A Semiparametric Network Formation Model with Multiple Linear Fixed Effects*

Luis E. Candelaria
University of Warwick
L.Candelaria@warwick.ac.uk

September 1st, 2017

Abstract

This paper analyzes a semiparametric model of network formation in the presence of multiple, unobserved, and agent-specific fixed effects. Given agents' observed attributes, the conditional distributions of these effects, as well as the disturbance terms associated with each linking decision are not parametrically specified. I give sufficient conditions for point identification of the coefficients on the observed covariates. This result relies on the existence of at least one continuous covariate with unbounded support. I provide partial identification results when all covariates have a bounded support. Specifically, I derive bounds for each component of the vector of parameters when all the covariates have a discrete support. I propose a semiparametric estimator for the vector of coefficients that is consistent and asymptotically normal as the number of individuals in the network increases. Monte Carlo experiments demonstrate that the estimator performs well in finite samples. Finally, in an empirical study, I analyze the determinants of a friendship network using the Add Health dataset.

JEL classification: C14; C25; C31; D85
Keywords: Network formation, semiparametric models, multiple fixed effects, irregular identification.

*First version: November 7th, 2016. This version: September 1st, 2017. I am deeply grateful to my co-advisors Shakeeb Khan and Arnaud Maurel, as well as Federico Bugni and Matthew Masten for their excellent guidance, constant encouragement, and helpful discussions. I also thank Rachel Kranton, Jia Li, Fu Ouyang, Adam Rosen, Takuya Ura, Yichong Zhang and seminar participants at Duke Microeconometrics Workshop for their comments. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. This research uses data from Add Health. See references for full citation. All errors are my own.

1 Introduction

People tend to connect with individuals whom they share similar observed attributes. This observation is known as homophily (McPherson, Smith-Lovin, and Cook 2001). Nonetheless, few investigate the role of homophily when individuals have preferences for unobserved attributes. Proper policy evaluation requires distinguishing among the contribution of these two factors since each has a distinct policy implication. For example, students forming friendships might link based on their similarities on observed socioeconomic attributes as well as on their preferences for high levels of unobserved ability. Whereas the socioeconomic attributes can be modified according to a policy, preferences for ability cannot be used as a policy instrument. In this paper, I develop a new identification strategy that recovers the preference parameters associated with the observed attributes in a model of network formation that accounts for valuations on unobserved agent-specific factors. The identification and estimation strategies that I develop do not depend on distributional assumptions of the unobserved random components. The existing studies that account for these two types of homophily rely on assumptions that restrict the distribution of the unobserved random components to belong to a parametric family. However, in Monte Carlo simulations, I show that their predictions can be biased if those assumptions fail.

In this paper, I consider a semiparametric model of network formation with multiple, unobserved and agent-specific factors. Specifically, a pair of agents (i, j) establish an undirected link according to the following network formation equation:[1]

$$D_{ij}^n = \mathbf{1}\left[X_{ij}^{n\prime}\beta_0 + \mu_i + \mu_j - \varepsilon_{ij}^n \geq 0\right], \tag{1}$$

where $\mathbf{1}[\cdot]$ is the indicator function, D_{ij}^n is a binary outcome variable that takes a value equal to 1 if agents (i, j) form a link and 0 otherwise, X_{ij}^n is a K-dimensional vector of pair-specific, observed, and exogenous attributes, β_0 is a K-dimensional vector of unknown parameters, μ_i and μ_j are unobserved and agent-specific random variables, and ε_{ij}^n is an unobserved and link-specific disturbance term.

Intuitively, equation (1) says that an undirected link between two agents is formed if the link net benefit is nonnegative.[2] The factors in the net benefit can be classified into three different

[1] A link between two agents is undirected if the connection is reciprocal. In other words, two agents are either connected or they are not. It excludes the case that one agent is related to a second one without the second being related to the first. I discuss directed networks in Candelaria (2016).

[2] In section 2, I derive the network formation decision in equation (1) as a stability condition in a random utility model with transferable utilities.

categories. The first class, given by the vector of pair-specific and exogenous attributes, captures the agents' preferences for establishing a link based on observed characteristics. For instance, this component is known as homophily in preferences when these factors capture similarity in observed characteristics. The second class, formed by the agent-specific and unobserved factors, captures the individual preferences for association based on agent-specific attributes. Finally, the third class, given by a link-specific disturbance term, captures the exogenous factors that influence the decision of forming a specific link. The last two factors are known to the agents but unobserved to the researcher.

The unobserved agent-specific factors in equation (1) allow for heterogeneous net benefits across each individual's decisions; this extends the model's capacity to predict network structures with heterogeneous individual connections. Moreover, under an unrestricted distribution of the unobserved and agent-specific factors, these components exhibit unrestricted dependence with the observed attributes. Therefore, these factors constitute agent-specific fixed effects in the network formation model. From here after these factors will be referred to as fixed effects.

This paper has two main contributions. The first contribution is to propose a new identification strategy to identify the coefficients on the observed covariates in a semiparametric network formation model with multiple fixed effects. These coefficients are empirically relevant, for example in the peer effects literature they characterize the preferences for homophily. Notably, these coefficients provide information about policy instruments that can be used to achieve an economic outcome. I provide sufficient conditions that guarantee point identification of the parameter β_0 in equation (1). Using a weaker set of assumptions, I characterize the identified set as a solution to a system of a finite number of linear inequalities and provide bounds for each component of the parameter of interest.

The second contribution is to introduce a consistent semiparametric estimator of β_0. I give conditions for asymptotic normality of this estimator. The rate of convergence of the estimator can be affected by the asymptotic probability of the set on which the unknown parameter is identified. Specifically, the convergence rate is slower than the parametric rate (square root of the sample size) if the probability of that set converges to zero. I perform inference in a setting when only one network that has a large number of agents is observed in the data. The asymptotic analysis is conducted by allowing the number of agents to grow. This framework is referred to as "large-market" asymptotics.

While it is not clear whether identification strategies based on parametric assumptions used in previous work extend to a model with a broader and more complex type of heterogeneity, my approach does. In Candelaria (2016), I study the formation of a directed network with interactive fixed effects. Specifically, agent i establishes a *directed* link with agent j according to the following equation:

$$D_{ij}^n = \mathbf{1}\left[X_{ij}^{n\prime}\beta_0 + \mu_i + g(\mu_i, \mu_j) - \varepsilon_{ij}^n \geq 0\right], \tag{2}$$

where $g(\cdot,\cdot)$ is a symmetric function of the unobserved fixed effects μ_i, μ_j.

Specification (2) considers an asymmetric formation of links and allows for simultaneous nonlinear correlation between the unobserved factors and the observed attributes. Furthermore, the agent-specific fixed effect μ_j may affect the linking decisions of individual i, differently for different j due to the unobserved complementarities on the fixed effects. Equation (2) nests the additive and agent-specific fixed effects model as a special case. Specifically, equation (2) degenerates to equation (1) when $g(\mu_i, \mu_j) = \mu_j$, and ε_{ij}^n is symmetric. In Candelaria (2016), I show that a generalization of the identification strategy, introduced in this paper, can be used to identify the coefficients β_0 in equation (2). This demonstrates the adaptability of the techniques discussed in this paper.

In an empirical application, I study the determinants that drive the formation of a friendship network. I use the National Longitudinal Study of Adolescent Health (Add Health) to construct a network of best friends using one high school with 469 students. The vector X_{ij}^n accounts for socioeconomic and demographic attributes of individuals i and j, such as gender, education level, race. The first attribute in X_{ij}^n is the household income of individuals i and j, which is recorded as a continuous variable and is necessary for the point identification result. I then estimate the parameter β_0 and find evidence for homophily in observed attributes in a model that also accounts for unobserved heterogeneity.

Literature Review

In the rest of the section, I discuss how my results compare to the related literature. The network formation model that I consider in this paper builds on the framework introduced by Graham (2017). While his paper aims to detect homophily in preferences in a model with agent heterogeneity, his approach is restricted to models where the disturbance terms have a parametric distribution — specifically a logistic distribution — and the fixed effects have an additive structure. In other words, the approach introduced by Graham (2017) will fail to partial out the fixed effects and therefore point identify β_0 if one of these assumptions does not hold. This is not the case for the method I develop in this paper. Specifically, the identification strategy and the estimator I propose are new and can be applied to models where the distribution of the disturbance terms is not parametrically specified and the heterogeneity does not follow an additive structure (in the extension work Candelaria 2016). In recent work, Dzemski (2014) studies a model of link formation with agent heterogeneity. However, his methodology differs completely to the one proposed in this paper since he analyzes the formation of a directed network and follows a conditional maximum likelihood approach.

My identification strategy consists of finding a sufficient statistic for the multiple fixed effects in equation (1), which does not depend on the parametric distribution of the disturbance terms (Andersen, 1970). The intuition behind this strategy is similar to the technique used by the maximum score estimator to identify the semiparametric binary choice models in a panel data framework (Manski 1987). Specifically, the sufficient statistic that I develop is characterized by within-individual and across-individuals variation in the link decisions to differentiate out the multiple fixed effects. This statistic differs from others previously used in the nonlinear panel data literature since the endogeneity entailed by the multiple fixed effects in the network formation equation is more complex. In section 3.1, I provide a more detailed discussion on the nature of the sufficient static. Moreover, I show that the sufficient statistic suggested by the panel data literature fails to identify β_0 in equation (1).

The sufficient statistic restricts the analysis to a set of subnetworks that exhibit sufficient link variation to differentiate out the multiple fixed effects. Depending on the relative tail conditions between the observed attributes and disturbance terms, the set of subnetworks consistent with the link variation might have a small probability. In this case, the coefficients of the observed attributes are said to be identified on a thin set. In section 4, I address the implications of the thin set identification on the convergence rate of the estimator (Andrews and Schafgans, 1998; Newey, 1990; Chamberlain, 2010 and Khan and Tamer, 2010). To shed some light into this point, I next briefly discuss the estimator that I develop.

I propose an M-estimator that minimizes a fourth order U-statistic. The estimator falls within the class of Maximum Rank estimators, which are commonly used to estimate monotonic transformation models (Han 1987). The estimator I propose is most closely related to the Leapfrog estimator in Abrevaya (1999b). The network formation model, given by equation (1), represents a weakly monotonic transformation model. This type of transformation is not nested in the models analyzed by Abrevaya (1999b) since his methodology is designed for transformation functions that are strictly increasing and invertible. This is the first paper to apply an estimator within this class to a network structure with multiple fixed effects. If β_0 is identified in a set with probability tending to zero, the convergence rate of the estimator is slower than the parametric rate (square root of the sample size). Hence, the estimator is said to be non-regular (Newey, 1990 and Chamberlain, 2010). I propose an inference method with an adaptive convergence rate as in Andrews and Schafgans (1998) and Khan and Tamer (2010). A detailed discussion of this result is provided in section 4.

The network formation model that I analyze is related to the empirical games literature. Specifically, the model in equation (1) can be derived as a stability condition in a static game. Some papers that study the strategic formation of a network as a static game include Sheng (2012); Goldsmith-Pinkham and Imbens (2013); Boucher and Mourifié (2013); Leung (2015a,b); Menzel (2015); Miyauchi (2016) and de Paula, Richards-Shubik, and Tamer (2016). These papers study network formation models that account for network externalities. Network externalities generate interdependencies in the linking decisions that depend on the structure of the network. The identification and estimation methods used in these papers are entirely different to the ones proposed in this paper. Specifically, all of these papers follow a parametric estimation approach. The only exception is de Paula et al. (2016), which focuses exclusively on the identification analysis. Furthermore, only Goldsmith-Pinkham and Imbens (2013) considers unobserved agent heterogeneity, but under their specification the agent-specific effects are parametrically distributed and independent from the vector of attributes. None of these assumptions are imposed in the method proposed in this paper.

There is a different approach that augments the network formation decision in equation (1) with a parametric meeting process that determines how the links are sequentially revised over time (Christakis, Fowler, Imbens, and Kalyanaraman 2010; Snijders, Koskinen, and Schweinberger 2010; Hsieh and Lee 2012; Chandrasekhar and Jackson 2014; Mele 2015 and Badev, 2014). This approach specifies a parametric distribution over the space of all potential networks and differs from distribution-free framework that is followed in this paper. Furthermore, some of these papers rely on computationally intensive Bayesian estimation techniques such as Markov Chain Monte Carlo (MCMC). The semiparametric estimator that I proposed is computationally tractable and can be computed in $O(n^3 log(n))$ calculations. This paper provides an alternative to recover the preferences for homophily in the formation of a network.

Finally, the network formation model considered in this paper is also related to the literature of structural matching models analyzed by Choo and Siow (2006); Fox (2010); Galichon and Salanié (2012) and Fox (2016). As a shared feature, both frameworks focus on transferable utility models. However, the network formation model is qualitatively different to the two-sided matching models since in a network model any pair of individuals can potentially form a link.

In contrast, in a two-sided matching model, only agents across markets can establish a link.

The rest of the paper is organized as follows. Section 2 formalizes the network formation model. Section 3 describes the identification strategy and states point identification and partial identification results. In section 4, I outline the semiparametric estimator and I show consistency and asymptotic normality. Section 5 reports some Monte Carlo simulations. Section 6 considers an empirical application analyzing a friendship network. Section 7 concludes by summarizing and suggesting areas for future research. The appendix B collects all the proofs of the paper.

2 Model

A network is an ordered pair $(\mathcal{N}_n, D^n)$ comprising a set $\mathcal{N}_n = \{1, \cdots, n\}$ of n nodes or agents together with an $n \times n$ adjacency matrix D^n of edges, which represents the links between the nodes in $\mathcal{N}_n$. Let D^n_{ij} denote the (i, j)th entry of the matrix D^n.

I assume the network is undirected and unweighted. A network is undirected if the adjacency matrix is symmetric, that is, if for any entries (i, j) and (j, i) the adjacency matrix has identical elements, $D^n_{ij} = D^n_{ji}$. A network is unweighted if any entry (i, j) of the adjacency matrix takes one of either two values. The values are normalized to be 0 and 1. In other words, $D^n_{ij} \in \{0, 1\}$, where $D^n_{ij} = 1$ if the agents i and j share a link and $D^n_{ij} = 0$ otherwise. Furthermore, I normalize the value of self-ties to zero, that is, $D^n_{ii} = 0$ for all $i \in \mathcal{N}_n$.

Example 1 (Undirected and Unweighted Network)
A friendships network of best friends is an important example of an undirected and unweighted network. Two agents are considered to be best friends, $D^n_{ij} = 1$, if and only if both agents list each other as friends. Also, this framework rules out the case of an agent reporting herself as a friend.

Given the set of agents in the network, a pair of agents (i, j) with $i, j \in \mathcal{N}_n$ and $i \neq j$, constitute a dyad. Let $\mathcal{N}^{(2)}_n \equiv \{(1, 2), \cdots\cdots, (n-1, n)\}$ denote the set of total unique dyads. $\mathcal{N}^{(2)}_n$ has cardinality

$$N \equiv \binom{n}{2} = O(n^2).$$

Each dyad $(i, j) \in \mathcal{N}^{(2)}_n$ is endowed with a $(K+1)$-dimensional vector of observed attributes $Z^n_{ij} = (D^n_{ij}, X^n_{ij})$, and an unobserved dyad-specific disturbance term ε^n_{ij}. The first element in the vector of observed attributes Z_{ij}^n denotes the link status in that specific dyad, D_{ij}^n. The second element in Z_{ij}^n is a vector of observed exogenous attributes at a dyad-level, $X_{ij}^n \in \mathbb{R}^K$. Common examples of observed attributes used to explain the formation of a friendships network among high school students are age, gender, race, parent's education level, and household's income (I provide a detailed discussion of this empirical motivation in section 6). Conditions on the support of the vector of exogenous attributes X_{ij}^n are discussed in section 3. The unobserved dyad-specific disturbance component ε_{ij}^n captures exogenous random factors that influence the decision of establishing a connection between agents i and j. These components are unobserved to the researcher.

Since the network is undirected, the random vector X_{ij}^n is symmetric, $X_{ij}^n = X_{ji}^n$. If the exogenous characteristics are measured at an agent-level, the dyad-level vector X_{ij}^n can be constructed by transforming the agent-specific covariates for agents i and j using a nonlinear function that is symmetric in each of its components. For instance, let X_i^n represent a vector of exogenous attributes of agent i. Then X_{ij}^n could be defined as $X_{ij}^n = g(X_i^n, X_j^n) = g(X_j^n, X_i^n)$. Different specifications of g can be used to capture similarity ($g(X_j^n, X_i^n) = (X_i^n - X_j^n)^2$) or complementarity ($g(X_j^n, X_i^n) = X_i^n \cdot X_j^n$) in attributes between agents i and j in dyad (i, j). The choice of $g(\cdot, \cdot)$ varies according to the empirical application.

Each individual i in the network $\mathcal{N}_n$ is endowed with an unobserved and agent-specific random factor $\mu_i \in \mathbb{R}$. This random component captures the individual preferences for establishing a link based on agent-specific attributes.

Let $\mathbf{X}^n \equiv \left(X_{12}^n, \cdots, X_{n-1,n}^n\right)$ be the profile of exogenous attributes for all dyads in the network, $\tilde{\mu} \equiv (\mu_1, \cdots, \mu_n)$ be the vector of unobserved agent-specific components and $\varepsilon^n \equiv \left(\varepsilon_{12}^n, \cdots, \varepsilon_{n-1,n}^n\right)$ be the profile of dyad-specific disturbance terms.

Agent i's latent marginal benefit of establishing a link with j is

$$V_{ij}(\mathbf{X}^n, \mu_j, \varepsilon_{ij}^n) = u_{ij}(\mathbf{X}^n) + \mu_j - \frac{1}{2}\varepsilon_{ij}^n, \tag{3}$$

where $u_{ij}(\mathbf{X}^n)$ denotes the observed marginal utility, and ε_{ij}^n is symmetric. Specifically, the observed marginal utility is defined as:

$$u_{ij}(\mathbf{X}^n) \equiv \frac{1}{2} X_{ij}^{n\prime} \beta_0, \tag{4}$$

where β_0 is a K-dimensional vector of unknown parameters that captures the effect of the observed attributes on the agent's preferences for establishing a link. This component represents the agent's preferences for homophily, $X_{ij}^{n\prime}\beta_0$.

Denote the joint net benefit of adding the link $\{ij\}$ to the network D^n by

$$V_{ij}(\mathbf{X}^n, \mu_j, \varepsilon_{ij}^n) + V_{ji}(\mathbf{X}^n, \mu_i, \varepsilon_{ij}^n) \equiv X_{ij}^{n\prime}\beta_0 + \mu_i + \mu_j - \varepsilon_{ij}^n. \tag{5}$$

In addition to preferences for observed attributes, the joint net benefit also accounts for preferences for association based on agent-specific factors, $\mu_i + \mu_j$, and for exogenous factors affecting the decision of establishing a link ε_{ij}^n.

Equation (5) implies that individuals i and j in the dyad (i, j) only have valuations for their own observed attributes and agent-specific factors. To clarify, in the link formation decision for dyad (i, j), the individuals do not account for observed or unobserved attributes of other individual's in the network, neither for the general structure of the network other than dyad (i, j). These effects are known as network externalities. Some examples of these effects are preferences for having friends in common or popularity effects. I leave this extension as future research.

Next, I introduce the definition of stability.

Definition 1 (Stability)
A network D^n is stable with transfers if for any $i, j \in \mathcal{N}_n$:

1. *for all $D_{ij}^n = 1$, $V_{ij}(\mathbf{X}^n, \mu_j, \varepsilon_{ij}^n) + V_{ji}(\mathbf{X}^n, \mu_i, \varepsilon_{ij}^n) \geq 0$;*
2. *for all $D_{ij}^n = 0$, $V_{ij}(\mathbf{X}^n, \mu_j, \varepsilon_{ij}^n) + V_{ji}(\mathbf{X}^n, \mu_i, \varepsilon_{ij}^n) < 0$.*

Note, that the definition of stability adapts the pairwise stability in Jackson and Wolinsky (1996) to allow for transfer utilities. A similar stability concept has been used in Sheng (2012). The stability condition provides a microeconomic foundation to the network formation rule in equation (1). Intuitively, this condition states that a link within dyad (i, j) is established if the net benefit of that connection is nonnegative.

To simplify notation, I will omit the dependence of the network on the sample size n and denote the vector of attributes as $Z_{ij} = (D_{ij}, X_{ij})$ and the dyad-specific disturbance term as ε_{ij} for any $(i, j) \in \mathcal{N}_n^{(2)}$.

3 Identification

In this section, I state the main point identification result for the semiparametric network formation model with multiple agent-specific factors, specified by equation (1). I then provide partial identification results under a weaker set of assumptions. In section 3.1, I describe the identification strategy. Section 3.2 establishes the main point identification result, which is achieved by conditioning on a set that ensures enough variation within and across individuals' links. In section 3.3, I discuss identification failure when the probability of this set is zero. Moreover, I show that, unlike my approach, the typical identification strategy implied by the panel data maximum score estimator fails to identify β_0. In section 3.4, I characterize the identified set when all the covariates have bounded support, as well as provide bounds for each component of the parameter of interest.

3.1 Identification Strategy

The intuition behind the identification strategy is summarized in figure 1. Consider the subnetwork formed between agents $i, j, k, l \in \mathcal{N}_n$. All the links represented in figure 1 are undirected. A solid line connecting two agents denotes that a link exists and a dashed line denotes that a link is absent. To simplify the intuition, in both diagrams below I omit the link status between the agents in the dyad (k, l). The realized outcome from that decision is non-informative for describing the intuition of the identification strategy.

Diagram 1 represents the subnetwork formed by dyads (i, k) and (i, l). Given $\{\mathbf{X}^n = x, \tilde{\mu} = \mu\}$, suppose that the conditional probabilities of establishing a link between dyads (i, k) and (i, l) are different. Without loss of generality, assume that:

$$\mathbb{P}\left[D_{il} = 1 \mid \mathbf{X}^n = x, \tilde{\mu} = \mu\right] < \mathbb{P}\left[D_{ik} = 1 \mid \mathbf{X}^n = x, \tilde{\mu} = \mu\right]. \tag{6}$$

If the dyad-specific unobserved random variables are identically distributed, then the equation (6) holds if and only if:

$$x'_{il}\beta_0 + \mu_i + \mu_l < x'_{ik}\beta_0 + \mu_i + \mu_k,$$

where agent i's individual-specific fixed effect μ_i is a common element. Therefore, the within-

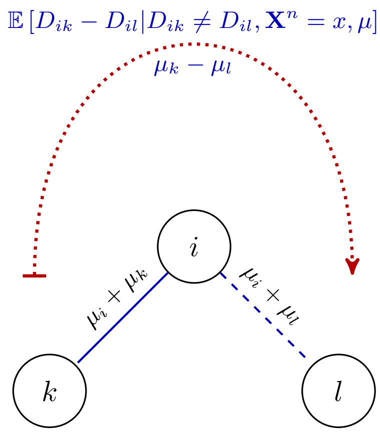


Diagram 1: Undirected links in dyads (i,k) and (i,l).

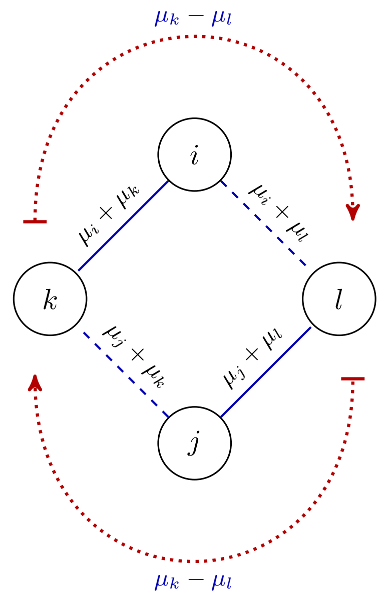


Diagram 2: Undirected links in tetrad (i,j,k,l).

Figure 1: Subnetwork formed by agents $i, j, k, l \in \mathcal{N}_n$.

individual difference implies:

$$0 < (x_{ik} - x_{il})'\beta_0 + (\mu_k - \mu_l).$$

The previous intuition suggests that for any individuals $i, j, l \in \mathcal{N}_n$: the conditional expectation of the within-individual difference $D_{ik} - D_{il}$ is characterized by the difference of the observed exogenous regressors, $(x_{ik} - x_{il})'\beta_0$, and the difference of the unobserved factors, $\mu_k - \mu_l$. Agent i's individual-specific factor is differenced out by computing the net difference. In diagram 1, the dotted line labeled as $\{\mathbb{E}\left[D_{ik} - D_{il} | D_{ik} \neq D_{il}, \mathbf{X}^n = x, \mu\right]\}$ depicts this intuition. This line shows that the contribution of the unobserved agent-specific factors on the conditional expectation of $D_{ik} - D_{il}$ is characterized exclusively by the composite factor $\mu_k - \mu_l$, and not agent i's individual-specific factor μ_i.

Specifically, the following equation holds for the conditional median of the net difference $D_{ik} - D_{il}$,

$$\text{Med}(D_{ik} - D_{il} | \mathbf{X}^n = \mathbf{x}, D_{il} \neq D_{ik}) = \text{sign}\left((x_{ik} - x_{il})'\beta_0 + (\mu_k - \mu_l)\right), \tag{7}$$

where sign$(\cdot)$ stands for the sign function, which is defined as sign$(x) = 1$ if $x \geq 0$ and sign$(x) = -1$ if $x < 0$ for any $x \in \mathbb{R}$. The proof is in the appendix B.

Equation (7) conveys two main points. First, agent i's individual-specific fixed effect is differenced out by conditioning on observing within-individual variation in the realized links. Second, the conditional median of the net difference $D_{ik} - D_{il}$ depends on the unobserved random factor $\mu_k - \mu_l$ due to the presence of multiple agent-specific fixed effects in the network formation model (1). Therefore, equation (7) does not identify β_0. In other words, the typical maximum score identification strategy fails to point identify β_0. I provide a more detailed explanation of this result in section 3.3.2.

The point-identification argument in the network formation model with multiple fixed effects is the following. Consider the links formed within the tetrad (i, j, k, l). Given $\{\mathbf{X}^n = x, \tilde{\mu} = \mu\}$, suppose that the conditional probability of establishing a link between dyad (j, l) is greater than the one for dyad (j, k). That is,

$$\mathbb{P}\left[D_{jk} = 1 \mid \mathbf{X}^n = x, \tilde{\mu} = \mu\right] < \mathbb{P}\left[D_{jl} = 1 \mid \mathbf{X}^n = x, \tilde{\mu} = \mu\right].$$

Analogously to above, the conditional expectation of the net difference between individual j's linking decisions $D_{jk} - D_{jl}$ is characterized by the difference of the observed exogenous regressors $(x_{jk} - x_{jl})'\beta_0$ and the difference of the unobserved factors $\mu_k - \mu_l$. The composite unobserved factor $\mu_k - \mu_l$ constitutes a common, unobserved fixed effect across the within-individual variations for agents i and j. Diagram 2 illustrates this point.

The previous intuition suggests that, with enough across-individuals variation, the composite fixed effect $\mu_k - \mu_l$ can be differenced out by computing the across-individuals difference of $D_{ik} - D_{il}$ and $D_{jk} - D_{jl}$. Specifically, in section 3.2, I show that the following equation for the conditional median of the pairwise difference holds:

$$\begin{aligned}\text{Med}\left\{[D_{ik} - D_{il}] - [D_{jk} - D_{jl}] \mid \mathbf{X}^n = \mathbf{x}, D_{ik} \neq D_{il}, D_{jk} \neq D_{jl}, D_{ik} \neq D_{jk}\right\} = \\ 2 \times \text{sign}\left\{\left[(x_{ik} - x_{il}) - (x_{jk} - x_{jl})\right]' \beta_0\right\},\end{aligned} \tag{8}$$

for any $\mathbf{X}^n = \mathbf{x}$ in a set of sufficient variation that will be defined below.

Equation (8) is fully characterized by the observed variables $(D^n, \mathbf{X}^n)$. In sections 3.2 and 3.4, I show that equation (8) can be used to point identify β_0 under support conditions on the exogenous attributes.

The conditioning event $\{D_{ik} \neq D_{il}, D_{jk} \neq D_{jl}, D_{ik} \neq D_{jk}\}$ in equation (8) ensures the sufficient within-individual and across-individuals variation in the linking decisions to identify β_0. The intuition behind the conditioning event is as follows. The components $D_{ik} \neq D_{il}, D_{jk} \neq D_{jl}$ capture the within-individual variation, which are used to partial out the individual-specific fixed effects that are constant within each individual's decisions. For example, for agent i the individual heterogeneity μ_i is partial out by the within-individual difference $D_{ik} - D_{il}$.

