathbf{W}'_{\mathcal{N}_K}))}\right) \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\pi_i((W, \mathbf{W}_{\mathcal{N}_K}), (W', \mathbf{W}'_{\mathcal{N}_K})))\frac{y_i(W, \mathbf{W}_{\mathcal{N}_K})y_i(W', \mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_i(W', \mathbf{W}'_{\mathcal{N}_K}))} \\
&\qquad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((W, \mathbf{W}_{\mathcal{N}_K}), (W', \mathbf{W}'_{\mathcal{N}_K})))\frac{y_i(W, \mathbf{W}_{\mathcal{N}_K})y_j(W', \mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_j(W', \mathbf{W}'_{\mathcal{N}_K}))} \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((W, \mathbf{W}_{\mathcal{N}_K}), (W', \mathbf{W}'_{\mathcal{N}_K})))\frac{y_i(W, \mathbf{W}_{\mathcal{N}_K})y_j(W', \mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_j(W', \mathbf{W}'_{\mathcal{N}_K}))} \qquad \text{(A.7)}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{E}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K}))\mathbf{E}(\bar{Y}^{obs}_{HT}(W',\mathbf{W}'_{\mathcal{N}_K}))) = \\
&\quad \mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N} I_i(W,\mathbf{W}_{\mathcal{N}_K})\frac{Y_i^{obs}(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}\right]\mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N} I_i(W',\mathbf{W}'_{\mathcal{N}_K}))\frac{Y_i^{obs}(W',\mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W',\mathbf{W}'_{\mathcal{N}_K}))}\right] \\
&\quad = \left[\frac{1}{N}\sum_{i=1}^{N} y_i(W,\mathbf{W}_{\mathcal{N}_K})\right]\left[\frac{1}{N}\sum_{i=1}^{N} y_i(W',\mathbf{W}'_{\mathcal{N}_K}))\right] \\
&\quad = \frac{1}{N^2}\sum_{i=1}^{N} y_i(W,\mathbf{W}_{\mathcal{N}_K})y_i(W',\mathbf{W}'_{\mathcal{N}_K})) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W',\mathbf{W}'_{\mathcal{N}_K})) \qquad \text{(A.8)}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K}),\bar{Y}^{obs}_{HT}(W',\mathbf{W}'_{\mathcal{N}_K}))) = \mathbf{E}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K})\bar{Y}^{obs}_{HT}(W',\mathbf{W}'_{\mathcal{N}_K}))) \\
&\qquad - \mathbf{E}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K}))\mathbf{E}(\bar{Y}^{obs}_{HT}(W',\mathbf{W}'_{\mathcal{N}_K}))) \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K})))\frac{y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W',\mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))} \\
&\qquad - \frac{1}{N^2}\sum_{i=1}^{N} y_i(W,\mathbf{W}_{\mathcal{N}_K})y_i(W',\mathbf{W}'_{\mathcal{N}_K})) - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W',\mathbf{W}'_{\mathcal{N}_K})) \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K}))) - \pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))\right]\frac{y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W',\mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))} \\
&\qquad - \frac{1}{N^2}\sum_{i=1}^{N} y_i(W,\mathbf{W}_{\mathcal{N}_K})y_i(W',\mathbf{W}'_{\mathcal{N}_K})). \qquad \text{(A.9)}
\end{aligned}
$$

A.4 Estimation of the Variance

There are two conditions under a measurable design: $\pi_i(W,\mathbf{W}_{\mathcal{N}_K}) > 0$ and $\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K}) > 0$. Non-measurable designs are the designs that do not meet one of these two conditions. The Horvitz–Thompson variance estimator of $\mathbf{Var}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K}))$ is provided as follows:

$$\begin{aligned}\widehat{\mathbf{Var}}_{HT}(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{\mathcal{N}_K})) &= \frac{1}{N^2}\widehat{\mathbf{Var}}_{HT}(\widehat{y_{HT}^{T}}(W,\mathbf{W}_{\mathcal{N}_K}))\\ &= \frac{1}{N^2}\sum_{i\in U} I_i(W,\mathbf{W}_{\mathcal{N}_K})[1-\pi_i(W,\mathbf{W}_{\mathcal{N}_K})]\left[\frac{Y_i^{obs}(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}\right]^2\\ + \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i} I_i(W,\mathbf{W}_{\mathcal{N}_K})I_j(W,\mathbf{W}_{\mathcal{N}_K})&\frac{[\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})-\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})]}{\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})}\\ &\times\frac{Y_i^{obs}(W,\mathbf{W}_{\mathcal{N}_K})Y_j^{obs}(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})}. \qquad (A.10)\end{aligned}$$

Under measurable designs,i.e., the joint probabilities $\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K}) > 0$ for all i and j, and this estimated variance is unbiased. However, under non measurable designs when $\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K}) = 0$ for some i and j, then this estimated variance will be biased.

Let's re-express the variance in equation A.6 as follows:

$$\begin{aligned}\mathbf{Var}(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{\mathcal{N}_K})) &= \frac{1}{N^2}\sum_{i=1}^{N}\pi_i(W,\mathbf{W}_{\mathcal{N}_K})[1-\pi_i(W,\mathbf{W}_{\mathcal{N}_K})]\left[\frac{y_i(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}\right]^2\\ +\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\in U, j\neq i:\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})>0}&[\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})-\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})]\frac{y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})}\\ &-\sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})=0} y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W,\mathbf{W}_{\mathcal{N}_K}). \qquad (A.11)\end{aligned}$$

If $\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K}) = 0$ for some i and j, then

$$\mathbf{E}(\widehat{\mathbf{Var}}_{HT}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K}))) = \mathbf{E}\left(\frac{1}{N^2}\sum_{i\in U} I_i(W,\mathbf{W}_{\mathcal{N}_K})[1-\pi_i(W,\mathbf{W}_{\mathcal{N}_K})]\left[\frac{Y^{obs}_i(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}\right]^2\right.$$
$$+\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})>0} I_i(W,\mathbf{W}_{\mathcal{N}_K})I_j(W,\mathbf{W}_{\mathcal{N}_K})\frac{[\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})-\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})]}{\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})}$$
$$\left.\times\frac{Y^{obs}_i(W,\mathbf{W}_{\mathcal{N}_K})Y^{obs}_j(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})}\right)$$
$$= \mathbf{Var}_{HT}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K})) + \sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})=0} y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W,\mathbf{W}_{\mathcal{N}_K}).$$
$$= \mathbf{Var}_{HT}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K})) + A, \quad \text{(A.12)}$$

where $A = \sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})=0} y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W,\mathbf{W}_{\mathcal{N}_K})$ and we can never observe $y_i(W,\mathbf{W}_{\mathcal{N}_K})$ and $y_j(W,\mathbf{W}_{\mathcal{N}_K})$ together because $\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K}) = 0$.

As derived in Aronow and Samii (2013, 2017), we have the following variance bias correction,

$$\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K})) = \widehat{\mathbf{Var}}_{HT}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K})) + \widehat{A}^*(W,\mathbf{W}_{\mathcal{N}_K})$$
$$= \frac{1}{N^2}\sum_{i\in U} I_i(W,\mathbf{W}_{\mathcal{N}_K})[1-\pi_i(W,\mathbf{W}_{\mathcal{N}_K})]\left[\frac{Y^{obs}_i(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}\right]^2$$
$$+\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})>0} I_i(W,\mathbf{W}_{\mathcal{N}_K})I_j(W,\mathbf{W}_{\mathcal{N}_K})\frac{[\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})-\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})]}{\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})}$$
$$\times\frac{Y^{obs}_i(W,\mathbf{W}_{\mathcal{N}_K})Y^{obs}_j(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W,\mathbf{W}_{\mathcal{N}_K})}$$
$$+\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}(W,\mathbf{W}_{\mathcal{N}_K})=0}\left[\frac{I_i(W,\mathbf{W}_{\mathcal{N}_K})Y^{2^{obs}}_i(W,\mathbf{W}_{\mathcal{N}_K})}{2\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}+\frac{I_j(W,\mathbf{W}_{\mathcal{N}_K})Y^{2^{obs}}_j(W,\mathbf{W}_{\mathcal{N}_K})}{2\pi_j(W,\mathbf{W}_{\mathcal{N}_K})}\right]$$
$$\text{(A.13)}$$

where

$$\widehat{A^*}(W, \mathbf{W}_{\mathcal{N}_K}) = \frac{1}{N^2} \sum_{i \in U} \sum_{j \in U, j \neq i : \pi_{ij}(W, \mathbf{W}_{\mathcal{N}_K}) = 0} \left[\frac{I_i(W, \mathbf{W}_{\mathcal{N}_K}) Y_i^{2^{obs}}(W, \mathbf{W}_{\mathcal{N}_K})}{2\pi_i(W, \mathbf{W}_{\mathcal{N}_K})} \right.$$
$$\left. + \frac{I_j(W, \mathbf{W}_{\mathcal{N}_K}) Y_j^{2^{obs}}(W, \mathbf{W}_{\mathcal{N}_K})}{2\pi_j(W, \mathbf{W}_{\mathcal{N}_K})} \right] \quad \text{(A.14)}$$

Then, $\widehat{\mathbf{Var}}_A(\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{\mathcal{N}_K}))$ is a conservative estimator for the variance of Horvitz–Thompson estimator of the average potential outcomes under exposure $(W, \mathbf{W}_{\mathcal{N}_K})$.

$$\mathbf{E}(\widehat{\mathbf{Var}}_A(\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{\mathcal{N}_K})) \geq \mathbf{Var}_{HT}(\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{\mathcal{N}_K})) \quad \text{(A.15)}$$

Proof. First, for simplicity, let $I_i = I_i(W, \mathbf{W}_{\mathcal{N}_K})$, $I_j = I_j(W, \mathbf{W}_{\mathcal{N}_K})$, $\pi_i = \pi_i(W, \mathbf{W}_{\mathcal{N}_K})$, $\pi_j = \pi_j(W, \mathbf{W}_{\mathcal{N}_K})$, $\pi_{ij} = \pi_{ij}(W, \mathbf{W}_{\mathcal{N}_K})$, $Y_i = Y_i^{obs}(W, \mathbf{W}_{\mathcal{N}_K})$ and $Y_j = Y_j^{obs}(W, \mathbf{W}_{\mathcal{N}_K})$ and instead of the average, consider the variance bias correction of the total as follows,

$$\widehat{\mathbf{Var}}_C(\widehat{y_{HT}^T}(W, \mathbf{W}_{\mathcal{N}_K})) = \sum_{i \in U} I_i[1 - \pi_i] \left[\frac{Y_i}{\pi_i}\right]^2 + \sum_{i \in U} \sum_{j \in U, j \neq i : \pi_{ij} > 0} I_i I_j \frac{[\pi_{ij} - \pi_i \pi_j]}{\pi_{ij}} \times \frac{Y_i Y_j}{\pi_i \pi_j}$$
$$+ \sum_{i \in U} \sum_{j \in U, j \neq i : \pi_{ij} = 0} \left[I_i \frac{|Y_i|^{a_{ij}}}{a_{ij} \pi_i} + I_j \frac{|Y_j|^{b_{ij}}}{b_{ij} \pi_j} \right]$$

where a_{ij} and b_{ij} are positive real numbers such that $\frac{1}{a_{ij}} + \frac{1}{b_{ij}} = 1$ for all pairs i and j with $\pi_{ij} = 0$.

By Young's inequality, if $\frac{1}{a_{ij}} + \frac{1}{b_{ij}} = 1$,

$$\frac{|y_i|^{a_{ij}}}{a_{ij}} + \frac{|y_j|^{b_{ij}}}{b_{ij}} \geq |y_i||y_j|.$$

Define A^* such that,

$$A^* = \sum_{i\in U} \sum_{j\in U, j\neq i:\pi_{ij}=0} \frac{|y_i|^{a_{ij}}}{a_{ij}} + \frac{|y_j|^{b_{ij}}}{b_{ij}} \geq \sum_{i\in U} \sum_{j\in U, j\neq i:\pi_{ij}=0} |y_i||y_j| \geq \sum_{i\in U} \sum_{j\in U, j\neq i:\pi_{ij}=0} y_i y_j = A$$

and

$$A^* \geq \sum_{i\in U} \sum_{j\in U, j\neq i:\pi_{ij}=0} |y_i||y_j| \geq \sum_{i\in U} \sum_{j\in U, j\neq i:\pi_{ij}=0} -y_i y_j = -A.$$

Therefore,

$$\mathbf{Var}_{HT}(\widehat{y^T_{HT}}(W, \mathbf{W}_{\mathcal{N}_K})) + A^* \geq \mathbf{Var}_{HT}(\widehat{y^T_{HT}}(W, \mathbf{W}_{\mathcal{N}_K})) - A$$

and the associated Horvitz–Thompson estimator of A^* is

$$\widehat{A^*} = \sum_{i\in U} \sum_{j\in U, j\neq i:\pi_{ij}=0} \left[I_i \frac{|Y_i|^{a_{ij}}}{a_{ij}\pi_i} + I_j \frac{|Y_j|^{b_{ij}}}{b_{ij}\pi_j} \right]$$

where unbiasedness of $\widehat{A^*}$ follows by $\mathbf{E}(I_i) = \pi_i$ and $\mathbf{E}(I_j) = \pi_j$. By Equation A.12 and by $\mathbf{E}(\widehat{A^*}) = A^*$,

$$\mathbf{E}(\widehat{\mathbf{Var}}_{HT}(\widehat{y^T_{HT}}(W, \mathbf{W}_{\mathcal{N}_K})) + \widehat{A^*}) = \mathbf{Var}_{HT}(\widehat{y^T_{HT}}(W, \mathbf{W}_{\mathcal{N}_K})) + A + A^*.$$

Hence,

$$\mathbf{E}(\widehat{\mathbf{Var}}_C(\widehat{y^T_{HT}}(W, \mathbf{W}_{\mathcal{N}_K}))) \geq \mathbf{Var}_{HT}(\widehat{y^T_{HT}}(W, \mathbf{W}_{\mathcal{N}_K})).$$

As a special case, assigning all $a_{ij} = b_{ij} = 2$ such that $\frac{1}{2} + \frac{1}{2} = 1$ for all pairs i and j with $\pi_{ij} = 0$ where

$$A^* = \sum_{i \in U} \sum_{j \in U, j \neq i : \pi_{ij} = 0} \frac{|y_i|^2}{2} + \frac{|y_j|^2}{2}$$

and

$$\widehat{A^*} = \sum_{i \in U} \sum_{j \in U, j \neq i : \pi_{ij} = 0} \left[I_i \frac{|Y_i|^2}{2\pi_i} + I_j \frac{|Y_j|^2}{2\pi_j} \right]$$

and for $\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K}) = \frac{1}{N} \widehat{y^T_{HT}}(W, \mathbf{W}_{\mathcal{N}_K})$,we have

$$\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K})) = \frac{1}{N^2} \sum_{i \in U} I_i(W, \mathbf{W}_{\mathcal{N}_K})[1 - \pi_i(W, \mathbf{W}_{\mathcal{N}_K})] \left[\frac{Y^{obs}_i(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K})} \right]^2$$
$$+ \frac{1}{N^2} \sum_{i \in U} \sum_{j \in U, j \neq i : \pi_{ij}(W, \mathbf{W}_{\mathcal{N}_K}) > 0} I_i(W, \mathbf{W}_{\mathcal{N}_K}) I_j(W, \mathbf{W}_{\mathcal{N}_K}) \frac{[\pi_{ij}(W, \mathbf{W}_{\mathcal{N}_K}) - \pi_i(W, \mathbf{W}_{\mathcal{N}_K})\pi_j(W, \mathbf{W}_{\mathcal{N}_K})]}{\pi_{ij}(W, \mathbf{W}_{\mathcal{N}_K})}$$
$$\times \frac{Y^{obs}_i(W, \mathbf{W}_{\mathcal{N}_K}) Y^{obs}_j(W, \mathbf{W}_{\mathcal{N}_K})}{\pi_i(W, \mathbf{W}_{\mathcal{N}_K}) \pi_j(W, \mathbf{W}_{\mathcal{N}_K})}$$
$$+ \frac{1}{N^2} \sum_{i \in U} \sum_{j \in U, j \neq i : \pi_{ij}(W, \mathbf{W}_{\mathcal{N}_K}) = 0} \left[\frac{I_i(W, \mathbf{W}_{\mathcal{N}_K}) Y^{2^{obs}}_i(W, \mathbf{W}_{\mathcal{N}_K})}{2\pi_i(W, \mathbf{W}_{\mathcal{N}_K})} + \frac{I_j(W, \mathbf{W}_{\mathcal{N}_K}) Y^{2^{obs}}_j(W, \mathbf{W}_{\mathcal{N}_K})}{2\pi_j(W, \mathbf{W}_{\mathcal{N}_K})} \right]$$

Therefore, we have proved that,

$$\mathbf{E}(\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K})) \geq \mathbf{Var}_{HT}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K}))$$

Hence, $\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K})$ is a conservative estimator of $\mathbf{Var}_{HT}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K}))$. ■

Moreover, last term in equation A.9 is unidentified because each unit receives only one exposure and can only be observed under this exposure. Hence, there is no unbiased estimator for the variance of the proposed estimators. However, if the joint probabilities $\pi_{ij}((W, \mathbf{W}_{\mathcal{N}_K}), (W', \mathbf{W}'_{\mathcal{N}_K}))) > 0$ for two different exposures $(W, \mathbf{W}_{\mathcal{N}_K})$ and $(W', \mathbf{W}'_{\mathcal{N}_K}))$ for all i and j, an estimator for the covariance in A.9 can be as follows:

$$\begin{aligned}
&\widehat{\mathbf{Cov}}(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{\mathcal{N}_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W}'_{\mathcal{N}_K}))) = \\
&\qquad \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U,j\neq i}\Bigg[\frac{I_i(W,\mathbf{W}_{\mathcal{N}_K})I_j(W',\mathbf{W}'_{\mathcal{N}_K}))}{\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K})))}\frac{Y_i(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}\frac{Y_j(W',\mathbf{W}'_{\mathcal{N}_K}))}{\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))} \\
&\qquad\qquad \times[\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K}))) - \pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))]\Bigg] \\
&\qquad -\frac{1}{N^2}\sum_{i\in U}\left[\frac{I_i(W,\mathbf{W}_{\mathcal{N}_K})Y_i^{2^{obs}}(W,\mathbf{W}_{\mathcal{N}_K})}{2\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}+\frac{I_i(W',\mathbf{W}'_{\mathcal{N}_K}))Y_i^{2^{obs}}(W',\mathbf{W}'_{\mathcal{N}_K}))}{2\pi_i(W',\mathbf{W}'_{\mathcal{N}_K}))}\right] \quad (A.16)
\end{aligned}$$

such that

$$\mathbf{E}(\widehat{\mathbf{Cov}}(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{\mathcal{N}_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W}'_{\mathcal{N}_K})))) \leq \mathbf{Cov}(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{\mathcal{N}_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W}'_{\mathcal{N}_K}))) \quad (A.17)$$

By Young's inequality, this can be proved by the fact that the expected value of last term in equation A.16 is less than or equal to the last term in equation A.9.

