Identification and Decomposition of Causal Effect Accounting for Network Change

Siwon Ryu

March 12, 2024

Abstract

This study analyzes an identification and estimation procedure for the causal treatment effect, explicitly considering network changes resulting from treatment. Compared to the classical approach in program evaluation literature that assumes independent units, recent empirical studies emphasize the significance of spillover effects. However, these studies often assume that the underlying network is fixed or unaffected by treatment. At the same time, there has been some empirical evidence that treatment can also affect the network. This study analyzes the identification of the causal effect of treatment with interference between units and also accounts for possible network changes. The main contribution of this study is the decomposition of the causal effect into two distinct parts: the effect of treatment when the network remains unchanged and the effect when only the network structure is changed by the treatment. This approach enhances our understanding of the mechanisms of a policy or program by explicitly considering counterfactual scenarios in which the network is changed or unchanged due to treatment. Additionally, the study addresses quasi-experimental situations, providing a Difference-in-Differences approach.

1 Introduction

Evaluating a program is one of main topics in empirical economics, typically involving the estimation of causal effects of a program. Many such methods are based on potential outcome framework (e.g., Rubin (1974)), with the Stable Unit Treatment Value Assumption (SUTVA), which exclude the inteference between units, or individuals. This framework is useful since it provides causal interpretation of parameters. However, considering that economic agents usually interact with one another, the SUTVA would be restrictive in many cases. Recent empirical evidence also highlight that the potential significance of spillover effects in program evaluations.

If the outcome of each unit is causally influenced by the treatment status of others, potential outcomes can be expressed as functions of the entire treatment vector. In such situations, researchers often face two primary challenges. The first relates to choose which treatment effects to focus on. In SUTVA situations, treatment status is expressed by a binary variable, making it straightforward to define treatment effects. However, in more general cases where treatments are represented by binary vectors, multiple scenarios can arise, such as the effect of one's treatment, a neighbor's treatment, or all others' treatments, among others. The second challenge is about the dimensionality, as the number of potential outcomes increases with the number of units. To address these challenges, studies in this field commonly employ the constant treatment response assumption proposed by Manski (2013), or utilize the exposure mapping approach to significantly reduce dimensionality.

In this literature, it has been commonly assumed that the underlying network structure is not causally affected by the treatment, or fixed. This assumption holds in situations where it is difficult to make significant changes in the network in the short term. Additionally, as exposure maps are typically functions of the underlying network, it is reasonable to fix the network structure and analyze the causal effect of different exposure levels. Nevertheless, recent empirical studies provide evidence that programs can indeed cause changes in the network structure (e.g., Comola and Prina (2021), Banerjee et al. (2023, forthcoming), Dupas, Keats, and Robinson (2019)). In particular, Comola and Prina (2021) argue estimating treatment effect without considering potential network changes due to the program could result in bias.

This study introduces a novel approach to estimate the causal effect of a program, considering the possibility of changing the network resulting from the treatment. Further, this study decomposes the treatment (or spillover) effect into two parts. The first part examines the impact of the treatment when the network remains unchanged, while the second part focuses on the effect of the treatment when it only changes the network structure. Therefore, if the treatment doesn't cause changes in the network, the second part has no effect, and the first part will align with existing methods that estimate treatment and spillover effects.

Furthermore, many studies in the literature often assume randomized experiments, where treatments are randomly assigned, to provide exogenous variation for identifying causal parameters. This approach is valid when using data from randomized field experiments. However, in some cases, treatment selection issues may exist. This is especially true when the variation in question arises from natural experiments, as it may not be entirely random, in terms of the correlation between the characteristics in the baseline period. This study addresses such quasi-experimental situations and, specifically, investigates to identify and estimate the causal effect of a program and its decompositions using the Difference-in-Differences approach, provided that we have the data over two or more periods, including pre and post-treament periods.

The main contribution of this study can be summarized in two aspects. First, this is the first attempt to analyze causal effects accounting for potential network changes with the potential outcome framework. The proposed decomposition allows for an in-depth examination of a program's mechanism. Additionally, if the network remains unchanged after get treatmented, the estimation reduces to the existing methods, and the decomposition will be trivial. In this regard, this study generalize the existing method to accommodate possible causal network changes due to treatment. Second, the methods introduced in this study are applicable to both randomized experiments and quasi-experimental datasets, particularly by using the Difference-in-Differences approach.

This paper is organized as follows. Section 2 describes the settings used in this study, defines the parameters of interest, and addresses the issue of their identification. In particular, the decomposition of causal effects is defined in Section 2.2. In Section

3, the estimation procedure is presented. Section 4 verifies the finite and asymptotic performance through Monte Carlo simulations. Section **??** is empirical illustration of the proposed method through an empirical example (in progress), and Section 5 concludes this study.

1.1 Related Literature

This study is closely related to the literature on the identification and estimation of causal effects of a program when units interact with each others, hence SUTVA is violated. A number of studies have tackled the challenge of estimating causal effects in scenarios involving network interference. In randomized experimental settings, Leung (2020) provide estimators and their asymptotic properties when treatment is randomly assigned, and researchers have exact knowledge of the exposure map. However, the accuracy of the specified exposure map relies on assumptions regarding the extent to which each unit relies on their neighbors. Leung (2022) introduce the concept of approximate neighborhood interference, where interference diminishes as network distance increases. This property is employed to correct potential exposure map misspecifications. As these studies have emphasized, while exposure mapping is useful for analyzing potential outcomes without assuming SUTVA by significantly reducing dimensionality, one must be careful for specification of the exposure map. Also, while the exposure mapping is useful to deal with dimensionality problem of potential outcome, It is not applicable for the potential network structure. Suppose potential network is determined by a exposure map. However, because the exposure map is usually determined by the underlying network structure, it is not straightforward to define appropriate exposure map for potential network, or potential links. To avoid this problem, I assumes dyadic network structure instead, without using exposure mapping approach.

Vazquez-Bare (2022b) propose interpretations of projection coefficients in the traditional difference-in-means and linear-in-means model in the presence of network interference. The approach is similar in that I initially interpret a projection coefficient of the outcome on treatment, the sum of others' treatment, and their interactions. However, rather than assuming the existence of an exposure mapping, I use a linear response function with dyadic potential network structure to account for network changes

driven by treatment. Additionally, to decompose the effect into pure treatment effects and network-changing effects, I introduce another layer that relates these projection parameters to the causal parameters.

Treatment could be endogenous in various directions, and one source of such endogeneity is the existence of imperfect compliance. That is, even if the treatment is randomly assigned by design, each unit could choose if they actually take it or not. Vazquez-Bare (2022a), DiTraglia et al. (2023), Hoshino and Yanagi (2021), Kormos, Lieli, and Huber (2023), Kang and Imbens (2016), Blackwell (2017) have studied such situations. Another endoegeneity arise from selection problem. For example, if a treatment is assigned based on the outcomes at the pre-treatment period, then that type of treatment is not entirely random. In this case, if two time periods before and after treatment are available, then one can use a quasi-experimental variation by assuming exogenous parallel trend to identify difference-in-differences estimand for average treatment effect on treated. Xu (2023 WP), Butts (2021) study Difference-in-differeces approach under network interference. Both studies assume fixed network, and assume they observe a correctly specified exposure map.

Some empirical studies provide evidence of how programs can impact underlying social networks and economic outcomes. Comola and Prina (2021) utilize data from a randomized experiment in Nepal and observe that providing savings accounts to female-headed households led to changes in network degrees. In particular, untreated households shifted from 0.8% to 1%, while households with at least one treated member shifted from 81% to 76%. Banerjee et al. (2023, forthcoming) examine data from Karnataka, India, and a field experiment in Hyderabad to investigate the effects of exposure to formal financial institutions on network density. They also develop a dynamic link formation model to explain their findings. Dupas, Keats, and Robinson (2019) use data from a field experiment in rural Kenya where households were provided with free savings accounts. Their results show that households became less reliant on distant family members and more supportive of neighbors and friends within their village. These studies emphasize the importance of considering network effects in program evaluations.

In this respect, Comola and Prina (2021) explicitly address how to estimate treat-

ment effect accounting for network changes. The authors estimate a two-period linear-in-means model (e.g., Manski (2013), Bramoullé, Djebbari, and Fortin (2009)) and define the direct and indirect treatment effects as the partial derivatives of conditional expectations with respect to the treatment vector. Their main findings suggest that the indirect effects could be underestimated if potential network changes resulting from the treatment are not considered. However, as their empirical findings, these effects can be interpreted as a combination of the treatment effect and the effect from network change due to the treatment. This leads to challenges in interpreting these estimated parameters causally. To address this issue, in this paper, I explicitly decompose each component of the effects and provide a clear causal interpretation for each of them.

When a program affect outcomes through an underlying interaction structure, the potential outcomes become a function of the entire treatment vector without any imposed restrictions. Consequently, the central challenge becomes the identification of causal effects in the presence of multiple treatment options. In quasi-experimental settings, Frölich (2004) provides a comprehensive overview of program evaluation methodologies when dealing with multiple treatment options. Furthermore, Fricke (2017) specializes in formalizing the identifying conditions for difference-in-differences settings. These contributions in the literature offer valuable insights into addressing the complexities of estimating causal effects in scenarios involving multiple treatment options and interaction structures.

2 Model and Identification

In this section, I provide an overview of the setting and define the parameters of interest. I begin by discussing response functions for potential links and potential outcomes. Then, I establish key causal parameters, specifically the direct and indirect effects. The direct effect focuses on how an individual's treatment status impacts their outcome, while the indirect effects consider the effects of changes in others' treatment statuses. Subsequently, I propose a decomposition of these effects into two parts. The first part examines the impact when the underlying network remains unchanged from the pre-treatment period, referred to as the "treatment effect," as the conventional treatment effect in the literature denotes this particular effect. The second part of the decomposition is about the effect solely arising from changes in the network structure, referred to as the "network effect" in this study.