Second, the component $D_{ik} \neq D_{jk}$ captures the across-individuals i and j variation.[3] This variation is used to partial out the composite and unobserved factor $\mu_k - \mu_l$. Therefore, this condition is crucial to point identify β_0. I show in subsection 3.3.1 that if the event $\{D_{ik} \neq D_{il}, D_{jl} \neq D_{jk}, D_{ik} \neq D_{jk}\}$ has probability zero then equation (8) is uninformative to identify β_0.

Some empirical examples of network topologies for which the event $\{D_{ik} \neq D_{il}, D_{jl} \neq D_{jk}, D_{ik} \neq D_{jk}\}$ has probability zero are close to empty networks, dense networks, and homogeneous networks. A network is homogeneous when the individuals establish similar connections with probability one. For example, in the network formation model given by equation (1), the equilibrium network structure will be homogeneous when agents establish their connections based mainly only on their preferences for individual-specific attributes.

3.2 Formal Point Identification Result

In this section, I formalize the main point identification result. The following set of assumptions are sufficient to prove point identification of β_0.

Assumption A1

The following hold for any n.

1. $\{\varepsilon_{ik}\}_{(i,k)\in\mathcal{N}_n^{(2)}}$ *are independent and identically distributed (i.i.d.) conditional on* $\{\mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu\}$. *That is for any* $(i,k),(j,l) \in \mathcal{N}_n^{(2)}$:

$$\varepsilon_{ik} \perp\!\!\!\perp \varepsilon_{jl} \mid \mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu, \quad \textit{and} \quad F_{\varepsilon_{ik}|\mathbf{x},\mu} = F_{\varepsilon_{jl}|\mathbf{x},\mu}.$$

2. *The probability density function* $f_{\varepsilon_{i1}|\mathbf{x},\mu}$ *is positive everywhere on* $\mathbb{R}^1$ *for all* $(\mathbf{x},\mu)$.

Here $F_{\varepsilon_{i1}|\mathbf{x},\mu}$ denotes the conditional distribution of ε_{i1} given $\{\mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu\}$.

[3] Equivalently, we could have considered $D_{il} \neq D_{jl}$. The information content in each of these two events, jointly with $\{D_{ik} \neq D_{il}, D_{jk} \neq D_{jl}\}$, is identical. Therefore, conditioning on $D_{ik} \neq D_{jk}$ is sufficient.

A1.1 states that the disturbance terms are i.i.d. across dyads. In other words, for any pair of dyads (i,k) and (j,l), the distributions of the disturbance terms in the network formation equations that are indexed by those dyads are conditionally invariant and independent. A1.1 is analogous to the standard "stationarity" assumption in panel data models because in a network model with a symmetric adjacency matrix, the dyads are the unit of observation which makes it irrelevant to focus on individual's labels.[4]

Although A1.1 requires the regressors to be strongly exogenous with respect to the disturbances, this specification allows for a flexible dependence structure between the unobserved agent-specific factors and the observed attributes. Specifically, the conditional distribution of the unobserved agent-specific factors $F_{\tilde{\mu}|\mathbf{x}}$ given the observed attributes $\mathbf{X}^n = \mathbf{x}$ is not assumed to belong to any parametric family. Consequently, the presence of the unobserved fixed effects in the network formation model generates a multiple incidental parameter problem with unobserved dependence across the dyads' linking decisions.

A1.2 requires the disturbance terms to have a large support given $\{\mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu\}$. Given any specification $\{\mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu\}$, A1.2 ensures that the event $\{D_{ik} \neq D_{il}\}$ happens with positive positive for any dyads $(i,k),(i,l) \in \mathcal{N}_n^{(2)}$. In other words, assumption A1.2 guarantees the existence of within-individual variation in the outcome linking decisions.

Assumption A1 is commonly used in semiparametric nonlinear panel data models, for example in Manski (1987); Han (1987); Abrevaya (1999b) and Arellano and Honoré (2001), as well as in network formation models, such as in Graham (2017); Leung (2015a) and Menzel (2015).

Let $\Delta_{kl} X_i \equiv X_{ik} - X_{il}$ for any $i,l,k \in \mathcal{N}_n$.

Assumption A2

The following hold for any n, and any $i,l,k \in \mathcal{N}_n$, with $l \neq k$.

1. *The support of $\Delta_{kl} X_i$ is not contained in any proper linear subspace of $\mathbb{R}^K$.*
2. *There exists at least one component $\Delta_{kl} X_i^{(s)}$, $s \in \{1,\cdots,K\}$, with $\beta_{0,s} \neq 0$ such that for almost every $\Delta_{kl} x_i^{(-s)} = (\Delta_{kl} x_i^{(1)},\cdots,\Delta_{kl} x_i^{(s-1)},\Delta_{kl} x_i^{(s+1)},\cdots,\Delta_{kl} x_i^{(K)})$, the distribution of $\Delta_{kl} X_i^{(s)}$ conditional on $\Delta_{kl} X_i^{(-s)} = \Delta_{kl} x_i^{(-s)}$ has a positive density almost everywhere with respect to the Lebesgue measure.*

Without loss of generality, I set $\beta_{0,1} = 1$ or -1. This is a scale normalization used to identify

[4] In the nonlinear panel data literature, the disturbance term component is said to have a stationary distribution if $F_{\varepsilon_{ik}|\mathbf{x},\mu} = F_{\varepsilon_{ij}|\mathbf{x},\mu}$ for any $(i,k),(i,j) \in \mathcal{N}_n^{(2)}$. Due to the symmetry of the network, the disturbance term satisfies $\varepsilon_{ij} = \varepsilon_{ji}$, which jointly with the stationarity assumption of the distribution of ε_{ji} implies that $F_{\varepsilon_{ik}|\mathbf{x},\mu} = F_{\varepsilon_{jl}|\mathbf{x},\mu}$.

β_0 instead of the scaled parameter $\beta_0/||\beta_0||$. The normalization is without loss of generality, since the sign of $\beta_{0,1}$ is identified from the limits:

$$\lim_{x_{ik}^{(1)}\to\infty} \mathbb{P}\left[D_{ik}=1 \mid \mathbf{X}^n = x\right],$$
$$\lim_{x_{ik}^{(1)}\to-\infty} \mathbb{P}\left[D_{ik}=1 \mid \mathbf{X}^n = x\right].$$

If the $\text{sign}(\beta_{0,1}) = -1$, then $\beta_{0,1}$ can be normalized to -1.

A2.1 is a full rank condition for the exogenous attributes. A2.2 requires the observed covariates to have a large support, which implies that $\Delta_{kl}X_i'b$ has everywhere a positive density for any $b \in \mathbb{R}^K$ with $b_1 \neq 0$. The existence of at least one continuous covariate is a necessary condition for achieving point identification since it guarantees the existence of a subset in the support of $\Delta_{kl}X_i - \Delta_{kl}X_j$ with positive probability over which β_0 is identified from any $b \in \mathbb{R}^K$.

Conditions A2 is frequently used in semiparametric nonlinear panel data models, for example in Manski (1987); Han (1987) and Abrevaya (1999b), and in the literature of empirical games with strategic interactions, for example in Tamer (2003) and Kline (2015). In section 3.4, I give alternative sufficient conditions for point identification when regressors are continuous with bounded support.

Assumption A3
For any $i \in \mathcal{N}_n$,

$$\sup(\mu_i \mid X_{ij} = x) \subseteq [B_L, B_U],$$

for any $x \in \text{supp}(X_{ij})$, and given $B_L, B_U < \infty$.

A3 states that the agent-specific fixed effects have bounded support. Furthermore, this assumption allows for the distribution of the fixed effects to be heterogeneous across individuals, as long as common bounds for their support exist. A3 doesn't restrict the dependence between the fixed effects and the exogenous covariates.

Assumptions A2 and A3, guarantees that the within-individual variation in the observed attributes dominates the magnitude of the variation in the fixed effects. A similar condition has been used in weakly separable models with endogenous dummy variables, Vytlacil and Yildiz (2007).

The following theorem states the main point identification result. To simplify notation, consider the following definitions. For any distinct $i, j, l, k \in \mathcal{N}_n$, let:

$$\Omega(ijlk) \equiv \{D_{ik} \neq D_{il}, D_{jl} \neq D_{jk}, D_{ik} \neq D_{jk}\},$$

and

$$Y_{kl}^{(i)} \equiv (D_{ik} - D_{il}).$$

Theorem 3.1 *1. Let assumptions A1 - A3 hold. Then, for any n, and any $i, j, l, k \in \mathcal{N}_n$:*

$$\text{Med}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\right] = 2 \times \text{sign}\left\{[(x_{ik} - x_{il}) - (x_{jk} - x_{jl})]' \beta_0\right\}, \quad (7)$$

where $\mathbf{x} \in \mathcal{X}_B$, and

$$\mathcal{X}_B = \{\mathbf{x} \in \mathbf{X}^n : \textit{ for any } i, j, k, l \in \mathcal{N}_n, \mid \Delta_{kl} x_i \beta_0 \mid \geq (B_U - B_L), \textit{ and } \\ \text{sign}\{\Delta_{kl} x_i \beta_0\} + \text{sign}\{\Delta_{kl} x_i \beta_0\} = 0\}.$$

2. Let assumptions A1 - A3 hold. Then β_0 is point identified.

Equation (7) is fully characterized in terms of the observed variables $(D^n, \mathbf{X}^n)$ and it represents an identifying condition for β_0. This equation conveys two main points. First, the event $\Omega(ijlk)$ constitutes a sufficient statistic for the agent-specific factors in the conditional median of $Y_{kl}^{(i)} - Y_{kl}^{(j)}$. In other words, the conditional median of the pairwise-difference of the links given $\Omega(ijlk)$ is fully characterized by the pairwise-variation in the observed attributes. Second, the set $\Omega(ijlk)$ ensures sufficient within-individual and across-individuals variation to identify β_0 under the support conditions on the exogenous attributes.

Intuitively, equation (7) holds because conditional on $\{\mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\}$ the random variable $Y_{kl}^{(i)} - Y_{kl}^{(j)}$ has a Bernoulli distribution with support $\{-2, 2\}$. This statement follows from two results. First, the random variable

$$\left\{Y_{kl}^{(m)} \middle| \mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\right\} \quad \text{for} \quad m = i, j,$$

has a Bernoulli distribution with support $\{-1, 1\}$ due to the within-individual variation $D_{mk} \neq D_{ml}$ for $m = i, j$ implied by the set $\Omega(ijlk)$. Second, the across-individuals variation ensures that the following equivalences hold:

$$Y_{kl}^{(i)} = 1 \quad \Leftrightarrow \quad Y_{kl}^{(j)} = -1, \\ Y_{kl}^{(i)} = -1 \quad \Leftrightarrow \quad Y_{kl}^{(j)} = 1. \quad (9)$$

For instance, suppose $Y_{kl}^{(i)} = 1$, then it follows that conditional on $\{\mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\}$:

$$Y_{kl}^{(i)} = 1 \Leftrightarrow \{D_{ik} = 1, D_{il} = 0\} \Leftrightarrow \{D_{jk} = 0, D_{jl} = 1\} \Leftrightarrow Y_{kl}^{(j)} = -1.$$

The first equivalence follows from the definition of $Y_{kl}^{(i)}$ and the within-individual i variation $D_{ik} \neq D_{il}$. The second equivalence holds because of the across-individuals variation $D_{ik} \neq D_{jk}$ and the within-individual j variation $D_{jl} \neq D_{jk}$. The last equivalence is symmetric to the first equivalence for agent j.

The proof of the second equivalence in (9) follows analogous arguments. Thus, the random variable $\left\{ Y_{kl}^{(i)} - Y_{kl}^{(j)} \middle| \mathbf{X}^n = \mathbf{x}, \Omega(ijlk) \right\}$ has a Bernoulli distribution with support $\{-2, 2\}$, and

$$\begin{aligned} \text{Med}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = x, \Omega(ijlk)\right] \\ = 2 \times \text{sign}\Big\{\mathbb{P}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} = 2 | \mathbf{X}^n = x, \Omega(ijlk)\right] \\ - \mathbb{P}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} = -2 | \mathbf{X}^n = x, \Omega(ijlk)\right]\Big\}. \end{aligned} \tag{10}$$

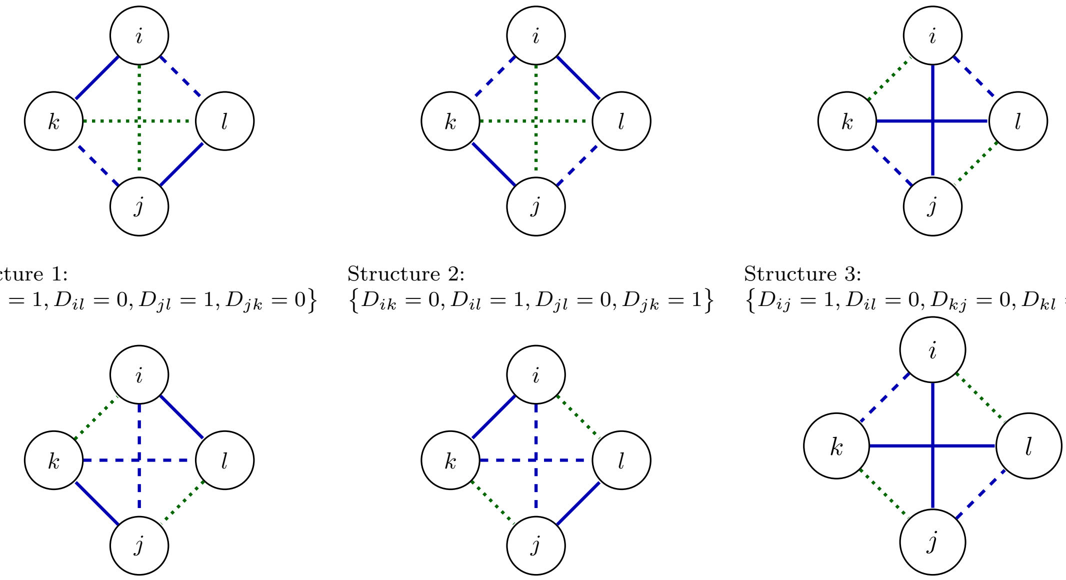
Finally, following the intuition described in section 3.1, given A1 and the set $\Omega(ijlk)$ the pairwise difference of the linking decisions for agents i and j implies:

$$\text{sign}\left\{\mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\right]\right\} = \text{sign}\left\{\left[(x_{ik} - x_{il}) - (x_{jk} - x_{jl})\right]' \beta_0\right\}, \tag{11}$$

for any $\mathbf{x} \in \mathcal{X}_B$.

The proof of part 1 in Theorem 3.1 is concluded by showing that the right-hand side of equation (10) is equal to twice the left-hand side of equation (11). This result follows from $\left\{ Y_{kl}^{(i)} - Y_{kl}^{(j)} \middle| \mathbf{X}^n = \mathbf{x}, \Omega(ijlk) \right\}$ being a Bernoulli random variable.

The point identification strategy relies on exploiting the within-individuals and across-individuals variations in the links formed. For any tetrad (i, j, k, l), the set $\Omega(ijlk)$ characterizes all possible subnetwork structures that generate sufficient variation to identify β_0. Figure 2 below depicts all the subnetworks contained in $\Omega(ijlk)$.



Structure 1: $\{D_{ik} = 1, D_{il} = 0, D_{jl} = 1, D_{jk} = 0\}$

Structure 2: $\{D_{ik} = 0, D_{il} = 1, D_{jl} = 0, D_{jk} = 1\}$

Structure 3: $\{D_{ij} = 1, D_{il} = 0, D_{kj} = 0, D_{kl} = 1\}$

Structure 4: $\{D_{ij} = 0, D_{il} = 1, D_{kj} = 1, D_{kl} = 0\}$

Structure 5: $\{D_{ik} = 1, D_{ij} = 0, D_{lk} = 0, D_{lj} = 1\}$

Structure 6: $\{D_{ik} = 0, D_{ij} = 1, D_{lk} = 1, D_{lj} = 0\}$

Note: A solid line indicates that a link exists, a dashed line indicates that a link is absent, and a slightly dotted line indicates that the link is either present or absent.

Figure 2: Subnetwork formed by agents $i, j, k, l \in \mathcal{N}_n$.

Consider the subnetwork structure 1 in figure 2, which is given by:

$$\{D_{ik} = 1, D_{il} = 0, D_{jl} = 1, D_{jk} = 0\}\,. \tag{12}$$

Under structure 1, the dyads (i, k) and (j, l) form an undirected link, depicted by solid lines. No link is formed by dyads (i, l) and (j, k), depicted by dashed lines. The decisions D_{ij} and $D_{k,l}$ could generate any outcome and the resulting structure will be consistent with $\Omega(ijkl)$. Note that if dyads (i, j) and (k, l) form undirected links, the resulting subnetwork structure is consistent with structure 3 in figure 2.

A2 is a sufficient condition for point identification of β_0. This assumption requires the existence of at least one covariate with full support. If this assumption fails, then equation (11) yields a collection of moment inequalities that can be used to partially identify β_0. Furthermore, these moments inequalities characterize the identified set of the network formation model with multiple fixed effects.

Specifically, let $\mathcal{B}_0$ denote the identified set. The identifying set is defined as the collection of $(b_2, \cdots, b_K) \in \mathbb{R}^{K-1}$ such that $b = (1, b_2, \cdots, b_K)$ satisfies the moment conditions in (11). That is

$$\begin{aligned} \mathcal{B}_0 = \Big\{ & b \in \mathbb{R}^k : \mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\right] \gtreqless 0 \Leftrightarrow [\Delta_{kl} x_i - \Delta_{kl} x_j]' b \gtreqless 0 \,, \\ & \forall\, i, j, k, l \in \mathcal{N}_n \text{ and} \\ & \forall \mathbf{x} = (x_{12}, \cdots, x_{ik}, \cdots, x_{il}, \cdots, x_{jk}, \cdots, x_{jl}, \cdots, x_{n-1,n}) \in \operatorname{supp}(\mathbf{X}^n) \cap \mathcal{X}_B \Big\} \,, \end{aligned} \tag{13}$$

The identifying set $\mathcal{B}_0$ will be used in section 3.4 to derive bounds for each elements in the vector of unknowns β_0 when assumption A2 fails.

3.3 Identification Failure

In this section, I discuss two cases of identification failure. First, I show that if the class

$$\Omega_n \equiv \{\Omega(ijlk) : i, j, k, l \in \mathcal{N}_n\} \,,$$

has probability zero, then the median of the pairwise difference of the links does not have identification power to recover β_0. Intuitively, if Ω_n has probability zero the underlying network does not exhibit sufficient within-individual and across-individuals variation to partial out the fixed effects.

Second, I show that the identification strategy implies by the panel data maximum score estimator (Manski 1987) does not identify β_0 in the network formation model with multiple fixed effects. In particular, computing the within-individual difference conditioning on the "switchers" fails to capture the contribution of the fixed effects along the longitudinal dimension.

3.3.1 Thin Set Identification

Intuitively, the set $\Omega(ijkl)$ ensures sufficient variation to partial out the agent-specific fixed effects. If the class Ω_n has probability zero, then $Y_{kl}^{(i)} - Y_{kl}^{(j)}$ will not have enough information to identify β_0. I formalize this result in the following theorem.

Theorem 3.2

Let assumptions A1 - A3. If the class Ω_n has probability zero, then the median of the random variable $\left\{ Y_{kl}^{(i)} - Y_{kl}^{(j)} \middle| \mathbf{X}^n = \mathbf{x} \right\}$ does not have identification power for any $\mathbf{x} \in \operatorname{supp}(\mathbf{X}^n)$. That *is, the set of parameters that are observationally equivalent to* β_0 *in terms of*

$$\text{Med}\left\{ Y_{kl}^{(i)} - Y_{kl}^{(j)} \middle| \mathbf{X}^n = \mathbf{x} \right\}$$

is $\mathbb{R}^K$*.*

The class Ω_n has probability zero if for any tetrad (i, j, k, l) the resulting subnetwork structure violates at least one condition in $\Omega(ijlk) \equiv \{D_{ik} \neq D_{il}, D_{jl} \neq D_{jk},\ \ D_{ik} \neq D_{jk}\}$. In appendix B.2, I characterize all the subnetwork structures that are not consistent with $\Omega(ijkl)$, and under which the class Ω_n does not have identification power.

Specifically, some examples of network structures for which the class Ω_n has probability zero include (i) dense networks where everybody is connected to everyone with probability one; (ii) empty networks where no links are formed with probability one; and (iii) homogeneous networks where individuals form similar connections with probability one. I formalize this intuition in the following proposition.

Proposition 3.1
Given the network formation model in (1), the class Ω_n *has probability zero if for any* n*, and any* $i, j \in \mathcal{N}_n$*:*

1. *(dense network) the conditional distribution of* $X_{ij}'\beta_0$ *given* $\tilde{\mu} = \mu$ *and* $\varepsilon_{ij} = e$ *has a probability density that is everywhere positive on the interval* $[\mu_i + \mu_j - e,\ \infty)$ *.*
2. *(empty network) the conditional distribution of* $X_{ij}'\beta_0$ *given* $\tilde{\mu} = \mu$ *and* $\varepsilon_{ij} = e$ *has a probability density that is everywhere positive on the interval* $(-\infty,\ \mu_i + \mu_j - e]$*.*
3. *(homogeneous network) the conditional distribution of* $\mu_i + \mu_j$ *given* $X_{ij}' = x$ *and* $\varepsilon_{ij} = e$ *has a probability density that is everywhere positive on the interval* $[\ e - x'\beta_0,\ \infty)$ *.*

The realized network under condition (1) in proposition 3.1 is dense since every link is created with probability 1. That is, for any dyad $(i, j) \in \mathcal{N}_n^{(2)}$: $\mathbb{P}[D_{ij} = 1 \mid \mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu] = 1$. Under condition (2), the realized network is empty since no link is created with probability one. That is, for any dyad $(i, j) \in \mathcal{N}_n^{(2)}$: $\mathbb{P}[D_{ij} = 1 \mid \mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu] = 0$. Finally, under condition (3) the individuals create similar connections driven by their preferences for agent-specific attributes. The resulting network structure is homogeneous. The class Ω_n has probability zero under the conditions (1), (2), and (3) in proposition 3.1 because the equilibrium network does not exhibit sufficient link variation either within or across individuals.

In tables 5 and 6 of section 5, I report Monte Carlo simulations which provide evidence on the probability of Ω_n been arbitrarily close to zero when the realized network structure is either dense or empty. Table 7 shows that Ω_n has probability arbitrarily close to zero when the realized network is homogeneous. Finally, the numerical evidence in tables 5 and 6 indicate that if the network structure is sparse, the probability of Ω_n is positive. A sparse network exhibits sufficient link variation.

The large support conditions in assumptions A1 and A2 guarantee that Ω_n has a probability greater than zero. Nonetheless, this probability could be arbitrarily small, which will complicate the inference procedure. In section 4, I discuss inference on β_0 under two scenarios: (i) the probability of Ω_n converges to zero as the network size grows, and (ii) the probability of Ω_n converges to a positive constant as the network size grows. Specifically, I show that the convergence rate of the estimator for β_0 can be slower than the square root of the sample size if the probability of Ω_n converges to zero.

3.3.2 Nonlinear Panel Data Identification Strategy: Maximum Score

In this section, I show that the incidental parameter problem in the network formation model in (1) is more complex than in a nonlinear panel data model with both cross-sectional and time fixed effects. Specifically, in a network formation model, the fixed effects across the longitudinal dimension may be arbitrarily correlated with the vector of observed attributes. Consequently, these fixed effects are not strongly exogenous as the time fixed effects are in nonlinear panel data models. To this end, I show that following a Maximum Score type identification strategy (Manski 1975, 1987) to identify the vector of parameters does not identify β_0 in this model.

The following proposition adapts Lemma 2 in Manski (1987) to the network formation model specified by equation (1). The result states that the median of $D_{ik} - D_{il}$ conditional on $\mathbf{X}^n = x$ and the "switchers" $D_{ik} \neq D_{il}$ does not identify β_0.

Proposition 3.2 *1. Let assumption A1 hold; then, for any n, and any $i, l, k \in \mathcal{N}_n$*

$$\text{Med}(D_{ik} - D_{il} | \mathbf{X}^n = x, D_{il} + D_{ik} = 1) = \text{sign}\left[(x_{ik} - x_{il})'\beta_0 + (\mu_k - \mu_l)\right]. \tag{6}$$

2. Let Assumptions A1 - A2 hold. Then, the set of parameters consistent with equation (6) is $\mathbb{R}^K$. That is, equation (6) does not have identification power.

In contrast to a nonlinear panel data model with a single individual fixed effect, conditioning on the within-individual variation $D_{il} \neq D_{ik}$ does not fully absorb the contribution of the multiple agent-specific fixed effects. The within-individual difference fails to partial out the fixed effects along the longitudinal dimension. This property is exhibited in the equation (6) by the presence of the composite factors $\mu_k - \mu_l$. In summary, the incidental parameter problem in a network formation model is more complex than in a nonlinear panel data model with both agent-specific and time fixed effects.

Remark 1
Proposition 3.2 formalizes the conjecture made by Charbonneau (2014) regarding the impossibility to generalize Maximum Score to the presence of multiple fixed effects. Proposition 3.2 states that the conditional median of $D_{ik} - D_{il}$ given $\{\mathbf{X}^n = \mathbf{x}, D_{ik} \neq D_{il}\}$, which is a (known) specific feature of the distribution of observables, does not have identification power to recover β_0. Nonetheless, this does not mean that β_0 is unidentified. Specifically, Theorem 3.1 proves that β_0 is point identified using a different (known) specific feature of the distribution of observables after conditioning on Ω_n.

3.4 Alternative Identifying Assumptions

The point identification result in section 3.2 relies on $\Delta_{kl}X_i^{(1)}$ having a large support conditional on $\Delta_{kl}X_i^{(-1)} = \Delta_{kl}x_i^{(-1)}$, for any $i, k, l \in \mathcal{N}_n$ (Assumption A2.2). In this section, I study the identification of the semiparametric network formation model in equation (1) when Assumption A2.2 is violated. Specifically, I consider two scenarios: (i) all the covariates have bounded support, and the conditional distribution of $\Delta_{kl}X_i^{(1)}$ given $\Delta_{kl}X_i^{(-1)} = \Delta_{kl}x_i^{(-1)}$ is continuous for any $\Delta_{kl}x_i^{(-1)}$; (ii) all the covariates have a discrete and finite support. The following results are especially relevant for empirical applications with datasets in which is hard to justify the existence of a covariate with large support.

3.4.1 Bounded Support and one Continuous Covariate

The main result of this section shows that β_0 can be point identified when the conditional distribution of $\Delta_{kl}X_i^{(1)}$ given $\Delta_{kl}X_i^{(-1)} = \Delta_{kl}x_i^{(-1)}$ is continuous, and all the covariates have a bounded support. In other words, I show that the existence of a covariate with large support is not a necessary condition to point identify β_0.

The next assumption weakens assumption A2.

Assumption A2′

The following hold for any n, and any $i, l, k \in \mathcal{N}_n$, with $l \neq k$.

1. *The random vector $\Delta_{kl}X_i$ has a bounded support on $\mathbb{R}^K$, and $\Delta_{kl}X_i^{(1)}$ is an absolutely continuous random variable.*

2. *For some $\eta > 0$, there exist an interval $S_\eta = [-\eta, \eta]$ and a set $A_\eta \in \mathbb{R}^{K-1}$ such that*

   (a) *A_η is not contained in any proper linear subspace of $\mathbb{R}^{K-1}$.*

   (b) *$\mathbb{P}\left(\Delta_{kl}X_i^{(-1)} \in A_\eta\right) > 0$.*

   (c) *For almost every $\Delta_{kl}x_i^{(-1)} \in A_\eta$, the distribution of $\Delta_{kl}X_i'\beta_0$ conditional on $\Delta_{kl}x_i^{(-1)} = \Delta_{kl}x_i^{(-1)}$ has a probability density that is everywhere positive on S_η.*

A2′.1 restricts the covariates to have a bounded support and therefore assumption A2 no longer holds. Part (c) in A2′.2 assumes that the linear index $\Delta_{kl}X_i'\beta_0$ has a continuous distribution in the interval S_η, which contains $\Delta_{kl}X_i'\beta_0 = 0$. This theorem, is a slightly modified version of the result obtained by Manski (1988) and Horowitz (2012) for semiparametric binary-response models.

Proposition 3.3

Let assumptions A1, A2′ and A3 hold; then β_0 is point identified.