For the case where the joint probabilities $\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K}))) = 0$ for some i and j, the covariance in A.9 can be refined as follows:

$$\begin{aligned}
&\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{\mathcal{N}_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W}'_{\mathcal{N}_K}))) = \\
&\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U,j\neq i:\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K})))>0}[\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K}))) - \pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))] \\
&\qquad\qquad \times\frac{y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W',\mathbf{W}'_{\mathcal{N}_K}))}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))} \\
&\qquad -\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U:\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K})))=0} y_i(W,\mathbf{W}_{\mathcal{N}_K})y_j(W',\mathbf{W}'_{\mathcal{N}_K})). \quad (A.18)
\end{aligned}$$

Consequently, the more general covariance estimator is

$$\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{\mathcal{N}_K})) =$$

$$\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K})))>0}\left[\frac{I_i(W,\mathbf{W}_{\mathcal{N}_K})I_j(W',\mathbf{W}'_{\mathcal{N}_K}))}{\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K})))}\frac{Y^{obs}_i(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}\frac{Y^{obs}_j(W',\mathbf{W}'_{\mathcal{N}_K}))}{\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))}\right.$$

$$\left.\times[\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K}))) - \pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))]\right]$$

$$-\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U:\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K})))=0}\left[\frac{I_i(W,\mathbf{W}_{\mathcal{N}_K})Y_i^2}{2\pi_i(W,\mathbf{W}_{\mathcal{N}_K})} + \frac{I_j(W',\mathbf{W}'_{\mathcal{N}_K}))Y_j^2}{2\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))}\right] \quad (A.19)$$

Proposition A.1.

$$\mathbf{E}(\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K}), \bar{Y}^{obs}_{HT}(W',\mathbf{W}'_{\mathcal{N}_K}))) \leq \mathbf{Cov}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K}), \bar{Y}^{obs}_{HT}(W',\mathbf{W}'_{\mathcal{N}_K}))). \quad (A.20)$$

Proof. First, for simplicity, let $I_i = I_i(W, \mathbf{W}_{\mathcal{N}_K})$, $I_j = I_j(W', \mathbf{W}'_{\mathcal{N}_K}))$, $\pi_i = \pi_i(W, \mathbf{W}_{\mathcal{N}_K})$, $\pi_j = \pi_j(W', \mathbf{W}'_{\mathcal{N}_K}))$, $\pi_{ij} = \pi_{ij}((W, \mathbf{W}_{\mathcal{N}_K}), (W', \mathbf{W}'_{\mathcal{N}_K})))$, $Y_i = Y^{obs}_i(W, \mathbf{W}_{\mathcal{N}_K})$ and $Y_j = Y^{obs}_j(W', \mathbf{W}'_{\mathcal{N}_K}))$ and re-express the covariance in A.18 as follows,

$$\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W,\mathbf{W}_{\mathcal{N}_K}), \bar{Y}^{obs}_{HT}(W',\mathbf{W}'_{\mathcal{N}_K}))) = \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U,j\neq i:\pi_{ij}>0}[\pi_{ij}-\pi_i\pi_j]\frac{y_iy_j}{\pi_i\pi_j} - \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U:\pi_{ij}=0}y_iy_j$$

$$= \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U,j\neq i:\pi_{ij}>0}[\pi_{ij}-\pi_i\pi_j]\frac{y_iy_j}{\pi_i\pi_j} - \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U:\pi_{ij}=0}A'. \quad (A.21)$$

By Young's inequality, if $\frac{1}{a_{ij}} + \frac{1}{b_{ij}} = 1$,

$$\frac{|y_i|^{a_{ij}}}{a_{ij}} + \frac{|y_j|^{b_{ij}}}{b_{ij}} \geq |y_i||y_j|. \quad (A.22)$$

Then,

$$A'^{*} = \sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}=0} \frac{|y_i|^{a_{ij}}}{a_{ij}} + \frac{|y_j|^{b_{ij}}}{b_{ij}} \geq \sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}=0} |y_i||y_j| \geq \sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}=0} y_i y_j = A' \tag{A.23}$$

and

$$A'^{*} \geq \sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}=0} |y_i||y_j| \geq \sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}=0} -y_i y_j = -A'.$$

Therefore,

$$\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}>0} \left[\pi_{ij} - \pi_i\pi_j\right]\frac{y_i y_j}{\pi_i\pi_j} - \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U:\pi_{ij}=0} A'^{*} \leq$$
$$\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}>0} \left[\pi_{ij} - \pi_i\pi_j\right]\frac{y_i y_j}{\pi_i\pi_j} - \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U:\pi_{ij}=0} A'.$$

and the associated Horvitz–Thompson estimator of A'^{*} is

$$\widehat{A'^{*}} = \sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}=0} \left[I_i\frac{|Y_i|^{a_{ij}}}{a_{ij}\pi_i} + I_j\frac{|Y_j|^{b_{ij}}}{b_{ij}\pi_j}\right]$$

and consider the variance bias correction of the average as follows,

$$\begin{aligned}\widehat{\mathbf{Cov}}_A(\bar{Y}_{HT}^{obs}(W, \mathbf{W}_{\mathcal{N}_K}), \bar{Y}_{HT}^{obs}(W', \mathbf{W}'_{\mathcal{N}_K})) &= \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}>0}\left[\frac{I_i I_j}{\pi_{ij}}\frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}\times\left[\pi_{ij} - \pi_i\pi_j\right]\right] \\ &\quad - \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}=0}\left[I_i\frac{|Y_i|^{a_{ij}}}{a_{ij}\pi_i} + I_j\frac{|Y_j|^{b_{ij}}}{b_{ij}\pi_j}\right] \\ &= \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}>0}\left[\frac{I_i I_j}{\pi_{ij}}\frac{Y_i}{\pi_i}\frac{Y_j}{\pi_j}\times\left[\pi_{ij} - \pi_i\pi_j\right]\right] - \frac{1}{N^2}\sum_{i\in U}\sum_{j\in U:\pi_{ij}=0}\widehat{A'^{*}}\end{aligned}$$

where a_{ij} and b_{ij} are positive real numbers such that $\frac{1}{a_{ij}} + \frac{1}{b_{ij}} = 1$ for all pairs i and j with $\pi_{ij} = 0$.

As a special case, assigning all $a_{ij} = b_{ij} = 2$ such that $\frac{1}{2} + \frac{1}{2} = 1$ for all pairs i and j with $\pi_{ij} = 0$,where

$$A'^{*} = \sum_{i \in U} \sum_{j \in U, j \neq i : \pi_{ij} = 0} \frac{|y_i|^2}{2} + \frac{|y_j|^2}{2}$$

and

$$\widehat{A'^{*}} = \sum_{i \in U} \sum_{j \in U, j \neq i : \pi_{ij} = 0} \left[I_i \frac{|Y_i|^2}{2\pi_i} + I_j \frac{|Y_j|^2}{2\pi_j} \right]$$

we have

$$\mathbf{E}(\widehat{\mathbf{Cov}}_A(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{\mathcal{N}_K}))) \leq \mathbf{Cov}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W}'_{\mathcal{N}_K}))).$$

■

Since $\widehat{\mathbf{Var}}_A(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{\mathcal{N}_K})$ is a conservative variance estimator, the covariance estimator in A.19, provides a conservative variance estimator of any estimator of the form $\widehat{\delta} = X - Y$ such that $\mathbf{Var}(X - Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$ -2$\mathbf{Cov}(X, Y)$ which apply to all estimators under Assumption 3.1. However, under Assumption 3.2 and Assumption 4.1, we have estimators of the form $\widehat{\delta} = (X - Y) + (W - Z)$ such that $\mathbf{Var}((X - Y) + (W - Z)) = \mathbf{Var}(X) + \mathbf{Var}(Y) + \mathbf{Var}(W) + \mathbf{Var}(Z)$ -2$\mathbf{Cov}(X, Y)$ + 2$\mathbf{Cov}(X, W)$ -2$\mathbf{Cov}(X, Z)$ -2$\mathbf{Cov}(Y, W)$ +2$\mathbf{Cov}(Y, Z)$ -2$\mathbf{Cov}(W, Z)$. To get conservative variance estimator of any estimator of the second form, $\widehat{\mathbf{Cov}}_A(X, Y)$ can be used to estimate covariance components with negative coefficient while covariance components with positive coefficient need another covariance estimator that is guaranteed to have expectation greater than or equal to the true covariance.

We provide the following covariance estimator,

$$\widehat{\mathbf{Cov}}_B(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{\mathcal{N}_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W}'_{\mathcal{N}_K}))=$$

$$\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U, j\neq i:\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K}))>0}\left[\frac{I_i(W,\mathbf{W}_{\mathcal{N}_K})I_j(W',\mathbf{W}'_{\mathcal{N}_K}))}{\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K})))}\frac{Y_i^{obs}(W,\mathbf{W}_{\mathcal{N}_K})}{\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}\frac{Y_j^{obs}(W',\mathbf{W}'_{\mathcal{N}_K}))}{\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))}\right.$$

$$\times[\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K})))-\pi_i(W,\mathbf{W}_{\mathcal{N}_K})\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))]]$$

$$+\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U:\pi_{ij}((W,\mathbf{W}_{\mathcal{N}_K}),(W',\mathbf{W}'_{\mathcal{N}_K})))=0}\left[\frac{I_i(W,\mathbf{W}_{\mathcal{N}_K})Y_i^2}{2\pi_i(W,\mathbf{W}_{\mathcal{N}_K})}+\frac{I_j(W',\mathbf{W}'_{\mathcal{N}_K}))Y_j^2}{2\pi_j(W',\mathbf{W}'_{\mathcal{N}_K}))}\right] \quad (A.24)$$

Proposition A.2.

$$\mathbf{E}(\widehat{\mathbf{Cov}}_B(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{\mathcal{N}_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W}'_{\mathcal{N}_K})))\geq\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W,\mathbf{W}_{\mathcal{N}_K}),\bar{Y}_{HT}^{obs}(W',\mathbf{W}'_{\mathcal{N}_K}))). \quad (A.25)$$

The proof follows by Young's inequality in equations A.22 and A.23.

Appendix B

Properties of Horvitz-Thompson Estimators in Chapter 3

First, let $\mathbf{W}_\ell^* = (W_{\ell,1}^*, W_{\ell,2}^*, \ldots, W_{\ell,K}^*) \in \{0,1\}^K$ denote the treatment vector assignment of length K where the first ℓ nearest neighbors are given treatment and the rest are control:

$$W_{\ell,j}^* = \begin{cases} 1, & j \leq \ell, \\ 0, & \text{otherwise.} \end{cases} \tag{B.1}$$

Note that $\mathbf{W}_K^* = \mathbf{1}$, and define $\mathbf{W}_0^* = \mathbf{0}$. The *average ℓ^{th}–nearest neighbor indirect effect* (AℓNNIE) is defined as

$$\delta_\ell = \frac{1}{N} \sum_{i=1}^{N} (y_i(0, \mathbf{W}_\ell^*) - y_i(0, \mathbf{W}_{\ell-1}^*)). \tag{B.2}$$

Note that $\mathbf{W}_\ell^*$ and $\mathbf{W}_{\ell-1}^*$ are identical except that $W_{\ell,\ell}^* = 1$ and $W_{\ell-1,\ell}^* = 0$. Hence, δ_ℓ may be interpreted as the average difference in response due to the treatment status of the ℓ^{th}–nearest-neighbor. Additionally, under KNNIM, the AIE is the sum of the AℓNNIEs.

Lemma B.1.

$$\delta_{ind} = \sum_{\ell=1}^{K} \delta_\ell. \tag{B.3}$$

Proof.

$$\begin{aligned}
\delta_{1^{st}} + \delta_{2^{nd}} + \delta_{3^{rd}} + \dots + \delta_{\ell} + \dots + \delta_K = \\
&\bar{y}(0, \mathbf{W}_1^*) - \bar{y}(0, \mathbf{W}_0^*) \\
&+ \bar{y}(0, \mathbf{W}_2^*) - \bar{y}(0, \mathbf{W}_1^*) \\
&+ \bar{y}(0, \mathbf{W}_3^*) - \bar{y}(0, \mathbf{W}_2^*) \\
&+ \dots \\
&+ \bar{y}(0, \mathbf{W}_{\ell}^*) - \bar{y}(0, \mathbf{W}_{\ell-1}^*) \\
&+ \dots \\
&+ \bar{y}(0, \mathbf{W}_K^*) - \bar{y}(0, \mathbf{W}_{K-1}^*) \\
&= \bar{y}(0, \mathbf{W}_K^*) - \bar{y}(0, \mathbf{W}_0^*) \\
&= \bar{y}(0, \mathbf{1}) - \bar{y}(0, \mathbf{0}) \\
&= \delta_{ind}.
\end{aligned} \tag{B.4}$$

■

Lemma B.2.

$$\delta_{tot} = \delta_{dir} + \delta_{ind} \tag{B.5}$$

Proof.

$$\delta_{dir} + \delta_{ind} = \bar{y}(1, \mathbf{1}) - \bar{y}(0, \mathbf{1}) + \bar{y}(0, \mathbf{1}) - \bar{y}(0, \mathbf{0}) = \bar{y}(1, \mathbf{1}) - \bar{y}(0, \mathbf{0}) = \delta_{tot} \tag{B.6}$$

■

B.1 Properties of HT-ATOTE under K-NIA

Now, we find the expected value and the variance of HT-ATOT estimator.

$$\widehat{\delta}_{HT,tot} = \bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0}). \tag{B.7}$$

B.1.1 The Expected Value of HT-ATOTE

$$\begin{aligned}\mathbf{E}(\widehat{\delta}_{HT,tot}) = \mathbf{E}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})) = \mathbf{E}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) - \mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{0}))\\ = \bar{y}(1,\mathbf{1}) - \bar{y}(0,\mathbf{0}) = \delta_{tot}.\end{aligned} \tag{B.8}$$

B.1.2 The Variance of HT-ATOTE

We derive the variance using the properties $\mathbf{Var}(X-Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$ -2$\mathbf{Cov}(X,Y)$ and $\mathbf{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$.

First, we derive $\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}))$ as follows:

$$\begin{aligned}\big(\mathbf{E}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})\big)^2 &= \left(\mathbf{E}\left(\frac{1}{N}\sum_{i=1}^{N} I_i(1,\mathbf{1})\frac{Y^{obs}_i(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right)\right)^2\\ &= \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\left(I_i(1,\mathbf{1})\right)\frac{y_i(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right)^2\\ &= \frac{1}{N^2}\left(\sum_{i=1}^{N} y_i(1,\mathbf{1})\right)^2\\ &= \frac{1}{N^2}\sum_{i=1}^{N} y_i^2(1,\mathbf{1}) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(1,\mathbf{1})y_j(1,\mathbf{1})\end{aligned} \tag{B.9}$$

$$
\begin{aligned}
\mathbf{E}((\bar{Y}^{obs}_{HT}(1,\mathbf{1}))^2) &= \mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} I_i(1,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right)^2\right] \\
&= \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i^2(1,\mathbf{1})\frac{Y_i^{2^{obs}}(1,\mathbf{1})}{\pi_i^2(1,\mathbf{1})}\right)\right. \\
&\quad \left.+\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(1,\mathbf{1})I_j(1,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})Y_j^{obs}(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})}\right)\right] \\
&= \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i^2(1,\mathbf{1})\frac{Y_i^{2^{obs}}(1,\mathbf{1})}{\pi_i^2(1,\mathbf{1})}\right)\right] \\
&\quad +\mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(1,\mathbf{1})I_j(1,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})Y_j^{obs}(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})}\right)\right] \\
&= \left(\frac{1}{N^2}\sum_{i=1}^{N} \mathbf{E}[I_i^2(1,\mathbf{1})]\frac{y_i^2(1,\mathbf{1})}{\pi_i^2(1,\mathbf{1})}\right) \\
&\quad +\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} \mathbf{E}[I_i(1,\mathbf{1})I_j(1,\mathbf{1})]\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})}\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \pi_i(1,\mathbf{1})\frac{y_i^2(1,\mathbf{1})}{\pi_i^2(1,\mathbf{1})} \\
&\quad +\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} \pi_{ij}(1,\mathbf{1})\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})} \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \frac{y_i^2(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} \pi_{ij}(1,\mathbf{1})\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})} \qquad \text{(B.10)}
\end{aligned}
$$

Hence, the variance of $\bar{Y}^{obs}_{HT}(1,\mathbf{1})$ is

$$
\begin{aligned}
\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) &= \mathbf{E}((\bar{Y}^{obs}_{HT}(1,\mathbf{1}))^2) - \big(\mathbf{E}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})\big)^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\frac{y_i^2(1,\mathbf{1})}{\pi_i(1,\mathbf{1})} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}(1,\mathbf{1})\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})} \\
&\quad - \frac{1}{N^2}\sum_{i=1}^{N}y_i^2(1,\mathbf{1}) - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}y_i(1,\mathbf{1})y_j(1,\mathbf{1})
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{N^2}\sum_{i=1}^{N}\pi_i(1,\mathbf{1})[1-\pi_i(1,\mathbf{1})]\left[\frac{y_i(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right]^2 \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(1,\mathbf{1}) - \pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})]\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})}. \qquad \text{(B.11)}
\end{aligned}
$$

Similarly, the variance of $\bar{Y}^{obs}_{HT}(0,\mathbf{0})$ is

$$
\begin{aligned}
\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) &= \mathbf{E}((\bar{Y}^{obs}_{HT}(0,\mathbf{0}))^2) - \big(\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})\big)^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\frac{y_i^2(0,\mathbf{0})}{\pi_i(0,\mathbf{0})} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}(0,\mathbf{0})\frac{y_i(0,\mathbf{0})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})} \\
&\quad - \frac{1}{N^2}\sum_{i=1}^{N}y_i^2(0,\mathbf{0}) - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}y_i(0,\mathbf{0})y_j(0,\mathbf{0})
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{N^2}\sum_{i=1}^{N}\pi_i(0,\mathbf{0})[1-\pi_i(0,\mathbf{0})]\left[\frac{y_i(0,\mathbf{0})}{\pi_i(0,\mathbf{0})}\right]^2 \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{0}) - \pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})]\frac{y_i(0,\mathbf{0})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})}. \qquad \text{(B.12)}
\end{aligned}
$$