2.1 Response Functions for Links and Potential Outcomes

Suppose there are G groups with N individuals in each group. Let $t \in \{0, 1\}$ denote the time periods, where some individual are assigned to a treatment group after $t = 0$. In this context, let $t = 0$ represent the pre-treatment period, and $t = 1$ the post-treatment period. Also, let $D_{ig} \in \{0, 1\}$ be the indicator for an individual receiving treatment, and suppose there is no imperfect compliance. Each individual interacts with others through an underlying network structure. Specifically, let $A_{ijtg} \in \{0, 1\}$ represent the link between individuals i and j in group g at time t, and there is no self link following the convention in the literature, i.e., $A_{ijtg} = 1$ if two different units i and j are linked at time t in group g, and $A_{iitg} = 0$ for all i. The network can be both directed and undirected in this setting. For each individual i in group g, let $Y_{itg} \in \mathbb{R}$ be the outcome of interest at time t.

Without any restrictions, potential outcomes and potential links can be expressed as functions of the entire treatment vector in each group. Let $\boldsymbol{d} \in \{0, 1\}^N$ represent a vector of potential treatments. Corresponding to this treatment $\boldsymbol{d}$, let $A_{ijtg}(\boldsymbol{d})$ denote the *potential* link between individuals i and j, and $Y_{itg}(\boldsymbol{d})$ denote the *potential* outcome of individual i at time period t. Given the existence of 2^N possible potential variables,

defining the causal effect of interest becomes challenging, particularly as the number of units increases. To address this dimensionality issue, I first introduce a restriction on the potential network that each network link is formed from a dyadic model:

Assumption 1 (Dyadic Response on Potential Network Links)**.** *For each pair* (i, j) *in group* g *at time period* t*, for any* $\boldsymbol{d}, \boldsymbol{d}' \in \{0,1\}^N$*, if* $d_i = d_i'$ *and* $d_j = d_j'$*, then* $A_{ijtg}(\boldsymbol{d}) = A_{ijtg}(\boldsymbol{d}')$ *with probability one. Therefore, by abusing notation, denote the potential link as follows: for* $\boldsymbol{d} = (d_1, ..., d_N) \in \{0,1\}^N$*,*

$$A_{ijtg}(\boldsymbol{d}) = \begin{cases} A_{ijtg}(d_i, d_j) & i \neq j, \\ 0 & i = j. \end{cases}$$

Assumption 1 impose the restriction that each pair's potential link is determined by the pair's treatment status (d_i, d_j) only, but not affected by the other's treatment status. For example, Assumption 1 is satisfied for a dyadic link formation model (e.g., Graham (2017)):

$$A_{ij}(\boldsymbol{d}) = \mathbb{1}\{\psi_1 d_i + \psi_2 d_j + \delta_i + \delta_j + u_{ij} > 0\}. \tag{1}$$

The potential outcome can be influenced by the treatment vector through a mapping, in which the range has a lower or fixed dimension. This property is referred to as constant treatment response (e.g., Manski (2013)), and the mapping is called as an exposure map in the literature. The existence or the exact functional form of an exposure mapping is largely unknown without restrictions on the response function. However, in some cases, it is possible to specify an appropriate exposure map for a potential outcome. For example, if the network is anonymous or exchangeable, meaning that the name of each individual has no meaning, then only own treatment status, the number of treated or untreated neighbors could impact the potential outcome. Leung (2020) demonstrated that assuming (i) local spillover, indicating that interference occurs within a network distance of 1, and (ii) exchangeability is equivalent to having the exposure map of one's own treatment (d_i), the number of treated neighbors $(\sum_j A_{ij} D_j)$, and the number of untreated neighbors $(\sum_j A_{ij}(1 - D_j))$. In other

words, for any $\boldsymbol{d}, \boldsymbol{d}' \in \{0,1\}^2$, whenever $d_i = d_i'$, $\sum_j A_{ij}(d_i, d_j)d_j = \sum_j A_{ij}(d_i', d_j')d_j'$, and $\sum_j A_{ij}(d_i, d_j)(1-d_j) = \sum_j A_{ij}(d_i', d_j')(1-d_j')$, then we have $Y_i(\boldsymbol{d}) = Y_i(\boldsymbol{d}')$ with probability one. In order to focus on the identification of causal effects and their subsequent decomposition, I assume that the response functions for potential outcomes are linear in such exposures.[1]

Assumption 2 (Linear Response on Potential Outcomes)**.** *For each i in group g, let $Y_{itg}(\boldsymbol{d})$ be potential outcome corresponding to $\boldsymbol{d} \in \{0,1\}^N$. Assume potential outcome is determined by the following linear response function:*

$$Y_{i1g}(\boldsymbol{d}) = \alpha_1 + \beta d_i + \varepsilon_{i1g}(d_i) + \gamma_1 \sum_{j=1, j\neq i}^{N} A_{ij1g}(d_i, d_j)d_j + \gamma_2 \sum_{j=1, j\neq i}^{N} A_{ij1g}(d_i, d_j)(1-d_j),$$

$$Y_{i0g}(\boldsymbol{d}) = \alpha_0 + \varepsilon_{i0g}(d_i) + \gamma_2 \sum_{j=1, j\neq i}^{N} A_{ij0g}(d_i, d_j).$$

The parameter β is the difference in potential outcomes at $t = 1$ when all links are fixed, but only d_i changes from zero to one. Therefore, β captures the direct treatment effect at $t = 1$ when the network remains unchanged. Moving on, since $\sum_{j\neq i} A_{ijtg}d_j$ represents the number of potentially treated neighbors, γ_1 captures the spillover (or exposure) effect from marginal changes in the number of treated neighbors. Similarly, γ_2 captures those from the untreated neighbors. γ_2 also appears in the pre-treated period to impose the condition that all individuals are untreated at $t = 0$.[2]

Because changes in d_i could also affect potential links $\{A_{ij1g}(d_i, d_j) : j = 1, ..., N\}$, the own treatment could have additional effect on outcomes, which is the effect from changed links driven by the own treatment. I define this type of effect as the "network effect". Similarly, changes in the other's treatment d_j can have similar two types of effects, which will be formally defined below.

Assumption 2 implies that the observed outcome at $t = 1$ can be expressed as a

[1] In other words, Assumption 2 is equivalent to assuming (i) local spillover, (ii) exchangeability, and (iii) the additive separability of potential outcomes with respect to the exposures.

[2] Here, we implicitly assume that the effect captured by γ_2 remains unchanged in both periods. This assumption is made for the sake of simplicity in the identifying assumptions, as detailed in Section 2.3. However, it is possible to allow for γ_2 to vary with time, i.e., γ_{20} and γ_{21}, with minor additional assumptions (see Section 2.3).

linear-in-sums model:

$$Y_{i1g} = \alpha_1 + \beta D_i + \gamma_1 \sum_{j=1}^{N} A_{ij} D_j + \gamma_2 \sum_{j=1}^{N} A_{ij}(1 - D_j) + \varepsilon_{i1g}.$$

In this regard, this model generalizes the standard potential outcome framework into two dimensions. Firstly, if the network is unaffected by treatment or fixed, i.e., $A_{ij}(d_i, d_j) = A_{ij}$, then this response function is the same as used in the literature (e.g., Leung (2020), Forastiere, Airoldi, and Mealli (2021)). Hence, this setting can be applied to investigate the causal effect with a predetermined or fixed underlying network. Secondly, if $\gamma_1 = \gamma_2 = 0$, then the response function reduces to the standard potential outcome model under SUTVA.

2.2 Causal Parameters and Decomposition

In this subsection, I suppress the group and time index subscripts (t, g) for simplicity to define parameters of interest and their decompositions. First, consider a case with $N = 2$, that there are only two individuals in each group. By Assumption 2, the response function of the outcome can be written as a function of own treatment d_i, neighbor's treatment d_j, and their potential link $A_{ij}(d_i, d_j)$:

$$Y_i(\boldsymbol{d}) = h(d_i, d_j, A_{ij}(d_i, d_j)), \quad \boldsymbol{d} \in \{0, 1\}^2.$$

Then, the effect of own treatment (d_i) on i's outcome is expressed as

$$\begin{aligned} &h(1, 0, A_{ij}(1, 0)) - h(0, 0, A_{ij}(0, 0)) \\ &= \underbrace{h(1, 0, A_{ij}(1, 0)) - h(1, 0, A_{ij}(0, 0))}_{\text{=Direct Network Effect}} + \underbrace{h(1, 0, A_{ij}(0, 0)) - h(0, 0, A_{ij}(0, 0))}_{\text{=Direct Treatment Effect}}. \end{aligned}$$

Also, the effect of neighbor j's treatment (d_j) on i's outcome is expressed as

$$\begin{aligned} &h(0,1,A_{ij}(0,1)) - h(0,0,A_{ij}(0,0)) \\ &= \underbrace{h(0,0,A_{ij}(0,1)) - h(0,0,A_{ij}(0,0))}_{\text{=Indirect Network Effect}} + \underbrace{h(0,1,A_{ij}(0,1)) - h(0,0,A_{ij}(0,1))}_{\text{=Indirect Treatment Effect}}. \end{aligned}$$

The direct and indirect network effects represent the causal impact resulting from changes in links driven by treatment. The second terms capture the effects of treatment when links remain unchanged from the pre-treatment period.