To understand the intuition behind proposition 3.3, consider the identified set $\mathcal{B}_0$ defined in equation (13). To simplify the exposition, I restate the identified set below.

$$\begin{aligned}\mathcal{B}_0 = \Big\{ & b \in \mathbb{R}^k : \mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\right] \gtreqless 0 \Leftrightarrow [\Delta_{kl}x_i - \Delta_{kl}x_j]' b \gtreqless 0\,, \\ & \forall\, i, j, k, l \in \mathcal{N}_n \text{ and} \\ & \forall \mathbf{x} = (x_{12}, \cdots, x_{ik}, \cdots, x_{il}, \cdots, x_{jk}, \cdots, x_{jl}, \cdots, x_{n-1,n}) \in \operatorname{supp}(\mathbf{X}^n) \cap \mathcal{X}_B \Big\}\,,\end{aligned}$$

Assumption A2′ implies that the linear index $\Delta_{kl}X_i'\beta_0$ has a positive density on the interval S_η conditional on $\Delta_{kl}X_i^{(1)} = \Delta_{kl}x_i^{(1)}$, for any $\Delta_{kl}x_i^{(1)} \in A_\eta$ and $i, k, l \in \mathcal{N}_n$. Let $W_{kl,ij} \equiv (\Delta_{kl}X_i - \Delta_{kl}X_j)$, for any $i, j, l, k \in \mathcal{N}_n$, which has a continuous density with respect to the Lebesgue measure since $\Delta_{kl}X_i^{(1)}$ is a continuous random variable with bounded support.

Define the sets:

$$S_1(b) \equiv \left\{w : w'\beta_0 < 0 \ \leq w'b\right\}$$
$$S_2(b) \equiv \left\{w : w'b \ \ < 0 \ \leq w'\beta_0\right\}.$$

A necessary and sufficient condition for identification is

$$\mathbb{P}\left[S_1(b) \cup S_2(b)\right] > 0. \tag{14}$$

If $\mathbb{P}\left[S_1(b)\right] > 0$, then there exists a subset in $\text{supp}(W'_{kl,ij}\beta_0)$ with nonzero probability in which

$$\mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\right] < 0,$$

with parameter value β_0, and

$$\mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\right] \geq 0$$

for the parameter value b. The argument is symmetric for $\mathbb{P}\left[S_2(b)\right] > 0$.

Let $b = (1, b^{(-1)})$, $\beta_0 = (1, \beta_0^{(-1)})$, and $w = (w^{(1)}, w^{(-1)}) \in \text{supp}\left(W_{kl,ij}\right)$, with $b^{(-1)}, \beta_0^{(-1)}, w^{(-1)} \in \mathbb{R}^{K-1}$. Then the sets $S_1(b)$, $S_2(b)$ can be written as:

$$S_1(b) \equiv \left\{w : -w^{(-1)\prime}(b^{(-1)} - \beta_0^{(-1)}) \leq w'\beta_0 < 0\right\},$$
$$S_2(b) \equiv \left\{w : 0 \leq w'\beta_0 < -w^{(-1)\prime}(b^{(-1)} - \beta_0^{(-1)})\right\}$$

Conditions (a) and (b) in A2′.2 ensure that $\mathbb{P}\left[w^{(-1)\prime}(b^{(-1)} - \beta_0^{(-1)}) \neq 0\right] > 0$ for any $b^{(-1)} \neq \beta_0^{(-1)}$, and $w^{(-1)} \in A_\eta$. Then, condition (c) in A2′.2 implies that at least one of the sets $S_1(b), S_2(b)$ has a nonzero probability since $W'_{kl,ij}\beta_0$ has a conditional probability density given $w_{kl,ij}^{(-1)} = \Delta_{kl} x_i^{(-1)}$ that is everywhere positive on $2 \times S_\eta$. In other words, A2′ guarantees that the sufficient condition for identification in (14) is satisfied. Therefore, β_0 is point identified.

3.4.2 Finite and Discrete Support

In this section, I show that the parameter of interest in the the network formation model can be partially identified when all the covariates have a bounded and discrete support. This result is known from the literature of binary choice models such as Manski (1975) and Komarova (2013).

I follow the recursive procedure introduced by Komarova (2013) for binary choice models to obtain bounds on each component $\beta_{0,k}$ of the parameter of interest β_0, for $k = 2, \cdots, K$. The bounds obtained are used to approximate the identified set by the smallest multidimensional rectangular superset that covers the identified set $\mathcal{B}_0$.

The following assumption replaces A2.

Assumption A2″
For any n, and any $i, k, l \in \mathcal{N}_n$, with $k \neq l$.

1. *The support of X_{ik} is not contained in any proper linear subspace of $\mathbb{R}^K$.*
2. *The profile vector of observed attributes $\mathbf{X}^n \equiv (X_{12}, \cdots, X_{n-1,n})$ has a discrete support given by*
$$\text{supp}(\mathbf{X}^n) = \left\{\mathbf{x}^1, \cdots, \mathbf{x}^D\right\},$$
for a finite D.

Under A1, A2″ and A3, the identified set is the following:

$$\mathcal{B}_0 = \Big\{ b \in \mathbb{R}^k : \mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}^d, \Omega(ijlk)\right] \gtreqless 0 \Leftrightarrow \left[\Delta_{kl} x_i^{(d)} - \Delta_{kl} x_j^{(d)}\right]' b \gtreqless 0, \\ \text{for any distinct } i, j, k, l \in \mathcal{N}_n, \text{ and } \mathbf{x}^d \in \text{supp}(\mathbf{X}^n) \cap \mathcal{X}_B \Big\}, \tag{15}$$

where

$$\mathbf{x}^d \equiv \left(x_{12}^{(d)}, \cdots, x_{ik}^{(d)} \cdots, x_{il}^{(d)}, \cdots, x_{jk}^{(d)}, \cdots, x_{jl}^{(d)}, \cdots, x_{n-1,n}^{(d)}\right),$$
$$\Delta_{kl} x_i^{(d)} \equiv x_{ik}^{(d)} - x_{il}^{(d)}.$$

The identified set is characterized by a system of linear inequalities, which is used to compute the bounds for the components $\{\beta_{0,k}\}_{k=2}^K$ of β_0. The system of inequalities in (15) contains both strict and non-strict inequalities.

To simplify notation, let $\mathbf{i}_{1,2}$ denote the dyad (i_1, i_2) for any unique dyad in $\mathcal{N}_n^{(2)}$. Furthermore, let $P_{n,4}$ denote the set of total tetrads with distinct elements (i_1, i_2, i_3, i_4) from

$\{1, 2, \cdots, n\}$. Index the elements in $P_{n,4}$ by the letter $\mathbf{i}$.

$$z_{\mathbf{i},d} \equiv \text{sign}\left\{\mathbb{E}\left[Y_{3,4}^{(1)} - Y_{3,4}^{(2)}|\mathbf{X}^n = \mathbf{x}^d, \Omega(\mathscr{I}_1)\right]\right\}\left[\Delta_{3,4}x_1^{(d)} - \Delta_{3,4}x_2^{(d)}\right],$$

with $\Delta_{3,4}x_s^{(d)} = x_{\mathbf{i}_{s,3}}^{(d)} - x_{\mathbf{i}_{s,4}}^{(d)}$, $s = 1, 2$. For any $\mathbf{i} = (i_1, i_2, i_3, i_4) \in P_{n,4}$ and $\mathbf{x}^d \in \text{supp}(\mathbf{X}^n) \cap \mathcal{X}_B$, denote the K-dimensional vector of pairwise differences of the observed attributes that preserves the sign of the corresponding inequality in (15).

Let $z_{\mathbf{i},d}^{(k)}$ denote the kth element of the signed vector of pairwise differences of observed attributes $z_{\mathbf{i},d}$, for any $k = 1, \cdots, K$.

Rewrite the conditions characterizing the identified set in (15) as the following system of linear inequalities with $K-1$ unknowns given by $b_2, \cdots, b_K$

$$\begin{aligned}
z_{1,1}^{(1)} + z_{1,1}^{(2)}b_2 + z_{1,1}^{(3)}b_3 + \cdots + z_{1,1}^{(K)}b_K &\geq 0,\\
z_{2,1}^{(1)} + z_{2,1}^{(2)}b_2 + z_{2,1}^{(3)}b_3 + \cdots + z_{2,1}^{(K)}b_K &\geq 0,\\
&\vdots\\
z_{M,1}^{(1)} + z_{M,1}^{(2)}b_2 + z_{M,1}^{(3)}b_3 + \cdots + z_{M,1}^{(K)}b_K &\geq 0,\\
&\vdots\\
z_{M,D}^{(1)} + z_{M,D}^{(2)}b_2 + z_{M,D}^{(3)}b_3 + \cdots + z_{M,D}^{(K)}b_K &\geq 0,
\end{aligned} \tag{S_1}$$

where $M \equiv |P_{n,4}|$, and $|P_{n,4}|$ denotes the cardinality of $P_{n,4}$. For simplicity, the system in (S_1) is written as system non-strict linear inequalities; although, the initial system in (15) contains both.

The solutions to the system (S_1) establish bounds for each component of the parameter β_0 by following the recursive procedure introduced by Komarova (2013). Her procedure recursively simplifies the system (S_1) by excluding one unknown variable at each iteration. The recursive elimination continues until it reaches a simplified system with only one unknown variable. The upper and lower bounds for the remaining unknown are then computed from the simplified system. Her algorithm is repeated, using different elimination sequences, until the bounds for all the elements $\{\beta_{0,k}\}_{k=2}^K$ are computed.

Denote by $\underline{b}_k$ (and $\bar{b}_k$) the lower (upper) bound for the unknown parameter $\beta_{0,k}$ for $k = 2, \cdots, K$. Komarova (2013) shows that the identified set can be approximated by the smallest multidimensional rectangle superset that covers $\mathcal{B}_0$. This superset, denoted by $R(\mathcal{B}_0)$, is defined as the Cartesian product of the intervals $\left\{[\underline{b}_k,\ \bar{b}_k]\right\}_{k=2}^{K}$that bound the elements $\{\beta_{0,k}\}_{k=2}^{K}$. That is,

$$R(\mathcal{B}_0) \equiv \prod_{k=2}^{K} [\underline{b}_k,\ \bar{b}_k].$$

I illustrate her recursive procedure in the next example.

Example 2 (Bounds)
In this example, I characterize the identified set and the smallest multidimensional rectangle superset that covers the identified set. I discuss the computation of the bounds for a general network formation model in appendix C.

For any n, consider the following network formation model: [5]

$$D_{ik} = \mathbf{1}\left[X_{ik}^{(1)} + X_{ik}^{(2)}\beta_{0,2} + X_{ik}^{(3)}\beta_{0,3} + \mu_i + \mu_k - \varepsilon_{ik} \geq 0\right] \quad \text{for any } (i,k) \in \mathcal{N}_n^{(2)}, \tag{16}$$

where $\beta_0 = (1, \beta_{0,2}, \beta_{0,3})' = (1, 1.5, -1.5)'$. For any $(i,k) \in \mathcal{N}_n^{(2)}$, the supports of the observed attributes are $\operatorname{supp}(X_{ik}^{(1)}) = \{-2,-1,0,1,2,3,4\}$, $\operatorname{supp}(X_{ik}^{(2)}) = \{-1,0,1\}$, and $\operatorname{supp}(X_{ik}^{(3)}) = \{0,1,2\}$. Hence, the support of X_{ik} contains 63 points.

To characterize the identified set is necessary to determine the set of vectors $b \in \mathbb{R}^K$ that are observationally equivalent to β_0 under the moment inequalities in (15).

Under the true DGP, the sign of each inequality in (S_1) is determined according to the rule:

If

$$(1, 1.5, -1.5)\left[(x_{ik}^{(d)} - x_{il}^{(d)}) - (x_{jk}^{(d)} - x_{jl}^{(d)})\right] \geq 0 \Rightarrow \mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}^d, \Omega(\mathbf{i})\right] \geq 0.$$

If

$$(1, 1.5, -1.5)\left[(x_{ik}^{(d)} - x_{il}^{(d)}) - (x_{jk}^{(d)} - x_{jl}^{(d)})\right] < 0 \Rightarrow \mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}^d, \Omega(\mathbf{i})\right] < 0,$$

for any $\mathbf{i} = (i,j,k,l) \in P_{n,4}$ and $\mathbf{x}^d \in \operatorname{supp}(\mathbf{X}^n) \cap \mathcal{X}_B$.

---

[5]This example uses the same data generating process (DGP) design for the network formation model as the Monte Carlo simulations in section 5 up to the discretization of the supports of $X_{ik}^{(1)}$ and $X_{ik}^{(3)}$. Assumption A2 requires $X_{ik}^{(1)}$ to have a large support conditional on the remaining exogenous covariates. The discretized support of $X_{ik}^{(1)}$ accounts for 95% of its original probability mass. The discretized support of $X_{ik}^{(3)}$ takes both end points of the original support, and the only integer value in between the end points.

The identified set $\mathcal{B}_0$ is defined as the set of vectors $b = (1, b_2, b_3) \in \mathbb{R}^3$ that satisfy

$$\begin{aligned}
\mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = \mathbf{x}^1, \Omega(\mathbf{i})\right] \gtreqless 0 &\Leftrightarrow \left[(x_{ik}^{(1)} - x_{il}^{(1)}) - (x_{jk}^{(1)} - x_{jl}^{(1)})\right]' b \gtreqless 0,\\
\mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = \mathbf{x}^2, \Omega(\mathbf{i})\right] \gtreqless 0 &\Leftrightarrow \left[(x_{ik}^{(2)} - x_{il}^{(2)}) - (x_{jk}^{(2)} - x_{jl}^{(2)})\right]' b \gtreqless 0,\\
&\vdots\\
\mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = \mathbf{x}^D, \Omega(\mathbf{i})\right] \gtreqless 0 &\Leftrightarrow \left[(x_{ik}^{(D)} - x_{il}^{(D)}) - (x_{jk}^{(D)} - x_{jl}^{(D)})\right]' b \gtreqless 0,
\end{aligned} \tag{E_1}$$

for any $\mathbf{i} = (i, j, k, l) \in P_{n,4}$.

The bounds for the components $\beta_{0,2}$ and $\beta_{0,3}$ are computed from the solutions to the following system of linear inequalities implied by (E_1). Let,

$$z_{\mathbf{i},d}^k = \operatorname{sign} \mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = \mathbf{x}^d, \Omega(\mathbf{i})\right]\left[(x_{ik}^{(d)} - x_{il}^{(d)}) - (x_{jk}^{(d)} - x_{jl}^{(d)})\right],$$

for $d = 1, \cdots, D$. System (E_1) can be written as:

$$\begin{aligned}
z_{1,1}^{(1)} + z_{1,1}^{(2)} b_2 + z_{1,1}^{(3)} b_3 &\geq 0,\\
z_{2,1}^{(1)} + z_{2,1}^{(2)} b_2 + z_{2,1}^{(3)} b_3 &\geq 0,\\
&\vdots\\
z_{M,D}^{(1)} + z_{M,D}^{(2)} b_2 + z_{M,D}^{(3)} b_3 &\geq 0.
\end{aligned} \tag{E_2}$$

To illustrate the recursive procedure, suppose the goal is to find the bounds for the component $\beta_{0,3}$. Consider, the ijth inequality in system (E_2)

$$z_{i,j}^{(1)} + z_{i,j}^{(2)} b_2 + z_{i,j}^{(3)} b_3 \geq 0.$$

Solving for b_2, the ijth linear inequality is equivalent to:

$$\begin{aligned}
\text{if} \quad z_{i,j}^{(2)} \geq 0 \quad &\Rightarrow \quad -\frac{z_{i,j}^{(1)}}{z_{i,j}^{(2)}} - \frac{z_{i,j}^{(3)}}{z_{i,j}^{(2)}} b_3 \leq b_2,\\
\text{if} \quad z_{i,j}^{(2)} \leq 0 \quad &\Rightarrow \quad -\frac{z_{i,j}^{(1)}}{z_{i,j}^{(2)}} - \frac{z_{i,j}^{(3)}}{z_{i}^{(2)}} b_3 \geq b_2.
\end{aligned}$$

The process is repeated on the $M \times D$ linear inequalities in (E_2). Suppose that the system (S_1) has N_1 inequalities with $z_{\cdot}^{(2)} \geq 0$, N_2 inequalities with $z_{\cdot}^{(2)} \leq 0$, and N_3 inequalities with

$z^{(2)}_{\cdot} = 0$; *then the system* (S_1) *is equivalent to*

$$\begin{aligned} L_i(b_3) &\leq b_2, \qquad i = 1, \cdots, N_1, \\ U_j(b_3) &\geq b_2, \qquad j = 1, \cdots, N_2, \\ Z_r(b_3) &\geq 0, \qquad r = 1, \cdots, N_3, \end{aligned} \tag{E_3}$$

where $L_i(\cdot), U_j(\cdot), Z_r(\cdot)$ *are linear functions of* b_3 *and do not depend on* b_2*.*

The system (E_3) *yields the simplified system:*

$$\begin{aligned} &U_j(b_3) \geq L_i(b_3), \quad i = 1, \cdots, N_1, \quad j = 1, \cdots, N_2, \\ &Z_r(b_3) \geq 0, \quad r = 1, \cdots, N_3, \end{aligned}$$

which can be written as

$$\begin{aligned} u_l + v_l b_3 &\geq 0, \quad l = 1, \cdots, L, \\ w_r &\geq 0, \quad r = 1, \cdots, N_3, \end{aligned} \tag{E_4}$$

where $L \equiv N_1 \times N_2$*. System* (E_4) *was obtained after simplifying* (E_1) *using the recursive procedure introduced by Komarova (2013). The system* (E_4) *has* b_3 *as the only unknown variable.*

The lower and upper bounds for $\beta_{0,3}$ *are derived from the simplified system* (E_4) *as follows:*

$$\begin{aligned} \underline{b}_3 &= \max_{l=1,\cdots,L} \left\{ -\frac{u_l}{v_l} : v_l > 0 \right\}, \\ \bar{b}_3 &= \min_{l=1,\cdots,L} \left\{ -\frac{u_l}{v_l} : v_l < 0 \right\}. \end{aligned}$$

The process to obtain the bounds for $\beta_{0,2}$ *is symmetric. Then, the smallest multidimensional rectangle superset* $R(\mathcal{B}_0)$ *that covers the identified set is*

$$R(\mathcal{B}_0) = [\underline{b}_2,\ \bar{b}_2] \times [\underline{b}_3,\ \bar{b}_3].$$

In figure 3, I depict the bounds for the components in β_0*, the identified set, and the smallest multidimensional rectangle superset that covers* $\mathcal{B}_0$*.*

The bounds obtained by the recursive procedure introduced by Komarova (2013) are unique and independent from the order used to simplify the system (15). That is, the bounds obtained are uniform over the order of elimination process. Notably, the identified set is char-

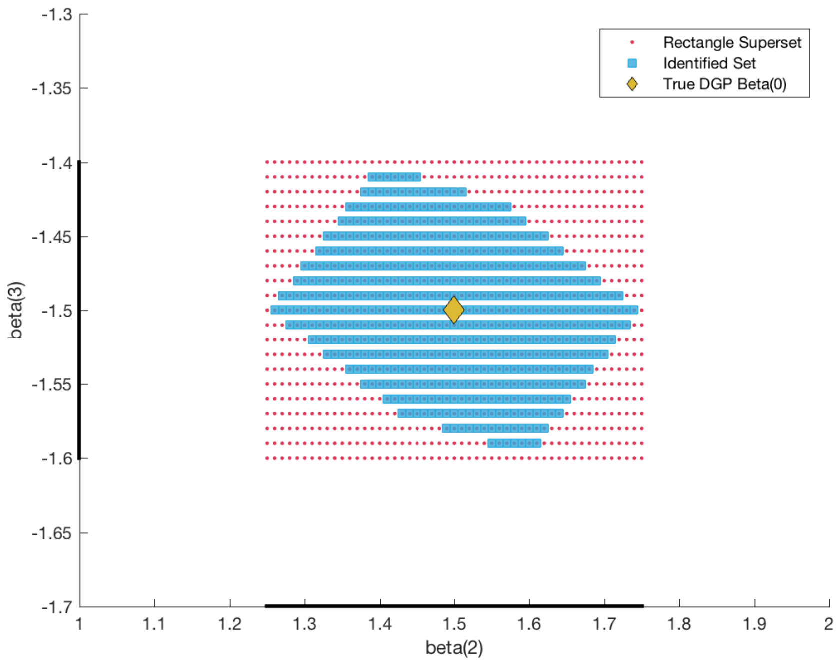


Figure 3: Bounds and Rectangular Superset

acterized using only the information contained in the conditional median of $Y_{kl}^{(i)} - Y_{kl}^{(j)}$ given $\{\mathbf{X}^n = \mathbf{x}, \Omega(ijkl)\}$. Sharpness of the identified set and the bounds $\left\{[\underline{b}_k,\ \bar{b}_k]\right\}_{k=2}^{K}$ is still an open question in the literature, and I leave this question as future research.

4 Estimation

In this section, I propose a semiparametric pairwise difference estimator for β_0 under the point identification assumptions of section 3. A semiparametric approach is attractive because it does not confine the distribution of the disturbance term to any specific parametric family. Furthermore, it allows for a flexible statistical dependence structure between the agent-specific factors and the exogenous attributes.

The pairwise difference estimator is an M-estimator that minimizes a 4th order U-process. The estimator generalizes the Leapfrog estimator introduced by Abrevaya (1999b) to a network structure with multiple and unobserved heterogeneity. In particular, Abrevaya (1999b) introduced the Leapfrog estimator as an estimation method for strictly monotonic transformation models, for both panel data and cross-sectional frameworks. Invertibility of the transformation function is a necessary condition for the models considered in that paper. The network formation model in equation (1) constitutes a weakly monotonic transformation function that is not invertible. Hence, the network model studied in this paper is not nested in the class of models considered by Abrevaya (1999b).[6]

I show that the estimator for β_0 is consistent and has an asymptotic normal distribution. If the probability of the class Ω_n converges to a positive constant, as the size of the network increases, the estimator has a parametric convergence rate (square root of the sample size). If the probability of the class Ω_n converges to zero, the convergence rate of the estimator is slower than the parametric rate. The slower rate of convergence is a consequence of identifying β_0 in a set with arbitrarily small probability, also referred to as a thin set. In this case, β_0 is said to be irregularly identified (Newey 1990; Andrews and Schafgans 1998 and Khan and Tamer 2010).

4.1 Pairwise difference Estimator

I propose an estimator for β_0 based on the identification condition described in (7). Consider following limiting objective function

$$Q(b) \equiv 2\mathbb{E}\left[S(\mathcal{X}_B) \times \text{ sign}\left\{[(X_{ik} - X_{il}) - (X_{jk} - X_{jl})]' \, b\right\} \times \left(Y_{kl}^{(i)} - Y_{kl}^{(j)}\right) \mid \Omega(ijlk)\right], \quad (17)$$

where, $S(\mathcal{X}_B)$ is an indicator function that is equal to 1 if $\mathbf{x} \in \mathcal{X}_B$, and 0 otherwise.

The next proposition states that this limiting objective function is uniquely maximized at the true parameter value, $b = \beta_0$.

Proposition 4.1
Let assumptions A1, A2 and A3 hold. Then, the limiting objective function $Q(b)$ is uniquely maximized at $b = \beta_0$. That is,

$$Q(\beta_0) > Q(b), \quad \textit{for all} \quad b \in \mathbb{R}^K \textit{ with } b \neq \beta_0.$$

[6] Abrevaya denotes the Leapfrog estimator as pairwise difference estimator in the cross-sectional framework. Although Abrevaya's estimator also minimizes a 4th order U-statistics, the semiparametric estimator introduced in this paper is qualitatively different from his estimator because is developed to estimate models of link formation among dyads. The two estimators have different asymptotic behaviors, which becomes visible from their distinct convergence rates.

Consider a sample of size n

$$\{z_{ij}\}_{(i,j)\in\mathcal{N}_n^{(2)}} \equiv \{D_{ij}, \mathbf{x}_{ij}\}_{(i,j)\in\mathcal{N}_n^{(2)}} .$$

Recall that $\mathbf{i}_{1,2}$ indexes the unique dyad $(1,2)$ in $\mathcal{N}_n^{(2)}$. The sample analog of the limiting objective function is a 4th order U-statistic defined as

$$Q_n(b) \equiv \binom{n}{4}^{-1} \sum_{C_{n,4}} h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b), \tag{18}$$

where $\sum_{C_{n,4}}$ denotes summation over the $\binom{n}{4}$ combinations of tetrads with distinct elements $(\mathbf{i}_{1,3}, \mathbf{i}_{1,4}, \mathbf{i}_{2,3}, \mathbf{i}_{2,4})$ from $\{1, 2, \cdots, n\}$. The function h, known as the kernel of the U-statistic, is defined as

$$\begin{aligned} h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) \equiv \\ \frac{2}{4!}\sum_{P_4} \left\{ S(x_{\mathbf{i}_{1,3}}, x_{\mathbf{i}_{1,4}}, x_{\mathbf{i}_{2,3}}, x_{\mathbf{i}_{2,4}}, B_U, B_L) \times \operatorname{sign}\left\{ [\Delta_{3,4}x_1 - \Delta_{3,4}x_2]' \, b \right\} \right. \\ \left. \times \left(y_{3,4}^{(1)} - y_{3.4}^{(2)} \right) \times \mathbf{1}\left\{ \left| \left(y_{3,4}^{(1)} - y_{3.4}^{(2)} \right) \right| = 2 \right\} \right\}, \end{aligned}$$

where

$$\begin{aligned} S(x_{\mathbf{i}_{1,3}}, x_{\mathbf{i}_{1,4}}, x_{\mathbf{i}_{2,3}}, x_{\mathbf{i}_{2,4}}, B_U, B_L) = \\ \mathbf{1}\left[\Delta_{3,4}x_1'b > (B_U - B_L)\right] \mathbf{1}\left[\Delta_{3,4}x_2'b < (B_L - B_U)\right] + \\ \mathbf{1}\left[\Delta_{3,4}x_1'b < (B_L - B_U)\right] \mathbf{1}\left[\Delta_{3,4}x_2'b > (B_U - B_L)\right], \end{aligned}$$

and $\sum_{P_4}$ denotes summation over the 4! permutations $\{\mathbf{i}_{1,3}, \mathbf{i}_{1,4}, \mathbf{i}_{2,3}, \mathbf{i}_{2,4}\}$ of $\{1, 2, 3, 4\}$. The kernel function is symmetric with respect to its argument (see Remark 2 below).

The semiparametric pairwise difference estimator is

$$\hat{\beta}_n = \underset{b\in\tilde{\mathcal{B}}\subset\mathbb{R}^K}{\arg\max}\, Q_n(b), \tag{19}$$

where the first dimension of the vector of unknown parameters is normalized equal to one in the parameter space, as a consequence of scale normalization used to point identify β_0 instead of the scaled parameter $\beta_0/\,||\beta_0||$. This normalization is discussed in section 3.2.

