Simultaneous Equations with Binary Outcomes and Social Interactions*

Xiaodong Liu

Department of Economics

University of Colorado Boulder

Boulder, CO 80309, USA

Email: xiaodong.liu@colorado.edu

March, 2018

Abstract

This paper introduces a discrete-choice simultaneous-equation social interaction model. We provide a microfoundation for the econometric model by considering an incomplete information game where individuals interact in multiple activities through a network. We characterize the sufficient condition for the existence of a unique BNE of the game. We discuss the identification of the econometric model and propose a two-stage estimation procedure, where the reduced form parameters are estimated by

*I would like to thank the associate editor, two anonymous referees, Lung-fei Lee, Zhongjian Lin, Haiqing Xu, and Chao Yang for valuable comments and suggestions. I gratefully acknowledge financial support from the CARTSS at University of Colorado Boulder. 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.

the NPL algorithm in the first stage and the structural parameters are recovered from the estimated reduced form parameters by the AGLS estimator in the second stage. Monte Carlo experiments show that the proposed estimation procedure performs well in finite samples and remains computationally feasible when networks are large. We also provide an empirical example to illustrate the empirical relevance of the proposed model.

JEL classification: C31, C35

Key words: discrete choices, networks, rational expectations, simultaneous equations

1 Introduction

In everyday life, people make interrelated choices in various activities, many of which are influenced by the choices of other people. This paper proposes a simultaneous-equation social interaction model with binary outcomes to characterize the interdependence of individual choices both across activities and across individuals.

In their seminal papers, Brock and Durlauf (2001a, 2001b) introduced a discrete-choice *group interaction* model, where individuals are partitioned into groups and an individual's choice depends on his/her expectation of the average choice of the group he/she belongs to.[1] In this model, the expectation is formed based on group-level information. As beliefs are rational, the subjective expectation on the average choice of the group is identical for every individual in the group, and coincides with the mathematical expectation. Brock and Durlauf (2007) treated the *homogenous* rational expectation as a data moment and discussed nonparametric identification of the model.

More recently, discrete-choice *network interaction* models, where an individual's choice

[1]A review of the recent development of social interaction models can be found in Blume et al. (2011).

is influenced by the expected outcomes of his/her direct connections (or peers) in the network, have attracted great interest. As every individual in the network usually has a different set of peers and the expectation is formed based on individual-level information, the rational expectation is *heterogenous* in this model. Lee et al. (2014), Lin and Xu (2017), Yang and Lee (2017) and Xu (2018) have studied different variations of this model.

The aforementioned works focus on social interactions within a certain activity. This paper generalizes the single-activity social interaction model with discrete choices to a simultaneous-equation model. To motivate the specification of the econometric model, we consider an incomplete information game where individuals interact in multiple activities through a network. We characterize the sufficient condition for the existence of a unique Bayesian Nash Equilibrium (BNE) of the game, which in turn guarantees the coherency and completeness of the econometric model (Tamer, 2003). We discuss the identification of the econometric model and propose a two-stage estimation procedure, where the reduced form parameters are estimated by the nested pseudo likelihood (NPL) algorithm (Aguirregabiria and Mira, 2007) in the first stage and the structural parameters are recovered from the estimated reduced form parameters by the Amemiya generalized least squares (AGLS) estimator (Amemiya, 1978) in the second stage. Monte Carlo experiments show that the proposed estimation procedure performs well in finite samples and remains computationally feasible when networks are large. Finally, to illustrate the empirical relevance of the proposed model, we apply it to study peer effects of participating in sports activities on underage smoking, using a representative sample of U.S. teenagers in the National Longitudinal Study of Adolescent Health (Add Health) data.

The rest of the paper is organized as follows. Sections 2 introduces an incomplete information network game with multiple activities. Section 3 presents the econometric model implied by the network game and discusses the identification and estimation of the model. Section 4 gives Monte Carlo simulation results on the finite sample performance of the proposed estimation procedure. Section 5 provides an empirical example. Section 6 briefly concludes. The proofs are collected in the appendix.

Throughout the paper we adopt the following notation. Let $\Phi(\cdot)$ denote the standard normal distribution function with the density function $\phi(\cdot)$. For an $n \times m$ matrix $\mathbf{A} = [a_{ij}]$, the vectorization of $\mathbf{A}$ is given by $\text{vec}(\mathbf{A}) = (a_{11}, \cdots, a_{n1}, a_{12}, \cdots, a_{nm})'$,[2] and the row sum and column sum matrix norms of $\mathbf{A}$ are denoted by $||\mathbf{A}||_\infty = \max_{i=1,\cdots,n} \sum_{j=1}^{m} |a_{ij}|$ and $||\mathbf{A}||_1 = \max_{j=1,\cdots,m} \sum_{i=1}^{n} |a_{ij}|$ respectively.

2 Incomplete Information Network Game

To provide a microfoundation for the proposed econometric model, we consider an incomplete information network game. Suppose a set of individuals $\mathcal{N} = \{1, \cdots, n\}$ interacts within a network. Let $\mathbf{W} = [w_{ij}]$ be an $n \times n$ predetermined adjacency matrix, where the (i, j)-th element w_{ij} is a known nonnegative constant that captures the proximity of individuals i and j in the network. As a normalization, $w_{ii} = 0$ for all i. We define the *peers* of individual i as the set of individuals $\mathcal{N}_i = \{j : w_{ij} > 0\}$.

The individuals in the network participate in m activities. Let $I(\cdot)$ be an indicator function that equals one if its argument is true and zero otherwise. The observed binary choice/outcome d_{ik} of individual i in activity k depends on a continuous *latent* variable y_{ik} such that $d_{ik} = I(y_{ik} > 0)$. We interpret y_{ik} as the underlying intention and d_{ik} as the actual choice/outcome (Maddala, 1983). Thus, it's natural to assume that individual i knows his/her own intentions $y_{i1}, \cdots, y_{im}$ but not those of his/her peers. The utility of

[2] If $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ are conformable matrices, then $\text{vec}(\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec}(\mathbf{B})$, where $\otimes$ denotes the Kronecker product.

individual i is a linear-quadratic function of $y_{i1}, \cdots, y_{im}$:[3]

$$\mathcal{U}_i = \underbrace{\sum_{k=1}^m (\sum_{l=1}^m \varrho_{lk} \sum_{j=1}^n w_{ij} d_{jl} + \varpi_{ik} - \varepsilon_{ik}) y_{ik}}_{\text{payoff}} - \underbrace{\frac{1}{2} \sum_{k=1}^m \sum_{l=1}^m \vartheta_{lk} y_{ik} y_{il}}_{\text{cost}}, \tag{2.1}$$

with $\vartheta_{kl} = \vartheta_{lk}$ and $\vartheta_{kk} \neq 0$ for all k and l. The utility has a payoff/cost structure. The cost is a quadratic function of $y_{i1}, \cdots, y_{im}$, where ϑ_{lk} $(k \neq l)$ represents the complementarity or substitutability (depending on the sign of ϑ_{lk}) between y_{ik} and y_{il}. The marginal payoff of y_{ik} depends on his/her peers' *actual choices/outcomes* $d_{j1}, \cdots, d_{jm}$ for $j \in \mathcal{N}_i$ and his/her own productivity $\varpi_{ik} - \varepsilon_{ik}$. The coefficient ϱ_{lk} represents the spillover effect of the peers' choices/outcomes in activity l on the marginal payoff of y_{ik}. The productivity of individual i in activity k has two components, where ϖ_{ik} is commonly known by all individuals in the network and ε_{ik} is an idiosyncratic shock that is privately observed by individual i. We assume that ε_{ik} is independently distributed over i according to some commonly known distribution and ε_{ik} is independent of $\{\varpi_{1l}, \cdots, \varpi_{nl}\}_{l=1,\cdots,m}$.

The utility (2.1) differs from that in Cohen-Cole et al. (2018) in two ways. First, in (2.1), y_{ik} represents the unobservable intention and the utility of individual i depends on the actual choices/outcomes of the peers given by the signs of $y_{j1}, \cdots, y_{jm}$ for $j \in \mathcal{N}_i$; while, in Cohen-Cole et al. (2018), y_{ik} represents the observable action/outcome in a continuous action/outcome space and the utility of individual i depends directly on $y_{j1}, \cdots, y_{jm}$ for $j \in \mathcal{N}_i$. Second, (2.1) has an privately observed random shock ε_{ik}, while Cohen-Cole et al. (2018) considered a complete information network game. As a result, maximization of the utility (2.1) motivates a *discrete-choice* simultaneous-equation econometric model of social interactions, while the network game in Cohen-Cole et al. (2018) leads to a *linear* simultaneous-equation econometric model of social interactions.

[3] The linear-quadratic specification of the utility function is common for network games (see, e.g., Ballester et al., 2006; Blume et al., 2015).

Given the network topology and the observable components of the productivity, the individuals simultaneously choose y_{ik}, for $k = 1, \cdots, m$, to maximize their *expected* utilities[4]

$$\begin{aligned}
&\mathbb{E}(\mathcal{U}_i | \{\varpi_{1k}, \cdots, \varpi_{nk}, \varepsilon_{ik}\}_{k=1,\cdots,m}) \\
= \quad & \textstyle\sum_{k=1}^m (\sum_{l=1}^m \varrho_{lk} \sum_{j=1}^n w_{ij} p_{jl} + \varpi_{ik} - \varepsilon_{ik}) y_{ik} - \frac{1}{2} \sum_{k=1}^m \sum_{l=1}^m \vartheta_{lk} y_{ik} y_{il}.
\end{aligned}$$

where $p_{jl} = \mathbb{E}(d_{jl} | \{\varpi_{1k}, \cdots, \varpi_{nk}\}_{k=1,\cdots,m})$. From the first order condition of utility maximization, we have

$$\textstyle\sum_{l=1}^m \theta_{lk} y_{il} = \sum_{l=1}^m \lambda_{lk} \sum_{j=1}^n w_{ij} p_{jl} + \pi_{ik} - \epsilon_{ik}, \tag{2.2}$$

where $\theta_{lk} = \vartheta_{lk}/\vartheta_{kk}$, $\lambda_{lk} = \varrho_{lk}/\vartheta_{kk}$, $\pi_{ik} = \varpi_{ik}/\vartheta_{kk}$, and $\epsilon_{ik} = \varepsilon_{ik}/\vartheta_{kk}$.[5] In matrix form, (2.2) can be written as

$$\textstyle\sum_{l=1}^m \theta_{lk} \mathbf{y}_l = \sum_{l=1}^m \lambda_{lk} \mathbf{W} \mathbf{p}_l + \boldsymbol{\pi}_k - \boldsymbol{\epsilon}_k, \tag{2.3}$$

where $\mathbf{y}_l = (y_{1l}, \cdots, y_{nl})'$, $\mathbf{p}_l = (p_{1l}, \cdots, p_{nl})'$, $\boldsymbol{\pi}_k = (\pi_{1k}, \cdots, \pi_{nk})'$, and $\boldsymbol{\epsilon}_k = (\epsilon_{1k}, \cdots, \epsilon_{nk})'$.

