# Essays on Econometrics of Network Models

by

Luis Enrique Candelaria Barrera

Department of Economics
Duke University

Date: ____________________
Approved:

Shakeeb Khan, Co-Supervisor

Arnaud Maurel, Co-Supervisor

Federico A. Bugni

Matthew A. Masten

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2017

# Abstract

Essays on Econometrics of Network Models

by

Luis Enrique Candelaria Barrera

Department of Economics
Duke University

Date: ____________________
Approved:

____________________
Shakeeb Khan, Co-Supervisor

____________________
Arnaud Maurel, Co-Supervisor

____________________
Federico A. Bugni

____________________
Matthew A. Masten

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2017

# Abstract

Social networks affect a broad class of economic activities. The three chapters of my dissertation study social networks from two different lines of research. The first line of research examines the formation process of a social network. In Chapter 2, I introduce a new identification strategy and a semiparametric estimator for the formation process of an undirected network with additive agent-specific fixed effects. In Chapter 3, I analyze the formation process of a directed network with a broader type of unobserved heterogeneity. This heterogeneity is modeled as interactive fixed effects. The second line of my research complements the first approach by exploring the influence that network structures have on different economic activities. In Chapter 4, I recover the endogenous and exogenous social effects in a high-dimensional panel data model with an unobserved network structure.

To my parents. A mis papás.

# Contents

# List of Tables

# List of Figures

# List of Abbreviations and Symbols

## Symbols

| | |
|---|---|
| $\xrightarrow{a.s.}$ | Almost sure convergence. |
| $\xrightarrow{p}$ | Convergence in probability. |
| $\xrightarrow{d}$ | Convergence in distribution. |
| $\|\|\cdot\|\|$ | Frobenius matrix norm. |
| $H(A, B)$ | Hausdorff metric. |
| $\mathcal{B}_0$ | Identified set. |
| $\text{Med}\{X\}$ | Median of the random variable $X$. |
| $\nabla_m$ | *mth* Partial derivative operator. |
| $\text{sign}(X)$ | Sign function. |
| $A \subseteq B$ | A is a subset of B. |
| $\text{supp}(X)$ | Support of the random variable $X$. |

# Acknowledgements

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 have received advice and comments from Rachel Kranton, Jia Li, Fu Ouyang, Adam Rosen, Takuya Ura and Yichong Zhang.

I am grateful to many wonderful friends at Duke, especially to Irem Altan, Vasco Botelho, Irving de Lira, Elisa Maffioli, Jose Martinez, Soroush Ghazi, Marat Kussainov, Catherine Moon, Zdravko Paskalev, and Takuya Ura. These great friends made my years in Durham more enjoyable.

I will be eternally grateful to Aslı Cansunar, for her continuous encouragement, advice, and for standing behind me through all this journey. Her unconditional friendship and kindness fill me with happiness every day and help me find the strength through the toughest moments. She inspires me to become a better person, and I would never have completed my Ph.D. degree without her.

Last but not least, I want to thank my family in Mexico. My brothers, Jose Alberto y Miguel Angel Candelaria Barrera, and my nephew, Gabriel A. Candelaria, for their friendship, patience, and all the lessons that they have taught me through the years. This work could not have been written without them.

To my parents, Alberto Candelaria and Josefina Barrera, for their unconditional love, support, and guidance throughout my life. Nothing of what I have achieved could have been possible without them. This work is dedicated to them.

# 1

# Introduction

This dissertation consists of three self-contained essays on the econometrics of social networks. Recent research in social sciences has provided extensive evidence of the influence that network structures have on social or economic behavior, and hence the outcomes. For example, the literature on peer effects has studied the effect of network interactions on crime (Calvo-Armengol and Jackson 2004; Bayer et al. 2009), educational achievements (Sacerdote 2001; Calvó-Armengol et al. 2009; Masten 2015), health (Christakis and Fowler 2007; Cohen-Cole and Fletcher 2008), the spread of information (Campbell 2013; Banerjee et al. 2013), job search (Calvo-Armengol and Jackson 2004) and risk sharing (Fafchamps and Gubert 2007; Kinnan and Townsend 2012).

The substantial impact that social networks have on describing economic activities has intensified their study in economics. The research on network structures has focused mainly on two aspects. The first aspect examined is the formation and evolution of social networks. The second aspect analyzed is the influence that network structures have on economic activities. I study the formation process of undirected and directed social networks in Chapter 2 and 3 of this dissertation. In Chapter 4, I explore the recovery of network effects when the underlying structure of interactions is unknown.

In Chapter 2, I analyze the formation process of an undirected network in the presence of multiple and agent-specific fixed effects with an additive structure. The formation model is semiparametric. That is, 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 characterize 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 introduce a semiparametric estimator for the vector of coefficients that is consistent and asymptotically normal as the number of individuals in the network increases. I illustrate the methods introduced by analyzing the determinants of a friendship network using the Add Health dataset.

In Chapter 3, I study the formation of a directed network when agents have preferences for similarity on observed characteristics, as well as complementarity on unobserved agent-specific factors. The complementarity on unobserved factors is modeled as the interaction of agent-specific fixed effects. Such directed networks are of interest in many economic applications, such as in the adoption and diffusion of technology. In this chapter, I recover the contribution of homophily on observed attributes on forming a directed link from the complementarity on the unobserved heterogeneous components. Furthermore, I show that the semiparametric methods developed in Chapter 2 can be extended to analyze a directed network, and used to identify the coefficients that measure homophily on observed attributes.

Chapter 4 of this dissertation focuses on understanding the effects that network structures have on economic activities. Specifically, I study a panel data model of strategic interactions when the structure of the network is unobserved and unrestricted. The endogenous and exogenous network effects are allowed to be heterogeneous across individuals. I develop an identification strategy that point identifies the endogenous and exogenous network effects. I use a high-dimensional approximately sparse model to consistently estimate the structure of interactions when the network is stable over time and sparse. Furthermore, I provide Oracle convergence rates for the estimates of the strategic interaction parameters.

# 2

# A Semiparametric Network Formation Model with Multiple Linear Fixed Effects

## 2.1 Introduction

People tend to connect with individuals whom they share similar observed attributes. This observation is known as homophily (McPherson et al. 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]

$$D_{ij}^{n} = \mathbf{1}\left[X_{ij}^{n\prime}\beta_0 + \mu_i + \mu_j - \varepsilon_{ij}^{n} \geqslant 0\right], \tag{2.1}$$

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, $\beta_0$ is a $K$-dimensional vector of unknown parameters, $\mu_i$ and $\mu_j$ are unobserved and agent-specific random variables, and $\varepsilon_{ij}^{n}$ is an unobserved and link-specific disturbance term.

Intuitively, equation (2.1) 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 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 ex-

[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 2.2, I derive the network formation decision in equation (2.1) as a stability condition in a random utility model with transferable utilities.

ogenous 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 (2.1) allow for heterogeneous net benefits across each individuals 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 semi-parametric 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 $\beta_0$ in equation (2.1). 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 $\beta_0$. 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:

$$D_{ij}^n = \mathbf{1}\left[X_{ij}^{n\prime}\beta_0 + \mu_i + g(\mu_i, \mu_j) - \varepsilon_{ij}^n \geqslant 0\right], \tag{2.2}$$

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

Specification (2.2) 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 $\mu_j$ may affect the linking decisions of individual $i$, differently for different $j$ due to the unobserved complementarities on the fixed effects. Equation (2.2) nests the additive and agent-specific fixed effects model as a special case. Specifically, equation (2.2) degenerates to equation (2.1) when $g(\mu_i, \mu_j) = \mu_j$, and $\varepsilon_{ij}^n$ 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 $\beta_0$ in equation (2.2). 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 $\beta_0$ 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 (2015). 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 (2015) will fail to partial out the fixed effects and therefore point identify $\beta_0$ 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 (2.1), 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 2.3.1, 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 $\beta_0$ in equation (2.1).

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 2.4, 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 (2.1), 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 $\beta_0$ 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 2.4.

The network formation model that I analyze is related to the empirical games literature. Specifically, the model in equation (2.1) 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 et al. (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 (2.1) with a parametric meeting process that determines how the links are sequentially revised over time (Christakis et al. 2010; Snijders et al. 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 2.2 formalizes the network formation model. Section 2.3 describes the identification strategy and states point identification and partial identification results. In section 2.4, I outline the semiparametric estimator and I show consistency and asymptotic normality. Section 2.5 reports some Monte Carlo simulations. Section 2.6 considers an empirical application analyzing a friendship network. Section 2.7 concludes by summarizing and suggesting areas for future research. The appendix 2.8 collects all the proofs of the paper.

## 2.2 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 characteristics $Z^n_{ij} = (D^n_{ij}, X^n_{ij})$, and an unobserved dyad-specific disturbance term $\varepsilon^n_{ij}$. The first element in the vector of observed attributes $Z^n_{ij}$ denotes the link status in that specific dyad, $D^n_{ij}$. The second element in $Z^n_{ij}$ 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 2.6). Conditions on the support of the vector of exogenous attributes $X_{ij}^n$ are discussed in section 2.3. The unobserved dyad-specific disturbance component $\varepsilon_{ij}^n$ 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 ($g(X_j^n, X_i^n) = \left(X_i^n - X_j^n\right)^2$) or complementarities ($g(X_j^n, X_i^n) = X_i^n \cdot X_j^n$) 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

$$V_{ij}(\mathbf{X}^n, \mu_j, \varepsilon_{ij}^n) = u_{ij}(\mathbf{X}^n) + \mu_j - \frac{1}{2}\varepsilon_{ij}^n, \tag{2.3}$$

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

$$u_{ij}(\mathbf{X}^n) \equiv \frac{1}{2} X_{ij}^{n\prime} \beta_0, \tag{2.4}$$

where $\beta_0$ 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

$$V_{ij}(\mathbf{X}^n, \mu_j, \varepsilon_{ij}^n) + V_{ji}(\mathbf{X}^n, \mu_i, \varepsilon_{ij}^n) \equiv X_{ij}^{n\prime} \beta_0 + \mu_i + \mu_j - \varepsilon_{ij}^n. \tag{2.5}$$

In addition of 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 $\varepsilon_{ij}^n$.

Equation (2.5) implies that individuals $i$ and $j$ in the dyad $(i, j)$ have valuations only 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 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 $D_{ij}^n = 1$, $V_{ij}(\mathbf{X}^n, \mu_j, \varepsilon_{ij}^n) + V_{ji}(\mathbf{X}^n, \mu_i, \varepsilon_{ij}^n) \geqslant 0$;*

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

## 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 (2.1). I then provide partial identification results under a weaker set of assumptions. In section 2.3.1, I describe the identification strategy. Section 2.3.2 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 2.3.3, 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 $\beta_0$. In section 2.3.4, 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.

### *2.3.1 Identification Strategy*

The intuition behind the identification strategy is summarized in figure 2.1. Consider the subnetwork formed by the undirected links between agents $i, j, k, l \in \mathcal{N}_n$. All the links represented in figure 2.1 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:

$$\mathbb{P}\left[D_{il} = 1 \mid \mathbf{X}^n = x, \tilde{\mu} = \mu\right] < \mathbb{P}\left[D_{ik} = 1 \mid \mathbf{X}^n = x, \tilde{\mu} = \mu\right]. \tag{2.6}$$

If the link-specific unobserved random variables are identically distributed, then

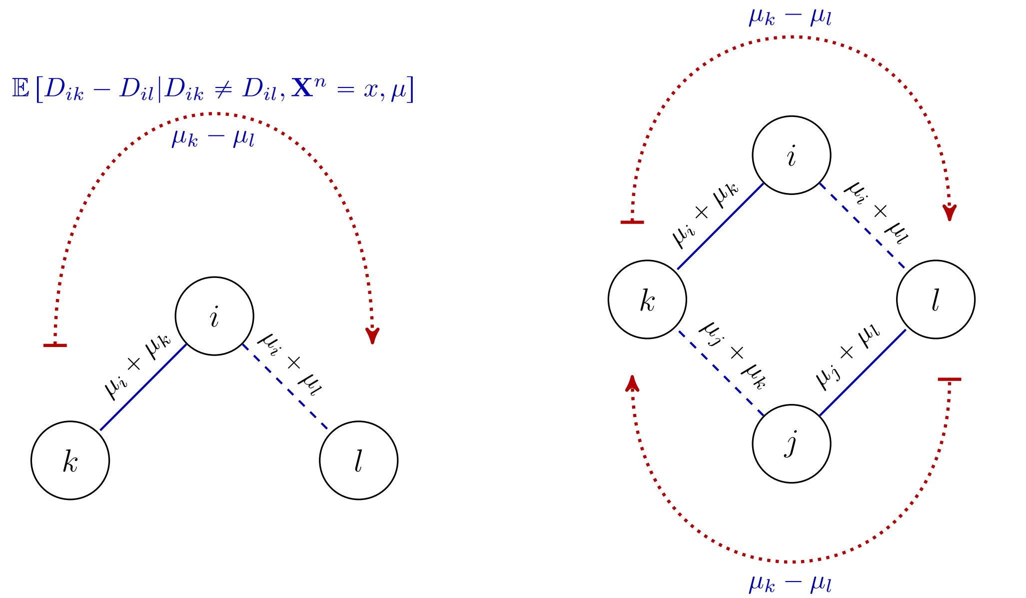



Diagram 1: Undirected links in dyads $(i, k)$ and $(i, l)$. Diagram 2: Undirected links in tetrad $(i, j, k, l)$.

FIGURE 2.1: Subnetwork formed by agents $i, j, k, l \in \mathcal{N}_n$.

the equation (2.6) 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 $\mu_i$ is a common element. Therefore, the within-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 $\{\mathbb{E}[D_{ik} - D_{il} | D_{ik} \neq D_{il}, \mathbf{X}^n = x, \mu]\}$ 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 $\mu_k - \mu_l$.

Specifically, the following equation holds for the conditional median of the net difference, $D_{ik} - D_{il}$. The proof is in the appendix 2.8

$$\mathrm{Med}(D_{ik} - D_{il} | \mathbf{X}^n = \mathbf{x}, D_{il} \neq D_{ik}) = \mathrm{sign}\left((x_{ik} - x_{il})'\beta_0 + (\mu_k - \mu_l)\right), \tag{2.7}$$

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

Equation (2.7) 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 (2.1). Therefore, equation (2.7) does not identify $\beta_0$. In other words, the typical maximum score identification strategy fails to point identify $\beta_0$. I provide a more detailed explanation of this result in section 2.3.3.

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 2.3.2, I show that the following equation for the conditional median of the pairwise difference holds:

$$\text{Med}\{[D_{ik} - D_{il}] - [D_{jk} - D_{jl}] \mid \mathbf{X}^n = \mathbf{x}, D_{ik} \neq D_{il}, D_{jk} \neq D_{jl}, D_{ik} \neq D_{jk}\} = \\ 2 \times \text{sign}\left\{[(x_{ik} - x_{il}) - (x_{jk} - x_{jl})]' \beta_0\right\}, \quad (2.8)$$

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

Equation (2.8) is fully characterized by the observed variables $(D^n, \mathbf{X}^n)$. In sections 2.3.2 and 2.3.4, I show that equation (2.8) can be used to point identify $\beta_0$ 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 (2.8) ensures the sufficient within-individual and across-individuals variation in the linking decisions to identify $\beta_0$. 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 $\mu_i$ 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 $\beta_0$. I show in subsection 2.3.3 that if the event $\{D_{ik} \neq D_{il}, D_{jl} \neq D_{jk}, D_{ik} \neq D_{jk}\}$ has probability zero then equation (2.8) is uninformative to identify $\beta_0$.

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 almost empty networks, dense networks, and homogeneous networks. A network is homogeneous when the individuals

[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.

establish the similar connections with probability one. For example, in the network formation model given by equation (2.1), the equilibrium network structure will be homogeneous when agents establish their connections based mainly only on their preferences for individual-specific attributes.

### *2.3.2 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 $\beta_0$.

**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 $\varepsilon_{i1}$ given $\{\mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu\}$.

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]

[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 $\varepsilon_{ji}$ implies that $F_{\varepsilon_{ik}|\mathbf{x},\mu} = F_{\varepsilon_{jl}|\mathbf{x},\mu}$.

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 (2015); Leung (2015a) and Menzel (2015).

Let $\Delta_{kl} X_i = X_{ik} - X_{il}$ 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 $\beta_0$ 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 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 $\beta_0$ 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 2.3.4, I give alternative sufficient conditions for point identification when regressors are continuous with bounded support.

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

$$\sup(\mu_i \mid \mathbf{X}^n = \mathbf{x}) \subseteq [B_L, B_U],$$

*for any* $\mathbf{x} \in \text{supp}(\mathbf{X}^n)$*, and given* $B_L, B_U < \infty$*.*

A3 states that the agent-specific fixed effects have bounded support. This assumption allows for discrete or continuously distributed fixed effects. Furthermore, their distribution could be heterogeneous as long as common bounds for their support exist.

This assumption, jointly with A2, 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 1.**

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

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

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

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

2. *Let assumptions A1 - A3 hold. Then $\beta_0$ 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 $\beta_0$. 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 $\beta_0$ 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

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

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:

$$\begin{aligned} Y_{kl}^{(i)} = 1 \quad &\Leftrightarrow \quad Y_{kl}^{(j)} = -1, \\ Y_{kl}^{(i)} = -1 \quad &\Leftrightarrow \quad Y_{kl}^{(j)} = 1. \end{aligned} \tag{2.9}$$

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 (2.9) follows analogous arguments. Thus, the random variable $Y_{kl}^{(i)} - Y_{kl}^{(j)}\Big| \mathbf{X}^n = \mathbf{x}, \Omega(ijlk)$ has a Bernoulli distribution with support $\{-2, 2\}$, and

$$\begin{aligned} &\text{Med}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = x, \Omega(ijlk)\right] \\ &\qquad = 2 \times \text{sign}\left\{\mathbb{P}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} = 2 | \mathbf{X}^n = x, \Omega(ijlk)\right]\right. \\ &\qquad\qquad \left. - \mathbb{P}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} = -2 | \mathbf{X}^n = x, \Omega(ijlk)\right]\right\}. \end{aligned} \tag{2.10}$$

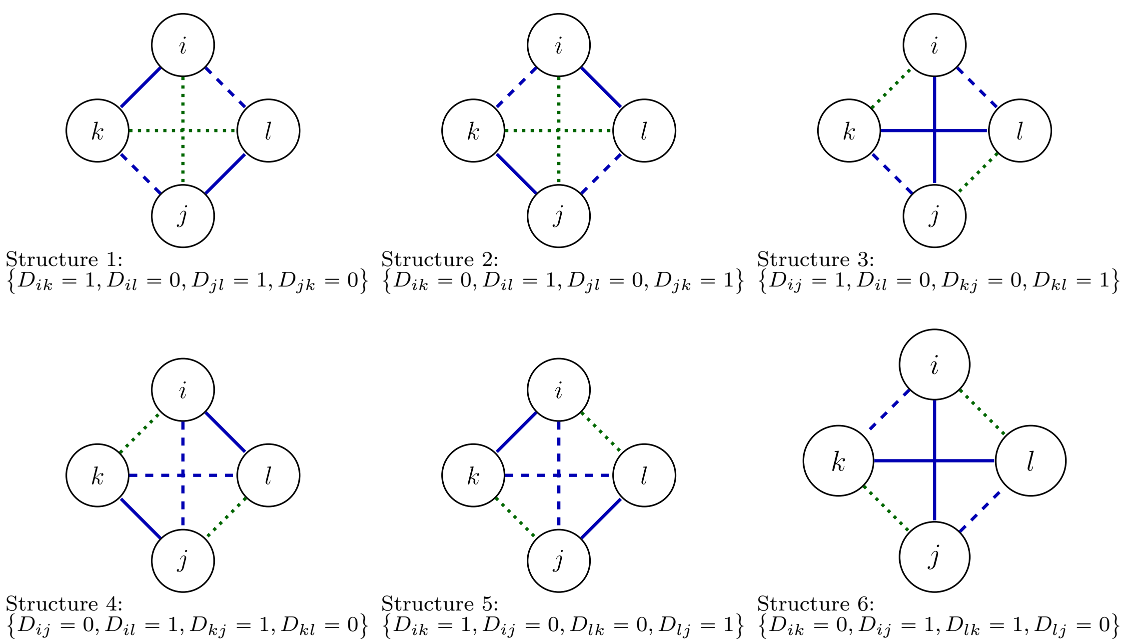
Finally, following the intuition described in section 2.3.1, given A1 and the set $\Omega(ijlk)$ the pairwise difference of the linking decisions for agents $i$ and $j$ implies:

$$\text{sign}\left\{\mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\right]\right\} = \text{sign}\left\{\left[(x_{ik} - x_{il}) - (x_{jk} - x_{jl})\right]' \beta_0\right\}, \tag{2.11}$$

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

The proof of part 1 in Theorem 1 is concluded by showing that the right-hand side of equation (2.10) is equal to twice the left-hand side of equation (2.11). This result follows from $Y_{kl}^{(i)} - Y_{kl}^{(j)} \Big| \mathbf{X}^n = \mathbf{x}, \Omega(ijlk)$ 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 $\beta_0$. Figure 2.2 below depicts all the subnetworks contained in $\Omega(ijlk)$.




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

Consider the subnetwork structure 1 in figure 2.2, which is given by:

$$\{D_{ik} = 1, D_{il} = 0, D_{jl} = 1, D_{jk} = 0\}\,. \tag{2.12}$$

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 2.2.

