Count Data Models with Social Interactions Under Rational Expectations

Elysée Aristide Houndetoungan*

This version: March 31, 2022

Please download the latest version at ahoundetoungan.com

Abstract

I propose a peer effect model for counting variables using a game of incomplete information. I provide sufficient conditions under which the game equilibrium is unique and estimate the model parameters using a Nested Partial Likelihood (NLP) approach. I generalize the NPL estimator to the case of endogenous networks and establish its asymptotic properties. I show that estimating peer effects on counting variables using models that ignore the counting nature of the outcome, such as the Tobit model, leads to inconsistent estimators. I use the model to evaluate peer effects on students' participation in extracurricular activities.

Keywords: Discrete model, Social networks, Bayesian game, Rational expectations, Network formation.

JEL Classification: C25, C31, C73, D84, D85.

---

*Thema, Cy Cergy Paris Université; Email: aristide.houndetoungan@cyu.fr.

I would like to thank Vincent Boucher for his helpful comments and insights. I would also like to thank Bernard Fortin, Yann Bramoullé, Arnaud Dufays, Luc Bissonnette, and Marion Goussé for helpful comments and discussions.

I provide an easy-to-use R package—named CDatanet—for implementing the model and methods used in this paper. The package is located at `https://github.com/ahoundetoungan/CDatanet`.

This research uses data from Add Health, a program 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 given to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain Add Health data files is available on the Add Health website (www.cpc.unc.edu/addhealth). No direct support was received from Grant P01-HD31921 for this research.

1 Introduction

There is a large and growing literature on peer effects in economics.[1] Recent contributions include, among others, models for limited dependent variables, including binary (e.g., Brock and Durlauf, 2001; Lee et al., 2014; Liu, 2019), ordered (e.g., Liu and Zhou, 2017), multinomial (e.g., Guerra and Mohnen, 2020), and censored (e.g., Xu and Lee, 2015b) variables. However, to the best of my knowledge, there are no existing models for counting variables with microeconomic foundations, despite these variables being prevalent in survey data (e.g., number of physician visits, frequency of consumption of a good/service, frequency of participation in an activity). Peer effects on those variables are often estimated using a linear-in-means model or a binary model after transforming the outcome into binary data (e.g., Liu et al., 2012; Patacchini and Zenou, 2012; Fujimoto and Valente, 2013; Liu et al., 2014; Fortin and Yazbeck, 2015; Boucher, 2016; Lee et al., 2020).

In both cases, the estimation strategy ignores the counting nature of the dependent variable. In the case of the linear-in-means model, this raises a microfoundation issue. The structural framework behind the linear-in-means model assumes a continuous outcome (see Ballester et al., 2006; Calvó-Armengol et al., 2009; Liu, 2019). Assuming a discrete outcome in the same framework would be the source of a multiple equilibria issue. Because this framework only supports continuous data, there is some doubt about what is being estimated from counting data using such an approach. On the other hand, transforming the outcome into binary data does not allow a peer effect interpretation in terms of intensive margin effects but only as extensive margin effects (e.g., Lee et al., 2014; Liu, 2019).

In this paper, I propose a network model under rational expectations (RE), in which the outcome is a counting variable. I show that the model's parameters can be estimated using the Nested Partial Likelihood (NPL) method proposed by Aguirregabiria and Mira (2007). I generalize this estimation strategy to the case where the network is endogenous. I show that estimating peer effects on counting variables using models that ignore the counting nature of the outcome, such as the spatial autoregressive (SAR) model (Lee, 2004; Bramoullé et al., 2009) or the SAR Tobit (SART) model (Xu and Lee, 2015b), leads to inconsistent estimators. I estimate peer effects on the number of extracurricular activities in which students are enrolled using the data set provided by the National Longitudinal Study of Adolescent Health (Add Health). Finally, I provide an easy-to-use R package—named `CDatanet`—for implementing the model.[2]

The model is based on a static game with incomplete information (see Harsanyi, 1967; Osborne and Rubinstein, 1994) similar to that of the linear models (e.g., Ballester et al., 2006; Calvó-Armengol et al., 2009; Blume et al., 2015; Liu, 2019). The assumption of incomplete information is extensively

[1] For recent reviews, see De Paula (2017) and Bramoullé et al. (2020).

[2] The package is available at github.com/ahoundetoungan/CDatanet.

considered in the literature on peer effect models for discrete outcomes (e.g., Brock and Durlauf, 2001; Bajari et al., 2010; Lee et al., 2014; Liu, 2019; Yang and Lee, 2017; Guerra and Mohnen, 2020; Boucher and Bramoullé, 2020). Although this assumption is well suited to many empirical applications, it also implies a unique game equilibrium under weak conditions. Individuals in the game interact through a directed network, simultaneously choose their strategy, and receive a payoff that depends on their belief about the choice of their peers. However, unlike the linear models, which assume a linear-quadratic payoff, the counting nature of the outcome allows for dealing with a more flexible payoff. Note that the linear-quadratic payoff is only used because it leads to an easy-to-estimate linear reduced form. I show that this linear-quadratic payoff implies a strong econometric restriction in the case of counting variables and leads to an inconsistent estimator of peer effects.