In a general case with N individuals, each indirect effect can be influenced by the treatment status of all neighbors. However, the potential outcome is affected by the treatment of others solely through the count of treated or untreated neighborhoods[3]. Therefore, I focus on the marginal effect of a neighbor's treatment, specifically the impact of "one additional" treated neighbor. Let $\{\boldsymbol{e}_1, ..., \boldsymbol{e}_N\}$ be standard Euclidean basis, i.e., for each i, $e_{ii} = 1$, and $e_{ij} = 0$ for all $j \neq i$. Then, the average direct effect of the treated (π^D) on the outcome is defined as

$$\begin{aligned} \text{Average Direct Effect } (\pi^D) &:= E[Y_i(\boldsymbol{e}_i) - Y_i(\mathbf{0}) | \boldsymbol{D} = \boldsymbol{e}_i] \\ &= \underbrace{\beta}_{:=\pi^{DT}} + \underbrace{\gamma_2 \sum_{j=1}^{N} E[A_{ij}(1,0) - A_{ij}(0,0) | D_i = 1, D_j = 0]}_{:=\pi^{DN}}. \end{aligned}$$

For the decomposition, Assumption 2 provides the causal interpretation as

$$\begin{aligned} &\text{Average Direct Treatment Effect } (\pi^{DT}) = \beta, \\ &\text{Average Direct Network Effect } (\pi^{DN}) = \gamma_2 \sum_{j=1}^{N} E[A_{ij}(1,0) - A_{ij}(0,0) | D_i = 1, D_j = 0], \end{aligned}$$

In the sense that π^{DT} is the effect of own treatment when the network is fixed, and π^{DN} is the effect of changed links driven by own treatment. Similarly, the average

[3]This is due to the exposure mapping used to construct the response function being implied by the exchangeability of networks.

indirect effect of the treated (π^I) on the outcome is defined as for some j,

$$\begin{aligned}\text{Average Indirect Effect} &:= E[Y_i(\boldsymbol{e}_j) - Y_i(\mathbf{0})|\boldsymbol{D} = \boldsymbol{e}_j] \\ &= \underbrace{(\gamma_1 - \gamma_2)E[A_{ij}(0,1)|D_i = 0, D_j = 1]}_{:=\pi^{IT}} \\ &\quad + \underbrace{\gamma_2 E[A_{ij}(0,1) - A_{ij}(0,0)|D_i = 0, D_j = 1]}_{:=\pi^{IN}},\end{aligned}$$

where each part is interpreted as follows:

$$\begin{aligned}\text{Average Indirect Treatment Effect } (\pi^{IT}) &= (\gamma_1 - \gamma_2)E[A_{ij}(0,1)|D_i = 0, D_j = 1], \\ \text{Average Indirect Network Effect } (\pi^{IN}) &= \gamma_2 E[A_{ij}(0,1) - A_{ij}(0,0)|D_i = 0, D_j = 1].\end{aligned}$$

From the linear response function, the direct and indirect effects for any scenarios (i.e., any $\boldsymbol{d} \in \{0,1\}^N$) can be expressed as the sum of $(\pi^{DT}, \pi^{DN}, \pi^{IT}, \pi^{IN})$s. To identify these four effects separately, we need the values of $\beta, \gamma_1, \gamma_2$, and the conditional distribution of $A_{ij}(0,1) - A_{ij}(0,0)$, $A_{ij}(1,1) - A_{ij}(0,0)$, and $A_{ij}(0,0)$.

2.3 Identification

The identification strategy for decomposing treatment effects consists of two main components. First, to address potential network changes in the post-treatment period, I employ the the difference-in-differences approaches to identify the distribution of potential link formation and changes. Second, I demonstrate an estimating equation for outcomes from Assumption 2. In this estimating equation, the coefficients are functions of the parameters in the response function $(\beta, \gamma_1, \gamma_2)$, as well as the potential distribution of link formation and link changes. By integrating these two components, the decomposition of average effects, as defined in Section 2.2, is identified.

2.3.1 Identifying Assumptions

To begin with, I list the identifying assumptions for the main identification results.

Assumption 3 (Local Exogeneity). *Let* $\boldsymbol{D}_g = (D_{1g}, ..., D_{Ng})$ *be observed treatment vector of group* g*. For each time* t *and group* g*,*

1. $E[\varepsilon_{itg}(d)|\boldsymbol{D}_g] = E[\varepsilon_{itg}(d)|D_{ig}]$*, for all individual* i *and* $d \in \{0,1\}$*,*
2. $E[A_{ijtg}(d,e)|\boldsymbol{D}_g] = E[A_{ijtg}(d,e)|D_{ig}, D_{jg}]$*, for all pairs* (i,j) *and* $(d,e) \in \{0,1\}^2$*.*

Assumption 4 (No Anticipation). *For each group* g*,*

1. $E[\varepsilon_{i0g}(0)|D_{ig} = 1] = E[\varepsilon_{i0g}(1)|D_{ig} = 1]$*, for all individual* i*,*
2. $E[A_{ij0g}(d,e)|D_i = d, D_j = e] = E[A_{ij0g}(0,0)|D_i = d, D_j = e]$*, for all pairs* (i,j)*, and* $(d,e) \in \{0,1\}^2$*.*

Assumption 5 (Parallel Trend). *Let* Δ *be the first-difference operator, i.e.,* $\Delta Z = Z_1 - Z_0$*. For each group* g*,*

1. $E[\Delta\varepsilon_{ig}(0)|D_{ig} = 1] = E[\Delta\varepsilon_{ig}(0)|D_{ig} = 0]$*, for each individual* i*,*
2. $E[\Delta A_{ijg}(0,0)|D_{ig} = d, D_{jg} = e] = E[\Delta A_{ijg}(0,0)|D_{ig} = 0, D_{jg} = 0]$*, for all pairs* (i,j) *and* $(d,e) \in \{0,1\}^2$*.*

Assumption 6 (Distributions). *The individual level data* $\{(Y_{i1g}, Y_{i0g}, D_{ig}) : i \in \{1, ..., N\}, g \in \{1, ..., G\}\}$ *are identically distributed over both individuals and groups, and independent across groups. And the dyadic data* $\{(D_{ig}, D_{jg}, A_{ij1g}, A_{ij0g}) : (i,j) \in \{1, ..., N\}^2, g \in \{1, ..., G\}\}$ *are identically distributed over all pairs* (i,j) *and groups* g*, and independent across groups.*

Assumption 7 (Overlap). *For all pairs* (i,j) *and* g*,* $\Pr(D_{ig} = d, D_{jg} = e) \in (0,1)$ *for all* $(d,e) \in \{0,1\}^2$*.*

Assumption 8 (Identification). *Let* $\tilde{\boldsymbol{D}}_{ig} = \left(1, D_{ig}, \sum_{j\neq i} D_{jg}, D_{ig}\sum_{j\neq i} D_{jg}\right)'$*. Then,* $E[\tilde{\boldsymbol{D}}_{ig}\tilde{\boldsymbol{D}}_{ig}']$ *is nonsingular.*

Assumption 4, 5 are additional restrictions on the response function in 2. In particular, by construction of outcome response function, Assumption 4 implies there is no

anticipation about treatment on potential outcomes at the pre-treatment period:

$$E[Y_{i0g}(\boldsymbol{d})|\boldsymbol{D}_g = \boldsymbol{d}] = E[Y_{i0g}(\boldsymbol{0})|\boldsymbol{D}_g = \boldsymbol{d}], \quad \forall \boldsymbol{d} \in \{0,1\}^N,$$

and Assumption 5 implies parallel trend on potential outcomes:[4]

$$E[Y_{i1g}(\boldsymbol{0}) - Y_{i0g}(\boldsymbol{0})|\boldsymbol{D}_g = \boldsymbol{d}] = E[Y_{i1g}(\boldsymbol{0}) - Y_{i0g}(\boldsymbol{0})|\boldsymbol{D}_g = \boldsymbol{0}], \quad \forall \boldsymbol{d} \in \{0,1\}^N.$$

Assumption 6 states potential links and outcomes are identically distributed over individual, pairs, and groups, and independently distributed over groups conditional on treatment assignments. Independence across groups will play a role in constructing the asymptotic properties of the estimators. Assumption 7 is a standard requirement for the existence of conditional expectations. For Assumption 8 to hold, we need between-group variation. Let $\bar{D}_g = \sum j = 1^N D_{jg}$. Therefore, as $\sum_{j \neq i} D_{jg} = \bar{D}_g - D_{ig}$, D_{ig} and $\sum_{j \neq i} D_{jg}$ are not linearly independent if only one group is available.

2.3.2 Identification

Because each group is independent and identical, I suppress the group index for simplicity in this subsection. For each $(d, e) \in \{0,1\}^2$, the average effect on the link between two individuals (i, j) with $D_i = d, D_j = e$ is defined as follows:

$$H(d, e) := E[A_{ij1}(d, e) - A_{ij1}(0, 0)|D_i = d, D_j = e].$$

Identifying this conditional expectation is considered equivalent to identifying the average treatment effect on the treated for multiple treatments (quaternary treatment in this case). It can be established under dyadic parallel trend and no-anticipation assumptions, as investigated in studies such as Frölich (2004) and Fricke (2017). Proposition 1 summarizes the identification results for dyadic difference-in-differences estimand:

Proposition 1 (Difference-in-differences of links)**.** *For each* $t \in \{0,1\}$*, and* $(d, e) \in$

[4]The coefficient γ_2 is the effect of untreated neighbor, and it is assumed to be same for each period $t \in \{0,1\}$ on this purpose. If we allow time varying coefficient for γ_2, the parallel trend for outcome holds if and only if $\gamma_{20} = \gamma_{21}$, or $A_{ij0g}(d, e) = A_{ij0g}$ with probability one.

$\{0,1\}^2$, *define* $m_t(d,e) := E[A_{ijt}|D_i = d, D_j = e]$ *as the probability of* i *and* j *being linked conditional on the treatment statuses, and* $\Delta m(d,e) := m_1(d,e) - m_0(d,e)$ *as the difference of that conditional probability over time. Under Assumptions 4-2, 5-2, 7, 6 we have for each* $(d,e) \in \{0,1\}^2$,

$$H(d,e) = \Delta m(d,e) - \Delta m(0,0), \tag{2}$$

and $H(0,0) = 0$ *by definition.*

Consider the following conditional expectations:

$$\begin{aligned} E[A_{ij1g}|D_{ig}, D_{jg}] &= \widetilde{\boldsymbol{W}}'_{ijg}\boldsymbol{\zeta}_0, \\ E[\Delta A_{ijg}|D_{ig}, D_{jg}] &= \widetilde{\boldsymbol{W}}'_{ijg}\boldsymbol{\xi}_0, \end{aligned} \tag{3}$$

where $\widetilde{\boldsymbol{W}}_{ijg} = (1, D_{ig}, D_{jg}, D_{ig}D_{jg})' \in \mathbb{R}^4$. Note that for any discrete random variable $X \in \{x_0, x_1, ..., x_M\}$ and a random vector Y, we have

$$E[Y|X] = E[Y|X = x_0] + \sum_{j=1}^{M} \mathbb{1}\{X = x_j\}(E[Y|X = x_j] - E[Y|X = x_0]).$$

Thus, Proposition 1 implies that the coefficients in (3) are given by

$$\begin{aligned} \boldsymbol{\zeta}_0 &= \begin{pmatrix} m_1(0,0) \\ m_1(1,0) - m_1(0,0) \\ m_1(0,1) - m_1(0,0) \\ m_1(1,1) - m_1(1,0) - m_1(0,1) + m_1(0,0) \end{pmatrix}, \\ \boldsymbol{\xi}_0 &= \begin{pmatrix} \Delta m(0,0) \\ H(1,0) \\ H(0,1) \\ H(1,1) - H(1,0) - H(0,1) \end{pmatrix}. \end{aligned} \tag{4}$$

In addition, define

$$\boldsymbol{\omega}_0 := \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \times \boldsymbol{\zeta}_0 = \begin{pmatrix} m_1(0,1) \\ m_1(1,1) - m_1(0,1) \end{pmatrix}. \tag{5}$$

Proposition 2 introduces an estimating equation for first-differentiated outcomes with a parameter vector $\boldsymbol{\delta}$. This coefficient vector $\boldsymbol{\delta}$ is expressed by a function of the parameters in the response function: $\boldsymbol{\theta} = (\Delta\alpha, \beta, \gamma_1, \gamma_2)'$, and the parameters related to the distributions of potential links: $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ in (4).