Let $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_m]$, $\mathbf{P} = [\mathbf{p}_1, \cdots, \mathbf{p}_m]$, $\boldsymbol{\Pi} = [\boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_m]$, and $\mathbf{E} = [\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_m]$. For all the m activities, it follows from (2.3) that

$$\mathbf{Y}\boldsymbol{\Theta} = \mathbf{W}\mathbf{P}\boldsymbol{\Lambda} + \boldsymbol{\Pi} - \mathbf{E}, \tag{2.4}$$

where $\boldsymbol{\Theta} = [\theta_{lk}]$ and $\boldsymbol{\Lambda} = [\lambda_{lk}]$ are $m \times m$ parameter matrices.[6] The off-diagonal element

[4]In this paper, we consider a static network game with a predetermined network. Although not the focus of this paper, it is worth pointing out rapid progress has recently been made on the modeling, identification and estimation of the network formation process (see, e.g., Goldsmith-Pinkham and Imbens, 2013; Graham, 2015; Mele, 2017; Hsieh and Lee, 2017).

[5]The parameters in (2.1) are identifiable up to a proportionality factor (Maddala, 1983). As we show in the next section, the scale-normalized parameters in (2.2) can be identified.

[6]$\boldsymbol{\Theta}$ has unitary diagonal elements by construction.

of $\mathbf{\Theta}$, θ_{lk}, represents *the simultaneity effect*, i.e., an individual's choice in an activity k may depend on his/her own choice in a related activity l. The diagonal element of $\mathbf{\Lambda}$, λ_{kk}, represents *the within-activity peer effect*, where an agent's choice in an activity k may depend on the expected choices/outcomes of the peers in the same activity. The off-diagonal element of $\mathbf{\Lambda}$, λ_{lk}, represents *the cross-activity peer effect*, where an agent's choice in an activity k may depend on the expected choices/outcomes of the peers in a related activity l.

If $\mathbf{\Theta}$ is nonsingular, then the reduced form of model (2.4) is

$$\mathbf{Y} = \mathbf{W}\mathbf{P}\mathbf{\Lambda}^* + \mathbf{\Pi}^* - \mathbf{E}^*, \tag{2.5}$$

where $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{\Theta}^{-1}$, $\mathbf{\Pi}^* = \mathbf{\Pi}\mathbf{\Theta}^{-1}$, and $\mathbf{E}^* = \mathbf{E}\mathbf{\Theta}^{-1}$. From (2.5),

$$y_{ik} = \sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl} + \pi_{ik}^* - \epsilon_{ik}^*, \tag{2.6}$$

where λ_{lk}^* is the (l,k)-th element of $\mathbf{\Lambda}^*$, and π_{ik}^* and ϵ_{ik}^* are the (i,k)-th elements of $\mathbf{\Pi}^*$ and $\mathbf{E}^*$ respectively. Then,

$$\Pr(d_{ik} = 1 | \{\varpi_{1k}, \cdots, \varpi_{nk}\}_{k=1,\cdots,m}) = F_k(\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl} + \pi_{ik}^*),$$

where $F_k(\cdot)$ is the distribution function of ϵ_{ik}^*.

Let $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$, $\mathbf{p} = \mathrm{vec}(\mathbf{P})$, $\boldsymbol{\pi}^* = \mathrm{vec}(\mathbf{\Pi}^*)$, and $\boldsymbol{\epsilon}^* = \mathrm{vec}(\mathbf{E}^*)$. Let

$$\vec{h}(\mathbf{p}) = [\vec{h}_1(\mathbf{p})', \cdots, \vec{h}_m(\mathbf{p})']',$$

where $\vec{h}_k(\mathbf{p}) = [F_k(u_{1k}), \cdots, F_k(u_{nk})]'$ with $u_{ik} = \sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl} + \pi_{ik}^*$. In the BNE, $\mathbf{p} = \vec{h}(\mathbf{p})$ (Osborne and Rubinstein, 1994). A sufficient condition for the existence of a unique solution of $\mathbf{p} = \vec{h}(\mathbf{p})$ is given as follows.

Assumption 1. (i) *$F_k(\cdot)$ is a continuous distribution function with a density function denoted by $f_k(\cdot)$.* (ii) *$\boldsymbol{\Theta}$ is nonsingular, and either $||\boldsymbol{\Lambda}^*||_1 < [||\mathbf{W}||_\infty \max_k \sup_u f_k(u)]^{-1}$ or $||\boldsymbol{\Lambda}^*||_\infty < [||\mathbf{W}||_1 \max_k \sup_u f_k(u)]^{-1}$.*

Assumption 1 (ii) suggests that, for the equilibrium to be unique, the social interaction effects cannot be too strong. If ϵ^*_{ik} follows the standard normal distribution, then $\max_k \sup_u f_k(u) = 1/\sqrt{2\pi}$. Furthermore, in some empirical studies of social networks, it may be reasonable to have $\mathbf{W}$ row-normalized (see, e.g., Lin, 2010; Boucher et al., 2014). When $\mathbf{W}$ is row-normalized, $\sum_{j=1}^m w_{ij} = 1$ for all i and hence $||\mathbf{W}||_\infty = 1$. In this case, Assumption 1 (ii) holds if $\max_{k=1,\cdots,m} \sum_{l=1}^m |\lambda^*_{lk}| < \sqrt{2\pi}$. It is worth pointing out that, when $m = 1$, Assumption 1 (ii) coincides with the sufficient condition for the existence of a unique rational expectation equilibrium for the single-activity social interaction model in Lee et al. (2014).

Under Assumption 1, the following proposition establishes the existence and uniqueness of the pure strategy BNE of this incomplete information network game by the contraction mapping theorem.

Proposition 2.1. *If Assumption 1 holds, then the incomplete information network game with the utility (2.1) has a unique pure strategy BNE with the equilibrium strategy profile $\mathbf{y}^\star$ given by*

$$\mathbf{y}^\star = (\boldsymbol{\Lambda}^{*\prime} \otimes \mathbf{W})\mathbf{p}^\star + \boldsymbol{\pi}^* - \boldsymbol{\epsilon}^*,$$

where the vector of equilibrium beliefs $\mathbf{p}^\star$ is the unique solution of

$$\mathbf{p} = \vec{h}(\mathbf{p}). \tag{2.7}$$

When Assumption 1 holds, the contraction mapping property of $\vec{h}(\mathbf{p})$ not only guarantees the coherency and completeness of the model (Tamer, 2003), but also suggests the NPL algorithm, based on the fixed point mapping (2.7), converges to a consistent estimator (Kasahara and Shimotsu, 2012). In the following section, we introduce the econometric model motivated by the network game, and discuss its identification and estimation.

3 Econometric Model

3.1 Model and Identification

Let $\boldsymbol{\pi}_k = \mathbf{X}\boldsymbol{\beta}_k$, for $k = 1, \cdots, m$, where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]'$ is an $n \times q$ matrix of observations on q exogenous variables and $\boldsymbol{\beta}_k$ is a $q \times 1$ vector of unknown parameters.[7] The $q \times 1$ vector of exogenous variables $\mathbf{x}_i$ contains exogenous characteristics of individual i, and may also include the exogenous characteristics of individual i's peers. For instance, let $\boldsymbol{\chi}$ be a matrix of observations on exogenous individual characteristics. Then, a possible specification of $\mathbf{X}$ is given by $\mathbf{X} = [\boldsymbol{\chi}, \mathbf{W}\boldsymbol{\chi}]$, with the coefficients of $\mathbf{W}\boldsymbol{\chi}$ representing *contextual effects* (Manski, 1993).

Substitution of $\mathbf{\Pi} = \mathbf{XB}$, with $\mathbf{B} = [\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_m]$, into (2.4) gives the structural econometric model

$$\mathbf{Y\Theta} = \mathbf{WP\Lambda} + \mathbf{XB} - \mathbf{E}. \tag{3.1}$$

We assume $\text{vec}(\mathbf{E})|\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I}_n)$, where $\mathbf{\Sigma} = [\sigma_{lk}]$ is an $m \times m$ covariance matrix. As we discuss in Section 2, what distinguishes this model from the single-activity social interaction model in Lee et al. (2014) is that it allows for *the simultaneity effect* captured by θ_{lk} $(k \neq l)$, *the cross-activity peer effect* captured by λ_{lk} $(k \neq l)$, besides *the within-activity peer effect* captured by λ_{kk}. The main purpose of this paper is to identify and estimate these different effects in model (3.1).

[7] As $\pi_{ik} = \varpi_{ik}/\vartheta_{kk}$, this can be considered as a parameterization of $\varpi_{ik} = \mathbf{x}_i'\mathbf{b}_k$ in the utility (2.1), with $\mathbf{b}_k = \vartheta_{kk}\boldsymbol{\beta}_k$.

If $\mathbf{\Theta}$ is nonsingular, the reduced form of model (3.1) is

$$\mathbf{Y} = \mathbf{WP\Lambda}^* + \mathbf{XB}^* - \mathbf{E}^*, \tag{3.2}$$

where $\mathbf{\Lambda}^* = \mathbf{\Lambda\Theta}^{-1}$, $\mathbf{B}^* = \mathbf{B\Theta}^{-1}$, and $\mathbf{E}^* = \mathbf{E\Theta}^{-1}$. As $\text{vec}(\mathbf{E})|\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I}_n)$, we have $\text{vec}(\mathbf{E}^*)|\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma}^* \otimes \mathbf{I}_n)$ with $\mathbf{\Sigma}^* = [\sigma^*_{lk}] = \mathbf{\Theta}'^{-1}\mathbf{\Sigma\Theta}^{-1}$. Following Maddala (1983), the diagonal elements of $\mathbf{\Sigma}^*$ are normalized such that $\sigma^*_{kk} = 1$ for all $k = 1, \cdots, m$. The assumption on model disturbances is summarized as follows.[8]

Assumption 2. $\text{vec}(\mathbf{E})|\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I}_n)$, *which implies* $\text{vec}(\mathbf{E}^*)|\mathbf{X} \sim N(\mathbf{0}, \mathbf{\Sigma}^* \otimes \mathbf{I}_n)$ *with* $\mathbf{\Sigma}^* = \mathbf{\Theta}'^{-1}\mathbf{\Sigma\Theta}^{-1}$. *We normalize the model parameters so that* $\mathbf{\Sigma}^*$ *has a unity diagonal.*

First, we consider the identification of the reduced form parameters $\mathbf{\Psi}^* = [\mathbf{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}]'$. Let λ^*_{lk} denote the (l,k)-th element of $\mathbf{\Lambda}^*$, and $\boldsymbol{\beta}^*_k$ denote the k-th column of $\mathbf{B}^*$. Given the observed adjacency matrix $\mathbf{W}$ and exogenous covariates $\mathbf{X}$, the parameters $\mathbf{\Psi}^* = [\mathbf{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}]'$ and the alternative parameters $\widetilde{\mathbf{\Psi}}^* = [\widetilde{\mathbf{\Lambda}}^{*\prime}, \widetilde{\mathbf{B}}^{*\prime}]'$ are observationally equivalent if

$$\Pr(y_{ik} = 1|\mathbf{W}, \mathbf{X}) = \Phi(\textstyle\sum_{l=1}^m \lambda^*_{lk} \sum_{j=1}^n w_{ij} p_{jl} + \mathbf{x}_i'\boldsymbol{\beta}^*_k) = \Phi(\sum_{l=1}^m \widetilde{\lambda}^*_{lk} \sum_{j=1}^n w_{ij} \widetilde{p}_{jl} + \mathbf{x}_i'\widetilde{\boldsymbol{\beta}}^*_k) \tag{3.3}$$

for all $i = 1, \cdots, n$ and $k = 1, \cdots, m$, where, under Assumption 1, p_{ik} and $\widetilde{p}_{ik}$ are uniquely