Remark 2

Given that underlying network is undirected, it can be shown that 18 *out of the* 4! *total permutations have an identical contribution to the kernel function. For example, the permutations* $(\mathbf{i}_{1,3}, \mathbf{i}_{1,4}, \mathbf{i}_{2,3}, \mathbf{i}_{2,4})$ *and* $(\mathbf{i}_{3,1}, \mathbf{i}_{3,2}, \mathbf{i}_{4,1}, \mathbf{i}_{4,2})$ *have identical contribution to the kernel since*

$$\left(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}}\right) - \left(x_{\mathbf{i}_{2,3}} - x_{\mathbf{i}_{2,4}}\right) = \left(x_{\mathbf{i}_{3,1}} - x_{\mathbf{i}_{3,2}}\right) - \left(x_{\mathbf{i}_{4,1}} - x_{\mathbf{i}_{4,2}}\right),$$
$$y_{3,4}^{(1)} - y_{3,4}^{(2)} = \left(D_{\mathbf{i}_{1,3}} - D_{\mathbf{i}_{1,4}}\right) - \left(D_{\mathbf{i}_{2,3}} - D_{\mathbf{i}_{2,4}}\right) = \left(D_{\mathbf{i}_{3,1}} - D_{\mathbf{i}_{3,2}}\right) - \left(D_{\mathbf{i}_{4,1}} - D_{\mathbf{i}_{4,2}}\right) = y_{1,2}^{(3)} - y_{1,2}^{(4)}.$$

Hence,

$$\begin{aligned}\operatorname{sign}&\left\{\left[\left(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}}\right) - \left(x_{\mathbf{i}_{2,3}} - x_{\mathbf{i}_{2,4}}\right)\right]' b\right\} \times \left(y_{3,4}^{(1)} - y_{3,4}^{(2)}\right) \\ &= \operatorname{sign}\left\{\left[\left(x_{\mathbf{i}_{3,1}} - x_{\mathbf{i}_{3,2}}\right) - \left(x_{\mathbf{i}_{4,1}} - x_{\mathbf{i}_{4,2}}\right)\right]' b\right\} \times \left(y_{1,2}^{(3)} - y_{1,2}^{(4)}\right).\end{aligned}$$

The proof for the remaining cases is analogous. Therefore, the kernel function simplifies to

$$\begin{aligned} & h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) \equiv \\ \frac{2}{6}\sum_{s=1}^{6} & \left\{ S(x_{\mathbf{s}_{1,3}}, x_{\mathbf{s}_{1,4}}, x_{\mathbf{s}_{2,3}}, x_{\mathbf{s}_{2,4}}, B_U, B_L) \times \operatorname{sign}\left\{\left[\left(x_{\mathbf{s}_{1,3}} - x_{\mathbf{s}_{1,4}}\right) - \left(x_{\mathbf{s}_{2,3}} - x_{\mathbf{s}_{2,4}}\right)\right]' b\right\} \right. \\ & \left. \times \left(y_{3,4}^{(1)} - y_{3,4}^{(2)}\right) \times \mathbf{1}\left\{\left|\left(y_{3,4}^{(1)} - y_{3,4}^{(2)}\right)\right| = 2\right\}\right\}, \end{aligned}$$

where $(\mathbf{s}_{1,3}, \mathbf{s}_{1,4}, \mathbf{s}_{2,3}, \mathbf{s}_{2,4})$ *denotes the permutation of the index* $(\mathbf{i}_{1,3}, \mathbf{i}_{1,4}, \mathbf{i}_{2,3}, \mathbf{i}_{2,4})$. *The 6 unique permutation are*

$$\begin{aligned}&\{(\mathbf{i}_{1,3}, \mathbf{i}_{1,4}, \mathbf{i}_{2,3}, \mathbf{i}_{2,4}), (\mathbf{i}_{1,4}, \mathbf{i}_{1,3}, \mathbf{i}_{2,4}, \mathbf{i}_{2,3}), (\mathbf{i}_{1,2}, \mathbf{i}_{1,4}, \mathbf{i}_{3,2}, \mathbf{i}_{3,4}), \\ &(\mathbf{i}_{1,4}, \mathbf{i}_{1,2}, \mathbf{i}_{3,4}, \mathbf{i}_{3,2}), (\mathbf{i}_{1,2}, \mathbf{i}_{1,3}, \mathbf{i}_{4,2}, \mathbf{i}_{4,3}), (\mathbf{i}_{1,3}, \mathbf{i}_{1,2}, \mathbf{i}_{4,3}, \mathbf{i}_{4,2})\}.\end{aligned}$$

4.2 Consistency

In this section, I provide sufficient conditions for the pairwise difference estimator, defined in equation (19), to be consistent. Assumptions B1 and B2 adapt those in Abrevaya (1999b) to a network formation model. Assumption B3 imposes a lower bound on how fast the probability of the class Ω_n can go to zero as the sample size increase.

Assumption B1

The researcher observes a random sample of n *agents, the link status and dyad-level observed* *attributes for all the unique dyads in the sample*

$$\left\{(D_{ij}, \mathbf{x}_{ij})\right\}_{(i,j)\in\mathcal{N}_n^{(2)}}, \text{ for } n \in \mathbb{N}.$$

Assumption B2
The parameter space $\tilde{\mathcal{B}}$ is compact and β_0 is an interior point of $\tilde{\mathcal{B}}$.

Assumption B3
Let $p_n \equiv \mathbb{P}(\Omega_n)$, where

1. $p_n \to p_0 \geq 0$, *as* $n \to \infty$.
2. $\sqrt{N}p_n \to \infty$, *as* $n \to \infty$.

Assumption B1 states that the researcher observes only one realization of the network with a large number of individuals. The asymptotic analysis is conducted by assuming the number of individuals in the sample increases. This framework is known as "large-market" asymptotics, and is suitable for applications in which only one large network is observed. In recent years, the "large-market" asymptotics paradigm has received an increasingly amount of attention in the network formation literature. Some notable papers that follow this approach are Boucher and Mourifié (2013); Chandrasekhar and Jackson (2014); Graham (2017); Leung (2015a,b); Menzel (2015) and de Paula et al. (2016). In the empirical application, I study the friendships network formed within one high school by a large number of students.

Assumption B2 is a regularity condition that is frequently used in the literature of semi-parametric methods. Compactness of the parameter space is used to prove consistency of the pairwise difference estimator. Assumption B2 also requires β_0 to be an interior point of the parameter space. This condition is used to derive the asymptotic distribution of the estimator. The methodology relies on finding a quadratic approximation for a smooth function of the kernel of the U-statistic. Condition B2 has also been used by Han (1987); Sherman (1993, 1994) and Abrevaya (1999b). For further references see Powell (1994).

Assumption B3 states that the probability of the identifying class Ω_n converges to a non-negative constant, which could be zero. Under this assumption, the subnetwork structures that satisfy the conditions in $\Omega(i, j, k, l)$, see figure 2, become more unlike as the number of individuals in the network increases. The probability of Ω_n will convergence to zero if there is not enough within-individual and across individuals variation in the links. Some examples of networks for which the probability of Ω_n converges to zero are networks that, with probability approaching one, become dense, empty, homogeneous, or if the subnetwork created by any sampled tetrad consists of a single edge, two edges that are adjacent to each other, or three edges. In Appendix B, I depict all the subnetworks that do not meet the conditions in $\Omega(ijkl)$.

Assumption B3.2 states that the probability of the class Ω_n cannot converge to zero at a faster rate than square root of the sample size $N = O(n^2)$. The unique dyads formed by the n individuals constitute the relevant sample in the network formation model. Therefore, the actual sample size needed to estimate β_0 is $p_n\sqrt{N}$. A similar property to B3.2 is used in Graham (2017).

Theorem 4.1
Let assumptions A1–A3, B1–B3 hold. Then,

$$\hat{\beta}_n - \beta_0 \overset{a.s.}{\to} 0$$

as $n \to \infty$.

Theorem 4.1 shows that the pairwise difference estimator converges to the true parameter value almost surely as the size of the network increases.

4.2.1 Asymptotic Normality

The main result of this section is that the pairwise difference estimator is asymptotically normal. The proof of this result follows similar arguments as in Sherman (1993, 1994). The following assumption provides sufficient conditions to derive the asymptotic distribution. First, I introduce additional notation to simplify the exposition.

For $b \in \tilde{\mathcal{B}}$, and each $z \in S$ with sampling distribution $\mathbf{P}$ on S, let

$$\tau(z,b) \equiv h(z,\mathbf{P},\mathbf{P},\mathbf{P},b) + h(\mathbf{P},z,\mathbf{P},\mathbf{P},b) + h(\mathbf{P},\mathbf{P},z,\mathbf{P},b) + h(\mathbf{P},\mathbf{P},z,\mathbf{P},b),$$

where $h(z,\mathbf{P},\mathbf{P},\mathbf{P},b)$ denotes the conditional expectation of $h(\cdot,\cdot,\cdot,\cdot,b)$ under $\mathbf{P}^4$, given its first argument. $\mathbf{P}^4$ denotes the product measure $\mathbf{P}\times\mathbf{P}\times\mathbf{P}\times\mathbf{P}$ for the sampling distribution $\mathbf{P}$ on S, and given $\mathbf{P}^4 < \infty$.

Although $h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{1,4}}, \cdot)$ is a discontinuous function for each $(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{1,4}}) \in S^4$, the function $\tau(z,\cdot)$ can be many times differentiable if the distribution of $[\Delta_{3,4}x_1 - \Delta_{3,4}x_2]'\,b$, is sufficiently smooth.

Let $||\cdot||$ denote the Frobenius matrix norm, $||(a_{ij})|| = (\sum_{i,j} a_{i,j}^2)^{1/2}$, ∇_m denote the mth partial derivative operator with respect to b, and let

$$|\nabla_m|g(b) \equiv \sum_{i_1,\cdots,i_m} \left| \frac{\partial^m}{\partial b_{i_1} \cdots \partial b_{i_m}} g(b) \right|,$$

for any differentiable function of b.

Assumption B4
Let $\mathscr{M}$ denote a neighborhood of β_0.

1. *For each $z \in S$, all mixed second partial derivatives of $\tau(z, \cdot)$ exist on $\mathscr{M}$.*
2. *There is an integrable function $M(z)$, such that for all z in S and b in $\mathscr{M}$*

$$||\nabla_2 \tau(z, b) - \nabla_2 \tau(z, \beta_0)|| \leq M(z)|b - \beta_0|.$$

3. $\mathbb{E}|\nabla_1 \tau(\cdot, \beta_0)|^2 < \infty$.
4. $\mathbb{E}|\nabla_2|\tau(\cdot, \beta_0) < \infty$.
5. *The matrix $\mathbb{E}\left[\nabla_2 \tau(\cdot, \beta_0) \mid \Omega_n\right]$ is negative definite.*

Theorem 4.2
Let $\hat{\beta}_n$ be a value that maximizes $Q_n(\beta)$ over the parameter space $\mathcal{B}$. If assumptions A1–A3, and B1–B4 hold, then:

$$p_n \sqrt{N}(\hat{\beta}_n - \beta_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, V^{-1} \Delta V^{-1}), \quad \text{as } n \to \infty \tag{20}$$

where

$$4V = \mathbb{E}\left[\nabla_2 \tau(\cdot, \beta_0) \mid \Omega_n\right],$$
$$\Delta = \mathbb{E}\left[\nabla_1 \tau(\cdot, \beta_0)\right]\left[\nabla_1 \tau(\cdot, \beta_0)\right]'.$$

If the probability of the class Ω_n converges to a positive constant, $p_n \to p_0 > 0$ as $n \to \infty$, then the pairwise difference estimator has a parametric convergence rate $\sqrt{N}$.

If the probability of the class Ω_n converges to zero, then the converge rate is slower than the parametric rate. This result is a consequence of identifying β_0 in a thin set. The next theorem shows that the information bound is zero for the network formation model if the probability of the Ω_0 converges to zero.

Theorem 4.3
In the network formation model characterized by equation (1), under assumptions A1–A3, B1–B4, and if $p_n \to 0$ as $n \to \infty$, then the information bound for β_0 is 0.

Given the varying rates, inference can be conducted using the approach proposed by Andrews and Schafgans (1998) and Khan and Tamer (2010). Let $\hat{\Sigma}_n$ denote the estimator of the asymptotic variance:

$$\hat{\Sigma}_n = \hat{V}_n^{-1}\hat{\Delta}_n\hat{V}_n^{-1}/\hat{p}_n^2,$$

where $\hat{V}_n$ and $\hat{\Delta}_n$ denote consistent estimators for V and Δ, and

$$\hat{p}_n \equiv \binom{n}{4}^{-1} \sum_{C_{n,4}} \mathbf{1}\left[h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{1,4}}, \hat{\beta}_n) \neq 0\right], \tag{21}$$

which is a consistent estimator for p_0.

Theorem 4.4
Let $\hat{\beta}_n$ be a value that maximizes $Q_n(\beta)$ over the parameter space $\mathcal{B}$. If assumptions A1–A3, and B1–B4 hold, then:

$$\hat{\Sigma}_n^{-1/2}\sqrt{N}(\hat{\beta}_n - \beta_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I), \quad \text{as } n \to \infty \tag{22}$$

I conduct inference by estimating the asymptotic covariance matrix and referring to the standard normal critical value. Section 7 in Sherman (1993) shows how to consistently estimate V and Δ using numerical derivatives. Alternatively, Subbotin (2007) shows that the nonparametric bootstrap is valid for inference for maximum rank estimators. In the empirical application, I use the alternative bootstrap method introduced in Honoré and Hu (2015). Their approach reduced the computing time by estimating only one-dimensional parameters instead of one $K \times K$ dimensional parameter.

4.3 Asymptotic Properties under Partial Identification

In this section, I show that if β_0 is partially identified, then the identified set can be consistently approximated from the system of linear inequalities in (15). Specifically, by replacing the conditional expectations in (15) with a consistent estimates, the resulting system can be used to approximate the identified set. The next theorem provides the main result in this section.

First, denote by $H(\cdot,\cdot)$ the Hausdorff metric. Specifically, for two non-empty sets A and B let

$$H(A,B) \equiv \max\left\{\sup_{a\in A}\inf_{b\in B}||a-b||\,,\sup_{b\in B}\inf_{a\in A}||a-b||\right\}.$$

Furthermore, let

$$\hat{\mathbb{E}}_n\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = \mathbf{x}^d, \Omega(ijlk)\right]$$

denote a consistent estimator of

$$\mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = \mathbf{x}^d, \Omega(ijlk)\right]$$

as the sample size grows, and for any distinct $i,j,k,l \in \mathcal{N}_n$ and $\mathbf{x}^d \in \operatorname{supp}\mathbf{X}^n \cap \mathcal{X}_B$.

Theorem 4.5

Let assumptions A1, A2′, A3, B1 and B3 hold. Then, if

$$r_n\left(\hat{\mathbb{E}}_n\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = \mathbf{x}^d, \Omega(ijlk)\right] - \mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = \mathbf{x}^d, \Omega(ijlk)\right]\right) = O_p(1),$$

for any distinct $i,j,k,l \in \mathcal{N}_n$ and $\mathbf{x}^d \in \operatorname{supp}\mathbf{X}^n \cap \mathcal{X}_B$, where $\{r_n\}_{n\in\mathbb{N}}$ is a nonnegative sequence such that $r_n \to \infty$ as $n \to \infty$. Let $\{\epsilon_n\}_{n\in\mathbb{N}}$ be a nonnegative sequence such that $\epsilon_n \to 0$ and $\epsilon_n r_n \to \infty$.

Let $\hat{\mathcal{B}}$ denote a solution to the following system of inequalities:

$$\hat{\mathbb{E}}_n\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = \mathbf{x}^d, \Omega(ijlk)\right] \gtreqless -\epsilon_n \Leftrightarrow \left[\left(x_{ik}^{(d)} - x_{il}^{(d)}\right) - \left(x_{jk}^{(d)} - x_{jl}^{(d)}\right)\right]' b \gtreqless 0. \tag{23}$$

for any distinct $i,j,k,l \in \mathcal{N}_n$ and $\mathbf{x}^d \in \operatorname{supp}\mathbf{X}^n \cap \mathcal{X}_B$. Then:

1. $H(\hat{\mathcal{B}}, \mathcal{B}_0) \xrightarrow{p} 0$,
2. $H(R(\hat{\mathcal{B}}), R(\mathcal{B}_0)) \xrightarrow{p} 0$,

as $n \to \infty$.

The previous theorem states that the estimated identified set, and interval bounds, consistently approximate the true identified set, and true interval bounds, for each component of β_0, respectively.

5 Monte Carlo Simulations

5.1 Computation

The objective function $Q_n(b)$ is a 4th order U-statistic, which requires $O(n^4)$ operations. The estimator $\hat{\beta}_n$ can be equivalently computed from the following objective function $\tilde{Q}_n(b)$, which can be computed in $O(n^3 log(n))$ operations by implementing sorting algorithms that uses binary search trees, as described in Abrevaya (1999a).

$$\tilde{Q}_n(b) \equiv \frac{1}{n(n-1)(n-2)} \sum_{P_3} S(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, b)\text{Rank}_{(\mathbf{i}_{3,4})} \left[(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}})'b\right] y^{(1)}_{3,4},$$

where $\sum_{P_3}$ denotes summation over the $n(n-1)(n-2)$ permutations of triads with distinct elements (i_1, i_3, i_4) from $\{1, 2, \cdots, n\}$

The function $\text{Rank}_{(\mathbf{i}_{3,4})}\left[(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}})'b\right]$ denotes the rank of agent i_1's within-individual variation of the linear index $\left[(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}})'b\right]$, among the remaining $(n-3)$ within-individual variations for agents i_2 other than i_1 within dyads (i_2, i_3) and (i_2, i_4). That is

$$\text{Rank}_{(\mathbf{i}_{3,4})} \left[(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}})'b\right] \equiv \sum_{i_2 \in \mathcal{N}_n \setminus \{i_1, i_3, i_4\}} \mathbf{1}\left\{(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}})'b \geq (x_{\mathbf{i}_{2,3}} - x_{\mathbf{i}_{2,4}})'b\right\}.$$

5.2 Finite Sample Performance

In this section, I study the finite sample properties of the pairwise difference semiparametric estimator introduced in section 4. I compare the performance of this estimator with the Tetrad Logit (TL) estimator introduced in Graham (2014). I consider two setups with different distributional assumptions on the link-specific disturbance term. In the first setup, the link-specific disturbance term has a logistic distribution. Under this setup, the TL estimator is correctly specified for the network formation model. In the second setup, the link-specific disturbance term has a standard normal distribution. Under this setup, the network formation model studied in Graham (2014) is not correctly specified.

I simulate the data from the network formation model in (1) with the following true DGP value for the unknown parameter

$$\beta_0 = [1, 1.5, -1.5]'.$$

I assume the vector of dyad-specific attributes X_{ij} is

$$X_{ij} = \left[X_{ij}^{(1)}, X_{ij}^{(2)}, X_{ij}^{(3)}\right], \quad \text{with} \quad X_{ij}^{(l)} = z_{il} z_{jl} \quad \text{for} \quad l = 1, 2, 3.$$

The agent-specific observed attributes z_{i1}, z_{i2}, z_{i3} are draw as follows:

$$\begin{aligned} z_{i1} \sim & \mathcal{N}(0, 3), \\ z_{i2} \sim & \text{Unif}\{-1, 1\} \text{ with } P(z_{i2} = -1) = P(z_{i2} = 0) = P(z_{i2} = 1) = 1/3, \\ z_{i3} \sim & \text{Unif}(-2, 2). \end{aligned}$$

I assume the latent agent-specific fixed effects are $\alpha_i = \frac{\lambda}{3}(z_{i1} + z_{i2} + z_{i3}) + (1 - \lambda)W$, where $W \sim \mathcal{N}(0, 1)$ and $\lambda \in \{0.25, 0.5, 0.75\}$. The parameter λ measures the degree of correlation between the agent-specific observed covariates and unobserved fixed effects. The bounded fixed effects are specified as:

$$\mu_i = \begin{cases} B_L & \text{if} \quad \alpha_i < B_L \\ \alpha_i & \text{if} \quad B_L \leq \alpha_i \leq B_U \\ B_U & \text{if} \quad B_U < \alpha_i \end{cases},$$

with $B_L = B_U = 1$.

The dyad-specific disturbance term is draw from two distribution. Specifically, I consider $\varepsilon_{ij}^{(1)} \sim$ Logistic$(0, 1)$, and $\varepsilon_{ij}^{(2)} \sim \mathcal{N}(0, 2)$.

The next tables report the estimates of β_0 obtained from 500 Monte Carlo simulations with correlation parameter $\lambda = 0.5$ and sample sizes $n = 100, 250$, and 500. I report the Monte Carlo simulations for λ equal to 0.25 and 0.75 in appendix E. The results are qualitatively similar.

Tables 1 and 2 report the finite sample properties of the pairwise difference and the TL estimators under the first and second setup, respectively. I report the scaled-normalized estimates $\beta_0/\beta_{0,1}$ for both estimators, which is consistent with assumption A2 and the definition of the parameter space in (19). I focus on the median, mean, bias in percentage points, and the root mean square error (RMSE) for each estimator. Tables 1 and 2 also report the probability of the class Ω_n and the average degree of the network.

Table 1 shows that under the logistic design and in a small network with 100 individuals, the estimation bias for the coefficient associated with the discrete covariate $\beta_{0,2}$ is approximately the same for both estimators, 5.9% for the pairwise difference estimator and 4.1% for the TL estimator. The performance of the pairwise difference estimator is as good as to the one of the TL for the coefficient associated with the continuous covariate regarding bias and RMSE.

In the larger network with 250 individuals. The performance of both estimators improves. Specifically, the estimation bias for $\beta_{0,2}$ of the pairwise difference estimator decreases significantly to 1.02%. The bias for $\beta_{0,3}$ remains small. The performance of the pairwise difference estimator regarding the RMSE also improves. I also report the estimates of the pairwise difference estimator for a larger network with size $n = 500$.[7] These results confirm the good asymptotic performance. Specifically, the bias and RMSE of the pairwise difference estimator become negligible as the network size increases. In additional tests, I have estimated the network formation model with sample sizes of 1000 and 2000, and the results are qualitatively similar. Notably, the pairwise difference estimator performs as well as TL when the model is link-specific disturbance terms are correctly specified as logistic.

Table 2 shows that under the standard normal design, the TL estimator for the continuous covariate can present a bias of 15% in a small network of size $n = 100$. This bias decreases to 13% in a larger network of size $n = 250$, yet it fails to disappear. These results suggest that the performance of the TL is undermined when the distribution of the dyad-disturbance term is misspecified. The pairwise difference estimate for the coefficient associated with discrete covariate presents a bias of 5.7% in a small network of size $n = 100$. However, this bias decreases to 5% and 4.7% as the network increases to $n = 250$ and $n = 500$. The pairwise difference estimator for $\beta_{0,3}$ presents a bias of 13% for a network with $n = 100$, which decreases to 7.2% and 5.2% as the network size increases. The pairwise difference estimator also has a good performs regarding RMSE.

The numerical evidence in tables 1 and 2 suggest that the pairwise difference estimator has a good performance in finite samples, independently of the distribution of the dyad-specific unobserved components. Furthermore, its properties improve considerably as network size increases. Finally, the TL can suffer a nonzero bias when the dyad-specific disturbance terms is different from logistic.

[7] I do not report the estimates of the TL estimator for a network with sample size $n = 500$ due to its computational complexity. Initial tests suggest that computing the TL estimator, in Matlab, for a network with 500 nodes requires a computing cluster with more than 100 gigabytes of memory. This challenge could be overcome by using other computing languages such as C++ or Python.

Table 1: Monte Carlo Simulations: Logistic(0,1)

	Pairwise Difference				Tetrad Logit				$P(\Omega_n)$	E(Degree)
	Median	Mean	Bias(%)	RMSE	Median	Mean	Bias(%)	RMSE		
$N = 100$									5.924%	47.134
$\beta_{0,2}/\beta_{0,1} = 1.5$	1.495	1.589	5.968	1.519	1.504	1.562	4.172	0.369		
$\beta_{0,3}/\beta_{0,1} = -1.5$	-1.494	-1.493	0.463	0.184	-1.731	-1.773	18.261	0.390		
$N = 250$									8.848%	118.799
$\beta_{0,2}/\beta_{0,1} = 1.5$	1.499	1.484	1.024	0.164	1.493	1.528	1.887	0.270		
$\beta_{0,3}/\beta_{0,1} = -1.5$	-1.486	-1.487	0.854	0.066	-1.680	-1.697	13.135	0.284		
$N = 500$									8.243%	236.443
$\beta_{0,2}/\beta_{0,1} = 1.5$	1.513	1.508	0.587	0.034						
$\beta_{0,3}/\beta_{0,1} = -1.5$	-1.504	-1.501	0.076	0.030						

Note: Number of Monte Carlo simulations M=500, correlation parameter $\lambda = 0.5$

Table 2: Monte Carlo Simulations: Normal(0,2)

	Pairwise Difference				Tetrad Logit				$P(\Omega_n)$	E(Degree)
	Median	Mean	Bias(%)	RMSE	Median	Mean	Bias(%)	RMSE		
$N = 100$									7.914%	47.125
$\beta_{0,2}/\beta_{0,1} = 1.5$	1.630	1.585	5.715	0.727	1.651	1.665	7.454	0.437		
$\beta_{0,3}/\beta_{0,1} = -1.5$	-1.734	-1.702	13.613	1.836	-1.735	-1.763	15.712	0.438		
$N = 250$									7.346%	117.700
$\beta_{0,2}/\beta_{0,1} = 1.5$	1.567	1.551	5.061	0.886	1.524	1.512	4.133	0.325		
$\beta_{0,3}/\beta_{0,1} = -1.5$	-1.677	-1.632	7.245	1.074	-1.691	-1.674	13.128	0.325		
$N = 500$									7.148%	236.301
$\beta_{0,2}/\beta_{0,1} = 1.5$	1.529	1.542	4.761	0.881						
$\beta_{0,3}/\beta_{0,1} = -1.5$	-1.572	-1.553	5.281	0.801						

Note: Number of Monte Carlo simulations M=500, correlation parameter $\lambda = 0.5$

6 Empirical Application

In this section, I use the methods developed in this paper to estimate the network formation model in equation (1) on a friendships network among high school students. The objective is to estimate the preference parameters associated with socio-demographic and educational factors. From an empirical perspective, these parameters represent the individuals' preferences towards homophily on observed attributes. I use the self-reported friendship links from the Add Health dataset (Harris, Halpern, Whitsel, Hussey, Tabor, Entzel, and Udry (2009)) to construct an undirected network of friends. I then the preference parameters using the pairwise difference estimator introduced in section 4.

6.1 Add Health dataset

The Add Health dataset is a national representative survey of adolescents in grades 7-12 in the United States during the 1994 to 1995 school year. This dataset has been designed to study the impact of the social environment; for example, friends, neighborhood, and school, on the adolescents' behavior. This survey is a longitudinal study collected in four waves of in-home interviews.[8] I use data from the Wave 1 in-home survey, which contains information on the total 90,118 participants and the friendships nominations.

These friendship data have been used to study the impact of social interactions on many different outcomes of interest, as in Bramoullé, Djebbari, and Fortin (2009); Calvó-Armengol, Patacchini, and Zenou (2009) and Christakis and Fowler (2008). This dataset has also been used to estimate network formation models on friendship relationships as in Christakis et al. (2010); Mele (2015), and Miyauchi (2016).

From the 132 schools in the sample, all the students enrolled in 16 high schools, known as saturated schools, were selected for in-home interviews. In these interviews, the students were asked to name up to five male and five female students.[9] The saturated high schools include two large schools with more than 700 enrolled students each and 14 small schools with less than 180 enrolled students each. I select one of the large saturated high schools for the empirical study.

A friendship link is formed if for any pair of students both agents named each other as friends, regardless of the order in which they do it. I use socio-demographic and educational factors to

[8] The Add Health website describes the data in detail, www.cpc.unc.edu/projects/addhealth.

[9] In all the remaining schools, the students have been asked to name only one male and one female friend.

model the formation of the friendships network. Specifically, I consider the household's income, the age, current academic grade, gender, race, overall GPA of the respondent, and the parent's level of education. Table 3 reports descriptive statistics for the exogenous covariates.

Household Income denotes the total income before taxes that the respondent's family perceived in the year 1994. This variable is recorded numerically, as opposed to being censored-coded. The minimum value for the household income in the sample is \$4,000, and the maximum value \$200,000. Female is a gender dummy variable that indicates if the respondent is a female. Grade denotes the current academic grade of the student. In this sample, this variable includes from 9 to 12 grade. Hispanic, White, Black, Asian, Indian, and Other Races are dummy variables that indicate the respondent's ethnicity. The high school considered is predominantly white with approximately 94% of the students being white. The variable Overall GPA is constructed as a sample average of the student's grades in English, History, Mathematics, and Science courses. Mother's Education and Father's Education are coded as 0 = never went to school, 1 = 8 grade or less, 2 = above 8 grade but not a high school graduate, 3 = professional training instead of high school, 4= high school graduate, 5 = GED, 6 = professional training after high school, 7 = attended college but did not graduate, 8 = college graduate, 9 = professional training after college.

I transform the covariates Household Income, Age, Grade, Overall GPA, Mother's education, and Father's education by subtracting their mean. Household Income is used as the covariate with large support, which after the transformation has a minimum value of -\$47,000 and a maximum value of \$148,000 in the sample. Although the support of Household Income is not unbounded, it is sufficient to contain the support of the remaining covariates as discussed in the point identification result with one continuous regressor with bounded support, section 3.4.

After dropping missing observations for age and household's income, the total number of observations in the sample is $n = 469$. In total 319 students named at least one friend. The probability of the class Ω_n in the sample is 2.24%. The probability of Ω_n in the remaining large high school is 2.59%. Although the probability of Ω_n is larger in the other high school, that high school was not selected because it suffers from a more severe missing observation problem.

6.2 Empirical Results

I estimate the network formation model in equation (1) using the pairwise difference estimator. I compare these results with the point estimates obtained from computing Graham (2017) Tetrad

Logit estimator and a naive logistic regression. The naive logistic regression ignores the presence of the fixed effects μ_i and μ_j in equation (1), and assumes ε_{ij} has a logistic distribution. I construct the dyad-level observed attributes as in the Monte Carlo designs. I use the socio-demographic and educational covariates described in Table 3.

There is a total of 319 students that named at least one friend in the sample, these students form 50,761 total unique dyads. The probability of the identifying class Ω_n in the sample is 2.24%. The average number of friends named by each student is 3.62. The total sample used by the logistic regression to estimate the preference parameters is equal to the total number of unique dyads. In contrast, the actual sample size used by the pairwise difference estimator is composed by 2.24% of all the tetrads.

The naive logistic regression suggests a negative homophily effect on the formation of friendship links of the covariates: age, female gender, white race, and overall GPA. In addition, it captures a positive homophily effect of the mother's and father's education level on the formation of friendships. This estimator drops the indicators for Asian and Indian races because the small number of observations in these covariates generates a close to perfect collinearity problem in the logistic regression at a dyad-level.