Proposition 2 (Identification of parameters in outcome model)**.** *Let* $\Delta Y_{ig} := Y_{i1g} - Y_{i0g}$. *Then, under Assumptions 1, 2, 3, 4, 5, 7, and 6,*

$$E[\Delta Y_{ig} | \boldsymbol{D}_g] = \tilde{\boldsymbol{D}}_{ig}' \boldsymbol{\delta}, \tag{6}$$

where $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3, \delta_4)' \in \mathbb{R}^4$, $\tilde{\boldsymbol{D}}_{ig} = \left(D_{ig}, \sum_{j \neq i} D_{jg}, D_{ig} \sum_{j \neq i} D_{jg} \right)' \in \mathbb{R}^4$, *and*

$$\begin{aligned}
\delta_1 &= \Delta\alpha + (N-1)\Delta m(0,0)\gamma_2, \\
\delta_2 &= \beta + (N-1)H(1,0)\gamma_2, \\
\delta_3 &= m_1(0,1)\gamma_1 + (H(0,1) - m_1(0,1))\gamma_2, \\
\delta_4 &= (m_1(1,1) - m_1(0,1))\gamma_1 + (H(1,1) - H(1,0) - H(0,1) - m_1(1,1) + m_1(0,1))\gamma_2.
\end{aligned}$$

Further, under Assumption 8, $\boldsymbol{\delta}$ *is identified by a least squares estimand:*

$$\boldsymbol{\delta} = E[\tilde{\boldsymbol{D}}_{ig} \tilde{\boldsymbol{D}}_{ig}']^{-1} E[\tilde{\boldsymbol{D}}_{ig} \Delta Y_{ig}].$$

$\boldsymbol{\delta}$ in (6) has no specific causal interpretations yet. However, it can be used to recover

causal parameters. Define the following matrix

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 & 0 & (N-1)\Delta m(0,0) \\ 0 & 1 & 0 & (N-1)H(1,0) \\ 0 & 0 & m_1(0,1) & H(0,1) - m_1(0,1) \\ 0 & 0 & m_1(1,1) - m_1(0,1) & H(0,1) - H(1,1) - H(1,0) - m_1(1,1) + m_1(0,1) \end{pmatrix}.$$

$$= \begin{pmatrix} 1 & 0 & 0 & (N-1)\xi_1 \\ 0 & 1 & 0 & (N-1)\xi_2 \\ 0 & 0 & \omega_1 & \xi_3 - \omega_1 \\ 0 & 0 & \omega_2 & \xi_4 - \omega_2 \end{pmatrix} \tag{7}$$

Note that $\boldsymbol{H}$ is computed using $\boldsymbol{\omega}$ (or $\boldsymbol{\zeta}$) and $\boldsymbol{\xi}$ defined in (5) and (4), which are identified by Proposition 1. Furthermore, Proposition 2 implies that $\boldsymbol{\delta} = \boldsymbol{H}\boldsymbol{\theta}$, where $\boldsymbol{\theta} = (\Delta\alpha, \beta, \gamma_1, \gamma_2)$ represents the parameters in the response function, as in Assumption 2. Consequently, if $\boldsymbol{H}$ is invertible, $\boldsymbol{\theta}$ is recovered. Further, once $\boldsymbol{\theta}$ is identified, the causal parameters $\boldsymbol{\pi} = (\pi^{DT}, \pi^{DN}, \pi^{IT}, \pi^{IN})'$ defined in Section 2.2 are also identified by using the identified parameters $(\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\zeta}, \boldsymbol{\xi})$. Proposition 3 summarizes this result of identification, representing the main contribution of this study.

Proposition 3 (Identification of Decompositions on Treatment Effects)**.** *Suppose* $\boldsymbol{\delta} = (\delta_1, \delta_1, \delta_2, \delta_3)'$ *in (6),* $\boldsymbol{\xi}$ *in (4), and* $\boldsymbol{\omega}$ *in (5) are identified. Then, the parameters* $\boldsymbol{\theta} := (\Delta\alpha, \beta, \gamma_1, \gamma_2)'$ *are identified by*

$$\boldsymbol{\theta} = \boldsymbol{H}^{-1}\boldsymbol{\delta},$$

where $\boldsymbol{H}$ *is defined in (7), provided that* $\omega_1\xi_4 \neq \omega_2\xi_3$ *(i.e.,* $\boldsymbol{H}$ *is invertible). In addition, the parameters of decompositions defined in Section 2.2 are identified by*

$$\boldsymbol{\pi} = \begin{pmatrix} \pi^{DT} \\ \pi^{DN} \\ \pi^{IT} \\ \pi^{IN} \end{pmatrix} = \begin{pmatrix} \beta \\ \gamma_2(N-1)H(1,0) \\ (\gamma_1 - \gamma_2)m_1(0,1) \\ \gamma_2 H(0,1) \end{pmatrix} = \begin{pmatrix} \theta_2 \\ (N-1)\theta_4\xi_2 \\ (\theta_3 - \theta_4)\omega_1 \\ \theta_4\xi_3 \end{pmatrix}. \tag{8}$$

Remark (Fixed Network)**.** If links are not affected by treatment, then $H(d,e) = 0$, indicating the absence of network effects. Thus, $\pi^{DT} = \beta$, $\pi^{IT} = (\gamma_1 - \gamma_2)m_1(0,1)$ capture the average direct, indirect effect, respectively.

Remark (Random Assignment)**.** If $\{(\varepsilon_{ig}(d), A_{ij1g}(d,e))\} \perp \boldsymbol{D}_g$, then we don't need Assumptions 5, 4. First, because $E[A_{ij1g}|D_i = d, D_j = e] = E[A_{ij1g}(d,e)]$, we have $H(d,e) = m_1(d,e) - m_1(0,0)$. Similarly, $\boldsymbol{\delta} = E[\tilde{\boldsymbol{D}}_{ig}\tilde{\boldsymbol{D}}_{ig}']^{-1}E[\tilde{\boldsymbol{D}}_{ig}Y_{i1g}]$. The identification of $\boldsymbol{\theta}$, $\boldsymbol{\pi}$ follows the same transformation with

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 & 0 & (N-1)m_1(0,0) \\ 0 & 1 & 0 & (N-1)(m_1(1,0) - m_1(0,0)) \\ 0 & 0 & m_1(0,1) & -m_1(0,0) \\ 0 & 0 & m_1(1,1) - m_1(0,1) & -m_1(1,0) + m_1(0,0) \end{pmatrix}.$$

Remark (Without Interactions)**.** If $A_{ijtg} = 0$ for all i, j, t, g, then the only parameter of interest is β, which is identified by the canonical difference-in-differences estimand.

3 Estimation and Inference

In this section, I present estimators for the parameters identified in Section 2.3 and the decompositions defined in Section 3. Since all identification arguments are constructive, the natural choice for estimators is plug-in estimators. Additionally, because each parameter is defined by a projection coefficient for a conditional expectation, the plug-in estimators are essentially least squares estimators. Hence, the estimation procedure is straightforward but requires three stages. For each estimator, clustered standard errors can be employed to conduct inference, taking into account the dependency within groups.

3.1 Estimators

I propose a three-stage procedure to estimate the parameters of interest. The first-stage estimators estimate the distributions of potential links $(\boldsymbol{\omega}, \boldsymbol{\xi})$ and the coefficient of outcome regression $(\boldsymbol{\delta})$. Subsequently, the parameters in the response function $(\boldsymbol{\theta})$

in Assumption 2 are estimated in the second stage, by using the first-stage estimates. Finally, the decompositions of causal effects ($\boldsymbol{\pi}$) are estimated in the third stage, by using the estimates from the first and second stages. To this end, define the following notations:

$$\tilde{\boldsymbol{W}}_{ijg} := (1, D_{ig}, D_{jg}, D_{ig}D_{jg})' \in \mathbb{R}^4, \qquad \tilde{\boldsymbol{W}}_g := (\tilde{\boldsymbol{W}}_{11g}, ..., \tilde{\boldsymbol{W}}_{NNg})' \in \mathbb{R}^{N(N-1)\times 4},$$

$$\tilde{\boldsymbol{D}}_{ig} := \left(1, D_{ig}, \sum_{j\neq i} D_{jg}, D_{ig}, \sum_{j\neq i} D_{jg}\right)' \in \mathbb{R}^4, \qquad \tilde{\boldsymbol{D}}_g := (\tilde{\boldsymbol{D}}_{1g}, ..., \tilde{\boldsymbol{D}}_{Ng})' \in \mathbb{R}^{N\times 4},$$

$$\boldsymbol{A}_{tg} := (A_{11tg}, ..., A_{NNg})' \in \mathbb{R}^{N(N-1)},$$

$$\boldsymbol{Y}_{tg} := (Y_{1tg}, ..., Y_{Ntg})' \in \mathbb{R}^N.$$

The first-stage parameters, denoted as $\boldsymbol{\xi}_0, \boldsymbol{\omega}_0, \boldsymbol{\delta}_0$, are defined in (4), (5), and (6). They are estimated using the following least squares estimators:

$$\hat{\boldsymbol{\omega}} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \left[\frac{1}{G}\sum_{g=1}^{G} \tilde{\boldsymbol{W}}_g' \tilde{\boldsymbol{W}}_g'\right]^{-1} \frac{1}{G}\sum_{g=1}^{G} \tilde{\boldsymbol{W}}_g' \boldsymbol{A}_{1g},$$

$$\hat{\boldsymbol{\xi}} = \left[\frac{1}{G}\sum_{g=1}^{G} \tilde{\boldsymbol{W}}_g' \tilde{\boldsymbol{W}}_g'\right]^{-1} \frac{1}{G}\sum_{g=1}^{G} \tilde{\boldsymbol{W}}_g' (\boldsymbol{A}_{1g} - \boldsymbol{A}_{0g}),$$

$$\hat{\boldsymbol{\delta}} = \left[\frac{1}{G}\sum_{g=1}^{G} \tilde{\boldsymbol{D}}_g' \tilde{\boldsymbol{D}}_g'\right]^{-1} \frac{1}{G}\sum_{g=1}^{G} \tilde{\boldsymbol{D}}_g' (\boldsymbol{Y}_{1g} - \boldsymbol{Y}_{0g}).$$