[8] Brock and Durlauf (2007) considered a discrete-choice *group interaction* model, where the rational expectation on the average choice of the group is *homogenous* (within a group) as it is formed based on *group-level* information. Brock and Durlauf (2007) treated the rational expectation as a data moment and established nonparametric identification of the model. However, in this paper, the rational expectation on the average choice of the peers is *heterogenous* as it is formed based on *individual-level* information. It might not be reasonable to assume the rational expectation can be directly identified from a sample moment as in Brock and Durlauf (2007). Therefore, in this paper, we consider parametric identification following the literature of discrete-choice *network interaction* models (e.g. Lee et al., 2014; Lin and Xu, 2017; Xu, 2018).

determined by the fixed point mappings

$$p_{ik} = \Phi(\sum_{l=1}^{m} \lambda_{lk}^* \sum_{j=1}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k^*) \tag{3.4}$$

$$\widetilde{p}_{ik} = \Phi(\sum_{l=1}^{m} \widetilde{\lambda}_{lk}^* \sum_{j=1}^{n} w_{ij} \widetilde{p}_{jl} + \mathbf{x}_i' \widetilde{\boldsymbol{\beta}}_k^*) \tag{3.5}$$

respectively. If (3.3)-(3.5) hold, then $p_{ik} = \widetilde{p}_{ik}$ for all $i = 1, \cdots, n$ and $k = 1, \cdots, m$. Substitution of $p_{ik} = \widetilde{p}_{ik}$ into the observational equivalence condition (3.3) yields

$$\sum_{l=1}^{m} (\lambda_{lk}^* - \widetilde{\lambda}_{lk}^*) \sum_{j=1}^{n} w_{ij} p_{jl} + \mathbf{x}_i' (\boldsymbol{\beta}_k^* - \widetilde{\boldsymbol{\beta}}_k^*) = 0, \qquad \forall i, k$$

or

$$[\mathbf{WP}, \mathbf{X}](\boldsymbol{\Psi}^* - \widetilde{\boldsymbol{\Psi}}^*) = \mathbf{0}.$$

If $[\mathbf{WP}, \mathbf{X}]$ has full column rank, then the observational equivalence of $\boldsymbol{\Psi}^*$ and $\widetilde{\boldsymbol{\Psi}}^*$ implies $\boldsymbol{\Psi}^* = \widetilde{\boldsymbol{\Psi}}^*$, i.e., the reduced form parameters $\boldsymbol{\Psi}^*$ are identifiable.

With the reduced form parameters identified, the structural parameters $\boldsymbol{\Theta}$, $\boldsymbol{\Lambda}$ and $\mathbf{B}$ can be identified from the equations $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\boldsymbol{\Theta}^{-1}$ and $\mathbf{B}^* = \mathbf{B}\boldsymbol{\Theta}^{-1}$ under the usual rank conditions. Let $\boldsymbol{\Gamma} = [\boldsymbol{\Theta}', -\boldsymbol{\Lambda}', -\mathbf{B}']'$ and $\boldsymbol{\gamma}_k$ denote the k-th column of $\boldsymbol{\Gamma}$. Suppose $\mathbf{R}_k$ is a conformable matrix representing zero restrictions on the coefficients of the k-th equation such that $\mathbf{R}_k \boldsymbol{\gamma}_k = \mathbf{0}$. The sufficient and necessary rank condition for the identification of the structural parameters $\boldsymbol{\Gamma}$ from the reduced form parameters is $\text{rank}(\mathbf{R}_k \boldsymbol{\Gamma}) = m - 1$ for $k = 1, \cdots, m$ (Schmidt, 1976). The above rank conditions are summarized in the following assumption.

Assumption 3. (i) $[\mathbf{WP}, \mathbf{X}]$ *has full column rank.* (ii) $\text{rank}(\mathbf{R}_k \boldsymbol{\Gamma}) = m - 1$ *for* $k = 1, \cdots, m$.

Remark 1. The matrix of equilibrium beliefs $\mathbf{P}$ is implicitly defined by

$$p_{ik} = \Phi(\textstyle\sum_{l=1}^{m} \lambda_{lk}^{*} \sum_{j=1}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k^{*}). \tag{3.6}$$

To help understand the rank condition of $[\mathbf{WP}, \mathbf{X}]$ in Assumption 3 (i), we consider a linear probability model, where Equation (3.6) becomes $p_{ik} = \sum_{l=1}^{m} \lambda_{lk}^{*} \sum_{j=1}^{n} w_{ij} p_{jl} + \mathbf{x}_i' \boldsymbol{\beta}_k^{*}$. In matrix form, the equilibrium beliefs of the linear probability model are given by

$$\mathbf{P} = \mathbf{WP}\boldsymbol{\Lambda}^{*} + \mathbf{XB}^{*},$$

which implies $\mathrm{vec}(\mathbf{P}) = (\mathbf{I}_{mn} - \boldsymbol{\Lambda}^{*\prime} \otimes \mathbf{W})^{-1} (\mathbf{I}_m \otimes \mathbf{X}) \mathrm{vec}(\mathbf{B}^{*})$ assuming the nonsingularity of $\mathbf{I}_{mn} - \boldsymbol{\Lambda}^{*\prime} \otimes \mathbf{W}$. If $\mathbf{X} = [\boldsymbol{\chi}, \mathbf{W}\boldsymbol{\chi}]$, where $\boldsymbol{\chi}$ is a matrix of observations on exogenous individual characteristics, and all off-diagonal elements of the zero-diagonal adjacency matrix $\mathbf{W}$ are $(n-1)^{-1}$, then, it follows by a similar argument as in Bramoullé et al. (2009) that, for the linear probability model, $[\mathbf{WP}, \mathbf{X}]$ does not have full column rank. This non-identification result due to the perfect collinearity of $\mathbf{WP}$ and $[\boldsymbol{\chi}, \mathbf{W}\boldsymbol{\chi}]$ is referred to as the "reflection problem" (Manski, 1993). By contrast, due to the nonlinearity of Equation (3.6), Brock and Durlauf (2007) pointed out that the "reflection problem" does not arise in the binary-choice social interaction model.

Remark 2. To help understand the rank condition in Assumption 3 (ii), we consider a "triangular" system. Suppose $m = 2$ and $\theta_{21} = \lambda_{21} = \lambda_{12} = 0$. Then, model (3.1) becomes

$$\mathbf{y}_1 = \lambda_{11} \mathbf{W} \mathbf{p}_1 + \mathbf{X} \boldsymbol{\beta}_1 - \boldsymbol{\epsilon}_1 \tag{3.7}$$

$$\mathbf{y}_2 = -\theta_{12} \mathbf{y}_1 + \lambda_{22} \mathbf{W} \mathbf{p}_2 + \mathbf{X} \boldsymbol{\beta}_2 - \boldsymbol{\epsilon}_2. \tag{3.8}$$

In this case, an individual's choice is influenced by the expectation on the peers' choices in the same activity. Furthermore, an individual's choice in activity 2 is influenced by his/her own choice in activity 1, but not the other way around. The exclusion restrictions $\theta_{21} = \lambda_{21} = 0$ can be represented by $\mathbf{R}_1\boldsymbol{\gamma}_1 = \mathbf{0}$ where

$$\mathbf{R}_1 = \begin{bmatrix} 0 & -1 & 0 & 0 & \mathbf{0}_{1\times q} \\ 0 & 0 & 0 & -1 & \mathbf{0}_{1\times q} \end{bmatrix}$$

and $\boldsymbol{\gamma}_1 = (1, -\theta_{21}, -\lambda_{11}, -\lambda_{21}, -\boldsymbol{\beta}_1')'$. The exclusion restriction $\lambda_{12} = 0$ can be represented by $\mathbf{R}_2\boldsymbol{\gamma}_2 = 0$ where $\mathbf{R}_2 = (0, 0, -1, 0, \mathbf{0}_{1\times q})$ and $\boldsymbol{\gamma}_2 = (-\theta_{21}, 1, -\lambda_{12}, -\lambda_{22}, -\boldsymbol{\beta}_2')'$. It is easy to check Assumption 3 (ii) holds because

$$\mathbf{R}_1\boldsymbol{\Gamma} = \begin{bmatrix} 0 & -1 \\ 0 & \lambda_{22} \end{bmatrix}$$

has rank one, and $\mathbf{R}_2\boldsymbol{\Gamma} = [\lambda_{11}, 0]$ also has rank one if $\lambda_{11} \neq 0$. When Assumption 3 (ii) holds, the structural parameters can be identified if the reduced form parameters are identifiable. The reduced form of the model is

$$\begin{aligned} \mathbf{y}_1 &= \lambda_{11}^*\mathbf{W}\mathbf{p}_1 + \mathbf{X}\boldsymbol{\beta}_1^* - \boldsymbol{\epsilon}_1^* \\ \mathbf{y}_2 &= \lambda_{12}^*\mathbf{W}\mathbf{p}_1 + \lambda_{22}^*\mathbf{W}\mathbf{p}_2 + \mathbf{X}\boldsymbol{\beta}_2^* - \boldsymbol{\epsilon}_2^*, \end{aligned}$$

where

$$\lambda_{11}^* = \lambda_{11}, \quad \boldsymbol{\beta}_1^* = \boldsymbol{\beta}_1, \quad \lambda_{12}^* = -\theta_{12}\lambda_{11}, \quad \lambda_{22}^* = \lambda_{22}, \quad \text{and} \quad \boldsymbol{\beta}_2^* = \boldsymbol{\beta}_2 - \theta_{12}\boldsymbol{\beta}_1. \qquad (3.9)$$

It is easy to see that, indeed, if the reduced form parameters λ_{11}^*, $\boldsymbol{\beta}_1^*$, λ_{12}^*, λ_{22}^*, $\boldsymbol{\beta}_2^*$ are identifiable, then structural parameters λ_{11}, $\boldsymbol{\beta}_1$, θ_{12}, λ_{22}, $\boldsymbol{\beta}_2$ can be identified from the linear restrictions given in (3.9).

Under Assumptions 1-3, the next section proposes a two-stage estimation procedure for the consistent estimation of the structural parameters in model (3.1).

3.2 Estimation

The estimation of model (3.1) follows the above identification strategy. First, we estimate the reduced form parameters $\mathbf{\Psi}^* = [\mathbf{\Lambda}^{*\prime}, \mathbf{B}^{*\prime}]'$ using the NPL algorithm. The NPL algorithm was proposed by Aguirregabiria and Mira (2007) for the estimation of dynamic discrete-choice games, and has recently been adopted by Lin and Xu (2017) for the estimation of large network games. For the estimation of the reduced form equation (3.2), the NPL algorithm starts from an arbitrary initial value $\mathbf{p}^{(0)} \in [0,1]^{nm}$ and takes the following iterative steps:

Step 1 Given $\mathbf{p}^{(j-1)}$, obtain $\widehat{\boldsymbol{\psi}}_k^{*(j)} \equiv (\widehat{\lambda}_{1k}^{*(j)}, \cdots, \widehat{\lambda}_{mk}^{*(j)}, \widehat{\boldsymbol{\beta}}_k^{*(j)\prime})' = \arg\max \ln L(\boldsymbol{\psi}_k^*; \mathbf{p}^{(j-1)})$, where

$$\begin{aligned}\ln L(\boldsymbol{\psi}_k^*; \mathbf{p}^{(j-1)}) &= \textstyle\sum_{i=1}^n d_{ik} \ln \Phi(\sum_{l=1}^m \lambda_{lk}^* \sum_{j=1}^n w_{ij} p_{jl}^{(j-1)} + \mathbf{x}_i' \boldsymbol{\beta}_k^*) \\ &\quad + \textstyle\sum_{i=1}^n (1-d_{ik}) \ln[1 - \Phi(\sum_{l=1}^m \lambda_{lk}^* \sum_{j=1}^n w_{ij} p_{jl}^{(j-1)} + \mathbf{x}_i' \boldsymbol{\beta}_k^*)],\end{aligned}$$

for $k = 1, \cdots, m$.