Next, we find $\mathbf{E}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})\bar{Y}^{obs}_{HT}(0,\mathbf{0}))$ and $\mathbf{E}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}))\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{0}))$:

$$
\begin{aligned}
&\mathbf{E}\left[\bar{Y}^{obs}_{HT}(1,\mathbf{1})\bar{Y}^{obs}_{HT}(0,\mathbf{0})\right] = \\
&\qquad \mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} I_i(1,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right)\left(\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{0})\frac{Y_i^{obs}(0,\mathbf{0})}{\pi_i(0,\mathbf{0})}\right)\right] \\
&\qquad = \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i(1,\mathbf{1})I_i(0,\mathbf{0})\frac{Y_i^{obs}(1,\mathbf{1})Y_i^{obs}(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_i(0,\mathbf{0})}\right)\right. \\
&\qquad \left. + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(1,\mathbf{1})I_j(0,\mathbf{0})\frac{Y_i^{obs}(1,\mathbf{1})Y_j^{obs}(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})}\right)\right] \\
&\qquad = \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i(1,\mathbf{1})I_i(0,\mathbf{0})\frac{Y_i^{obs}(1,\mathbf{1})Y_i^{obs}(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_i(0,\mathbf{0})}\right)\right] \\
&\qquad + \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(1,\mathbf{1})I_j(0,\mathbf{0})\frac{Y_i^{obs}(1,\mathbf{1})Y_j^{obs}(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})}\right)\right] \\
&\qquad = \left(\frac{1}{N^2}\sum_{i=1}^{N}\mathbf{E}\left[I_i(1,\mathbf{1})I_i(0,\mathbf{0})\right]\frac{y_i(1,\mathbf{1})y_i(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_i(0,\mathbf{0})}\right) \\
&\qquad + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\mathbf{E}\left[I_i(1,\mathbf{1})I_j(0,\mathbf{0})\right]\frac{y_i(1,\mathbf{1})y_i(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})}\right) \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\pi_i((1,\mathbf{1}),(0,\mathbf{0}))\frac{y_i(1,\mathbf{1})y_i(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_i(0,\mathbf{0})} \\
&\qquad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{1}),(0,\mathbf{0}))\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})} \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{1}),(0,\mathbf{0}))\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})} \qquad \text{(B.13)}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{E}(\bar{Y}_{HT}^{obs}(1,\mathbf{1}))\mathbf{E}(\bar{Y}_{HT}^{obs}(0,\mathbf{0})) = \\
&\qquad \mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N} I_i(1,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right]\mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{0})\frac{Y_i^{obs}(0,\mathbf{0})}{\pi_i(0,\mathbf{0})}\right] \\
&\qquad = \left[\frac{1}{N}\sum_{i=1}^{N} y_i(1,\mathbf{1})\right]\left[\frac{1}{N}\sum_{i=1}^{N} y_i(0,\mathbf{0})\right] \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{1})y_i(0,\mathbf{0}) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(1,\mathbf{1})y_j(0,\mathbf{0}) \qquad \text{(B.14)}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{Cov}(\bar{Y}_{HT}^{obs}(1,\mathbf{1}),\bar{Y}_{HT}^{obs}(0,\mathbf{0})) = \mathbf{E}\left[\bar{Y}_{HT}^{obs}(1,\mathbf{1})\bar{Y}_{HT}^{obs}(0,\mathbf{0})\right] - \mathbf{E}(\bar{Y}_{HT}^{obs}(1,\mathbf{1}))\mathbf{E}(\bar{Y}_{HT}^{obs}(0,\mathbf{0})) \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{1}),(0,\mathbf{0}))\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})} \\
&\qquad - \frac{1}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{1})y_i(0,\mathbf{0}) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(1,\mathbf{1})y_j(0,\mathbf{0}) \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{1}),(0,\mathbf{0})) - \pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})\right]\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})} \\
&\qquad\qquad - \frac{1}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{1})y_i(0,\mathbf{0}). \qquad \text{(B.15)}
\end{aligned}
$$

Then, we have the following:

$$
\begin{aligned}
\mathbf{Var}(\widehat{\delta}_{HT,tot}) = \mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})) &= \mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) + \mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) \\
&\quad - 2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \pi_i(1,\mathbf{1})[1-\pi_i(1,\mathbf{1})]\left[\frac{y_i(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right]^2 \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(1,\mathbf{1}) - \pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})]\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})} \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N} \pi_i(0,\mathbf{0})[1-\pi_i(0,\mathbf{0})]\left[\frac{y_i(0,\mathbf{0})}{\pi_i(0,\mathbf{0})}\right]^2 \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{0}) - \pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})]\frac{y_i(0,\mathbf{0})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})} \\
&\quad - 2\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{1}),(0,\mathbf{0})) - \pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})\right]\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})}\right. \\
&\qquad \left. -\frac{1}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{1})y_i(0,\mathbf{0})\right). \qquad \text{(B.16)}
\end{aligned}
$$

B.2 Properties of HT-ADEE under K-NIA

Now, we find the expected value and the variance of HT-ADE estimator.

$$
\widehat{\delta}_{HT,dir} = \bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{1}). \qquad \text{(B.17)}
$$

B.2.1 The Expected Value of HT-ADEE

$$
\begin{aligned}
\mathbf{E}(\widehat{\delta}_{HT,dir}) = \mathbf{E}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{1})) &= \mathbf{E}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) - \mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{1})) \\
&= \bar{y}(1,\mathbf{1}) - \bar{y}(0,\mathbf{1}) = \delta_{dir}. \qquad \text{(B.18)}
\end{aligned}
$$

B.2.2 The Variance of HT-ADEE

We derive the variance using the properties $\mathbf{Var}(X-Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$ -2$\mathbf{Cov}(X,Y)$ and $\mathbf{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$.

First, we derive $\mathbf{Var}(\bar{Y}_{HT}^{obs}(1,\mathbf{1}))$ as follows:

$$\begin{aligned}
\left(\mathbf{E}(\bar{Y}_{HT}^{obs}(1,\mathbf{1})\right)^2 &= \left(\mathbf{E}\left(\frac{1}{N}\sum_{i=1}^{N} I_i(1,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right)\right)^2 \\
&= \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\left(I_i(1,\mathbf{1})\right)\frac{y_i(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right)^2 \\
&= \frac{1}{N^2}\left(\sum_{i=1}^{N} y_i(1,\mathbf{1})\right)^2 \\
\\
&= \frac{1}{N^2}\sum_{i=1}^{N} y_i^2(1,\mathbf{1}) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(1,\mathbf{1})y_j(1,\mathbf{1})
\end{aligned} \tag{B.19}$$

$$
\begin{aligned}
\mathbf{E}((\bar{Y}_{HT}^{obs}(1,\mathbf{1}))^2) &= \mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} I_i(1,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right)^2\right] \\
&= \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i^2(1,\mathbf{1})\frac{Y_i^{2^{obs}}(1,\mathbf{1})}{\pi_i^2(1,\mathbf{1})}\right)\right. \\
&\quad \left. + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(1,\mathbf{1})I_j(1,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})Y_j^{obs}(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})}\right)\right] \\
&= \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i^2(1,\mathbf{1})\frac{Y_i^{2^{obs}}(1,\mathbf{1})}{\pi_i^2(1,\mathbf{1})}\right)\right] \\
&\quad + \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(1,\mathbf{1})I_j(1,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})Y_j^{obs}(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})}\right)\right] \\
&= \left(\frac{1}{N^2}\sum_{i=1}^{N} \mathbf{E}[I_i^2(1,\mathbf{1})]\frac{y_i^2(1,\mathbf{1})}{\pi_i^2(1,\mathbf{1})}\right) \\
&\quad + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} \mathbf{E}[I_i(1,\mathbf{1})I_j(1,\mathbf{1})]\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})}\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \pi_i(1,\mathbf{1})\frac{y_i^2(1,\mathbf{1})}{\pi_i^2(1,\mathbf{1})} \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} \pi_{ij}(1,\mathbf{1})\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})} \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \frac{y_i^2(1,\mathbf{1})}{\pi_i(1,\mathbf{1})} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} \pi_{ij}(1,\mathbf{1})\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})} \qquad \text{(B.20)}
\end{aligned}
$$

Hence, the variance of $\bar{Y}_{HT}^{obs}(1,\mathbf{1})$ is

$$\begin{aligned}\mathbf{Var}(\bar{Y}_{HT}^{obs}(1,\mathbf{1})) &= \mathbf{E}((\bar{Y}_{HT}^{obs}(1,\mathbf{1}))^2) - \big(\mathbf{E}(\bar{Y}_{HT}^{obs}(1,\mathbf{1})\big)^2 \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\frac{y_i^2(1,\mathbf{1})}{\pi_i(1,\mathbf{1})} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}(1,\mathbf{1})\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})} \\ &\quad - \frac{1}{N^2}\sum_{i=1}^{N}y_i^2(1,\mathbf{1}) - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}y_i(1,\mathbf{1})y_j(1,\mathbf{1})\end{aligned}$$

$$\begin{aligned}&= \frac{1}{N^2}\sum_{i=1}^{N}\pi_i(1,\mathbf{1})[1-\pi_i(1,\mathbf{1})]\left[\frac{y_i(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right]^2 \\ &\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(1,\mathbf{1}) - \pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})]\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})}. \qquad \text{(B.21)}\end{aligned}$$

Similarly, the variance of $\bar{Y}_{HT}^{obs}(0,\mathbf{1})$ is

$$\begin{aligned}\mathbf{Var}(\bar{Y}_{HT}^{obs}(0,\mathbf{1})) &= \mathbf{E}((\bar{Y}_{HT}^{obs}(0,\mathbf{1}))^2) - \big(\mathbf{E}(\bar{Y}_{HT}^{obs}(0,\mathbf{1})\big)^2 \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\frac{y_i^2(0,\mathbf{1})}{\pi_i(0,\mathbf{1})} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}(0,\mathbf{1})\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})} \\ &\quad - \frac{1}{N^2}\sum_{i=1}^{N}y_i^2(0,\mathbf{1}) - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}y_i(0,\mathbf{1})y_j(0,\mathbf{1})\end{aligned}$$

$$\begin{aligned}&= \frac{1}{N^2}\sum_{i=1}^{N}\pi_i(0,\mathbf{1})[1-\pi_i(0,\mathbf{1})]\left[\frac{y_i(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right]^2 \\ &\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{1}) - \pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})]\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})}. \qquad \text{(B.22)}\end{aligned}$$

Next, we find $\mathbf{E}(\bar{Y}_{HT}^{obs}(1,\mathbf{1})\bar{Y}_{HT}^{obs}(0,\mathbf{1}))$ and $\mathbf{E}(\bar{Y}_{HT}^{obs}(1,\mathbf{1}))\mathbf{E}(\bar{Y}_{HT}^{obs}(0,\mathbf{1}))$:

$$\begin{aligned}
&\mathbf{E}\left[\bar{Y}^{obs}_{HT}(1,\mathbf{1})\bar{Y}^{obs}_{HT}(0,\mathbf{1})\right] = \\
&\qquad \mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} I_i(1,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right)\left(\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{1})\frac{Y_i^{obs}(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right)\right] \\
&\qquad = \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i(1,\mathbf{1})I_i(0,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})Y_i^{obs}(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_i(0,\mathbf{1})}\right)\right. \\
&\qquad \left. + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(1,\mathbf{1})I_j(0,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})Y_j^{obs}(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})}\right)\right] \\
&\qquad = \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i(1,\mathbf{1})I_i(0,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})Y_i^{obs}(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_i(0,\mathbf{1})}\right)\right] \\
&\qquad + \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(1,\mathbf{1})I_j(0,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})Y_j^{obs}(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})}\right)\right] \\
&\qquad = \left(\frac{1}{N^2}\sum_{i=1}^{N}\mathbf{E}\left[I_i(1,\mathbf{1})I_i(0,\mathbf{1})\right]\frac{y_i(1,\mathbf{1})y_i(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_i(0,\mathbf{1})}\right) \\
&\qquad + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\mathbf{E}\left[I_i(1,\mathbf{1})I_j(0,\mathbf{1})\right]\frac{y_i(1,\mathbf{1})y_i(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})}\right) \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\pi_i((1,\mathbf{1}),(0,\mathbf{1}))\frac{y_i(1,\mathbf{1})y_i(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_i(0,\mathbf{1})} \\
&\qquad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{1}),(0,\mathbf{1}))\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})} \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{1}),(0,\mathbf{1}))\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})} \qquad \text{(B.23)}
\end{aligned}$$

$$
\begin{aligned}
&\mathbf{E}(\bar{Y}_{HT}^{obs}(1,\mathbf{1}))\mathbf{E}(\bar{Y}_{HT}^{obs}(0,\mathbf{1})) = \\
&\qquad \mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N} I_i(1,\mathbf{1})\frac{Y_i^{obs}(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right]\mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{1})\frac{Y_i^{obs}(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right] \\
&\qquad = \left[\frac{1}{N}\sum_{i=1}^{N} y_i(1,\mathbf{1})\right]\left[\frac{1}{N}\sum_{i=1}^{N} y_i(0,\mathbf{1})\right] \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{1})y_i(0,\mathbf{1}) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(1,\mathbf{1})y_j(0,\mathbf{1}) \qquad \text{(B.24)}
\end{aligned}
$$

$$
\begin{aligned}
&\mathbf{Cov}(\bar{Y}_{HT}^{obs}(1,\mathbf{1}),\bar{Y}_{HT}^{obs}(0,\mathbf{1})) = \mathbf{E}\left[\bar{Y}_{HT}^{obs}(1,\mathbf{1})\bar{Y}_{HT}^{obs}(0,\mathbf{1})\right] - \mathbf{E}(\bar{Y}_{HT}^{obs}(1,\mathbf{1}))\mathbf{E}(\bar{Y}_{HT}^{obs}(0,\mathbf{1})) \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{1}),(0,\mathbf{1}))\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})} \\
&\qquad - \frac{1}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{1})y_i(0,\mathbf{1}) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(1,\mathbf{1})y_j(0,\mathbf{1}) \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{1}),(0,\mathbf{1})) - \pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})\right]\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})} \\
&\qquad - \frac{1}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{1})y_i(0,\mathbf{1}). \qquad \text{(B.25)}
\end{aligned}
$$

Then, we have the following:

$$\begin{aligned}\mathbf{Var}(\widehat{\delta}_{HT,dir}) = \mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{1})) &= \mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) + \mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{1})) \\ &\quad - 2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{1})) \\ &= \frac{1}{N^2}\sum_{i=1}^{N} \pi_i(1,\mathbf{1})[1-\pi_i(1,\mathbf{1})]\left[\frac{y_i(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right]^2 \\ &\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(1,\mathbf{1}) - \pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})]\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})} \\ &\quad + \frac{1}{N^2}\sum_{i=1}^{N} \pi_i(0,\mathbf{1})[1-\pi_i(0,\mathbf{1})]\left[\frac{y_i(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right]^2 \\ &\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{1}) - \pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})]\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})} \\ &\quad - 2\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{1}),(0,\mathbf{1})) - \pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})\right]\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})}\right. \\ &\qquad \left. - \frac{1}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{1})y_i(0,\mathbf{1})\right). \qquad \text{(B.26)}\end{aligned}$$

B.3 Properties of HT-AIEE under K-NIA

Here, we find the expected value and the variance of HT-AIE estimator

$$\widehat{\delta}_{HT,ind} = \bar{Y}^{obs}_{HT}(0,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0}). \qquad \text{(B.27)}$$

Note that, the average total effect estimator is the sum of the average direct and indirect estimators:

Lemma B.3.

$$\widehat{\delta}_{HT,tot} = \widehat{\delta}_{HT,dir} + \widehat{\delta}_{HT,ind} \qquad \text{(B.28)}$$

Proof.

$$\begin{aligned}\widehat{\delta}_{HT,dir} + \widehat{\delta}_{HT,ind} &= \bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{1}) + \bar{Y}^{obs}_{HT}(0,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0}) \\ &= \bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0}) \\ &= \widehat{\delta}_{HT,tot} \quad \text{(B.29)}\end{aligned}$$

■

B.3.1 The Expected Value of HT-AIEE

$$\begin{aligned}\mathbf{E}(\widehat{\delta}_{HT,ind}) = \mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})) &= \mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{1})) - \mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) \\ &= \bar{y}(0,\mathbf{1}) - \bar{y}(0,\mathbf{0}) = \delta_{ind}. \quad \text{(B.30)}\end{aligned}$$

B.3.2 The Variance of HT-AIEE

We derive the variance using the properties $\mathbf{Var}(X-Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$ -2$\mathbf{Cov}(X,Y)$ and $\mathbf{Cov}(X,Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$.