The estimator $\hat{\boldsymbol{\omega}}$ estimates the conditional expectations of links at $t = 1$. On the other hand, $\hat{\boldsymbol{\xi}}$ is the difference-in-differences estimator for links, using dyadic data to estimate the average treatment effect of the treated on network links. $\hat{\boldsymbol{\delta}}$ represents the coefficient of the conditional mean model of the outcome in (6). Once these estimators, $\hat{\boldsymbol{\omega}}$, $\hat{\boldsymbol{\xi}}$, and $\hat{\boldsymbol{\delta}}$, are computed, the parameter $\boldsymbol{\theta}_0$ is estimated by $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{H}}^{-1}\hat{\boldsymbol{\delta}}$, where

$$\hat{\boldsymbol{H}} = \begin{pmatrix} 1 & 0 & 0 & \hat{\xi}_1 \\ 0 & 1 & 0 & \hat{\xi}_2 \\ 0 & 0 & \hat{\omega}_1 & \hat{\xi}_3 - \hat{\omega}_1 \\ 0 & 0 & \hat{\omega}_2 & \hat{\xi}_4 - \hat{\omega}_2 \end{pmatrix}.$$

And lastly, the decompositions $\boldsymbol{\pi}_0$ is estimated by the following plug-in estimator:

$$\hat{\boldsymbol{\pi}} = \begin{pmatrix} \hat{\delta}_2 \\ (N-1)\hat{\gamma}_2\hat{\xi}_2 \\ (\hat{\delta}_3 - \hat{\delta}_4)\hat{\omega}_1 \\ \hat{\delta}_4\hat{\xi}_3 \end{pmatrix}.$$

3.2 Inference

Since the proposed estimators are least squares estimators for projection coefficients, standard large sample theories apply. The random sample of i.i.d. groups plays a key role in the application of asymptotic theories. Additionally, the following assumptions are required as standard regularity conditions.

Assumption 9 (Regularity Conditions)**.**

- $E[\tilde{\boldsymbol{W}}_g\tilde{\boldsymbol{W}}_g']$, $E[\tilde{\boldsymbol{D}}_g\tilde{\boldsymbol{D}}_g']$ *are nonsingular.*
- $E\left[|Y_{itg}|^4\right] < \infty$

The first part of Assumption 9 is a regularity condition to ensure the uniqueness and thus identification of parameters as the projection coefficients. The second part is employed to derive the asymptotic distribution as appropriate bounded moments are required. Since indicator variables are bounded, only the boundedness of outcome is stated. Let "$\Rightarrow \mathcal{N}$" denote convergence in distribution for some multivariate normal distribution. Proposition 4 summarizes the influence functions for parameters in each stage:

Proposition 4. *Let* $\boldsymbol{V}_g = (\tilde{\boldsymbol{W}}_g, \boldsymbol{A}_{1g}, \boldsymbol{A}_{0g}, \boldsymbol{Y}_{1g}, \boldsymbol{Y}_{0g}, \tilde{\boldsymbol{D}}_g)$, $\boldsymbol{M} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$, $\boldsymbol{Q} =$

$E[\tilde{\boldsymbol{W}}_g\tilde{\boldsymbol{W}}_g']$, *and* $\boldsymbol{R} = E[\tilde{\boldsymbol{D}}_g\tilde{\boldsymbol{D}}_g']$. *Under Assumptions 2, 1, 3, 5, 4, 7, 6, 9, we have*

$$\sqrt{G}\begin{pmatrix}\hat{\boldsymbol{\omega}}-\boldsymbol{\omega}_0\\ \hat{\boldsymbol{\xi}}-\boldsymbol{\xi}_0\\ \hat{\boldsymbol{\delta}}-\boldsymbol{\delta}_0\\ \hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\\ \hat{\boldsymbol{\pi}}-\boldsymbol{\pi}_0\end{pmatrix} = \frac{1}{\sqrt{G}}\sum_{g=1}^{G}\begin{pmatrix}\boldsymbol{M}\psi_{\boldsymbol{\zeta}}(\boldsymbol{V}_g)\\ \psi_{\boldsymbol{\xi}}(\boldsymbol{V}_g)\\ \psi_{\boldsymbol{\delta}}(\boldsymbol{V}_g)\\ \psi_{\boldsymbol{\theta}}(\boldsymbol{V}_g)\\ \psi_{\boldsymbol{\pi}}(\boldsymbol{V}_g)\end{pmatrix} + o_p(1) \Rightarrow \mathcal{N},$$

where

$$\begin{aligned}
\psi_{\boldsymbol{\zeta}}(\boldsymbol{V}_g) &= \boldsymbol{Q}^{-1}\tilde{\boldsymbol{W}}_g'(\boldsymbol{A}_{1g}-\tilde{\boldsymbol{W}}_g\boldsymbol{\zeta}_0),\\
\psi_{\boldsymbol{\xi}}(\boldsymbol{V}_g) &= \boldsymbol{Q}^{-1}\tilde{\boldsymbol{W}}_g'(\Delta\boldsymbol{A}_g-\tilde{\boldsymbol{W}}_g\boldsymbol{\xi}_0),\\
\psi_{\boldsymbol{\delta}}(\boldsymbol{V}_g) &= \boldsymbol{R}^{-1}\tilde{\boldsymbol{D}}_g'(\Delta\boldsymbol{Y}_g-\tilde{\boldsymbol{D}}_g\boldsymbol{\delta}_0),\\
\psi_{\boldsymbol{\theta}} &= \boldsymbol{H}_0^{-1}(\psi_{\boldsymbol{\delta}}-\psi_{\boldsymbol{H}}\boldsymbol{\theta}_0),\ \text{where } \psi_{\boldsymbol{H}} = \begin{pmatrix}1&0&0&\psi_{\xi_1}\\ 0&1&0&\psi_{\xi_2}\\ 0&0&\psi_{\omega_1}&\psi_{\xi_3}-\psi_{\omega_1}\\ 0&0&\psi_{\omega_2}&\psi_{\xi_4}-\psi_{\omega_2}\end{pmatrix},\\
\psi_{\boldsymbol{\pi}} &= \begin{pmatrix}\psi_{\theta_2}\\ (N-1)(\psi_{\theta_4}H(1,0)+\gamma_2\psi_{\xi_2})\\ (\psi_{\theta_3}-\psi_{\theta_4})\,m_1(0,1)+(\gamma_1-\gamma_2)\psi_{\omega_1}\\ \psi_{\theta_3}H(0,1)+\gamma_2\psi_{\xi_3}\end{pmatrix}.
\end{aligned}$$

Proposition 4 implies that $\hat{\boldsymbol{\pi}}$ has asymptotic normal distribution with zero mean and the following asymptotic variance

$$\text{Avar}(\sqrt{G}(\hat{\boldsymbol{\pi}}-\boldsymbol{\pi}_0)) = E[\psi_{\boldsymbol{\pi}}(\boldsymbol{V}_g)\psi_{\boldsymbol{\pi}}(\boldsymbol{V}_g)'].$$

Let $\hat{\psi}_{\boldsymbol{\pi}}$ be the empirical influence function (plug-in estimator) for $\psi_{\boldsymbol{\pi}}$. Then, the clustered standard error is computed as the square root of the diagonal elements of the

following matrix:

$$\frac{1}{G^2}\sum_{g=1}^{G}\hat{\psi}_{\boldsymbol{\pi}}(\boldsymbol{Z}_g)\hat{\psi}_{\boldsymbol{\pi}}(\boldsymbol{Z}_g)'. \tag{9}$$

Note that

$$\frac{1}{G}\sum_{g=1}^{G}\hat{\psi}_{\boldsymbol{\pi}}(\boldsymbol{Z}_g)\hat{\psi}_{\boldsymbol{\pi}}(\boldsymbol{Z}_g)' \xrightarrow{p} E[\psi_{\boldsymbol{\pi}}(\boldsymbol{Z}_g)\psi_{\boldsymbol{\pi}}(\boldsymbol{Z}_g)'],$$

and therefore the inference based on the clustered standard error is asymptotically valid. Similar clustered standard errors are defined similarly for the other parameters.

4 Monte Carlo Simulation

To investigate the finite and asymptotic characteristics of the estimators intoduced in Section 3, I conduct simulations using data generated based on the Assumptions studied in Section 2 with different sample sizes.