Step 2 Given $\widehat{\mathbf{\Psi}}^{*(j)} = [\widehat{\boldsymbol{\psi}}_1^{*(j)}, \cdots, \widehat{\boldsymbol{\psi}}_m^{*(j)}]$, obtain $\mathbf{p}^{(j)} = \vec{h}(\mathbf{p}^{(j-1)}; \widehat{\mathbf{\Psi}}^{*(j)})$, where

$$\vec{h}(\mathbf{p}^{(j-1)}; \widehat{\mathbf{\Psi}}^{*(j)}) = [\vec{h}_1(\mathbf{p}^{(j-1)}; \widehat{\mathbf{\Psi}}^{*(j)})', \cdots, \vec{h}_m(\mathbf{p}^{(j-1)}; \widehat{\mathbf{\Psi}}^{*(j)})']',$$

with

$$\vec{h}_k(\mathbf{p}^{(j-1)}; \widehat{\mathbf{\Psi}}^{*(j)}) = \begin{bmatrix} \Phi(\sum_{l=1}^m \widehat{\lambda}_{lk}^{*(j)} \sum_{j=1}^n w_{1j} p_{jl}^{(j-1)} + \mathbf{x}_1' \widehat{\boldsymbol{\beta}}_k^{*(j)}) \\ \vdots \\ \Phi(\sum_{l=1}^m \widehat{\lambda}_{lk}^{*(j)} \sum_{j=1}^n w_{nj} p_{jl}^{(j-1)} + \mathbf{x}_n' \widehat{\boldsymbol{\beta}}_k^{*(j)}) \end{bmatrix},$$

for $k = 1, \cdots, m$. Update $\mathbf{p}^{(j-1)}$ in Step 1 to $\mathbf{p}^{(j)}$. Repeat Steps 1 and 2 until the process converges.

Kasahara and Shimotsu (2012) have shown that a key determinant of the convergence of the NPL algorithm is the contraction property of the fixed point mapping (2.7), which is ensured by Assumption 1. When the NPL algorithm converges, the NPL estimator $\widehat{\mathbf{\Psi}}^* = [\widehat{\boldsymbol{\psi}}_1^*, \cdots, \widehat{\boldsymbol{\psi}}_m^*]$ satisfies $\widehat{\boldsymbol{\psi}}_k^* = \arg\max \ln L(\boldsymbol{\psi}_k^*; \widehat{\mathbf{p}})$, for $k = 1, \cdots, m$, where $\widehat{\mathbf{p}}$ is implicitly defined in $\widehat{\mathbf{p}} = \vec{h}(\widehat{\mathbf{p}}; \widehat{\mathbf{\Psi}}^*)$. Under regularity conditions, it follows by a similar argument as in Aguirregabiria and Mira (2007) and Lin and Xu (2017) that the NPL estimator is root-n consistent and asymptotically normal. The asymptotic distribution of the NPL estimator is given in Appendix A.

Remark 3. Alternatively, the reduced form parameters can be estimated by the nested fixed point (NFXP) algorithm (Rust, 1987), which has been adopted by Lee et al. (2014) and Yang and Lee (2017) to estimate discrete-choice social interaction models with rational expectations. In our case, the NFXP estimator is given by $\widetilde{\mathbf{\Psi}}^* = \arg\max \ln L(\mathbf{\Psi}^*; \mathbf{p}(\mathbf{\Psi}^*))$, where $\ln L(\mathbf{\Psi}^*; \mathbf{p}(\mathbf{\Psi}^*))$ is the log-likelihood function based on the multivariate normal distribution of $(\mathbf{y}_1, \cdots, \mathbf{y}_m)$ and $\mathbf{p}(\mathbf{\Psi}^*)$ is the (unique) solution of $\mathbf{p} = \vec{h}(\mathbf{p}; \mathbf{\Psi}^*)$. Compared to the proposed NPL algorithm, the NFXP algorithm is computationally demanding because it repeatedly solves the fixed point mapping $\mathbf{p} = \vec{h}(\mathbf{p}; \mathbf{\Psi}^*)$ for $\mathbf{p}(\mathbf{\Psi}^*)$ at each candidate parameter value of $\mathbf{\Psi}^*$ in the search for the maximum of the log-likelihood function. Furthermore, the NFXP algorithm has to estimate the reduced form equations

$$\mathbf{y}_k = \sum_{l=1}^{m} \lambda_{lk}^{*} \mathbf{W}\mathbf{p}(\boldsymbol{\Psi}^{*}) + \mathbf{X}\boldsymbol{\beta}_k^{*} - \boldsymbol{\epsilon}_k^{*}, \qquad \text{for } k = 1, \cdots, m, \tag{3.10}$$

jointly, because the k-th reduced form equation depends not only on $\boldsymbol{\psi}_k^{*} = (\lambda_{1k}^{*}, \cdots, \lambda_{mk}^{*}, \boldsymbol{\beta}_k^{*\prime})'$ but also on $\boldsymbol{\Psi}^{*} = [\boldsymbol{\psi}_1^{*}, \cdots, \boldsymbol{\psi}_m^{*}]$ via $\mathbf{p}(\boldsymbol{\Psi}^{*})$. On the other hand, as the NPL algorithm maximizes the "pseudo" log-likelihood functions of the reduced form equations

$$\mathbf{y}_k = \sum_{l=1}^{m} \lambda_{lk}^{*} \mathbf{W}\mathbf{p}^{(j-1)} + \mathbf{X}\boldsymbol{\beta}_k^{*} - \boldsymbol{\epsilon}_k^{*}, \qquad \text{for } k = 1, \cdots, m, \tag{3.11}$$

for a given $\mathbf{p}^{(j-1)}$ at each iteration, it can estimate (3.11) equation-by-equation for $\boldsymbol{\psi}_k^{*}$. Therefore, the proposed NPL algorithm is asymptotically less efficient than the NFXP algorithm but computationally much less costly, especially when the number of simultaneous equations m is large.

Let $\mathbf{Z} = [\mathbf{WP}, \mathbf{X}]$. Then, the reduced form (3.2) can be written more compactly as

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\Psi}^{*} - \mathbf{E}^{*}. \tag{3.12}$$

With the estimated reduced form parameters $\widehat{\boldsymbol{\Psi}}^{*}$, we can estimate the structural parameters $\boldsymbol{\Gamma} = [\boldsymbol{\Theta}', -\boldsymbol{\Psi}']'$ with $\boldsymbol{\Psi} = [\boldsymbol{\Lambda}', \mathbf{B}']'$ by the AGLS procedure (Amemiya, 1978). Amemiya (1978) and Lee (1981) have shown that the AGLS estimator is more efficient than a Heckman-type two-stage estimator for the simultaneous-equation probit model. Without loss of generality, we describe the AGLS procedure for the first equation of the structural model (3.1). According to Assumption 3 (ii), the identification of the structural parameters $\boldsymbol{\Gamma}$ from the reduced form parameters $\boldsymbol{\Psi}^{*}$ requires exclusion restrictions on $\boldsymbol{\Gamma}$. Under appropriate exclusion restrictions, the first equation of the structural model (3.1) can be written as, with a slight abuse of notation,

$$\mathbf{y}_1 = -\mathbf{Y}_1\boldsymbol{\theta}_1 + \mathbf{Z}_1\boldsymbol{\psi}_1 - \boldsymbol{\epsilon}_1, \tag{3.13}$$

where $\mathbf{Y}_1$ and $\mathbf{Z}_1$ denote, respectively, submatrices of observations on the variables (other than $\mathbf{y}_1$) in $\mathbf{Y}$ and $\mathbf{Z}$ that are included in the first equation, with their coefficients $\boldsymbol{\theta}_1$ and $\boldsymbol{\psi}_1$ being, respectively, vectors of unrestricted parameters in the first columns of $\boldsymbol{\Theta}$ and $\boldsymbol{\Psi}$. Let $\mathbf{J}_Y$ and $\mathbf{J}_Z$ be selection matrices such that $\mathbf{Y}_1 = \mathbf{Y}\mathbf{J}_Y$ and $\mathbf{Z}_1 = \mathbf{Z}\mathbf{J}_Z$. Then, (3.13) can be rewritten as

$$\mathbf{y}_1 = -\mathbf{Y}\mathbf{J}_Y\boldsymbol{\theta}_1 + \mathbf{Z}\mathbf{J}_Z\boldsymbol{\psi}_1 - \boldsymbol{\epsilon}_1. \tag{3.14}$$

Substitution of (3.12) into (3.14) gives

$$\mathbf{y}_1 = -\mathbf{Z}(\boldsymbol{\Psi}^*\mathbf{J}_Y\boldsymbol{\theta}_1 - \mathbf{J}_Z\boldsymbol{\psi}_1) + \mathbf{E}^*\mathbf{J}_Y\boldsymbol{\theta}_1 - \boldsymbol{\epsilon}_1. \tag{3.15}$$

Comparing (3.15) with (3.12), we have

$$\boldsymbol{\psi}_1^* = -\boldsymbol{\Psi}^*\mathbf{J}_Y\boldsymbol{\theta}_1 + \mathbf{J}_Z\boldsymbol{\psi}_1, \tag{3.16}$$

where $\boldsymbol{\psi}_1^*$ is the first column of $\boldsymbol{\Psi}^*$. (3.16) implies the following regression equation

$$\widehat{\boldsymbol{\psi}}_1^* = -\widehat{\boldsymbol{\Psi}}^*\mathbf{J}_Y\boldsymbol{\theta}_1 + \mathbf{J}_Z\boldsymbol{\psi}_1 + \mathbf{v}_1, \tag{3.17}$$

where $\mathbf{v}_1 = (\widehat{\boldsymbol{\psi}}_1^* - \boldsymbol{\psi}_1^*) + (\widehat{\boldsymbol{\Psi}}^* - \boldsymbol{\Psi}^*)\mathbf{J}_Y\boldsymbol{\theta}_1$. Let $\boldsymbol{\Omega}_{11}$ denote the asymptotic covariance matrix of $\mathbf{v}_1$, and $\widehat{\boldsymbol{\Omega}}_{11}$ denote a consistent estimator of $\boldsymbol{\Omega}_{11}$. The AGLS estimator of $\boldsymbol{\delta}_1 = (\boldsymbol{\theta}_1', \boldsymbol{\psi}_1')'$ is given by

$$\widehat{\boldsymbol{\delta}}_1 = (\widehat{\mathbf{H}}_1'\widehat{\boldsymbol{\Omega}}_{11}^{-1}\widehat{\mathbf{H}}_1)^{-1}\widehat{\mathbf{H}}_1'\widehat{\boldsymbol{\Omega}}_{11}^{-1}\widehat{\boldsymbol{\psi}}_1^*,$$

where $\widehat{\mathbf{H}}_1 = [-\widehat{\boldsymbol{\Psi}}^* \mathbf{J}_Y, \mathbf{J}_Z]$. The AGLS procedure estimates the m equations in the form of (3.17) equation by equation. Lee (1982) extends the AGLS estimator to a seemingly unrelated regressions (SUR) type estimator that estimate the m equations jointly. The detailed derivation of the SUR-type AGLS estimator and its asymptotic covariance matrix is given in Appendix A.