I provide sufficient conditions under which the game has a unique Bayesian Nash Equilibrium (BNE). However, the econometric specification of the model raises an identification issue. Parameter identification is generally established by setting a rank condition on the matrix of explanatory variables (e.g., Lee, 2004; Yang and Lee, 2017). In the case of rational expectation models, this matrix contains the expected average outcome, which is an unobserved variable. Therefore, the rank condition cannot be verified empirically. Using the fact that the expected outcome is unbounded in the case of counting variables, I provide verifiable identification conditions. In particular, I show that the parameters are identified under the same conditions set by Bramoullé et al. (2009) for linear-in-means models.

I show that the model parameters can be estimated using the NPL algorithm proposed by Aguirregabiria and Mira (2007). I generalize this algorithm to the case where the network is endogenous. Endogeneity is due to unobservable individual attributes, which influence both link formation in the network and the outcome (see Johnsson and Moon, 2015; Graham, 2017). To control for the endogeneity, I use a two-stage estimation strategy. In the first stage, I consider a dyadic linking model in which the probability of link formation between two agents depends on, among other things, the unobservable attributes. Using a Gibbs sampler, I estimate the posterior distribution of those unobservable attributes. In the second stage, the estimator of the unobservable attributes is included in the count data model as supplementary explanatory variables. However, because I use the estimate of the unobservable attributes and not the true unobservable attributes, this complicates the asymptotic of the NPL estimator at the second stage. To circumvent this issue, I establish a new limiting distribution that accounts for the uncertainty related to the first-stage estimation. I assess the finite sample performance of the estimation strategy using Monte Carlo simulations.

I provide an empirical application. I use the Add Health data set to estimate peer effects on the number of extracurricular activities in which students are enrolled. Participation in extracurricular activities is associated with positive educational, social, and developmental outcomes such as increased achievement, improved interpersonal skills, reduced levels of delinquency, reduced likelihood of dropping out, and improved self-esteem (see Holland and Andre, 1987; McNeal Jr, 1999; Darling, 2005). Controlling for network endogeneity, I find that increasing the expected number of activities in which a student's friends are enrolled by one implies an increase in the expected number of activities in which the student is enrolled by 0.256. As in the Monte Carlo study, I also find that the SART model overestimates these marginal peer effects, as the corresponding result is given by 0.325.

This paper contributes to the literature on social interaction models for limited dependent variables by being the first to deal with counting outcomes. The existing models deal with binary (e.g., Brock and Durlauf, 2001; Soetevent and Kooreman, 2007; Lee et al., 2014; Xu and Lee, 2015a; Liu, 2019; Boucher and Bramoullé, 2020), censored (e.g., Xu and Lee, 2015b), ordered (e.g., Liu and Zhou, 2017), and multinomial outcomes (e.g., Guerra and Mohnen, 2020). Moreover, my model generalizes the rational expectation model for binary data developed by Lee et al. (2014). When the outcome is bounded and only takes two values, I show that the structure of my model game and the BNE are similar to those of Lee et al. (2014).
Importantly, in the literature on spatial autoregressive models for limited dependent variables, cases of count data have been studied (e.g, Karlis, 2003; Liesenfeld et al., 2016; Inouye et al., 2017; Glaser, 2017). These papers consider reduced form equations in which the dependent counting variable is spatially autocorrelated. However, the models are not based on any process (game) that explains how the individuals choose their strategy and thus how they are influenced by their peers. Therefore, the reduced form cannot be interpreted as a best-response function, and the spatial dependence parameter cannot be interpreted as peer effects.

The paper contributes to the literature on peer effect models with endogenous networks. Goldsmith-Pinkham and Imbens (2013) as well as Hsieh and Lee (2016) consider a Bayesian hierarchical model to control for endogeneity. They use an MCMC approach to jointly simulate from the posterior distribution of the network formation model parameters and the outcome model parameters. Although this method would be more efficient as the estimation is done in a single step, it can be cumbersome to implement with a discrete data model. Johnsson and Moon (2015) also develop a strategy to estimate the linear-in-means peer effect model by controlling for the endogeneity of the network. Their estimation method is semiparametric and relies on a control function approach. My strategy to control for endogeneity can be readily implemented with discrete outcome models, since the network formation model is estimated, in a first stage, separately from the outcome model estimation. I establish a new limiting distribution that takes into account the uncertainty of the estimation in the first stage.

The paper also contributes to the extensive empirical literature on social interactions. Existing papers studying peer effects using count data rely on linear-in-means models estimated by the maximum likelihood approach of Lee (2004) or the two-stage least squares method of Kelejian and Prucha (1998) that ignores the counting nature of the outcome (e.g., Liu et al., 2012; Patacchini and Zenou, 2012;

Fujimoto and Valente, 2013; Liu et al., 2014; Fortin and Yazbeck, 2015; Boucher, 2016; Lee et al., 2020). I show that peer effects estimated in this way are inconsistent. My empirical application to students' participation in extracurricular activities accounts for the counting nature of the outcome.

The remainder of the paper is organized as follows. Section 2 presents the microeconomic foundation of the model on the basis of an incomplete information network game. Section 3 addresses the identification and estimation of the model parameters. Section 4 documents the Monte Carlo experiments. Section 5 presents the empirical results and the method used to control for the endogeneity of the network. Section 6 concludes this paper.