First, the treatment indicators D_{ig} are generated by a Bernoulli distribution with the probability $P_D = 0.5$. Next, the potential network links are generated according to the following binary choice model: For each $(i,j) \in \{1,...,N\}^2$, $t \in \{0,1\}$, $g \in \{1,...,G\}$, $A_{iitg}(d,e) = 0$ for all $(d,e) \in \{0,1\}^2$, and

$$A_{ijtg}(d,e) = \mathbb{1}\{f_t(d,e) \geq u_{ijtg}\}, \quad i \neq j,$$

where $f_t(d,e) = (1,d,e,de)\boldsymbol{a}_t$, $\boldsymbol{a}_0 = (0.1,0,0,0)'$, $\boldsymbol{a}_1 = (-0.3,0.5,0.1,0.6)'$, and $u_{ijtg} = v_{ijg} - g_t(D_{ig}, D_{jg})$, where $v_{ijg} \sim F_v$. Consequently, we have $E[A_{ijtg}(d,e)|D_{ig} = d', D_{jg} = e'] = F_v(f_t(d,e) + g_t(d',e'))$. Assumption 4 is satisfied as $f_0(d,e) = f_0(0,0)$ for all $(d,e) \in \{0,1\}^2$. The values for function g are set by $g_0(1,1) = 0.4$, $g_0(1,0) = 0.3$, $g_0(0,1) = 0.2$, $g_0(0,0) = 0.1$, $g_1(0,0) = 0.01$, and the remaining values are constructed for the potential links to satisfy Assumption 5.[5] Lastly, the outcome is generated from

[5]across all $(d,e) \in \{0,1\}^2$, then the values of g_1 must also be same for Assumption 5, and hence u_{ijtg} independent of the treatment. Consequently, the difference-in-differences approach becomes unnecessary, as the differences in means are enough to identify the causal effect of treatment on potential links. To avoid this scenario, I initially specify the values of g_0 and $g_1(0,0)$ and then construct the

the response function defined in Assumption 2 :

$$Y_{i1g} = \alpha_1 + \beta D_{ig} + \gamma_1 \sum_{j=1}^{N} A_{ij1g} D_{jg} + \gamma_2 \sum_{j=1}^{N} A_{ij1g}(1 - D_{jg}) + u_{i1g},$$
$$Y_{i0g} = \alpha_2 + \gamma_2 \sum_{j=1}^{N} A_{ij0g} + u_{i0g},$$

where $u_{itg} \sim N(0,1) + s_b(D_{ig} - P_D)$ to consider both exogenous and endogenous treatment. The parameters of the response function are set by $\theta := (\alpha_0, \alpha_1, \beta, \gamma_1, \gamma_2) = (1, 1.2, 5, 0.6, 0.3)$.

The generated data consists of a dyadic level data $\{D_{ig}D_{jg}, A_{ij0g}, A_{ij1g} : (i,j) \in \{1,...,N\}^2, i \neq j, g \in \{1,...,G\}\}$, and individual level data $\{D_{ig}, Y_{i0g}, Y_{i1g} : i \in \{1,...,N\}, g \in \{1,...,G\}\}$. The estimators defined in Section 3, and clustered standard errors for each estimators are computed. The coverage rate is computed by the proportion of the cases that estimate is included in the 95% confidence interval, computed by the clusterd standard error, among all simulations. For instance, for a parameter $\phi \in \mathbb{R}^K$, let $\hat{\phi}_b$ be the estimator for a replication b, and $CI_b(\phi) = [\hat{\phi}_b \pm z_{\alpha/2} SE_b(\phi)]$, where $z_{\alpha/2} = \Phi^{-1}(1-\alpha/2)$, $\alpha = 0.05$. Then,

$$\text{MSE} = \frac{1}{B}\sum_{b=1}^{B}(\hat{\phi}_b - \phi_0)(\hat{\phi}_b - \phi_0)',$$
$$\text{Coverage} = \frac{1}{B}\sum_{b=1}^{B}\mathbb{1}\{\phi_0 \in CI_b(\phi)\}.$$

Tables 1 and 2 display results for each parameters. The suggested asymptotic theory and the clustered standard errors work well, as all coverage rates are close to 95%, and the mean squared errors converges to zero. For detailed results on each parameter, see Tables **??** in Appendix B.

---

remaining values as

$$g_1(d,e) = \Phi^{-1}(F_v(a_{10} + g_1(0,0)) - F_v(a_{00} + g_0(0,0)) + F_v(a_{00} + g_0(d,e)) - a_{10}),$$

which ensures that Assumption 5 is satisfied by construction.

Table 1: Root Mean Squared Errors

		Parameteres					
G	N	$\boldsymbol{\xi}$	$\boldsymbol{\zeta}$	$\boldsymbol{\omega}$	$\boldsymbol{\delta}$	$\boldsymbol{\theta}$	$\boldsymbol{\pi}$
100	15	0.014	0.019	0.011	0.387	2.781	2.55
500	15	0.006	0.009	0.005	0.162	1.096	1.004
1,000	15	0.005	0.006	0.003	0.121	0.82	0.751
5,000	15	0.002	0.003	0.001	0.055	0.377	0.344
10,000	15	0.001	0.002	0.001	0.037	0.247	0.226

notes. For each sample size N, G, the number of simulations is set to $B = 500$.

Table 2: Coverage Rates

		Parameteres					
G	N	$\boldsymbol{\xi}$	$\boldsymbol{\zeta}$	$\boldsymbol{\omega}$	$\boldsymbol{\delta}$	$\boldsymbol{\theta}$	$\boldsymbol{\pi}$
100	15	0.944	0.946	0.936	0.92	0.926	0.932
500	15	0.944	0.93	0.932	0.958	0.954	0.958
1,000	15	0.938	0.936	0.954	0.94	0.932	0.934
5,000	15	0.944	0.946	0.956	0.928	0.946	0.94
10,000	15	0.946	0.944	0.948	0.952	0.95	0.956

notes. For each sample size N, G, the number of simulations is set to $B = 500$.

5 Conclusion

This study contributes to the literature by introducing a novel approach that considers the possibility of changing the network resulting from treatment. The proposed decomposition of treatment effects distinguishes between the impact when the network remains unchanged and the effect of the treatment on the network structure. This generalization of existing approaches provides a more comprehensive understanding of program effects, accommodating scenarios where the network changes due to treatment. Moreover, the proposed identification addressess quasi-experimental situations, particularly using the Difference-in-Differences approach.

References

Banerjee, Abhijit et al. (2023, forthcoming). "Changes in social network structure in response to exposure to formal credit markets". In: *Review of Economic Studies*.

Blackwell, Matthew (2017). "Instrumental variable methods for conditional effects and causal interaction in voter mobilization experiments". In: *Journal of the American Statistical Association* 112.518, pp. 590–599.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009). "Identification of peer effects through social networks". In: *Journal of econometrics* 150.1, pp. 41–55.

Butts, Kyle (2021). "Difference-in-differences estimation with spatial spillovers". In: *arXiv preprint arXiv:2105.03737*.

Chernozhukov, Victor et al. (2022). "Locally robust semiparametric estimation". In: *Econometrica* 90.4, pp. 1501–1535.

Comola, Margherita and Silvia Prina (2021). "Treatment effect accounting for network changes". In: *Review of Economics and Statistics* 103.3, pp. 597–604.

DiTraglia, Francis J et al. (2023). "Identifying causal effects in experiments with spillovers and non-compliance". In: *Journal of Econometrics*.

Dupas, Pascaline, Anthony Keats, and Jonathan Robinson (2019). "The effect of savings accounts on interpersonal financial relationships: Evidence from a field experiment in rural Kenya". In: *The Economic Journal* 129.617, pp. 273–310.

Forastiere, Laura, Edoardo M Airoldi, and Fabrizia Mealli (2021). "Identification and estimation of treatment and interference effects in observational studies on networks". In: *Journal of the American Statistical Association* 116.534, pp. 901–918.

Fricke, Hans (2017). "Identification based on difference-in-differences approaches with multiple treatments". In: *Oxford Bulletin of Economics and Statistics* 79.3, pp. 426–433.

Frölich, Markus (2004). "Programme evaluation with multiple treatments". In: *Journal of Economic Surveys* 18.2, pp. 181–224.

Graham, Bryan S (2017). "An econometric model of network formation with degree heterogeneity". In: *Econometrica* 85.4, pp. 1033–1063.

Hoshino, Tadao and Takahide Yanagi (2021). "Causal inference with noncompliance and unknown interference". In: *arXiv preprint arXiv:2108.07455*.

Ichimura, Hidehiko and Whitney K Newey (2022). "The influence function of semiparametric estimators". In: *Quantitative Economics* 13.1, pp. 29–61.

Kang, Hyunseung and Guido Imbens (2016). "Peer encouragement designs in causal inference with partial interference and identification of local average network effects". In: *arXiv preprint arXiv:1609.04464*.

Kormos, Mate, Robert P Lieli, and Martin Huber (2023). "Treatment Effect Analysis for Pairs with Endogenous Treatment Takeup". In: *arXiv preprint arXiv:2301.04876*.

Leung, Michael P (2020). "Treatment and spillover effects under network interference". In: *Review of Economics and Statistics* 102.2, pp. 368–380.

— (2022). "Causal inference under approximate neighborhood interference". In: *Econometrica* 90.1, pp. 267–293.

Manski, Charles F (2013). "Identification of treatment response with social interactions". In: *The Econometrics Journal* 16.1, S1–S23.

Rubin, Donald B (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." In: *Journal of educational Psychology* 66.5, p. 688.

Vazquez-Bare, Gonzalo (2022a). "Causal spillover effects using instrumental variables". In: *Journal of the American Statistical Association*, pp. 1–12.

— (2022b). "Identification and estimation of spillover effects in randomized experiments". In: *Journal of Econometrics*.

Xu, Ruonan (2023 WP). "Difference-in-Differences with Interference: A Finite Population Perspective". In: *Working Paper*.

A Proofs

Proof of Proposition 1. Let $(d, e) \in \{(0,1),(1,0),(1,1)\}$ be given. Then,

$$\begin{aligned}
&E[A_{ij1g}(0,0)|D_{ig}=d, D_{jg}=e] \\
&= E[A_{ij0g}(0,0)|D_{ig}=d, D_{jg}=e] + E[A_{ij1g}(0,0) - A_{ij0g}(0,0)|D_{ig}=0, D_{jg}=0] \\
&= E[A_{ij0g}(d,e)|D_{ig}=d, D_{jg}=e] + E[A_{ij1g} - A_{ij0g}|D_{ig}=0, D_{jg}=0] \\
&= E[A_{ij0g}|D_{ig}=d, D_{jg}=e] + E[A_{ij1g} - A_{ij0g}|D_{ig}=0, D_{jg}=0], \qquad (10)
\end{aligned}$$

where the first equation is by Assumption 5-2, and the second equation is by Assumption 4-2. It follows that

$$\begin{aligned}
H(d,e) &= E[A_{ij1g}(d,e) - A_{ij1g}(0,0)|D_{ig}=d, D_{jg}=e] \\
&= E[A_{ij1g}(d,e)|D_{ig}=d, D_{jg}=e] - E[A_{ij1g}(0,0)|D_{ig}=d, D_{jg}=e] \\
&= E[A_{ij1g}|D_{ig}=d, D_{jg}=e] - E[A_{ij0g}|D_{ig}=d, D_{jg}=e] \\
&\quad - E[A_{ij1g} - A_{ij0g}|D_{ig}=0, D_{jg}=0] \\
&= E[A_{ij1g} - A_{ij0g}|D_{ig}=d, D_{jg}=e] - E[A_{ij1g} - A_{ij0g}|D_{ig}=0, D_{jg}=0],
\end{aligned}$$

and $H_g(0,0) = 0$ by construction. Assumption 7 guarantees the existence of conditional expectations.