A2 is a sufficient condition for point identification of $\beta_0$. This assumption requires the existence of at least one covariate with full support. If this assumption fails, then equation (2.11) yields a collection of moment inequalities that can be used to partially identify $\beta_0$. 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 (2.11). That is

$$\begin{aligned}\mathcal{B}_0 = \Big\{ & b \in \mathbb{R}^k : \mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\right] \gtreqless 0 \Leftrightarrow [\Delta_{kl} x_i - \Delta_{kl} x_j]' \, b \gtreqless 0\,, \\ & \forall\; i, j, k, l \in \mathcal{N}_n \text{ and} \\ & \forall \mathbf{x} = (x_{12}, \cdots, x_{ik}, \cdots, x_{il}, \cdots, x_{jk}, \cdots, x_{jl}, \cdots, x_{n-1,n}) \in \operatorname{supp}(\mathbf{X}^n) \cap \mathcal{X}_B \Big\}\,,\end{aligned} \tag{2.13}$$

The identifying set $\mathcal{B}_0$ will be used in section 2.3.4 to derive bounds for each elements in the vector of unknowns $\beta_0$ when assumption A2 fails.

### 2.3.3 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 $\beta_0$. Intuitively, if $\Omega_n$ 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 $\beta_0$ 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.

*Thin Set Identification*

Intuitively, the set $\Omega(ijkl)$ ensures sufficient variation to partial out the agent-specific fixed effects. If the class $\Omega_n$ has probability zero, then $Y_{kl}^{(i)} - Y_{kl}^{(j)}$ will not have enough information to identify $\beta_0$. I formalize this result in the following theorem.

**Theorem 2.** *Let assumptions A1 - A3. If the class $\Omega_n$ has probability zero, then the median of the random variable $Y_{kl}^{(i)} - Y_{kl}^{(j)} \Big| \mathbf{X}^n = \mathbf{x}$ 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 $\beta_0$ in terms of*

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

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

The class $\Omega_n$ 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 2.8.2, I characterize all the subnetwork structures that are not consistent with $\Omega(ijkl)$, and under which the class $\Omega_n$ does not have identification power.

Specifically, some examples of network structures for which the class $\Omega_n$ 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.** *Given the network formation model in* (2.1), *the class* $\Omega_n$ *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 is a dense network since a link is established within any dyad $(i,j) \in \mathcal{N}_n^{(2)}$ with probability one,

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

Under condition (2), the realized network is empty since no link is created within any dyad $(i,j) \in \mathcal{N}_n^{(2)}$ with probability one,

$$\mathbb{P}\left[D_{ij} = 1 \mid \mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu\right] = 0.$$

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 $\Omega_n$ has probability zero under the conditions (1), (2), and (3) in proposition 3 because the equilibrium network does not exhibit sufficient link variation either within or across individuals.

In tables 2.5 and 2.6 of section 2.5, I report Monte Carlo simulations which provide evidence on the probability of $\Omega_n$ been arbitrarily close to zero when the realized network structure is either dense or empty. Table 2.7 shows that $\Omega_n$ has probability arbitrarily close to zero when the realized network is homogeneous. Finally, the numerical evidence in 2.5 and 2.6 indicate that if the network structure is sparse, the probability of $\Omega_n$ is positive. A sparse network exhibits sufficient link variation.

The large support conditions in assumptions A1 and A2 guarantee that $\Omega_n$ has a probability greater than zero. Nonetheless, this probability could be arbitrarily small, which will complicate the inference procedure. In section 2.4, I discuss inference on $\beta_0$ under two scenarios: (i) the probability of $\Omega_n$ converges to zero as the network size grows, and (ii) the probability of $\Omega_n$ converges to a positive constant as the network size grows. Specifically, I show that the convergence rate of the estimator for $\beta_0$ can be slower than the square root of the sample size if the probability of $\Omega_n$ converges to zero.

*Nonlinear Panel Data Identification Strategy: Maximum Score*

In this section, I show that the incidental parameter problem in the network formation model in (2.1) is more complex than in a nonlinear panel data model with both cross-sectional and time fixed effects. Specifically, in the network model, the fixed effects across the longitudinal dimension may be arbitrarily correlated with the vector of observed attributes. Consequently, these fixed effects do not satisfy a strong exogeneity condition as the time fixed effects do 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 $\beta_0$ in this model.

The following proposition adapts Lemma 2 in Manski (1987) to the network formation model specified by equation (2.1). Furthermore, it 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 $\beta_0$.

**Proposition 4.**

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

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

2. *Let Assumptions A1 and 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 4 formalizes the conjecture made by Charbonneau (2014) regarding the impossibility to generalize Maximum Score to the presence of multiple fixed effects. Proposition 4 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* $\beta_0$*. Nonetheless, this does not mean that* $\beta_0$ *is unidentified. Specifically, Theorem 1 proves that* $\beta_0$ *is point identified using a different (known) specific feature of the distribution of observables after conditioning on* $\Omega_n$*.*

### *2.3.4 Alternative Identifying Assumptions*

The point identification result in section 2.3.2 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 (2.1) when Assumption A2.2 is violated. Specifically, I consider two scenarios: (1) 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 $\Delta_{kl}x_i^{(-1)}$; (2) 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.

*Bounded Support and one Continuous Covariate*

The main result of this section shows that $\beta_0$ 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 $\beta_0$.

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) *$A_\eta$ 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* $S_\eta$*.*

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 $S_\eta$, 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 5.** *Let assumptions A1, A2′ and A3 hold; then* $\beta_0$ *is point identified.*

To understand the intuition behind proposition 5, consider the identified set $\mathcal{B}_0$ defined in equation (2.13). To simplify the exposition, I restate the identified set below.

$$\mathcal{B}_0 = \Big\{ b \in \mathbb{R}^k : \mathbb{E}\left[ Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}, \Omega(ijlk) \right] \gtreqless 0 \Leftrightarrow [\Delta_{kl}x_i - \Delta_{kl}x_j]'\, b \gtreqless 0\,,$$

$$\forall\, i, j, k, l \in \mathcal{N}_n \text{ and}$$

$$\forall \mathbf{x} = (x_{12}, \cdots, x_{ik}, \cdots, x_{il}, \cdots, x_{jk}, \cdots, x_{jl}, \cdots, x_{n-1,n}) \in \text{supp}(\mathbf{X}^n) \cap \mathcal{X}_B \Big\}\,,$$

Assumption A2′ implies that the linear index $\Delta_{kl}X_i'\beta_0$ has a positive density on the interval $S_\eta$ 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:

$$S_1(b) \equiv \{ w : w'\beta_0 < 0 \ \leqslant w'b \}$$

$$S_2(b) \equiv \{ w : w'b \ \ < 0 \ \leqslant w'\beta_0 \}\,.$$

A necessary and sufficient condition for identification is

$$\mathbb{P}\left[ S_1(b) \cup S_2(b) \right] > 0. \tag{2.14}$$

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

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

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 $w = (w^{(1)}, w^{(-1)}) \in \operatorname{supp}(W_{kl,ij})$, 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:

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

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

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 (2.14) is satisfied. Therefore, $\beta_0$ is point identified.

*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 $\beta_0$, 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:

$$\mathcal{B}_0 = \Big\{ b \in \mathbb{R}^k : \mathbb{E}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}^d, \Omega(ijlk)\right] \gtreqless 0 \Leftrightarrow \left[\Delta_{kl} x_i^{(d)} - \Delta_{kl} x_j^{(d)}\right]' b \gtreqless 0,$$
$$\text{for any distinct } i, j, k, l \in \mathcal{N}_n, \text{ and } \mathbf{x}^d \in \text{supp}(\mathbf{X}^n) \cap \mathcal{X}_B \Big\}, \quad (2.15)$$

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 $\beta_0$. The system of inequalities in (2.15) 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 boldface letter $\mathbf{i}$.

Let

$$z_{\mathbf{i},d} \equiv \text{sign}\left\{\mathbb{E}\left[Y_{\mathbf{i}_{3,4}}^{(i_1)} - Y_{\mathbf{i}_{3,4}}^{(i_2)} | \mathbf{X}^n = \mathbf{x}^d, \Omega(\mathbf{i})\right]\right\}\left[\left(x_{\mathbf{i}_{1,3}}^{(d)} - x_{\mathbf{i}_{1,4}}^{(d)}\right) - \left(x_{\mathbf{i}_{2,3}}^{(d)} - x_{\mathbf{i}_{2,4}}^{(d)}\right)\right],$$

for any $\mathbf{i} = (i_1, i_2, i_3, i_4) \in P_{n,4}$ and $\mathbf{x}^d \in \text{supp}(\mathbf{X}^n)$, denote the $K$-dimensional vector of pairwise differences of the observed attributes that preserves the sign of the corresponding inequality in (2.15). Let $z_{\mathbf{i},d}^{(k)}$ denote the $k$th 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 (2.15) as the following system of linear inequalities with $K-1$ unknowns given by $b_2, \cdots, b_K$

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

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 (2.15) contains both.

The solutions to the system $(S_1)$ establish bounds for each component of the parameter $\beta_0$ 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 2.9.*

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

$$D_{ik} = \mathbf{1}\left[X_{ik}^{(1)} + X_{ik}^{(2)}\beta_{0,2} + X_{ik}^{(3)}\beta_{0,3} + \mu_i + \mu_k - \varepsilon_{ik} \geqslant 0\right] \quad \text{for any } (i,k) \in \mathcal{N}_n^{(2)}, \tag{2.16}$$

*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* $\text{supp}\,(X_{ik}^{(1)}) = \{-2,-1,0,1,2,3,4\}$, $\text{supp}\,(X_{ik}^{(2)}) = \{-1,0,1\}$, *and* $\text{supp}\,(X_{ik}^{(3)}) = \{0,1,2\}$. *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 $\beta_0$ under the moment inequalities in* (2.15).

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

*If*

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

*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 $\mathbf{x}^d \in \text{supp}(\mathbf{X}^n)$.*

[5] This example uses the same data generating process (DGP) design for the network formation model as the Monte Carlo simulations in section 2.5 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*

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

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

$$\vdots$$

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

*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:*

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

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

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

*Solving for* $b_2$*, the* $ij$*th linear inequality is equivalent to:*

$$\text{if} \quad z_{i,j}^{(2)} \geqslant 0 \quad \Rightarrow \quad -\frac{z_{i,j}^{(1)}}{z_{i,j}^{(2)}} - \frac{z_{i,j}^{(3)}}{z_{i,j}^{(2)}} b_3 \leqslant b_2,$$

$$\text{if} \quad z_{i,j}^{(2)} \leqslant 0 \quad \Rightarrow \quad -\frac{z_{i,j}^{(1)}}{z_{i,j}^{(2)}} - \frac{z_{i,j}^{(3)}}{z_{i}^{(2)}} b_3 \geqslant b_2.$$

*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 $z_{\cdot}^{(2)} \geqslant 0$, $N_2$ inequalities with $z_{\cdot}^{(2)} \leqslant 0$, and $N_3$ inequalities with $z_{\cdot}^{(2)} = 0$; then the system $(S_1)$ is equivalent to*

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

*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:*

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

*which can be written as*

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

*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:*

$$\underline{b}_3 = \max_{l=1,\cdots,L} \left\{ -\frac{u_l}{v_l} : v_l > 0 \right\},$$

$$\bar{b}_3 = \min_{l=1,\cdots,L} \left\{ -\frac{u_l}{v_l} : v_l < 0 \right\}.$$

*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 2.3, I depict the bounds for the components in $\beta_0$, the identified set, and the smallest multidimensional rectangle superset that covers $\mathcal{B}_0$.*

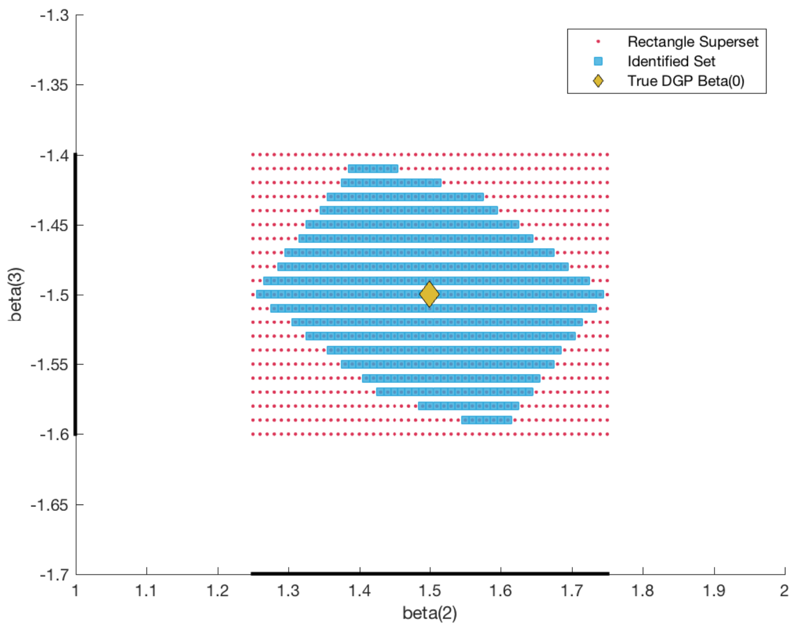



FIGURE 2.3: Bounds and Rectangular Superset

The bounds obtained by the recursive procedure introduced by Komarova (2013) are unique and independent from the order used to simplify the system (2.15). That is, the bounds obtained are uniform over the order of elimination process. Notably, the identified set is characterized 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.

## 2.4 Estimation

In this section, I propose a semiparametric pairwise difference estimator for $\beta_0$ under the point identification assumptions of section 2.3. 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 (2.1) 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 $\beta_0$ is consistent and has an asymptotic normal distribution. If the probability of the class $\Omega_n$ 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 $\Omega_n$ 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 $\beta_0$ in a set with arbitrarily small probability, also referred to as a thin set. In this case, $\beta_0$ is said to be irregularly

[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.

identified (Newey 1990; Andrews and Schafgans 1998 and Khan and Tamer 2010).

### *2.4.1 Pairwise difference Estimator*

I propose an estimator for $\beta_0$ based on the identification condition described in (7). Consider following limiting objective function

$$Q(b) \equiv 2\mathbb{E}\left[S(\mathcal{X}_B) \times \ \text{sign}\left\{[(X_{ik} - X_{il}) - (X_{jk} - X_{jl})]' b\right\} \times \left(Y_{kl}^{(i)} - Y_{kl}^{(j)}\right) \mid \Omega(ijlk)\right], \tag{2.17}$$

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

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

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 $(i_1, i_2) \in \mathcal{N}_n^{(2)}$. The sample analog of the limiting objective function is a 4th order U-statistic defined as

$$Q_n(b) \equiv \binom{n}{4}^{-1} \sum_{C_{n,4}} h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b), \tag{2.18}$$

where $\sum_{C_{n,4}}$ denotes summation over the $\binom{n}{4}$ combinations of tetrads with distinct elements $(i_1, i_2, i_3, i_4)$ from $\{1, 2, \cdots, n\}$. The function $h$, known as the kernel of the U-statistic, is defined as

$$h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) \equiv$$

$$\frac{2}{4!} \sum_{P_4} \Big\{ S(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) \times \text{sign}\left\{\left[\left(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}}\right) - \left(x_{\mathbf{i}_{2,3}} - x_{\mathbf{i}_{2,4}}\right)\right]' b\right\}$$

$$\times \left(y_{\mathbf{i}_{3,4}}^{(i_1)} - y_{\mathbf{i}_{3,4}}^{(i_2)}\right) \times \mathbf{1}\left\{\left|\left(y_{\mathbf{i}_{3,4}}^{(i_1)} - y_{\mathbf{i}_{3,4}}^{(i_2)}\right)\right| = 2\right\}\Big\},$$

where

$$S(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) =$$
$$\mathbf{1}\left[\left(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}}\right) b > B_U - B_L\right] \mathbf{1}\left[\left(x_{\mathbf{i}_{2,3}} - x_{\mathbf{i}_{2,4}}\right) b < B_L - B_U\right] +$$
$$\mathbf{1}\left[\left(x_{\mathbf{i}_{3,1}} - x_{\mathbf{i}_{3,2}}\right) b > B_U - B_L\right] \mathbf{1}\left[\left(x_{\mathbf{i}_{4,1}} - x_{\mathbf{i}_{4,2}}\right) b < B_L - B_U\right],$$

and $\sum_{P_4}$ denotes summation over the 4! permutations $\{i_1, i_2, i_3, i_4\}$ 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

$$\hat{\beta}_n = \arg\max_{b \in \tilde{\mathcal{B}} \subset \mathbb{R}^K} Q_n(b), \tag{2.19}$$

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 $\beta_0$ instead of the scaled parameter $\beta_0/\,||\beta_0||$. This normalization is discussed in section 2.3.2.

**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* $(i_1, i_2, i_3, i_4)$ *and* $(i_3, i_4, i_1, i_2)$ *have identical contribution to the kernel since*

$$\left(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}}\right) - \left(x_{\mathbf{i}_{2,3}} - x_{\mathbf{i}_{2,4}}\right) = \left(x_{\mathbf{i}_{3,1}} - x_{\mathbf{i}_{3,2}}\right) - \left(x_{\mathbf{i}_{4,1}} - x_{\mathbf{i}_{4,2}}\right),$$
$$y^{(i_1)}_{\mathbf{i}_{3,4}} - y^{(i_2)}_{\mathbf{i}_{3,4}} = y^{(i_3)}_{\mathbf{i}_{1,2}} - y^{(i_4)}_{\mathbf{i}_{1,2}}.$$

*Hence,*

$$\operatorname{sign}\left\{\left[\left(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}}\right) - \left(x_{\mathbf{i}_{2,3}} - x_{\mathbf{i}_{2,4}}\right)\right]' b\right\} \times \left(y^{(i_1)}_{\mathbf{i}_{3,4}} - y^{(i_2)}_{\mathbf{i}_{3,4}}\right)$$
$$= \operatorname{sign}\left\{\left[\left(x_{\mathbf{i}_{3,1}} - x_{\mathbf{i}_{3,2}}\right) - \left(x_{\mathbf{i}_{4,1}} - x_{\mathbf{i}_{4,2}}\right)\right]' b\right\} \times \left(y^{(i_3)}_{\mathbf{i}_{1,2}} - y^{(i_4)}_{\mathbf{i}_{1,2}}\right)$$

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

$$h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) \equiv$$

$$\frac{2}{6}\sum_{s=1}^{6}\left\{S(z_{\mathbf{i}_{s_1,s_3}}, z_{\mathbf{i}_{s_1,s_4}}, z_{\mathbf{i}_{s_2,s_3}}, z_{\mathbf{i}_{s_2,s_4}}, b)\,\mathrm{sign}\left\{\left[\left(x_{\mathbf{i}_{s_1,s_3}} - x_{\mathbf{i}_{s_1,s_4}}\right) - \left(x_{\mathbf{i}_{s_2,s_3}} - x_{\mathbf{i}_{s_2,s_4}}\right)\right]' b\right\}\right.$$
$$\left. \times \left(y^{(i_{s_1})}_{\mathbf{i}_{s_3,s_4}} - y^{(i_{s_2})}_{\mathbf{i}_{s_3,s_4}}\right) \times \mathbf{1}\left\{\left|\left(y^{(i_{s_1})}_{\mathbf{i}_{s_3,s_4}} - y^{(i_{s_@})}_{\mathbf{i}_{s_3,s_4}}\right)\right| = 2\right\}\right\},$$

*where* $(i_{s_1}, i_{s_2}, i_{s_3}, i_{s_4})$ *denotes the permutation of the index* $(i_1, i_2, i_3, i_4)$. *The 6 unique permutation are*

$$\{(i_1, i_2, i_3, i_4), (i_1, i_2, i_4, i_3), (i_1, i_3, i_2, i_4), (i_1, i_3, i_4, i_2), (i_1, i_4, i_2, i_3), (i_1, i_4, i_3, i_2)\}.$$

### *2.4.2 Consistency*

In this section, I provide sufficient conditions for the pairwise difference estimator, defined in equation (2.19), 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 $\Omega_n$ 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*

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

**Assumption B2.** *The parameter space* $\tilde{\mathcal{B}}$ *is compact and* $\beta_0$ *is an interior point of* $\tilde{\mathcal{B}}$.

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

1. $p_n \to p_0 \geqslant 0$, *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 (2015); 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 semiparametric methods. Compactness of the parameter space is used to prove consistency of the pairwise difference estimator. Assumption B2 also requires $\beta_0$ 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 $\Omega_n$ converges to a nonnegative constant, which could be zero. Under this assumption, the subnetwork structures that satisfy the conditions in $\Omega(i, j, k, l)$, see figure 2.2, become more unlike as the number of individuals in the network increases. The probability of $\Omega_n$ 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 $\Omega_n$ 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 2.8, I depict all the subnetworks that do not meet the conditions in $\Omega(ijkl)$.

Assumption B3.2 states that the probability of the class $\Omega_n$ 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 $\beta_0$ is $p_n\sqrt{N}$. A similar property to B3.2 is used in Graham (2015).

**Theorem 7.** *Let assumptions A1–A3, B1–B3 hold. Then,*

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

*as* $n \to \infty$.

Theorem 7 shows that the pairwise difference estimator converges to the true parameter value almost surely as the size of the network increases.

*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 $h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, \cdot)$ is a discontinuous function for each $(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}) \in S^4$, the function $\tau(z, \cdot)$ can be many times differentiable if the distribution of

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

is sufficiently smooth.