2 Microeconomic Foundations

This section presents the microfoundations of the model. Let $\mathcal{V} = \{1, \dots, n\}$ be a population of n agents partitioned into M sub-groups $\mathcal{V}^1, \dots, \mathcal{V}^M$ with n_m the size of the m-th subgroup. Agents' choice is denoted by $y_i \in \mathbb{N}$, an integer variable also called a *counting variable* (e.g., the number of cigarettes smoked per day or per week). Let $s(i)$ be the group of individual i (observable by all agents and the econometrician). Agents interact through a directed network. Let $\mathbf{G} = [g_{ij}]$ be an $n \times n$ adjacency matrix, where the (i, j)-th element is non-negative and captures the proximity of the individuals i and j in the network. Interactions are restricted to individuals from the same group.[3] I define the peers of individual i as the set of individuals $\mathcal{V}_i = \{j, g_{ij} > 0\}$. By convention, nobody interacts with himself/herself, that is $g_{ii} = 0 \ \forall \ i \in \mathcal{V}$.

2.1 Incomplete Information Network Game

I use a game of incomplete information to rationalize the model (see Osborne and Rubinstein, 1994). Agents act noncooperatively. As a common assumption in the literature, agent i's decision is influenced by their own observable characteristics, denoted ψ_i (eventually their peers' observable characteristics), unobservable individual characteristics interpreted as the agent's type (private information), and other individuals' choice (see e.g., Brock and Durlauf, 2001; Bajari et al., 2010; Yang and Lee, 2017; De Paula, 2017).[4] Specifically, following Brock and Durlauf (2001, 2007), I assume that individual preferences

[3] Such a restriction is known as *maximality* (see Calvó-Armengol et al., 2009; Lee et al., 2014; Liu, 2019).

[4] It is well known that when agent's type is observed by other players (complete information), the game equilibrium is not unique, especially when the outcome is discrete. Multiple equilibria is a challenging issue both theoretically and empirically (see De Paula, 2013). The assumption of incomplete information is interesting as it implies a unique equilibrium under reasonable conditions. This assumption is extensively considered in the literature (see e.g., Brock and Durlauf, 2001; Bajari et al., 2010; Lee et al., 2014; Liu, 2019; Yang and Lee, 2017; Guerra and Mohnen, 2020). It also suits well many empirical applications like the one I present in Section 5.

about the choice of y_i are described by an additive discrete payoff function defined by

$$U\left(y_i, \mathbf{y}_{-i}\right) = \underbrace{\psi_i y_i - c(y_i)}_{\text{private sub-payoff}} - \underbrace{\frac{\lambda}{2}\left(y_i - \bar{y}_i\right)^2}_{\text{social cost}} + \underbrace{e_i(y_i)}_{\text{type}}, \quad (1)$$

where $\lambda \geqslant 0$, $\mathbf{y}_{-i} = (y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$, and $\bar{y}_i = \sum_{j \in \mathcal{V}_i} g_{ij} y_j$.[5] Throughout the paper, the vector's subscript $-i$ is used to denote the vector excluding the i-th component. In the payoff (1), the term $\psi_i y_i - c(y_i)$ is a private subpayoff that depends on individual choice y_i and on individual observable characteristics ψ_i.[6] $c(y_i)$ is the cost associated with the choice of y_i. I let the cost function $c(.)$ be flexible. The cost function is generally defined as a quadratic function in many peer effect models (e.g., Ballester et al., 2006; Calvó-Armengol et al., 2009; Blume et al., 2015; Liu, 2019). As shown in Section 3.1, a quadratic cost function implies a strong restriction on the econometric model. The term $\frac{\lambda}{2}\left(y_i - \bar{y}_i\right)^2$ is a social cost that increases with the gap between the agent and peers' choices. Such a specification of the social cost implies conformist preferences (see Akerlof, 1997).

Agent's type is described by $\left(e_i(r)\right)_{r \in \mathbb{N}}$, a sequence of random variables. Each agent observes their own type; that is, i observes $e_i(r)$ for any $r \in \mathbb{N}$; however, they do not observe others' type and therefore, they do not observe the others' choice $\mathbf{y}_{-i}$. Under this consideration, the agent maximizes, not the random payoff (1), but its expectation, where the expectation is taken with respect to their beliefs over $\mathbf{y}_{-i}$. As common in Bayesian game literature, I assume that the private information, $e_i(r)$, $r \in \mathbb{N}$, is distributed identically among agents and that this distribution is common knowledge to all the agents (see e.g., Brock and Durlauf, 2001; Bajari et al., 2010; Lee, 2004; Yang and Lee, 2017). Thus, agents form rational expectations; that is, their expectation of the payoff is the *true* mathematical expectation and can be expressed as

$$U^e\left(y_i, \mathbf{y}_{-i}\right) = \psi_i y_i - c(y_i) - \frac{\lambda}{2}\,\mathbb{E}_{\mathbf{y}_{-i}}\left[\,\left(y_i - \bar{y}_i\right)^2\,\right] + e_i(y_i). \quad (2)$$

As Equation (2) is only defined for non-negative y_i, I set, by convention, that $c(-1) = +\infty$, which implies $U^e\left(-1, \mathbf{y}^e_{-i}\right) = -\infty$. This will be helpful to simplify many equations.[7]

To show that there is a unique count choice y_i that maximizes the expected payoff (2), I restrict the game to some representations of the payoff terms. These representations make the model tractable both theoretically and econometrically. The first restriction is about the cost function.