4 Monte Carlo Experiment

To investigate the finite sample performance of the proposed estimation procedure, we conduct a limited simulation study on the following model

$$\mathbf{y}_1 = -\theta_{21}\mathbf{y}_2 + \lambda_{11}\mathbf{W}\mathbf{p}_1 + \lambda_{21}\mathbf{W}\mathbf{p}_2 + \mathbf{X}_1\boldsymbol{\beta}_1 - \boldsymbol{\epsilon}_1 \tag{4.1}$$

$$\mathbf{y}_2 = -\theta_{12}\mathbf{y}_1 + \lambda_{12}\mathbf{W}\mathbf{p}_1 + \lambda_{22}\mathbf{W}\mathbf{p}_2 + \mathbf{X}_2\boldsymbol{\beta}_2 - \boldsymbol{\epsilon}_2, \tag{4.2}$$

where $\mathbf{X}_1 = [\boldsymbol{\chi}_1, \mathbf{W}\boldsymbol{\chi}_1]$ and $\mathbf{X}_2 = [\boldsymbol{\chi}_2, \mathbf{W}\boldsymbol{\chi}_2]$. $\boldsymbol{\chi}_1$ and $\boldsymbol{\chi}_2$ are $n \times 1$ vectors of random variables representing observable individual characteristics. The coefficients of $\mathbf{W}\boldsymbol{\chi}_1$ and $\mathbf{W}\boldsymbol{\chi}_2$ represent contextual effects (Manski, 1993). In the data generating process, we consider two specifications of the adjacency matrix $\mathbf{W} = [w_{ij}]$. The first one is based on a circular network where the n individuals are equidistantly located around a circle and are only connected with the nearest neighbors. The non-zero elements of the corresponding adjacency matrix are $w_{1,2} = w_{1,n} = w_{n,1} = w_{n,n-1} = 1/2$ and $w_{i,i-1} = w_{i,i+1} = 1/2$ for $i = 2, \cdots, n-1$. The second one is based on a random network, where each individual randomly nominates five friends. $w_{ij} = 1/5$ if i nominates j as a friend.

We conduct 1000 repetitions in the simulation with $n \in \{500, 1000, 2000\}$. In each repetition, $\boldsymbol{\chi}_1$ and $\boldsymbol{\chi}_2$ are generated from $N(\mathbf{0}, \mathbf{I}_n)$ and the reduced form disturbances given by $(\boldsymbol{\epsilon}_1^{*\prime}, \boldsymbol{\epsilon}_2^{*\prime})'$ are generated from $N(\mathbf{0}, \boldsymbol{\Sigma}^* \otimes \mathbf{I}_n)$, where

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} 1 & \sigma_{12}^* \\ \sigma_{12}^* & 1 \end{bmatrix}.$$

We set $\theta_{21} = \theta_{12} = 0.5$, $\lambda_{11} = \lambda_{22} = 0.9$, $\lambda_{21} = \lambda_{12} = 0.6$, and $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = (1,1)'$, and experiment with different values for σ_{12}^*.

[Tables 1 and 2 approximately here]

For the estimation of the structural parameters from the reduced form parameters, we consider both the equation-by-equation AGLS estimator (AGLS-1) and the SUR-type AGLS estimator (AGLS-2). We report the mean and standard deviation (SD) of the empirical distributions of the estimates. The estimation results for Equation (4.1) are reported in Tables 1 and 2. The estimates are essentially unbiased when the sample size is large. The standard deviations of all estimators reduce as the sample size increases. The AGLS-2 estimator reduces the standard deviation of the AGLS-1 estimator. In general, the reduction in the standard deviation is more prominent when σ_{12}^* is larger. For the circular network, when $n = 2000$ and $\sigma_{12}^* = 0.9$, the AGLS-2 estimators of λ_{11}, λ_{21}, β_{11} and β_{21} reduce the standard deviations of the AGLS-1 estimators by, respectively, 7.3%, 6.7%, 7.4%, and 27.9%. Furthermore, the proposed estimation procedure remains computationally feasible when the network size is large. For the circular network, when $n = 2000$ and $\sigma_{12}^* = 0.5$, the average computation time of one simulation repetition is less than 5 seconds.[9]

[9] The computation is conducted on a PC with an Intel(R) Core(TM) i7-6700 CPU @ 3.4 GHz and 32 GB RAM.

5 Empirical Example

To illustrate the empirical relevance of the proposed model, we study the peer effect of participating in sports activities on underage smoking using a unique and now widely used data set provided by the National Longitudinal Study of Adolescent Health. The Add Health data provides national representative information on 7th-12th graders in the United States. The in-school survey was conducted during the 1994-1995 year with four follow-up in-home interviews. Here we only use the first wave of Add Health data.

In this empirical example, we consider the estimation of model (3.1) with $d_{i1} = I(y_{i1} > 0)$ and $d_{i2} = I(y_{i2} > 0)$ being indicators of smoking and participating in sports activities respectively. To be more specific, d_{i1} is one if student i smokes once a week or more in the last 12 months, and zero otherwise. d_{i2} is one if student i participates in cheerleading/dance, baseball/softball, basketball, field hockey, football, ice hockey, soccer, swimming, tennis, track, volleyball, wrestling, and/or other sport team this year, and zero otherwise.

The adjacency matrix $\mathbf{W} = [w_{ij}]$ is constructed based on the friend-nomination information in the Add Health data. In the survey, students were asked to nominate their best friends (up to 5 male friends and 5 female friends) from a school roster.[10] Let n_i denote the number of friends that student i nominates. We set $w_{ij} = 1/n_i$ if student i nominates j as a friend and $w_{ij} = 0$ otherwise. After removing students with missing observations on the dependent variables, the sample consists of 71,918 students distributed over 141 schools. A summary of the data can be found in Table 3.

[Table 3 approximately here]

As discussed in Section 3.1, the identification of model (3.1) requires exclusion restric-

[10] In the data, the average number of friends of a student is 3.29 with the standard deviation 2.40. Less than 1% of the students nominated 10 friends and thus the bound on the number of friend-nominations is not binding.

tions. In this empirical application, we impose the following exclusion restrictions. First, whether or not a student's parent smokes cigarettes and whether or not cigarettes are easily available in a student's home only indirectly affect his/her decision to participate in sport teams through his/her smoking behavior and the smoking behavior of his/her friends.[11] Second, whether or not a student is well coordinated only indirectly affects a student's smoking behavior through his/her participation in sport teams and the sports participation decisions of his/her friends.

[Table 4 approximately here]

Under the above exclusion restrictions, the rank condition given by Assumption 3 is satisfied. We apply the estimation procedure described in Section 3.2 and Appendix A to estimate the model. The estimation results are reported in Table 4.[12] The left panel of Table 4 reports the estimates from Lee et al.'s (2014) single-equation network model without taking into account the potential interdependence of smoking and sports participation decisions. The right panel of Table 4 reports the SUR-type AGLS estimates from the simultaneous-equation network model proposed in this paper.

First of all, for the single-equation network model, the estimated within-activity peer effect parameters $\widehat{\lambda}_{kk}$ lie within the parameter space $|\lambda_{kk}| < \sqrt{2\pi}$ that guarantees the existence of a unique BNE (Lee et al., 2014). For the simultaneous-equation network model, the estimated reduced form parameters by the NPL algorithm are $\widehat{\lambda}^*_{11} = 1.4016$, $\widehat{\lambda}^*_{21} = -0.4666$, $\widehat{\lambda}^*_{12} = -1.2111$, and $\widehat{\lambda}^*_{22} = 0.8487$. The estimated reduced form parameters satisfy

[11] The information on whether or not a student's parent smokes cigarettes and whether or not cigarettes are easily available in a student's home is only available for a subsample of the students who are randomly selected to participate in the in-home interviews. Hence, we introduce dummy variables "parent smokes: missing" and "cigarettes in the home: missing" to indicate missing observations.

[12] In the estimation, we control for exogenous contextual effects (i.e. the influence on a student's behavior from his/her friends' exogenous characteristics, see, Manski, 1993) and school fixed effects. To save space, the estimates of contextual effects and school fixed effects are not reported in Table 4.

the parameter constraint, $\max\{|\lambda_{11}^*| + |\lambda_{21}^*|, |\lambda_{12}^*| + |\lambda_{22}^*|\} < \sqrt{2\pi}$, implied by Assumption 1 (ii).

For both single-equation and simultaneous-equation network models, we find evidence of positive within-activity peer effects, i.e., a student is more likely to smoke when his/her friends smoke and a student is more likely to participate in sport teams when his/her friend participate in sport teams. However, compared with the simultaneous-equation network model, the single-equation network model tends to overestimate the within-activity peer effect. More interestingly, from the simultaneous-equation network model, we also find that participation in sport teams not only reduces a student's own smoking behavior (a negative simultaneity effect) but also reduces the likelihood for his/her friend to smoke (a negative cross-activity peer effect).

6 Summary

In this paper, we consider a simultaneous-equation social interaction model with binary outcomes. The specification of the model is based on an incomplete information network game with individuals interacting in multiple activities. We give a sufficient condition for the existence of a unique BNE of the game. Following the identification strategy of the econometric model, we propose a two-stage estimation procedure and investigate its finite sample performance by Monte Carlo simulation experiments. We also provide an empirical example to illustrate the proposed estimation procedure.

Some possible extensions of the current work are in order. For example, people may form different social networks for different activities they participate. So it would be a natural extension to introduce activity-specific networks to the current model. However, the formation of activity-specific networks is likely to be correlated with the choices of actions, which may render a problem of endogenous networks. We leave this extension to future research.

References

Aguirregabiria, V. and Mira, P. (2007). Sequential estimation of dynamic discrete games, *Econometrica* **75**: 1–53.

Amemiya, T. (1978). The estimation of a simultaneous equation generalized probit model, *Econometrica* **46**: 1193–1205.

Ballester, C., Calvó-Armengol, A. and Zenou, Y. (2006). Who's who in networks. Wanted: the key player, *Econometrica* **74**: 1403–1417.

Blume, L. E., Brock, W. A., Durlauf, S. N. and Ioannides, Y. M. (2011). Identification of social interactions, *in* J. Benhabib, A. Bisin and M. O. Jackson (eds), *Handbook of Social Economics*, Vol. 1B, North-Holland, pp. 855–966.

Blume, L. E., Brock, W. A., Durlauf, S. N. and Jayaraman, R. (2015). Linear social interactions models, *Journal of Political Economy* **123**: 444–496.

Boucher, V., Bramoullé, Y., Djebbari, H. and Fortin, B. (2014). Do peers affect student achievement? Evidence from Canada using group size variation, *Journal of Applied Econometrics* **29**: 91–109.

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. (2001a). Discrete choice with social interaction, *Review of Economic Studies* **68**: 235–260.