Let $||\cdot||$ denote the Frobenius matrix norm, $||(a_{ij})|| = (\sum_{i,j} a_{i,j}^2)^{1/2}$, $\nabla_m$ denote the $m$th 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 $\beta_0$.*

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 $\mathscr{M}$*

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

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 8.** *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:*

$$p_n \sqrt{N} (\hat{\beta}_n - \beta_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, V^{-1} \Delta V^{-1}), \quad \text{as } n \to \infty \tag{2.20}$$

*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 $\Omega_n$ 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 $\Omega_n$ converges to zero, then the converge rate is slower than the parametric rate. This result is a consequence of identifying $\beta_0$ in a thin set. The next theorem shows that the information bound is zero for the network formation model if the probability of the $\Omega_0$ converges to zero.

**Theorem 9.** *In the network formation model characterized by equation* (2.1)*, under assumptions A1–A3, B1–B4, and if $p_n \to 0$ as $n \to \infty$, then the information bound for $\beta_0$ 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 $\Delta$, and

$$\hat{p}_n \equiv \binom{n}{4}^{-1} \sum_{C_{n,4}} \mathbf{1}\left[h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, \hat{\beta}_n) \neq 0\right], \tag{2.21}$$

which is a consistent estimator for $p_0$.

**Theorem 10.** *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:*

$$\hat{\Sigma}_n^{-1/2}\sqrt{N}(\hat{\beta}_n - \beta_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I), \quad \text{as } n \to \infty \tag{2.22}$$

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 $\Delta$ 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.

### 2.4.3 Asymptotic Properties under Partial Identification

In this section, I show that if $\beta_0$ is partially identified, then the identified set can be consistently approximated from the system of linear inequalities in (2.15). Specifically, by replacing the conditional expectations in (2.15) 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 $\mathbf{x}^d \in \operatorname{supp}\mathbf{X}^n$.

**Theorem 11.** *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 $\mathbf{x}^d \in \operatorname{supp}\mathbf{X}^n$, 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:*

$$\hat{\mathbb{E}}_n\left[Y_{kl}^{(i)} - Y_{kl}^{(j)}|\mathbf{X}^n = \mathbf{x}^d, \Omega(ijlk)\right] \gtreqless -\epsilon_n \Leftrightarrow \left[\left(x_{ik}^{(d)} - x_{il}^{(d)}\right) - \left(x_{jk}^{(d)} - x_{jl}^{(d)}\right)\right]' b \gtreqless 0. \tag{2.23}$$

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

1. $H(\hat{\mathcal{B}}, \mathcal{B}_0) \xrightarrow{p} 0$ *as* $n \to \infty$.

2. $H(R(\hat{\mathcal{B}}), R(\mathcal{B}_0)) \xrightarrow{p} 0$, *as* $n \to \infty$ *as* $n \to \infty$.

*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 $\beta_0$, respectively.

## 2.5 Monte Carlo Simulations

### *2.5.1 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).

$$\tilde{Q}_n(b) \equiv \frac{1}{n(n-1)(n-2)} \sum_{P_3} S(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, b) \text{Rank}_{(\mathbf{i}_{1,3},\mathbf{i}_{1,4})} \left[ (x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}})' b \right] y_{\mathbf{i}_{1,3}}^{(i_1)},$$

where $\sum_{P_3}$ denotes summation over the $n(n-1)(n-2)$ permutations of triads with distinct elements $(i_1, i_3, i_2)$ from $\{1, 2, \cdots, n\}$

The function $\text{Rank}_{(\mathbf{i}_{1,3},\mathbf{i}_{1,4})} \left[ (x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}})' b \right]$ denotes the rank of agent $i_1$'s within-individual variation of the linear index $\left[ (x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}})' b \right]$ within dyads $(i_1, i_3)$ and $(i_1, i_4)$, 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

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

### *2.5.2 Finite Sample Performance*

In this section, I study the finite sample properties of the pairwise difference semiparametric estimator introduced in section 2.4. 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 (2.1) 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:

$$z_{i1} \sim \mathcal{N}(0, 3),$$

$$z_{i2} \sim \text{Unif}\{-1, 1\} \text{ with } P(z_{i2} = -1) = P(z_{i2} = 0) = P(z_{i2} = 1) = 1/3,$$

$$z_{i3} \sim \text{Unif}(-2, 2).$$

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 $\lambda$ measures the degree of correlation between the agent-specific observed covariates and unobserved fixed effects. The bounded fixed effects are specified as:

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

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 $\beta_0$ 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 $\lambda$ equal to 0.25 and 0.75 in appendix 2.11. The results are qualitatively similar.

Tables 2.1 and 2.2 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 (2.19). I focus on the median, mean, bias in percentage points, and the root mean square error (RMSE) for each estimator. Tables 2.1 and 2.2 also report the probability of the class $\Omega_n$ and the average degree of the network.

Table 2.1 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.

[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 2.2 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 2.1 and 2.2 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.

Table 2.1: 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 2.2: 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 $\lambda = 0.5$

## 2.6 Empirical Application

In this section, I use the methods developed in this paper to estimate the network formation model in equation (2.1) 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 et al. (2009)) to construct an undirected network of friends. I then the preference parameters using the pairwise difference estimator introduced in section 2.4.

### *2.6.1 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 et al. (2009); Calvó-Armengol et al. (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,

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

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 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 2.3 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.

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

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 2.3.4.

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 $\Omega_n$ in the sample is 2.24%. The probability of $\Omega_n$ in the remaining large high school is 2.59%. Although the probability of $\Omega_n$ is larger in the other high school, that high school was not selected because it suffers from a more severe missing observation problem.

### *2.6.2 Empirical Results*

I estimate the network formation model in equation (2.1) using the pairwise difference estimator. I compare these results with the point estimates obtained from computing Graham (2015) Tetrad Logit estimator and a naive logistic regression. The naive logistic regression ignores the presence of the fixed effects $\mu_i$ and $\mu_j$ in equation (2.1), and assumes $\varepsilon_{ij}$ 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 2.3.

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 $\Omega_n$ 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 2.3: 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 2.4: 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.

## 2.7 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 (2.1), 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

$$D_{ij}^{n} = \mathbf{1}\left[X_{ij}^{n\prime}\beta_0 + \mu_i + g(\mu_i, \mu_j) - \varepsilon_{ij}^{n} \geqslant 0\right], \tag{2.24}$$

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

Under the specification in (2.24), individuals create connections based on both homophily on observed and unobserved characteristics. Furthermore, the agent-specific fixed effect $\mu_j$ 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.

*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.

## 2.8 Proofs

### *2.8.1 Point Identification*

The following two lemmas formalize the intuition behind the identification strategy and they are used to prove Theorem 1.

Lemma 12 formalizes the intuition of Diagram 2 in Figure 2.1. 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 13 specifies a median condition for the random variable $Y^{(i)}_{kl} - Y^{(j)}_{kl}\Big| \mathbf{X}^n, \Omega(ijlk)$.

**Lemma 12.** *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* $\mathbf{x} \in \mathcal{X}_B$*,*

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

*and for* $B$ *defined in A2.*

***Preliminaries:***

Let

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

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

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,

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

if and only if

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

$$\geqslant$$

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

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

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

then:

$$\begin{aligned} \mid \Delta_{kl} x_i \beta_0 \mid \ &\geqslant \ 2B \\ \mid \Delta_{kl} x_j \beta_0 \mid \ &\geqslant \ 2B, \end{aligned} \tag{2.25}$$

and

$$\text{sign}\{\Delta_{kl} x_i \beta_0\} + \text{sign}\{\Delta_{kl} x_i \beta_0\} = 0. \tag{2.26}$$

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 (2.25) imply:

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

Then by A3, it follows that:

$$\Delta_{kl} x_i \beta_0 \ \geqslant \ \ 2B \ \geqslant (\mu_l - \mu_k), \tag{2.27}$$

$$\Delta_{kl} x_j \beta_0 \ \leqslant \ -2B \ \leqslant (\mu_l - \mu_k). \tag{2.28}$$

Equation (2.27) can be equivalently written as:

$$x_{ik} \beta_0 + \mu_i + \mu_k \geqslant x_{il} \beta_0 + \mu_i + \mu_l$$

$$w_{ik}(\beta_0) \geqslant w_{il}(\beta_0)$$

Equation (2.28) can be equivalently written as:

$$x_{jk}\beta_0 + \mu_j + \mu_k \leqslant x_{jl}\beta_0 + \mu_j + \mu_l$$

$$w_{jk}(\beta_0) \leqslant w_{jl}(\beta_0)$$

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 (2.25) imply:

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

Then by A3, it follows that:

$$\Delta_{kl}x_i\beta_0 \leqslant -2B \leqslant (\mu_l - \mu_k), \tag{2.29}$$

$$\Delta_{kl}x_j\beta_0 \geqslant 2B \geqslant (\mu_l - \mu_k). \tag{2.30}$$

Equations (2.29) and (2.30) can be equivalently written as:

$$w_{ik}(\beta_0) \leqslant w_{il}(\beta_0)$$

$$w_{jk}(\beta_0) \geqslant w_{jl}(\beta_0).$$

*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:

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

that is:

$$\Delta_{kl}x_i - \Delta_{kl}x_j > 0$$

$$\Leftrightarrow$$

$$\Delta_{kl}x_i + (\mu_k - \mu_l) - \Delta_{kl}x_j - (\mu_k - \mu_l) > 0$$

$$\Leftrightarrow$$

$$\Delta_{kl}x_i + (\mu_k - \mu_l) > 0 \text{ and } \Delta_{kl}x_j + (\mu_k - \mu_l) < 0$$

$$\Leftrightarrow$$

$$w_{ik}(\beta_0) > w_{il}(\beta_0) \text{ and } w_{jk}(\beta_0) < w_{jl}(\beta_0)$$

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

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

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:

$$w_{ik}(\beta_0) \geqslant w_{il}(\beta_0)$$

$$w_{jk}(\beta_0) \leqslant w_{jl}(\beta_0).$$

Alternatively, if

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

is true, then:

$$w_{ik}(\beta_0) \leqslant w_{il}(\beta_0)$$

$$w_{jk}(\beta_0) \geqslant w_{jl}(\beta_0).$$

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 13.** *For any $n$, and any $i, j, k, l \in \mathcal{N}_n$,*

$$\text{Med}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = x, \Omega(ijlk)\right] = 2 \times \text{sign}\left\{\mathbb{P}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} = 2 | \mathbf{X}^n = x, \Omega(ijlk)\right] - \mathbb{P}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} = -2 | \mathbf{X}^n = x, \Omega(ijlk)\right]\right\}. \tag{2.31}$$

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

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

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

with $p + q = 1$. Then,

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

the results follows from this observation. ■

*Proof of Theorem 1*

**Part I**

Note,

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

Therefore:

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

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 $\Delta^2 X \equiv [(X_{ik} - X_{il}) - (X_{jk} - X_{jl})]$ 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 $\text{sign}\,(xb) \neq (x\beta_0)$. 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 $\beta_0$ is identified up to scale. ■

### *2.8.2 Identification Failure: Thin Set*

*Proof of Theorem 2*

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

$$\left\{D_{ik} \neq D_{il}, D_{jk} \neq D_{jl}, D_{ik} \neq D_{jk} : (i,j,k,l) \in P_{n,4}\right\}, \tag{2.32}$$

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 (2.32) 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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 (2.32). 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 13 no longer holds. Specifically,

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

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

Since Lemma 13 fails, equation (7) is misspecified and does not have identification power. Therefore, the conclusion in Theorem 1 is wrong.

Now, I show that the Med $\left[\Delta Y_{ij} | \mathbf{X}^n = x, \Omega(ijlk)^c\right]$ does not have power to identify $\beta_0$. 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 $\beta_0$,

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

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

$$\mu_i + \mu_k = \sup X_{ik}^{(1)} \Rightarrow D_{ik} = 1$$

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

$$\mu_i + \mu_l = \inf X_{il}'\theta, \text{ for any } \theta \in R^K$$

$$\mu_j + \mu_l = \inf X_{jl}'\theta, \text{ for any } \theta \in R^K$$

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

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

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

Thus $p_3(\tilde{\beta}, x, \mu) \geqslant 0.5$ 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

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

Thus $p_3(\tilde{\beta}, x, \mu) \geqslant 0.5$ 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

$$\mu_i + \mu_k = \sup X'_{ik}\theta, \text{ for any } \theta \in R^K,$$

$$\mu_j + \mu_k = \inf X^{(1)}_{jk} \Rightarrow D_{jk} = 0,$$

$$\mu_i + \mu_l = \inf X^{(1)}_{il} \Rightarrow D_{ik} = 0,$$

$$\mu_j + \mu_l = \sup X'_{jl}\theta, \text{ for any } \theta \in R^K.$$

Therefore,

$$p_1(\tilde{\beta}, x, \mu) = 0,$$

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

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

and

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

Thus $p_3(\tilde{\beta}, x, \mu) \geqslant 0.5$ 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

$$\mu_i + \mu_k = \sup X'_{ik}\theta, \text{ for any } \theta \in R^K,$$

$$\mu_j + \mu_k = \sup X'_{jk}\theta, \text{ for any } \theta \in R^K,$$

$$\mu_i + \mu_l = \inf X^{(1)}_{il} \Rightarrow D_{il} = 0,$$

$$\mu_j + \mu_l = \sup X^{(1)}_{jl} \Rightarrow D_{jl} = 1.$$

Therefore,

$$p_1(\tilde{\beta}, x, \mu) = 0,$$

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

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

and

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

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

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

$$\mu_i + \mu_k = \sup X_{ik}^{(1)} \Rightarrow D_{ik} = 1,$$

$$\mu_i + \mu_l = \inf X_{il}^{(1)} \Rightarrow D_{il} = 0.$$

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:

$$\text{If } s = k: \quad \mu_j + \mu_k = \inf X_{jk}^{(1)} \Rightarrow D_{jk} = 0,$$

$$\text{If } s = l: \quad \mu_j + \mu_l = \sup X_{jl}^{(1)} \Rightarrow D_{jk} = 1.$$

Therefore,

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

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

**Case 6:** This case is analogous to case 5.

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

$$\mu_i + \mu_k = \sup X_{ik}^{(1)} \Rightarrow D_{ik} = 1,$$

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

$$\mu_i + \mu_l = \inf X_{il}^{(1)} \Rightarrow D_{il} = 0,$$

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

II. Suppose under the true model, characterized by $\beta_0$,

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

**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,

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

and

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

Therefore:

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

Which implies, that

$$p_1(\tilde{\beta}, x, \mu) \leqslant 0.5$$
$$p_1(\tilde{\beta}, x, \mu) + p_2(\tilde{\beta}, x, \mu) \geqslant 0.5$$

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

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

Therefore,

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

Note,

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

and

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

Therefore:

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

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,

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

and

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

$$\leqslant \quad \mathbb{P}\left[D_{ik} = 1, D_{il} = 1, D_{jk} = 0, D_{jl} = 0 | \mathbf{X}^n = x, \Omega(ijlk)^c\right]$$

$$\leqslant \quad p_2(\tilde{\beta}, x, \mu)$$

Therefore:

$$p_1(\tilde{\beta}, x, \mu) \leqslant \quad p_2(\tilde{\beta}, x, \mu)$$

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

Which implies, that

$$p_1(\tilde{\beta}, x, \mu) \leqslant 0.5$$

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

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

$$\mu_i + \mu_k = \sup X'_{ik}\theta, \text{ for any } \theta \in R^K,$$

$$\mu_j + \mu_k = \sup X'_{jk}\theta, \text{ for any } \theta \in R^K,$$

$$\mu_i + \mu_l = \sup X^{(1)}_{il} \Rightarrow D_{il} = 1,$$

$$\mu_j + \mu_l = \sup X^{(1)}_{jl} \Rightarrow D_{jl} = 1.$$

Therefore,

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

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

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

Note,

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

and

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

Therefore:

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

Which implies, that

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

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,

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

and

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

Therefore:

$$p_1(\tilde{\beta}, x, \mu) \leqslant p_2(\tilde{\beta}, x, \mu)$$

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

Which implies, that

$$p_1(\tilde{\beta}, x, \mu) \leqslant 0.5$$

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

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:

$$\mu_i + \mu_k = \sup X_{ik}^{(1)} \Rightarrow D_{ik} = 1,$$

$$\mu_j + \mu_k = \sup X_{jk}^{(1)} \Rightarrow D_{jk} = 1,$$

$$\mu_i + \mu_l = \sup X_{il}^{(1)} \Rightarrow D_{il} = 1,$$

$$\mu_j + \mu_l = \sup X_{jl}^{(1)} \Rightarrow D_{jl} = 1.$$

III. Suppose under the true model, characterized by $\beta_0$

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

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 $\beta_0$ in terms of $\text{Med}\left[\Delta Y_{ij}|\mathbf{X}^n = x, \Omega(ijlk)^c\right]$ 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*

*Proof.* Note that,

$$0 = \mathbb{P}\left[\Omega_n\right] \Leftrightarrow \mathbb{P}\left[\{D_{ik} \neq D_{il}, D_{jl} \neq D_{jk}, D_{ik} \neq D_{jk}\} : i, j, k, l \in \mathcal{N}_n\right] = 0. \tag{2.33}$$

Hence, it is sufficient to show that one of conditions in (2.33) 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 $\Omega_n$ 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

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

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,

$$\begin{aligned}\mathbb{P}\left[D_{ij} = 1 \mid \tilde{\mu} = \mu\right] =& \int \int \mathbf{1}\left\{x'\beta_0 + \mu_i + \mu_j - e \geqslant 0\right\} \lambda(dx) dG_{\varepsilon|\mu}(e) \\ =& 1 - \int F_{X'_{ij}\beta_0|\mu,e}(e - \mu_i - \mu_j) dG_{\varepsilon|\mu}(e). \end{aligned} \tag{2.34}$$

**Part 1.**

Rewrite equation (2.34) 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,

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

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 3.2,

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

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

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

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,

$$\begin{aligned}\mathbb{P}\left[D_{ij} = 1 \mid X_{ij}^{'} = x\right] &= \int\int \mathbf{1}\left\{x'\beta_0 + \mu_i + \mu_j - e \geqslant 0\right\} \nu(d\mu) dG_{\varepsilon|x}(e) \\ &= 1 - \int F_{\tilde{\mu}_i+\tilde{\mu}_j|x,e}(e - x'\beta_0)\, dG_{\varepsilon|x}(e). \qquad (2.35)\end{aligned}$$

Rewrite equation (2.35) 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,

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

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,

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

■

### 2.8.3 Identification Failure: Maximum Score

The following lemma is used to prove proposition 4. The next lemma adapts Lemma 1 in Manski (1987) to the multiple fixed effects case.

**Lemma 14.** *Let assumption A1 hold. For any $n$, and any $i, l, k \in \mathcal{N}_n$.*

$$\operatorname{sign}\left[(x_{ik} - x_{il})'\beta_0 + (\mu_k - \mu_l)\right] = \operatorname{sign}\left(\mathbb{E}\left[D_{ik} - D_{il} \mid \mathbf{X}^n = \mathbf{x}\right]\right) \tag{2.36}$$

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

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

Note,

$$x'_{il}\beta_0 + \mu_i + \mu_l \leqslant x'_{ik}\beta_0 + \mu_i + \mu_k \quad \Leftrightarrow \quad x'_{il}\beta_0 + \mu_l \leqslant x'_{ik}\beta_0 + \mu_k$$

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

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

Equivalently for

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

In summary,

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

thus,

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

∎

*Proof of Proposition 4*

*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,

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

By Lemma 14, the result follows.

**Part 2.**

Denote by $P_{\beta_0}$ 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 4, we will assume that $\tilde{\mu}$ is a known function of the exogenous attributed. Note that if $\beta_0$ 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:

$$\mu_i - \mu_k = X_{ik}' (\beta_0 - K)$$

$$\tilde{\beta} = \beta_0 + cK$$

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

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

and

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

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 $\beta_0$ we can find a $\tilde{\beta} \in \mathcal{B}$ with $\tilde{\beta} \neq \beta_0$ such that they are observationally equivalent. ■

*Proof of Proposition 5*

The proof of this proposition is the main text.

### *2.8.4 Inference*

*Proof of Proposition 6*

*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 $\beta_0$ uniquely maximizes $Q(b)$.

■

*Proof of Theorem 7*

*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 $\beta_0$.

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

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

$$h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) \equiv$$

$$\frac{1}{4!}\sum_{P_4}\left\{2 \times \operatorname{sign}\left\{\left[\left(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}}\right) - \left(x_{\mathbf{i}_{2,3}} - x_{\mathbf{i}_{2,4}}\right)\right]' b\right\} \times \left(y_{\mathbf{i}_{3,4}}^{(i_1)} - y_{\mathbf{i}_{3,4}}^{(i_2)}\right)\right.$$

$$\left. \times \mathbf{1}\left\{\left|\left(y_{\mathbf{i}_{3,4}}^{(i_1)} - y_{\mathbf{i}_{3,4}}^{(i_2)}\right)\right| = 2\right\}\right\},$$

Let

$$\tilde{Q}_n(b) \equiv\ p_n^{-1}Q_n(b),$$

$$\tilde{h}(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b) \equiv\ p_n^{-1}h(z_{\mathbf{i}_{1,3}}, z_{\mathbf{i}_{1,4}}, z_{\mathbf{i}_{2,3}}, z_{\mathbf{i}_{2,4}}, b),$$

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

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

Then

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

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

- *Condition 1:* Follows from assumption B2.

- *Condition 2:* Assumption A2 implies that $\left(x_{\mathbf{i}_{1,3}} - x_{\mathbf{i}_{1,4}}\right)' b$ is continuously distributed for each $b \in \tilde{\mathcal{B}}$, for any $(i_1, i_3, i_4)$ from $\{1, 2, \cdots, n\}$. Therefore,

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

  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 6.