[5] The use of additive payoff has been a popular simplification in discrete choice literature since McFadden (1973).

[6] In the econometric model, $\psi_i = \alpha_{s(i)} + \mathbf{x}_i'\boldsymbol{\beta} + \bar{\mathbf{x}}_i'\boldsymbol{\gamma}$, where $\mathbf{x}_i$ and $\bar{\mathbf{x}}_i$ are vectors of observable individual-specific characteristics (control variables) and peers' average characteristics, respectively, $\alpha_{s(i)}$ is a group-specific effect, and $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ are unknown parameters.

[7] Note also that since the strategy space of y_i is $\mathbb{N}$, the expectation $\mathbb{E}_{\mathbf{y}_{-i}}\left[\,(y_i - \bar{y}_i)^2\,\right]$ involves an infinite summation that may not be finite. Under Assumptions 2.1 and 2.2 stated later, all infinite summations used in the paper are finite (see Appendix A.2).

Assumption 2.1. *$c(.)$ is a strictly increasing and convex function.*

The assumption of convex cost function means increasing difference in the cost: $\Delta c(r+1) - \Delta c(r) \geqslant 0$, $\forall\ r \in \mathbb{N}$, where Δ is the first difference operator; i.e., for any sequence $(b_r)_r$, $\Delta b_r = b_r - b_{r-1}$. Under conditions set on the distribution of agent's type in Assumption 2.2, the convexity of the cost function implies a strictly concave expected payoff in y_i. As discussed in Section 3.3, the assumption of a convex cost function can be more flexible and generalized to a larger class of functions. Moreover, note that Assumption 2.1 is weaker than the linear-quadratic payoff function broadly imposed in the literature in the case of linear models (see Ballester et al., 2006; Calvó-Armengol et al., 2009; Liu, 2019).

I also set a second restriction on the distribution of the agent's type. As comparisons in discrete games are done using the increase in the payoff for an additional unit of y_i, I set the assumption for the distribution of $\Delta e_i(r) := e_i(r) - e_i(r-1)$, for any $r \in \mathbb{N}^*$.

Assumption 2.2. *For all $i \in \mathcal{V}$, $r \in \mathbb{N}^*$, $e_i(r) = e_i(r-1) + \varepsilon_i$, where ε_i's are independent and identically distributed according to a continuous symmetric distribution with a cumulative distribution function (cdf) F_ε and a probability density function (pdf) f_ε, such that $f_\varepsilon(x) = o\left(|x|^{-\kappa}\right)$ at ∞, for some $\kappa > 3$.*

Assumption 2.2 implies that ε_i, the first difference of $e_i(y_i)$, does not depend on y_i. The agent associates the same information, ε_i, with any additional unit; that is, $e_i(y_i) = \varepsilon_i y_i + e_i(0)$. This condition simplifies the econometric model. Moreover, the assumption that ε_i's are independent is a classic simplification.[8] Assumption 2.2 also states that ε_i has a continuous symmetric distribution and $f_\varepsilon(x) = o\left(|x|^{-\kappa}\right)$ at ∞, for some $\kappa > 3$. The assumption of continuity is necessary so that ε_i has a continuous density function. The symmetry of this density function simplifies many equations. The condition $f_\varepsilon = o(1/x^\kappa)$ at ∞ for some $\kappa > 3$ implies that the probability that y_i takes the value r will decrease exponentially as r grows to infinity.[9] Many usual distributions suit Assumption 2.2 (e.g., normal, logistic, and student). Both Assumptions 2.1 and 2.2 imply that there is a unique count choice that maximizes the payoff.

Proposition 2.1. *Under Assumptions 2.1 and 2.2, there is a unique $r_0 \in \mathbb{N}$ at which $U^e\left(., \mathbf{y}_{-i}\right)$ is maximized. Moreover, $U^e\left(r, \mathbf{y}_{-i}\right) \geqslant \max\left\{U^e\left(r-1, \mathbf{y}_{-i}\right),\ U^e\left(r+1, \mathbf{y}_{-i}\right)\right\}$ iff $r = r_0$.*

Proposition 2.1 implies that the agents' choice $y_i = r$ if and only if $\Delta U^e\left(r+1, \mathbf{y}_{-i}\right) \leqslant 0 \leqslant \Delta U^e\left(r, \mathbf{y}_{-i}\right)$, which is equivalent, by Equation (2), to $-\psi_i - \lambda \bar{y}_i^e + a_r \leqslant \varepsilon_i \leqslant -\psi_i - \lambda \bar{y}_i^e + a_{r+1}$, where $a_r = \Delta c(r) + \lambda r - \dfrac{\lambda}{2}$, $\bar{y}_i^e = \sum_{j \in \mathcal{V}_i} g_{ij} y_j^e$, and y_i^e is the *rational* expected (*true* expectation of the) choice given

[8] Such a restriction can be released because one can account for the correlation between ε_i's as correlated effects (see Manski, 1993).