Brock, W. A. and Durlauf, S. N. (2001b). Interactions-based models, *in* J. J. Heckman and E. Leamer (eds), *Handbook of Econometrics*, Vol. 5, North-Holland, pp. 3297–3380.

Brock, W. A. and Durlauf, S. N. (2007). Identification of binary choice models with social interactions, *Journal of Econometrics* **140**: 52–75.

Cohen-Cole, E., Liu, X. and Zenou, Y. (2018). Multivariate choices and identification of social interactions, *Journal of Applied Econometrics* **33**: 165–178.

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

Graham, B. S. (2015). Methods of identification in social networks, *Annual Review of Economics* **7**: 465–485.

Greene, W. H. (2012). *Econometric Analysis*, Prentice Hall, Upper Saddle River, NJ.

Hsieh, C. S. and Lee, L. F. (2017). Specification and estimation of network formation and network interaction models with the exponential probability distribution. Working paper, The Chinese University of Hong Kong.

Kasahara, H. and Shimotsu, K. (2012). Sequential estimation of structural models with a fixed point constraint, *Econometrica* **80**: 2303–2319.

Lee, L. F. (1981). Simultaneous equations models with discrete and censored variables, *in* C. F. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press, Cambridge.

Lee, L. F. (1982). Health and wage: A simultaneous equation model with multiple discrete indicators, *International Economic Review* **23**: 199–221.

Lee, L. F., Li, J. and Lin, X. (2014). Binary choice models with social network under heterogeneous rational expectations, *The Review of Economics and Statistics* **96**: 402–417.

Lin, X. (2010). Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables, *Journal of Labor Economics* **28**: 825–860.

Lin, Z. and Xu, H. (2017). Estimation of social-influence-dependent peer pressure in a large network game, *The Econometrics Journal* **20**: S86–S102.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*, Cambridge University Press, New York.

Manski, C. F. (1993). Identification of endogenous social effects: the reflection problem, *The Review of Economic Studies* **60**: 531–542.

Mele, A. (2017). A structural model of dense network formation, *Econometrica* **85**: 825–850.

Osborne, M. J. and Rubinstein, A. (1994). *A Course in Game Theory*, The MIT Press, Cambridge and London.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher, *Econometrica* **55**: 999–1033.

Schmidt, P. (1976). *Econometrics*, Marcel Dekker, New York.

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

Xu, H. (2018). Social interactions in large networks: A game theoretic approach, *International Economic Review* **59**: 257–284.

Yang, C. and Lee, L. F. (2017). Social interactions under incomplete information with heterogeneous expectations, *Journal of Econometrics* **198**: 65–83.

Appendices

A AGLS Estimator

For exposition purpose, we consider the case with $m = 2$. The result can be generalized to the case with an arbitrary m in a similar manner as in Lee (1981, 1982). When $m = 2$, the structural model (3.1) is

$$\begin{aligned} \mathbf{y}_1 &= -\mathbf{y}_2\theta_{21} + \mathbf{Z}_1\boldsymbol{\psi}_1 - \boldsymbol{\epsilon}_1 \\ \mathbf{y}_2 &= -\mathbf{y}_1\theta_{12} + \mathbf{Z}_2\boldsymbol{\psi}_2 - \boldsymbol{\epsilon}_2, \end{aligned} \tag{A.1}$$

with the reduced form equations

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{Z}\boldsymbol{\psi}_1^* - \boldsymbol{\epsilon}_1^* \\ \mathbf{y}_2 &= \mathbf{Z}\boldsymbol{\psi}_2^* - \boldsymbol{\epsilon}_2^*, \end{aligned} \tag{A.2}$$

where $\mathbf{Z} = [\mathbf{W}\mathbf{p}_1, \mathbf{W}\mathbf{p}_2, \mathbf{X}]$, and $\mathbf{Z}_1$ and $\mathbf{Z}_2$ contain columns of $\mathbf{Z}$ that appear, respectively, in the two structural equations (A.1). The i-th row of $\mathbf{Z}$ is $\mathbf{z}_i' = (\mathbf{w}_i\mathbf{p}_1, \mathbf{w}_i\mathbf{p}_2, \mathbf{x}_i')$, where $\mathbf{w}_i = (w_{i1}, \cdots, w_{in})$ denotes the i-th row of $\mathbf{W}$. We assume $(\epsilon_{i1}^*, \epsilon_{i2}^*)' \sim$ i.i.d.$N(\mathbf{0}, \boldsymbol{\Sigma}^*)$ with

$$\boldsymbol{\Sigma}^* = \begin{bmatrix} 1 & \sigma_{12}^* \\ \sigma_{12}^* & 1 \end{bmatrix}.$$

The reduced form parameters $\boldsymbol{\psi}_1^*$ and $\boldsymbol{\psi}_2^*$ can be estimated by the NPL algorithm described in Section 3.2. Let the NPL estimator of $\boldsymbol{\psi}^* = (\boldsymbol{\psi}_1^{*\prime}, \boldsymbol{\psi}_2^{*\prime})'$ be denoted by $\widehat{\boldsymbol{\psi}}^* = (\widehat{\boldsymbol{\psi}}_1^{*\prime}, \widehat{\boldsymbol{\psi}}_2^{*\prime})'$. Let $\widehat{\mathbf{p}} = (\widehat{\mathbf{p}}_1', \widehat{\mathbf{p}}_2')'$ be a NPL fixed point implicitly defined by $\widehat{\mathbf{p}}_k = \vec{h}_k(\widehat{\mathbf{p}}; \widehat{\boldsymbol{\psi}}_k^*) = [\Phi(\widehat{\mathbf{z}}_1'\widehat{\boldsymbol{\psi}}_k^*), \cdots, \Phi(\widehat{\mathbf{z}}_n'\widehat{\boldsymbol{\psi}}_k^*)]'$ for $k = 1, 2$, where $\widehat{\mathbf{z}}_i = (\mathbf{w}_i\widehat{\mathbf{p}}_1, \mathbf{w}_i\widehat{\mathbf{p}}_2, \mathbf{x}_i')'$. Then, the NPL estimator is given by $\widehat{\boldsymbol{\psi}}_k^* = \arg\max \ln L(\boldsymbol{\psi}_k^*; \widehat{\mathbf{p}})$, where

$$\ln L(\boldsymbol{\psi}_k^*; \widehat{\mathbf{p}}) = \textstyle\sum_{i=1}^n d_{ik} \ln \Phi(\widehat{\mathbf{z}}_i' \boldsymbol{\psi}_k^*) + \sum_{i=1}^n (1 - d_{ik}) \ln[1 - \Phi(\widehat{\mathbf{z}}_i' \boldsymbol{\psi}_k^*)], \qquad \text{for } k = 1, 2.$$

Let $\mathbf{J}_1$ and $\mathbf{J}_2$ be selection matrices such that $\mathbf{Z}_1 = \mathbf{Z}\mathbf{J}_1$ and $\mathbf{Z}_2 = \mathbf{Z}\mathbf{J}_2$. Then, the structural model can be rewritten as

$$\begin{aligned} \mathbf{y}_1 &= -\mathbf{y}_2\theta_{21} + \mathbf{Z}\mathbf{J}_1\boldsymbol{\psi}_1 - \boldsymbol{\epsilon}_1 \\ \mathbf{y}_2 &= -\mathbf{y}_1\theta_{12} + \mathbf{Z}\mathbf{J}_2\boldsymbol{\psi}_2 - \boldsymbol{\epsilon}_2. \end{aligned} \tag{A.3}$$

Substitution of (A.2) into (A.3) gives

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{Z}(-\boldsymbol{\psi}_2^*\theta_{21} + \mathbf{J}_1\boldsymbol{\psi}_1) + \boldsymbol{\epsilon}_2^*\theta_{21} - \boldsymbol{\epsilon}_1 \\ \mathbf{y}_2 &= \mathbf{Z}(-\boldsymbol{\psi}_1^*\theta_{12} + \mathbf{J}_2\boldsymbol{\psi}_2) + \boldsymbol{\epsilon}_1^*\theta_{12} - \boldsymbol{\epsilon}_2. \end{aligned} \tag{A.4}$$

By comparing (A.4) with (A.2), we have

$$\begin{aligned} \boldsymbol{\psi}_1^* &= -\boldsymbol{\psi}_2^*\theta_{21} + \mathbf{J}_1\boldsymbol{\psi}_1 \\ \boldsymbol{\psi}_2^* &= -\boldsymbol{\psi}_1^*\theta_{12} + \mathbf{J}_2\boldsymbol{\psi}_2, \end{aligned}$$

which implies the regression equations

$$\begin{aligned} \widehat{\boldsymbol{\psi}}_1^* &= -\widehat{\boldsymbol{\psi}}_2^*\theta_{21} + \mathbf{J}_1\boldsymbol{\psi}_1 + \mathbf{v}_1 \\ \widehat{\boldsymbol{\psi}}_2^* &= -\widehat{\boldsymbol{\psi}}_1^*\theta_{12} + \mathbf{J}_2\boldsymbol{\psi}_2 + \mathbf{v}_2 \end{aligned}$$

with

$$
\begin{aligned}
\mathbf{v}_1 &= (\widehat{\boldsymbol{\psi}}_1^* - \boldsymbol{\psi}_1^*) + (\widehat{\boldsymbol{\psi}}_2^* - \boldsymbol{\psi}_2^*)\theta_{21} \\
\mathbf{v}_2 &= (\widehat{\boldsymbol{\psi}}_2^* - \boldsymbol{\psi}_2^*) + (\widehat{\boldsymbol{\psi}}_1^* - \boldsymbol{\psi}_1^*)\theta_{12}.
\end{aligned}
$$

Let the asymptotic covariance matrix of $\mathbf{v} = (\mathbf{v}_1', \mathbf{v}_2')'$ be denoted by $\boldsymbol{\Omega}$. Let $\boldsymbol{\delta}_1 = (\theta_{21}, \boldsymbol{\psi}_1')'$ and $\boldsymbol{\delta}_2 = (\theta_{12}, \boldsymbol{\psi}_2')'$. Then, the SUR-type AGLS estimator of $\boldsymbol{\delta} = (\boldsymbol{\delta}_1', \boldsymbol{\delta}_2')'$ is given by

$$
\widehat{\boldsymbol{\delta}} = (\widehat{\mathbf{H}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\mathbf{H}})^{-1}\widehat{\mathbf{H}}'\widehat{\boldsymbol{\Omega}}^{-1}\widehat{\boldsymbol{\psi}}^*,
$$

where

$$
\widehat{\mathbf{H}} = \begin{bmatrix} \widehat{\mathbf{H}}_1 & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{H}}_2 \end{bmatrix}
$$

with $\widehat{\mathbf{H}}_1 = [-\widehat{\boldsymbol{\psi}}_2^*, \mathbf{J}_1]$ and $\widehat{\mathbf{H}}_2 = [-\widehat{\boldsymbol{\psi}}_1^*, \mathbf{J}_2]$, and $\widehat{\boldsymbol{\Omega}}$ is a consistent estimator of $\boldsymbol{\Omega}$.