Notably, the pairwise difference estimator and Graham's Tetrad Logit predict opposite signs for the parameters associated with the covariates: female gender, white race, overall GPA, and Mother's education. This insight suggests that the estimates obtained with the naive logistic regression are biased due to the omission of the fixed effects. Specifically, the logistic regression underestimates the preferences for homophily on the covariates: female gender, white gender, and overall GPA. Furthermore, it overestimates the preferences for homophily on the mother's education.

Both estimators, the pairwise difference and the Tetrad Logit indicate positive preferences for homophily on the covariates: current academic grade, Hispanic race, White race, and overall GPA. Similarly, both estimators imply preferences for heterogeneity in Asian race and Mother's education. Distinctively, the pairwise difference estimator also predicts strong preferences for homophily on Female gender and Father's Education. These results imply the presence of strong homophily effects among Female, Hispanic, and White students, as well as among students within the same academic grade and with high level of academic performance. Finally, the empirical results suggest that the level of Father's Education is more important for the formation of a friendships network among High School students than the level of the Mother's education.

Table 3: Descriptive Statistics

Variable	Count	Mean	Std. Dev.	Min	Max
Household Income	24109	51.405	29.68	4	200
Age	7367	15.707	1.183	14	19
Female	676	0.441	0.497	0	1
Grade	4810	10.255	1.085	9	12
Hispanic	12	0.025	0.150	0	1
White	442	0.942	0.233	0	1
Black	3	0.006	0.079	0	1
Asian	7	0.014	0.121	0	1
Indian	14	0.029	0.170	0	1
Other races	17	0.036	0.187	0	1
Overall GPA	1100	2.346	0.956	0	4
Mother's Education	1989	4.240	2.419	0	9
Father's Education	1945	4.147	2.794	0	9
Sample size = 469.					

Table 4: Estimation Results

	Logistic	Pairwise Difference	Graham (2015)
Age	−1.245***	−0.826	−1.088
Female	−1.875***	0.635**	0.032
Grade	0.764***	1.264**	0.553*
Hispanic	0.772	1.322***	1.100***
White	−3.758***	1.661**	1.544***
Black		0.382	0.085
Asian		−1.172**	−1.491**
Indian	−0.597	−0.318	−0.742
Other races	−0.461	−0.553	−1.061
Overall GPA	−0.102***	2.436**	2.350**
Mother's Education	0.276***	−0.352*	−0.615*
Father's Education	0.240***	1.549***	0.748

$P(\Omega_n) = 2.24\%$
Average Degree = 3.62.
Number of Students = 319.
Number of dyads = 50,721.

*,**,*** represents the significant at 10%, 5%, and 1% level.
Observations with any missing data are dropped.

7 Conclusion

In this paper, I have studied a network formation model with multiple additive fixed effects. I propose a new identification strategy that point identifies the vector of coefficients on the observed covariates, which accounts for observed homophily. This result relies on the existence of at least one continuous covariate with large support. Under a weaker set of assumptions, I show that point identification can still be obtained if at least one continuous covariate exists. If all the covariates have bounded and discrete support, I derive bounds for each component of the vector of coefficients. Under the assumptions that guarantee point identification, I introduce a semiparametric estimator and show in Monte Carlo simulations that it performs well in finite samples.

As an extension to the network formation model in equation (1), I study the formation of a directed network with interactive fixed effects in the accompanying paper Candelaria (2016). Specifically, a directed link is formed according to the equation

$$D_{ij}^{n} = \mathbf{1}\left[X_{ij}^{n\prime}\beta_0 + \mu_i + g(\mu_i, \mu_j) - \varepsilon_{ij}^{n} \geq 0\right], \tag{24}$$

where $g(\cdot,\cdot)$ is a symmetric function of the unobserved fixed effects μ_i, μ_j.

Under the specification in (24), individuals create connections based on both homophily on observed and unobserved characteristics. Furthermore, the agent-specific fixed effect μ_j may affect the linking decisions of individual i, differently for different agents j due to the unobserved complementarities on the fixed effects. Disentangling the effect of homophily on observed from unobserved attributes is empirically relevant from a policy perspective.

The network formation model studied in this paper excludes network externalities. Network externalities generate interdependencies in the linking decisions that depend on the structure of the network. Recent papers that have studied network formation model with network externalities do not account for unobserved heterogeneity. This is an important extension that I leave for future research.

References

Abrevaya, J. (1999a). Computation of the maximum rank correlation estimator. *Economics letters 62*(3), 279–285.

Abrevaya, J. (1999b). Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics 93*(2), 203–228.

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society. Series B (Methodological)*, 283–301.

Andrews, D. W. and M. M. Schafgans (1998). Semiparametric estimation of the intercept of a sample selection model. *The Review of Economic Studies 65*(3), 497–517.

Arellano, M. and B. Honoré (2001). Panel data models: some recent developments. *Handbook of Econometrics 5*, 3229–3296.

Badev, A. (2014). Discrete games in endogenous networks: Theory and policy. *Working Paper*.

Boucher, V. and I. Mourifié (2013). My friend far far away: Asymptotic properties of pairwise stable networks. *Working Paper*.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *The Review of Economic Studies 76*(4), 1239–1267.

Candelaria, L. E. (2016). Network formation models with interactive fixed effects. *Working Paper*.

Chamberlain, G. (2010). Binary response models for panel data: Identification and information. *Econometrica 78*(1), 159–168.

Chandrasekhar, A. G. and M. O. Jackson (2014). Tractable and consistent random graph models. *Working Paper*.

Charbonneau, K. (2014). Multiple fixed effects in nonlinear panel data models. *Working Paper*.

Choo, E. and A. Siow (2006). Who marries whom and why. *Journal of Political Economy 114*(1), 175–201.

Christakis, N. A. and J. H. Fowler (2008). The collective dynamics of smoking in a large social network. *New England Journal of Medicine 358*(21), 2249–2258.

Christakis, N. A., J. H. Fowler, G. W. Imbens, and K. Kalyanaraman (2010). An empirical model for strategic network formation. *Working Paper*.

de Paula, A., S. Richards-Shubik, and E. T. Tamer (2016). Identification of preferences in network formation games. *Working Paper*.

Dzemski, A. (2014). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Working Paper*.

Fox, J. T. (2010). Identification in matching games. *Quantitative Economics 1*(2), 203–254.

Fox, J. T. (2016). Estimating matching games with transfers. *Working Paper*.

Galichon, A. and B. Salanié (2012). Cupid's Invisible Hand: Social Surplus and Identification in Matching Models. *Working Paper*.

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*(3), 253–264.

Graham, B. S. (2014). Methods of identification in social networks.

Graham, B. S. (2017). An econometric model of link formation with degree heterogeneity. *Working Paper*.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics 35*(2), 303–316.

Harris, K. M., C. T. Halpern, E. Whitsel, J. Hussey, J. Tabor, P. Entzel, and J. R. Udry (2009). The national longitudinal study of adolescent health: Research design. *WWW document*.

Honoré, B. E. and L. Hu (2015). Poor (wo)man's bootstrap. *Working Paper*.

Horowitz, J. L. (2012). *Semiparametric methods in econometrics*, Volume 131. Springer Science & Business Media.

Hsieh, C.-S. and L.-F. Lee (2012). A structural modeling approach for network formation and social interactions with applications to students' friendship choices and selectivity on activities. *Working Paper*.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of economic theory 71*(1), 44–74.

Khan, S. and E. Tamer (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica 78*(6), 2021–2042.

Kline, B. (2015). Identification of complete information games. *Journal of Econometrics 189*(1), 117–131.

Komarova, T. (2013). Binary choice models with discrete regressors: Identification and misspecification. *Journal of Econometrics 177*(1), 14–33.

Leung, M. (2015a). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics 188*(1), 182–195.

Leung, M. (2015b). A weak law for moments of pairwise-stable networks. *Working Paper*.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*(3), 205–228.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, 357–362.

Manski, C. F. (1988). Identification of binary response models. *Journal of the American Statistical Association 83*(403), 729–738.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 415–444.

Mele, A. (2015). A structural model of segregation in social networks. *Working Paper*.

Menzel, K. (2015). Stategic network formation with many agents. *Working Paper*.

Miyauchi, Y. (2016). Structural estimation of a pairwise stable network with nonnegative externality. *Journal of Econometrics, Forthcoming*.

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics 5*(2), 99–135.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics 4*, 2111–2245.

Powell, J. L. (1994). Estimation of semiparametric models. *Handbook of econometrics 4*, 2443–2521.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*, Volume 162. John Wiley & Sons.

Sheng, S. (2012). Identification and estimation of network formation games. *Working Paper*.

Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica*, 123–137.

Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *The Annals of Statistics*, 439–459.

Snijders, T. A., J. Koskinen, and M. Schweinberger (2010). Maximum likelihood estimation for social network dynamics. *The Annals of Applied Statistics 4*(2), 567.

Subbotin, V. (2007). Asymptotic and bootstrap properties of rank regressions. *Working Paper*.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies 70*(1), 147–165.

Towns, J., T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. R. Scott, and N. Wilkins-Diehr (2014, Sept.-Oct). XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering 16*(5), 62–74.

Vytlacil, E. and N. Yildiz (2007). Dummy endogenous variables in weakly separable models. *Econometrica 75*(3), 757–779.

Data References

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (*http : //www.cpc.unc.edu /addhealth*). No direct support was received from grant P01-HD31921 for this analysis.

A Proofs

B Proofs

B.1 Point Identification

The following two lemmas formalize the intuition behind the identification strategy and they are used to prove Theorem 3.1.

Lemma 1 formalizes the intuition of Diagram 2 in Figure 1. That is, a pairwise difference between the net difference of the decisions linking individuals i and j cancels out the agent-specific fixed effects. Lemma 2 specifies a median condition for the random variable $Y_{kl}^{(i)} - Y_{kl}^{(j)}\Big| \mathbf{X}^n, \Omega(ijlk)$.

Lemma 1
Let assumption A1-A3 hold. Then for any n, and any different $i, j, k, l \in \mathcal{N}_n$ the following condition holds:

$$\text{sign}\left\{\left[\Delta_{kl}x_i - \Delta_{kl}x_j\right]\beta_0\right\} = \text{sign}\left\{\mathbb{E}\left[(D_{ik} - D_{il}) - (D_{jk} - D_{jl})|\mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\right]\right\}$$

where $\mathbf{x} \in \mathcal{X}_B$*,*

$$\mathcal{X}_B = \{\mathbf{x} \in \mathbf{X}^n : \textit{ for any } i, j, k, l \in \mathcal{N}_n, \mid \Delta_{kl}x_i\beta_0 \mid \geq 2B, \textit{ and } \\ \text{sign}\left\{\Delta_{kl}x_i\beta_0\right\} + \text{sign}\left\{\Delta_{kl}x_i\beta_0\right\} = 0\}$$

and for B defined in A2.

Preliminaries:

Let

$$w_{ik}(\beta_0) = x_{ik}\beta_0 + \mu_i + \mu_k$$

for any $(i,k) \in \mathcal{N}_n^{(2)}$.

Let $\mathbb{E}\left[Z|x,\mu,\Omega\right]$, denote the conditional expectation of any random variable Z given $\{\mathbf{X}^n = x, \tilde{\mu} = \mu, \Omega(ijlk)\}$, i.e.

$$\mathbb{E}\left[Z|\mathbf{X}^n = x, \tilde{\mu} = \mu, \Omega(ijlk)\right].$$

Note that

$$\begin{aligned}
&\mathbb{E}\left[(D_{ik} - D_{il}) - (D_{jk} - D_{jl})|x,\mu,\Omega\right] \\
&= 2\left[\mathbb{P}\left[(D_{ik} - D_{il}) - (D_{jk} - D_{jl}) = 2|x,\mu,\Omega\right]\right. \\
&\left. - \mathbb{P}\left[(D_{ik} - D_{il}) - (D_{jk} - D_{jl}) = -2|x,\mu,\Omega\right]\right] \\
&= 2\left[\mathbb{P}\left[D_{ik} = 1, D_{il} = 0, D_{jl} = 1, D_{jk} = 0|x,\mu,\Omega\right]\right. \\
&\left. - \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jl} = 0, D_{jk} = 1|x,\mu,\Omega\right]\right] \\
&= \frac{2}{\mathbb{P}\left[\Omega|x,\mu\right]}\left[\mathbb{P}\left[D_{ik} = 1, D_{il} = 0, D_{jl} = 1, D_{jk} = 0|x,\mu\right]\right. \\
&\left. - \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jl} = 0, D_{jk} = 1|x,\mu\right]\right]
\end{aligned}$$

Then,

$$\mathbb{E}\left[(D_{ik} - D_{il}) - (D_{jk} - D_{jl})|x,\mu,\Omega\right] \geq 0$$

if and only if

$$\begin{aligned}
&\mathbb{P}\left[D_{ik} = 1, D_{jl} = 1, D_{il} = 0, D_{jk} = 0|x,\mu\right] \\
&\geq \\
&\mathbb{P}\left[D_{il} = 1, D_{jk} = 1, D_{ik} = 0, D_{jl} = 0|x,\mu\right]
\end{aligned}$$

Where

$$\mathbb{P}\left[D_{ik}=1,D_{jl}=1,D_{il}=0,D_{jk}=0|x,\mu\right]=$$
$$\int_{-\infty}^{w_{ik}(\beta_0)}\int_{-\infty}^{w_{jl}(\beta_0)}\int_{w_{il}(\beta_0)}^{\infty}\int_{w_{jk}(\beta_0)}^{\infty} f_{\varepsilon|\mathbf{x},\mu}(s_1)f_{\varepsilon|\mathbf{x},\mu}(s_2)f_{\varepsilon|\mathbf{x},\mu}(s_3)f_{\varepsilon|\mathbf{x},\mu}(s_4)ds_1ds_2ds_3ds_4$$

and

$$\mathbb{P}\left[D_{il}=1,D_{jk}=1,D_{ik}=0,D_{jl}=0|x,\mu\right]=$$
$$\int_{-\infty}^{w_{il}(\beta_0)}\int_{-\infty}^{w_{jk}(\beta_0)}\int_{w_{ik}(\beta_0)}^{\infty}\int_{w_{jl}(\beta_0)}^{\infty} f_{\varepsilon|\mathbf{x},\mu}(s_1)f_{\varepsilon|\mathbf{x},\mu}(s_2)f_{\varepsilon|\mathbf{x},\mu}(s_3)f_{\varepsilon|\mathbf{x},\mu}(s_4)ds_1ds_2ds_3ds_4$$

For any $\mathbf{x}\in\mathcal{X}_B$, with

$$\mathcal{X}_B=\{\mathbf{x}\in\mathbf{X}^n:\text{ for any } i,j,k,l\in\mathcal{N}_n,\ \mid\Delta_{kl}x_i\beta_0\mid\geq 2B,\text{ and }\\ \operatorname{sign}\{\Delta_{kl}x_i\beta_0\}+\operatorname{sign}\{\Delta_{kl}x_i\beta_0\}=0\}$$

then:

$$\begin{aligned}\mid\Delta_{kl}x_i\beta_0\mid &\geq 2B\\ \mid\Delta_{kl}x_j\beta_0\mid &\geq 2B,\end{aligned}\tag{25}$$

and

$$\operatorname{sign}\{\Delta_{kl}x_i\beta_0\}+\operatorname{sign}\{\Delta_{kl}x_i\beta_0\}=0.\tag{26}$$

Two mutually exclusive cases are possible,

$$\{\Delta_{kl}x_i\beta_0>0,\Delta_{kl}x_j\beta_0<0\}$$

or

$$\{\Delta_{kl}x_i\beta_0<0,\Delta_{kl}x_j\beta_0>0\}.$$

Assume the first case to be true, that is $\Delta_{kl}x_i\beta_0>0,\Delta_{kl}x_j\beta_0<0$. Given that $\mathbf{x}\in\mathcal{X}_B$, then the conditions in (25) imply:

$$\begin{aligned} \Delta_{kl} x_i \beta_0 &\geq 2B \\ -\Delta_{kl} x_j \beta_0 &\geq 2B. \end{aligned}$$

Then by A3, it follows that:

$$\Delta_{kl} x_i \beta_0 \geq \quad 2B \geq (\mu_l - \mu_k), \tag{27}$$

$$\Delta_{kl} x_j \beta_0 \leq -2B \leq (\mu_l - \mu_k). \tag{28}$$

Equation (27) can be equivalently written as:

$$\begin{gathered} x_{ik}\beta_0 + \mu_i + \mu_k \geq x_{il}\beta_0 + \mu_i + \mu_l \\ w_{ik}(\beta_0) \geq w_{il}(\beta_0) \end{gathered}$$

Equation (28) can be equivalently written as:

$$\begin{gathered} x_{jk}\beta_0 + \mu_j + \mu_k \leq x_{jl}\beta_0 + \mu_j + \mu_l \\ w_{jk}(\beta_0) \leq w_{jl}(\beta_0) \end{gathered}$$

Assume the second case to be true, that is $\Delta_{kl} x_i \beta_0 < 0, \Delta_{kl} x_j \beta_0 > 0$. Given that $\mathbf{x} \in \mathcal{X}_B$, then the conditions in (25) imply:

$$\begin{aligned} -\Delta_{kl} x_i \beta_0 &\geq 2B \\ \Delta_{kl} x_j \beta_0 &\geq 2B. \end{aligned}$$

Then by A3, it follows that:

$$\Delta_{kl} x_i \beta_0 \leq -2B \leq (\mu_l - \mu_k), \tag{29}$$

$$\Delta_{kl} x_j \beta_0 \geq \quad 2B \geq (\mu_l - \mu_k). \tag{30}$$

Equations (29) and (30) can be equivalently written as:

$$w_{ik}(\beta_0) \leq w_{il}(\beta_0)$$
$$w_{jk}(\beta_0) \geq w_{jl}(\beta_0).$$

Proof of Lemma 1:

Part I:

Fix $\mathbf{X}^n = \mathbf{x}$ and $\tilde{\mu} = \mu$ with $\mathbf{x} \in \mathcal{X}_B$, and suppose:

$$\text{sign}\left\{\left[\Delta_{kl}x_i - \Delta_{kl}x_j\right]\beta_0\right\} = 1,$$

that is:

$$\begin{aligned} \Delta_{kl}x_i - \Delta_{kl}x_j > 0 \\ \Leftrightarrow \\ \Delta_{kl}x_i + (\mu_k - \mu_l) - \Delta_{kl}x_j - (\mu_k - \mu_l) > 0 \\ \Leftrightarrow \\ \Delta_{kl}x_i + (\mu_k - \mu_l) > 0 \text{ and } \Delta_{kl}x_j + (\mu_k - \mu_l) < 0 \\ \Leftrightarrow \\ w_{ik}(\beta_0) > w_{il}(\beta_0) \text{ and } w_{jk}(\beta_0) < w_{jl}(\beta_0) \end{aligned}$$

The first and second equivalences follow from the definition of $\mathcal{X}_B$; and the third one from the definition of $w_{jk}(\beta_0)$.

Then, by A1, these conditions imply:

$$\int_{-\infty}^{w_{ik}(\beta_0)} f_{\varepsilon|\mathbf{x},\mu}(s_1)ds_1 > \int_{-\infty}^{w_{il}(\beta_0)} f_{\varepsilon|\mathbf{x},\mu}(s_1)ds_1$$
$$\int_{-\infty}^{w_{jl}(\beta_0)} f_{\varepsilon|\mathbf{x},\mu}(s_2)ds_2 > \int_{-\infty}^{w_{jk}(\beta_0)} f_{\varepsilon|\mathbf{x},\mu}(s_2)ds_2,$$

which are sufficient conditions for

$$\begin{aligned}
\mathbb{P}\left[D_{ik}=1, D_{jl}=1, D_{il}=0, D_{jk}=0|x,\mu\right] \\
> \\
\mathbb{P}\left[D_{il}=1, D_{jk}=1, D_{ik}=0, D_{jl}=0|x,\mu\right],
\end{aligned}$$

and therefore for

$$\mathbb{E}\left[(D_{ik}-D_{il})-(D_{jk}-D_{jl})|x,\mu,\Omega(ijlk)\right]>0,$$

for any $\mathbf{X}^n=\mathbf{x}$ and $\tilde{\mu}=\mu$ with $\mathbf{x}\in\mathcal{X}_B$.

Part II: Fix $\mathbf{X}^n=\mathbf{x}$ and $\tilde{\mu}=\mu$ with $\mathbf{x}\in\mathcal{X}_B$, and suppose:

$$\mathbb{E}\left[(D_{ik}-D_{il})-(D_{jk}-D_{jl})|x,\mu,\Omega(ijlk)\right]>0.$$

Given that $\mathbf{x}\in\mathcal{X}_B$, then either

$$\{\Delta_{kl}x_i\beta_0>0, \Delta_{kl}x_j\beta_0<0\}$$

or

$$\{\Delta_{kl}x_i\beta_0<0, \Delta_{kl}x_j\beta_0>0\}$$

is true. However, note that if

$$\{\Delta_{kl}x_i\beta_0>0, \Delta_{kl}x_j\beta_0<0\}$$

is true, then:

$$\begin{aligned}
w_{ik}(\beta_0) &\geq w_{il}(\beta_0) \\
w_{jk}(\beta_0) &\leq w_{jl}(\beta_0).
\end{aligned}$$

Alternatively, if

$$\{\Delta_{kl}x_i\beta_0<0, \Delta_{kl}x_j\beta_0>0\},$$

is true, then:

$$\begin{aligned}
w_{ik}(\beta_0) &\leq w_{il}(\beta_0) \\
w_{jk}(\beta_0) &\geq w_{jl}(\beta_0).
\end{aligned}$$

By hypothesis,

$$\mathbb{E}\left[(D_{ik} - D_{il}) - (D_{jk} - D_{jl})|x, \mu, \Omega(ijlk)\right] > 0$$

then is the case that $\{\Delta_{kl} x_i \beta_0 > 0, \Delta_{kl} x_j \beta_0 < 0\}$ is true, which implies:

$$\left[\Delta_{kl} x_i - \Delta_{kl} x_j\right] \beta_0 > 0,$$

and alternatively

$$\text{sign}\left\{\left[\Delta_{kl} x_i - \Delta_{kl} x_j\right] \beta_0\right\} = 1.$$

Similar arguments can be used to show that

$$\text{sign}\left\{\left[\Delta_{kl} x_i - \Delta_{kl} x_j\right] \beta_0\right\} = -1$$

if and only if

$$\text{sign}\left\{\mathbb{E}\left[(D_{ik} - D_{il}) - (D_{jk} - D_{jl})|x, \mu, \Omega(ijlk)\right]\right\} = -1,$$

For any $\mathbf{X}^n = \mathbf{x}$ and $\tilde{\mu} = \mu$ with $\mathbf{x} \in \mathcal{X}_B$ ■

Lemma 2
For any n, and any $i, j, k, l \in \mathcal{N}_n$,

$$\begin{aligned} \text{Med}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = x, \Omega(ijlk)\right] \\ = 2 \times \text{sign}\Big\{\mathbb{P}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} = 2|\mathbf{X}^n = x, \Omega(ijlk)\right] \\ - \mathbb{P}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} = -2|\mathbf{X}^n = x, \Omega(ijlk)\right]\Big\}. \end{aligned} \tag{31}$$

Proof. Note that $Y_{kl}^{(i)} - Y_{kl}^{(j)}|\left[\mathbf{X}^n = x, \Omega(ijlk)\right]$ is a Bernoulli random variable with support $\{-2, 2\}$. Let

$$\begin{aligned} p \equiv& \mathbb{P}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} = 2|\mathbf{X}^n = x, \Omega(ijlk)\right] \\ q \equiv& \mathbb{P}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} = -2|\mathbf{X}^n = x, \Omega(ijlk)\right] \end{aligned}$$

with $p+q=1$. Then,

$$\text{Med}\left(Y_{kl}^{(i)}-Y_{kl}^{(j)}|\mathbf{X}^n=x,\Omega(ijlk)\right)=\begin{cases}2 & \text{if } p\geq q\\ -2 & \text{otherwise}\end{cases}$$

the results follows from this observation. ■

Proof of Theorem 3.1

Part I

Note,

$$\begin{aligned}\mathbb{E}\left[(D_{ik}-D_{il})-(D_{jk}-D_{jl})|\mathbf{X}^n=\mathbf{x},\Omega(ijlk)\right]=\\ 2\mathbb{P}\left[Y_{kl}^{(i)}-Y_{kl}^{(j)}=2|\mathbf{X}^n=x,\Omega(ijlk)\right]-2\mathbb{P}\left[Y_{kl}^{(i)}-Y_{kl}^{(j)}=-2|\mathbf{X}^n=x,\Omega(ijlk)\right]\\ 2\left\{\mathbb{P}\left[Y_{kl}^{(i)}-Y_{kl}^{(j)}=2|\mathbf{X}^n=x,\Omega(ijlk)\right]-\mathbb{P}\left[Y_{kl}^{(i)}-Y_{kl}^{(j)}=-2|\mathbf{X}^n=x,\Omega(ijlk)\right]\right\}\end{aligned}$$

Therefore:

$$\begin{aligned}\text{sign}\left\{\mathbb{E}\left[(D_{ik}-D_{il})-(D_{jk}-D_{jl})|\mathbf{X}^n=\mathbf{x},\Omega(ijlk)\right]\right\}=\\ \text{sign}\left\{\mathbb{P}\left[Y_{kl}^{(i)}-Y_{kl}^{(j)}=2|\mathbf{X}^n=x,\Omega(ijlk)\right]-\mathbb{P}\left[Y_{kl}^{(i)}-Y_{kl}^{(j)}=-2|\mathbf{X}^n=x,\Omega(ijlk)\right]\right\}\end{aligned}$$

Therefore,

$$\text{Med}\left[Y_{kl}^{(i)}-Y_{kl}^{(j)}|\mathbf{X}^n=x,\Omega(ijlk)\right]=2\times\text{sign}\left\{\left[\Delta_{kl}x_i-\Delta_{kl}x_j\right]\beta_0\right\}$$

for any $\mathbf{x}\in\mathcal{X}_B$.

Part II

Fix any $i,j,k,l\in\mathcal{N}_n$, by assumption A2, $(X_{sk}-X_{sl})$, for $s=i,j$, has everywhere positive density. Let $\Delta^2X\equiv[(X_{ik}-X_{il})-(X_{jk}-X_{jl})]$ has a continuous density with respect to the

Lebesgue measure on $\mathbb{R}^K$ given by

$$f_{\Delta^2 X}(x) = \int_{\mathbb{R}^K} f(w)f(x+w)dw$$

where f is the density function of the distribution of $(X_{sk} - X_{sl})$, for $s = i, j$

For any $b \in \mathbb{R}^K$ such that $b_1 \neq 0$ and $b \neq \beta_0$, we can find a set of values of $\Delta^2 X = x$ with positive measure such that sign $(xb) \neq (x\beta_0)$. In other words, let

$$\mathcal{X}_{(b)} \equiv \left[x \in \mathbb{R}^K : \text{sign}\,(xb) \neq (x\beta_0)\right],$$

then assumption A2 guarantees that

$$\int_{\mathcal{X}_b} f_{\Delta^2 X}(x)dx > 0$$

Therefore β_0 is identified up to scale. ■

B.2 Identification Failure: Thin Set

Proof of Theorem 3.2

Proof. Suppose $P\left(\Omega(ijlk)\right) = 0$, then the following event has measure zero:

$$\{D_{ik} \neq D_{il}, D_{jk} \neq D_{jl}, D_{ik} \neq D_{jk} : (i,j,k,l) \in P_{n,4}\}, \tag{32}$$

where $P_{n,4}$ stands for the collection of all permutation of 4 elements $\{i,j,k,l\}$ from $\{1, 2, \cdots, n\}$. Equivalently, the set $\Omega(i,j,k,l)$ has measure zero if for any tetrad, (i,j,k,l), at least one of the conditions in (32) holds as an equality with probability one.