□

Lemma 1. *Let $\boldsymbol{D} = (D_1, ..., D_N) \in \{0,1\}^N$ be a random vector, and $Y_{it} : \{0,1\}^N \to \mathbb{R}$, $f_{ijt} : \{0,1\}^2 \to \mathbb{R}$, $h_{it} : \{0,1\} \to \mathbb{R}$ are functions. Assume*

- *(LR) $Y_{it}(\boldsymbol{d}) = h_{it}(d_i) + \sum_{j=1, j\neq i}^{n} f_{ijt}(d_i, d_j)$, for all $(d_i, d_j) \in \{0,1\}^2$.*
- *(LE-h) $E[h_{it}(d)|\boldsymbol{D}] = E[h_{it}(d)|D_i]$, for all $d \in \{0,1\}$.*
- *(LE-f) $E[f_{ijt}(d,e)|\boldsymbol{D}] = E[f_{ijt}(d,e)|D_i, D_j]$, for all $(d,e) \in \{0,1\}^2$.*

Then, $E[Y_{i1} - Y_{i0}|\boldsymbol{D}] = \tilde{\boldsymbol{D}}_i'\boldsymbol{\theta}$*, where*

$$\tilde{\boldsymbol{D}}_i = \begin{pmatrix} 1 \\ D_i \\ \sum_{j-1,j\neq i}^{n} D_j \\ D_i \sum_{j-1,j\neq i}^{n} D_j \end{pmatrix}, \quad \boldsymbol{\theta} = \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \theta_3 \end{pmatrix} = \begin{pmatrix} \alpha + \sum_{j=1,j\neq i}^{n} K_{ij}(0,0) \\ \beta + \sum_{j=1,j\neq i}^{n} (K_{ij}(1,0) - K_{ij}(0,0)) \\ K_{ij}(0,1) - K_{ij}(0,0) \\ K_{ij}(1,1) - K_{ij}(1,0) - K_{ij}(0,1) + K_{ij}(0,0) \end{pmatrix},$$

and $K_{ij}(d,e) := E[f_{ij1} - f_{ij0}|D_i = d, D_j = e]$ *for* $(d,e) \in \{0,1\}^2$.

Proof of Lemma 1. Let $\alpha := E[h_{i1} - h_{i0}|D_i = 0]$ and $\beta := E[h_{i1} - h_{i0}|D_i = 1] - E[h_{i1} - h_{i0}|D_i = 0]$. Then, Assumption (LE-h) implies

$$E[h_{i1} - h_{i0}|\boldsymbol{D}] = E[h_{i1} - h_{i0}|D_i] = \alpha + \beta D_i. \tag{11}$$

Next, let $K_{ij} = K_{ij}(D_i, D_j) = E[f_{ij1} - f_{ij0}|D_i, D_j]$. Then, (LE-f) implies

$$\begin{aligned} E[f_{ij1} - f_{ij0}|\boldsymbol{D}] = K_{ij} &= \sum_{(d,e)\in\{0,1\}^2} K_{ij}(d,e) \\ &= K_{ij}(0,0) + D_i(K_{ij}(1,0) - K_{ij}(0,0)) \\ &\quad + D_j(K_{ij}(0,1) - K_{ij}(0,0)) \\ &\quad + D_iD_j(K_{ij}(1,1) - K_{ij}(1,0) - K_{ij}(0,1) + K_{ij}(0,0)). \end{aligned} \tag{12}$$

By combining (LR), (11), and (12), we have

$$\begin{aligned} E[Y_{i1} - Y_{i0}|\boldsymbol{D}] &= E[h_{i1} - h_{i0}|\boldsymbol{D}] + \sum_{j=1,j\neq i}^{n} E[f_{ij1} - f_{ij0}|\boldsymbol{D}] \\ &= \alpha + \sum_{j=1,j\neq i}^{n} K_{ij}(0,0) \\ &\quad + D_i \left(\beta + \sum_{j=1,j\neq i}^{n} (K_{ij}(1,0) - K_{ij}(0,0)) \right) + \sum_{j=1,j\neq i}^{n} D_j(K_{ij}(0,1) - K_{ij}(0,0)) \\ &\quad + D_i \sum_{j=1,j\neq i}^{n} D_j(K_{ij}(1,1) - K_{ij}(1,0) - K_{ij}(0,1) + K_{ij}(0,0)) = \tilde{\boldsymbol{D}}_i'\boldsymbol{\theta}. \end{aligned}$$

□

Proof of Proposition 2. Let

$$f_{ijtg}(d_i, d_j) = \begin{cases} \gamma_1 A_{ij1g}(d_i, d_j) d_j + \gamma_2 A_{ij1g}(d_i, d_j)(1 - d_j) & t = 1, \\ \gamma_3 A_{ij0g}(d_i, d_j) & t = 0. \end{cases}$$

$$h_{it}(d) = \begin{cases} \alpha_{1g} + \gamma_0 d_i + \varepsilon_{i1g} & t = 1, \\ \alpha_{0g} + \varepsilon_{i0g} & t = 0. \end{cases}$$

Then, Assumptions 2, 1, 3 imply Assumptions in the lemma 1. Also,

$$f_{ij1g}(d, e) - f_{ij0g}(d, e) = A_{ij1g}(d, e)\left(\gamma_2 + (\gamma_1 - \gamma_2)e\right) - \gamma_3 A_{ij0g}(d, e).$$

It follows that

$$\begin{aligned} &E[f_{ij1g}(0,0) - f_{ij0g}(0,0) | D_i = d, D_j = e] \\ &= E[A_{ij1g}(0,0) | D_i = d, D_j = e]\gamma_2 - E[A_{ij0g}(0,0) | D_i = d, D_j = e]\gamma_3, \\ &= E[A_{ij1g}(0,0) - A_{ij0g}(0,0) | D_i = d, D_j = e]\gamma_2 + E[A_{ij0g}(0,0) | D_i = d, D_j = e](\gamma_2 - \gamma_3). \end{aligned}$$

by Assumption 5, the first term doesn't depend on d, e. Hence, for the potential outcome to satisfy Assumption 5, we need to assume $\gamma_2 = \gamma_3$ or $E[A_{ij0g}(0,0) | D_i = d, D_j = e] = E[A_{ij0g}(0,0)]$ for all $(d, e) \in \{0, 1\}^2$.

Case 1. Assume $\gamma_2 = \gamma_2$.

In this case, it follows that

$$f_{ij1g}(d, e) - f_{ij0g}(d, e) = (A_{ij1g}(d, e) - A_{ij0g}(d, e))\gamma_2 + A_{ij1g}(d, e)(\gamma_1 - \gamma_2)e,$$

and therefore

$$\begin{aligned} K_{ij}(d, e) &= \Delta m(d, e)\gamma_2 + m_1(d, e)(\gamma_1 - \gamma_2)e, \\ K_{ij}(d, e) - K_{ij}(0, 0) &= H(d, e)\gamma_2 + m_1(d, e)(\gamma_1 - \gamma_2)e. \end{aligned}$$

Thus we have the desired result

$$\boldsymbol{\theta} = \begin{pmatrix} \alpha + (N-1)\Delta m(0,0)\gamma_2 \\ \beta + (N-1)H(1,0)\gamma_2 \\ H(0,1)\gamma_2 + m_1(0,1)(\gamma_1 - \gamma_2) \\ (H(1,1) - H(1,0) - H(0,1))\gamma_2 + (m_1(1,1) - m_1(0,1))(\gamma_1 - \gamma_2) \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 & 0 & (N-1)\Delta m(0,0) \\ 0 & 1 & 0 & (N-1)H(1,0) \\ 0 & 0 & m_1(0,1) & H(0,1) - m_1(0,1) \\ 0 & 0 & m_1(1,1) - m_1(0,1) & m_0(1,1) - H(1,0) - m_0(0,1) \end{pmatrix} \begin{pmatrix} \Delta\alpha \\ \beta \\ \gamma_1 \\ \gamma_2 \end{pmatrix}.$$

Case 2. Assume $E[A_{ij0g}(0,0)|D_i = d, D_j = e] = E[A_{ij0g}(0,0)]$ for all $(d,e) \in \{0,1\}^2$.

In this case,

$$\begin{aligned} K_{ij}(d,e) &= m_1(d,e)\gamma_2 + m_1(d,e)(\gamma_1 - \gamma_2)e - m_0\gamma_3 \\ &= \Delta m(d,e)\gamma_2 + m_1(d,e)(\gamma_1 - \gamma_2)e + (\gamma_2 - \gamma_3)m_0, \\ K_{ij}(d,e) - K_{ij}(0,0) &= H(d,e)\gamma_2 + m_1(d,e)(\gamma_1 - \gamma_2)e. \end{aligned}$$

Thus we have

$$\boldsymbol{\theta} = \begin{pmatrix} \alpha + (N-1)m_1(0,0)\gamma_2 - (N-1)m_0\gamma_3 \\ \beta + (N-1)H(1,0)\gamma_2 \\ H(0,1)\gamma_2 + m_1(0,1)(\gamma_1 - \gamma_2) \\ (H(1,1) - H(1,0) - H(0,1))\gamma_2 + (m_1(1,1) - m_1(0,1))(\gamma_1 - \gamma_2) \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 0 & 0 & (N-1)m_1(0,0) \\ 0 & 1 & 0 & (N-1)H(1,0) \\ 0 & 0 & m_1(0,1) & H(0,1) - m_1(0,1) \\ 0 & 0 & m_1(1,1) - m_1(0,1) & m_0(1,1) - H(1,0) - m_0(0,1) \end{pmatrix} \begin{pmatrix} \Delta\alpha - (N-1)m_0\gamma_3 \\ \beta \\ \gamma_1 \\ \gamma_2 \end{pmatrix}.$$