To derive the explicit form of $\boldsymbol{\Omega}$, we need to find the asymptotic covariance matrix of $\widehat{\boldsymbol{\psi}}^*$. The first order condition of the NPL estimator is

$$
\frac{\partial \ln L(\widehat{\boldsymbol{\psi}}_k^*; \widehat{\mathbf{p}})}{\partial \boldsymbol{\psi}_k^*} = \textstyle\sum_{i=1}^n \frac{[d_{ik} - \Phi(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_k^*)]\phi(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_k^*)}{\Phi(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_k^*)[1 - \Phi(\widehat{\mathbf{z}}_i'\widehat{\boldsymbol{\psi}}_k^*)]}\widehat{\mathbf{z}}_i = 0.
$$

A Taylor expansion of the above equation around $\boldsymbol{\psi}^*$ gives

$$
\textstyle\sum_{i=1}^n \frac{(d_{ik} - \Phi_{ik})\phi_{ik}}{\Phi_{ik}(1 - \Phi_{ik})}\mathbf{z}_i - \textstyle\sum_{i=1}^n \frac{\phi_{ik}^2}{\Phi_{ik}(1 - \Phi_{ik})}\mathbf{z}_i[\mathbf{z}_i' + \lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \boldsymbol{\psi}_k^{*\prime}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \boldsymbol{\psi}_k^{*\prime}}](\widehat{\boldsymbol{\psi}}_k^* - \boldsymbol{\psi}_k^*) = O_p(1)
$$

where $\Phi_{ik} = \Phi(\mathbf{z}_i'\boldsymbol{\psi}_k^*)$ and $\phi_{ik} = \phi(\mathbf{z}_i'\boldsymbol{\psi}_k^*)$. As $\mathbf{p}_k = \vec{h}_k(\mathbf{p}) = (\Phi_{1k}, \cdots, \Phi_{nk})'$, the implicit function theorem implies that

$$\begin{aligned}
\frac{\partial \mathbf{p}_1}{\partial \boldsymbol{\psi}_1^{*\prime}} &= \mathbf{S}_1^{-1}\mathbf{D}_1\mathbf{Z} \\
\frac{\partial \mathbf{p}_2}{\partial \boldsymbol{\psi}_2^{*\prime}} &= \mathbf{S}_2^{-1}\mathbf{D}_2\mathbf{Z} \\
\frac{\partial \mathbf{p}_1}{\partial \boldsymbol{\psi}_2^{*\prime}} &= \lambda_{21}^*\mathbf{S}_1^{-1}\mathbf{D}_1\mathbf{W}(\mathbf{I}_n - \lambda_{22}^*\mathbf{D}_2\mathbf{W})^{-1}\mathbf{D}_2\mathbf{Z} \\
\frac{\partial \mathbf{p}_2}{\partial \boldsymbol{\psi}_1^{*\prime}} &= \lambda_{12}^*\mathbf{S}_2^{-1}\mathbf{D}_2\mathbf{W}(\mathbf{I}_n - \lambda_{11}^*\mathbf{D}_1\mathbf{W})^{-1}\mathbf{D}_1\mathbf{Z}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{S}_1 &= \mathbf{I}_n - \lambda_{11}^*\mathbf{D}_1\mathbf{W} - \lambda_{12}^*\lambda_{21}^*\mathbf{D}_1\mathbf{W}(\mathbf{I}_n - \lambda_{22}^*\mathbf{D}_2\mathbf{W})^{-1}\mathbf{D}_2\mathbf{W} \\
\mathbf{S}_2 &= \mathbf{I}_n - \lambda_{22}^*\mathbf{D}_2\mathbf{W} - \lambda_{12}^*\lambda_{21}^*\mathbf{D}_2\mathbf{W}(\mathbf{I}_n - \lambda_{11}^*\mathbf{D}_1\mathbf{W})^{-1}\mathbf{D}_1\mathbf{W}
\end{aligned}$$

with $\mathbf{D}_k = \mathrm{diag}(\phi_{1k}, \cdots, \phi_{nk})$ for $k = 1, 2$. Therefore, by the standard asymptotic theory,

$$\sqrt{n}(\widehat{\boldsymbol{\psi}}_k^* - \boldsymbol{\psi}_k^*) \overset{A}{=} \{\frac{1}{n}\sum_{i=1}^n \frac{\phi_{ik}^2}{\Phi_{ik}(1-\Phi_{ik})}\mathbf{z}_i[\mathbf{z}_i' + \lambda_{1k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_1}{\partial \boldsymbol{\psi}_k^{*\prime}} + \lambda_{2k}^*\mathbf{w}_i\frac{\partial \mathbf{p}_2}{\partial \boldsymbol{\psi}_k^{*\prime}}]\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{(d_{ik}-\Phi_{ik})\phi_{ik}}{\Phi_{ik}(1-\Phi_{ik})}\mathbf{z}_i,$$

where $\mathbf{a} \overset{A}{=} \mathbf{b}$ means vectors $\mathbf{a}$ and $\mathbf{b}$ have the same asymptotic distribution. The asymptotic covariance matrix of $\widehat{\boldsymbol{\psi}}^* = (\widehat{\boldsymbol{\psi}}_1^{*\prime}, \widehat{\boldsymbol{\psi}}_2^{*\prime})'$ is

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{12}' & \mathbf{V}_{22} \end{bmatrix},$$

where

$$\mathbf{V}_{kl} = [\mathbf{Z}'\mathbf{A}_k(\mathbf{Z} + \lambda_{1k}^*\mathbf{W}\frac{\partial \mathbf{p}_1}{\partial \boldsymbol{\psi}_k^{*\prime}} + \lambda_{2k}^*\mathbf{W}\frac{\partial \mathbf{p}_2}{\partial \boldsymbol{\psi}_k^{*\prime}})]^{-1}\mathbf{Z}'\mathbf{B}_{kl}\mathbf{Z}[(\mathbf{Z} + \lambda_{1l}^*\mathbf{W}\frac{\partial \mathbf{p}_1}{\partial \boldsymbol{\psi}_l^{*\prime}} + \lambda_{2l}^*\mathbf{W}\frac{\partial \mathbf{p}_2}{\partial \boldsymbol{\psi}_l^{*\prime}})'\mathbf{A}_l\mathbf{Z}]^{-1},$$

with $\mathbf{A}_k = \text{diag}_{i=1}^n(\frac{\phi_{ik}^2}{\Phi_{ik}(1-\Phi_{ik})})$ and $\mathbf{B}_{kl} = \text{diag}_{i=1}^n(\frac{\phi_{ik}\phi_{il}\mathbb{E}[(d_{ik}-\Phi_{ik})(d_{il}-\Phi_{il})]}{\Phi_{ik}\Phi_{il}(1-\Phi_{ik})(1-\Phi_{il})})$, for $k = 1, 2$. Note $\mathbb{E}[(d_{i1} - \Phi_{i1})^2] = \Phi_{i1}(1 - \Phi_{i1})$, $\mathbb{E}[(d_{i2} - \Phi_{i2})^2] = \Phi_{i2}(1 - \Phi_{i2})$, and

$$\mathbb{E}[(d_{i1} - \Phi_{i1})(d_{i2} - \Phi_{i2})] = \sum_{t=1}^{2}\sum_{s=1}^{2} t^* s^* \Phi(t^*\mathbf{z}_i'\boldsymbol{\psi}_1^*)\Phi(s^*\mathbf{z}_i'\boldsymbol{\psi}_2^*)\Phi_2(-t^*\mathbf{z}_i'\boldsymbol{\psi}_1^*, -s^*\mathbf{z}_i'\boldsymbol{\psi}_2^*, t^*s^*\sigma_{12}^*)$$

where $t^* = 2t - 3$, $s^* = 2s - 3$, and $\Phi_2(\cdot, \cdot, \sigma_{12}^*)$ denotes the standardized bivariate normal distribution function with correlation coefficient σ_{12}^*. The unknown tetrachoric correlation coefficient σ_{12}^* can be estimated by the bivariate probit ML method (see, e.g., Greene, 2012). Then, the asymptotic covariance matrix of $\mathbf{v} = (\mathbf{v}_1', \mathbf{v}_2')'$ is

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{12}' & \boldsymbol{\Omega}_{22} \end{bmatrix}$$

where $\boldsymbol{\Omega}_{11} = \mathbf{V}_{11} + \theta_{21}^2\mathbf{V}_{22} + \theta_{21}(\mathbf{V}_{12} + \mathbf{V}_{12}')$, $\boldsymbol{\Omega}_{22} = \mathbf{V}_{22} + \theta_{12}^2\mathbf{V}_{11} + \theta_{12}(\mathbf{V}_{12} + \mathbf{V}_{12}')$, and $\boldsymbol{\Omega}_{12} = \theta_{12}\mathbf{V}_{11} + \theta_{21}\mathbf{V}_{22} + \mathbf{V}_{12} + \theta_{12}\theta_{21}\mathbf{V}_{12}'$.

B Proof

Proof of Proposition 2.1. As $\vec{h}(\cdot)$ is continuous Assumption 1 (i), it follows by the Brouwer fixed-point theorem that $\mathbf{p} = \vec{h}(\mathbf{p})$ has at least one solution. By the contraction mapping theorem, $\mathbf{p} = \vec{h}(\mathbf{p})$ has a unique solution if $||\partial\vec{h}(\mathbf{p})/\partial\mathbf{p}'|| < 1$ for some matrix norm $||\cdot||$.

$$\frac{\partial\vec{h}(\mathbf{p})}{\partial\mathbf{p}'} = \begin{bmatrix} \partial\vec{h}_1(\mathbf{p})/\partial\mathbf{p}_1' & \cdots & \partial\vec{h}_1(\mathbf{p})/\partial\mathbf{p}_m' \\ \vdots & & \vdots \\ \partial\vec{h}_m(\mathbf{p})/\partial\mathbf{p}_1' & \cdots & \partial\vec{h}_m(\mathbf{p})/\partial\mathbf{p}_m' \end{bmatrix},$$

where

$$\frac{\partial \vec{h}_k(\mathbf{p})}{\partial \mathbf{p}_l'} = \lambda_{lk}^* \begin{bmatrix} w_{11} f_k(u_{1k}) & \cdots & w_{1n} f_k(u_{1k}) \\ \vdots & & \vdots \\ w_{n1} f_k(u_{nk}) & \cdots & w_{nn} f_k(u_{nk}) \end{bmatrix}.$$

It follows that

$$\left\| \frac{\partial \vec{h}(\mathbf{p})}{\partial \mathbf{p}'} \right\|_\infty \leq \max_{k=1,\cdots,m} \sum_{l=1}^m |\lambda_{lk}^*| \max_{i=1,\cdots,n} \sum_{j=1}^m |w_{ij}| \max_k \sup_u f_k(u) = ||\mathbf{\Lambda}^*||_1 ||\mathbf{W}||_\infty \max_k \sup_u f_k(u),$$

and

$$\left\| \frac{\partial \vec{h}(\mathbf{p})}{\partial \mathbf{p}'} \right\|_1 \leq \max_{l=1,\cdots,m} \sum_{k=1}^m |\lambda_{lk}^*| \max_{j=1,\cdots,n} \sum_{i=1}^m |w_{ij}| \max_k \sup_u f_k(u) = ||\mathbf{\Lambda}^*||_\infty ||\mathbf{W}||_1 \max_k \sup_u f_k(u).$$

Hence, the desired result follows if Assumption 1 (ii) holds. $\square$

Table 1: Monte Carlo Simulation Results (Circular Network)