First, we derive $\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}))$ as follows:

$$\begin{aligned}\left(\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{1})\right)^2 &= \left(\mathbf{E}\left(\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{1})\frac{Y_i^{obs}(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right)\right)^2\\ &= \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{E}\left(I_i(0,\mathbf{1})\right)\frac{y_i(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right)^2\\ &= \frac{1}{N^2}\left(\sum_{i=1}^{N} y_i(0,\mathbf{1})\right)^2\end{aligned}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} y_i^2(0,\mathbf{1}) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(0,\mathbf{1})y_j(0,\mathbf{1}) \tag{B.31}$$

$$\begin{aligned}
\mathbf{E}((\bar{Y}^{obs}_{HT}(0,\mathbf{1}))^2) &= \mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{1})\frac{Y_i^{obs}(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right)^2\right] \\
&= \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i^2(0,\mathbf{1})\frac{Y_i^{2^{obs}}(0,\mathbf{1})}{\pi_i^2(0,\mathbf{1})}\right)\right. \\
&\quad \left. + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(0,\mathbf{1})I_j(0,\mathbf{1})\frac{Y_i^{obs}(0,\mathbf{1})Y_j^{obs}(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})}\right)\right] \\
&= \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i^2(0,\mathbf{1})\frac{Y_i^{2^{obs}}(0,\mathbf{1})}{\pi_i^2(0,\mathbf{1})}\right)\right] \\
&\quad + \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(0,\mathbf{1})I_j(0,\mathbf{1})\frac{Y_i^{obs}(0,\mathbf{1})Y_j^{obs}(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})}\right)\right] \\
&= \left(\frac{1}{N^2}\sum_{i=1}^{N} \mathbf{E}[I_i^2(0,\mathbf{1})]\frac{y_i^2(0,\mathbf{1})}{\pi_i^2(0,\mathbf{1})}\right) \\
&\quad + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} \mathbf{E}[I_i(0,\mathbf{1})I_j(0,\mathbf{1})]\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})}\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \pi_i(0,\mathbf{1})\frac{y_i^2(0,\mathbf{1})}{\pi_i^2(0,\mathbf{1})} \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} \pi_{ij}(0,\mathbf{1})\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})} \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \frac{y_i^2(0,\mathbf{1})}{\pi_i(0,\mathbf{1})} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} \pi_{ij}(0,\mathbf{1})\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})} \qquad \text{(B.32)}
\end{aligned}$$

Hence, the variance of $\bar{Y}^{obs}_{HT}(0,\mathbf{1})$ is

$$
\begin{aligned}
\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{1})) &= \mathbf{E}((\bar{Y}^{obs}_{HT}(0,\mathbf{1}))^2) - \big(\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{1})\big)^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\frac{y_i^2(0,\mathbf{1})}{\pi_i(0,\mathbf{1})} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}(0,\mathbf{1})\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})} \\
&\quad - \frac{1}{N^2}\sum_{i=1}^{N}y_i^2(0,\mathbf{1}) - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}y_i(0,\mathbf{1})y_j(0,\mathbf{1})
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{N^2}\sum_{i=1}^{N}\pi_i(0,\mathbf{1})[1-\pi_i(0,\mathbf{1})]\left[\frac{y_i(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right]^2 \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{1}) - \pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})]\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})}. \qquad \text{(B.33)}
\end{aligned}
$$

Similarly, the variance of $\bar{Y}^{obs}_{HT}(0,\mathbf{0})$ is

$$
\begin{aligned}
\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) &= \mathbf{E}((\bar{Y}^{obs}_{HT}(0,\mathbf{0}))^2) - \big(\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})\big)^2 \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\frac{y_i^2(0,\mathbf{0})}{\pi_i(0,\mathbf{0})} + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}(0,\mathbf{0})\frac{y_i(0,\mathbf{0})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})} \\
&\quad - \frac{1}{N^2}\sum_{i=1}^{N}y_i^2(0,\mathbf{0}) - \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}y_i(0,\mathbf{0})y_j(0,\mathbf{0})
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{N^2}\sum_{i=1}^{N}\pi_i(0,\mathbf{0})[1-\pi_i(0,\mathbf{0})]\left[\frac{y_i(0,\mathbf{0})}{\pi_i(0,\mathbf{0})}\right]^2 \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{0}) - \pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})]\frac{y_i(0,\mathbf{0})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})}. \qquad \text{(B.34)}
\end{aligned}
$$

Next, we find $\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{1})\bar{Y}^{obs}_{HT}(0,\mathbf{0}))$ and $\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}))\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{0}))$:

$$
\begin{aligned}
&\mathbf{E}\left[\bar{Y}^{obs}_{HT}(0,\mathbf{1})\bar{Y}^{obs}_{HT}(0,\mathbf{0})\right] = \\
&\qquad \mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{1})\frac{Y_i^{obs}(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right)\left(\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{0})\frac{Y_i^{obs}(0,\mathbf{0})}{\pi_i(0,\mathbf{0})}\right)\right] \\
&\qquad = \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i(0,\mathbf{1})I_i(0,\mathbf{0})\frac{Y_i^{obs}(0,\mathbf{1})Y_i^{obs}(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_i(0,\mathbf{0})}\right)\right. \\
&\qquad \left. + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(0,\mathbf{1})I_j(0,\mathbf{0})\frac{Y_i^{obs}(0,\mathbf{1})Y_j^{obs}(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})}\right)\right] \\
&\qquad = \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i(0,\mathbf{1})I_i(0,\mathbf{0})\frac{Y_i^{obs}(0,\mathbf{1})Y_i^{obs}(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_i(0,\mathbf{0})}\right)\right] \\
&\qquad + \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(0,\mathbf{1})I_j(0,\mathbf{0})\frac{Y_i^{obs}(0,\mathbf{1})Y_j^{obs}(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})}\right)\right] \\
&\qquad = \left(\frac{1}{N^2}\sum_{i=1}^{N}\mathbf{E}\left[I_i(0,\mathbf{1})I_i(0,\mathbf{0})\right]\frac{y_i(0,\mathbf{1})y_i(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_i(0,\mathbf{0})}\right) \\
&\qquad + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\mathbf{E}\left[I_i(0,\mathbf{1})I_j(0,\mathbf{0})\right]\frac{y_i(0,\mathbf{1})y_i(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})}\right) \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\pi_i((0,\mathbf{1}),(0,\mathbf{0}))\frac{y_i(0,\mathbf{1})y_i(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_i(0,\mathbf{0})} \\
&\qquad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((0,\mathbf{1}),(0,\mathbf{0}))\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})} \\
&\qquad = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((0,\mathbf{1}),(0,\mathbf{0}))\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})} \qquad \text{(B.35)}
\end{aligned}
$$

$$\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}))\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) =$$

$$\begin{aligned}
&\mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{1})\frac{Y_i^{obs}(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right]\mathbf{E}\left[\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{0})\frac{Y_i^{obs}(0,\mathbf{0})}{\pi_i(0,\mathbf{0})}\right]\\
&= \left[\frac{1}{N}\sum_{i=1}^{N} y_i(0,\mathbf{1})\right]\left[\frac{1}{N}\sum_{i=1}^{N} y_i(0,\mathbf{0})\right]\\
&= \frac{1}{N^2}\sum_{i=1}^{N} y_i(0,\mathbf{1})y_i(0,\mathbf{0}) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(0,\mathbf{1})y_j(0,\mathbf{0}) \quad \text{(B.36)}
\end{aligned}$$

$$\begin{aligned}
\mathbf{Cov}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}),\bar{Y}^{obs}_{HT}(0,\mathbf{0})) &= \mathbf{E}\left[\bar{Y}^{obs}_{HT}(0,\mathbf{1})\bar{Y}^{obs}_{HT}(0,\mathbf{0})\right] - \mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}))\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{0}))\\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((0,\mathbf{1}),(0,\mathbf{0}))\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})}\\
&- \frac{1}{N^2}\sum_{i=1}^{N} y_i(0,\mathbf{1})y_i(0,\mathbf{0}) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(0,\mathbf{1})y_j(0,\mathbf{0})\\
&= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((0,\mathbf{1}),(0,\mathbf{0})) - \pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})\right]\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})}\\
&- \frac{1}{N^2}\sum_{i=1}^{N} y_i(0,\mathbf{1})y_i(0,\mathbf{0}). \quad \text{(B.37)}
\end{aligned}$$

Then, we have the following:

$$
\begin{aligned}
\mathbf{Var}(\widehat{\delta}_{HT,ind}) = \mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})) &= \mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{1})) + \mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) \\
&\quad - 2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \pi_i(0,\mathbf{1})[1-\pi_i(0,\mathbf{1})]\left[\frac{y_i(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right]^2 \\
&+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{1}) - \pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})]\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})} \\
&+ \frac{1}{N^2}\sum_{i=1}^{N} \pi_i(0,\mathbf{0})[1-\pi_i(0,\mathbf{0})]\left[\frac{y_i(0,\mathbf{0})}{\pi_i(0,\mathbf{0})}\right]^2 \\
&+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{0}) - \pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})]\frac{y_i(0,\mathbf{0})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})} \\
&- 2\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((0,\mathbf{1}),(0,\mathbf{0})) - \pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})\right]\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})}\right. \\
&\qquad \left. - \frac{1}{N^2}\sum_{i=1}^{N} y_i(0,\mathbf{1})y_i(0,\mathbf{0})\right). \qquad (B.38)
\end{aligned}
$$

B.4 Properties of HT-AℓNNIEE under K-NIA

Now, we find the expected value and the variance of HT-AℓNNIEE under $K-NIA$ assumption.

$$
\widehat{\delta}_{HT,\ell} = \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1}). \qquad (B.39)
$$

First, note that the indirect effect estimator $\widehat{\delta}_{HT,ind}$ is the sum of all ℓ-nearest neighbors indirect effects estimators ,i.e.,

Lemma B.4.

$$
\widehat{\delta}_{HT,ind} = \sum_{\ell=1}^{K} \widehat{\delta}_{HT,\ell}. \qquad (B.40)
$$

Proof. For $\ell \in \mathcal{N}_{ik}$, if we define $\mathbf{W}^*_\ell$ as previously, then we have the following:

$$\begin{aligned}
\widehat{\delta}_{HT,1^{st}} + \widehat{\delta}_{HT,2^{nd}} + \widehat{\delta}_{HT,3^{rd}} + \dots + \widehat{\delta}_{HT,\ell} + \dots + \widehat{\delta}_{HT,K} = \\
\bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_1) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_0) \\
+ \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_2) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_1) \\
+ \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_3) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_2) \\
+ \dots \\
+ \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_{\ell-1}) \\
+ \dots \\
+ \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_K) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_{K-1}) \\
= \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_K) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_0) \\
= \bar{Y}^{obs}_{HT}(0, \mathbf{1}) - \bar{Y}^{obs}_{HT}(0, \mathbf{0}) \\
= \widehat{\delta}_{HT,ind}
\end{aligned} \tag{B.41}$$

■

B.4.1 The Expected Value of HT-AℓNNIEE

$$\begin{aligned}
\mathbf{E}(\widehat{\delta}_{HT,\ell}) = \mathbf{E}(\bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_{\ell-1})) = \mathbf{E}(\bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_\ell)) - \mathbf{E}(\bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_{\ell-1})) \\
= \bar{y}(0, \mathbf{W}^*_\ell) - \bar{y}(0, \mathbf{W}^*_{\ell-1}) = \delta_\ell.
\end{aligned} \tag{B.42}$$

B.4.2 The Variance of HT-AℓNNIEE

We find the variance by using the property $\mathbf{Var}(X - Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$ -2$\mathbf{Cov}(X, Y)$ and the property $\mathbf{Cov}(X, Y) = \mathbf{E}(XY) - \mathbf{E}(X)\mathbf{E}(Y)$.

First, we derive $\mathbf{Var}(\bar{Y}_{HT}^{obs}(0, \mathbf{W}_\ell^*))$ as follows:

$$\begin{aligned}
\left(\mathbf{E}(\bar{Y}_{HT}^{obs}(0, \mathbf{W}_\ell^*)\right)^2 &= \left(\mathbf{E}\left(\frac{1}{N}\sum_{i=1}^{N} I_i(0, \mathbf{W}_\ell^*)\frac{Y_i^{obs}(0, \mathbf{W}_\ell^*)}{\pi_i(0, \mathbf{W}_\ell^*)}\right)\right)^2 \\
&= \left(\frac{1}{N}\sum_{i=1}^{N} \mathbf{E}\left(I_i(0, \mathbf{W}_\ell^*)\right)\frac{y_i(0, \mathbf{W}_\ell^*)}{\pi_i(0, \mathbf{W}_\ell^*)}\right)^2 \\
&= \frac{1}{N^2}\left(\sum_{i=1}^{N} y_i(0, \mathbf{W}_\ell^*)\right)^2
\end{aligned}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} y_i^2(0, \mathbf{W}_\ell^*) + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(0, \mathbf{W}_\ell^*)y_j(0, \mathbf{W}_\ell^*) \tag{B.43}$$

$$
\begin{aligned}
\mathbf{E}((\bar{Y}_{HT}^{obs}(0,\mathbf{W}_\ell^*))^2) &= \mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{W}_\ell^*)\frac{Y_i^{obs}(0,\mathbf{W}_\ell^*)}{\pi_i(0,\mathbf{W}_\ell^*)}\right)^2\right] \\
&= \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i^2(0,\mathbf{W}_\ell^*)\frac{Y_i^{2^{obs}}(0,\mathbf{W}_\ell^*)}{\pi_i^2(0,\mathbf{W}_\ell^*)}\right)\right. \\
&\left.+\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(0,\mathbf{W}_\ell^*)I_j(0,\mathbf{W}_\ell^*)\frac{Y_i^{obs}(0,\mathbf{W}_\ell^*)Y_j^{obs}(0,\mathbf{W}_\ell^*)}{\pi_i(0,\mathbf{W}_\ell^*)\pi_j(0,\mathbf{W}_\ell^*)}\right)\right] \\
&= \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i^2(0,\mathbf{W}_\ell^*)\frac{Y_i^{2^{obs}}(0,\mathbf{W}_\ell^*)}{\pi_i^2(0,\mathbf{W}_\ell^*)}\right)\right] \\
&+\mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(0,\mathbf{W}_\ell^*)I_j(0,\mathbf{W}_\ell^*)\frac{Y_i^{obs}(0,\mathbf{W}_\ell^*)Y_j^{obs}(0,\mathbf{W}_\ell^*)}{\pi_i(0,\mathbf{W}_\ell^*)\pi_j(0,\mathbf{W}_\ell^*)}\right)\right] \\
&= \left(\frac{1}{N^2}\sum_{i=1}^{N} \mathbf{E}[I_i^2(0,\mathbf{W}_\ell^*)]\frac{y_i^2(0,\mathbf{W}_\ell^*)}{\pi_i^2(0,\mathbf{W}_\ell^*)}\right) \\
&+\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} \mathbf{E}[I_i(0,\mathbf{W}_\ell^*)I_j(0,\mathbf{W}_\ell^*)]\frac{y_i(0,\mathbf{W}_\ell^*)y_j(0,\mathbf{W}_\ell^*)}{\pi_i(0,\mathbf{W}_\ell^*)\pi_j(0,\mathbf{W}_\ell^*)}\right) \\
&= \frac{1}{N^2}\sum_{i=1}^{N} \pi_i(0,\mathbf{W}_\ell^*)\frac{y_i^2(0,\mathbf{W}_\ell^*)}{\pi_i^2(0,\mathbf{W}_\ell^*)} \\
&+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} \pi_{ij}(0,\mathbf{W}_\ell^*)\frac{y_i(0,\mathbf{W}_\ell^*)y_j(0,\mathbf{W}_\ell^*)}{\pi_i(0,\mathbf{W}_\ell^*)\pi_j(0,\mathbf{W}_\ell^*)} \\
&= \frac{1}{N^2}\sum_{i=1}^{N}\frac{y_i^2(0,\mathbf{W}_\ell^*)}{\pi_i(0,\mathbf{W}_\ell^*)}+\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} \pi_{ij}(0,\mathbf{W}_\ell^*)\frac{y_i(0,\mathbf{W}_\ell^*)y_j(0,\mathbf{W}_\ell^*)}{\pi_i(0,\mathbf{W}_\ell^*)\pi_j(0,\mathbf{W}_\ell^*)} \qquad \text{(B.44)}
\end{aligned}
$$

Hence, the variance of $\bar{Y}_{HT}^{obs}(0,\mathbf{W}_\ell^*)$ is

$$\begin{aligned}\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)) &= \mathbf{E}((\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell))^2) - \left(\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)\right)^2 \\ &= \frac{1}{N^2}\sum_{i=1}^N \frac{y_i^2(0,\mathbf{W}^*_\ell)}{\pi_i(0,\mathbf{W}^*_\ell)} + \frac{1}{N^2}\sum_{i=1}^N\sum_{j\neq i}\pi_{ij}(0,\mathbf{W}^*_\ell)\frac{y_i(0,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_\ell)}{\pi_i(0,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_\ell)} \\ &\quad - \frac{1}{N^2}\sum_{i=1}^N y_i^2(0,\mathbf{W}^*_\ell) - \frac{1}{N^2}\sum_{i=1}^N\sum_{j\neq i} y_i(0,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_\ell)\end{aligned}$$

$$\begin{aligned}&= \frac{1}{N^2}\sum_{i=1}^N \pi_i(0,\mathbf{W}^*_\ell)[1-\pi_i(0,\mathbf{W}^*_\ell)]\left[\frac{y_i(0,\mathbf{W}^*_\ell)}{\pi_i(0,\mathbf{W}^*_\ell)}\right]^2 \\ &+ \frac{1}{N^2}\sum_{i=1}^N\sum_{j\neq i}[\pi_{ij}(0,\mathbf{W}^*_\ell)-\pi_i(0,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_\ell)]\frac{y_i(0,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_\ell)}{\pi_i(0,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_\ell)}.\end{aligned} \tag{B.45}$$

Similarly, the variance of $\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})$ is

$$\begin{aligned}\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) &= \mathbf{E}((\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1}))^2) - \left(\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})\right)^2 \\ &= \frac{1}{N^2}\sum_{i=1}^N \frac{y_i^2(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_{\ell-1})} + \frac{1}{N^2}\sum_{i=1}^N\sum_{j\neq i}\pi_{ij}(0,\mathbf{W}^*_{\ell-1})\frac{y_i(0,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_{\ell-1})} \\ &\quad - \frac{1}{N^2}\sum_{i=1}^N y_i^2(0,\mathbf{W}^*_{\ell-1}) - \frac{1}{N^2}\sum_{i=1}^N\sum_{j\neq i} y_i(0,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_{\ell-1})\end{aligned}$$

$$\begin{aligned}&= \frac{1}{N^2}\sum_{i=1}^N \pi_i(0,\mathbf{W}^*_{\ell-1})[1-\pi_i(0,\mathbf{W}^*_{\ell-1})]\left[\frac{y_i(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_{\ell-1})}\right]^2 \\ &+ \frac{1}{N^2}\sum_{i=1}^N\sum_{j\neq i}[\pi_{ij}(0,\mathbf{W}^*_{\ell-1})-\pi_i(0,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_{\ell-1})]\frac{y_i(0,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_{\ell-1})}.\end{aligned} \tag{B.46}$$