[9] This condition is important so that the infinite summations defined in the paper (e.g., the expected choice y_i^e and the expected payoff U^e) exist (see Appendix A.2).

the agents' observable characteristics. This characterization of y_i links agent i's decision to a random event. This is useful as it allows to write the probability that $y_i = r$ given the agents' characteristics. Let $p_{ir} = \mathbb{P}(y_i = r|\psi_1,\ \dots,\ \psi_n)$ this probability. Using the symmetry of the distribution of ε_i, p_{ir} can be written as

$$p_{ir} = F_\varepsilon\left(\lambda\bar{y}_i^e + \psi_i - a_r\right) - F_\varepsilon\left(\lambda\bar{y}_i^e + \psi_i - a_{r+1}\right). \tag{3}$$

Equation (3) is similar to the specification of an ordered model (see Amemiya, 1981; Baetschmann et al., 2015). One can get the same characterization by assuming a latent variable $y_i^* = \lambda\bar{y}_i^e + \psi_i + \varepsilon_i$, such that $y_i = r$ if and only if $y_i^* \in (a_r,\ a_{r+1})$. However, the mechanisms behind both specifications are different. In the case of an ordered model, agents choose the latent variable y_i^* and not the counting variable y_i directly. The game is not the one described by the payoff (1). Moreover, unlike a classical ordered model, y_i is unbounded and there is then an infinite number of cut points a_r, $r \in \mathbb{N}$.

Equation (3) also shows how assuming a quadratic cost function implies a strong econometric restriction. As $a_r = \Delta c(r) + \lambda r - \dfrac{\lambda}{2}$, the quadratic cost implies that a_r is linear in r. Put differently, $a_{r+1} - a_r$ is constant $\forall\ r \in \mathbb{N}^*$. This condition is too restrictive empirically. For instance, the ordered model does not set any restriction on the distance between the cut points. This justifies why estimating peer effects on counting variables using a classical SAR or SART model leads to biased estimations. Indeed, these models are based on a game similar to that described by the payoff (1) with a quadratic cost function (see Ballester et al., 2006; Calvó-Armengol et al., 2009; Xu and Lee, 2015b). With a discrete outcome, one can release this restriction and get an econometric model more general than the SAR model.

Equation (3) gives the consistency condition of any rational belief system with respect to the distribution of the agent's type. A belief system $\mathbf{p} = (p_{ir})$ is said to be *rational* or *consistent* (with respect the distribution of ε_i) if and only if it verifies Equation (3), where y_i^e is the expected outcome associated with that belief system and can be written as $y_i^e = \sum_{r=1}^{\infty} r p_{ir}$.

Equation (3) generalizes the case of binary outcomes under RE studied by Lee et al. (2014). To see why, let us consider a particular cost function such that $\Delta c(r) = +\infty$ for any $r \geqslant 2$. This implies that $a_r = +\infty$ for any $r \geqslant 2$. Therefore, $p_{i0} = 1 - F_\varepsilon\left(\lambda\bar{y}_i^e + \psi_i - a_1\right)$ and $p_{i1} = F_\varepsilon\left(\lambda\bar{y}_i^e + \psi_i - a_1\right)$. Thus, y_i only takes the values 0 and 1, almost surely, and the expected choice y_i^e is equal to p_{i1}; i.e., the probability of the event $\{y_i = 1\}$. The condition $p_{i1} = F_\varepsilon\left(\lambda\sum_{j=1}^{n} g_{ij}p_{j1} + \psi_i - a_1\right)$ is the characterization of the rational beliefs in the case of the binary choice.

2.2 Expected Outcome Uniqueness

Proposition 2.1 states that there is a unique Bayesian Nash equilibrium given the expected outcome $\mathbf{y}^e = (y_1^e, \dots, y_n^e)$. However, there may exist more than one expected outcome and belief system $\mathbf{p} = (p_{ir})$, which verifies the RE condition (3). As $y_i^e = \sum_{r=1}^{\infty} r p_{ir}$, Equation (3) can also be expressed as $\mathbf{p} = \mathbf{H}(\mathbf{p})$, where $\mathbf{H}$ is some mapping defined from $[0,1]^{\infty}$ to itself. Finding a rational belief system amounts to computing the fixed points of $\mathbf{H}$. The existence of $\mathbf{p}$ is guaranteed by Schauder's fixed point theorem.[10] However, because $\mathbf{H}$ is defined in an unbounded dimensional space, the uniqueness of $\mathbf{p}$ cannot be established by the contraction mapping theorem as in other discrete models under RE (e.g., Brock and Durlauf, 2001; Lee et al., 2014; Guerra and Mohnen, 2020).

Equation (3) also implies that knowledge of the rational expected outcome $\mathbf{y}^e$ is sufficient to compute the underlying rational beliefs $\mathbf{p}$ and vice versa. This has a very useful implication: if the rational expected outcome $\mathbf{y}^e$ is unique, then the rational belief system is also unique. Moreover, because the expected outcome $\mathbf{y}^e$ is a finite-dimensional vector, this result simplifies the proof of a unique consistent belief system. I show that the rational expected outcome also verifies a fixed point equation as stated by the following proposition.