		$\theta_{21} = 0.5$	$\lambda_{11} = 0.9$	$\lambda_{21} = 0.6$	$\beta_{11} = 1.0$	$\beta_{21} = 1.0$
$n = 500$						
$\sigma_{12}^{*} = 0.1$	AGLS-1	0.497(0.107)	0.928(0.307)	0.597(0.383)	1.026(0.137)	1.024(0.193)
	AGLS-2	0.501(0.105)	0.921(0.294)	0.604(0.368)	1.030(0.126)	1.026(0.154)
$\sigma_{12}^{*} = 0.5$	AGLS-1	0.497(0.114)	0.925(0.319)	0.597(0.401)	1.028(0.140)	1.025(0.203)
	AGLS-2	0.502(0.111)	0.916(0.305)	0.607(0.382)	1.032(0.130)	1.027(0.153)
$\sigma_{12}^{*} = 0.9$	AGLS-1	0.495(0.115)	0.927(0.331)	0.592(0.415)	1.029(0.143)	1.025(0.205)
	AGLS-2	0.499(0.113)	0.916(0.319)	0.603(0.401)	1.033(0.132)	1.027(0.155)
$n = 1000$						
$\sigma_{12}^{*} = 0.1$	AGLS-1	0.501(0.074)	0.906(0.215)	0.607(0.273)	1.015(0.090)	1.018(0.134)
	AGLS-2	0.502(0.071)	0.907(0.202)	0.606(0.258)	1.018(0.086)	1.015(0.103)
$\sigma_{12}^{*} = 0.5$	AGLS-1	0.503(0.078)	0.906(0.231)	0.608(0.291)	1.014(0.094)	1.017(0.138)
	AGLS-2	0.504(0.075)	0.905(0.213)	0.608(0.273)	1.017(0.087)	1.014(0.103)
$\sigma_{12}^{*} = 0.9$	AGLS-1	0.501(0.080)	0.906(0.244)	0.607(0.307)	1.014(0.096)	1.015(0.147)
	AGLS-2	0.503(0.076)	0.903(0.223)	0.609(0.283)	1.016(0.089)	1.014(0.102)
$n = 2000$						
$\sigma_{12}^{*} = 0.1$	AGLS-1	0.498(0.053)	0.903(0.154)	0.597(0.194)	1.006(0.064)	1.007(0.096)
	AGLS-2	0.499(0.051)	0.901(0.145)	0.600(0.182)	1.007(0.060)	1.008(0.074)
$\sigma_{12}^{*} = 0.5$	AGLS-1	0.497(0.054)	0.904(0.161)	0.595(0.203)	1.006(0.065)	1.006(0.101)
	AGLS-2	0.498(0.052)	0.901(0.151)	0.599(0.190)	1.007(0.062)	1.008(0.073)
$\sigma_{12}^{*} = 0.9$	AGLS-1	0.496(0.054)	0.905(0.165)	0.594(0.209)	1.006(0.068)	1.008(0.104)
	AGLS-2	0.497(0.052)	0.903(0.153)	0.595(0.195)	1.008(0.063)	1.007(0.075)

Mean(SD)

Table 2: Monte Carlo Simulation Results (Random Network)

		$\theta_{21} = 0.5$	$\lambda_{11} = 0.9$	$\lambda_{21} = 0.6$	$\beta_{11} = 1.0$	$\beta_{21} = 1.0$
$n = 500$						
$\sigma_{12}^{*} = 0.1$	AGLS-1	0.495(0.081)	0.932(0.400)	0.588(0.424)	1.023(0.120)	1.017(0.271)
	AGLS-2	0.498(0.081)	0.926(0.360)	0.594(0.384)	1.024(0.116)	1.021(0.200)
$\sigma_{12}^{*} = 0.5$	AGLS-1	0.495(0.082)	0.930(0.417)	0.589(0.442)	1.025(0.126)	1.017(0.281)
	AGLS-2	0.498(0.081)	0.926(0.371)	0.594(0.398)	1.026(0.121)	1.019(0.198)
$\sigma_{12}^{*} = 0.9$	AGLS-1	0.496(0.083)	0.933(0.438)	0.586(0.463)	1.027(0.129)	1.019(0.292)
	AGLS-2	0.498(0.083)	0.932(0.376)	0.585(0.399)	1.029(0.124)	1.016(0.188)
$n = 1000$						
$\sigma_{12}^{*} = 0.1$	AGLS-1	0.502(0.057)	0.898(0.277)	0.614(0.294)	1.005(0.085)	1.016(0.186)
	AGLS-2	0.503(0.057)	0.903(0.248)	0.609(0.263)	1.006(0.080)	1.010(0.139)
$\sigma_{12}^{*} = 0.5$	AGLS-1	0.501(0.059)	0.896(0.291)	0.615(0.309)	1.005(0.088)	1.015(0.197)
	AGLS-2	0.502(0.059)	0.902(0.258)	0.608(0.272)	1.007(0.083)	1.008(0.137)
$\sigma_{12}^{*} = 0.9$	AGLS-1	0.500(0.060)	0.901(0.304)	0.608(0.325)	1.005(0.089)	1.013(0.208)
	AGLS-2	0.501(0.059)	0.907(0.265)	0.601(0.281)	1.008(0.084)	1.005(0.133)
$n = 2000$						
$\sigma_{12}^{*} = 0.1$	AGLS-1	0.500(0.040)	0.885(0.203)	0.618(0.215)	1.003(0.058)	1.012(0.130)
	AGLS-2	0.500(0.040)	0.886(0.185)	0.617(0.198)	1.004(0.056)	1.010(0.094)
$\sigma_{12}^{*} = 0.5$	AGLS-1	0.500(0.041)	0.884(0.215)	0.619(0.227)	1.004(0.061)	1.013(0.135)
	AGLS-2	0.500(0.041)	0.889(0.190)	0.615(0.203)	1.005(0.058)	1.008(0.093)
$\sigma_{12}^{*} = 0.9$	AGLS-1	0.499(0.041)	0.883(0.229)	0.619(0.244)	1.004(0.064)	1.011(0.143)
	AGLS-2	0.500(0.041)	0.885(0.199)	0.617(0.215)	1.005(0.061)	1.008(0.091)

Mean(SD)

Table 3: Data Summary

	Definition	Mean	SD
Dependent variables			
Smoking	1 if smokes once a week or more in the last 12 months	0.17	0.37
Sports	1 if participates in cheerleading/dance, baseball/softball, basketball, field hockey, football, ice hockey, soccer, swimming, tennis, track, volleyball, wrestling, and/or other sport team this year	0.56	0.50
Control variables			
Age	Age	15.09	1.67
Years in school	Years in current school	2.53	1.43
Female	1 if female	0.52	0.50
(White)	1 if White American	0.59	0.49
African American	1 if African American	0.15	0.36
Other races	1 if race is not White or African American	0.26	0.44
Health	1 if health is excellent	0.31	0.46
Live with both parents	1 if live with both parents	0.74	0.44
Parental care	1 if parents care about the student very much	0.83	0.37
(Parent education: less than HS)	1 if parent's education is less than high school	0.11	0.31
Parent education: HS grad	1 if parent's education is high school or higher but no college degree	0.43	0.49
Parent education: college grad	1 if parent's education is college or higher	0.31	0.46
Parent education: missing	1 if parent's education level is unknown	0.16	0.37
(Parent does not work for pay)	1 if parent does not work for pay	0.05	0.22
Parent works for pay	1 if parent works for pay	0.87	0.34
Parent works for pay: missing	1 if parent's work status is unknown	0.08	0.27
(Parents do not smoke)	1 if neither parents smoke	0.06	0.24
Parent smokes	1 if either parent smokes	0.11	0.32
Parent smokes: missing	1 if "parent smokes" is missing	0.82	0.38
(No cigarettes in the home)	1 if cigarettes are not easily available at home	0.13	0.33
Cigarettes in the home	1 if cigarettes are easily available at home	0.05	0.23
Cigarettes in the home: missing	1 if "cigarettes in the home" is missing	0.82	0.38
Coordination	1 if well coordinated	0.35	0.48

The variable in the parentheses is the reference category.
If both parents are in the household, the education and work status of the father is considered.
The data includes 141 schools with 71,918 observations.

Table 4: Estimation of the Binary Choice Network Model

	Single Equation		**Simultaneous Equations**	
	Smoking	Sports	Smoking	Sports
Within-activity peer effect (λ_{kk})	1.7353***	1.0214***	1.2639***	0.8359***
	(0.0850)	(0.0703)	(0.1206)	(0.0784)
Cross-activity peer effect (λ_{lk})			-0.2390***	-1.0845***
			(0.1014)	(0.1435)
Simultaneity effect ($-\theta_{lk}$)			-0.1543***	-0.0737
			(0.0513)	(0.0659)
Age	0.1141***	-0.0592***	0.0957***	-0.0533***
	(0.0057)	(0.0053)	(0.0068)	(0.0085)
Years in school	-0.0241***	-0.0016	-0.0259***	-0.0058
	(0.0061)	(0.0051)	(0.0061)	(0.0054)
Female	-0.0472***	-0.2217***	-0.0889***	-0.2256***
	(0.0137)	(0.0113)	(0.0186)	(0.0119)
African American	-0.6711***	0.3159***	-0.6190***	0.2588***
	(0.0289)	(0.0210)	(0.0329)	(0.0490)
Other races	-0.0974***	0.0490***	-0.0939***	0.0397***
	(0.0167)	(0.0139)	(0.0167)	(0.0154)
Health	-0.5224***	0.2790***	-0.4697***	0.2304***
	(0.0151)	(0.0114)	(0.0226)	(0.0354)
Live with both parents	-0.1585***	0.0374***	-0.1557***	0.0168
	(0.0148)	(0.0123)	(0.0148)	(0.0162)
Parental care	-0.2646***	0.0503***	-0.2551***	0.0243
	(0.0164)	(0.0146)	(0.0166)	(0.0226)
Parent education: HS grad	-0.0507***	0.0788***	-0.0378*	0.0708***
	(0.0206)	(0.0174)	(0.0209)	(0.0177)
Parent education: college grad	-0.1491***	0.2153***	-0.1148***	0.1923***
	(0.0227)	(0.0189)	(0.0251)	(0.0215)
Parent education: missing	-0.1277***	-0.0105	-0.1322***	-0.0249
	(0.0248)	(0.0203)	(0.0247)	(0.0222)
Parent works for pay	-0.0015	0.1633***	0.0245	0.1653***
	(0.0268)	(0.0223)	(0.0281)	(0.0222)
Parent works for pay: missing	-0.0847***	0.2123***	-0.0487	0.2077***
	(0.0358)	(0.0294)	(0.0375)	(0.0299)
Parent smokes	0.1635***		0.1607***	
	(0.0339)		(0.0334)	
Parent smokes: missing	0.0483		0.0448	
	(0.0965)		(0.0948)	
Cigarettes in the home	0.2386***		0.2382***	
	(0.0316)		(0.0312)	
Cigarettes in the home: missing	0.1071		0.1098	
	(0.0939)		(0.0921)	
Coordination		0.2442***		0.2418***
		(0.0110)		(0.0114)
σ^*_{12}			-0.0923***	
			(0.0077)	
Contextual effects	Yes	Yes	Yes	Yes
School fixed effects	Yes	Yes	Yes	Yes

To save space, estimates of contextual effects and school fixed effects are not reported.
Standard errors in parentheses. Statistical significance: ***p<0.01; **p<0.05; *p<0.1.