Next, we find $\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1}))$ and $\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell))\mathbf{E}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1}))$:

$$
\begin{aligned}
&\mathbf{E}\left[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})\right] = \\
&\quad \mathbf{E}\left[\left(\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{W}^*_\ell)\frac{Y_i^{obs}(0,\mathbf{W}^*_\ell)}{\pi_i(0,\mathbf{W}^*_\ell)}\right)\left(\frac{1}{N}\sum_{i=1}^{N} I_i(0,\mathbf{W}^*_{\ell-1})\frac{Y_i^{obs}(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_{\ell-1})}\right)\right] \\
&\quad = \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i(0,\mathbf{W}^*_\ell)I_i(0,\mathbf{W}^*_{\ell-1})\frac{Y_i^{obs}(0,\mathbf{W}^*_\ell)Y_i^{obs}(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_\ell)\pi_i(0,\mathbf{W}^*_{\ell-1})}\right)\right. \\
&\quad + \left.\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(0,\mathbf{W}^*_\ell)I_j(0,\mathbf{W}^*_{\ell-1})\frac{Y_i^{obs}(0,\mathbf{W}^*_\ell)Y_j^{obs}(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_{\ell-1})}\right)\right] \\
&\quad = \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N} I_i(0,\mathbf{W}^*_\ell)I_i(0,\mathbf{W}^*_{\ell-1})\frac{Y_i^{obs}(0,\mathbf{W}^*_\ell)Y_i^{obs}(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_\ell)\pi_i(0,\mathbf{W}^*_{\ell-1})}\right)\right] \\
&\quad + \mathbf{E}\left[\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} I_i(0,\mathbf{W}^*_\ell)I_j(0,\mathbf{W}^*_{\ell-1})\frac{Y_i^{obs}(0,\mathbf{W}^*_\ell)Y_j^{obs}(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_{\ell-1})}\right)\right] \\
&\quad = \left(\frac{1}{N^2}\sum_{i=1}^{N}\mathbf{E}\left[I_i(0,\mathbf{W}^*_\ell)I_i(0,\mathbf{W}^*_{\ell-1})\right]\frac{y_i(0,\mathbf{W}^*_\ell)y_i(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_\ell)\pi_i(0,\mathbf{W}^*_{\ell-1})}\right) \\
&\quad + \left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\mathbf{E}\left[I_i(0,\mathbf{W}^*_\ell)I_j(0,\mathbf{W}^*_{\ell-1})\right]\frac{y_i(0,\mathbf{W}^*_\ell)y_i(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_{\ell-1})}\right) \\
&\quad = \frac{1}{N^2}\sum_{i=1}^{N}\pi_i((0,\mathbf{W}^*_\ell),(0,\mathbf{W}^*_{\ell-1}))\frac{y_i(0,\mathbf{W}^*_\ell)y_i(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_\ell)\pi_i(0,\mathbf{W}^*_{\ell-1})} \\
&\quad + \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((0,\mathbf{W}^*_\ell),(0,\mathbf{W}^*_{\ell-1}))\frac{y_i(0,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_{\ell-1})} \\
&\quad = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((0,\mathbf{W}^*_\ell),(0,\mathbf{W}^*_{\ell-1}))\frac{y_i(0,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_{\ell-1})} \qquad \text{(B.47)}
\end{aligned}
$$

$$\mathbf{E}(\bar{Y}_{HT}^{obs}(0,\mathbf{W}_\ell^*))\mathbf{E}(\bar{Y}_{HT}^{obs}(0,\mathbf{W}_{\ell-1}^*)) =$$

$$\mathbf{E}\left[\frac{1}{N}\sum_{i=1}^N I_i(0,\mathbf{W}_\ell^*)\frac{Y_i^{obs}(0,\mathbf{W}_\ell^*)}{\pi_i(0,\mathbf{W}_\ell^*)}\right]\mathbf{E}\left[\frac{1}{N}\sum_{i=1}^N I_i(0,\mathbf{W}_{\ell-1}^*)\frac{Y_i^{obs}(0,\mathbf{W}_{\ell-1}^*)}{\pi_i(0,\mathbf{W}_{\ell-1}^*)}\right]$$

$$= \left[\frac{1}{N}\sum_{i=1}^N y_i(0,\mathbf{W}_\ell^*)\right]\left[\frac{1}{N}\sum_{i=1}^N y_i(0,\mathbf{W}_{\ell-1}^*)\right]$$

$$= \frac{1}{N^2}\sum_{i=1}^N y_i(0,\mathbf{W}_\ell^*)y_i(0,\mathbf{W}_{\ell-1}^*) + \frac{1}{N^2}\sum_{i=1}^N\sum_{j\neq i} y_i(0,\mathbf{W}_\ell^*)y_j(0,\mathbf{W}_{\ell-1}^*) \quad \text{(B.48)}$$

$$\mathbf{Cov}(\bar{Y}_{HT}^{obs}(0,\mathbf{W}_\ell^*),\bar{Y}_{HT}^{obs}(0,\mathbf{W}_{\ell-1}^*)) = \mathbf{E}(\bar{Y}_{HT}^{obs}(0,\mathbf{W}_\ell^*)\bar{Y}_{HT}^{obs}(0,\mathbf{W}_{\ell-1}^*)) - \mathbf{E}(\bar{Y}_{HT}^{obs}(0,\mathbf{W}_\ell^*))\mathbf{E}(\bar{Y}_{HT}^{obs}(0,\mathbf{W}_{\ell-1}^*))$$

$$= \frac{1}{N^2}\sum_{i=1}^N\sum_{j\neq i}\pi_{ij}((0,\mathbf{W}_\ell^*),(0,\mathbf{W}_{\ell-1}^*))\frac{y_i(0,\mathbf{W}_\ell^*)y_j(0,\mathbf{W}_{\ell-1}^*)}{\pi_i(0,\mathbf{W}_\ell^*)\pi_j(0,\mathbf{W}_{\ell-1}^*)}$$

$$- \frac{1}{N^2}\sum_{i=1}^N y_i(0,\mathbf{W}_\ell^*)y_i(0,\mathbf{W}_{\ell-1}^*) - \frac{1}{N^2}\sum_{i=1}^N\sum_{j\neq i} y_i(0,\mathbf{W}_\ell^*)y_j(0,\mathbf{W}_{\ell-1}^*)$$

$$= \frac{1}{N^2}\sum_{i=1}^N\sum_{j\neq i}\left[\pi_{ij}((0,\mathbf{W}_\ell^*),(0,\mathbf{W}_{\ell-1}^*)) - \pi_i(0,\mathbf{W}_\ell^*)\pi_j(0,\mathbf{W}_{\ell-1}^*)\right]\frac{y_i(0,\mathbf{W}_\ell^*)y_j(0,\mathbf{W}_{\ell-1}^*)}{\pi_i(0,\mathbf{W}_\ell^*)\pi_j(0,\mathbf{W}_{\ell-1}^*)}$$

$$- \frac{1}{N^2}\sum_{i=1}^N y_i(0,\mathbf{W}_\ell^*)y_i(0,\mathbf{W}_{\ell-1}^*). \quad \text{(B.49)}$$

Then, we have the following:

$$\begin{aligned}\mathbf{Var}(\widehat{\delta}_{HT,\ell}) = \mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) &= \mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)) + \mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) \\ &- 2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell), \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1}) \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\pi_i(0,\mathbf{W}^*_\ell)[1-\pi_i(0,\mathbf{W}^*_\ell)]\left[\frac{y_i(0,\mathbf{W}^*_\ell)}{\pi_i(0,\mathbf{W}^*_\ell)}\right]^2 \\ &+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{W}^*_\ell) - \pi_i(0,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_\ell)]\frac{y_i(0,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_\ell)}{\pi_i(0,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_\ell)} \\ &+ \frac{1}{N^2}\sum_{i=1}^{N}\pi_i(0,\mathbf{W}^*_{\ell-1})[1-\pi_i(0,\mathbf{W}^*_{\ell-1})]\left[\frac{y_i(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_{\ell-1})}\right]^2 \\ &+ \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{W}^*_{\ell-1}) - \pi_i(0,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_{\ell-1})]\frac{y_i(0,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_{\ell-1})} \\ &- 2\left(\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((0,\mathbf{W}^*_\ell),(0,\mathbf{W}^*_{\ell-1})) - \pi_i(0,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_{\ell-1})\right]\frac{y_i(0,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_{\ell-1})}\right. \\ &\left. - \frac{1}{N^2}\sum_{i=1}^{N}y_i(0,\mathbf{W}^*_\ell)y_i(0,\mathbf{W}^*_{\ell-1})\right). \end{aligned} \tag{B.50}$$

B.5 Properties of HT-ATOTE under the No-Interaction between Direct and Indirect Effects Assumption

Under Assumption 3.2, if we assume that there is no interaction between direct and indirect effects, the unbiased Horvitz–Thompson estimator of ATOT is provided as follows.

$$\widehat{\delta^*}_{HT,tot} = \widehat{\delta^*}_{HT,dir} + \widehat{\delta^*}_{HT,ind}. \tag{B.51}$$

Lemma B.5. *For* $C_1 = C_2 = \frac{1}{2}$,

$$\widehat{\delta^*}_{HT,tot} = \widehat{\delta}_{HT,tot}. \tag{B.52}$$

Proof.

$$\begin{aligned}\widehat{\delta^*}_{HT,tot} = \widehat{\delta^*}_{HT,dir} + \widehat{\delta^*}_{HT,ind} &= C_1[\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{1})] + C_2[\bar{Y}^{obs}_{HT}(1,\mathbf{0}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})] \\ &+ C_1[\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(1,\mathbf{0})] + C_2[\bar{Y}^{obs}_{HT}(0,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})] \\ &= C_1\bar{Y}^{obs}_{HT}(1,\mathbf{1}) + C_1\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - C_2\bar{Y}^{obs}_{HT}(0,\mathbf{0}) - C_2\bar{Y}^{obs}_{HT}(0,\mathbf{0}) \\ &= 2C_1\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - 2C_2\bar{Y}^{obs}_{HT}(0,\mathbf{0})\end{aligned}$$

and if $C_1 = C_2 = \frac{1}{2}$, then $\widehat{\delta^*}_{HT,tot} = \widehat{\delta}_{HT,tot}$.

■

B.5.1 The Expected Value of HT-ATOTE

$$\mathbf{E}(\widehat{\delta^*}_{HT,tot}) = \mathbf{E}(\widehat{\delta^*}_{HT,dir}) + \mathbf{E}(\widehat{\delta^*}_{HT,ind}) = \delta_{dir} + \delta_{ind} = \delta_{tot}. \tag{B.53}$$

where $\mathbf{E}(\widehat{\delta^*}_{HT,dir})$ and $\mathbf{E}(\widehat{\delta^*}_{HT,ind})$ are provided in the next two sections.

B.5.2 The Variance of HT-ATOTE

$$\begin{aligned}\mathbf{Var}(\widehat{\delta^*}_{HT,tot}) &= \mathbf{Var}(\widehat{\delta^*}_{HT,dir} + \widehat{\delta^*}_{HT,ind}) \\ &= \mathbf{Var}(C_1[\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{1})] + C_2[\bar{Y}^{obs}_{HT}(1,\mathbf{0}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})] \\ &+ C_1[\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(1,\mathbf{0})] + C_2[\bar{Y}^{obs}_{HT}(0,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})]) \\ &= \mathbf{Var}(2C_1\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - 2C_2\bar{Y}^{obs}_{HT}(0,\mathbf{0})) \\ &= 4C_1^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) + 4C_2^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2(4C_1C_2)\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0}))\end{aligned} \tag{B.54}$$

and for $C_1 = C_2 = \frac{1}{2}$,

$$\begin{aligned}\mathbf{Var}(\widehat{\delta^*}_{HT,tot}) &= \mathbf{Var}(\widehat{\delta}_{HT,tot}) \\ &= \mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) + \mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})) \quad \text{(B.55)}\end{aligned}$$

We can also find the variance as follows:

$$
\begin{aligned}
\mathbf{Var}(\widehat{\delta^*}_{HT,tot}) &= \mathbf{Var}(\widehat{\delta^*}_{HT,dir} + \widehat{\delta^*}_{HT,ind}) \\
&= \mathbf{Var}(C_1[\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}) - \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})] + C_2[\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}) - \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})] \\
&\quad + C_1[\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}) - \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})] + C_2[\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}) - \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})]) \\
&= C_1^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}) + C_1^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}) + C_2^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}) + C_2^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0}) \\
&\quad + C_1^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}) + C_1^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}) + C_2^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}) + C_2^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0}) \\
&\quad - 2C_1^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})) + 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) \\
&\quad - 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) + 2C_1^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1})) \\
&\quad - 2C_1^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) + 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})) \\
&\quad - 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) \\
&\quad - 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) \\
&\quad + 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) - 2C_1^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1})) \\
&\quad + 2C_1^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) - 2C_1C_2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})) \\
&\quad + 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) \\
&\quad - 2C_2^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) + 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1})) \\
&\quad - 2C_1C_2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) + 2C_2^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})) \\
&\quad - 2C_2^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) \\
&\quad - 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1})) \\
&\quad + 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) - 2C_2^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})) \\
&\quad + 2C_2^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) \\
&\quad - 2C_1^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) + 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})) \\
&\quad - 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) \\
&\quad - 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})) \\
&\quad + 2C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) - 2C_2^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0}))
\end{aligned}
\tag{B.56}
$$

This can be simplified as follows

$$
\begin{aligned}
\mathbf{Var}(\widehat{\delta^*}_{HT,tot}) &= \mathbf{Var}(\widehat{\delta^*}_{HT,dir} + \widehat{\delta^*}_{HT,ind}) \\
&= \mathbf{Var}(C_1[\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}) - \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})] + C_2[\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}) - \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})] \\
&\quad + C_1[\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}) - \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})] + C_2[\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}) - \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})]) \\
&= 4C_1^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}) + C_1^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}) + C_2^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}) + 4C_2^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0}) \\
&\quad + C_1^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}) + C_2^2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}) \\
&\quad - 4C_1^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})) + 4C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) \\
&\quad - 8C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) - 2C_1^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) \\
&\quad + 4C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})) - 4C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) \\
&\quad + 4C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) + 2C_1^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) \\
&\quad - 2C_1C_2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})) - 4C_2^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) \\
&\quad - 2C_1C_2\mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) + 2C_2^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})) \\
&\quad + 4C_1C_2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) - 4C_2^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{1})) \\
&\quad - 2C_1^2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{0})) \qquad \text{(B.57)}
\end{aligned}
$$

and for $C_1 = C_2 = \frac{1}{2}$,

$$
\begin{aligned}
\mathbf{Var}(\widehat{\delta^*}_{HT,tot}) &= \mathbf{Var}(\widehat{\delta}_{HT,tot}) \\
&= \mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1})) + \mathbf{Var}(\bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) - 2\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(1,\mathbf{1}), \bar{Y}^{\boldsymbol{obs}}_{HT}(0,\mathbf{0})) \qquad \text{(B.58)}
\end{aligned}
$$

B.6 Properties of HT-ADEE under the No-Interaction between Direct and Indirect Effects Assumption

Next, we compute the expected value and variance of Horvitz–Thompson estimator $\widehat{\delta^*}_{HT,dir}$ under the no-interaction between direct and indirect effects assumption.

$$\widehat{\delta^*}_{HT,dir} = C_1[\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{1})] + C_2[\bar{Y}^{obs}_{HT}(1,\mathbf{0}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})] \tag{B.59}$$

B.6.1 The Expected Value of HT-ADEE

Note that under the no-interaction between direct and indirect effects assumption and for $C_1 + C_2 = 1$:

$$\begin{aligned}\mathbf{E}(\widehat{\delta^*}_{HT,dir}) &= \mathbf{E}(C_1[\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{1})] + C_2[\bar{Y}^{obs}_{HT}(1,\mathbf{0}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})]) \\ &= \mathbf{E}(C_1[\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{1})]) + \mathbf{E}(C_2[\bar{Y}^{obs}_{HT}(1,\mathbf{0}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})]) \\ &= C_1[\bar{y}(1,\mathbf{1}) - \bar{y}(0,\mathbf{1})] + C_2[\bar{y}(1,\mathbf{0}) - \bar{y}(0,\mathbf{0})] \\ &= C_1[\bar{y}(1,\mathbf{1}) - \bar{y}(0,\mathbf{1})] + C_2[\bar{y}(1,\mathbf{1}) - \bar{y}(0,\mathbf{1})] = \delta_{dir}\end{aligned} \tag{B.60}$$

Hence, $\widehat{\delta^*}_{HT,dir}$ is an unbiased estimator for δ_{dir}.

B.6.2 The Variance of HT-ADEE

The variance can simply be computed using the property that $\mathbf{Var}(\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{a_i X_i}) = \sum_{i=1}^{N} a_i \mathbf{Var}(\mathbf{X_i}) +$ $2\sum_{j\neq i} a_i a_j \mathbf{Cov}(X_i X_j)$ as follows.

$$\mathbf{Var}(\widehat{\delta}^*{}_{HT,dir}) = \mathbf{Var}(C_1[\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{1})] + C_2[\bar{Y}^{obs}_{HT}(1,\mathbf{0}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})])$$

$$= C_1^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) + C_1^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{1})) + C_2^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{0}) + C_2^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{0}))$$

$$- 2C_1^2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{1})) + 2C_1C_2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(1,\mathbf{0}))$$

$$- 2C_1C_2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2C_1C_2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}), \bar{Y}^{obs}_{HT}(1,\mathbf{0}))$$

$$+ 2C_1C_2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2C_2^2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{0}), \bar{Y}^{obs}_{HT}(0,\mathbf{0}))$$

$$= \frac{C_1^2}{N^2}\sum_{i=1}^{N}\pi_i(1,\mathbf{1})[1-\pi_i(1,\mathbf{1})]\left[\frac{y_i(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right]^2$$

$$+ \frac{C_1^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(1,\mathbf{1}) - \pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})]\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})}.$$

$$+ \frac{C_1^2}{N^2}\sum_{i=1}^{N}\pi_i(0,\mathbf{1})[1-\pi_i(0,\mathbf{1})]\left[\frac{y_i(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right]^2$$

$$+ \frac{C_1^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{1}) - \pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})]\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})}$$

$$+ \frac{C_2^2}{N^2}\sum_{i=1}^{N}\pi_i(1,\mathbf{0})[1-\pi_i(1,\mathbf{0})]\left[\frac{y_i(1,\mathbf{0})}{\pi_i(1,\mathbf{0})}\right]^2$$

$$+ \frac{C_2^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(1,\mathbf{0}) - \pi_i(1,\mathbf{0})\pi_j(1,\mathbf{0})]\frac{y_i(1,\mathbf{0})y_j(1,\mathbf{0})}{\pi_i(1,\mathbf{0})\pi_j(1,\mathbf{0})}.$$

$$+ \frac{C_2^2}{N^2}\sum_{i=1}^{N}\pi_i(0,\mathbf{0})[1-\pi_i(0,\mathbf{0})]\left[\frac{y_i(0,\mathbf{0})}{\pi_i(0,\mathbf{0})}\right]^2$$

$$+ \frac{C_2^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{0}) - \pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})]\frac{y_i(0,\mathbf{0})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})}$$

$$- 2\left(\frac{C_1^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{1}),(0,\mathbf{1})) - \pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})\right]\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})}\right.$$

$$\left. - \frac{C_1^2}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{1})y_i(0,\mathbf{1})\right)$$

$$+ 2\left(\frac{C_1C_2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{1}),(1,\mathbf{0})) - \pi_i(1,\mathbf{1})\pi_j(1,\mathbf{0})\right]\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{0})}\right.$$

$$\left. - \frac{C_1C_2}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{1})y_i(1,\mathbf{0})\right). \quad \text{(B.61)}$$

$$
\begin{aligned}
&-2\left(\frac{C_1C_2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{1}),(0,\mathbf{0}))-\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})\right]\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})}\right.\\
&\qquad\left.-\frac{C_1C_2}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{1})y_i(0,\mathbf{0})\right)\\
&-2\left(\frac{C_1C_2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((0,\mathbf{1}),(1,\mathbf{0}))-\pi_i(0,\mathbf{1})\pi_j(1,\mathbf{0})\right]\frac{y_i(0,\mathbf{1})y_j(1,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_j(1,\mathbf{0})}\right.\\
&\qquad\left.-\frac{C_1C_2}{N^2}\sum_{i=1}^{N}y_i(0,\mathbf{1})y_i(1,\mathbf{0})\right)\\
&+2\left(\frac{C_1C_2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((0,\mathbf{1}),(0,\mathbf{0}))-\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})\right]\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})}\right.\\
&\qquad\left.-\frac{C_1C_2}{N^2}\sum_{i=1}^{N}y_i(0,\mathbf{1})y_i(0,\mathbf{0})\right)\\
&-2\left(\frac{C_2^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{0}),(0,\mathbf{0}))-\pi_i(1,\mathbf{0})\pi_j(0,\mathbf{0})\right]\frac{y_i(1,\mathbf{0})y_j(0,\mathbf{0})}{\pi_i(1,\mathbf{0})\pi_j(0,\mathbf{0})}\right.\\
&\qquad\left.-\frac{C_2^2}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{0})y_i(0,\mathbf{0})\right) \qquad \text{(B.62)}
\end{aligned}
$$

B.7 Properties of HT-AIEE under the No-Interaction between Direct and Indirect Effects Assumption

Now, we compute the expected value and variance of Horvitz–Thompson estimator $\widehat{\delta^*}_{HT,ind}$ under the no-interaction between direct and indirect effects assumption.

$$
\widehat{\delta^*}_{HT,ind}=C_1[\bar{Y}_{HT}^{obs}(1,\mathbf{1})-\bar{Y}_{HT}^{obs}(1,\mathbf{0})]+C_2[\bar{Y}_{HT}^{obs}(0,\mathbf{1})-\bar{Y}_{HT}^{obs}(0,\mathbf{0})] \qquad \text{(B.63)}
$$

B.7.1 The Expected Value of HT-AIEE

Note that under the no-interaction between direct and indirect effects assumption and for $C_1 + C_2 = 1$:

$$\begin{aligned}\mathbf{E}(\widehat{\delta^*}_{HT,ind}) &= \mathbf{E}(C_1[\bar{Y}_{HT}^{obs}(1,\mathbf{1}) - \bar{Y}_{HT}^{obs}(1,\mathbf{0})] + C_2[\bar{Y}_{HT}^{obs}(0,\mathbf{1}) - \bar{Y}_{HT}^{obs}(0,\mathbf{0})]) \\ &= \mathbf{E}(C_1[\bar{Y}_{HT}^{obs}(1,\mathbf{1}) - \bar{Y}_{HT}^{obs}(1,\mathbf{0})]) + \mathbf{E}(C_2[\bar{Y}_{HT}^{obs}(0,\mathbf{1}) - \bar{Y}_{HT}^{obs}(0,\mathbf{0})]) \\ &= C_1[\bar{y}(1,\mathbf{1}) - \bar{y}(1,\mathbf{0})] + C_2[\bar{y}(0,\mathbf{1}) - \bar{y}(0,\mathbf{0})] \\ &= C_1[\bar{y}(0,\mathbf{1}) - \bar{y}(0,\mathbf{0})] + C_2[\bar{y}(0,\mathbf{1}) - \bar{y}(0,\mathbf{0})] = \delta_{ind} \quad \text{(B.64)}\end{aligned}$$

Hence, $\widehat{\delta^*}_{HT,ind}$ is an unbiased estimator for δ_{ind}.