Proposition 2.2. *Let $\mathbf{L}(\mathbf{y}^e) = (\ell_1(\mathbf{y}^e)\ \dots\ \ell_n(\mathbf{y}^e))'$, where $\ell_i(\mathbf{y}^e) = \sum_{r=1}^{\infty} F_{\varepsilon}(\lambda \bar{y}_i^e + \psi_i - a_r)$ for all $i \in \mathcal{V}$. Any rational expected outcome $\mathbf{y}^{e*}$ verifies $\mathbf{y}^{e*} = \mathbf{L}(\mathbf{y}^{e*})$.*

Proposition 2.2 reduces the resolution of the game from an infinite-dimensional space to a finite-dimensional space. The uniqueness of $\mathbf{p}$ can now be established if $\mathbf{L}$ is a contracting mapping. I then make the following assumption.

Assumption 2.3. $\lambda < \dfrac{B_c}{||\mathbf{G}||_{\infty}}$, *where* $B_c = \left(\max_{u \in \mathbb{R}} \sum_{r=1}^{\infty} f_{\varepsilon}(u - a_r)\right)^{-1}$.

The multiple RE equilibria issue generally arises in peer effect models when the peer effect parameter exceeds some threshold (see Yang and Lee, 2017; Lee et al., 2014). Assumption 2.3 defines this threshold for the case of my model.[11] Assumption 2.3 also generalizes the restriction imposed on λ in the binary model proposed by Lee et al. (2014). If $\Delta c(r) = +\infty$ for $r \geqslant 2$, Assumption 2.3 implies that $\lambda < \dfrac{1}{||\mathbf{G}||_{\infty} f_{\varepsilon}(0)}$, which is the restriction set on λ in the binary data model.

When the network matrix is row normalized ($||\mathbf{G}||_{\infty} = 1$), Assumption 2.3 implies that $\lambda < B_c$. This is equivalent to assuming that the marginal peer effects are less than one for any value of ψ_i. Indeed, from the expected choice expression, $y_i^e = \ell_i(\mathbf{y}^e)$, the marginal expected choice with respect

[10] Generalization of Brouwer's fixed point theorem to an infinite-dimensional space (see Smart, 1980, Chapter 2).

[11] It is important that $\max_{u \in \mathbb{R}} \sum_{r=1}^{\infty} f_{\varepsilon}(u - a_r)$ be finite so that there exists a nonempty convex and compact set of λ's that verify Assumption 2.3. I state and prove a general lemma in Appendix A.2 on the convergence of all the infinite summations used in this paper.

to average expected peers' choice $\partial \ell_i(\mathbf{y}^e)/\partial \bar{y}_i^e$ is given by $\lambda \sum_{r=1}^{\infty} f_{\varepsilon}(\lambda \bar{y}_i^e + \psi_i - a_r)$, which is less than one by Assumption 2.3. Put differently, agents do not increase their expected choice greater than the increase in their average expected peers' choice, *ceteris paribus*. This is a standard uniqueness condition in peer effect models and will be verified in most cases (see Bramoullé et al., 2009).
The following theorem established the uniqueness of the rational belief system.

Theorem 2.1. *Under Assumptions 2.1–2.3, the game of incomplete information associated with payoff (1) has a unique rational belief system* $\mathbf{p}^* = (p_{ir}^*)$*, where the associated expected outcome* $\mathbf{y}^{e*}$ *is the unique solution of* $\mathbf{y}^e = \mathbf{L}(\mathbf{y}^e)$.

In practice, the econometrician does not observe the rational expected outcome $\mathbf{y}^{e*}$, nor do they observe the rational belief system $\mathbf{p}^*$. However, $\mathbf{y}^{e*}$ can be computed as the unique $\mathbf{L}$'s fixed point under Assumption 2.3. Besides, $\mathbf{p}^*$ can also be computed from $\mathbf{y}^{e*}$ using Equation (3). The rational belief system $\mathbf{p}^*$ defines the distribution of $\mathbf{y}^*$. This implies that the likelihood of the observed outcome $\mathbf{y}^*$ can be computed. In the next section, I study the parameter identification and present the model estimation strategy.

3 Econometric Model

In this section, I present the econometric specification of the model, study the parameter identification, and propose a strategy to estimate the model. The estimation strategy relies on the likelihood approach. I also account for network endogeneity by allowing unobservable individual characteristics to affect both the error term of the counting variable model and the adjacency matrix.