- *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):

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

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

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

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 $\hat{\beta}_n \stackrel{a.s.}{\to} \beta_0$ as $n \to \infty$. ■

*Proof of Theorem 8*

*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 $\beta_0$. That is

$$\left\|\hat{\beta}-\beta_0\right\| = O_p(1/p_n\sqrt{N}). \tag{2.37}$$

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

$$\begin{aligned} \tilde{Q}_n(b)-\tilde{Q}_n(\beta_0) &= \\ \frac{1}{2}(b-\beta_0)'V(b-\beta_0)+\frac{1}{p_n\sqrt{N}}(b-\beta_0)'\mathbf{W}_n & \\ +o_p\left(\|(b-\beta_0)\|^2/p_n\right)+o_p\left(\frac{1}{p_nN}\right) & \end{aligned} \tag{2.38}$$

uniformly in $o_p(1)$ neighborhoods of $\beta_0$, 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 (2.38) 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

$$\tilde{Q}_n(b)-\tilde{Q}_n(\beta_0)=Q(b)-Q(\beta_0)+\mathbf{P}_nf_1(\cdot,b)+\sum_{i=2}^{4}U_n^if_i(\cdot,b), \tag{2.39}$$

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 $\beta_0$

$$\tau(z,b) = \tau(z,\beta_0) + (b-\beta_0)'\nabla_1\tau(z,\beta_0) + \frac{1}{2}(b-\beta_0)'\nabla_2\tau(z,b^*)(b-\beta_0), \qquad (2.40)$$

for any $z \in S$ and $b \in \mathscr{M}$ and $b^*$ between $b$ and $\beta_0$.

Assumption B4.2 implies

$$||(b-\beta_0)'\left[\nabla_2\tau(z,b) - \nabla_2\tau(z,\beta_0)\right](b-\beta_0)|| \leqslant M(z)\,||(b-\beta_0)||^3 \quad \text{as} \quad b \to \beta_0 \qquad (2.41)$$

From inequality (2.41) and the integrability of $M$, the expected value of equation (2.40) implies:

$$\mathbb{E}\left[(\tau(\cdot,b) - \tau(\cdot,\beta_0))\right] = (b-\beta_0)'\mathbb{E}\nabla_1\tau(\cdot,\beta_0) + \frac{1}{2}(b-\beta_0)'\mathbb{E}\nabla_2\tau(\cdot,b^*)(b-\beta_0)$$

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

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

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 $\beta_0$, 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,

$$\mathbf{P}_n f_1(\cdot, b) = \frac{1}{p_n \sqrt{N}} (b - \beta_0)' \mathbf{W}_n + o_p\left(||(b - \beta_0)||^2 / p_n\right) \tag{2.42}$$

uniformly over $o_p(1)$ neighborhoods of $\beta_0$, where $\mathbf{W}_n$ converges in distribution to a $\mathcal{N}(\mathbf{0}, \boldsymbol{\Delta})$ 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

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

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:

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

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 (2.38) we need to show that

$$U_n^2 f_2(\cdot, b) + U_n^3 f_3(\cdot, b) + U_n^4 f_4(\cdot, b) = o_p(1/p_n N) \tag{2.43}$$

uniformly over $o_p(1)$ neighborhoods of $\beta_0$.

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 (2.43) is proved by using Corollary 8 in Sherman (1994), if the following conditions hold:

$$\mathbf{P}^i\,||f_i(\cdot, b)|| \to 0 \quad \text{as} \quad b \to \beta_0, \tag{2.44}$$

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 $\beta_0$ 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 $\beta_0$ 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:

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

Similar steps can be used to prove (2.44) for $i = 2, 3$. Then by Corollary 8 in Sherman (1994), follows:

$$U_n^2 f_2(\cdot, b) + U_n^3 f_3(\cdot, b) + U_n^4 f_4(\cdot, b) = o_p(1/p_n N) \tag{2.45}$$

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

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

Then, equation (2.37) has been established.

*Proof of Part 2:*

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

$$t_n^* = -V^{-1}\mathbf{W}_n + \beta_0$$

$$t_n = p_n \sqrt{N}(\hat{\beta} - \beta_0) + \beta_0$$

Then, by definition of $t_n^*$

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

By applying (2.38) twice in the last expression, multiplying by $p_n^2 N$, collecting terms, and using that $V$ is negative definite.

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

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 9*

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

Suppose, $F_\epsilon$ 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 $\beta_0$ is zero. ■

*Proof of Theorem 11*

*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,

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

There are two cases three consider:

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

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

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

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

as $n \to \infty$.

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

as $n \to \infty$.

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

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

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

Finally,

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

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

**Part 2:**

For any $\delta > 0$, if

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

Thus

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

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

## 2.9 Sharp Bounds

In this section, I illustrate the recursive procedure introduced in Komarova (2013) to derive sharp bounds for each element in $\beta_0$ of the network formation model given by equation (2.1). Suppose that the goal is to find the sharp bounds for the $K$th component of $\beta_0$, i.e., $\beta_K$. The recursive procedure then starts by excluding one unknown variable at each iteration from the following system.

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

Consider excluding the unknown $\beta_2$. The $ij$th inequality in system $(A_1)$

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

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

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

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

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

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

$$L_i(b_3, \cdots, b_K) < b_2, \qquad i = 1, \cdots, N_1,$$

$$U_j(b_3, \cdots, b_K) > b_2, \qquad j = 1, \cdots, N_2,$$

$$Z_r(b_3, \cdots, b_K) > 0, \qquad r = 1, \cdots, N_3,$$

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 $\beta_K$ are derived from the simplified system $(A_{K-1})$ as follows:

$$\underline{b}_K = \max_{l=1,\cdots,L} \left\{ -\frac{u_l}{v_l} : v_l > 0 \right\},$$

$$\bar{b}_K = \min_{l=1,\cdots,L} \left\{ -\frac{u_l}{v_l} : v_l < 0 \right\}.$$

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,

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

## 2.10 Thin Set

Table 2.5: 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 2.6: 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 2.7: 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.

## 2.11 Monte Carlo Simulations

Table 2.8: 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 2.9: 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 2.10: 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 2.11: 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 $\lambda = 0.75$

# 3

# Identification of Network Formation Models with Interactive Unobserved Agent-Specific Heterogeneity

## 3.1 Introduction

In this paper, I study the formation of a directed network when agents have a latent utility that accounts for the similarity of observed characteristics, as well as the complementarity of unobserved agent-specific factors. The complementarity of the unobserved factors is modeled as multiple interactive fixed effects. The objective of this paper is to recover and differentiate the contribution of homophily on observed attributes on forming a directed link from the similarity of the unobserved heterogeneity components. I show that the semiparametric methods developed in Chapter 2 can be extended to analyze a directed network with a more sophisticated type of heterogeneity. These methods are used to identify the coefficients that measure homophily on observed attributes. This is the first paper that addresses the formation of a directed network with interactive fixed effects and following a semiparametric approach.

## 3.2 Model

I study the formation of a directed network involving $n$ agents. Index by $i$ a random sample of agents, with $i = 1, \cdots, n$. An agent decides to form a link with another individual in the network if the net benefit of forming the link is nonnegative. The net benefit depends on exogenous attributes, agent-specific unobserved components and on an idiosyncratic shock. I proceed to formally describe the model.

**Sample:** Let $\mathcal{N}_n = \{1, 2, \cdots, n\}$ be the set of potentially connected agents. The agents could be firms, countries, students, etc. Each pair of sampled agents $(i, j)$ with $i, j \in \mathcal{N}_n$ and $i \neq j$ constitute a dyad. Denote the total sample of dyads by.

$$\mathcal{N}_n^{(2)} \equiv \{(1, 2), \cdots\cdots, (n-1, n)\}.$$

In what follows, we will refer to the set of agents equivalently as nodes or individuals.

**Network:** Let $D^n \in \mathcal{D}^n$ represents the adjacency matrix of a directional network. Denote the $(i, j)th$ characteristic element of $D^n$ as $D_{ij}^n$, where $D_{ij}^n = 1$ if agent $i$ establishes a directed link with agent $j$ and $D_{ij}^n = 0$ otherwise. Note that the adjacency matrix considered in this model is asymmetric, i.e. $D_{ij}^n \neq D_{ji}^n$. For example, in a friendships network agent $i$ might list agent $j$ as a friend. However, agent $j$ might not reciprocate this listing. I rule out self-ties, i.e. $D_{ii}^n = 0$ for all $i \in \mathcal{N}_n$. To simplify the notation, we will omit the dependence of the network on the sample size, $n$.

**Exogenous attributes:** Each sampled dyad $(i, j) \in \mathcal{N}_n$ is endowed with a vector of observed exogenous attributes $X_{ij} \in \mathbb{R}^K$. The vector $X_{ij}^n$ can contain discrete or continuous attributes, e.g. race, sex and income, etc. If the exogenous attributes are observed 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 $X_{ij}^n = g(X_i^n, X_j^n) = g(X_j^n, X_i^n)$, with a function $g(\cdot,\cdot)$ that is symmetric and nonlinear. Different specifications of $g$ can be used to capture similarity ($g(X_j^n, X_i^n) = \left(X_i^n - X_j^n\right)^2$) 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.

**Agent-specific heterogeneity:** Each agent is endowed with individual-specific components that are unobserved to the researcher. The unobserved components are $\{\mu_i, \gamma_i\}_{i=1}^n$. Both unobserved factors will influence the agent's decision. However, they have a different effect on the net link benefit. The factor $\mu_i$ captures the individual node heterogeneity. Meanwhile, the interactions between the factors $\gamma_i, \gamma_j$, captures the preferences of forming a directed link based on homophily or complementarity on unobserved factors. The dependence between the agent-specific components and the exogenous attributes is left unrestricted. 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. I will explain the intuition behind the fixed effects in depth below.

**Idiosyncratic Shocks:** The unobserved dyad-specific disturbance component $\varepsilon_{ij}^n$ captures exogenous random factors that influence the decision of establishing a direct link between agents $i$ and $j$. These components are unobserved to the researcher. An example of these shocks are search costs associated with forming a link.

**Network Formation Rule:** Agent $i$ forms a directional link with agent $j$ if the total net benefit of forming that link is nonnegative. In particular, for any $(i,j) \in \mathcal{N}_n^{(2)}$, it is optimal for agent $i$ to form a directional link with agent $j$ if

$$D_{ij}^n = \mathbf{1}\left[X_{ij}^{n\prime}\beta_0 + \mu_i + \gamma_i\gamma_j - \varepsilon_{ij}^n \geqslant 0\right] \tag{3.1}$$

Intuitively, equation (2.1) says that a directed link between agents $i$ and $j$ is formed if the link net benefit is nonnegative. The factors in the net benefit can be classified into four different 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, characterized by the additive agent-specific and unobserved factor, represents the individual preferences for creating links. The third class, given by the interaction of the agent-specific components captures the assortative association based on similarity on unobserved factors. Finally, the last class, given by a link-specific disturbance term, captures the exogenous factors that influence the decision of forming a specific link. The last three factors are known to the agents but unobserved to the researcher.

The unobserved agent-specific factors in equation (3.1) allow for heterogeneous net benefits across individuals. This observation extends the model's capacity to predict network structures with different individual connections. Moreover, due to the presence of the interactive fixed effects the valuation of any agent-specific fixed effects is different across linking decisions. Intuitively, the interactive component provides information regarding homophilic preferences based on unobserved components.

If $\gamma_i = 1$, equation (3.1) degenerates to a nonlinear model with multiple additive unobserved agent-specific components, in a similar fashion as in Chapter 2, in Charbonneau (2014) and Graham (2015). However, a main difference is that equation (3.1) analyzes a directed network instead of an undirected network as the above references do.

Equation 3.1 could be extended to include endogenous network effects. That is, an individual's utility is affected by the overall structure of the network. For example, an individual might have preference for forming connections with individuals that have common friends or that have a large number of friends. Although, this is an interesting extension, I left it as further research.

Some recent papers that have studied endogenous network preferences are Menzel (2015); Leung (2015a); Mele (2015); Goldsmith-Pinkham and Imbens (2013); Christakis et al. (2010); Sheng (2012). Only Goldsmith-Pinkham and Imbens (2013) allows for unobserved node heterogeneity. However this paper follows a random effects approach. The remaining papers, except for Sheng (2012) which follows a partial identification approach and assumes the presence of multiple observations of a single network, rely up to some extend on parametric distributional assumptions.

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 $\varepsilon_{ij}$ . In the following section we formalize the econometric model and provide the identification results.

## 3.3 Identification

In this section, I state the main point identification result for the semiparametric network formation model with interactive fixed effects, specified by equation (3.1). In section 3.3.1, I describe the identification strategy. Section 3.3.2 establishes the main point identification result, which is achieved by conditioning on a set that ensures enough variation within and across individuals' links.

### *3.3.1 Identification Strategy*

The intuition behind the identification strategy is summarized in figure 3.1. Consider the subnetwork formed by the undirected links between agents $i, k, l \in \mathcal{N}_n$. The direction of the links are represented by arrows. A solid line connecting two agents denotes that a link exists and a dashed line denotes that a link is absent.

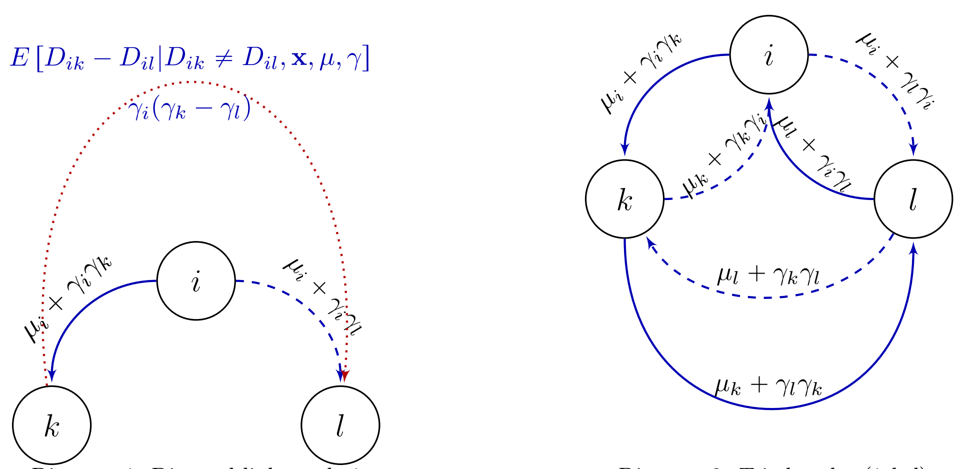



Diagram 1: Directed links node $i$ Diagram 2: Triad nodes $(i, k, l)$

FIGURE 3.1: Directed network for nodes $i, k, l \in \mathcal{N}_n$.

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

$$\mathbb{P}\left[D_{il} = 1 \mid \mathbf{X}^n = x, \tilde{\mu} = \mu, \tilde{\gamma} = \gamma\right] < \mathbb{P}\left[D_{ik} = 1 \mid \mathbf{X}^n = x, \tilde{\mu} = \mu, \tilde{\gamma} = \gamma\right]. \tag{3.2}$$

If the link-specific unobserved random variables are identically distributed, then the equation (3.2) holds if and only if:

$$x'_{il}\beta_0 + \mu_i + \gamma_i\gamma_l < x'_{ik}\beta_0 + \mu_i + \gamma_i\gamma_k,$$

where agent $i$'s individual-specific fixed effects $\mu_i, \gamma_i$ are common elements. Therefore, the within-individual difference implies:

$$0 < (x_{ik} - x_{il})'\beta_0 + \gamma_i(\gamma_k - \gamma_l).$$

The previous observation 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, $\gamma_i(\gamma_k - \gamma_l)$. The net difference has differenced out the common factor $\mu_i$.

In diagram 1, the dotted line labeled as $\{\mathbb{E}\left[D_{ik} - D_{il} | D_{ik} \neq D_{il}, \mathbf{X}^n = x, \mu, \gamma\right]\}$ depicts this intuition, it 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 $\gamma_i(\gamma_k - \gamma_l)$.

Specifically, the following equation holds for the conditional median of the net difference, $D_{ik} - D_{il}$. The proof is in the appendix 3.6

$$\text{Med}(D_{ik} - D_{il} | \mathbf{X}^n = \mathbf{x}, D_{il} \neq D_{ik}) = \text{sign}\left((x_{ik} - x_{il})'\beta_0 + \gamma_i(\gamma_k - \gamma_l)\right), \tag{3.3}$$

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

Equation (3.3) implies that the conditional median of the net difference $D_{ik} - D_{il}$ depends on the unobserved composite factor $\gamma_i(\gamma_k - \gamma_l)$ due to the presence of multiple agent-specific fixed effects in the network formation model (3.1). Therefore, equation (3.3) cannot be used to identify $\beta_0$.

The point-identification argument in the network formation model with interactive fixed effects is the following. Consider the direct linking decisions of individual $k$ as described in Diagram 2. Specifically, given $\{\mathbf{X}^n = x, \tilde{\mu} = \mu, \tilde{\gamma} = \gamma\}$, suppose that the conditional probability of establishing a link between dyad $(k, l)$ is greater than the one for dyad $(k, i)$. That is:

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

Analogously to above, the conditional expectation of the net difference between individual $k$'s linking decisions $D_{kl} - D_{ki}$ is characterized by the difference of the observed exogenous regressors $(x_{kl} - x_{ki})'\beta_0$ and the difference of the unobserved factors $\gamma_k(\gamma_l - \gamma_i)$. That is:

$$0 < (x_{kl} - x_{ki})'\beta_0 + \gamma_k(\gamma_l - \gamma_i).$$

The composite factor $\gamma_k\gamma_i$ is a common and unobserved fixed effect across the within-individual variation for agents $i$ and $k$. The previous observation suggests that, with enough across-individuals variation, the composite fixed effect $\gamma_k\gamma_i$ can be differenced out by computing the across-individuals difference of $D_{ik} - D_{il}$ and $D_{kl} - D_{ki}$.

However, this difference fails to partial out the composite effects $\gamma_i\gamma_l$ and $\gamma_k\gamma_l$. The within-individual $l$ variation can be used to controlled for these composite effects if the linking decisions exhibit the sufficient variation as described in Diagram 2. That is

$$\mathbb{P}\left[D_{lk} = 1 \mid \mathbf{X}^n = x, \tilde{\mu} = \mu, \tilde{\gamma} = \gamma\right] < \mathbb{P}\left[D_{li} = 1 \mid \mathbf{X}^n = x, \tilde{\mu} = \mu, \tilde{\gamma} = \gamma\right],$$

is characterized by the difference of the observed exogenous regressors $(x_{li} - x_{lk})'\beta_0$ and the difference of the unobserved factors $\gamma_l(\gamma_i - \gamma_k)$. In other words,

$$0 < (x_{li} - x_{lk})'\beta_0 + \gamma_l(\gamma_i - \gamma_k).$$

The previous intuition suggests that, with enough across-individuals variation, the composite fixed effect $\gamma_i(\gamma_k - \gamma_l)$ can be differenced out by computing the across-individuals difference of $D_{ik} - D_{il}$, $D_{kl} - D_{ki}$ and $D_{li} - D_{lk}$. Specifically, in the next section, I show that the following equation for the conditional median of the pairwise difference holds:

$$\begin{aligned}\text{Med}\{[D_{ik} - D_{il}] + [D_{kl} - D_{ki}] + [D_{li} - D_{lk}] \mid \mathbf{X}^n = \mathbf{x}, \Omega(ilk)_1, \Omega(ilk)_2\} = \\ 3 \times \text{sign}\left\{[(x_{ik} - x_{il}) + (x_{kl} - x_{ki}) + (x_{li} - x_{lk})]' \beta_0\right\}, \quad (3.4)\end{aligned}$$

where

$$\Omega(ilk)_1 = \{D_{ik} \neq D_{il}, D_{kl} \neq D_{ki}, D_{li} \neq D_{lk}\},$$

$$\Omega(ilk)_2 = \{D_{ik} \neq D_{ki}, D_{il} \neq D_{li}, D_{lk} \neq D_{kl}\},$$

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

Equation (3.4) is fully characterized by the observed variables $(D^n, \mathbf{X}^n)$. In the next section, I show that equation (3.4) can be used to point identify $\beta_0$ under support conditions on the exogenous attributes.

The conditioning events $\Omega(ilk)_1$ and $\Omega(ilk)_2$ in equation (3.4) ensure the sufficient within-individual and across-individuals variation in the linking decisions to identify $\beta_0$. The intuition behind the conditioning event $\Omega(ilk)_1$ is that the components $D_{ik} \neq D_{il}, D_{ki} \neq D_{kl}, D_{li} \neq D_{lk}$ 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 $\mu_i$ is partial out by the within-individual difference $D_{ik} - D_{il}$.

Second, the event $\Omega(ilk)_2$ captures the across-individuals $i$ and $j$ variation. This variation is used to partial out the composite and unobserved factor $\gamma_i\gamma_k$. Therefore, this condition is crucial to point identify $\beta_0$. The set $\Omega(ilk)_2$ uses the directed structure of the networks. Intuitively, the conditions in $\Omega(ilk)_2$ suggest that all the links formed by the triad (i,k,l) are directed.

### 3.3.2 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 $\beta_0$.

Denote by $\tilde{\mu} \equiv (\mu_1, \cdots, \mu_n)$, $\tilde{\gamma} \equiv (\gamma_1, \cdots, \gamma_n)$, and $\mathbf{X}^n \equiv \left(X_{12}^n, \cdots, X_{n-1,n}^n\right)$.

For any $i, k \in \mathcal{N}_n$, let $F_{\varepsilon_{ik}|\mathbf{X}^n=\mathbf{x},\tilde{\mu}=\mu,\tilde{\gamma}=\gamma}$ denote the conditional distribution of $\varepsilon_{ik}$ given $\mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu, \tilde{\gamma} = \gamma$.