If the event $\Omega(i,j,k,l)$ has measure zero, then the subnetwork formed by any tetrad (i,j,k,l), could be classified into one of the following five structures:

1. **Zero links are formed**, characterized by the following decisions:

$$D_{ik} = 0, D_{il} = 0, D_{jk} = 0, D_{jl} = 0 \quad \Rightarrow \quad Y_{kl}^{(i)} - Y_{kl}^{(j)} = 0$$

2. **One link is formed**, characterized by the following decisions:

$$\begin{aligned}
D_{ik}=1, D_{il}=0, D_{jk}=0, D_{jl}=0 \quad &\Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=1\\
D_{ik}=0, D_{il}=1, D_{jk}=0, D_{jl}=0 \quad &\Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=-1\\
D_{ik}=0, D_{il}=0, D_{jk}=1, D_{jl}=0 \quad &\Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=-1\\
D_{ik}=0, D_{il}=0, D_{jk}=0, D_{jl}=1 \quad &\Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=1
\end{aligned}$$

3. **Two links are formed**, characterized by the following decisions:

$$\begin{aligned}
D_{ik}=1, D_{il}=1, D_{jk}=0, D_{jl}=0 \quad &\Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=0\\
D_{ik}=0, D_{il}=0, D_{jk}=1, D_{jl}=1 \quad &\Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=0\\
D_{ik}=1, D_{il}=0, D_{jk}=1, D_{jl}=0 \quad &\Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=0\\
D_{ik}=0, D_{il}=1, D_{jk}=0, D_{jl}=1 \quad &\Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=0
\end{aligned}$$

4. **Three links are formed**, characterized by the following decisions:

$$\begin{aligned}
D_{ik}=1, D_{il}=1, D_{jk}=1, D_{jl}=0 \quad &\Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=-1\\
D_{ik}=1, D_{il}=1, D_{jk}=0, D_{jl}=1 \quad &\Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=1\\
D_{ik}=1, D_{il}=0, D_{jk}=1, D_{jl}=1 \quad &\Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=1\\
D_{ik}=0, D_{il}=1, D_{jk}=1, D_{jl}=1 \quad &\Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=-1
\end{aligned}$$

5. **Four links are formed**, characterized by the following decisions:

$$D_{ik}=1, D_{il}=1, D_{jk}=1, D_{jl}=1 \quad \Rightarrow \quad Y_{kl}^{(i)}-Y_{kl}^{(j)}=0$$

Let

$$\begin{aligned}
p_1(\beta_0,x,\mu) =&\mathbb{P}\left[Y_{kl}^{(i)}-Y_{kl}^{(j)}=-1|\mathbf{X}^n=x,\Omega(ijlk)^c\right],\\
p_2(\beta_0,x,\mu) =&\mathbb{P}\left[Y_{kl}^{(i)}-Y_{kl}^{(j)}=0|\mathbf{X}^n=x,\Omega(ijlk)^c\right],\\
p_3(\beta_0,x,\mu) =&\mathbb{P}\left[Y_{kl}^{(i)}-Y_{kl}^{(j)}=1|\mathbf{X}^n=x,\Omega(ijlk)^c\right],
\end{aligned}$$

where

$$\begin{aligned}
p_1(\beta_0, x, \mu) =& \mathbb{P}\left[D_{ik}=0, D_{il}=1, D_{jk}=0, D_{jl}=0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik}=0, D_{il}=0, D_{jk}=1, D_{jl}=0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik}=1, D_{il}=1, D_{jk}=1, D_{jl}=0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik}=0, D_{il}=1, D_{jk}=1, D_{jl}=1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right],
\end{aligned}$$

$$\begin{aligned}
p_2(\beta_0, x, \mu) =& \mathbb{P}\left[D_{ik}=1, D_{il}=1, D_{jk}=0, D_{jl}=0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik}=0, D_{il}=0, D_{jk}=1, D_{jl}=1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik}=1, D_{il}=0, D_{jk}=1, D_{jl}=0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik}=0, D_{il}=1, D_{jk}=0, D_{jl}=1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik}=0, D_{il}=0, D_{jk}=0, D_{jl}=0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik}=1, D_{il}=1, D_{jk}=1, D_{jl}=1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right],
\end{aligned}$$

$$\begin{aligned}
p_3(\beta_0, x, \mu) =& \mathbb{P}\left[D_{ik}=1, D_{il}=0, D_{jk}=0, D_{jl}=0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik}=0, D_{il}=0, D_{jk}=0, D_{jl}=1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik}=1, D_{il}=1, D_{jk}=0, D_{jl}=1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik}=1, D_{il}=0, D_{jk}=1, D_{jl}=1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right].
\end{aligned}$$

None of the previous network structures satisfies all the conditions in the event (32). As a consequence, the support of $\left\{Y_{kl}^{(i)} - Y_{kl}^{(j)} \mid \mathbf{x}, \Omega(ijlk)^c\right\}$ is $\{-1, 0, 1\}$, where $\Omega(ijlk)^c$ stands for the complement of $\Omega(ijlk)$. Therefore, the pairwise difference is no longer a Bernoulli random variable with support $\{-2, 2\}$, as is the case when $\Omega(ijlk)$ has positive measure.

Since the support of $\left\{Y_{kl}^{(i)} - Y_{kl}^{(j)} \mid \mathbf{x}, \Omega(ijlk)^c\right\}$ is not equal to $\{-2, 2\}$, Lemma 2 no longer holds. Specifically,

$$\begin{aligned}
&\text{Med}\left[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&= \mathbf{1}\left\{0.5 - (p_1(\beta_0, x, \mu) + p_2(\beta_0, x, \mu)) \geq 0\right\} - \mathbf{1}\left\{p_1(\beta_0, x, \mu) - 0.5 \geq 0\right\} \\
&\neq 2 \times \text{sign}\left\{\mathbb{P}\left[\Delta Y_{ij} = 2 | \mathbf{X}^n = x, \Omega(ijlk)\right] - \mathbb{P}\left[\Delta Y_{ij} = -2 | \mathbf{X}^n = x, \Omega(ijlk)\right]\right\},
\end{aligned}$$

where $\Delta Y_{ij} = Y_{kl}^{(i)} - Y_{kl}^{(j)}$.

Since Lemma 2 fails, equation (7) is misspecified and does not have identification power. Therefore, the conclusion in Theorem 3.1 is wrong.

Now, I show that the Med $[\Delta Y_{ij}|\mathbf{X}^n = x, \Omega(ijlk)^c]$ does not have power to identify β_0. Given assumption A2, for any i, j, k, l with $k \neq l$, the random variables

$$X_{ik}^{(1)} - X_{il}^{(1)} \quad \& \quad X_{jk}^{(1)} - X_{jl}^{(1)}$$

have large support conditional on $\Delta\tilde{\mathbf{x}}_i$, $\Delta\tilde{\mathbf{x}}_j$, respectively, almost everywhere with respect to the Lebesgue measure. Thus, the following cases arise:

- **Case 1:** $X_{ik}^{(1)}$ and $X_{jk}^{(1)}$ have large support conditional on $\Delta\tilde{\mathbf{x}}_i$, $\Delta\tilde{\mathbf{x}}_j$, respectively.
- **Case 2:** $X_{ik}^{(1)}$ and $X_{jl}^{(1)}$ have large support conditional on $\Delta\tilde{\mathbf{x}}_i$, $\Delta\tilde{\mathbf{x}}_j$, respectively.
- **Case 3:** $X_{il}^{(1)}$ and $X_{jk}^{(1)}$ have large support conditional on $\Delta\tilde{\mathbf{x}}_i$, $\Delta\tilde{\mathbf{x}}_j$, respectively.
- **Case 4:** $X_{il}^{(1)}$ and $X_{jl}^{(1)}$ have large support conditional on $\Delta\tilde{\mathbf{x}}_i$, $\Delta\tilde{\mathbf{x}}_j$, respectively.
- **Case 5:** $X_{ik}^{(1)}$ and $X_{il}^{(1)}$ have large support conditional on $\Delta\tilde{\mathbf{x}}_i$ and $X_{js}^{(1)}$ has large support conditional on $\Delta\tilde{\mathbf{x}}_j$, for either $s = k, l$.
- **Case 6:** $X_{jk}^{(1)}$ and $X_{jl}^{(1)}$ have large support conditional on $\Delta\tilde{\mathbf{x}}_j$ and $X_{is}^{(1)}$ has large support conditional on $\Delta\tilde{\mathbf{x}}_i$, for either $s = k, l$.
- **Case 7:** $X_{is}^{(1)}$ and $X_{js}^{(1)}$ have large support conditional on $\Delta\tilde{\mathbf{x}}_i$, $\Delta\tilde{\mathbf{x}}_j$, respectively, for both $s = k, l$.

I. Suppose under the true model, characterized by β_0,

$$\text{Med}\,[\Delta Y_{ij}|\mathbf{X}^n = x, \Omega(ijlk)^c] = 1 \Leftrightarrow \{0.5 - (p_1(\beta_0, x, \mu) + p_2(\beta_0, x, \mu)) \geq 0\}\,.$$

Case 1: Consider $\tilde{\beta} \neq \beta_0$ and

$$\begin{aligned}
\mu_i + \mu_k &= \sup X_{ik}^{(1)} \Rightarrow D_{ik} = 1 \\
\mu_j + \mu_k &= \inf X_{jk}^{(1)} \Rightarrow D_{jk} = 0 \\
\mu_i + \mu_l &= \inf X_{il}'\theta, \text{ for any } \theta \in R^K \\
\mu_j + \mu_l &= \inf X_{jl}'\theta, \text{ for any } \theta \in R^K
\end{aligned}$$

Therefore,

$$\begin{aligned}
p_1(\tilde{\beta}, x, \mu) =& 0\\
p_2(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right]\\
p_3(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 1, D_{il} = 0, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right]\\
&+ \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right]
\end{aligned}$$

and

$$\begin{aligned}
p_2(\tilde{\beta}, x, \mu) = \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \leq\\
\mathbb{P}\left[D_{ik} = 1, D_{il} = 0, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \leq p_3(\tilde{\beta}, x, \mu)
\end{aligned}$$

Thus $p_3(\tilde{\beta}, x, \mu) \geq 0.5$ and $\text{Med}\left[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c\right] = 1$.

Case 2: Consider $\tilde{\beta} \neq \beta_0$ and

$$\begin{aligned}
\mu_i + \mu_k =& \sup X_{ik}^{(1)} \Rightarrow D_{ik} = 1,\\
\mu_j + \mu_k =& \inf X_{jl}'\theta, \text{ for any } \theta \in R^K,\\
\mu_i + \mu_l =& \inf X_{il}'\theta, \text{ for any } \theta \in R^K,\\
\mu_j + \mu_l =& \sup X_{jk}^{(1)} \Rightarrow D_{jl} = 1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
p_1(\tilde{\beta}, x, \mu) =& 0,\\
p_2(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 1, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right],\\
p_3(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right]\\
&+ \mathbb{P}\left[D_{ik} = 1, D_{il} = 0, D_{jk} = 1, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right],
\end{aligned}$$

and

$$p_2(\tilde{\beta}, x, \mu) \leq p_3(\tilde{\beta}, x, \mu)$$

Thus $p_3(\tilde{\beta}, x, \mu) \geq 0.5$ and $\text{Med}\left[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c\right] = 1$.

Case 3: Consider $\tilde{\beta} \neq \beta_0$ and

$$\begin{aligned}
\mu_i + \mu_k =& \sup X_{ik}'\theta, \text{ for any } \theta \in R^K, \\
\mu_j + \mu_k =& \inf X_{jk}^{(1)} \Rightarrow D_{jk} = 0, \\
\mu_i + \mu_l =& \inf X_{il}^{(1)} \Rightarrow D_{ik} = 0, \\
\mu_j + \mu_l =& \sup X_{jl}'\theta, \text{ for any } \theta \in R^K.
\end{aligned}$$

Therefore,

$$\begin{aligned}
p_1(\tilde{\beta}, x, \mu) =& 0, \\
p_2(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 0, D_{il} = 0, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right], \\
p_3(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 1, D_{il} = 0, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik} = 0, D_{il} = 0, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right],
\end{aligned}$$

and

$$p_2(\tilde{\beta}, x, \mu) \leq p_3(\tilde{\beta}, x, \mu).$$

Thus $p_3(\tilde{\beta}, x, \mu) \geq 0.5$ and $\text{Med}\left[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c\right] = 1$.

Case 4: Consider $\tilde{\beta} \neq \beta_0$ and

$$\begin{aligned}
\mu_i + \mu_k =& \sup X_{ik}'\theta, \text{ for any } \theta \in R^K, \\
\mu_j + \mu_k =& \sup X_{jk}'\theta, \text{ for any } \theta \in R^K, \\
\mu_i + \mu_l =& \inf X_{il}^{(1)} \Rightarrow D_{il} = 0, \\
\mu_j + \mu_l =& \sup X_{jl}^{(1)} \Rightarrow D_{jl} = 1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
p_1(\tilde{\beta}, x, \mu) =& 0, \\
p_2(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 0, D_{il} = 0, D_{jk} = 1, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right], \\
p_3(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 0, D_{il} = 0, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik} = 1, D_{il} = 0, D_{jk} = 1, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right],
\end{aligned}$$

and

$$p_2(\tilde{\beta}, x, \mu) \leq p_3(\tilde{\beta}, x, \mu).$$

Thus $p_3(\tilde{\beta}, x, \mu) \geq 0.5$ and $\operatorname{Med}[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c] = 1$.

Case 5: Consider $\tilde{\beta} \neq \beta_0$ and

$$\begin{aligned} \mu_i + \mu_k =& \sup X_{ik}^{(1)} \Rightarrow D_{ik} = 1, \\ \mu_i + \mu_l =& \inf X_{il}^{(1)} \Rightarrow D_{il} = 0. \end{aligned}$$

Therefore,

$$\begin{aligned} p_1(\tilde{\beta}, x, \mu) =& 0, \\ p_2(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 1, D_{il} = 0, D_{jk} = 1, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right], \\ p_3(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 1, D_{il} = 0, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\ &+ \mathbb{P}\left[D_{ik} = 1, D_{il} = 0, D_{jk} = 1, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right], \end{aligned}$$

Since $X_{js}^{(1)}$ has large support for either $s = k, l$. Depending on the case, choose:

$$\begin{aligned} \text{If } s = k: \quad & \mu_j + \mu_k = \inf X_{jk}^{(1)} \Rightarrow D_{jk} = 0, \\ \text{If } s = l: \quad & \mu_j + \mu_l = \sup X_{jl}^{(1)} \Rightarrow D_{jk} = 1. \end{aligned}$$

Therefore,

$$p_2(\tilde{\beta}, x, \mu) \leq p_3(\tilde{\beta}, x, \mu).$$

Thus $p_3(\tilde{\beta}, x, \mu) \geq 0.5$ and $\operatorname{Med}[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c] = 1$.

Case 6: This case is analogous to case 5.

Case 7: This is the easiest case, one example is:

$$\begin{aligned} \mu_i + \mu_k =& \sup X_{ik}^{(1)} \Rightarrow D_{ik} = 1, \\ \mu_j + \mu_k =& \inf X_{jk}^{(1)} \Rightarrow D_{jk} = 0, \\ \mu_i + \mu_l =& \inf X_{il}^{(1)} \Rightarrow D_{il} = 0, \\ \mu_j + \mu_l =& \inf X_{jl}^{(1)} \Rightarrow D_{jl} = 0. \end{aligned}$$

II. Suppose under the true model, characterized by β_0,

$$\begin{aligned}\text{Med}\left[\Delta Y_{ij}|\mathbf{X}^n = x, \Omega(ijlk)^c\right] = 0 \Leftrightarrow &\{0.5 - (p_1(\beta_0, x, \mu) + p_2(\beta_0, x, \mu)) \leq 0\} \\ &\cap \{p_1(\beta_0, x, \mu) - 0.5 \leq 0\}.\end{aligned}$$

Case 1: Consider $\tilde{\beta} \neq \beta_0$ and

$$\begin{aligned}\mu_i + \mu_k =& \inf X_{ik}^{(1)} \Rightarrow D_{ik} = 0 \\ \mu_j + \mu_k =& \inf X_{jk}^{(1)} \Rightarrow D_{jk} = 0 \\ \mu_i + \mu_l =& \sup X_{il}'\theta, \text{ for any } \theta \in R^K \\ \mu_j + \mu_l =& \sup X_{jl}'\theta, \text{ for any } \theta \in R^K\end{aligned}$$

Therefore,

$$\begin{aligned}p_1(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 0, D_{jl} = 0|\mathbf{X}^n = x, \Omega(ijlk)^c\right], \\ p_2(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 0, D_{jl} = 1|\mathbf{X}^n = x, \Omega(ijlk)^c\right] \\ &+ \mathbb{P}\left[D_{ik} = 0, D_{il} = 0, D_{jk} = 0, D_{jl} = 0|\mathbf{X}^n = x, \Omega(ijlk)^c\right], \\ p_3(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 0, D_{il} = 0, D_{jk} = 0, D_{jl} = 1|\mathbf{X}^n = x, \Omega(ijlk)^c\right].\end{aligned}$$

Note,

$$\begin{aligned}p_1(\tilde{\beta}, x, \mu) &= \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 0, D_{jl} = 0|\mathbf{X}^n = x, \Omega(ijlk)^c\right] \\ &\leq \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 0, D_{jl} = 1|\mathbf{X}^n = x, \Omega(ijlk)^c\right] \\ &\leq p_2(\tilde{\beta}, x, \mu)\end{aligned}$$

and

$$\begin{aligned}p_3(\tilde{\beta}, x, \mu) &= \mathbb{P}\left[D_{ik} = 0, D_{il} = 0, D_{jk} = 0, D_{jl} = 1|\mathbf{X}^n = x, \Omega(ijlk)^c\right] \\ &\leq \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 0, D_{jl} = 1|\mathbf{X}^n = x, \Omega(ijlk)^c\right] \\ &\leq p_2(\tilde{\beta}, x, \mu)\end{aligned}$$

Therefore:

$$p_1(\tilde{\beta}, x, \mu) \leq \quad p_2(\tilde{\beta}, x, \mu)$$
$$p_3(\tilde{\beta}, x, \mu) \leq \quad p_1(\tilde{\beta}, x, \mu) + p_2(\tilde{\beta}, x, \mu)$$

Which implies, that

$$p_1(\tilde{\beta}, x, \mu) \leq 0.5$$
$$p_1(\tilde{\beta}, x, \mu) + p_2(\tilde{\beta}, x, \mu) \geq 0.5$$

Therefore, $\text{Med}\left[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c\right] = 0$.

Case 2: Consider $\tilde{\beta} \neq \beta_0$ and

$$\begin{aligned}
\mu_i + \mu_k =& \inf X_{ik}^{(1)} \Rightarrow D_{ik} = 0,\\
\mu_j + \mu_k =& \inf X_{jl}'\theta, \text{ for any } \ \theta \in R^K,\\
\mu_i + \mu_l =& \sup X_{il}'\theta, \text{ for any } \ \theta \in R^K,\\
\mu_j + \mu_l =& \sup X_{jl}^{(1)} \Rightarrow D_{jl} = 1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
p_1(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 1, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right],\\
p_2(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 0, D_{il} = 0, D_{jk} = 1, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right]\\
&+ \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right],\\
p_3(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 0, D_{il} = 0, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right].
\end{aligned}$$

Note,

$$\begin{aligned}
p_1(\tilde{\beta}, x, \mu) &= \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 1, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right]\\
&\leq \quad \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right]\\
&\leq \quad p_2(\tilde{\beta}, x, \mu)
\end{aligned}$$

and

$$\begin{aligned}
p_3(\tilde{\beta}, x, \mu) &= \mathbb{P}\left[D_{ik} = 0, D_{il} = 0, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&\leq \quad \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&\leq \quad p_2(\tilde{\beta}, x, \mu)
\end{aligned}$$

Therefore:

$$\begin{aligned}
p_1(\tilde{\beta}, x, \mu) \leq & \quad p_2(\tilde{\beta}, x, \mu) \\
p_3(\tilde{\beta}, x, \mu) \leq & \quad p_1(\tilde{\beta}, x, \mu) + p_2(\tilde{\beta}, x, \mu)
\end{aligned}$$

and $\text{Med}\left[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c\right] = 0$.

Case 3: Consider $\tilde{\beta} \neq \beta_0$ and

$$\begin{aligned}
\mu_i + \mu_k =& \sup X_{ik}' \theta, \text{ for any } \theta \in R^K, \\
\mu_j + \mu_k =& \inf X_{jk}^{(1)} \Rightarrow D_{jk} = 0, \\
\mu_i + \mu_l =& \sup X_{il}^{(1)} \Rightarrow D_{il} = 1, \\
\mu_j + \mu_l =& \inf X_{jl}' \theta, \text{ for any } \theta \in R^K.
\end{aligned}$$

Therefore,

$$\begin{aligned}
p_1(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right], \\
p_2(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
& + \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right], \\
p_3(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right].
\end{aligned}$$

Note,

$$\begin{aligned}
p_1(\tilde{\beta}, x, \mu) &= \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&\leq \quad \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&\leq \quad p_2(\tilde{\beta}, x, \mu)
\end{aligned}$$

and

$$\begin{aligned}
p_3(\tilde{\beta}, x, \mu) &= \mathbb{P}\left[D_{ik}=1, D_{il}=1, D_{jk}=0, D_{jl}=1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&\leq \mathbb{P}\left[D_{ik}=1, D_{il}=1, D_{jk}=0, D_{jl}=0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&\leq p_2(\tilde{\beta}, x, \mu)
\end{aligned}$$

Therefore:

$$\begin{aligned}
p_1(\tilde{\beta}, x, \mu) &\leq p_2(\tilde{\beta}, x, \mu) \\
p_3(\tilde{\beta}, x, \mu) &\leq p_1(\tilde{\beta}, x, \mu) + p_2(\tilde{\beta}, x, \mu)
\end{aligned}$$

Which implies, that

$$\begin{aligned}
p_1(\tilde{\beta}, x, \mu) &\leq 0.5 \\
p_1(\tilde{\beta}, x, \mu) + p_2(\tilde{\beta}, x, \mu) &\geq 0.5
\end{aligned}$$

Therefore, $\text{Med}\left[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c\right] = 0$.

Case 4: Consider $\tilde{\beta} \neq \beta_0$ and

$$\begin{aligned}
\mu_i + \mu_k &= \sup X_{ik}'\theta, \text{ for any } \theta \in R^K, \\
\mu_j + \mu_k &= \sup X_{jk}'\theta, \text{ for any } \theta \in R^K, \\
\mu_i + \mu_l &= \sup X_{il}^{(1)} \Rightarrow D_{il} = 1, \\
\mu_j + \mu_l &= \sup X_{jl}^{(1)} \Rightarrow D_{jl} = 1.
\end{aligned}$$

Therefore,

$$\begin{aligned}
p_1(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik}=0, D_{il}=1, D_{jk}=1, D_{jl}=1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right], \\
p_2(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik}=0, D_{il}=1, D_{jk}=0, D_{jl}=1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
&+ \mathbb{P}\left[D_{ik}=1, D_{il}=1, D_{jk}=1, D_{jl}=1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right], \\
p_3(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik}=1, D_{il}=1, D_{jk}=0, D_{jl}=1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right].
\end{aligned}$$

Note,

$$
\begin{aligned}
p_1(\tilde{\beta}, x, \mu) = \mathbb{P}\left[D_{ik} = 0, D_{il} = 1, D_{jk} = 1, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
\leq \quad & \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 1, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
\leq \quad & p_2(\tilde{\beta}, x, \mu)
\end{aligned}
$$

and

$$
\begin{aligned}
p_3(\tilde{\beta}, x, \mu) = \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
\leq \quad & \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 1, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
\leq \quad & p_2(\tilde{\beta}, x, \mu)
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
p_1(\tilde{\beta}, x, \mu) \leq \quad & p_2(\tilde{\beta}, x, \mu) \\
p_3(\tilde{\beta}, x, \mu) \leq \quad & p_1(\tilde{\beta}, x, \mu) + p_2(\tilde{\beta}, x, \mu)
\end{aligned}
$$

Which implies, that

$$
\begin{aligned}
p_1(\tilde{\beta}, x, \mu) &\leq 0.5 \\
p_1(\tilde{\beta}, x, \mu) + p_2(\tilde{\beta}, x, \mu) &\geq 0.5
\end{aligned}
$$

Therefore, $\text{Med}\left[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c\right] = 0$.

Case 5: Consider $\tilde{\beta} \neq \beta_0$ and

$$
\begin{aligned}
\mu_i + \mu_k &= \sup X_{ik}^{(1)} \Rightarrow D_{ik} = 1, \\
\mu_i + \mu_l &= \sup X_{il}^{(1)} \Rightarrow D_{il} = 1.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
p_1(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 1, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right], \\
p_2(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\
& + \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 1, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right], \\
p_3(\tilde{\beta}, x, \mu) =& \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right].
\end{aligned}
$$

$X_{js}^{(1)}$ has large support for either $s = k, l$. For both cases, set:

$$\mu_j + \mu_k = \inf X_{jk}^{(1)} \Rightarrow D_{jk} = 0,$$
$$\mu_j + \mu_l = \inf X_{jl}^{(1)} \Rightarrow D_{jk} = 1.$$

Note,

$$\begin{aligned} p_1(\tilde{\beta}, x, \mu) &= \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 1, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\ &\leq \quad \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\ &\leq \quad p_2(\tilde{\beta}, x, \mu) \end{aligned}$$

and

$$\begin{aligned} p_3(\tilde{\beta}, x, \mu) &= \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 1 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\ &\leq \quad \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right] \\ &\leq \quad p_2(\tilde{\beta}, x, \mu) \end{aligned}$$

Therefore:

$$p_1(\tilde{\beta}, x, \mu) \leq p_2(\tilde{\beta}, x, \mu)$$
$$p_3(\tilde{\beta}, x, \mu) \leq p_1(\tilde{\beta}, x, \mu) + p_2(\tilde{\beta}, x, \mu)$$

Which implies, that

$$p_1(\tilde{\beta}, x, \mu) \leq 0.5$$
$$p_1(\tilde{\beta}, x, \mu) + p_2(\tilde{\beta}, x, \mu) \geq 0.5$$

Therefore, $\text{Med}\left[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c\right] = 0$.

Case 6: This case is the opposite to case 5.

Case 7: This is the easiest case, one example is:

$$\begin{aligned}
\mu_i + \mu_k &= \sup X_{ik}^{(1)} \Rightarrow D_{ik} = 1, \\
\mu_j + \mu_k &= \sup X_{jk}^{(1)} \Rightarrow D_{jk} = 1, \\
\mu_i + \mu_l &= \sup X_{il}^{(1)} \Rightarrow D_{il} = 1, \\
\mu_j + \mu_l &= \sup X_{jl}^{(1)} \Rightarrow D_{jl} = 1.
\end{aligned}$$

III. Suppose under the true model, characterized by β_0

$$\begin{aligned}
\text{Med}\left[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c\right] = -1 \Leftrightarrow &\{0.5 - (p_1(\beta_0, x, \mu) + p_2(\beta_0, x, \mu)) \leq 0\} \\
&\cap \{p_1(\beta_0, x, \mu) - 0.5 \geq 0\}.
\end{aligned}$$

This part is analogous to case I.

Given any tetrad (i, j, k, l), we have shown that any $\tilde{\beta} \in \mathbb{R}^K$, with $\tilde{\beta}_1 = 1$ and $\tilde{\beta} \neq \beta_0$, is observationally equivalent to β_0 in terms of Med $[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c]$ if the event $\Omega(ijlk)$ has measure zero. Therefore, the median of the random variable $\{\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c\}$ does not have identification power. ■

Proof of Proposition 3.1

Proof. Note that,

$$0 = \mathbb{P}\left[\Omega_n\right] \Leftrightarrow \mathbb{P}\left[\{D_{ik} \neq D_{il}, D_{jl} \neq D_{jk}, D_{ik} \neq D_{jk}\} : i, j, k, l \in \mathcal{N}_n\right] = 0. \tag{33}$$

Hence, it is sufficient to show that one of conditions in (33) fails almost everywhere. Specifically, in the proofs of Parts 1 and 2, I show that if the across-individuals variation does not hold, then the class Ω_n has probability zero. That is,

$$\begin{aligned}
\mathbb{P}\left[\{D_{ik} = D_{jk}\} : i, j, k \in \mathcal{N}_n\right] = 1 \\
\Rightarrow \\
\mathbb{P}\left[\{D_{ik} \neq D_{il}, D_{jl} \neq D_{jk}, D_{ik} \neq D_{jk}\} : i, j, k, l \in \mathcal{N}_n\right] = 0,
\end{aligned}$$

where

$$\begin{aligned}&\mathbb{P}\left[\{D_{ik}=D_{jk}\}:i,j,k\in\mathcal{N}_n\right]\\ =&\mathbb{P}\left[D_{ik}=D_{jk}=1:i,j,k\in\mathcal{N}_n\right]+\mathbb{P}\left[D_{ik}=D_{jk}=0:i,j,k\in\mathcal{N}_n\right].\end{aligned}$$

The proof of Part 1 shows that

$$\begin{aligned}\mathbb{P}\left[D_{ik}=D_{jk}=1:i,j,k\in\mathcal{N}_n\right]=1,\\ \mathbb{P}\left[D_{ik}=D_{jk}=0:i,j,k\in\mathcal{N}_n\right]=0.\end{aligned}$$

Alternatively, the Part 2 shows that

$$\begin{aligned}\mathbb{P}\left[D_{ik}=D_{jk}=1:i,j,k\in\mathcal{N}_n\right]=0,\\ \mathbb{P}\left[D_{ik}=D_{jk}=0:i,j,k\in\mathcal{N}_n\right]=1.\end{aligned}$$

Consider the following convolution argument; fix $i,l\in\mathcal{N}_n$ and $\tilde{\mu}=\mu$, then

$$\mathbb{P}\left[D_{ij}=1\right]=\int\mathbb{P}\left[D_{ij}=1\mid\tilde{\mu}=\mu\right]dF_{\tilde{\mu}}(\mu).$$

Let

$$\int\mathbf{1}\left\{x'\beta_0+\mu_i+\mu_j-e\geq 0\right\}\lambda(dx)=1-F_{X'_{ij}\beta_0|\mu,e}(e-\mu_i-\mu_j),$$

where $F_{X'_{ij}\beta_0|\mu,e}(w)$ is the conditional density of $X'_{ij}\beta_0$ given $\tilde{\mu}=\mu$ and $\varepsilon_{ij}=e$.