3.1 Specification

Let $\psi_i = \alpha_{s(i)} + \mathbf{x}_i'\boldsymbol{\beta} + \bar{\mathbf{x}}_i'\boldsymbol{\gamma}$, where $\mathbf{x}_i$ and $\bar{\mathbf{x}}_i$ are K-vectors of observable individual-specific characteristics and peers' average characteristics, respectively, and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_M)'$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are unknown parameters to be estimated. The parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are respectively interpreted as own effects and contextual effects (Manski, 1993). $\alpha_{s(i)}$ is a group-specific effect. Let $\mathbf{X} = [\mathbf{x}_1 \ \dots \ \mathbf{x}_n]'$ and $\mathbf{W}$ be an $n \times M$ matrix, where the (i, m)-th entry is one if i belongs to the m-th group and zero otherwise. The vector of ψ_i's can be written as $\boldsymbol{\psi} = \mathbf{Z}\boldsymbol{\theta}$, where $\mathbf{Z} = [\mathbf{W} \ \mathbf{X} \ \mathbf{GX}]$ and $\boldsymbol{\theta} = \left(\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\gamma}'\right)'$. Because the model includes group heterogeneity as fixed effects, I assume that the number of groups M is bounded. This avoids an incidental parameter issue when n grows to infinity. When n grows to infinity, the number of groups M is fixed but the number of individuals in each group grows to infinity (see Lancaster, 2000).

The likelihood approach requires being specific about the distribution of ε_i. Given that the rational expected outcome depends on the cdf F_ε, it is challenging to estimate the model parameters without assuming this cdf.[12]

Assumption 3.1. $\varepsilon_i \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right).$

Whereas most papers assume a logistic distribution, Assumption 3.1 sets that ε_i follows a normal distribution. This choice allows me to deal with the endogeneity of the network (see Section 3.3.2). The RE Equation (3) can now be expressed as

$$p_{ir} = \Phi\left(\frac{\lambda \bar{y}_i^e + \mathbf{z}_i'\boldsymbol{\theta} - a_r}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\lambda \bar{y}_i^e + \mathbf{z}_i'\boldsymbol{\theta} - a_{r+1}}{\sigma_\varepsilon}\right), \tag{4}$$

where Φ is the cdf of $\mathcal{N}(0,1)$, $\mathbf{z}_i'$ is the i-th row of $\mathbf{Z}$, and $\boldsymbol{\theta} = \left(\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\gamma}'\right)'$. The mapping $\mathbf{L}$ in proposition 2.2 is given by $\mathbf{L}(\mathbf{y}^e, \boldsymbol{\Gamma}) = (\ell_1(\mathbf{y}^e, \boldsymbol{\Gamma}) \ \dots \ell_n(\mathbf{y}^e, \boldsymbol{\Gamma}))'$, where

$$\ell_i(\mathbf{y}^e, \boldsymbol{\Gamma}) = \sum_{r=1}^{\infty} \Phi\left(\frac{\lambda \bar{y}_i^e + \mathbf{z}_i'\boldsymbol{\theta} - a_r}{\sigma_\varepsilon}\right). \tag{5}$$

I denote by δ_r the increment of the cut points a_r's for any $r \geqslant 2$; that is $\delta_r = a_r - a_{r-1}$ and by $\delta_1 = 0$. As $a_r = \Delta c(r) + \lambda r - \dfrac{\lambda}{2}$, for any $r \geqslant 2$, $\delta_r = \Delta\Delta c(r) + \lambda$. This means that $\delta_r \geqslant \lambda$, because $c(.)$ is convex. The role of the convex cost assumption is to have a strictly positive increment in the cut points. The same result can be obtained with a weaker condition than Assumption 2.1.[13]
In specification (4), there is an infinite number of cut points (or δ_r) to be estimated because the cost function is nonparametric. Without additional restrictions on δ_r's, the model's identification would be challenging. My identification strategy is based on the assumption that the increment δ_r is bounded. Therefore, δ_r converges to some $\bar{\delta}$ as r grows to ∞. In particular, I assume that this limit is reached for large values of r.

Assumption 3.2. *There exists* $\bar{R} \in \mathbb{N}$ *such that* $\forall\ r \geqslant \bar{R}$, $\delta_r = \delta_{\bar{R}}$, *where* $\delta_{\bar{R}} > 0$.

Assumption 3.2 means that the cost function has a quadratic representation only when $y \geqslant \bar{\mathrm{R}} - 1$. This restriction is flexible because $\bar{\mathrm{R}}$ can be set high, depending on the dependent variable. In practice, I found that setting $\bar{\mathrm{R}}$ to the 90th or 95th quantile of y is sufficient to get very good estimates even when the true value of $\bar{\mathrm{R}}$ is higher (see Section 4). Under Assumption 3.2, the cut points can be written as $a_r = a_1 + \sum_{k=1}^{r} \delta_k$ if $1 \leqslant r < \bar{\mathrm{R}}$, and $a_r = a_1 + (r - \bar{\mathrm{R}})\delta_{\bar{\mathrm{R}}} + \sum_{k=1}^{\bar{\mathrm{R}}} \delta_k$ if $r \geqslant \bar{\mathrm{R}}$. Therefore, the unknown parameters to be estimated in specification (4) are λ, $\boldsymbol{\theta}$, $\boldsymbol{\delta} = (\delta_2, \dots, \delta_{\bar{\mathrm{R}}})'$, a_1, and σ_ε.

[12] In general, all peer effect models under RE assume the distribution of the agent's type (e.g., Brock and Durlauf, 2001, 2002; Lee et al., 2014; Liu, 2019; Guerra and Mohnen, 2020).