**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* $\{\mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu, \tilde{\gamma} = \gamma\}$. *That is for any* $(i, k), (j, l) \in \mathcal{N}_n^{(2)}$*:*

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

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

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 $F_{\tilde{\mu},\tilde{\gamma}|\mathbf{x}}$ 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 $\{\mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu, \tilde{\gamma} = \gamma\}$. Given any specification $\{\mathbf{X}^n = \mathbf{x}, \tilde{\mu} = \mu, \tilde{\gamma} = \gamma\}$, 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 (2015); Leung (2015a) and Menzel (2015).

Let $\Delta_{kl}X_i = X_{ik} - X_{il}$ 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 $\beta_0$ 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. If the 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 $\beta_0$ 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 2.3.4, I give alternative sufficient conditions for point identification when regressors are continuous with bounded support.

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

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

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

A3 states that the agent-specific fixed effects have bounded support. This assumption allows for discrete or continuously distributed fixed effects. Furthermore, their distribution could be heterogeneous as long as common bounds for their support exist.

This assumption, jointly with A2, 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:

$$\begin{aligned} Z_{kl}^{(i)} &\equiv (D_{ik} - D_{il}), \\ Z_{ik}^{(l)} &\equiv (D_{li} - D_{lk}), \\ Z_{li}^{(k)} &\equiv (D_{kl} - D_{ki}), \end{aligned}$$

**Theorem 15.** *1. Let assumptions A1 - A3 hold. Then, for any* $n$*, and any* $i, l, k \in \mathcal{N}_n$*:*

$$\begin{aligned} &Med\left[Z_{kl}^{(i)} + Z_{ik}^{(l)} + Z_{li}^{(k)} | \mathbf{X}^n = x, \Omega(ilk)_1, \Omega(ilk)_2\right] \\ &\qquad = 3 \times \operatorname{sign}\left\{\left[(X_{ik} - X_{il}) + (X_{li} - X_{lk}) + (X_{kl} - X_{ki})\right]' \beta_0\right\} \quad (3.5) \end{aligned}$$

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

$$\mathcal{X}_B = \Big\{\mathbf{x} \in \mathbf{X}^n : \text{ for any } i, j, k, l \in \mathcal{N}_n, \ | \Delta_{kl} x_i \beta_0 | \geqslant (B_U^2 - B_L^2), \text{ and } \operatorname{sign}\{\Delta_{kl} x_i \beta_0\} = \operatorname{sign}\{\Delta_{ik} x_l \beta_0\} = \operatorname{sign}\{\Delta_{li} x_k \beta_0\}\Big\}.$$

*2. Let assumptions A1 - A3 hold. Then $\beta_0$ is point identified.*

Equation (3.5) is fully characterized in terms of the observed variables $(D^n, \mathbf{X}^n)$ and it represents an identifying condition for $\beta_0$. This equation conveys two main points. First, the events $\Omega(ilk)_1, \Omega(ilk)_2$ constitute 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 difference of the links given $\Omega(ilk)$ is fully characterized by the variation in the observed attributes. Second, the set $\Omega(ilk)_1, \Omega(ilk)_2$ ensures sufficient within-individual and across-individuals variation to identify $\beta_0$ under the support conditions on the exogenous attributes.

## 3.4 Inference

In this section, I propose a semiparametric estimator for $\beta_0$ under the point identification assumptions of section 3.3.2. 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 semiparametric estimator is an M-estimator that minimizes a 3rd order U-process. This estimator adapts the methodology introduced in Chapter 2 of this dissertation to estimate a directed network. The methodology generalizes the Leapfrog estimator introduced by Abrevaya (1999b) to a network structure with multiple and unobserved interactive fixed effects.

The estimator for $\beta_0$ is consistent and has an asymptotic normal distribution. If the probability of the class $\Omega_n$ 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 $\Omega_n$ 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 $\beta_0$ in a set with arbitrarily small probability, also referred to as a thin set. In this case, $\beta_0$ is said to be irregularly identified (Newey 1990; Andrews and Schafgans 1998 and Khan and Tamer 2010).

### *3.4.1 Semiparametric Estimator*

Let $\Omega(ilk) = \Omega(ilk)_1 \cup \Omega(ilk)_2$. Consider the following limiting objective function

$$\begin{aligned} Q(b) =& 3\mathbb{E}\left[S(\mathcal{X}_B) \times \text{ sign}\left\{[(X_{ik} - X_{il}) + (X_{li} - X_{lk}) + (X_{kl} - X_{ki})]' b\right\}\right. \\ & \left. \times \left(Z_{kl}^{(i)} + Z_{ik}^{(l)} + Z_{li}^{(k)}\right) \mid \Omega(ijlk)\right], \end{aligned} \tag{3.6}$$

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

Consider a sample of size $n$

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

he semiparametric pairwise difference estimator is

$$\hat{\beta}_n = \underset{b\in\tilde{\mathcal{B}}\subset\mathbb{R}^K}{\arg\max}\, Q_n(b), \tag{3.7}$$

where $Q_n(b)$ is the sample analog of the limiting objective function. This is 3rd order U-statistic.

The following set of assumptions are sufficient to show that the semiparametric estimator, defined in equation (3.7), is consistent and distributed asymptotically normal. 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 $\Omega_n$ 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*

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

**Assumption B2.** *The parameter space $\tilde{\mathcal{B}}$ is compact and $\beta_0$ is an interior point of $\tilde{\mathcal{B}}$.*

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

1. $p_n \to p_0 \geqslant 0$, *as* $n \to \infty$.

2. $\sqrt{N}p_n \to \infty$, *as* $n \to \infty$.

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

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

where $h(z, \mathbf{P}, \mathbf{P}, b)$ denotes the conditional expectation of $h(\cdot, \cdot, \cdot, b)$ under $\mathbf{P}^3$, given its first argument. $\mathbf{P}^3$ denotes the product measure $\mathbf{P} \times \mathbf{P} \times \mathbf{P}$ for the sampling distribution $\mathbf{P}$ on $S$, and given $\mathbf{P}^3 < \infty$.

Let $||\cdot||$ denote the Frobenius matrix norm, $||(a_{ij})|| = (\sum_{i,j} a_{i,j}^2)^{1/2}$, $\nabla_m$ denote the $m$th 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 $\beta_0$.*

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 $\mathscr{M}$*

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

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 16.** *Let assumptions A1–A3, B1–B3 hold. Then,*

$$\hat{\beta}_n \xrightarrow{a.s.} \beta_0$$

*as $n \to \infty$.*

**Theorem 17.** *Let $\hat{\beta}_n$ be a value that maximizes $Q_n(\beta)$ over the parameter space $\mathcal{B}$. If assumptions A1–A3, B1–B4 hold, then:*

$$p_n \sqrt{N}(\hat{\beta}_n - \beta_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, V^{-1} \Delta V^{-1}), \quad \text{as } n \to \infty \tag{3.8}$$

*where*

$$3V = \mathbb{E}\left[\nabla_2 \tau(\cdot, \beta_0) \mid \Omega_n\right],$$

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

Theorem 16 shows that the semiparametric estimator converges to the true parameter value almost surely as the size of the network increases. Theorem 17 that the semiparametric estimator is asymptotically normal. If the probability of the class $\Omega_n$ 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 $\Omega_n$ converges to zero, then the converge rate is slower than the parametric rate. This result is a consequence of identifying $\beta_0$ in a thin set.

## 3.5 Conclusion

This is the first paper that studies the formation process of a directed network with unobserved and interactive fixed effects. I show that the methodology introduced in Chapter 2 can be generalized to analyze a broader type of heterogeneity. Specifically, the identification strategy can be adapted to point identify the vector of coefficients when the distribution of the unobserved components is not specified. Furthermore, the semiparametric estimator is consistent and asymptotically normal distributed.

As a future extension in would be interesting to consider a network formation model that accounts for network externalities as well as homophily of unobserved components. Additionally, it would be desirable to compare the results obtained under the current semiparametric frameworks with a model of random coefficients. The latter one has an advantage that it allows the identification of marginal effects.

## 3.6 Appendix

*Preliminaries*

Let

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

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

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

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

**Corollary 18.** *Let assumption A1 hold. For any $n$, and any $i,k,l \in \mathcal{N}_n$.*

$$\begin{aligned} & sgn\left\{\left[(X_{ik} - X_{il}) + (X_{li} - X_{lk}) + (X_{kl} - X_{ki})\right]' \beta_0\right\} \\ =& sgn\left\{\mathbb{E}[D_{ik} - D_{il}|\mathbf{W} = w] + \mathbb{E}[D_{li} - D_{lk}|\mathbf{W} = w] + \mathbb{E}[D_{kl} - D_{ki}|\mathbf{W} = w]\right\} \end{aligned} \tag{3.9}$$

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

$$\begin{aligned} \mathcal{X}_B = \big\{\mathbf{x} \in \mathbf{X}^n : \textit{ for any } i,j,k,l \in \mathcal{N}_n, \mid \Delta_{kl}x_i\beta_0 \mid \geqslant (B_U^2 - B_L^2), \textit{ and} \\ \text{sign}\{\Delta_{kl}x_i\beta_0\} = \text{sign}\{\Delta_{ik}x_l\beta_0\} = \text{sign}\{\Delta_{li}x_k\beta_0\}\big\}. \end{aligned}$$

*Proof of Corollary 18.*

$$\begin{aligned} & \mathbb{E}[(D_{ik} - D_{il}) + (D_{li} - D_{lk}) + (D_{kl} - D_{ki})|x,\eta,\Omega] \\ &= 3\mathbb{P}[(D_{ik} - D_{il}) + (D_{li} - D_{lk}) + (D_{kl} - D_{ki}) = 3|x,\eta,\Omega] \\ &-3\mathbb{P}[(D_{ik} - D_{il}) + (D_{li} - D_{lk}) + (D_{kl} - D_{ki}) = -3|x,\eta,\Omega] \\ =& \frac{3}{\mathbb{P}[\Omega|x,\mu]}\left[\mathbb{P}[D_{ik} = 1, D_{il} = 0, D_{li} = 1, D_{lk} = 0, D_{kl} = 1, D_{ki} = 0|x,\eta]\right. \\ &\left. - \mathbb{P}[D_{ik} = 0, D_{il} = 1, D_{li} = 0, D_{lk} = 1, D_{kl} = 0, D_{ki} = 1|x,\eta]\right]. \end{aligned}$$

Then

$$\mathbb{E}[(D_{ik} - D_{il}) + (D_{li} - D_{lk}) + (D_{kl} - D_{ki})|x,\eta,\Omega] \geqslant 0$$

if and only if

$$\mathbb{P}\left[D_{ik}=1, D_{il}=0, D_{li}=1, D_{lk}=0, D_{kl}=1, D_{ki}=0|x,\eta\right]$$

$$\geqslant$$

$$\mathbb{P}\left[D_{ik}=0, D_{il}=1, D_{li}=0, D_{lk}=1, D_{kl}=0, D_{ki}=1|x,\eta\right]$$

Where,

$$\mathbb{P}\left[D_{ik}=1, D_{il}=0, D_{li}=1, D_{lk}=0, D_{kl}=1, D_{ki}=0|x,\eta\right]=$$

$$F(w_{ik}(\beta_0))F(w_{li}(\beta_0))F(w_{kl}(\beta_0))(1-F(w_{il}(\beta_0)))(1-F(w_{lk}(\beta_0)))(1-F(w_{ki}(\beta_0)))$$

and

$$\mathbb{P}\left[D_{ik}=0, D_{il}=1, D_{li}=0, D_{lk}=1, D_{kl}=0, D_{ki}=1|x,\eta\right]=$$

$$F(w_{il}(\beta_0))F(w_{lk}(\beta_0))F(w_{ki}(\beta_0))(1-F(w_{ik}(\beta_0)))(1-F(w_{li}(\beta_0)))(1-F(w_{kl}(\beta_0)))$$

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

$$\mathcal{X}_B = \Big\{\mathbf{x} \in \mathbf{X}^n : \text{ for any } \ i,j,k,l \in \mathcal{N}_n, \ | \ \Delta_{kl} x_i \beta_0 \ | \geqslant (B_U^2 - B_L^2), \text{ and}$$
$$\operatorname{sign}\{\Delta_{kl}x_i\beta_0\} = \operatorname{sign}\{\Delta_{ik}x_l\beta_0\} = \operatorname{sign}\{\Delta_{li}x_k\beta_0\}\Big\}.$$

then:

$$\begin{aligned} |\ (x_{ik}-x_{il})\beta_0\ | \ &\geqslant \ (B_U^2 - B_L^2) \\ |\ (x_{li}-x_{lk})\beta_0\ | \ &\geqslant \ (B_U^2 - B_L^2), \\ |\ (x_{kl}-x_{ki})\beta_0\ | \ &\geqslant \ (B_U^2 - B_L^2), \end{aligned} \tag{3.10}$$

and

$$\operatorname{sign}\{\Delta_{kl}x_i\beta_0\} = \operatorname{sign}\{\Delta_{ik}x_l\beta_0\} = \operatorname{sign}\{\Delta_{li}x_k\beta_0\}. \tag{3.11}$$

Two mutually exclusive cases are possible,

$$\{\Delta_{kl}x_i\beta_0 > 0, \Delta_{ik}x_l\beta_0 > 0, \Delta_{li}x_k\beta_0 > 0\}$$

or

$$\{\Delta_{kl}x_i\beta_0 < 0, \Delta_{ik}x_l\beta_0 < 0, \Delta_{li}x_k\beta_0 < 0\}.$$

Assume the first case to be true: $\Delta_{kl}x_i\beta_0 > 0, \Delta_{ik}x_l\beta_0 > 0, \Delta_{li}x_k\beta_0 > 0$. Given that $\mathbf{x} \in \mathcal{X}_B$, if follows from A3 that:

$$\Delta_{kl}x_i\beta_0 \geqslant (B_U^2 - B_L^2) \geqslant \gamma_i(\gamma_l - \gamma_k), \quad (3.12)$$

$$\Delta_{li}x_k\beta_0 \geqslant (B_U^2 - B_L^2) \geqslant \gamma_k(\gamma_i - \gamma_l). \quad (3.13)$$

$$\Delta_{ik}x_l\beta_0 \geqslant (B_U^2 - B_L^2) \geqslant \gamma_l(\gamma_i - \gamma_k). \quad (3.14)$$

Equation (3.11) can be equivalently written as:

$$x_{ik}\beta_0 + \mu_i + \gamma_i\gamma_k \geqslant x_{il}\beta_0 + \mu_i + \gamma_i\gamma_l$$

$$w_{ik}(\beta_0) \geqslant w_{il}(\beta_0)$$

Equation (3.12) can be equivalently written as:

$$x_{kl}\beta_0 + \mu_k + \gamma_k\gamma_l \geqslant x_{ki}\beta_0 + \mu_k + \gamma_k\gamma_i$$

$$w_{kl}(\beta_0) \geqslant w_{ki}(\beta_0)$$

Equation (3.13) can be equivalently written as:

$$x_{li}\beta_0 + \mu_l + \gamma_l\gamma_i \geqslant x_{lk}\beta_0 + \mu_l + \gamma_l\gamma_l$$

$$w_{li}(\beta_0) \geqslant w_{lk}(\beta_0)$$

Therefore:

$$\mathbb{E}\left[(D_{ik} - D_{il}) + (D_{li} - D_{lk}) + (D_{kl} - D_{ki})|x, \eta, \Omega\right] \geqslant 0.$$

Equivalently, for the second case:

$$\{\Delta_{kl}x_i\beta_0 < 0, \Delta_{ik}x_l\beta_0 < 0, \Delta_{li}x_k\beta_0 < 0\}.$$

$$\Delta_{kl}x_i\beta_0 \leqslant (B_L^2 - B_U^2) \leqslant \gamma_i(\gamma_l - \gamma_k), \quad (3.15)$$

$$\Delta_{li}x_k\beta_0 \leqslant (B_L^2 - B_U^2) \leqslant \gamma_k(\gamma_i - \gamma_l). \quad (3.16)$$

$$\Delta_{ik}x_l\beta_0 \leqslant (B_L^2 - B_U^2) \leqslant \gamma_l(\gamma_i - \gamma_k). \quad (3.17)$$

Equation (3.14) can be equivalently written as:

$$x_{ik}\beta_0 + \mu_i + \gamma_i\gamma_k \leqslant x_{il}\beta_0 + \mu_i + \gamma_i\gamma_l$$

$$w_{ik}(\beta_0) \leqslant w_{il}(\beta_0)$$

Equation (3.15) can be equivalently written as:

$$x_{kl}\beta_0 + \mu_k + \gamma_k\gamma_l \leqslant x_{ki}\beta_0 + \mu_k + \gamma_k\gamma_i$$

$$w_{kl}(\beta_0) \leqslant w_{ki}(\beta_0)$$

Equation (3.16) can be equivalently written as:

$$x_{li}\beta_0 + \mu_l + \gamma_l\gamma_i \leqslant x_{lk}\beta_0 + \mu_l + \gamma_l\gamma_l$$

$$w_{li}(\beta_0) \leqslant w_{lk}(\beta_0)$$

Therefore:

$$\mathbb{E}\left[(D_{ik} - D_{il}) + (D_{li} - D_{lk}) + (D_{kl} - D_{ki})|x, \eta, \Omega\right] \leqslant 0.$$

■

**Corollary 19.** *Let assumption A1 hold. For any $n$, and any $i, k, l \in \mathcal{N}_n$.*

$$\begin{aligned} Med\left[Z_{kl}^{(i)} + Z_{ik}^{(l)} + Z_{li}^{(k)}|\mathbf{X}^n = x, \Omega(ilk)\right] \\ = 3 \times \text{sign}\Big\{\mathbb{P}\left[Z_{kl}^{(i)} + Z_{ik}^{(l)} + Z_{li}^{(k)} = 3|x, \eta, \Omega(ilk)\right] \\ - \mathbb{P}\left[Z_{kl}^{(i)} + Z_{ik}^{(l)} + Z_{li}^{(k)} = -3|x, \eta, \Omega(ilk)\right]\Big\} \end{aligned} \tag{3.18}$$

*Proof of Corollary 19.* Follows from $Z_{kl}^{(i)} + Z_{ik}^{(l)} + Z_{li}^{(k)} \,|\, [\mathbf{X}^n = x, \Omega(ilk)]$, being a Bernoulli random variable with support $\{-3, 3\}$. ■

*Proof of Theorem 15.* Given corollaries 18 and 19, it suffices to show that

$$\begin{aligned} &\text{sign}\left\{\mathbb{E}\left[(D_{ik} - D_{il}) + (D_{li} - D_{lk}) + (D_{kl} - D_{ki})|x, \eta, \Omega\right]\right\} \\ &= \text{sign}\left\{\mathbb{P}\left[(D_{ik} - D_{il}) + (D_{li} - D_{lk}) + (D_{kl} - D_{ki}) = 3|x, \eta, \Omega(ilk)\right]\right. \\ &\left. - \mathbb{P}\left[(D_{ik} - D_{il}) + (D_{li} - D_{lk}) + (D_{kl} - D_{ki}) = 3|x, \eta, \Omega(ilk)\right]\right\} \end{aligned} \tag{3.19}$$

This result follows from

$$\begin{aligned} &\mathbb{E}\left[(D_{ik}-D_{il})+(D_{li}-D_{lk})+(D_{kl}-D_{ki})|x,\eta,\Omega\right] \\ =&\frac{3}{\mathbb{P}\left[\Omega|x,\mu\right]}\left[\mathbb{P}\left[D_{ik}=1,D_{il}=0,D_{li}=1,D_{lk}=0,D_{kl}=1,D_{ki}=0|x,\eta\right]\right. \\ &\left.-\,\mathbb{P}\left[D_{ik}=0,D_{il}=1,D_{li}=0,D_{lk}=1,D_{kl}=0,D_{ki}=1|x,\eta\right]\right]. \end{aligned}$$

■

**Part II.**

Take any $b \neq \beta_0$, by assumption A2 we can find a set of values of $x \equiv (x_{ik}-x_{il}) + (x_{li}-x_{lk}) + (x_{kl}-x_{ki})$ such that $sgn\left(x'b\right) \neq \left(x'\beta_0\right)$ and such that has positive measure. In other words, let

$$\mathcal{X}_b \equiv \left[x \in \mathbb{R}^K : sgn\left(x'b\right) \neq \left(x'\beta_0\right)\right]$$

Assumption A2 guarantees that $\int_{\mathcal{X}_b} dF > 0$ and therefore $\beta_0$ is identified up to scale.

*Proof of Theorem 16.* Given assumptions A1–A3, B1–B3, it follows from Theorem 7 in Chapter 2. ■

*Proof of Theorem 17.* Given assumptions A1–A3, B1–B3, it follows from Theorem 8 in Chapter 2. ■

# 4

# Identification of Endogenous Effects in Panel Data Models with Unobserved Network Structure and a Large Number of Covariates

## 4.1 Introduction

The literature of social interactions has focused on understanding the effects that strategic interactions have over several economic outcomes. The approach usually implemented in previous studies assumes knowledge of the relevance group or restricts the type of interactions, such as symmetry or interactions within groups (e.g. Manski 1993; Brock and Durlauf 2002). However, Chandrasekhar and Lewis (2014) shows that the predicted social effects are biased if the true network of interactions is not correctly measured or observed. Furthermore, the literature of network theory provides evidence that different agents within a network have different centrality roles and therefore there is an implicit heterogeneity (Jackson 2010). This paper overcomes these difficulties by introducing a methodology that recovers consistently the reference groups and the social interaction effects when the network is unobserved, and the social effects are heterogeneous across individuals.