As a result, $\Delta\alpha$ and γ_3 are not separately identified in this case. However, $\boldsymbol{\pi}$ is identified by the same transformation $\boldsymbol{H}$. □

Proof of Proposition 3. This follows directly by invertibility of $\boldsymbol{H}$. □

Proof of Proposition 4. Let $\boldsymbol{Q} = E[\tilde{\boldsymbol{W}}_g'\tilde{\boldsymbol{W}}_g]$, $\hat{\boldsymbol{Q}} = \frac{1}{G}\sum_{g=1}^{G}\tilde{\boldsymbol{W}}_g'\tilde{\boldsymbol{W}}_g$, $\boldsymbol{R} = E[\tilde{\boldsymbol{D}}_g'\tilde{\boldsymbol{D}}_g]$ and $\hat{\boldsymbol{R}} = \frac{1}{G}\sum_{g=1}^{G}\tilde{\boldsymbol{D}}_g'\tilde{\boldsymbol{D}}_g$. Then, $\hat{\boldsymbol{Q}}^{-1} = \boldsymbol{Q}^{-1} + o_p(1)$ and $\hat{\boldsymbol{R}}^{-1} = \boldsymbol{R}^{-1} + o_p(1)$ by a law of large numbers and the continuous mapping theorem, since Q, E are nonsingular and data is i.i.d. across groups. Hence,

$$\begin{aligned}
\sqrt{G}(\hat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}_0) &= \frac{1}{\sqrt{G}}\sum_{g=1}^{G}\boldsymbol{Q}^{-1}\tilde{\boldsymbol{W}}_g'\varepsilon_g^{(\zeta)} + o_p(1)\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\tilde{\boldsymbol{W}}_g'\varepsilon_g^{(\zeta)},\\
\sqrt{G}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) &= \frac{1}{\sqrt{G}}\sum_{g=1}^{G}\boldsymbol{Q}^{-1}\tilde{\boldsymbol{W}}_g'\varepsilon_g^{(\xi)} + o_p(1)\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\tilde{\boldsymbol{W}}_g'\varepsilon_g^{(\xi)},\\
\sqrt{G}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}_0) &= \frac{1}{\sqrt{G}}\sum_{g=1}^{G}\boldsymbol{R}^{-1}\tilde{\boldsymbol{D}}_g'\varepsilon_g^{(\delta)} + o_p(1)\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\tilde{\boldsymbol{D}}_g'\varepsilon_g^{(\delta)},
\end{aligned}$$

where

$$\begin{aligned}
\varepsilon_g^{(\zeta)} &= \boldsymbol{A}_{1g} - \tilde{\boldsymbol{W}}_g\boldsymbol{\zeta}_0,\\
\varepsilon_g^{(\xi)} &= \Delta\boldsymbol{A}_g - \tilde{\boldsymbol{W}}_g\boldsymbol{\xi}_0,\\
\varepsilon_g^{(\delta)} &= \Delta\boldsymbol{Y}_g - \tilde{\boldsymbol{D}}_g\boldsymbol{\delta}_0.
\end{aligned}$$

Data are i.i.d. across groups, and by the condition $E\left[|Y_{itg}|^4\right] < \infty$, the second term in each equation is $O_p(1)$. Therefore the influence functions for $\boldsymbol{\zeta}, \boldsymbol{\xi}, \boldsymbol{\delta}$ are given by

$$\begin{aligned}
\psi_{\boldsymbol{\zeta}}(\boldsymbol{Z}_g) &= \boldsymbol{Q}^{-1}\tilde{\boldsymbol{W}}_g'\varepsilon_g^{(\zeta)},\\
\psi_{\boldsymbol{\xi}}(\boldsymbol{Z}_g) &= \boldsymbol{Q}^{-1}\tilde{\boldsymbol{W}}_g'\varepsilon_g^{(\xi)},\\
\psi_{\boldsymbol{\delta}}(\boldsymbol{Z}_g) &= R^{-1}\tilde{\boldsymbol{D}}_g'\varepsilon_g^{(\delta)}.
\end{aligned}$$

Because $\boldsymbol{\omega} = \boldsymbol{M}\boldsymbol{\zeta}$, we have $\psi_{\boldsymbol{\omega}} = \boldsymbol{M}\psi_{\boldsymbol{\zeta}}$ by delta method. Next, let $\boldsymbol{Z}_g \sim F$. When an estimand ψ is considered as a functional on the space of distributions that maps $F \mapsto \psi(F)$, then the influence function of that estimand is obtained by the Gáteaux derivative of the functional at the true distribution of data, on the direction to dirac measure to a specific observation (e.g., Ichimura and Newey (2022) Chernozhukov et al. (2022)), i.e., $\psi_\phi = \phi'(F)(\delta_{\boldsymbol{Z}_g} - F)$. Thus, the influence function for $\boldsymbol{H}$ is the matrix

valued function in which each element is replaced by its influence function. Also, note that we have $\boldsymbol{H\theta} = \boldsymbol{\delta}$, and this holds for all underlying distribution of data by construction. Thus, it follows from the chain rule of Gáteaux derivative that

$$\psi_{\boldsymbol{H}}\boldsymbol{\theta}_0 + \boldsymbol{H}_0\psi_{\boldsymbol{\theta}} = \psi_{\boldsymbol{\delta}},$$

and hence if $\boldsymbol{H}_0$ is invertible, we have the influence function of $\boldsymbol{\theta}$ as

$$\psi_{\boldsymbol{\theta}} = \boldsymbol{H}_0^{-1}(\psi_{\boldsymbol{\delta}} - \psi_{\boldsymbol{H}}\boldsymbol{\theta}_0).$$

By the same argument, we have the influence function for $\boldsymbol{\pi}$ as in the Proposition 4. □

Proof. Note that both $\boldsymbol{Q}, \boldsymbol{R}$ are nonsingular:

$$\boldsymbol{Q} = E\begin{bmatrix} 1 & D_i & D_j & D_iD_j \\ D_i & D_i & D_iD_j & D_iD_j \\ D_j & D_iD_j & D_j & D_iD_j \\ D_iD_j & D_iD_j & D_iD_j & D_iD_j \end{bmatrix},$$

$$\boldsymbol{R} = E\begin{bmatrix} 1 & D_i & \sum_{j\neq i} D_j & D_i\sum_{j\neq i} D_j \\ D_i & D_i & D_i\sum_{j\neq i} D_j & D_i\sum_{j\neq i} D_j \\ \sum_{j\neq i} D_j & D_i\sum_{j\neq i} D_j & \left(\sum_{j\neq i} D_j\right)^2 & D_i\left(\sum_{j\neq i} D_j\right)^2 \\ D_i\sum_{j\neq i} D_j & D_i\sum_{j\neq i} D_j & D_i\left(\sum_{j\neq i} D_j\right)^2 & D_i\left(\sum_{j\neq i} D_j\right)^2 \end{bmatrix}.$$

□

B Tables

Table 3: MSE, Variance, Bias for each parameters for different group size and $N = 30$

			1st stage			**2nd stage**		
Cat	N	G	θ_1	θ_2	θ_3	β	γ_1	γ_2
Mean	30	100	2.087	0.258	0.281	0.72	0.83	0.24
	30	500	2.147	0.261	0.276	1.024	0.797	0.196
	30	1000	2.144	0.26	0.277	1.014	0.799	0.198
	30	5000	2.15	0.261	0.276	1.034	0.796	0.195
True Values			0.214	0.26	0.277	1	0.8	0.2
MSE	30	100	0.126	<1e-5	0.001	2.898	0.031	0.057
	30	500	0.024	<1e-5	<1e-5	0.506	0.006	0.01
	30	1000	0.01	<1e-5	<1e-5	0.249	0.003	0.005
	30	5000	0.003	<1e-5	<1e-5	0.073	0.001	0.001
Variance	30	100	0.123	<1e-5	0.001	2.82	0.03	0.056
	30	500	0.024	<1e-5	<1e-5	0.506	0.006	0.01
	30	1000	0.01	<1e-5	<1e-5	0.249	0.003	0.005
	30	5000	0.003	<1e-5	<1e-5	0.072	0.001	0.001
Bias	30	100	-0.056	-0.002	0.004	-0.28	0.03	0.04
	30	500	0.003	<1e-5	<1e-5	0.024	-0.003	-0.004
	30	1000	0.001	<1e-5	<1e-5	0.014	-0.001	-0.002
	30	5000	0.007	<1e-5	<1e-5	0.034	-0.004	-0.005

notes. For each sample size N, G, the number of simulations is set to $B = 100$.

Table 4: MSE, Variance, Bias for each parameters for different group size and $N = 50$

			1st stage			**2nd stage**		
Cat	N	G	θ_1	θ_2	θ_3	β	γ_1	γ_2
Mean	50	100	2.927	0.26	0.277	0.952	0.803	0.204
	50	500	2.929	0.26	0.277	0.96	0.802	0.204
	50	1000	2.923	0.26	0.277	0.977	0.802	0.201
	50	5000	2.931	0.26	0.277	0.996	0.8	0.2
True Values			2.932	0.26	0.277	1	0.8	0.2
MSE	50	100	0.139	<1e-5	<1e-5	3.463	0.013	0.024
	50	500	0.029	<1e-5	<1e-5	0.624	0.002	0.004
	50	1000	0.014	<1e-5	<1e-5	0.304	0.001	0.002
	50	5000	0.004	<1e-5	<1e-5	0.091	<1e-5	0.001
Variance	50	100	0.139	<1e-5	<1e-5	3.461	0.013	0.024
	50	500	0.029	<1e-5	<1e-5	0.623	0.002	0.004
	50	1000	0.013	<1e-5	<1e-5	0.304	0.001	0.002
	50	5000	0.004	<1e-5	<1e-5	0.091	<1e-5	0.001
Bias	50	100	-0.005	<1e-5	<1e-5	-0.048	0.003	0.004
	50	500	-0.003	-0.001	<1e-5	-0.04	0.002	0.004
	50	1000	-0.009	<1e-5	<1e-5	-0.023	0.002	0.001
	50	5000	-0.002	<1e-5	<1e-5	-0.004	<1e-5	<1e-5

notes. For each sample size N, G, the number of simulations is set to $B = 100$.