B.7.2 The Variance of HT-AIEE

The variance can simply be computed using the property that $\mathbf{Var}(\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{a_i X_i}) = \sum_{i=1}^{N} a_i \mathbf{Var}(\mathbf{X_i}) + 2\sum_{j\neq i} a_i a_j \mathbf{Cov}(X_i X_j)$ as follows.

$$
\begin{aligned}
\mathbf{Var}(\widehat{\delta}^*{}_{HT,ind}) &= \mathbf{Var}(C_1[\bar{Y}^{obs}_{HT}(1,\mathbf{1}) - \bar{Y}^{obs}_{HT}(1,\mathbf{0})] + C_2[\bar{Y}^{obs}_{HT}(0,\mathbf{1}) - \bar{Y}^{obs}_{HT}(0,\mathbf{0})]) \\
&= C_1^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{1})) + C_1^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{0})) + C_2^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{1})) + C_2^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{0})) \\
&\quad - 2C_1^2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(1,\mathbf{0})) + 2C_1C_2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{1})) \\
&\quad - 2C_1C_2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2C_1C_2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{0}), \bar{Y}^{obs}_{HT}(0,\mathbf{1})) \\
&\quad + 2C_1C_2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{0}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})) - 2C_2^2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(0,\mathbf{1}), \bar{Y}^{obs}_{HT}(0,\mathbf{0})) \\
&= \frac{C_1^2}{N^2}\sum_{i=1}^{N}\pi_i(1,\mathbf{1})[1-\pi_i(1,\mathbf{1})]\left[\frac{y_i(1,\mathbf{1})}{\pi_i(1,\mathbf{1})}\right]^2 \\
&\quad + \frac{C_1^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(1,\mathbf{1}) - \pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})]\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{1})}. \\
&\quad + \frac{C_1^2}{N^2}\sum_{i=1}^{N}\pi_i(1,\mathbf{0})[1-\pi_i(1,\mathbf{0})]\left[\frac{y_i(1,\mathbf{0})}{\pi_i(1,\mathbf{0})}\right]^2 \\
&\quad + \frac{C_1^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(1,\mathbf{0}) - \pi_i(1,\mathbf{0})\pi_j(1,\mathbf{0})]\frac{y_i(1,\mathbf{0})y_j(1,\mathbf{0})}{\pi_i(1,\mathbf{0})\pi_j(1,\mathbf{0})} \\
&\quad + \frac{C_2^2}{N^2}\sum_{i=1}^{N}\pi_i(0,\mathbf{1})[1-\pi_i((0,\mathbf{1})]\left[\frac{y_i(0,\mathbf{1})}{\pi_i(0,\mathbf{1})}\right]^2 \\
&\quad + \frac{C_2^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{1}) - \pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})]\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{1})} \\
&\quad + \frac{C_2^2}{N^2}\sum_{i=1}^{N}\pi_i(0,\mathbf{0})[1-\pi_i(0,\mathbf{0})]\left[\frac{y_i(0,\mathbf{0})}{\pi_i(0,\mathbf{0})}\right]^2 \\
&\quad + \frac{C_2^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{0}) - \pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})]\frac{y_i(0,\mathbf{0})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{0})\pi_j(0,\mathbf{0})} \\
&\quad - 2\left(\frac{C_1^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{1}),(1,\mathbf{0})) - \pi_i(1,\mathbf{1})\pi_j(1,\mathbf{0})\right]\frac{y_i(1,\mathbf{1})y_j(1,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_j(1,\mathbf{0})}\right. \\
&\qquad \left. - \frac{C_1^2}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{1})y_i(1,\mathbf{0})\right) \\
&\quad + 2\left(\frac{C_1C_2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{1}),(0,\mathbf{1})) - \pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})\right]\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{1})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{1})}\right. \\
&\qquad \left. - \frac{C_1C_2}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{1})y_i(0,\mathbf{1})\right). \qquad \text{(B.65)}
\end{aligned}
$$

$$
\begin{aligned}
&-2\left(\frac{C_1C_2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{1}),(0,\mathbf{0}))-\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})\right]\frac{y_i(1,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(1,\mathbf{1})\pi_j(0,\mathbf{0})}\right.\\
&\qquad\qquad\left.-\frac{C_1C_2}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{1})y_i(0,\mathbf{0})\right)\\
&-2\left(\frac{C_1C_2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{0}),(0,\mathbf{1}))-\pi_i(1,\mathbf{0})\pi_j(0,\mathbf{1})\right]\frac{y_i(1,\mathbf{0})y_j(0,\mathbf{1})}{\pi_i(1,\mathbf{0})\pi_j(0,\mathbf{1})}\right.\\
&\qquad\qquad\left.-\frac{C_1C_2}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{0})y_i(0,\mathbf{1})\right)\\
&+2\left(\frac{C_1C_2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{0}),(0,\mathbf{0}))-\pi_i(1,\mathbf{0})\pi_j(0,\mathbf{0})\right]\frac{y_i(1,\mathbf{0})y_j(0,\mathbf{0})}{\pi_i(1,\mathbf{0})\pi_j(0,\mathbf{0})}\right.\\
&\qquad\qquad\left.-\frac{C_1C_2}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{0})y_i(0,\mathbf{0})\right)\\
&-2\left(\frac{C_2^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((0,\mathbf{1}),(0,\mathbf{0}))-\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})\right]\frac{y_i(0,\mathbf{1})y_j(0,\mathbf{0})}{\pi_i(0,\mathbf{1})\pi_j(0,\mathbf{0})}\right.\\
&\qquad\qquad\qquad\qquad\left.-\frac{C_2^2}{N^2}\sum_{i=1}^{N}y_i(0,\mathbf{1})y_i(0,\mathbf{0})\right)
\end{aligned}
\tag{B.66}
$$

B.8 Properties of HT-AℓNNIEE under the No-Interaction between Direct and Indirect Effects Assumption

Next, we compute the expected value and variance of Horvitz–Thompson estimator $\widehat{\delta^*}_{HT,\ell}$ under the no-interaction between direct and indirect effects assumption.

$$
\widehat{\delta^*}_{HT,\ell} = C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})] + C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell}) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})] \tag{B.67}
$$

Lemma B.6. *For $C_{\ell 1} + C_{\ell 2} = 1$, $C_{\ell i} = C_{\ell' i}$ for $\ell, \ell' = 1, 2, \ldots, K$ and i = 1, 2.*

$$\widehat{\delta^*}_{HT,ind} = \sum_{\ell=1}^{K} \widehat{\delta^*}_{HT,\ell} \tag{B.68}$$

Proof. Under the no-interaction between direct and indirect effects assumption, and for $C_{\ell 1} + C_{\ell 2} = 1$, $C_{\ell i} = C_{\ell' i}$ with $\ell, \ell' = 1, 2, \ldots, K$ and i = 1, 2, we have:

$$\begin{aligned}
\sum_{\ell=1}^{K} \widehat{\delta^*}_{HT,\ell} &= C_{11}[\bar{Y}^{obs}_{HT}(1, \mathbf{W}^*_1) - \bar{Y}^{obs}_{HT}(1, \mathbf{W}^*_0)] + C_{12}[\bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_1) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_0)] \\
&+ C_{21}[\bar{Y}^{obs}_{HT}(1, \mathbf{W}^*_2) - \bar{Y}^{obs}_{HT}(1, \mathbf{W}^*_1)] + C_{22}[\bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_2) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_1)] \\
&+ C_{31}[\bar{Y}^{obs}_{HT}(1, \mathbf{W}^*_3) - \bar{Y}^{obs}_{HT}(1, \mathbf{W}^*_2)] + C_{32}[\bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_3) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_2)] \\
&+ \ldots \\
&+ C_{\ell 1}[\bar{Y}^{obs}_{HT}(1, \mathbf{W}^*_{\ell}) - \bar{Y}^{obs}_{HT}(1, \mathbf{W}^*_{\ell-1})] + C_{\ell 2}[\bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_{\ell}) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_{\ell-1})] \\
&+ \ldots \\
&+ C_{K1}[\bar{Y}^{obs}_{HT}(1, \mathbf{W}^*_K) - \bar{Y}^{obs}_{HT}(1, \mathbf{W}^*_{K-1})] + C_{K2}[\bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_K) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_{K-1})] \\
&= C_{K1}\bar{Y}^{obs}_{HT}(1, \mathbf{W}^*_K) - C_{11}\bar{Y}^{obs}_{HT}(1, \mathbf{W}^*_0) + C_{K2}\bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_K) - C_{12}\bar{Y}^{obs}_{HT}(0, \mathbf{W}^*_0) \\
&= C_{K1}\bar{Y}^{obs}_{HT}(1, \mathbf{1}) - C_{11}\bar{Y}^{obs}_{HT}(1, \mathbf{0}) + C_{K2}\bar{Y}^{obs}_{HT}(0, \mathbf{1}) - C_{12}\bar{Y}^{obs}_{HT}(0, \mathbf{0}) = \widehat{\delta^*}_{HT,ind} \qquad \text{(B.69)}
\end{aligned}$$

■

B.8.1 The Expected Value of HT-AℓNNIEE

$$\begin{aligned}
\mathbf{E}(\widehat{\delta^*}_{HT,\ell}) &= \mathbf{E}(C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})] + C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})]) \\
&= \mathbf{E}(C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})]) + \mathbf{E}(C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})]) \\
&= C_{\ell 1}[\bar{y}(1,\mathbf{W}^*_\ell) - \bar{y}(1,\mathbf{W}^*_{\ell-1})] + C_{\ell 2}[\bar{y}(0,\mathbf{W}^*_\ell) - \bar{y}(0,\mathbf{W}^*_{\ell-1})] \\
&= C_{\ell 1}[\bar{y}(0,\mathbf{W}^*_\ell) - \bar{y}(0,\mathbf{W}^*_{\ell-1})] + C_{\ell 2}[\bar{y}(0,\mathbf{W}^*_\ell) - \bar{y}(0,\mathbf{W}^*_{\ell-1})] = \delta_\ell \quad \text{(B.70)}
\end{aligned}$$

Hence, $\widehat{\delta^*}_{HT,\ell}$ is an unbiased estimator for δ_ℓ.

B.8.2 The Variance of HT-AℓNNIEE

The variance can be computed where we first find $\mathbf{Var}(C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})])$, $\mathbf{Var}(C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})])$ and $\mathbf{Cov}(C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})], C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})])$ as follows.

First we find:

$$
\begin{aligned}
\mathbf{Var}(C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})]) &= C_1^2 \big(\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell})) + \mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})) \\
&\quad -2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}), \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})\big) \\
&= \frac{C_1^2}{N^2}\sum_{i=1}^{N} \pi_i(1,\mathbf{W}^*_{\ell})[1-\pi_i(1,\mathbf{W}^*_{\ell})]\left[\frac{y_i(1,\mathbf{W}^*_{\ell})}{\pi_i(1,\mathbf{W}^*_{\ell})}\right]^2 \\
&+ \frac{C_1^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(1,\mathbf{W}^*_{\ell}) - \pi_i(1,\mathbf{W}^*_{\ell})\pi_j(1,\mathbf{W}^*_{\ell})]\frac{y_i(1,\mathbf{W}^*_{\ell})y_j(1,\mathbf{W}^*_{\ell})}{\pi_i(1,\mathbf{W}^*_{\ell})\pi_j(1,\mathbf{W}^*_{\ell})} \\
&+ \frac{C_1^2}{N^2}\sum_{i=1}^{N} \pi_i(1,\mathbf{W}^*_{\ell-1})[1-\pi_i(1,\mathbf{W}^*_{\ell-1})]\left[\frac{y_i(1,\mathbf{W}^*_{\ell-1})}{\pi_i(1,\mathbf{W}^*_{\ell-1})}\right]^2 \\
&+ \frac{C_1^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(1,\mathbf{W}^*_{\ell-1}) - \pi_i(1,\mathbf{W}^*_{\ell-1})\pi_j(1,\mathbf{W}^*_{\ell-1})]\frac{y_i(1,\mathbf{W}^*_{\ell-1})y_j(1,\mathbf{W}^*_{\ell-1})}{\pi_i(1,\mathbf{W}^*_{\ell-1})\pi_j(1,\mathbf{W}^*_{\ell-1})} \\
&- 2\left(\frac{C_1^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{W}^*_{\ell}),(1,\mathbf{W}^*_{\ell-1})) - \pi_i(1,\mathbf{W}^*_{\ell})\pi_j(1,\mathbf{W}^*_{\ell-1})\right]\frac{y_i(1,\mathbf{W}^*_{\ell})y_j(1,\mathbf{W}^*_{\ell-1})}{\pi_i(1,\mathbf{W}^*_{\ell})\pi_j(1,\mathbf{W}^*_{\ell-1})}\right. \\
&\left. -\frac{C_1^2}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{W}^*_{\ell})y_i(1,\mathbf{W}^*_{\ell-1})\right). \qquad \text{(B.71)}
\end{aligned}
$$

Next, we find $\mathbf{Cov}(C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})], C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell}) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})])$. Note that:

$$
\begin{aligned}
&\mathbf{E}\left[C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})]C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})]\right] \\
&= C_{\ell 1}C_{\ell 2}\mathbf{E}\left[\big(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell)\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)\big) - \big(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell)\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})\big) - \big(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)\big)\right. \\
&\qquad \left. + \big(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})\big)\right] \\
&= C_{\ell 1}C_{\ell 2}\left[\mathbf{E}\left[\big(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell)\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)\big)\right] - \mathbf{E}\left[\big(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell)\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})\big)\right]\right. \\
&\qquad \left. -\mathbf{E}\left[\big(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)\big)\right] + \mathbf{E}\left[\big(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})\big)\right]\right] \\
&= \frac{C_{\ell 1}C_{\ell 2}}{N^2}\left(\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{W}^*_\ell),(0,\mathbf{W}^*_\ell))\frac{y_i(1,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_\ell)}{\pi_i(1,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_\ell)}\right) \\
&\quad - \frac{C_{\ell 1}C_{\ell 2}}{N^2}\left(\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{W}^*_\ell),(0,\mathbf{W}^*_{\ell-1}))\frac{y_i(1,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(1,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_{\ell-1})}\right) \\
&\quad - \frac{C_{\ell 1}C_{\ell 2}}{N^2}\left(\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{W}^*_{\ell-1}),(0,\mathbf{W}^*_\ell))\frac{y_i(1,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_\ell)}{\pi_i(1,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_\ell)}\right) \\
&\quad + \frac{C_{\ell 1}C_{\ell 2}}{N^2}\left(\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{W}^*_{\ell-1}),(0,\mathbf{W}^*_{\ell-1}))\frac{y_i(1,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(1,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_{\ell-1})}\right) \qquad \text{(B.72)}
\end{aligned}
$$