Then,

$$\begin{aligned}\mathbb{P}\left[D_{ij}=1\mid\tilde{\mu}=\mu\right]=&\int\int\mathbf{1}\left\{x'\beta_0+\mu_i+\mu_j-e\geq 0\right\}\lambda(dx)dG_{\varepsilon|\mu}(e)\\ =&1-\int F_{X'_{ij}\beta_0|\mu,e}(e-\mu_i-\mu_j)dG_{\varepsilon|\mu}(e).\end{aligned}\tag{34}$$

Part 1.

Rewrite equation (34) as

$$\mathbb{P}\left[D_{ij}=1 \mid \tilde{\mu}=\mu\right]=1-\int\int_{-\infty}^{e-\mu_i-\mu_j} f_{X_{ij}'\beta_0|\mu,e}(x)dx\; dG_{\varepsilon|\mu}(e).$$

Under the support condition in 3.1.1,

$$\int_{-\infty}^{e-\mu_i-\mu_j} f_{X_{ij}'\beta_0|\mu,e}(x)dx=0.$$

Hence,

$$\mathbb{P}\left[D_{ij}=1 \mid \tilde{\mu}=\mu\right]=1 \Rightarrow \mathbb{P}\left[D_{ij}=1\right]=1$$

for any $(i,j)\in\mathcal{N}_n^{(2)}$. Therefore,

$$1=\mathbb{P}\left[D_{ik}=D_{jk}=1: i,j,k\in\mathcal{N}_n\right]=\mathbb{P}\left[\{D_{ik}=D_{jk}\}: i,j,k\in\mathcal{N}_n\right]\Rightarrow\mathbb{P}\left[\Omega_n\right]=0.$$

Part 2.

Under the support condition in 3.1.2,

$$\int_{-\infty}^{e-\mu_i-\mu_j} f_{X_{ij}'\beta_0|\mu,e}(x)dx=1.$$

Hence,

$$\mathbb{P}\left[D_{ij}=1 \mid \tilde{\mu}=\mu\right]=0 \Rightarrow \mathbb{P}\left[D_{ij}=1\right]=0$$

for any $(i,j)\in\mathcal{N}_n^{(2)}$. Therefore,

$$1=\mathbb{P}\left[D_{ik}=D_{jk}=0: i,j,k\in\mathcal{N}_n\right]=\mathbb{P}\left[\{D_{ik}=D_{jk}\}: i,j,k\in\mathcal{N}_n\right]\Rightarrow\mathbb{P}\left[\Omega_n\right]=0.$$

Part 3.

Consider the following convolution argument; fix $i,l\in\mathcal{N}_n$ and $\tilde{\mu}=\mu$, then

$$\mathbb{P}\left[D_{ij}=1\right]=\int\mathbb{P}\left[D_{ij}=1 \mid X_{ij}'=x\right]dF_{X_{ij}'}(x).$$

Let

$$\int \mathbf{1}\left\{x'\beta_0 + \mu_i + \mu_j - e \geq 0\right\} \nu(d\mu) = 1 - F_{\tilde{\mu}_i+\tilde{\mu}_j|x,e}(e - x'\beta_0),$$

where $F_{\tilde{\mu}_i+\tilde{\mu}_j|x,e}(w)$ is the conditional density of $\tilde{\mu}_i + \tilde{\mu}_j$ given $X_{ij}^{'} = x$ and $\varepsilon_{ij} = e$.

Then,

$$\begin{aligned}\mathbb{P}\left[D_{ij} = 1 \mid X_{ij}^{'} = x\right] =& \int\int \mathbf{1}\left\{x'\beta_0 + \mu_i + \mu_j - e \geq 0\right\} \nu(d\mu) dG_{\varepsilon|x}(e) \\ =& 1 - \int F_{\tilde{\mu}_i+\tilde{\mu}_j|x,e}(e - x'\beta_0)\, dG_{\varepsilon|x}(e). \end{aligned} \tag{35}$$

Rewrite equation (35) as

$$\mathbb{P}\left[D_{ij} = 1 \mid X_{ij}^{'} = x\right] = 1 - \int\int_{-\infty}^{e-x'\beta_0} f_{\tilde{\mu}_i+\tilde{\mu}_j|x,e}(\mu) d\mu\, dG_{\varepsilon|\mu}(e).$$

Under the support condition in 3.1.1,

$$\int_{-\infty}^{e-x'\beta_0} f_{\tilde{\mu}_i+\tilde{\mu}_j|x,e}(\mu) d\mu = 0.$$

Hence,

$$\mathbb{P}\left[D_{ij} = 1 \mid X_{ij}^{'} = x\right] = 1 \Rightarrow \mathbb{P}\left[D_{ij} = 1\right] = 1$$

for any $(i, j) \in \mathcal{N}_n^{(2)}$. Therefore,

$$1 = \mathbb{P}\left[D_{ik} = D_{jk} = 1 : i, j, k \in \mathcal{N}_n\right] = \mathbb{P}\left[\{D_{ik} = D_{jk}\} : i, j, k \in \mathcal{N}_n\right] \Rightarrow \mathbb{P}\left[\Omega_n\right] = 0.$$

■

B.3 Identification Failure: Maximum Score

The following lemma is used to prove proposition 3.2. The next lemma adapts Lemma 1 in Manski (1987) to the multiple fixed effects case.

Lemma 3

Let assumption A1 hold. For any n, and any $i, l, k \in \mathcal{N}_n$.

$$\text{sign}\left[(x_{ik} - x_{il})'\beta_0 + (\mu_k - \mu_l)\right] = \text{sign}\left(\mathbb{E}\left[D_{ik} - D_{il} \mid \mathbf{X}^n = \mathbf{x}\right]\right) \tag{36}$$

Proof. Fix $i, k, l \in \mathcal{N}_n$ and $\mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu$,

$$\begin{aligned}
\mathbb{P}\left[D_{ik} = 1 \mid \mathbf{x}, \mu\right] =& F_{\varepsilon_i|\mathbf{x},\mu}\left[x'_{ik}\beta_0 + \mu_i + \mu_k\right] \\
\mathbb{P}\left[D_{il} = 1 \mid \mathbf{x}, \mu\right] =& F_{\varepsilon_i|\mathbf{x},\mu}\left[x'_{il}\beta_0 + \mu_i + \mu_l\right]
\end{aligned}$$

Note,

$$x'_{il}\beta_0 + \mu_i + \mu_l \ \leq \ x'_{ik}\beta_0 + \mu_i + \mu_k \qquad \Leftrightarrow \qquad x'_{il}\beta_0 + \mu_l \ \leq \ x'_{ik}\beta_0 + \mu_k$$

then it follows that for all $\mathbf{x}, \mu$,

$$\begin{aligned}
x'_{il}\beta_0 + \mu_l \ \leq \ x'_{ik}\beta_0 + \mu_k \ &\Leftrightarrow \ \mathbb{P}\left[D_{il} = 1 \mid \mathbf{x}, \mu\right] \ \leq \ \mathbb{P}\left[D_{ik} = 1 \mid \mathbf{x}, \mu\right] \\
&\Leftrightarrow \ \mathbb{E}\left[D_{il} \mid \mathbf{x}, \mu\right] \ \leq \ \mathbb{E}\left[D_{ik} \mid \mathbf{x}, \mu\right] \\
&\Leftrightarrow \ 0 \ \leq \ \mathbb{E}\left[D_{ik} - D_{il} \mid \mathbf{x}, \mu\right]
\end{aligned}$$

Equivalently for

$$x'_{ik}\beta_0 + \mu_k \ < \ x'_{il}\beta_0 + \mu_l \ \ \Leftrightarrow \ \ \mathbb{E}\left[D_{ik} - D_{il} \mid \mathbf{x}, \mu\right] < 0$$

In summary,

$$\begin{aligned}
(x_{ik} - x_{il})'\beta + [\mu_k - \mu_l] \ &\geq \ 0 \ \ \Leftrightarrow \ \ \mathbb{E}\left[D_{ik} - D_{il} \mid \mathbf{x}, \mu\right] \ \geq \ 0 \\
(x_{ik} - x_{il})'\beta + [\mu_k - \mu_l] \ &< \ 0 \ \ \Leftrightarrow \ \ \mathbb{E}\left[D_{ik} - D_{il} \mid \mathbf{x}, \mu\right] \ < \ 0
\end{aligned}$$

thus,

$$\begin{aligned}
\text{sign}\left\{(x_{ik} - x_{il})'\beta + [\mu_k - \mu_l]\right\} =& \,\text{sign}\left\{\mathbb{E}\left[D_{ik} - D_{il} \mid \mathbf{x}, \mu\right]\right\} \\
=& \,\text{sign}\left\{\mathbb{E}\left[D_{ik} - D_{il} \mid \mathbf{x}\right]\right\}
\end{aligned}$$

■

Proof of Proposition 3.2

Proof. **Part 1.**

Fix $i, k, l \in \mathcal{N}_n$, $\mathbf{X}^n = x$ and $\tilde{\mu} = \mu$. Note that, $D_{ik} - D_{il}|x, D_{il} + D_{ik} = 1$ is a Bernoulli random variable with support given by $\{-1, 1\}$ and probability distribution.

$$\mathbb{P}\left[D_{ik} - D_{il} = 1|x, D_{il} + D_{ik} = 1\right] = \frac{\mathbb{P}\left[D_{ik} = 1, D_{il} = 0|x\right]}{\mathbb{P}\left[D_{il} \neq D_{ik}|x\right]}$$
$$\mathbb{P}\left[D_{ik} - D_{il} = -1|x, D_{il} + D_{ik} = 1\right] = \frac{\mathbb{P}\left[D_{ik} = 0, D_{il} = 1|x\right]}{\mathbb{P}\left[D_{il} \neq D_{ik}|x\right]}$$

Then,

$$\begin{aligned}\text{Med}\left(D_{ik} - D_{il}|x, D_{il} + D_{ik} = 1\right) = &\,\text{sign}\left\{\mathbb{P}\left[D_{ik} = 1, D_{il} = 0|x\right]\right.\\ &\left.- \mathbb{P}\left[D_{ik} = 0, D_{il} = 1|x\right]\right\}\\ = &\,\text{sign}\left\{\mathbb{E}\left[D_{ik} - D_{il}|\mathbf{X}^n = x\right]\right\}\end{aligned}$$

By Lemma 3, the result follows.

Part 2.

Denote by P_{β_0} the distribution of the observables $Z = (D, X)$ under the true $\beta_0 \in \mathcal{B}$.

Denote by $G(P_{\beta_0})$ a (known) specific feature of the distribution of the observables given the true model. In particular, let $G(P_{\beta_0}) = \text{Med}(D_{ik} - D_{il}|\mathbf{X}^n = x, D_{il} + D_{ik} = 1)$. Then, equation (6) states that

$$G(P_{\beta_0}) = \text{sign}\left[(x_{ik} - x_{il})'\beta_0 + (\mu_k - \mu_l)\right]$$

Assumption 1 allows for a flexible representation of the conditional distribution of $\tilde{\mu}$ given $\mathbf{X}^n = \mathbf{x}$. That is the conditional distribution $F_{\tilde{\mu}|\mathbf{X^n=x}}$ is unrestricted. In order to prove proposition 3.2, we will assume that $\tilde{\mu}$ is a known function of the exogenous attributed. Note that if β_0 cannot be identified under this restriction of $\tilde{\mu}$, it will also fail to be identified either under a more flexible representation of $\tilde{\mu}$.

Consider $K \in \mathbb{R}^K$ and $c \in \mathbb{R}_+$, with $K \neq \mathbf{0}$ and $c \neq 0$. For any $i, k, l \in \mathcal{N}_n$, define:

$$\begin{aligned} \mu_i - \mu_k =& X'_{ik} (\beta_0 - K) \\ \tilde{\beta} =& \beta_0 + cK \end{aligned}$$

with $\tilde{\beta} \in \mathcal{B}$. Hence,

$$\begin{aligned} G(P_{\beta_0}) =& \operatorname{sign}\left[(x_{ik} - x_{il})'\beta_0 + (\mu_k - \mu_l)\right] \\ =& \operatorname{sign}\left[(x_{ik} - x_{il})'\beta_0 - [(\mu_i - \mu_k) - (\mu_i - \mu_l)]\right] \\ =& \operatorname{sign}\left[(x_{ik} - x_{il})'\beta_0 - (x_{ik} - x_{il})'\beta_0 + (x_{ik} - x_{il})' K\right] \\ =& \operatorname{sign}\left[(x_{ik} - x_{il})' K\right] \end{aligned}$$

and

$$\begin{aligned} G(P_{\tilde{\beta}}) =& \operatorname{sign}\left[(x_{ik} - x_{il})'\tilde{\beta} + (\mu_k - \mu_l)\right] \\ =& \operatorname{sign}\left[(x_{ik} - x_{il})' \{\beta_0 + cK\} - [(\mu_i - \mu_k) - (\mu_i - \mu_l)]\right] \\ =& \operatorname{sign}\left[(x_{ik} - x_{il})'\beta_0 + (x_{ik} - x_{il})'cK - (x_{ik} - x_{il})'\beta_0 + (x_{ik} - x_{il})' K\right] \\ =& \operatorname{sign}\left[(c+1)(x_{ik} - x_{il})' K\right] \end{aligned}$$

Given that $c > 0$, then $c + 1 > 0$. This implies that $G(P_{\beta_0}) = G(P_{\tilde{\beta}})$. In other words, we have shown that given β_0 we can find a $\tilde{\beta} \in \mathcal{B}$ with $\tilde{\beta} \neq \beta_0$ such that they are observationally equivalent. ■

Proof of Proposition 3.3

The proof of this proposition is the main text.

B.4 Inference

Proof of Proposition 4.1

Proof. Fix any $b \in \tilde{\mathcal{B}}$ such that $b \neq \beta$. For any, $\mathbf{x} \in \mathcal{X}_B$, Assumption A3 guarantees the existence of a set of values of $(x_{ik}, x_{il}, x_{jk}, x_{jl})$ with positive probability for which

$$h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0) \neq h(x_{ik}, x_{il}, x_{jk}, x_{jl}, b).$$

Denote this set by

$$\mathcal{H}_{(b)} \equiv \left[(x_{13}, x_{14}, x_{23}, x_{24}) \in \mathbb{R}^K \times \cdots \times \mathbb{R}^K : h(x_{ik}, x_{il}, x_{jk}, x_{jl}, b) \neq h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0)\right],$$

for any $b \in \tilde{\mathcal{B}}$ with $b \neq \beta_0$. Then,

$$\begin{aligned}
&Q(\beta_0) - Q(b) \\
&= \mathbb{E}\left[S(\mathcal{X}_B)\left\{\Delta_{kl}D_i - \Delta_{kl}D_j\right\}\left\{h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0) - h(x_{ik}, x_{il}, x_{jk}, x_{jl}, b)\right\} \mid \Omega\right] \\
&= \mathbb{E}_{\mathcal{X}_B}\left[\left\{\Delta_{kl}D_i - \Delta_{kl}D_j\right\} \times \left\{h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0) - h(x_{ik}, x_{il}, x_{jk}, x_{jl}, b)\right\} \mid \Omega\right] \\
&= \mathbb{E}_{\mathcal{X}_B \cap \mathcal{H}_{(b)}}\left[\left\{h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0) - h(x_{ik}, x_{il}, x_{jk}, x_{jl}, b)\right\}\mathbb{E}\left[\Delta_{kl}D_i - \Delta_{kl}D_j \mid x, \Omega\right]\right] \\
&+ \mathbb{E}_{\mathcal{X}_B \cap \mathcal{H}^c_{(b)}}\left[\left\{h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0) - h(x_{ik}, x_{il}, x_{jk}, x_{jl}, b)\right\}\mathbb{E}\left[\Delta_{kl}D_i - \Delta_{kl}D_j \mid x, \Omega\right]\right].
\end{aligned}$$

Then,

$$\begin{aligned}
&Q(\beta_0) - Q(b) \\
&= \mathbb{E}_{\mathcal{X}_B \cap \mathcal{H}_{(b)}}\left[\left\{h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0) - h(x_{ik}, x_{il}, x_{jk}, x_{jl}, b)\right\}\mathbb{E}\left[\left\{\Delta_{kl}D_i - \Delta_{kl}D_j\right\} \mid x, \Omega\right]\right] \\
&= 2\mathbb{E}_{\mathcal{X}_B \cap \mathcal{H}_{(b)}}\left[h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0)\mathbb{E}\left[\left\{\Delta_{kl}D_i - \Delta_{kl}D_j\right\} \mid x, \Omega\right]\right],
\end{aligned}$$

the last equality follows from the fact that in the set $\mathcal{H}_(b)$

$$h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0) - h(x_{ik}, x_{il}, x_{jk}, x_{jl}, b) = 2\left[h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0)\right],$$

because of the relationship

$$h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0) = 1 \Leftrightarrow h(x_{ik}, x_{il}, x_{jk}, x_{jl}, b) = -1$$
$$h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0) = -1 \Leftrightarrow h(x_{ik}, x_{il}, x_{jk}, x_{jl}, b) = 1.$$

We have shown that:

$$h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0) = \text{sign}\left\{\mathbb{E}\left[(D_{ik} - D_{il}) - (D_{jk} - D_{jl}) \mid x, \Omega\right]\right\}$$

in $\mathcal{X}_B$. Hence,

$$\begin{aligned} h(x_{ik}, x_{il}, x_{jk}, x_{jl}, \beta_0)\mathbb{E}\left[(D_{ik} - D_{il}) - (D_{jk} - D_{jl}) \mid x, \Omega\right] \\ = \mid \mathbb{E}\left[(D_{ik} - D_{il}) - (D_{jk} - D_{jl}) \mid x, \Omega\right] \mid > 0. \end{aligned}$$

for any $\mathbf{x} \in \left\{\mathcal{X}_B \cap \mathcal{H}_{(b)}\right\}$. Ass 3 guarantees that $\left\{\mathcal{X}_B \cap \mathcal{H}_{(b)}\right\}$ has positive measure. Therefore,

$$Q(\beta_0) - Q(b) = 2\mathbb{E}_{\mathcal{X}_B \cap \mathcal{H}_{(b)}}\left[\mid \mathbb{E}\left[(D_{ik} - D_{il}) - (D_{jk} - D_{jl}) \mid x, \Omega\right] \mid\right] > 0$$

Since $b \in \tilde{\mathcal{B}}$ was chosen arbitrarily, it follows that β_0 uniquely maximizes $Q(b)$.

■

Proof of Theorem 4.1

Proof. Theorem 2.1. in Newey and McFadden (1994) implies that the following conditions are sufficient to prove strong consistency.

1. $\tilde{\mathcal{B}}$ is compact.

2. $Q(b)$ is continuous.

3. $Q(b)$ is uniquely maximized at β_0.

4. $p_n^{-1}Q_n(b)$ converges almost surely to $Q(b)$, i.e. $\sup_{b\in\tilde{\mathcal{B}}}\left|\left|p_n^{-1}Q_n(b) - Q(b)\right|\right| \overset{a.s.}{\to} 0$.

Consider the scaled sample analog of the objective function $p_n^{-1}Q(b)$:

$$p_n^{-1}Q_n(b) \equiv p_n^{-1}\binom{n}{4}^{-1}\sum_{C_{n,4}} h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b),$$

For each $(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}) \in S^4 = S \times S \times S \times S$, the kernel function of the U-statistic is given by

$$\begin{aligned} &h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) \equiv \\ \frac{1}{4!}\sum_{P_4}\Big\{2 \times \text{sign}\Big\{\big[\big(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}}\big) - \big(x_{\mathbf{i}_{2,3}} - x_{\mathbf{i}_{2,4}}\big)\big]' b\Big\} &\times \Big(y^{(i_1)}_{\mathbf{i}_{3,4}} - y^{(i_2)}_{\mathbf{i}_{3,4}}\Big) \\ &\times \mathbf{1}\Big\{\Big|\Big(y^{(i_1)}_{\mathbf{i}_{3,4}} - y^{(i_2)}_{\mathbf{i}_{3,4}}\Big)\Big| = 2\Big\}\Big\}, \end{aligned}$$

Let

$$\begin{aligned} \tilde{Q}_n(b) \equiv&\ p_n^{-1}Q_n(b), \\ \tilde{h}(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) \equiv&\ p_n^{-1}h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b), \end{aligned}$$

for any $(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}) \in S^4$.

Let

$$\tilde{g}(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) \equiv\ \tilde{h}(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) - \mathbb{E}\Big[\tilde{h}(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b)\Big].$$

Then

$$\begin{aligned} \tilde{Q}_n(b) - Q(b) =&\binom{n}{4}^{-1}\sum_{C_{n,4}}\Big\{\tilde{h}(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) - \mathbb{E}\Big[\tilde{h}(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b)\Big]\Big\} \\ =&\binom{n}{4}^{-1}\sum_{C_{n,4}}\tilde{g}(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) \\ =&U_n^4\tilde{g}(\cdot, b), \end{aligned}$$

where $\Big\{U_n^4 g(\cdot, b); b \in \tilde{\mathcal{B}}\Big\}$ is a zero-mean U-process of order 4.

- *Condition 1:* Follows from assumption B2.
- *Condition 2:* Assumption A2 implies that $\big(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}}\big)' b$ is continuously distributed for

each $b \in \tilde{\mathcal{B}}$, for any (i_1, i_3, i_4) from $\{1, 2, \cdots, n\}$. Therefore,

$$\text{sign}\left\{\left[(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}}) - (x_{\mathbf{i}_{2,3}} - x_{\mathbf{i}_{2,4}})\right]' b\right\}$$

and $\tilde{h}(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b)$ are continuous $b \in \tilde{\mathcal{B}}$ for almost all

$$(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}) \in S^4$$

.

Since, $\tilde{h}(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, \cdot)$ is uniformly bounded in all of its arguments, then by the dominated convergence theorem $\mathbb{E}\left[h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b)\right]$ is continuous. Hence, $Q(b)$ is continuous.

- *Condition 3:* This condition follows from Proposition 4.1.
- *Condition 4:* The class of functions $\left\{\tilde{g}(\cdot, b) : b \in \tilde{\mathcal{B}}\right\}$ is Euclidean for the constant envelope 1. The process $\left\{U_n^4 \tilde{g}(\cdot, b) : b \in \tilde{\mathcal{B}}\right\}$ is a zero-mean U-process of order 4, then by Corollary 4 in Sherman (1994):

$$\begin{aligned}\sup_{\tilde{\mathcal{B}}} \left|\left|U_n^4 \tilde{g}(\cdot, b)\right|\right| =& O_p(1/\sqrt{N}) \\ \Rightarrow \sup_{\tilde{\mathcal{B}}} \left|\left|U_n^4 g(\cdot, b)\right|\right| =& O_p(1/p_n\sqrt{N})\end{aligned}$$

$$\Leftrightarrow \sup_{\tilde{\mathcal{B}}} \left|\left|\tilde{Q}_n(b) - Q(b)\right|\right| = O_p\left(\frac{1}{p_n\sqrt{N}}\right).$$

Assumption B3 guarantees that $p_n\sqrt{N} \to 0$. Therefore,

$$\sup_{\tilde{\mathcal{B}}} \left|\left|\tilde{Q}_n(b) - Q(b)\right|\right| = o_p(1).$$

Conditions 1-4 imply $\hat{\beta}_n \overset{a.s.}{\to} \beta_0$ as $n \to \infty$. ■

Proof of Theorem 4.2

Proof. This proof is divided in two parts.

Part 1: This part shows that the estimator $\hat{\beta}$ is $p_n\sqrt{N}$-consistent for β_0. That is

$$\left|\left|\hat{\beta} - \beta_0\right|\right| = O_p(1/p_n\sqrt{N}). \tag{37}$$

This result follows from Theorem 1 in Sherman (1993), and the following quadratic approximation to $\tilde{Q}(b)$:

$$\begin{aligned} \tilde{Q}_n(b) - \tilde{Q}_n(\beta_0) = \\ \frac{1}{2}(b-\beta_0)'V(b-\beta_0) + \frac{1}{p_n\sqrt{N}}(b-\beta_0)'\mathbf{W}_n \\ + o_p\left(\left|\left|(b-\beta_0)\right|\right|^2/p_n\right) + o_p\left(\frac{1}{p_nN}\right) \end{aligned} \tag{38}$$

uniformly in $o_p(1)$ neighborhoods of β_0, where $\mathbf{W}_n$ converges in distribution to a $\mathcal{N}(\mathbf{0}, \Delta)$ random vector.

Part 2: The asymptotic distribution of $(\hat{\beta} - \beta_0)$ is established from Part 1, equation (38) and Theorem 2 in Sherman (1993).

Proof of Part 1: This proof consists of three steps.

Let

$$f(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) = \tilde{h}(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) - \tilde{h}(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, \beta_0),$$

for each $(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b)$ in S^4 and each $b \in \tilde{\mathcal{B}}$.

Assume $\mathbf{P}^4 < \infty$, where $\mathbf{P}^4$ denotes the product measure $\mathbf{P}\times\mathbf{P}\times\mathbf{P}\times\mathbf{P}$ for the sampling distribution $\mathbf{P}$ on S. Given $Q_n(b)$, the Hoeffding decomposition of U-statistics guarantees that there exist functions $f_1, \cdots, f_4$ such that for each i, f_i is $\mathbf{P}$-degenerate on S^i, $i = 2, 3, 4$ and

$$\tilde{Q}_n(b) - \tilde{Q}_n(\beta_0) = Q(b) - Q(\beta_0) + \mathbf{P}_n f_1(\cdot, b) + \sum_{i=2}^{4} U_n^i f_i(\cdot, b), \tag{39}$$

where $Q(b) = \mathbf{P}^4\tilde{h}(\cdot,\cdot,\cdot,\cdot,b)$. For each $z \in S$, $f_1(\cdot, b)$ is defined as:

$$\begin{aligned} f_1(z,b) =& f(z,\mathbf{P},\mathbf{P},\mathbf{P},b) + f(\mathbf{P},z,\mathbf{P},\mathbf{P},b) + f(\mathbf{P},\mathbf{P},z,\mathbf{P},b) + f(\mathbf{P},\mathbf{P},\mathbf{P},z) \\ & - 4(Q(b) - Q(\beta_0)), \end{aligned}$$

where $f(z,\mathbf{P},\mathbf{P},\mathbf{P},b)$ denotes the conditional expectation of $\tilde{f}(\cdot, b)$ under $\mathbf{P}^4$, given its first argument. The remaining terms have analogous interpretations. The expressions for f_2, f_3, f_4 can be found in Serfling (2009, pp. 177-178).

Recall,

$$\tau(z,b) \equiv h(z,\mathbf{P},\mathbf{P},\mathbf{P},b) + h(\mathbf{P},z,\mathbf{P},\mathbf{P},b) + h(\mathbf{P},\mathbf{P},z,\mathbf{P},b) + h(\mathbf{P},\mathbf{P},\mathbf{P},z),$$

for each $z \in S$, and each $b \in \tilde{\mathcal{B}}$.

Step 1:

Given assumption B4, consider a Taylor expansion of $\tau(\cdot,b)$ around β_0

$$\tau(z,b) = \tau(z,\beta_0) + (b-\beta_0)'\nabla_1\tau(z,\beta_0) + \frac{1}{2}(b-\beta_0)'\nabla_2\tau(z,b^*)(b-\beta_0), \tag{40}$$

for any $z \in S$ and $b \in \mathscr{M}$ and b^* between b and β_0.