In this paper, I consider a high-dimensional panel data model in which the decision of each agent could be influenced by her exogenous characteristics, as well as the characteristics and decisions of her peers. The main features of this economic model are that the network of interactions is unknown, and the social interaction parameters are heterogeneous across individuals. That is, the researcher doesn't have prior information regarding the reference group of each agent, and the strategic interaction effects could be asymmetric across agents. This is the first paper that allows for both heterogeneous endogenous effects and heterogeneous exogenous effects when the network of interactions is unobserved (Manresa 2013).

This paper has two main contributions. The first contribution is to propose a new methodology to identify and estimate the reference group of interactions for each individual. The reference group represents the set of agents that influence directly a particular individual, and these effects are known as first order effects. The reference groups determine the type of interactions within a network. I give sufficient conditions to identify the parameters associated with the first order effects from the remaining higher order effects. I then use elements from the literature of regularization methods and approximately sparse models to consistently estimate the parameters associated with the first order effects, treating the remaining effects as nuisance parameters (Tibshirani 1996; Belloni et al. 2011, 2013). The econometric framework considered in this paper is suitable to analyze networks that are sparse and stable over time. These are structures of interactions for which the decision of any agent is influenced just by a small group of agents in the network, and this group remains relatively constant across time. This class of networks has been broadly studied in network theory and they are relevant for many economic applications. Jackson and Watts (2002).

The second contribution is to provide sufficient conditions to point identify the endogenous and exogenous social effects. Disentangling these effects are empirically relevant because they have different policy implications. Specifically, the peers' exogenous attributes could be modified by changing the composition of the reference group. However, the endogenous effects cannot be designed or controlled by the scientist. The identification strategy uses the exogenous variation across time of observed attributes associated with the second order effects. That is the exogenous attributes of the peers that do not have a first order effect (Bramoullé et al. 2009).

The paper is structured as follows. In section 2, I formalize the model of strategic interactions. Section 3 develops the identification argument and motivates the estimation method. Section 4 describes the estimation method. In particular, I detail a high-dimensional panel estimation method to recover the network of interactions and the Panel Post-Lasso estimation of the endogenous and exogenous effects. In Section 6, I explore the finite sample performance of the method by Monte Carlo simulations. Finally, Section 7 concludes. All the proofs are provided in the Appendix.

## 4.2 Model

In this section, I characterize the strategic interaction model as a Bayesian Game. The theoretical model builds on the strategic model introduced by Blume et al. (2013). Let $G$ denote a Bayesian game characterized by:

$$G = \left[\mathcal{N}, \{Y_i\}_{i=1}^{N}, \{u_i(\cdot)\}_{i=1}^{N}, \Gamma, \mathcal{F}(\cdot)\right],$$

where $\mathcal{N} = \{1, \cdots, N\}$ is the set of players in the network. Assume that at time $t \in T$ any agent $i \in \mathcal{N}$ is fully characterized by the triplet $(x_{it}, \epsilon_{it}, \alpha_i)$, where $x_{it}$ is an observed attribute that is publicly known. To simplicity the exposition, let $x_{it}$ be a scalar. The generalization to a $p \times 1$ dimensional vector with $p \geqslant 2$ is trivial. $\alpha_i$ is an agent-specific characteristic that also is publicly known by the agents, but unobserved to the researcher. The random component $\epsilon_{it}$ is privately observed by agent $i$, but unobserved to the researcher. This component captures idiosyncratic factors that are individual and time specific. Agent $i$'s type at time $t \in T$ is characterized by the triplet $(X_t, \alpha, \epsilon_{it}) \in \mathbb{R}^{2N+1}$, where $X_t = (x_{1t}, \ldots, x_{Nt})'$ is the profile of observed attributes, and $\alpha = (\alpha_1, \cdots, \alpha_N)$ is the profile of agent-specific attributes.

Let $\Gamma \subseteq \mathbb{R}^{3N}$ denote the subspace of types, with a characteristic element $(X_t, \epsilon_t, \alpha) \in \Gamma$, where $\epsilon_t = (\epsilon_{1t}, \cdots, \epsilon_{Nt})$. The probability distribution over the types $\Gamma$ is given by $\mathcal{F}$.

The set of action of individual $i$ is denoted by $Y_i$. Let $u_i(\cdot)$ be the utility function of $i$, which has the following nonlinear specification.

$$u_i\left[y_{it}, y_{-it}; X_t, \alpha, \epsilon_{it}\right] = \left[\alpha_i + \beta_i x_{it} + \epsilon_{it} + \rho_i \sum_{\mathcal{A}_i} a_{ij} x_{jt}\right] y_{it} - \frac{1}{2} y_{it}^2 - \frac{\theta_i}{2}\left(y_{it} - \sum_{\mathcal{A}_i} a_{ij} y_{jt}\right)^2, \tag{4.1}$$

where $y_{it}$ denotes the strategy of individual $i$ at time $t$, $y_{-it}$ denotes the strategy of all others $N-1$ individuals at time $t$, and $\mathcal{A}_i \subseteq \mathcal{N}$ denotes individual $i$'s reference group.

The utility function in equation (4.1) conveys three main insights. First, only the agents in $\mathcal{A}_i$ influence agent $i$'s utility. This feature captures the sparsity of the network, which implies that the decision of each individual in the network is affected only by a small number agents and not by all the individuals in the network. In this paper, the reference groups are known to the agents but are unknown to the researcher. Furthermore, equation (4.1) assumes that agent $i$ assumes group of interactions remains stable across time. This observation is consistent with an equilibrium network from a network formation model (Jackson and Watts 2002).

Second, the exogenous attributes of agent $i$'s peers generate spillover effects. This effect is known as the exogenous social effect (Manski 1993), and is captured by

$$\rho_i \sum_{\mathcal{A}_i} a_{ij} x_{jt}.$$

Finally, the decisions of agent $i$'s peers have a direct effect on agent $i$'s utility. This effect is known as the endogenous social effect, and is captured by

$$\sum_{\mathcal{A}_i}^{N} a_{ij} y_{jt}.$$

Note that in equation (4.1), the social interaction effect is modeled as peer pressure, which is captured by the loss of benefit due to the deviations of $y_i$ from the decisions taken by the rest of agents in $i$'s reference group.

In this paper, I provide sufficient conditions to identify and estimate the unobserved structure of interactions given by the adjacency matrix defined as $A \equiv \{a_{ij}\}_{i,j=1}^{N}$, as well as the social interactions parameters $\{\rho_i, \theta_i\}_{i=1}^{N}$. I assume the network is directed and weighted. A network is directed if the adjacency matrix is asymmetric, that is, if for any entries $(i,j)$th and $(j,i)$th the adjacency matrix can potentially be different, $a_{ij} \neq a_{ji}$. A network is weighted if any entry $(i,j)$th of the adjacency matrix takes a value within the interval $[0,1]$. In other words, $a_{ij} \in [0,1]$, where $a_{ij} \neq 0$ if the agents $i$ is influenced by the exogenous attributes or decisions of agent $j$, and $a_{ij} = 0$ otherwise. Furthermore, I normalize the value of self-ties to zero, that is, $a_{ij} = 0$ for all $i \in \mathcal{N}$.

As opposed to the existing literature, I allowed for both heterogeneous endogenous effects and exogenous effect. The identification strategy relies on recovering the adjacency matrix consistently and does not impose symmetry or homogeneity restrictions on the adjacency matrix. This approach suitable for networks which are unobserved or measured with error.

### *4.2.1 Existence of Bayesian Nash Equilibrium*

The next assumption is used to guarantee the existence of the Bayesian Nash Equilibrium. If the following condition does not hold, a corner solution is obtained.

**Assumption A1.** *The following holds for any $N$,*

1. *For any agent in the network, $\theta_i > 0$.*
2. *For any agent in the network, $||A_{i\cdot}||_0 \equiv \sum_{j \neq i} |a_{ij}| < 1$.*
3. *The strategy function is square integrable for any individual in the network. That is $\sigma_i : \mathcal{R}^{N+1} \rightarrow Y_i$ is in $\mathbb{L}^2$ for any $i \in \mathcal{N}$.*

A1.1 guarantees the existence of a contraction mapping, which is the technique used to prove the existence of the equilibrium. This condition implies that the peer effects arise due to strategic complementarity interactions. A1.2 imposes an upper bound on the strength of the peer influences. This assumption rules out explosive solutions of the model, and it has been commonly used in the literature of social interactions Brock and Durlauf (2005); Xu (2010).

Finally, A1.3 is a regularity condition which guarantees the existence of the expected utility.

The following proposition states the existence and uniqueness of a Bayesian Bayesian Nash Equilibrium. In order to simplify its exposition, consider the following notation. Denote the $i$th row of any $n \times m$ matrix $W$ by $W_{i\cdot} \in \mathbb{R}^m$. For any parameters $\kappa$ and $\phi$, define the diagonal matrix $\Lambda_{\kappa,\phi}$ with $(i,i)$th element $\frac{\kappa_i}{1+\phi_i}$. That is,

$$\Lambda_{\kappa,\phi} \equiv \text{diag}\left[\frac{\kappa_1}{1+\phi_1}, \cdots, \frac{\kappa_N}{1+\phi_N}\right].$$

**Proposition 20** (Existence of a BNE)**.** *Given Assumption A1, the Bayesian game $G$ has a unique Bayesian Nash Equilibrium (BNE). Furthermore, the Equilibrium Strategy Profile is given by,*

$$Y_t = \left[\boldsymbol{I}_N - \Lambda_{\theta,\theta} A\right]^{-1} \left(\Lambda_{\beta,\theta} + \Lambda_{\rho,\theta} A\right) X_t + \Lambda_{1,\theta}\left(\alpha + \epsilon_t\right) + H(X_t, \alpha, \epsilon_t) \tag{4.2}$$

*where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix, and $H(X_t, \alpha, \epsilon_t)$ is an $N \times 1$ vector with $i$th element*

$$H_i(X_t, \alpha, \epsilon_t) = \left(\frac{1}{1+\theta_i}\right) \sum_{j=0}^{N} \left[(\Lambda_{\theta,\theta} A)^j\right]_{i\cdot} \left(\alpha + \mathbb{E}\left[\epsilon_t | X_t, \alpha, \epsilon_{i,t}\right]\right) \tag{4.3}$$

*for each $t = 1, \cdots, T$.*

The component $H_i(X_t, \alpha, \epsilon_t)$ captures the higher-order beliefs about the idiosyncratic components. The next assumption implies that the observed characteristics are independent from the idiosyncratic components. This condition allows for unrestricted dependence between the observed covariates and the agent-specific fixed effects. Furthermore, it leaves unrestricted the specification of $H_i(\alpha_i, \epsilon_{it})$.

**Assumption A2.** *For any $t, t' \in T$: $\epsilon_{t'} \perp\!\!\!\perp X_t$.*

Given A2, $\mathbb{E}\left[\epsilon_t \mid X_t, \alpha, \epsilon_{i,t}\right] = \mathbb{E}\left[\epsilon_t \mid \alpha, \epsilon_{i,t}\right]$. Hence, the beliefs simplify to:

$$H_i(X_t, \alpha, \epsilon_t) \equiv H_i(\alpha_i, \epsilon_{i,t}). \tag{4.4}$$

Let

$$\begin{aligned} \Pi(\beta, \theta, \rho) &\equiv \left[\mathbf{I}_N - \Lambda_{\theta,\theta} A\right]^{-1} \left\{\Lambda_{\beta,\theta} + \Lambda_{\rho,\theta} A\right\}, \quad (4.5) \\ \alpha_\theta &\equiv \Lambda_{1,\theta}\alpha, \\ \epsilon_\theta &\equiv \Lambda_{1,\theta}\epsilon_t, \end{aligned}$$

where $\Pi(\beta, \theta, \rho)$ is an $N \times N$ matrix of path connected effects. To simplify exposition, I will denote $\Pi(\beta, \theta, \rho)$ by $\Pi$. The Bayesian Nash Equilibrium in equation 4.2 is written as,

$$Y_t = \alpha_\theta + \Pi X_t + v_t \tag{4.6}$$

where $v_t$ is a composite idiosyncratic component given by $v_t \equiv \epsilon_\theta + H(\alpha, \epsilon_t)$, with mean value equal to zero.

Consider the $i$th equation of reduced form equilibrium in equation (4.6)

$$y_{it} = \frac{\alpha_i}{1 + \theta_i} + \sum_{j \neq i} \pi_{ij} x_{jt} + v_{it}, \tag{4.7}$$

and note that this is a high-dimensional econometric model, where the number of regressors, and hence parameters, is much larger than the number of periods of time. That is $N - 1 >> T$. In the next section, I explain in detail the approximate sparse specification of the Bayesian Nash Equilibrium.

## 4.3 Approximate Sparse Model

In this section, I highlight the identification strategy used to recover the structural parameters $\left\{\theta_i, \beta_i, \rho_i, \{a_{i,j}\}_{j=1}^{N}\right\}_{i=1}^{N}$. In equation (4.6), the parameters interact in a difficult nonlinear way which complicates the identification of these coefficients using the reduced form equation. Alternatively, consider the next specification which is derived from the first-order conditions of the expected utility maximization problem.

$$y_{i,t} = \alpha_i + \frac{\beta_i}{1+\theta_i} x_{i,t} + \frac{\rho_i}{1+\theta_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} a_{ij} x_{j,t} + \frac{\theta_i}{1+\theta_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} a_{ij} \mathbb{E}\left[y_{j,t} | X_1, \cdots, X_T\right] + \epsilon_{i,t}, \quad (4.8)$$

for $i = 1, \cdots, N$ and $t = 1, \cdots, T$, where $N > T$.

Equation (4.8) is not the usual specification studied in the social interactions literature, e.g Manski (1993); Bramoullé et al. (2009); De Giorgi et al. (2010) and Blume et al. (2013). In this paper, the endogenous effect is characterized by the weighted average of the conditional expectations of each agent's peers. As opposed to the actual decisions taken by the peers. This conceptual difference is obtained from the strategic interaction model. The conditional expectation provides the enough information to identify the endogenous effect.

Given that $N >> T$, there are more parameters to recover in the model than observations. The following sparsity assumption reduces the dimensionality of the parameters to estimate. Sparse networks have been discussed in Jackson (2010) and Manresa (2013). Intuitively, the next condition suggests that each agents' decision in the network is affected by a small number of individuals. Note that the $N \times N$ matrix $\Pi$ represents the matrix of effects due to path connections. Specifically, denote the $(i, j)$th element of the matrix $\Pi$ by $\pi_{ij}$. The coefficient $\pi_{ij}$ represents the cumulative effect of agent $j$'s decision over agent $i$'s decision. That is, $\pi_{ij}$ captures the directed influence that agent $j$ has over agent $i$ if they have directed friends, in addition of the indirect effects through friends in common. As shown in Bramoullé and Kranton (2007), for any game of strategic complementaries the his matrix will be dense and will not satisfy a sparsity condition. Note, that $\Pi$ can be written as the geometric expansion:

$$\begin{aligned}\Pi =& \sum_{l=0}^{\infty} [\Lambda_{\theta,\theta} A]^l \{\Lambda_{\beta,\theta} + \Lambda_{\rho,\theta} A\} \\ =& \Lambda_{\beta,\theta} + \Lambda_{\beta,\rho,\theta^2} A + \Lambda_{\beta,\rho,\theta^2} \Lambda_{\theta,\theta} A^2 + \Lambda_{\beta,\rho,\theta^2} \Lambda_{\theta,\theta}^2 A^3 + \cdots\end{aligned}$$

where $\Lambda_{\beta,\rho,\theta^2} \equiv diag\left[\left\{\frac{\rho_i(1+\theta_i)+\theta_i\beta_i}{(1+\theta_i)^2}\right\}\right]$.

A1.1 and A1.2 ensure that the higher order effects dissipate as $l$ grows to infinity, at a rate given by $\Lambda_{\theta,\theta}^l A^l$, since the magnitude of the network effects is bounded. This insight is used to impose an upper bound in the magnitude of the second order effects and hence higher order effects.

The next assumption restricts the number of first-order effects. Intuitively, it states that any given agent is connected with only a small number of agents in the network. This assumption implies that the network of interactions is sparse (Jackson 2010).

**Assumption A3.** *The following holds for any* $N$*,*

$$\max_{i\in\{1,\cdots,N\}} ||A_{i\cdot}||_0 = \max_{i\in\{1,\cdots,N\}} \sum_{j=1}^{N} \mathbb{1}\{a_{ij} \neq 0\} \leqslant s_{T,N} = o(T). \quad (4.9)$$

The rate of sparsity is determined by the number of time periods in the panel since the identification power is obtained from the exogenous variations across the time dimension. Asymptotically, I allow for both $s_{T,N}$ to grow with $N$ and $T$. A3 does not impose sparsity over the path connected effects $\Pi$. The following assumption guarantee that the matrix $\Pi$ has only a small number of large coefficients and the remaining coefficients are approximately close to zero.

**Assumption A4.** *The following hold for any N,*

$$\max_{i\in\{1,\cdots,N\}}\left[\sum_{k=1}^{N}\pi_{ik}^{2}\right]=O_p\left(\sqrt{\frac{s_{T,N}}{N^2}}\right). \tag{4.10}$$

$$\max_{i\in\{1,\cdots,N\}}\left[\sum_{k=1}^{N}\pi_{ki}^{2}\right]=O_p\left(1\right). \tag{4.11}$$

*where* $\frac{s}{N}\rightarrow 0$ *as* $s,N,T\rightarrow\infty$.

A4 imposes an upper bound in the magnitude of the second and higher order effects. In particular, it implies that the higher order effects are approximately close to zero, although not equal to zero. Assumption bounds (4.10) the magnitude of the in-degree path connected effects of length 2. Meanwhile, condition (4.11) bounds the out-degree path connected effects of the same size. A4 does not restrict the in-degree or out-degree of the adjacency matrix for each particular agent. In other words, assumption A4 do not restrict the number of agents to which an agent can affect or the number of agents that affects that agent's decision.

Given A3 and A4, it is possible to approximate the reduced form equation by a small number of covariates that capture the first order effects. That is, the reduced form equation 4.6 is approximately sparse (Belloni et al. 2011, 2013, 2012). A4 suggests that a large number of regressors have coefficients close to zero.

Before introducing the approximately sparse model, consider the following notation. For any random variable $Z_{i,t}$, denote by $\tilde{Z}_{i,t}$ its deviation from the within-individual $i$'s sample average over time.

$$\tilde{Z}_{i,t}=Z_{i,t}-\frac{1}{T}\sum_{t=1}^{T}Z_{i,t}.$$

Using a within-group transformation to difference out the agent-specific fixed effect, the reduced form Bayesian Nash Equilibrium can be expressed as:

$$\tilde{Y}_t=\Pi\tilde{X}_t+\tilde{v}_t \tag{4.12}$$

The model in (4.12) is written using assumptions A3 and A4 to emphasize its approximately sparse specification. The matrix of path connected effects $\Pi$ is decomposed as $\Pi \equiv \Pi^{(s)} + \Pi^{(a)}$, where

$$\Pi^{(s)} = \Lambda_{\beta,\theta} + \Lambda_{\beta,\rho,\theta^2} A,$$

$$\Pi^{(a)} = \Lambda_{\beta,\rho,\theta^2} \left( \sum_{l=1}^{\infty} \Lambda_{\theta,\theta}^{l} A^{l+1} \right).$$

Then for any $t \in T$,

$$\tilde{Y}_t = \left[\Pi^{(s)} + \Pi^{(a)}\right] \tilde{X}_t + \tilde{v}_t,$$

$$\tilde{Y}_t = \Pi^{(s)} \tilde{X}_t + \Psi(\tilde{X}_t) + \tilde{v}_t. \tag{4.13}$$

where $\Psi(\tilde{X}_t) = \Pi^{(a)} \tilde{X}_t$.

Let $\Psi_i(\tilde{X}_t)$ denote the $i$th entry of the $N \times 1$ vector $\Psi(\tilde{X}_t)$. The following proposition shows that the approximation error is relatively small.

**Proposition 21.** *Given Assumptions A3 and A4, the magnitude of the approximation error, $\Psi_i(\tilde{X}_t)$ is bounded by*

$$\left[ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \Psi_i(\tilde{X}_t)^2 \right]^{1/2} \leqslant c_s = O_p\left(\frac{s_{T,N}}{NT}\right)$$

Proposition (21) implies that equation (4.13) can be approximated by at most $s_{T,N}$ coefficients, and the approximation errors is no larger than the conjectured size of the estimation error $\sqrt{\frac{s_{N,T}}{T}}$. The previous result allows us to recover $\Pi^{(s)}$, and $\mathcal{A}_i$ for any $i \in \mathcal{N}$.

## 4.4 Identification

In this section, I show point identification of the structural parameters under the sparsity and bounded effects assumptions, A3 and A4. Consider the linear in means specification in equation (4.8) under the sparsity assumption.

$$\tilde{y}_{i,t} = \frac{\beta_i}{1+\theta_i}\tilde{x}_{i,t} + \frac{\rho_i}{1+\theta_i}\sum_{j\in\mathcal{A}_i} a_{ij}\tilde{x}_{j,t} + \frac{\theta_i}{1+\theta_i}\sum_{j\in\mathcal{A}_i} a_{ij}\mathbb{E}\left[\tilde{y}_{j,t}|X_1,\cdots,X_T\right] + \tilde{\epsilon}_{i,t},$$

$$\tilde{y}_{i,t} = \omega_i^1\ \tilde{x}_{i,t} + \sum_{j\in\mathcal{A}_i}\omega_{i,j}^2\tilde{x}_{j,t} + \sum_{j\in\mathcal{A}_i}\omega_{i,j}^3\left\{\sum_{k\in\mathcal{A}_j}\pi_{jk}\tilde{x}_{k,t}\right\} + \tilde{\epsilon}_{i,t}, \tag{4.14}$$

where $\omega_i^1 \equiv \frac{\beta_i}{1+\theta_i}$, $\omega_{i,j}^2 \equiv \frac{\rho_i}{1+\theta_i}$, and $\omega_{i,j}^3 \equiv \frac{\theta_i}{1+\theta_i}$. The main insight in equation (4.14) is that the beliefs of peers actions can be approximated by a small subset of all the regressors under an approximation error. This result is used to prove point identification of the structural parameters.