$$\begin{aligned}
&\mathbf{E}(C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell)-\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})])\mathbf{E}(C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell)-\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})])\\
&\qquad = C_{\ell 1}\left[\bar{y}(1,\mathbf{W}^*_\ell)-\bar{y}(1,\mathbf{W}^*_{\ell-1})\right]C_{\ell 2}\left[\bar{y}(0,\mathbf{W}^*_\ell)-\bar{y}(0,\mathbf{W}^*_{\ell-1})\right]\\
&= C_{\ell 1}C_{\ell 2}\left[\bar{y}(1,\mathbf{W}^*_\ell)\bar{y}(0,\mathbf{W}^*_\ell)-\bar{y}(1,\mathbf{W}^*_\ell)\bar{y}(0,\mathbf{W}^*_{\ell-1})-\bar{y}(1,\mathbf{W}^*_{\ell-1})\bar{y}(0,\mathbf{W}^*_\ell)+\bar{y}(1,\mathbf{W}^*_{\ell-1})\bar{y}(0,\mathbf{W}^*_{\ell-1})\right]\\
&= \left(\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^N y_i(1,\mathbf{W}^*_\ell)\sum_{i=1}^N y_i(0,\mathbf{W}^*_\ell)\right)-\left(\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^N y_i(1,\mathbf{W}^*_\ell)\frac{1}{N}\sum_{i=1}^N y_i(0,\mathbf{W}^*_{\ell-1})\right)\\
&-\left(\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^N y_i(1,\mathbf{W}^*_{\ell-1})\frac{1}{N}\sum_{i=1}^N y_i(0,\mathbf{W}^*_\ell)\right)+\left(\frac{1}{N}\sum_{i=1}^N y_i(1,\mathbf{W}^*_{\ell-1})\frac{1}{N}\sum_{i=1}^N y_i(0,\mathbf{W}^*_{\ell-1})\right)\\
&\qquad = \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^N y_i(1,\mathbf{W}^*_\ell)y_i(0,\mathbf{W}^*_\ell)+\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^N\sum_{j\neq i} y_i(1,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_\ell)\\
&\qquad -\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^N y_i(1,\mathbf{W}^*_\ell)y_i(0,\mathbf{W}^*_{\ell-1})-\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^N\sum_{j\neq i} y_i(1,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_{\ell-1})\\
&\qquad -\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^N y_i(1,\mathbf{W}^*_{\ell-1})y_i(0,\mathbf{W}^*_\ell)-\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^N\sum_{j\neq i} y_i(1,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_\ell)\\
&\qquad +\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^N y_i(1,\mathbf{W}^*_{\ell-1})y_i(0,\mathbf{W}^*_{\ell-1})+\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^N\sum_{j\neq i} y_i(1,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_{\ell-1})
\end{aligned}\tag{B.73}$$

$$\begin{aligned}
&\mathbf{Cov}(C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})], C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})]) = \\
&\quad \mathbf{E}(C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})] C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})]) \\
&\quad - \mathbf{E}(C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})])\mathbf{E}(C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})]) \\
&\quad = \frac{C_{\ell 1}C_{\ell 2}}{N^2}\left(\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{W}^*_\ell),(0,\mathbf{W}^*_\ell))\frac{y_i(1,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_\ell)}{\pi_i(1,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_\ell)}\right) \\
&\quad - \frac{C_{\ell 1}C_{\ell 2}}{N^2}\left(\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{W}^*_\ell),(0,\mathbf{W}^*_{\ell-1}))\frac{y_i(1,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(1,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_{\ell-1})}\right) \\
&\quad - \frac{C_{\ell 1}C_{\ell 2}}{N^2}\left(\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{W}^*_{\ell-1}),(0,\mathbf{W}^*_\ell))\frac{y_i(1,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_\ell)}{\pi_i(1,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_\ell)}\right) \\
&\quad + \frac{C_{\ell 1}C_{\ell 2}}{N^2}\left(\sum_{i=1}^{N}\sum_{j\neq i}\pi_{ij}((1,\mathbf{W}^*_{\ell-1}),(0,\mathbf{W}^*_{\ell-1}))\frac{y_i(1,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(1,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_{\ell-1})}\right) \\
&\quad - \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{W}^*_\ell)y_i(0,\mathbf{W}^*_\ell) - \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(1,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_\ell) \\
&\quad + \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{W}^*_\ell)y_i(0,\mathbf{W}^*_{\ell-1}) + \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(1,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_{\ell-1}) \\
&\quad + \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{W}^*_{\ell-1})y_i(0,\mathbf{W}^*_\ell) + \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(1,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_\ell) \\
&\quad - \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{W}^*_{\ell-1})y_i(0,\mathbf{W}^*_{\ell-1}) - \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}\sum_{j\neq i} y_i(1,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_{\ell-1})
\end{aligned} \tag{B.74}$$

We can simplify the covariance further as follows:

$$\begin{aligned}
&\mathbf{Cov}(C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})], C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_\ell) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})]) \\
&= \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}((1,\mathbf{W}^*_\ell),(0,\mathbf{W}^*_\ell)) - \pi_i(1,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_\ell)]\frac{y_i(1,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_\ell)}{\pi_i(1,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_\ell)} \\
&\qquad - \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{W}^*_\ell)y_i(0,\mathbf{W}^*_\ell) \\
&- \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}((1,\mathbf{W}^*_\ell),(0,\mathbf{W}^*_{\ell-1})) - \pi_i(1,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_{\ell-1})]\frac{y_i(1,\mathbf{W}^*_\ell)y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(1,\mathbf{W}^*_\ell)\pi_j(0,\mathbf{W}^*_{\ell-1})} \\
&\qquad + \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{W}^*_\ell)y_i(0,\mathbf{W}^*_{\ell-1}) \\
&- \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}((1,\mathbf{W}^*_{\ell-1}),(0,\mathbf{W}^*_\ell)) - \pi_i(1,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_\ell)]\frac{y_i(1,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_\ell)}{\pi_i(1,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_\ell)} \\
&\qquad + \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{W}^*_{\ell-1})y_i(0,\mathbf{W}^*_\ell) \\
&= \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}((1,\mathbf{W}^*_{\ell-1}),(0,\mathbf{W}^*_{\ell-1})) - \pi_i(1,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_{\ell-1})]\frac{y_i(1,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(1,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_{\ell-1})} \\
&\qquad - \frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N} y_i(1,\mathbf{W}^*_{\ell-1})y_i(0,\mathbf{W}^*_{\ell-1}) \qquad \text{(B.75)}
\end{aligned}$$

Then, the variance can be computed using the property $\mathbf{Var}(X-Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$ -2$\mathbf{Cov}(X,Y)$.

Additionally, the variance can simply be computed using the property that $\mathbf{Var}(\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{a_i X_i}) = \sum_{i=1}^{N} a_i \mathbf{Var}(\mathbf{X_i}) + 2\sum_{j\neq i} a_i a_j \mathbf{Cov}(X_i X_j)$ as follows.

$$
\begin{aligned}
\mathbf{Var}(\widehat{\delta^*}_{HT,\ell}) &= \mathbf{Var}(C_{\ell 1}[\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}) - \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})] + C_{\ell 2}[\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell}) - \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})]) \\
&= \mathbf{Var}(C_{\ell 1}\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}) - C_{\ell 1}\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1}) + C_{\ell 2}\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell}) - C_{\ell 2}\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) \\
&= C_{\ell 1}^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell})) + C_{\ell 1}^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})) + C_{\ell 2}^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell})) + C_{\ell 2}^2\mathbf{Var}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) \\
&\quad - 2C_{\ell 1}^2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}), \bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1})) + 2C_{\ell 1}C_{\ell 2}\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}), \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell})) \\
&\quad - 2C_{\ell 1}C_{\ell 2}\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell}), \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) - 2C_{\ell 1}C_{\ell 2}\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1}), \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell})) \\
&\quad + 2C_{\ell 1}C_{\ell 2}\mathbf{Cov}(\bar{Y}^{obs}_{HT}(1,\mathbf{W}^*_{\ell-1}), \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) - 2C_{\ell 2}^2\mathbf{Cov}(\bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell}), \bar{Y}^{obs}_{HT}(0,\mathbf{W}^*_{\ell-1})) \\
&= \frac{C_{\ell 1}^2}{N^2}\sum_{i=1}^{N}\pi_i(1,\mathbf{W}^*_{\ell})[1-\pi_i(1,\mathbf{W}^*_{\ell})]\left[\frac{y_i(1,\mathbf{W}^*_{\ell})}{\pi_i(1,\mathbf{W}^*_{\ell})}\right]^2 \\
&\quad + \frac{C_{\ell 1}^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(1,\mathbf{W}^*_{\ell}) - \pi_i(1,\mathbf{W}^*_{\ell})\pi_j(1,\mathbf{W}^*_{\ell})]\frac{y_i(1,\mathbf{W}^*_{\ell})y_j(1,\mathbf{W}^*_{\ell})}{\pi_i(1,\mathbf{W}^*_{\ell})\pi_j(1,\mathbf{W}^*_{\ell})}. \\
&\quad + \frac{C_{\ell 1}^2}{N^2}\sum_{i=1}^{N}\pi_i(1,\mathbf{W}^*_{\ell-1})[1-\pi_i(1,\mathbf{W}^*_{\ell-1})]\left[\frac{y_i(1,\mathbf{W}^*_{\ell-1})}{\pi_i(1,\mathbf{W}^*_{\ell-1})}\right]^2 \\
&\quad + \frac{C_{\ell 1}^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(1,\mathbf{W}^*_{\ell-1}) - \pi_i(1,\mathbf{W}^*_{\ell-1})\pi_j(1,\mathbf{W}^*_{\ell-1})]\frac{y_i(1,\mathbf{W}^*_{\ell-1})y_j(1,\mathbf{W}^*_{\ell-1})}{\pi_i(1,\mathbf{W}^*_{\ell-1})\pi_j(1,\mathbf{W}^*_{\ell-1})} \\
&\quad + \frac{C_{\ell 2}^2}{N^2}\sum_{i=1}^{N}\pi_i(0,\mathbf{W}^*_{\ell})[1-\pi_i(0,\mathbf{W}^*_{\ell})]\left[\frac{y_i(0,\mathbf{W}^*_{\ell})}{\pi_i(0,\mathbf{W}^*_{\ell})}\right]^2 \\
&\quad + \frac{C_{\ell 2}^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{W}^*_{\ell}) - \pi_i(0,\mathbf{W}^*_{\ell})\pi_j(0,\mathbf{W}^*_{\ell})]\frac{y_i(0,\mathbf{W}^*_{\ell})y_j(0,\mathbf{W}^*_{\ell})}{\pi_i(0,\mathbf{W}^*_{\ell})\pi_j(0,\mathbf{W}^*_{\ell})}. \\
&\quad + \frac{C_{\ell 2}^2}{N^2}\sum_{i=1}^{N}\pi_i(0,\mathbf{W}^*_{\ell-1})[1-\pi_i(0,\mathbf{W}^*_{\ell-1})]\left[\frac{y_i(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_{\ell-1})}\right]^2 \\
&\quad + \frac{C_{\ell 2}^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}[\pi_{ij}(0,\mathbf{W}^*_{\ell-1}) - \pi_i(0,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_{\ell-1})]\frac{y_i(0,\mathbf{W}^*_{\ell-1})y_j(0,\mathbf{W}^*_{\ell-1})}{\pi_i(0,\mathbf{W}^*_{\ell-1})\pi_j(0,\mathbf{W}^*_{\ell-1})} \\
&\quad - 2\left(\frac{C_{\ell 1}^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{W}^*_{\ell}),(1,\mathbf{W}^*_{\ell-1})) - \pi_i(1,\mathbf{W}^*_{\ell})\pi_j(1,\mathbf{W}^*_{\ell-1})\right]\frac{y_i(1,\mathbf{W}^*_{\ell})y_j(1,\mathbf{W}^*_{\ell-1})}{\pi_i(1,\mathbf{W}^*_{\ell})\pi_j(1,\mathbf{W}^*_{\ell-1})}\right. \\
&\qquad \left. -\frac{C_{\ell 1}^2}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{W}^*_{\ell})y_i(1,\mathbf{W}^*_{\ell-1})\right) \\
&\quad + 2\left(\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{W}^*_{\ell}),(0,\mathbf{W}^*_{\ell})) - \pi_i(1,\mathbf{W}^*_{\ell})\pi_j(0,\mathbf{W}^*_{\ell})\right]\frac{y_i(1,\mathbf{W}^*_{\ell})y_j(0,\mathbf{W}^*_{\ell})}{\pi_i(1,\mathbf{W}^*_{\ell})\pi_j(0,\mathbf{W}^*_{\ell})}\right. \\
&\qquad \left. -\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{W}^*_{\ell})y_i(0,\mathbf{W}^*_{\ell})\right). \qquad \text{(B.76)}
\end{aligned}
$$

$$
\begin{aligned}
&-2\left(\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{W}_{\ell}^*),(0,\mathbf{W}_{\ell-1}^*))-\pi_i(1,\mathbf{W}_{\ell}^*)\pi_j(0,\mathbf{W}_{\ell-1}^*)\right]\frac{y_i(1,\mathbf{W}_{\ell}^*)y_j(0,\mathbf{W}_{\ell-1}^*)}{\pi_i(1,\mathbf{W}_{\ell}^*)\pi_j(0,\mathbf{W}_{\ell-1}^*)}\right.\\
&\qquad\left.-\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{W}_{\ell}^*)y_i(0,\mathbf{W}_{\ell-1}^*)\right)\\
&-2\left(\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{W}_{\ell-1}^*),(0,\mathbf{W}_{\ell}^*))-\pi_i(1,\mathbf{W}_{\ell-1}^*)\pi_j(0,\mathbf{W}_{\ell}^*)\right]\frac{y_i(1,\mathbf{W}_{\ell-1}^*)y_j(0,\mathbf{W}_{\ell}^*)}{\pi_i(1,\mathbf{W}_{\ell-1}^*)\pi_j(0,\mathbf{W}_{\ell}^*)}\right.\\
&\qquad\left.-\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{W}_{\ell-1}^*)y_i(0,\mathbf{W}_{\ell}^*)\right)\\
&+2\left(\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((1,\mathbf{W}_{\ell-1}^*),(0,\mathbf{W}_{\ell-1}^*))-\pi_i(1,\mathbf{W}_{\ell-1}^*)\pi_j(0,\mathbf{W}_{\ell-1}^*)\right]\frac{y_i(1,\mathbf{W}_{\ell-1}^*)y_j(0,\mathbf{W}_{\ell-1}^*)}{\pi_i(1,\mathbf{W}_{\ell-1}^*)\pi_j(0,\mathbf{W}_{\ell-1}^*)}\right.\\
&\qquad\left.-\frac{C_{\ell 1}C_{\ell 2}}{N^2}\sum_{i=1}^{N}y_i(1,\mathbf{W}_{\ell-1}^*)y_i(0,\mathbf{W}_{\ell-1}^*)\right)\\
&-2\left(\frac{C_{\ell 2}^2}{N^2}\sum_{i=1}^{N}\sum_{j\neq i}\left[\pi_{ij}((0,\mathbf{W}_{\ell}^*),(0,\mathbf{W}_{\ell-1}^*))-\pi_i(0,\mathbf{W}_{\ell}^*)\pi_j(0,\mathbf{W}_{\ell-1}^*)\right]\frac{y_i(0,\mathbf{W}_{\ell}^*)y_j(0,\mathbf{W}_{\ell-1}^*)}{\pi_i(0,\mathbf{W}_{\ell}^*)\pi_j(0,\mathbf{W}_{\ell-1}^*)}\right.\\
&\qquad\left.-\frac{C_{\ell 2}^2}{N^2}\sum_{i=1}^{N}y_i(0,\mathbf{W}_{\ell}^*)y_i(0,\mathbf{W}_{\ell-1}^*)\right) \qquad \text{(B.77)}
\end{aligned}
$$

B.8.3 Weights in HT-AℓNNIEE under Assumption 2

Using Lagrange multiplier, we can find the values of $C_{\ell 1}$ and $C_{\ell 2}$ in B.67 that gives the minimum variance of HT-AℓNNIEE under Assumption 3.2 as follows.

First, let $S_{11}^2 = \mathbf{Var}(\bar{Y}_{HT}^{obs}(1,\mathbf{W}_{\ell}^*))$, $S_{12}^2 = \mathbf{Var}(\bar{Y}_{HT}^{obs}(1,\mathbf{W}_{\ell-1}^*))$, $S_{21}^2 = \mathbf{Var}(\bar{Y}_{HT}^{obs}(0,\mathbf{W}_{\ell}^*))$ and $S_{22}^2 = \mathbf{Var}(\bar{Y}_{HT}^{obs}(0,\mathbf{W}_{\ell-1}^*))$ and assume that the covariance components in B.76 are equal to zero.

Then, we want to minimize the variance

$$\mathbf{Var}(\widehat{\delta^*}_{HT,\ell}) = \sum_{i=1}^{2} C_{\ell i}^2(S_{i1}^2 + S_{i2}^2) \qquad \text{(B.78)}$$

subject to the constraint $\sum_{i=1}^{2} C_{\ell i} = 1$

Using the method of Lagrange multiplier, we minimize the function:

$$\phi \;=\; \mathbf{Var}(\widehat{\delta^*}_{HT,\ell}) \;-\; \lambda(\sum_{i=1}^{2} C_{\ell i}^2 \;-\; 1) \;=\; \sum_{i=1}^{2} C_{\ell i}^2(S_{i1}^2 \;+\; S_{i2}^2) \;-\; \lambda(\sum_{i=1}^{2} C_{\ell i}^2 \;-\; 1) \quad \text{(B.79)}$$

Then, $\frac{\partial \phi}{\partial C_{\ell i}} = 0$ gives

$$\frac{\partial[\sum_{i=1}^{2} C_{\ell i}^2(S_{i1}^2 + S_{i2}^2) - \lambda(\sum_{i=1}^{2} C_{\ell i}^2 - 1)}{\partial C_{\ell i}}] = 0$$

$$2C_{\ell i}(S_{i1}^2 + S_{i2}^2) - \lambda = 0$$

$$C_{\ell i} = \frac{\lambda}{2(S_{i1}^2 + S_{i2}^2)} \quad \text{(B.80)}$$

Now, we find Lagrange multiplier λ as follows.

$$\sum_{i=1}^{2} C_{\ell i} = 1 \Rightarrow \frac{\lambda}{2(S_{11}^2 + S_{12}^2)} + \frac{\lambda}{2(S_{21}^2 + S_{22}^2)} = 1$$

$$\Rightarrow \frac{\lambda}{2}\left(\frac{S_{11}^2 + S_{12}^2 + S_{21}^2 + S_{22}^2}{(S_{11}^2 + S_{12}^2)(S_{21}^2 + S_{22}^2)}\right) = 1 \Rightarrow \lambda = \frac{2(S_{11}^2 + S_{12}^2)(S_{21}^2 + S_{22}^2)}{S_{11}^2 + S_{12}^2 + S_{21}^2 + S_{22}^2}. \quad \text{(B.81)}$$

Substituting in B.80, we have

$$C_{\ell 1} = \frac{S_{21}^2 + S_{22}^2}{S_{11}^2 + S_{12}^2 + S_{21}^2 + S_{22}^2}, C_{\ell 2} = \frac{S_{11}^2 + S_{12}^2}{S_{11}^2 + S_{12}^2 + S_{21}^2 + S_{22}^2}, \quad \text{(B.82)}$$

which gives the minimum variance of HT-AℓNNIEE under Assumption 3.2.

Appendix C

Properties of Horvitz-Thompson Estimators in Chapter 4

Here, we present proofs and properties of estimators provided under the no-interaction between the indirect effect assumption.