Assumption B4.2 implies

$$\left|\left|(b-\beta_0)'\left[\nabla_2\tau(z,b) - \nabla_2\tau(z,\beta_0)\right](b-\beta_0)\right|\right| \leq M(z)\,||(b-\beta_0)||^3 \quad \text{as} \quad b \to \beta_0 \tag{41}$$

From inequality (41) and the integrability of M, the expected value of equation (40) implies:

$$\begin{aligned}\mathbb{E}\left[(\tau(\cdot,b) - \tau(\cdot,\beta_0))\right] &= (b-\beta_0)'\mathbb{E}\nabla_1\tau(\cdot,\beta_0) + \frac{1}{2}(b-\beta_0)'\mathbb{E}\nabla_2\tau(\cdot,b^*)(b-\beta_0)\\ 4p_n(Q(b) - Q(\beta_0)) &= (b-\beta_0)'\mathbb{E}\nabla_1\tau(\cdot,\beta_0) + \frac{1}{2}(b-\beta_0)'\mathbb{E}\nabla_2\tau(\cdot,\beta_0)(b-\beta_0)\\ &\quad + o\left(||(b-\beta_0)||^2\right)\end{aligned}$$

$$\begin{aligned}Q(b) - Q(\beta_0) &= \frac{1}{4}(b-\beta_0)'p_n^{-1}\mathbb{E}\nabla_1\tau(\cdot,\beta_0) + \frac{1}{2}(b-\beta_0)'V(b-\beta_0)\\ &\quad + o\left(||(b-\beta_0)||^2/p_n\right)\end{aligned}$$

where $4V = \mathbb{E}\left[\nabla_2\tau(\cdot,\beta_0) \mid \Omega_n\right]$ is a negative definite matrix.

Since $Q(b) - Q(\beta_0)$ is maximized at β_0, then it necessarily holds $\mathbb{E}\nabla_1\tau(\cdot,\beta_0) = 0$.

$$Q(b) - Q(\beta_0) = \frac{1}{2}(b-\beta_0)'V(b-\beta_0) + o\left(||(b-\beta_0)||^2/p_n\right)$$

Step 2: To show that,

$$\mathbf{P}_n f_1(\cdot, b) = \frac{1}{p_n\sqrt{N}}(b-\beta_0)'\mathbf{W}_n + o_p\left(||(b-\beta_0)||^2/p_n\right) \tag{42}$$

uniformly over $o_p(1)$ neighborhoods of β_0, where $\mathbf{W}_n$ converges in distribution to a $\mathcal{N}(\mathbf{0}, \mathbf{\Delta})$ random vector.

Note

$$\begin{aligned} f_1(z,b) =& p_n^{-1}(\tau(z,b)-\tau(z,\beta_0)) - 4(Q(b)-Q(\beta_0)) \\ f_1(z,b) =& (b-\beta_0)'p_n^{-1}\nabla_1\tau(z,\beta_0) + \frac{1}{2}(b-\beta_0)'p_n^{-1}\nabla_2\tau(z,\beta_0)(b-\beta_0) \\ &+ p_n^{-1}M(Z)\,||(b-\beta_0)||^3 \\ &- 4\left\{\frac{1}{2}(b-\beta_0)'V(b-\beta_0) + o\left(||(b-\beta_0)||^2/p_n\right)\right\} \end{aligned}$$

Hence,

$$\begin{aligned} \mathbf{P}_n f_1(\cdot,b) = (b-\beta_0)'p_n^{-1}\mathbf{P}_n\nabla_1\tau(\cdot,\beta_0) + \frac{1}{2}(b-\beta_0)'\left\{p_n^{-1}\mathbf{P}_n\nabla_2\tau(\cdot,\beta_0) - 4V\right\}(b-\beta_0) \\ + o\left(||(b-\beta_0)||^2/p_n\right) + \mathbf{R}_n(p_n,(b-\beta_0)), \end{aligned}$$

where

$$|\mathbf{R}_n(p_n,(b-\beta_0))| \le ||(b-\beta_0)||^3\, p_n^{-1}\mathbf{P}_n M(\cdot).$$

Let,

$$\begin{aligned} \mathbf{W}_n =& \sqrt{N}\mathbf{P}_n\nabla_1\tau(\cdot,\beta_0), \\ \mathbf{D}_n =& p_n^{-1}\mathbf{P}_n\nabla_2\tau(\cdot,\beta_0) - 4V. \end{aligned}$$

Then:

$$\begin{aligned} \mathbf{P}_n f_1(\cdot,b) = \frac{1}{p_n\sqrt{N}}(b-\beta_0)'\mathbf{W}_n + \frac{1}{2}(b-\beta_0)'\mathbf{D}_n(b-\beta_0) + o\left(||(b-\beta_0)||^2/p_n\right) \\ + \mathbf{R}_n(p_n,(b-\beta_0)) \end{aligned}$$

Assumption B4.3 implies $\mathbf{P}|\nabla_1\tau(\cdot,\beta_0)|^2 < \infty$, given $\mathbf{P}\nabla_1\tau(\cdot,\beta_0) = 0$, then $\mathbf{W}_n$ converges in distribution to a $\mathcal{N}(\mathbf{0},\Delta)$.

Assumption B4.4 and a weak law of large numbers, imply that $\mathbf{D}_n \xrightarrow{p} 0$ as N tends to infinity.

Since the function M is integrable, then:

$$\mathbf{R}_n(p_n, (b - \beta_0)) = o_p\left(||(b - \beta_0)||^2 / p_n\right)$$

Hence, it has been proved that

$$\mathbf{P}_n f_1(\cdot, b) = \frac{1}{p_n\sqrt{N}}(b - \beta_0)'\mathbf{W}_n + o_p\left(||(b - \beta_0)||^2 / p_n\right)$$

as N tends to infinity.

Step 3: In order to prove equation (38) we need to show that

$$U_n^2 f_2(\cdot, b) + U_n^3 f_3(\cdot, b) + U_n^4 f_4(\cdot, b) = o_p(1/p_n N) \tag{43}$$

uniformly over $o_p(1)$ neighborhoods of β_0.

For each $i = 2, 3, 4$, the class $\left\{f_i(\cdot, b) : b \in \tilde{\mathcal{B}}\right\}$ is Euclidean for the constant envelope 1. Equation (43) is proved by using Corollary 8 in Sherman (1994), if the following conditions hold:

$$\mathbf{P}^i \, ||f_i(\cdot, b)|| \to 0 \quad \text{as} \quad b \to \beta_0, \tag{44}$$

for $i = 2, 3, 4$.

I will show the result for $i = 4$. The rest are analogous.

By assumption A2, the distribution of $\left(X_{\mathbf{i}_{1,3}} - X_{\mathbf{i}_{1,4}}\right)' b$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$. Thus,

$$P^4 \left\{\left[\left(X_{\mathbf{i}_{1,3}} - X_{\mathbf{i}_{1,4}}\right)' \beta_0 - \left(X_{\mathbf{i}_{2,3}} - X_{\mathbf{i}_{2,4}}\right)' \beta_0\right]\right\} = 0$$

Henceforth, $f(z_{\mathbf{i}_{1,3}}, \cdots, z_{\mathbf{i}_{2,4}}, \cdot)$ is continuous at β_0 for P^4 almost all $(z_{\mathbf{i}_{1,3}}, \cdots, z_{\mathbf{i}_{2,4}})$. The function f is uniformly bounded in each of its arguments, by the dominated convergence theorem the function $f_4(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, \cdot)$ is continuous at β_0 for P^4 almost all $(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}})$, since f_4 is an additive function on the expected value of f. Finally, the function f_4 is also uniformly bounded in all of its arguments, then by the dominated convergence theorem it holds:

$$\mathbf{P}^4 \, ||f_4(\cdot, b)|| \to 0 \quad \text{as} \quad b \to \beta_0$$

Similar steps can be used to prove (44) for $i = 2, 3$. Then by Corollary 8 in Sherman (1994), follows:

$$U_n^2 f_2(\cdot, b) + U_n^3 f_3(\cdot, b) + U_n^4 f_4(\cdot, b) = o_p(1/p_n N) \tag{45}$$

As in Theorem 1 of Sherman (1993), we can show that

$$\sqrt{p_n}\left|\left|\hat{\beta} - \beta_0\right|\right| \leq O_p\left(\frac{1}{\sqrt{p_n N}}\right)$$

Then, equation (37) has been established.

Proof of Part 2:

This result follows from similar arguments as the proof of Theorem 2 in Sherman (1994). Let

$$\begin{aligned} t_n^* &= \; -V^{-1}\mathbf{W}_n + \beta_0 \\ t_n &= p_n\sqrt{N}(\hat{\beta} - \beta_0) + \beta_0 \end{aligned}$$

Then, by definition of t_n^*

$$\tilde{Q}_n(t_n^*/p_n\sqrt{N}) - \tilde{Q}_n(\beta_0) \leq \tilde{Q}_n(t_n/p_n\sqrt{N}) - \tilde{Q}_n(\beta_0)$$

By applying (38) twice in the last expression, multiplying by $p_n^2 N$, collecting terms, and using that V is negative definite.

$$0 \leq -\frac{1}{2}(t_n - t_n^*)' V (t_n - t_n^*) \leq o_p\left(1\right)$$

Hence, it has been established $t_n = t_n^* + o_p(1)$. Equivalently,

$$p_n\sqrt{N}(\hat{\beta} - \beta_0) = -V^{-1}\mathbf{W}_n + o_p(1)$$

■

Proof of Theorem 4.3

Proof. Theorem 2 in Chamberlain (2010) states that the information bound is zero for any $\beta_0 \in \tilde{\mathcal{B}}$ unless F_ϵ is logistic.

Suppose, F_ϵ is logistic, then the information bounds is:

$$I(\beta_0) = \frac{p_0}{36}\mathbb{E}\left[-\nabla_{bb} \ln f_\epsilon(\beta_0)|\Omega_n\right] \Delta^{-1}\mathbb{E}\left[-\nabla_{bb} \ln f_\epsilon(\beta_0)|\Omega_n\right],$$

where $\Delta = \mathbb{E}\left[\nabla_b \ln f_\epsilon(\beta_0)\nabla_b \ln f_\epsilon(\beta_0)'\right]$. Hence, if $p_n \to p_0 = 0$ the information bound of β_0 is zero. ■

Proof of Theorem 4.5

Proof. This proof is divided in two parts. In part 1, I show

- $\mathbf{P}\left(\hat{\mathcal{B}} \neq \mathcal{B}\right) \to 0$, as $n \to \infty$.
- $\mathbf{P}\left(R(\hat{\mathcal{B}}) \neq R(\mathcal{B})\right) \to 0$, as $n \to \infty$.

I use this results in part 2 to show:

- $s_n H(\hat{\mathcal{B}}, \mathcal{B}) \xrightarrow{p} 0$, for any nonnegative sequence s_n and as $n \to \infty$.
- $s_n H(R(\hat{\mathcal{B}}), R(\mathcal{B})) \xrightarrow{p} 0$, as $n \to \infty$, for any nonnegative sequence s_n and as $n \to \infty$.

To simplify notation, denote by

$$\hat{\mathbb{E}}_n \equiv \hat{\mathbb{E}}_n\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = \mathbf{x}^d, \Omega(ijlk)\right]$$
$$\mathbb{E} \equiv \mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = \mathbf{x}^d, \Omega(ijlk)\right]$$

Part 1: Note,

$$\mathbf{P}\left(\hat{\mathcal{B}} \neq \mathcal{B}\right) \leq \sum_{\mathbf{x}^d \in \operatorname{supp} \mathbf{X}^n} \mathbf{P}\left(\operatorname{sign}\left(\hat{\mathbb{E}}_n + \epsilon_n\right) \neq \operatorname{sign}\left(\mathbb{E}\right)\right)$$

There are two cases three consider:

Case 1: $\mathbb{E} > 0$.

$$\begin{aligned}\mathbf{P}\left(\text{sign}\left(\hat{\mathbb{E}}_n + \epsilon_n\right) \neq \text{sign}\left(\mathbb{E}\right)\right) =& \mathbf{P}\left(\left(\hat{\mathbb{E}}_n + \epsilon_n < 0\right) \cap \left(\mathbb{E} > 0\right)\right) \\ =& \mathbf{P}\left(\hat{\mathbb{E}}_n - \mathbb{E} + \mathbb{E} + \epsilon_n < 0\right) \\ =& \mathbf{P}\left(\mathbb{E} + \epsilon_n < \mathbb{E} - \hat{\mathbb{E}}_n\right) \\ \leq& \mathbf{P}\left(0 < \mathbb{E} - \hat{\mathbb{E}}_n\right) \to 0\end{aligned}$$

as $n \to 0$, since $\hat{\mathbb{E}}_n \xrightarrow{p} \mathbb{E}$.

Case 2: $\mathbb{E} < 0$.

Since $\mathbb{E} < 0$ and given that $\epsilon_n \to 0$ as $n \to \infty$. Then, there exists a $\delta > 0$ such that

$$-\mathbb{E} - \epsilon_n \geq \delta > 0$$

as $n \to \infty$.

$$\begin{aligned}\mathbf{P}\left(\text{sign}\left(\hat{\mathbb{E}}_n + \epsilon_n\right) \neq \text{sign}\left(\mathbb{E}\right)\right) =& \mathbf{P}\left(\left(\hat{\mathbb{E}}_n + \epsilon_n \geq 0\right) \cap \left(\mathbb{E} < 0\right)\right) \\ =& \mathbf{P}\left(\hat{\mathbb{E}}_n - \mathbb{E} \geq -\mathbb{E} - \epsilon_n\right) \\ \leq& \mathbf{P}\left(\hat{\mathbb{E}}_n - \mathbb{E} \geq \delta > 0\right) \to 0\end{aligned}$$

as $n \to \infty$.

Case 3: $\mathbb{E} = 0$.

$$\begin{aligned}\mathbf{P}\left(\text{sign}\left(\hat{\mathbb{E}}_n + \epsilon_n\right) \neq \text{sign}\left(\mathbb{E}\right)\right) =& \mathbf{P}\left(\left(\hat{\mathbb{E}}_n + \epsilon_n < 0\right) \cap \left(\mathbb{E} = 0\right)\right) \\ =& \mathbf{P}\left(\hat{\mathbb{E}}_n - \mathbb{E} > \epsilon_n\right) \\ =& \mathbf{P}\left(\epsilon_n^{-1} r_n^{-1} r_n(\hat{\mathbb{E}}_n - \mathbb{E}) > 1\right) \to 0\end{aligned}$$

since $\epsilon_n^{-1} r_n^{-1} = o_p(1)$ and $r_n(\hat{\mathbb{E}}_n - \mathbb{E}) = O_p(1)$.

Finally,

$$\mathbf{P}\left(R(\hat{\mathcal{B}}) \neq R(\mathcal{B})\right) \leq \mathbf{P}\left(\hat{\mathcal{B}} \neq \mathcal{B}\right) \to 0$$

as $n \to 0$. This concludes the proof of Part 1.

Part 2:

For any $\delta > 0$, if

$$H(\hat{\mathcal{B}}, \mathcal{B}) \geq \delta \Rightarrow H(\hat{\mathcal{B}}, \mathcal{B}) \neq 0 \Rightarrow \hat{\mathcal{B}} \neq \mathcal{B}$$

Thus

$$\mathbf{P}\left(H(\hat{\mathcal{B}}, \mathcal{B}) \geq \delta\right) \leq \mathbf{P}\left(\hat{\mathcal{B}} \neq \mathcal{B}\right) \to 0$$

as $n \to \infty$. The same argument is used to prove $H(R(\hat{\mathcal{B}}), R(\mathcal{B})) \xrightarrow{p} 0$ as $n \to \infty$. ■

C Sharp Bounds

In this section, I illustrate the recursive procedure introduced in Komarova (2013) to derive sharp bounds for each element in β_0 of the network formation model given by equation (1). Suppose that the goal is to find the sharp bounds for the Kth component of β_0, i.e., β_K. The recursive procedure then starts by excluding one unknown variable at each iteration from the following system.

$$\begin{aligned}
z_{1,1}^{(1)} + z_{1,1}^{(2)} b_2 + z_{1,1}^{(3)} b_3 + \cdots + z_{1,1}^{(K)} b_K &\geq 0, \\
z_{2,1}^{(1)} + z_{2,1}^{(2)} b_2 + z_{2,1}^{(3)} b_3 + \cdots + z_{2,1}^{(K)} b_K &\geq 0, \\
&\vdots \\
z_{M,1}^{(1)} + z_{M,1}^{(2)} b_2 + z_{M,1}^{(3)} b_3 + \cdots + z_{M,1}^{(K)} b_K &\geq 0, \\
&\vdots \\
z_{M,D}^{(1)} + z_{M,D}^{(2)} b_2 + z_{M,D}^{(3)} b_3 + \cdots + z_{M,D}^{(K)} b_K &\geq 0,
\end{aligned} \tag{A_1}$$

Consider excluding the unknown β_2. The ijth inequality in system (A_1)

$$z_{i,j}^{(1)} + z_{i,j}^{(2)} b_2 + z_{i,j}^{(3)} b_3 + \cdots + z_{i,j}^{(K)} b_K \ \geq 0.$$

If $z_{i,j}^{(2)} \geq 0$, then the ijth inequality is equivalent to

$$-\frac{z_{i,j}^{(1)}}{z_{i,j}^{(2)}} - \frac{z_{i,j}^{(3)}}{z_{i,j}^{(2)}} b_3 - \cdots - \frac{z_{i,j}^{(K)}}{z_{i,j}^{(2)}} b_K \leq b_2$$

Alternatively, if $z_{i,j}^2 < 0$ then the ijth inequality is equivalent to

$$-\frac{z_{i,j}^{(1)}}{z_{i,j}^{(2)}} - \frac{z_{i,j}^{(3)}}{z_{i}^{(2)}} b_3 - \cdots - \frac{z_{i,j}^{(K)}}{z_{i}^{(2)}} b_K \geq b_2$$

Suppose the system (A_1) has N_1 inequalities with $z_{\cdot}^2 > 0$, N_2 inequalities with $z_{\cdot}^2 < 0$ and N_3 inequalities with $z_{\cdot}^2 = 0$. Then, the system (A_1) is equivalent to

$$\begin{aligned} L_i(b_3, \cdots, b_K) &< b_2, \qquad i = 1, \cdots, N_1, \\ U_j(b_3, \cdots, b_K) &> b_2, \qquad j = 1, \cdots, N_2, \\ Z_r(b_3, \cdots, b_K) &> 0, \qquad r = 1, \cdots, N_3, \end{aligned}$$

where $L_i(\cdot), U_j(\cdot), Z_r(\cdot)$ are linear functions of $b_3, \cdots, b_K$ and do not depend on b_2.

The previous system implies the following simplified system with $K - 2$ unknown variables and $(N_1 * N_2) + N_3$ inequalities

$$\begin{aligned} &U_j(b_3, \cdots, b_K) > L_i(b_3, \cdots, b_K), && i = 1, \cdots, N_1, \quad j = 1, \cdots, N_2, \\ &Z_r(b_3, \cdots, b_K) > 0, && r = 1, \cdots, N_3. \end{aligned} \tag{A_2}$$

By excluding an additional unknown variable from system (A_2), other than b_K, a simplified system with $K - 3$ unknown variables is obtained. The order of elimination is arbitrary. The process is repeated on the simplified systems until a system that has b_K as the only unknown variable is reached. The last simplified system has the following form:

$$\begin{aligned} u_l + v_l b_K &> 0, \quad l = 1, \cdots, L, \\ w_m &> 0, \quad m = 1, \cdots, M, \end{aligned} \tag{A_{K-1}}$$

with $u_l, v_l, w_m \in \mathbb{R}$ for $l = 1, \cdots, L$; $m = 1, \cdots, M$, and $v_l \neq 0$, $l = 1, \cdots, L$.

Then, the lower and upper bounds for β_K are derived from the simplified system (A_{K-1}) as follows:

$$\begin{aligned} \underline{b}_K &= \max_{l=1,\cdots,L} \left\{ -\frac{u_l}{v_l} : v_l > 0 \right\}, \\ \bar{b}_K &= \min_{l=1,\cdots,L} \left\{ -\frac{u_l}{v_l} : v_l < 0 \right\}. \end{aligned}$$

As previously mentioned, the sharp bounds for the rest of the components in the parameter of interest, $\beta_j, j \neq k$ are computed by repeating the same recursive procedure.

The identification set can be approximated by the smallest multidimensional rectangle superset that covers the identification set. The multidimensional rectangle superset is defined as the Cartesian product of the sets specified by the sharp bounds of each component of the parameter of interest. That is,

$$R(\mathcal{B}_0) \equiv \prod_{k=2}^{K} [\underline{b}_k,\ \bar{b}_k].$$

D Thin Set

Table 5: Thin Set Simulations: Stochastic Dominance and Sparsity

	Empty		Sparse		Dense	
	E [Degree]	$P[\Omega_n]$ (%)	E [Degree]	$P[\Omega_n]$ (%)	E [Degree]	$P[\Omega_n]$ (%)
$\lambda = 0.25$						
Log	20.30	4.32	49.53	16.71	97.15	0.06
LnN	9.34	1.01	36.98	13.73	95.88	0.11
N	19.47	3.84	49.52	18.11	98.56	0.00
Gam	19.54	3.87	49.36	19.63	87.12	1.56
T	28.59	8.30	49.45	18.25	90.54	1.03
$\lambda = 0.5$						
Log	23.56	5.71	49.44	16.95	95.48	0.21
LnN	10.58	1.28	36.62	13.72	92.34	0.47
N	22.44	5.03	49.39	18.58	98.13	0.01
Gam	23.11	5.41	49.32	21.04	76.73	4.72
T	33.90	11.29	49.30	18.84	84.53	2.71
$\lambda = 0.75$						
Log	27.81	7.88	49.30	17.14	91.75	0.86
LnN	12.38	1.74	36.06	13.64	80.39	3.52
N	26.38	6.92	49.21	18.82	96.75	0.07
Gam	27.08	7.34	49.20	22.42	54.40	11.08
T	40.51	15.00	49.26	19.29	72.11	7.27

Notes: N=100, M=250.

Table 6: Thin Set Simulations: Stochastic Dominance and Sparsity

	Empty		Sparse		Dense	
	E [Degree]	$P[\Omega(ijkl)]$ (%)	E [Degree]	$P[\Omega(ijkl)]$ (%)	E [Degree]	$P[\Omega(ijkl)]$ (%)
$\lambda = 0.25$						
Logistic	51.17	4.31	124.53	16.74	244.45	0.05
Lognormal	23.21	0.99	93.02	13.69	241.07	0.10
Normal	49.12	3.89	124.43	18.12	247.97	0.00
Gamma	49.19	3.90	124.47	19.74	219.29	1.54
T-student	72.69	8.48	124.33	18.29	227.69	1.03
$\lambda = 0.5$						
Logistic	59.68	5.82	124.33	16.91	240.29	0.20
Lognormal	26.93	1.32	92.53	13.85	232.05	0.48
Normal	56.84	5.14	124.31	18.54	246.72	0.01
Gamma	57.77	5.33	123.90	20.93	192.65	4.74
T-student	85.04	11.24	124.22	18.89	212.86	2.67
$\lambda = 0.75$						
Logistic	70.192	7.976	124.03	17.15	230.96	0.84
Lognormal	31.93	1.840	91.239	13.68	203.22	3.38
Normal	66.914	7.036	123.85	18.84	243.42	0.07
Gamma	67.846	7.274	123.57	22.49	137.28	11.02
T-student	101.83	14.92	123.91	19.46	181.38	7.27

Notes: N=250, M=250.

Table 7: Thin Set Simulations: Homogeneous Network

$\mu = 10 * \text{Bernoulli(p)} + (-5) * (1 - \text{Bernoulli(p)})$				
N=100	$E\,[\text{Degree}]$	$P\,[\Omega(ijkl)]\,(\%)$ (%)	Jaccard SI (Mean) (Mean)	Cosine SI (Mean) (Mean)
$p = 0.2$				
Log	37.66	0.38	0.55	0.70
LnN	20.52	0.83	0.35	0.53
N	36.66	0.31	0.60	0.73
Gam	31.14	0.42	0.56	0.70
T	27.30	0.34	0.57	0.70
$p = 0.8$				
Log	92.56	0.12	0.87	0.93
LnN	83.46	1.16	0.74	0.85
N	95.10	0.01	0.91	0.95
Gam	94.42	0.05	0.90	0.94
T	93.26	0.10	0.88	0.93

Notes: Number of Monte Carlo Simulations, M=250.

E Monte Carlo Simulations

Table 8: Monte Carlo Simulations: Logistic(0,1)

	Pairwise Difference				Tetrad Logit				$P(\Omega_n)$	E(Degree)
	Median	Mean	Bias(%)	RMSE	Median	Mean	Bias(%)	RMSE		
$N = 100$									5.924%	47.1344
$\beta_2/\beta_1 = 1.5$	1.508	1.675	11.687	1.540	1.472	1.528	1.864	0.468		
$\beta_3/\beta_1 = -1.5$	-1.476	-1.488	0.747	0.231	-1.685	-1.724	14.938	0.4686		
$N = 250$									6.661%	120.103
$\beta_2/\beta_1 = 1.5$	1.499	1.491	0.551	0.134	1.490	1.513	0.915	0.224		
$\beta_3/\beta_1 = -1.5$	-1.485	-1.496	0.269	0.077	-1.662	-1.675	11.668	0.224		
$N = 500$									5.890%	233.264
$\beta_2 = 1.5$	1.502	1.502	0.182	0.037						
$\beta_3 = -1.5$	-1.497	-1.489	0.674	0.118						

Note: Number of Monte Carlo simulations M=500, correlation parameter $\lambda = 0.25$

Table 9: Monte Carlo Simulations: Logistic(0,1)

	Pairwise Difference				Tetrad Logit				$P(\Omega_n)$	E(Degree)
	Median	Mean	Bias(%)	RMSE	Median	Mean	Bias(%)	RMSE		
$N = 100$									10.710 %	48.102
$\beta_2/\beta_1 = 1.5$	1.517	1.688	12.578	1.726	1.499	1.563	4.246	0.393		
$\beta_3/\beta_1 = -1.5$	-1.515	-1.510	0.697	0.131	-1.727	-1.775	18.335	0.393		
$N = 250$									10.826%	116.820
$\beta_2/\beta_1 = 1.5$	1.504	1.622	8.147	1.611	1.504	1.525	1.681	0.274		
$\beta_3/\beta_1 = -1.5$	-1.495	-1.501	0.065	0.061	-1.694	-1.698	13.217	0.274		
$N = 500$									10.694%	240.765
$\beta_2 = 1.5$	1.507	1.507	0.510	0.030						
$\beta_3 = -1.5$	-1.499	-1.500	0.030	0.026						

Note: Number of Monte Carlo simulations M=500, correlation parameter $\lambda = 0.75$

Table 10: Monte Carlo Simulations: Normal(0,2)

	Pairwise Difference				Tetrad Logit				$P(\Omega_n)$	E(Degree)
	Median	Mean	Bias(%)	RMSE	Median	Mean	Bias(%)	RMSE		
$N = 100$									4.577%	45.958
$\beta_2/\beta_1 = 1.5$	1.506	1.702	13.516	2.248	1.471	1.523	1.567	0.359		
$\beta_3/\beta_1 = -1.5$	-1.484	-1.493	0.455	0.237	-1.888	-1.931	28.796	0.359		
$N = 250$									4.607%	116.432
$\beta_2/\beta_1 = 1.5$	1.518	1.514	0.979	0.062	1.499	1.494	0.376	0.216		
$\beta_3/\beta_1 = -1.5$	-1.512	-1.508	0.556	0.064	-1.901	-1.903	26.883	0.216		
$N = 500$									4.607%	233.855
$\beta_2/\beta_1 = 1.5$	1.502	1.503	0.199	0.032						
$\beta_3/\beta_1 = -1.5$	-1.499	-1.502	0.139	0.032						

Note: Number of Monte Carlo simulations M=500, correlation parameter $\lambda = 0.25$

Table 11: Monte Carlo Simulations: Normal(0,2)

	Pairwise Difference				Tetrad Logit				$P(\Omega_n)$	E(Degree)
	Median	Mean	Bias(%)	RMSE	Median	Mean	Bias(%)	RMSE		
$N = 100$									10.269%	47.445
$\beta_2/\beta_1 = 1.5$	1.492	1.714	14.298	2.362	1.537	1.595	6.383	0.427		
$\beta_3/\beta_1 = -1.5$	-1.483	-1.494	0.363	0.115	-1.976	-2.001	33.430	0.427		
$N = 250$									10.275%	120.105
$\beta_2/\beta_1 = 1.5$	1.561	1.495	0.326	1.844	1.489	1.493	0.448	0.222		
$\beta_3/\beta_1 = -1.5$	-1.503	-1.502	0.169	0.041	-1.916	-1.917	27.825	0.222		
$N = 500$									10.270%	239.686
$\beta_2/\beta_1 = 1.5$	1.504	1.503	0.230	0.020						
$\beta_3/\beta_1 = -1.5$	-1.499	-1.498	0.101	0.021						

Note: Number of Monte Carlo simulations M=500, correlation parameter $\lambda = 0.75$