**Proposition 22.** *The following holds for any $N$. Under Assumptions A1-A4, the structural parameters $\left\{\beta_i, \theta_i, \gamma_i, \{a_{ij}\}_{j\neq i}\right\}_{i\in\mathcal{N}}$ are point identified if*

$$\mathcal{A}_i \subset \mathcal{A}_j \tag{4.15}$$

*for all $j \in \mathcal{A}_i$, and $i \in \mathcal{N}$.*

Condition (4.15) is sufficient to point identify the structural parameters. Intuitively, this condition holds when the reference groups are not equal. Specifically, this holds when agent $i$'s peers are connected to individuals that are outside agent $i$'s reference groups. If this is the case, the exogenous attributes of those peers that are not within individual $i$'s reference group can be used as instruments to generate exogenous variation in the endogenous and exogenous effects. If this condition does not hold for any $j \in A_i$, then the structural parameters for that individual cannot be disentangled into exogenous and endogenous effects. A similar identification strategy has been proposed by Bramoullé et al. (2009) and De Giorgi et al. (2010).

## 4.5 Inference

In this section, I discuss an estimation method that is suitable for the approximately sparse models introduced in section 3.3. The estimation method falls within the class of Lasso regularization methods (Tibshirani 1996; Candes and Tao 2007; Gautier and Tsybakov 2011; Fan and Liao 2011; Belloni et al. 2012; Horowitz and Huang 2012; Caner et al. 2013 and Caner et al. 2013.). Next, I introduce the estimator and develop it's asymptotic distribution.

### *4.5.1 Approximately Sparse Panel Data Models*

Consider the following moment conditions derived from assumption A2

$$\frac{1}{T}\sum_{t=1}^{T}\left(\tilde{X}_t\left(\tilde{y}_{i,t}-\tilde{X}_t'\Pi_i\right)\right)=0. \tag{4.16}$$

Since $N >> T$, there are more parameters to estimate than the number of moment conditions in 4.16. Therefore, the model is unfeasible as it is posed, and a penalization needs to be implemented. Consider the Lasso estimation technique, which solves the following penalized minimization model:

The Lasso estimate, $\hat{\Pi}^L$ is defined as the solution to the following penalized minimization model.

$$\hat{\Pi}^L = \arg\min_{\Pi}\left\{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\tilde{y}_{i,t}-\tilde{X}_t'\Pi_i\right]^2+\frac{\lambda}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\phi}_{ij}\left|\pi_{ij}\right|\right\}, \tag{4.17}$$

where $\hat{\Pi}^L$ denotes the Lasso estimate. The first component of the loss function in (4.17) is the usual of the panel data OLS loss function given by

$$Q_{N,T}(\Pi)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\tilde{y}_{i,t}-\tilde{X}_t'\Pi_i\right]^2$$

In addition, the loss function in (4.17) includes an $L_1$ penalization which forces a sparse solution.

$$\frac{\lambda}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{\phi}_{ij}\left|\pi_{ij}\right|$$

The Lasso problem is characterized by two main regularization parameters. A main penalty parameter given by $\lambda$, and covariate-specific penalty loadings given by $\left\{\hat{\phi}_{ij}\right\}_{i,j=1}^{N}$. The main penalty dictates the amount of regularization in the Lasso procedure. The covariate-specific penalty loadings provide information regarding the regressors that are relevant for explaining the endogenous variable.

The intuition behind the penalty loadings is to self-normalizations the first order conditions of the Lasso problem by using data-dependent penalty loadings. In order to shed some light into this point, consider the Kuhn-Tucker conditions with respect to $\pi_{ij}$, that minimize the function (4.17)

$$2\left|\sum_{t=1}^{T}\left(\hat{\phi}_{ij}\right)^{-1}\left[\tilde{v}_{it}\tilde{x}_{jt}\right]\right|\leqslant\lambda. \tag{4.18}$$

Let $S_{ij}$ denote the score given by

$$S_{ij}=\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\phi}_{ij}\right)^{-1}\left[\tilde{v}_{it}\tilde{x}_{jt}\right], \tag{4.19}$$

which is a $N\times 1$ vector. The main penalty parameter $\lambda$ is set to guarantee the good performance of Lasso. For this purpose, $\lambda$ is set large enough to dominate the noise represented by the score vector of the $N$ regression problems simultaneously. That is, the penalty parameter should satisfy

$$P\left(\frac{\lambda}{T}\geqslant K\max_{1\leqslant j\leqslant N}\left|\left|S_j\right|\right|_{\infty}\right)\to 1 \tag{4.20}$$

For some constant $K\geqslant 1$.

Belloni et al. (2012) conclude that:

$$\mathbb{P}\left(\sqrt{T}\max_{1\leqslant j\leqslant N}||S_j||_\infty \leqslant 2\Phi^{-1}\left(1-\gamma/(2N^2)\right)\right) \leqslant 1-\gamma+o(1) \tag{4.21}$$

Then the inequality in (4.21) implies that the bound in equation (4.20) can be satisfied by setting the penalty parameter as follows:

$$\lambda = 2K\sqrt{T}\Phi^{-1}(1-\gamma/(2N^2)) \tag{4.22}$$

The additional regularity conditions are required so that the Lasso estimator obtains a good performance. In particular the next assumptions control the good behavior of the Empirical Gram Matrix. Let M denote the empirical Gram matrix, $M \equiv \frac{1}{T}\sum_{t=1}^{T}\tilde{X}_t\tilde{X}_t'$. Note that if $N > T$ the Empirical Gram matrix is singular. However, the Lasso estimator will have a good performance if the sub matrices of the Empirical Gram matrix satisfy the next condition.

**Definition 2** (m-Sparse Eigenvalues)**.** *Let $M$ be an $N \times N$ matrix, then the minimal and maximal m-sparse eigenvalues of matrix $M$ are defined as,*

$$\phi_{min}(M) = \min_{\delta\in\Delta(m)} \delta' M\delta$$

$$\phi_{min}(M) = \max_{\delta\in\Delta(m)} \delta' M\delta$$

*where* $\Delta(m) = \left\{\delta \in \mathbb{R}^N : ||\delta||_0 \leqslant m, ||\delta||_2 = 1\right\}$.

where $\Delta(m)$ represents the subspace of a unit sphere formed by m-sparse vectors.

**Assumption A5.** *For any matrix $M$, there exist two constants $0 < k' \leqslant k'' < \infty$ that do not depend on $T$, but might depend on $C$ such that with probability approaching one*

$$k' \leqslant \phi_{min}(Cs)(M) \leqslant \phi_{max}(C_s)(M) \leqslant k''$$

*as $T \to \infty$.*

This condition implies that the minimum restricted eigenvalue can be bound away from zero. To derive the convergence rates, we require the following moment conditions on the reduced form errors and the regressors.

**Assumption A6.** *Assume the following*

*i.* *There is a random sample* $\{y_{it}, w_{it}\}_{i,t}$ .

*ii.* $\max_{i\leqslant N, j\leqslant N} \bar{\mathbb{E}}\left[\tilde{y}_{it}^2\right] + \bar{\mathbb{E}}\left[||\tilde{x}_{jt}^2\tilde{y}_{it}^2||\right] + 1/\bar{\mathbb{E}}\left[\tilde{x}_{jt}^2\tilde{v}_{it}^2\right] \leqslant 1$

*iii.* $\max_{i\leqslant N, j\leqslant N} \bar{\mathbb{E}}\left[||\tilde{x}_{it}^3\tilde{v}_{it}^3||\right] \leqslant K_n$

*iv.* $K_n^2 \log^3(N) = o(T)$ *and* $s\log(N) = o(T)$.

**Proposition 23** (Oracle bounds)**.** *Under Assumptions A1-A6, the Lasso estimates specified by 4.17 with Penalty level and loadings 4.22 obtain the Oracle rates*

$$\max_{1\leqslant i\leqslant N} \left|\left|\hat{\Pi}_{i\cdot} - \Pi_{i\cdot}^0\right|\right|_1 \leqslant O_p\left(\frac{\mu^2}{(k_C)^2}\left\{\sqrt{\frac{2s^2 log(N)}{T}}\right\}\right) \tag{4.23}$$

$$\max_{1\leqslant i\leqslant N} \left|\left|\tilde{X}_t'\hat{\Pi}_i - \tilde{X}_t'\Pi_i\right|\right|_{2,n} \leqslant O_p\left(\frac{\mu}{k_C}\left\{\sqrt{\frac{2slog(N)}{T}}\right\}\right) \tag{4.24}$$

Under a beta-min condition and neighborhood stability, I follow Meinshausen and Bühlmann (2006) to show variable selection consistency. In other words, under

$$\inf_{j\in\mathcal{A}_i} \left|\left|\pi_{ij}^0\right|\right| >> \sqrt{s\log(p)/NT}.$$

Then for a suitable $\lambda = \lambda_n >> \sqrt{\log(p)/NT}$,

$$\mathbb{P}\left[\hat{\mathcal{A}}_i = \mathcal{A}_i\right] \rightarrow 1, \quad \text{as} \quad N, T \rightarrow \infty.$$

From the previous result, the structural parameters can be estimated using the Post-Lasso estimation method in Belloni et al. (2012). As discuss in Belloni et al. (2012), the Post-Lasso alleviates the bias introduced by Lasso when selecting the parameters.

The post Lasso estimate estimates the structural parameters using the follow specification.

$$\tilde{y}_{i,t} = \omega_i^1 \, \tilde{x}_{i,t} + \sum_{j \in \hat{\mathcal{A}}_i} \omega_{i,j}^2 \tilde{x}_{j,t} + \sum_{j \in \hat{\mathcal{A}}_i} \omega_{i,j}^3 \left\{ \sum_{k \in \hat{\mathcal{A}}_j} \hat{\pi}_{jk} \tilde{x}_{k,t} \right\} + \tilde{\epsilon}_{i,t} \tag{4.25}$$

Inferring the set $\mathcal{A}_i$ required fairly strong assumptions. Alternatively, a less restrictive approach can be followed. In particular, the Lasso procedure can be used to recover a superset of the relevant covariates. In other words:

$$\mathbb{P}\left[\hat{\mathcal{A}}_i(\lambda) \supseteq \mathcal{A}_i\right] \to 1, \quad \text{as} \quad N, T \to \infty.$$

$$\limsup_{N,T \to \infty} \mathbb{P}\left[\hat{\mathcal{A}}_i(\lambda) = \mathcal{A}_i\right] < 1, \quad \text{as} \quad N, T \to \infty.$$

Identification of the structural parameters remain valid as depicted in Proposition 22, as long as condition (4.15) remains valid. Bühlmann and Van De Geer (2011) describes different approaches to achieve variable selection and control for the number of false positives. In other words, $\mid \hat{\mathcal{A}}_i \backslash \mathcal{A}_o \mid = O(\mid \mathcal{A}_o \mid)$, where $\mathcal{A}_o \subset \hat{\mathcal{A}}_i$ is the subset of coefficients that the oracle would select.

## 4.6 Conclusions

In this paper, I study a panel data model of strategic interactions. I develop an identification strategy that point identifies the endogenous and exogenous network effects when the topology of the network is unknown and unrestricted. This result relies on bounding the magnitude of the in-degree path connected effects of length 2. This restriction allows disentangling the strategic interactions effects that are direct from the ones generated through path connections. This result is used to approximate the equilibrium condition using high-dimensional approximately sparse panel models and to consistently estimate the network of interactions. Furthermore, I provide Oracle convergence rates for estimates of the strategic interaction parameters.

As future extensions, it would be interesting to develop inference theory for the vector of Lasso estimates (see Belloni et al. (2011)). Furthermore, it would be desirable to consider a more general model of peer effects. For example, a nonlinear model of strategic interactions beyond the linear panel data specification considered in this chapter. Finally, it would be interesting to propose a unified framework to model network formation models and strategic interactions simultaneously. By doing this, the methodology proposed could address peer effects and network endogeneity simultaneously.

## 4.7 Appendix A

*Proof of Proposition 20.* Consider the utility of individual $i$,

$$u_i = \left\{ \beta_i x_{i,t} + \alpha_i + \epsilon_{i,t} + \theta_i \sum_{\substack{j=1 \\ j \neq i}}^{N} a_{ij} x_{j,t} \right\} y_{i,t} - \frac{1}{2} y_{i,t}^2 - \frac{\gamma_i}{2} \left( y_{i,t} - \sum_{\substack{j=1 \\ j \neq i}}^{N} a_{ij} y_{j,t} \right)^2$$

let $\xi_{i,t} = \beta_i x_{i,t} + \alpha_i + \epsilon_{i,t} + \theta_i \sum_{\substack{j=1 \\ j \neq i}}^{N} a_{ij} x_{j,t}$. The first order condition of individual $i$ given the optimal strategy profile $y^*_{-i}$ is the following.

$$\frac{\partial \mathbb{E}\left[u_i | X_t, \alpha_i, \epsilon_{i,t}\right]}{\partial y_i} = \xi_{i,t} - y_{i,t}$$

$$-\gamma_i \left[ y_{i,t} - \sum_{\substack{j=1 \\ j \neq i}}^{N} a_{i,j} \mathbb{E}\left[ y_j^*(\xi_{j,t}, \alpha_j, \epsilon_{j,t}) | X_t, \alpha_i, \epsilon_{i,t} \right] \right] = 0$$

Hence,

$$y_{i,t} = \frac{\xi_{i,t}}{1+\gamma_i} + \frac{\gamma_i}{1+\gamma_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} a_{i,j} \mathbb{E}\left[ y_j^*(\xi_{j,t}, \alpha_j, \epsilon_{j,t}) | X_t, \alpha_i, \epsilon_{i,t} \right]$$

Which can be written as the following mapping,

$$(SY)_i(\xi_t, \alpha, \epsilon_t) = \frac{\xi_{i,t}}{1+\gamma_i} + \frac{\gamma_i}{1+\gamma_i} \sum_{\substack{j=1 \\ j \neq i}}^{N} a_{i,j} \mathbb{E}\left[ y_j^*(\xi_{j,t}, \alpha_j, \epsilon_{j,t}) | X_t, \alpha_i, \epsilon_{i,t} \right]$$

It will be shown that is a contraction mapping. Given that the operator satisfies the first order condition and given the Contraction Mapping Theorem, existence and uniqueness of the equilibrium is proved.

Consider any $i \in \mathcal{N}$ and fix $\xi_t, \alpha$ and $\epsilon$. Consider two strategy profiles $Y$ and $Y'$, with $Y \neq Y^{'}$. Hence,

$$\left\|(SY)_i(\xi_t, \alpha, \epsilon_t) - (SY^{'})_i(\xi_t, \alpha, \epsilon_t)\right\| =$$

$$\left\|\frac{\gamma_i}{1+\gamma_i}\sum_{\substack{j=1\\ j\neq i}}^{N} a_{ij}\left\{\mathbb{E}\left[y_j(\xi_{j,t}, \alpha_j, \epsilon_{j,t}) - y_j^{'}(\xi_{j,t}, \alpha_j, \epsilon_{j,t})|X_t, \alpha_i, \epsilon_{i,t}\right]\right\}\right\|$$

$$\leqslant \frac{\gamma_i}{1+\gamma_i}\sum_{\substack{j=1\\ j\neq i}}^{N} |a_{ij}|\left\{\mathbb{E}\left[\left\|y_j(\xi_{j,t}, \alpha_j, \epsilon_{j,t}) - y_j^{'}(\xi_{j,t}, \alpha_j, \epsilon_{j,t})\right\||X_t, \alpha_i, \epsilon_{i,t}\right]\right\}$$

$$\leqslant \frac{\gamma_i}{1+\gamma_i}\sum_{\substack{j=1\\ j\neq i}}^{N} |a_{ij}|\left\|y_j(\xi_{j,t}, \alpha_j, \epsilon_{j,t}) - y_j^{'}(\xi_{j,t}, \alpha_j, \epsilon_{j,t})\right\|$$

$$= \frac{\gamma_i}{1+\gamma_i}|A_i| \cdot \left\|Y(\xi_t, \alpha, \epsilon_t) - Y^{'}(\xi_t, \alpha, \epsilon_t)\right\|$$

Where the in the last equality, $A_i$ is the $i$th row of matrix A. The operator will be a contraction if

$$\frac{\gamma_i}{1+\gamma_i}|A_i| \leqslant \mathbf{1}$$

equivalently, for any $j \neq i$.

$$|a_{ij}| \leqslant \frac{1+\gamma_i}{\gamma_i}$$

which is guarantee by Assumption 1. Therefore the mapping is a contraction with coefficient $\max_{i\in\mathcal{N}} \frac{\gamma_i}{1+\gamma_i}$.

Finally aggregating the FOC for all individuals,

$$\left\|(SY)(\xi_t, \alpha, \epsilon_t) - (SY^{'})(\xi_t, \alpha, \epsilon_t)\right\| \leqslant$$

$$diag\left[\frac{\gamma_1}{1+\gamma_1}, \ldots, \frac{\gamma_N}{1+\gamma_N}\right]|A|\left\|Y(\xi_t, \alpha, \epsilon_t) - Y^{'}(\xi_t, \alpha, \epsilon_t)\right\|$$

$$\leqslant \left\{\max_{i\in\mathbb{N}} \frac{\gamma_i}{1+\gamma_i}\right\} \cdot \mathbf{I}_N \left\|Y(\xi_t, \alpha, \epsilon_t) - Y^{'}(\xi_t, \alpha, \epsilon_t)\right\|$$

Therefore the operator is a contraction and existence and uniqueness of the equilibrium is shown.

Finally, note the equilibrium strategy profile can be written as (abusing notation in the conditioning sigma-algebra),

$$Y^* = diag\left[\frac{1}{1+\gamma_i}, \cdots, \frac{1}{1+\gamma_N}\right]\xi_t + diag\left[\frac{\gamma_1}{1+\gamma_1}, \cdots, \frac{\gamma_N}{1+\gamma_N}\right]A\mathbb{E}\left[Y^*|X_t, \alpha, \epsilon_t\right]$$

where,

$$\xi_t = diag\left[\beta_1, \cdots, \beta_N\right]X_t + \alpha + \epsilon + diag\left[\theta_1, \cdots, \theta_N\right]AX_t$$

Then,

$$\begin{aligned} Y^* = & \left[\mathbf{I}_N - diag\left[\frac{\gamma_1}{1+\gamma_1}, \cdots, \frac{\gamma_N}{1+\gamma_N}\right]A\right]^{-1} \\ & \left\{diag\left[\frac{\beta_1}{1+\gamma_i}, \cdots, \frac{\beta_N}{1+\gamma_N}\right] + diag\left[\frac{\theta_1}{1+\gamma_1}, \cdots, \frac{\theta_N}{1+\gamma_N}\right]A\right\}X_t \\ & + diag\left[\frac{1}{1+\gamma_1}, \cdots, \frac{1}{1+\gamma_N}\right]\{\alpha + \epsilon_t\} + h(X_t, \alpha, \epsilon_t) \end{aligned}$$

where the $i$th element in $h(X_t, \alpha, \epsilon_t)$ is given by,

$$h(X_t, \alpha, \epsilon_t)_i = \sum_{j=0}^{N}\left[\left(diag\left[\frac{\gamma_1}{1+\gamma_1}, \cdots, \frac{\gamma_N}{1+\gamma_N}\right]A\right)^j\right]_{i\cdot}\left(\frac{1}{1+\gamma_i}\right)\mathbb{E}\left[\alpha + \epsilon_t|X_t, \alpha_i, \epsilon_{i,t}\right]$$

where for any matrix $W$, $[W]_{i\cdot}$ represents the $i$th row of the matrix. Using the same contraction mapping argument, we can show that $h(X_t, \alpha, \epsilon_t)$ is well-defined and the desired result has been achieved. ■

*Proof of Proposition 21.* Let $i \in \mathcal{N}$, note

$$\begin{aligned} \mathbb{E}_T\left[\psi_i(\tilde{X}_t)\right] \equiv & \frac{1}{T}\sum_{t=1}^{T}\left\{\psi_i(\tilde{X}_t)\right\} = \frac{1}{T}\sum_{t=1}^{T}\left\{\sum_{j=1}^{N}\pi_{i,j}^{(a)}\tilde{x}_{j,t}\right\} \\ & = \sum_{j=1}^{N}\pi_{i,j}^{(a)}\left\{\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{j,t}\right\} = \sum_{j=1}^{N}\pi_{i,j}^{(a)}\mathbb{E}_T\left[\tilde{x}_{j,t}\right] \end{aligned}$$

where $\mathbb{E}_T\left[\tilde{x}_{j,t}\right] = O_p(1)$, by LLN and Empirical Processes. Therefore, it is enough to prove

$$\sum_{j=1}^{N} \pi_{i,j}^{(a)} = O_p\left(\frac{s\log(N)}{T}\right)$$

Recall,

$$\Pi^{(a)} = \sum_{l=1}^{\infty} \Lambda_{\kappa,\theta}\Lambda_{\theta,\theta}^{l} A^{l+1}$$

Hence, the $i$th row of matrix $\Pi^{(a)}$ is given by

$$\Pi_{i}^{(a)} = \sum_{l=1}^{\infty} \left\{ \kappa_i \left(\frac{\theta_i}{1+\theta_i}\right)^{l} \right\} a_{i\cdot}^{(l+1)}$$

For any $l \in \mathbb{N}$, $a_{i\cdot}^{(l+1)}$ represents the $i$th row of the matrix of interactions, $A$, raised to the power of $(l+1)$, i.e. the $i$th row of $A^{l+1}$. Also, $\kappa_i$ is defined as $\kappa_i = \left\{\frac{\theta_i(1+\gamma_i)+\gamma_i\beta_i}{(1+\gamma_i)^2}\right\}$, which is a fixed constant.