C.1 Properties of HT-ADEE under the No-Interaction between the Indirect Effects Assumption

The unbiased Horvitz–Thompson estimator of ADE under the no-interaction between the indirect effects assumption is as follows:

$$\widehat{\delta^{**}}_{HT,dir} = \sum_{e=1}^{2^K} C_e[\bar{Y}^{obs}_{HT}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}})] \tag{C.1}$$

where $\mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}$ is the e^{th} treatment assignment vector of the K-nearest neighbors with $e = 1, 2, \ldots, 2^K$

C.1.1 The Expected Value of HT-ADEE

Lemma C.1.

$$\mathbf{E}(\widehat{\delta^{**}}_{HT,dir}) = \delta_{dir} \tag{C.2}$$

Proof. Note that under the no-interaction between the indirect effects assumption and for $\sum_{e=1}^{2^K} C_e = 1$,

$$\begin{aligned}
\mathbf{E}[\widehat{\delta^{**}}_{HT,dir}] = \mathbf{E}&\left[\sum_{e=1}^{2^K} C_e[\bar{Y}_{HT}^{obs}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}) - \bar{Y}_{HT}^{obs}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}})]\right] \\
&= \sum_{e=1}^{2^K} C_e[\mathbf{E}[\bar{Y}_{HT}^{obs}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}})] - \mathbf{E}[\bar{Y}_{HT}^{obs}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}})]] \\
&= \sum_{e=1}^{2^K} C_e[\bar{y}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}) - \bar{y}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}})] \\
&= \sum_{e=1}^{2^K} C_{\ell e}[\bar{y}(1, \mathbf{1}) - \bar{y}(0, \mathbf{1})] = \delta_{dir}
\end{aligned} \tag{C.3}$$

■

Hence, $\widehat{\delta^{**}}_{HT,dir}$ is an unbiased estimator for δ_{dir} and the variance can be computed using the property that $\mathbf{Var}(\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{a_i X_i}) = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j \mathbf{Cov}(X_i X_j)$ as follows.

C.1.2 The Variance of HT-ADEE

$$\begin{aligned}
\mathbf{Var}(\widehat{\delta^{**}}_{HT,dir}) = &\sum_{e,e',W_i=1} C_e C_{e'} \mathbf{Cov}(\bar{Y}_{HT}^{obs}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}_{HT}^{obs}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\
&+ \sum_{e,e',W_i=0} C_e C_{e'} \mathbf{Cov}(\bar{Y}_{HT}^{obs}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}_{HT}^{obs}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\
&- 2 \sum_{e,e',W_i=1,W_i=0} C_e C_{e'} \mathbf{Cov}(\bar{Y}_{HT}^{obs}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}_{HT}^{obs}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})).
\end{aligned} \tag{C.4}$$

C.2 Properties of HT-AℓNNIEE under the No-Interaction between the Indirect Effects Assumption

In this section we derive properties of HT-AℓNNIEE under the no-interaction between the indirect effects assumption. Here, we use $(W_i, W_\ell = 1, \mathbf{W_{e,K-1}})$ instead of $(W_i, W_{\mathbb{N}_{ik}})$ in the previous sections where the unbiased Horvitz–Thompson estimator of the average potential outcomes of units under any exposure $(W_i, W_\ell, \mathbf{W_{e,K-1}})$ is

$$\bar{Y}^{obs}_{HT}(W_i, W_\ell, \mathbf{W_{e,K-1}}) = \frac{1}{N}\sum_{i=1}^{N} I_i(W_i, W_\ell, \mathbf{W_{e,K-1}})\frac{Y_i^{obs}}{\pi_i(W_i, W_\ell, \mathbf{W_{e,K-1}})}. \tag{C.5}$$

The unbiased Horvitz–Thompson estimator of AℓNNIE under the no-interaction between the indirect effects assumption as follows:

$$\widehat{\delta^{**}}_{HT,\ell} = \sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})]. \tag{C.6}$$

C.2.1 The Expected Value of HT-AℓNNIEE

Lemma C.2.

$$\mathbf{E}(\widehat{\delta^{**}}_{HT,\ell}) = \delta_\ell. \tag{C.7}$$

Proof. Because all differences in Equation C.6 are equal in expectation, the proof follows by noting that one of the differences is $\widehat{\delta}_{\ell th}$ and by Equation B.42 as follows,

$$
\begin{aligned}
\mathbf{E}[\widehat{\delta^{**}}_{HT,\ell}] &= \mathbf{E}\left[\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})\right] \\
&= \sum_{e=1}^{2^K} C_{\ell e}\mathbf{E}\left[\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})\right] \\
&= \sum_{e=1}^{2^K} C_{\ell e}[\bar{y}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{y}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})] \\
&= \sum_{e=1}^{2^K} C_{\ell e}[\bar{y}(0, \mathbf{W}_\ell^*) - \bar{y}(0, \mathbf{W}_{\ell-\mathbf{1}}^*)] = \delta_\ell. \quad \text{(C.8)}
\end{aligned}
$$

■

Hence, $\widehat{\delta^{**}}_{HT,\ell}$ is an unbiased estimator for δ_ℓ and the variance can be computed using the property that $\mathbf{Var}(\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{a_i X_i}) = \sum_{i=1}^N \sum_{j=1}^N a_i a_j \mathbf{Cov}(X_i X_j)$ as follows.

C.2.2 The Variance of HT-AℓNNIEE

For $\widehat{\delta^{**}}_{HT,\ell} = \sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})]$ where $\sum_{e=1}^{2^K} C_{\ell e} = 1$, the variance is

$$
\begin{aligned}
\mathbf{Var}(\widehat{\delta^{**}}_{HT,\ell}) &= \mathbf{Var}\left(\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})]\right) \\
&= \sum_{e,e',W_\ell=1} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e',K-1}})) \\
&+ \sum_{e,e',W_\ell=0} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}}), \bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e',K-1}})) \\
&- 2\sum_{e,e',W_\ell=1,W_\ell=0} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e',K-1}})), \quad \text{(C.9)}
\end{aligned}
$$

C.3 Properties of HT-AIEE under the No-Interaction between the Indirect Effects Assumption

The unbiased Horvitz–Thompson estimator of AIE under the no-interaction between the indirect effects assumption is as follows:

Definition C.1.

$$\widehat{\delta^{**}}_{HT,ind} = \sum_{\ell=1}^{K} \widehat{\delta^{**}}_{HT,\ell} = \sum_{\ell=1}^{K} \sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})] \tag{C.10}$$

C.3.1 The Expected Value of HT-AIEE

For HT-AIE estimator under Assumption 4.1 with $\sum_{e=1}^{2^K} C_e = 1$, the expected value is provided in the following lemma.

Lemma C.3.

$$\mathbf{E}(\widehat{\delta^{**}}_{HT,ind}) = \delta_{ind}. \tag{C.11}$$

Proof. Under the no-interaction between the indirect effects assumption with $\sum_{e=1}^{2^K} C_e = 1$,

$$\mathbf{E}[\widehat{\delta^{**}}_{HT,ind}] = \mathbf{E}\left[\sum_{\ell=1}^{K}\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})]\right]$$

$$= \sum_{\ell=1}^{K}\sum_{e=1}^{2^K} C_{\ell e}(\mathbf{E}[\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})])$$

$$= \sum_{\ell=1}^{K}\sum_{e=1}^{2^K} C_{\ell e}[\bar{y}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{y}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})]$$

$$= \sum_{\ell=1}^{K}\sum_{e=1}^{2^K} C_{\ell e}[\bar{y}(0, \mathbf{W}^*_\ell) - \bar{y}(0, \mathbf{W}^*_{\ell-\mathbf{1}})]$$

$$= \sum_{\ell=1}^{K}\sum_{e=1}^{2^K} C_{\ell e}(\delta_\ell) = \delta_{ind} \quad \text{(C.12)}$$

■

Hence, $\widehat{\delta^{**}}_{HT,ind}$ is an unbiased estimator for δ_{ind} and the variance can be computed using the property that $\mathbf{Var}(\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{a_i X_i}) = \sum_{i=1}^{N} a_i \mathbf{Var}(\mathbf{X_i}) + 2\sum_{j \neq i} a_i a_j \mathbf{Cov}(X_i X_j)$ as follows.

C.3.2 The Variance of HT-AIEE

$$
\begin{aligned}
\mathbf{Var}(\widehat{\delta^{**}}_{HT,ind}) &= \mathbf{Var}\left(\sum_{e=1}^{2^K}\sum_{\ell,\ell'=1}^{K} C_{\ell e}[\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}_{HT}^{obs}(W_i, W_{\ell'} = 0, \mathbf{W_{e,K-1}})]\right) \\
&= \sum_{e=1}^{2^K}\sum_{\ell,\ell'=1}^{K} C_{\ell e}^2 \mathbf{Var}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}})) \\
&+ 2\sum_{e<e',W_\ell=1}^{2^K}\sum_{\ell,\ell=1}^{K} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e',K-1}})) \\
&+ \sum_{e=1}^{2^K}\sum_{\ell,\ell'=1}^{K} C_{\ell e}^2 \mathbf{Var}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})) \\
&+ 2\sum_{e<e',W_\ell=0}^{2^K}\sum_{\ell,\ell'=1}^{K} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}}), \bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e',K-1}})) \\
&- 2\sum_{e,e',W_\ell=1,W_\ell=0}\sum_{\ell,\ell'=1}^{K} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e',K-1}}))
\end{aligned}
\tag{C.13}
$$

Or we can simply rewrite $\mathbf{Var}(\widehat{\delta^{**}}_{HT,ind})$ as follows:

$$
\begin{aligned}
\mathbf{Var}(\widehat{\delta^{**}}_{HT,ind}) &= \sum_{e,e',W_\ell=W_{\ell'}=1}\sum_{\ell,\ell'=1}^{K} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}_{HT}^{obs}(W_i, W_{\ell'} = 1, \mathbf{W_{e',K-1}})) \\
&+ \sum_{e,e',W_\ell=W_{\ell'}=0}\sum_{\ell,\ell'=1}^{K} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}}), \bar{Y}_{HT}^{obs}(W_i, W_{\ell'} = 0, \mathbf{W_{e',K-1}})) \\
&- 2\sum_{e,e',W_\ell=1,W_{\ell'}=0}\sum_{\ell,\ell'=1}^{K} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}_{HT}^{obs}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}_{HT}^{obs}(W_i, W_{\ell'} = 0, \mathbf{W_{e',K-1}}))
\end{aligned}
\tag{C.14}
$$

The covariance between any two indirect effect estimators is as follows:

$$\begin{aligned}
&\mathbf{Cov}(\widehat{\delta^{**}}_{HT,\ell_{th}}, \widehat{\delta^{**}}_{HT,\ell'_{th}}) = \\
&\qquad \sum_{e,e',W_\ell=W_{\ell'}=1} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 1, \mathbf{W_{e',K-1}})) \\
&\qquad + \sum_{e,e',W_\ell=W_{\ell'}=0} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e',K-1}})) \\
&\qquad - 2 \sum_{e,e',W_\ell=1,W_{\ell'}=0} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e',K-1}})),
\end{aligned} \tag{C.15}$$

where $\mathbf{Cov}(\widehat{\delta^{**}}_{HT,\ell_{th}}, \widehat{\delta^{**}}_{HT,\ell'_{th}}) = \mathbf{E}(\widehat{\delta^{**}}_{HT,\ell_{th}}\widehat{\delta^{**}}_{HT,\ell'_{th}}) - \mathbf{E}(\widehat{\delta^{**}}_{HT,\ell_{th}})\mathbf{E}(\widehat{\delta^{**}}_{HT,\ell'_{th}})$ as follows,

$$\begin{aligned}
&\mathbf{E}(\widehat{\delta^{**}}_{HT,\ell_{th}}\widehat{\delta^{**}}_{HT,\ell'_{th}}) = \\
&\qquad\quad \mathbf{E}(\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})] \\
&\qquad\quad \times \sum_{e=1}^{2^K} C_{\ell' e}[\bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 1, \mathbf{W_{e,K-1}}) - \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e,K-1}})]) \\
&\qquad = \mathbf{E}(\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) \sum_{e=1}^{2^K} C_{\ell' e}[\bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 1, \mathbf{W_{e,K-1}})) \\
&\qquad - \mathbf{E}(\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) \sum_{e=1}^{2^K} C_{\ell' e}[\bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e,K-1}})) \\
&\qquad - \mathbf{E}(\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}}) \sum_{e=1}^{2^K} C_{\ell' e}[\bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 1, \mathbf{W_{e,K-1}})) \\
&\qquad + \mathbf{E}(\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}}) \sum_{e=1}^{2^K} C_{\ell' e}[\bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e,K-1}}))
\end{aligned} \tag{C.16}$$

and

$$
\begin{aligned}
&\mathbf{E}(\widehat{\delta^{**}}_{HT,\ell_{th}})\mathbf{E}(\widehat{\delta^{**}}_{HT,\ell'_{th}})) = \\
&\qquad \mathbf{E}(\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}) - \bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}})]) \\
&\qquad \times \mathbf{E}(\sum_{e=1}^{2^K} C_{\ell' e}[\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_{\ell'} = 1, \mathbf{W_{e,K-1}}) - \bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e,K-1}})]) \\
&\quad = \mathbf{E}(\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}))\mathbf{E}(\sum_{e=1}^{2^K} C_{\ell' e}[\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_{\ell'} = 1, \mathbf{W_{e,K-1}})) \\
&\quad - \mathbf{E}(\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}))\mathbf{E}(\sum_{e=1}^{2^K} C_{\ell' e}[\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e,K-1}})) \\
&\quad - \mathbf{E}(\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}}))\mathbf{E}(\sum_{e=1}^{2^K} C_{\ell' e}[\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_{\ell'} = 1, \mathbf{W_{e,K-1}})) \\
&\quad + \mathbf{E}(\sum_{e=1}^{2^K} C_{\ell e}[\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}}))\mathbf{E}(\sum_{e=1}^{2^K} C_{\ell' e}[\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e,K-1}})).
\end{aligned}
\tag{C.17}
$$

Hence,

$$
\begin{aligned}
&\mathbf{Cov}(\widehat{\delta^{**}}_{HT,\ell_{th}}, \widehat{\delta^{**}}_{HT,\ell'_{th}}) = \mathbf{E}(\widehat{\delta^{**}}_{HT,\ell_{th}}\widehat{\delta^{**}}_{HT,\ell'_{th}}) - \mathbf{E}(\widehat{\delta^{**}}_{HT,\ell_{th}})\mathbf{E}(\widehat{\delta^{**}}_{HT,\ell'_{th}}) \\
&\quad = \sum_{e,e',W_\ell = W_{\ell'} = 1} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_{\ell'} = 1, \mathbf{W_{e',K-1}})) \\
&\quad + \sum_{e,e',W_\ell = W_{\ell'} = 0} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_\ell = 0, \mathbf{W_{e,K-1}}), \bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e',K-1}})) \\
&\quad - 2\sum_{e,e',W_\ell = 1, W_{\ell'} = 0} C_{\ell e}C_{\ell e'}\mathbf{Cov}(\bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_\ell = 1, \mathbf{W_{e,K-1}}), \bar{Y}^{\boldsymbol{obs}}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W_{e',K-1}})).
\end{aligned}
\tag{C.18}
$$

C.4 Properties of HT-ATOTE under the No-Interaction between the Indirect Effects Assumption

Next, we compute the expected value and variance of the unbiased Horvitz–Thompson estimator $\widehat{\delta^{**}}_{HT,tot}$ under the no-interaction between the indirect effects assumption as provided in the following lemma,

Definition C.2.

$$\widehat{\delta^{**}}_{HT,tot} = \widehat{\delta^{**}}_{HT,dir} + \widehat{\delta^{**}}_{HT,ind}. \tag{C.19}$$

C.4.1 The Expected Value of HT-ATOTE

Lemma C.4.

$$\mathbf{E}(\widehat{\delta^{**}}_{HT,tot}) = \delta_{dir} + \delta_{ind} = \delta_{Total}. \tag{C.20}$$

The proof follows by lemmas C.1 and C.3.

C.4.2 The Variance of HT-ATOTE

$$
\begin{aligned}
\mathbf{Var}(\widehat{\delta^{**}}_{HT,tot}) = \mathbf{Var}(\widehat{\delta^{**}}_{HT,dir} + \widehat{\delta^{**}}_{HT,ind}) \\
= \mathbf{Var}\Bigg[\sum_{e=1}^{2^K} C_e[\bar{Y}^{obs}_{HT}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}) - \bar{Y}^{obs}_{HT}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}})] \\
+ \sum_{e=1}^{2^K}\sum_{\ell=1}^{K} C_{\ell e}[\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W}_{\mathbf{e},\mathbf{K-1}}) - \bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W}_{\mathbf{e},\mathbf{K-1}})]\Bigg] \\
= \sum_{e,e'} C_e C_{e'} \mathbf{Cov}(\bar{Y}^{obs}_{HT}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}^{obs}_{HT}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\
+ \sum_{e,e'} C_e C_{e'} \mathbf{Cov}(\bar{Y}^{obs}_{HT}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}^{obs}_{HT}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\
- 2 \sum_{e,e',W_i=1,W_i=0} C_e C_{e'} \mathbf{Cov}(\bar{Y}^{obs}_{HT}(1, \mathbf{W}_{\mathcal{N}_{\mathbf{ike}}}), \bar{Y}^{obs}_{HT}(0, \mathbf{W}_{\mathcal{N}_{\mathbf{ike'}}})) \\
+ \sum_{e,e',W_\ell=W_{\ell'}=1} \sum_{\ell,\ell'=1}^{K} C_{\ell e} C_{\ell e'} \mathbf{Cov}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W}_{\mathbf{e},\mathbf{K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 1, \mathbf{W}_{\mathbf{e'},\mathbf{K-1}})) \\
+ \sum_{e,e',W_\ell=W_{\ell'}=0} \sum_{\ell,\ell'=1}^{K} C_{\ell e} C_{\ell e'} \mathbf{Cov}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 0, \mathbf{W}_{\mathbf{e},\mathbf{K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W}_{\mathbf{e'},\mathbf{K-1}})) \\
- 2 \sum_{e,e',W_\ell=1,W_{\ell'}=0} \sum_{\ell,\ell'=1}^{K} C_{\ell e} C_{\ell e'} \mathbf{Cov}(\bar{Y}^{obs}_{HT}(W_i, W_\ell = 1, \mathbf{W}_{\mathbf{e},\mathbf{K-1}}), \bar{Y}^{obs}_{HT}(W_i, W_{\ell'} = 0, \mathbf{W}_{\mathbf{e'},\mathbf{K-1}})) \\
+ 2I(C\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), C'\bar{Y}^{obs}_{HT}(W', \mathbf{W'}_{N_K})) \sum_{e,e'} CC'\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W'}_{N_K})) \\
-2I(-C\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), C'\bar{Y}^{obs}_{HT}(W', \mathbf{W'}_{N_K})) \sum_{e,e'} CC'\mathbf{Cov}(\bar{Y}^{obs}_{HT}(W, \mathbf{W}_{N_K}), \bar{Y}^{obs}_{HT}(W', \mathbf{W'}_{N_K})).
\end{aligned}
\tag{C.21}
$$