Given Assumption A4, it follows from the Cauchy-Schwarz inequality that any characteristic element $(i,j)$th of the matrix $A^2$ is bounded.

$$a_{ij}^{(2)} = \left[\sum_{k=1}^{N} a_{ik}a_{kj}\right] \leqslant \left[\sum_{k=1}^{N} a_{ik}^2\right]^{1/2} \left[\sum_{k=1}^{N} a_{kj}^2\right]^{1/2} \leqslant O_p\left(\sqrt{\frac{s\log(N)}{T}}\right)$$

Subsequently, note that higher order effects can be bound recursively.

$$\begin{aligned} a_{ij}^{(3)} &= \left[\sum_{k=1}^{N} a_{ik}a_{kj}^{(2)}\right] \leqslant \left[\sum_{k=1}^{N} a_{ik}^2\right]^{1/2} \left[\sum_{k=1}^{N} \left(a_{kj}^{(2)}\right)^2\right]^{1/2} \\ &\leqslant O_p\left(\sqrt[4]{\frac{s\log(N)}{T}}\right) O_p\left(\sqrt{\frac{s\log(N)}{T}}\right) \\ &\leqslant O_p\left(\left\{\frac{s\log(N)}{T}\right\}^{3/4}\right) \end{aligned}$$

In general for any $k \geqslant 2$,

$$a_{ij}^{(k)} \leqslant O_p\left(\left\{\frac{s\log(N)}{T}\right\}^{k/4}\right)$$

Define

$$S_{a_i^{(2)}}(N) \equiv \left(\frac{\theta_i}{1+\theta_i}\right)\kappa_i\sum_{j=1}^{N} a_{ij}^{(2)} = O_p\left(\sqrt{\frac{s\log(N)}{T}}\right)$$

Equivalently, for any $k \geqslant 2$

$$S_{a_i^{(k)}}(N) \equiv \left(\frac{\theta_i}{1+\theta_i}\right)^{k-1}\kappa_i\sum_{j=1}^{N} a_{ij}^{(k)} \tag{4.26}$$

$$= \left(\frac{\theta_i}{1+\theta_i}\right)^{k-1} O_p\left(\left\{\frac{s\log(N)}{T}\right\}^{k/4}\right) \tag{4.27}$$

Hence,

$$\sum_{j=1}^{N}\pi_{ij}^{(a)} = \sum_{j=1}^{\infty} S_{a_i^{(j+1)}}(N) = \sum_{j=1}^{\infty}\left(\frac{\gamma_i}{1+\gamma_i}\right)^{j} S_{a_i^{(j+1)}}(N)$$

where for each $j$, $S_{a_i^{(j+1)}}(N)$ is bounded and given $\left(\frac{\gamma_i}{1+\gamma_i}\right)^j$ converges to zero the result follows. ∎

*Proof of Proposition 22.* The proof of proposition 22 consists of two steps.

**I.** Prove identification of the composite Parameters.

Differentiate out the fixed effects.

$$\tilde{y}_{i,t} = \frac{\beta_i}{1+\theta_i}\tilde{x}_{i,t} + \frac{\rho_i}{1+\theta_i}\sum_{\substack{j=1\\ j\neq i}}^{N} a_{ij}\tilde{x}_{j,t} + \frac{\theta_i}{1+\theta_i}\sum_{\substack{j=1\\ j\neq i}}^{N} a_{ij}\mathbb{E}\left[\tilde{y}_{j,t}|X_1,\cdots,X_T\right] + \tilde{v}_{i,t}$$

$$\tilde{y}_{i,t} = \omega_i^1\tilde{x}_{i,t} + \sum_{\substack{j=1\\ j\neq i}}^{N}\omega_{i,j}^2\tilde{x}_{j,t} + \sum_{\substack{j=1\\ j\neq i}}^{N}\omega_{i,j}^3\mathbb{E}\left[\tilde{y}_{j,t}|X_1,\cdots,X_T\right] + \tilde{\epsilon}_{i,t} \tag{4.28}$$

$$\tilde{y}_{i,t} = \omega_i^1\,\tilde{x}_{i,t} + \sum_{j\in\mathcal{A}_i}\omega_{i,j}^2\tilde{x}_{j,t} + \sum_{j\in\mathcal{A}_i}\omega_{i,j}^3\left\{\sum_{k\in\mathcal{A}_j}\pi_{jk}\tilde{x}_{k,t}\right\} + \tilde{\epsilon}_{i,t} \tag{4.29}$$

Identification of the composite parameters $[\omega_i^1, \{\omega_{i,j}^2, \omega_{i,j}^3\}]$ is achieved if:

For any $i \in \mathcal{N}$, $||\mathcal{A}_i||_0 \leqslant s$ and

$$\mathcal{A}_j \nsubseteq A_i \cup \{i\}$$

The main result of the paper proves that we can recover $\mathcal{A}_i$ for any $i \in \mathcal{N}$ with high accuracy as $N, T \to \infty$.

**II.** Second, we recover the structural parameters: $\left[\beta_i, \theta_i, \rho_i, \{a_{ij}\}_{j \neq i}\right]$.

1. Identification of $\theta_i$: $\sum_{j \neq i} \omega_{ij}^3 = \frac{\theta_i}{1+\theta_i} \sum_{j \neq i} a_{ij} = \frac{\theta_i}{1+\theta_i}$

   Hence: $\theta_i = \frac{\sum_{j \neq i} \omega_{ij}^4}{1 - \sum_{j \neq i} \omega_{ij}^3}$

2. Identification of $\{a_{ij}\}_{j \neq i}$: $\forall j \neq i$: $a_{ij} = \frac{1+\theta_i}{\theta_i} \omega_{ij}^3$.

3. Identification of $\beta i$: $\alpha_i = (1 + \gamma_i)\omega_i^1$

4. Identification of $\rho_i$: $\theta_i = w_{ij}^2(1 + \gamma_i)/a_{ij}$.

■

*Proof.* Proof Proposition 23 Consider the Lasso Estimate $\hat{\Pi}^L$ and denote by $\hat{\Pi}_i$ its $i$th column. Then by optimality of $\hat{\Pi}_i$.

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\tilde{y}_{i,t} - \tilde{X}_t' \hat{\Pi}_i\right]^2 - \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\tilde{y}_{i,t} - \tilde{X}_t' \Pi_i\right]^2$$

$$\leqslant \frac{\lambda}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\omega}_{ij} |\pi_{ij}| - \frac{\lambda}{NT} \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\omega}_{ij} |\hat{\pi}_{ij}|$$

Expanding the quadratic form.

$$\left[\tilde{y}_{i,t} - \tilde{X}_t' \hat{\Pi}_i\right]^2 = \tilde{y}_{i,t}^2 + \left(\tilde{X}_t' \hat{\Pi}_i\right)^2 - 2\tilde{y}_{i,t} \tilde{X}_t' \hat{\Pi}_i$$

Let $\delta_i = \hat{\Pi}_i - \Pi_i$, we can write the left-hand side of the inequality as.

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\tilde{X}_t'\hat{\Pi}_i\right)^2 - \left(\tilde{X}_t'\Pi_i\right)^2 - 2\tilde{y}_{i,t}\tilde{X}_t'\delta_i\right\}$$

Using the specification: $\tilde{y}_{i,t} = \tilde{X}_t'\hat{\Pi}_i + \Psi_i(\tilde{X}_t) + \tilde{v}_{i,t}$.

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\tilde{X}_t'\delta_i\right)^2\right\} + 2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\tilde{v}_{i,t}\tilde{X}_t'\delta_i\right\} + 2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\Psi_i(\tilde{X}_t)\tilde{X}_t'\delta_i\right\}$$

By Hölder's Inequality.

$$\left|2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\tilde{v}_{i,t}\tilde{X}_t'\delta_i\right\}\right| \leqslant \max_{1\leqslant i\leqslant N}||S_i||_\infty \frac{1}{N}\sum_{i=1}^{N}||\omega_i\delta_i||_1$$

$$\left|2\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\Psi_i(\tilde{X}_t)\tilde{X}_t'\delta_i\right\}\right| \leqslant c_s\frac{1}{N}\sum_{i=1}^{N}\left|\left|\tilde{X}_t'\delta_i\right|\right|_2$$

where $S_{ij} = \frac{1}{T}\sum_{t=1}^{T}(\hat{\omega}_{ij})^{-1}[\tilde{v}_{it}\tilde{x}_{jt}]$ and $c_s = \left[\frac{1}{TN}\sum_{i=1}^{N}\sum_{t=1}^{T}\Psi_i(\tilde{X}_t)^2\right]^{1/2}$. Then:

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\tilde{X}_t'\delta_i\right)^2\right\} \leqslant \frac{\lambda}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\{\hat{\omega}_{ij}|\pi_{ij}| - \hat{\omega}_{ij}|\hat{\pi}_{ij}|\}$$

$$+ \max_{1\leqslant i\leqslant N}||S_i||_\infty \frac{1}{N}\sum_{i=1}^{N}||\omega_i\delta_i||_1 + c_s\frac{1}{N}\sum_{i=1}^{N}\left|\left|\tilde{X}_t'\delta_i\right|\right|_2$$

By decomposability of the penalization function.

$$\frac{\lambda}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N}\{\hat{\omega}_{ij}|\pi_{ij}| - \hat{\omega}_{ij}|\hat{\pi}_{ij}|\} =$$

$$\frac{\lambda}{NT}\sum_{i=1}^{N}\left\{\sum_{j\in Supp(\Pi_i^0)}\{\hat{\omega}_{ij}|\delta_{ij}|\} - \sum_{j\in Supp^C(\Pi_i^0)}\{\hat{\omega}_{ij}|\delta_{ij}|\}\right\}$$

Using the appropriate choice of $\lambda$ such that: $\frac{\lambda}{T} \geqslant K\max_{1\leqslant j\leqslant N}||S_j||_\infty$.

$$\max_{1\leqslant i\leqslant N}||S_i||_\infty \frac{1}{N}\sum_{i=1}^{N}||\omega_i\delta_i||_1 \leqslant \frac{\lambda}{KTN}\sum_{i=1}^{N}\left|\left|\hat{\omega}_i^0\delta_i\right|\right|_1$$

Aggregating terms and using the condition: $l\hat{w}_i \leqslant \hat{w}_i^0 \leqslant u\hat{w}_i$

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\tilde{X}_t'\delta_i\right)^2\right\} \leqslant c_s\frac{1}{N}\sum_{i=1}^{N}\left|\left|\tilde{X}_t'\delta_i\right|\right|_2$$

$$\frac{\lambda}{NT}\sum_{i=1}^{N}\left\{\left(u+\frac{1}{K}\right)\left|\left|\hat{\omega}_i^0\delta_i\right|\right|_{T,1} - \left(l-\frac{1}{K}\right)\left|\left|\hat{\omega}_i^0\delta_i\right|\right|_{T^C,1}\right\}$$

Given that,

$$c_s(i) \leqslant \frac{1}{N}\sum_{i=1}^{N}\left|\left|\tilde{X}_t'\delta_i\right|\right|_2$$

The previous equation implies:

$$\frac{\lambda}{NT}\sum_{i=1}^{N}\left\{\left|\left|\hat{\omega}_i^0\delta_i\right|\right|_{T^C,1}\right\} \leqslant \frac{\lambda}{NT}\bar{C}_0\sum_{i=1}^{N}\left\{\left|\left|\hat{\omega}_i^0\delta_i\right|\right|_{T,1}\right\}$$

Using the definition of the restricted eigenvalue: $\kappa_{\bar{C}_0} = \min_{\delta\in\Delta}\frac{\sqrt{s}\left|\left|\tilde{X}_t'\delta_i\right|\right|_2}{\left|\left|\hat{\omega}_i^0\delta_i\right|\right|_{T,1}}$, we obtain the following inequality:

$$\frac{\lambda}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\tilde{X}_t'\delta_i\right)^2\right\} \leqslant \frac{\lambda\sqrt{s}}{NT\kappa_{\bar{C}_0}}\left(u+\frac{1}{K}\right)\sum_{i=1}^{N}\left\{\left|\left|\tilde{X}_t'\delta_i\right|\right|_2\right\} + c_s\frac{1}{N}\sum_{i=1}^{N}\left|\left|\tilde{X}_t'\delta_i\right|\right|_2$$

which implies:

$$\frac{\lambda}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\tilde{X}_t'\delta_i\right)^2\right\} \leqslant \left[\frac{\lambda\sqrt{s}}{NT\kappa_{\bar{C}_0}}\left(u+\frac{1}{K}\right)+c_s\right]^2$$

It follows,

$$\frac{\lambda}{NT}\sum_{i=1}^{N}\left\{\left|\left|\hat{\omega}_i^0\delta_i\right|\right|_{T^C,1}\right\} \leqslant \frac{\lambda}{NT}\bar{C}_0\sum_{i=1}^{N}\left\{\left|\left|\hat{\omega}_i^0\delta_i\right|\right|_{T,1}\right\}$$

which by definition of $\kappa_{\bar{C}_0}$

$$\frac{1}{NT}\sum_{i=1}^{N}\{\left|\left|\hat{\omega}_i^0\delta_i\right|\right|_1\} \leqslant \frac{\bar{C}_0\sqrt{s}}{\kappa_{\bar{C}_0}}\left[\frac{\lambda\sqrt{s}}{NT\kappa_{\bar{C}_0}}\left(u+\frac{1}{K}\right)+c_s\right]^2 + \frac{\bar{C}_0}{\lambda}c_s^2$$

By choosing, $\lambda = O(\sqrt{NT\log(N)})$ and $\sqrt{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Psi_i(\tilde{X}_t)^2} = O_p(\sqrt{\frac{s}{NT}})$, we obtain.

$$\frac{\lambda}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\tilde{X}_t'\hat{\Pi}_i - \tilde{X}_t'\Pi_i\right)^2\right\} = O_p\left(\frac{1}{\kappa^2_{\bar{C}_0}}\frac{s\log(N)}{NT} + \frac{s}{nT}\right)$$

$$\frac{\lambda}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left(\tilde{y}_{it} - \tilde{X}_t'\hat{\Pi}_i\right)^2\right\} = O_p\left(\frac{1}{\kappa^2_{\bar{C}_0}}\frac{s\log(N)}{NT}\right)$$

$$\frac{1}{N}\sum_{i=1}^{N}\left\{\left|\left|\hat{\omega}_i^0\delta_i\right|\right|_1\right\} \leqslant O_p\left(\frac{1}{\kappa^2_{\bar{C}_0}}\sqrt{\frac{s^2\log(N)}{NT}}\right)$$

■

**Proposition 24.** *Let assumptions 1-6 in Meinshausen and Bühlmann (2006) hold. Then*

$$\mathbb{P}\left[\hat{\mathcal{A}}_i = \mathcal{A}_i\right] \to 1,$$

*as* $T, N \to \infty$.

*Proof.* The proof follows from Theorem 1 and 2 in Meinshausen and Bühlmann (2006). ■

# Bibliography

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

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

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

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

Arellano, M. and Honoré, B. (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*.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013), "The diffusion of microfinance," *Science*, 341, 1236498.

Bayer, P., Hjalmarsson, R., and Pozen, D. (2009), "Building Criminal Capital Behind Bars: Peer Effects in Juvenile Corrections," *Quarterly Journal of Economics*, 124.

Belloni, A., Chernozhukov, V., and Hansen, C. (2011), "Inference for high-dimensional sparse econometric models," *arXiv preprint arXiv:1201.0220*.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012), "Sparse models and methods for optimal instruments with an application to eminent domain," *Econometrica*, 80, 2369–2429.

Belloni, A., Chernozhukov, V., et al. (2013), "Least squares after model selection in high-dimensional sparse models," *Bernoulli*, 19, 521–547.

Blume, L. E., Brock, W. A., Durlauf, S. N., and Jayaraman, R. (2013), "Linear social interactions models," Tech. rep., National Bureau of Economic Research.

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

Bramoullé, Y. and Kranton, R. (2007), "Public goods in networks," *Journal of Economic Theory*, 135, 478–494.

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

Brock, W. A. and Durlauf, S. N. (2002), "A multinomial-choice model of neighborhood effects," *American Economic Review*, pp. 298–303.

Brock, W. A. and Durlauf, S. N. (2005), "Multinomial Choice with Social Interactions," *The Economy As an Evolving Complex System, III: Current Perspectives and Future Directions*, p. 175.

Bühlmann, P. and Van De Geer, S. (2011), *Statistics for high-dimensional data: methods, theory and applications*, Springer Science & Business Media.

Calvo-Armengol, A. and Jackson, M. O. (2004), "The effects of social networks on employment and inequality," *The American Economic Review*, 94, 426–454.

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

Campbell, A. (2013), "Word-of-mouth communication and percolation in social networks," *The American Economic Review*, 103, 2466–2498.

Candelaria, L. E. (2016), "Network Formation Models with Interactive Fixed Effects," *Working Paper.*

Candes, E. and Tao, T. (2007), "The Dantzig selector: Statistical estimation when p is much larger than n," *The Annals of Statistics*, pp. 2313–2351.

Caner, M., Han, X., and Lee, Y. (2013), "Adaptive Elastic Net GMM Estimator with Many Invalid Moment Conditions: A Simultaneous Model and Moment Selection," .

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

Chandrasekhar, A. and Lewis, R. (2014), "Econometrics of Sampled Networks," *Working Paper.*

Chandrasekhar, A. G. and Jackson, M. O. (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 Siow, A. (2006), "Who marries whom and why," *Journal of Political Economy*, 114, 175–201.

Christakis, N. A. and Fowler, J. H. (2007), "The spread of obesity in a large social network over 32 years," *New England Journal of Medicine*, 357, 370–379.

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

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

Cohen-Cole, E. and Fletcher, J. M. (2008), "Is obesity contagious? Social networks vs. environmental factors in the obesity epidemic," *Journal of Health Economics*, 27, 1382–1387.

De Giorgi, G., Pellizzari, M., and Redaelli, S. (2010), "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics*, pp. 241–275.

de Paula, A., Richards-Shubik, S., and Tamer, E. T. (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*.

Fafchamps, M. and Gubert, F. (2007), "The formation of risk sharing networks," *Journal of development Economics*, 83, 326–350.

Fan, J. and Liao, Y. (2011), "Ultra high dimensional variable selection with endogenous covariates," *Manuscript. Princeton University*.

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

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

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

Gautier, E. and Tsybakov, A. B. (2011), "High-dimensional instrumental variables regression and confidence sets," *arXiv preprint arXiv:1105.2454*.

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

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

Graham, B. S. (2015), "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, 303–316.

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

Honoré, B. E. and Hu, L. (2015), "Poor (Wo)man's Bootstrap," *Working Paper*.

Horowitz, J. and Huang, J. (2012), "Penalized estimation of high-dimensional models under a generalized sparsity condition," .

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

Hsieh, C.-S. and Lee, L.-F. (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. (2010), *Social and economic networks*, Princeton University Press.

Jackson, M. O. and Watts, A. (2002), "The evolution of social and economic networks," *Journal of Economic Theory*, 106, 265–295.

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

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

Kinnan, C. and Townsend, R. (2012), "Kinship and financial networks, formal financial access, and risk reduction," *The American Economic Review*, 102, 289–293.

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

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

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

Leung, M. (2015b), "A Weak Law for Moments of Pairwise-Stable Networks," *Working Paper*.

Manresa, E. (2013), "Estimating the Structure of Social Interactions Using Panel Data," .

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

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

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

Manski, C. F. (1993), "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 60, 531–542.

Masten, M. A. (2015), "Random Coefficients on Endogenous Variables in Simultaneous Equations Models," .

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

Meinshausen, N. and Bühlmann, P. (2006), "High-dimensional graphs and variable selection with the lasso," *The annals of statistics*, pp. 1436–1462.

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, 99–135.

Newey, W. K. and McFadden, D. (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.

Sacerdote, B. (2001), "Peer Effects with Random Assignment: Results for Darthmouth Roomates," *The Quarterly Journal of Economics*, p. 681.

Serfling, R. J. (2009), *Approximation theorems of mathematical statistics*, vol. 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*, pp. 123–137.

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

Snijders, T. A., Koskinen, J., and Schweinberger, M. (2010), "Maximum likelihood estimation for social network dynamics," *The Annals of Applied Statistics*, 4, 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, 147–165.

Tibshirani, R. (1996), "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.

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

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

Xu, H. (2010), "Social Interactions in Large Networks: A Game Theoretic Approach," *Working Paper*.

# Biography

Luis Enrique Candelaria Barrera was born on March 10, 1986, in Mexico City. He is a Doctor of Philosophy Candidate in Economics at Duke University. He holds a B.A. in Economics from Instituto Tecnologico y de Estudios Superiores de Monterrey, Mexico in 2008, an M.A. in Economic Theory from Instituto Tecnologico Autonomo de Mexico, Mexico in 2010, and an M.A. in Economics from Duke University in 2012. He will start his academic career as an Assistant Professor of Economics at the University of Warwick, England.