[13] For instance, I can set that, $c(.)$ is a strictly increasing function that verifies $\Delta\Delta c(r) + \lambda > \epsilon$, for some $\epsilon > 0$. In this case, δ_r can be less than λ. However, in some empirical applications, the uniqueness RE equilibrium set in Assumption 2.3 would be violated if $\delta_r < \lambda$ for large r.

3.2 Identification

Let us first define the identification of the model.

Definition 3.1. *Assume $\mathbf{\Gamma} = \left(\lambda, \boldsymbol{\theta}', a_1, \boldsymbol{\delta}', \sigma_\varepsilon\right)'$. Let $\mathbf{\Gamma}_{(1)}$ and $\mathbf{\Gamma}_{(2)}$ be two values of $\mathbf{\Gamma}$. Let also $(p_{ir}^{(k)})$ and $\mathbf{y}^e_{(k)}$ be the rational beliefs and the rational expected outcome associated with $\mathbf{\Gamma}_{(k)}$ respectively, $k \in \{1, 2\}$. $\mathbf{\Gamma}$ is identified if $\mathbf{\Gamma}_{(1)} \neq \mathbf{\Gamma}_{(2)}$ implies that $\mathbf{y}^e_{(1)} \neq \mathbf{y}^e_{(2)}$ and $p_{ir}^{(1)} \neq p_{ir}^{(2)}$ for any r.*

Equation (4) does not change when λ, $\boldsymbol{\theta}$, a_1, $\boldsymbol{\delta}$, and σ_ε are multiplied by any positive number. This raises a first classical identification issue. In addition, $\boldsymbol{\alpha}$ and a_1 cannot be identified because they enter the equation only through their difference. As in an ordered model, these issues can be fixed by setting $\sigma_\varepsilon = 1$ and $a_1 = 0$. Then, the RE equation is

$$p_{ir} = \Phi\left(\lambda \bar{y}_i^e + \mathbf{z}_i'\boldsymbol{\theta} - a_r\right) - \Phi\left(\lambda \bar{y}_i^e + \mathbf{z}_i'\boldsymbol{\theta} - a_{r+1}\right), \tag{6}$$

where $a_0 = -\infty$, $a_r = \displaystyle\sum_{k=1}^{r} \delta_k$ if $1 \leqslant r < \bar{\mathrm{R}}$, $\delta_1 = 0$, $a_r = (r - \bar{\mathrm{R}})\delta_{\bar{\mathrm{R}}} + \displaystyle\sum_{k=1}^{\bar{\mathrm{R}}} \delta_k$ if $r \geqslant \bar{\mathrm{R}}$, and y_i^e is given by

$$y_i^e = \sum_{r=1}^{\infty} \Phi\left(\lambda \bar{y}_i^e + \mathbf{z}_i'\boldsymbol{\theta} - a_r\right). \tag{7}$$

As Φ is monotonic, according to Definition 3.1, the model parameters are identified if and only if $\tilde{\mathbf{Z}} = [\mathbf{G}\mathbf{y}^e \ \mathbf{Z}]$ is a full rank matrix. In the literature, $\tilde{\mathbf{Z}}$ is generally assumed to be a full rank matrix (e.g., see Brock and Durlauf, 2001; Lee et al., 2014; Liu, 2019; Yang and Lee, 2017; Guerra and Mohnen, 2020). This assumption is weak when the outcome is bounded. In this case, $\mathbf{G}\mathbf{y}^e$ is also bounded and cannot be linearly dependent on $\mathbf{Z}$ in general (except for special cases of $\mathbf{Z}$). In the current framework, $\mathbf{y}^e$ is unbounded. The full rank condition on $\tilde{\mathbf{Z}}$ is strong. Moreover, it cannot be tested given that $\mathbf{y}^e$ is not observed by the econometrician and depends on the true value of the parameters.
To prove that $\tilde{\mathbf{Z}}$ is a full rank matrix, I set the following assumption.

Assumption 3.3. $\boldsymbol{\gamma} \neq \mathbf{0}$, $\breve{\mathbf{Z}} = \left[\mathbf{W} \ \mathbf{X} \ \mathbf{G}\mathbf{X} \ \mathbf{G}^2\mathbf{X}\right]$ *is a full rank matrix if $n < \infty$ and $\displaystyle\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} \breve{\mathbf{z}}_i \breve{\mathbf{z}}_i'\right)$ is a finite nonsingular matrix, where $\breve{\mathbf{z}}_i$ is the i-th row of $\breve{\mathbf{Z}}$.*

Similar restrictions are also set by Bramoullé et al. (2009) in the case of linear-in-means models. The condition $\boldsymbol{\gamma} \neq \mathbf{0}$ means that the contextual effects matter. As argued by Bramoullé et al. (2009), with several characteristics, at least one component in $\boldsymbol{\gamma}$ must be different from 0. Moreover, setting that $\breve{\mathbf{Z}}$ is a full rank matrix is a testable restriction and must be verified for most network structures. For example, Assumption 3.3 is verified if agents from the same group are connected to each other (self-friendship excluded) and at least two groups have different sizes (see Lee, 2007). On the other hand, Assumption 3.3 is not verified if agents from the same group are connected to each other and self-friendship is included or when self-friendship is excluded but all the groups have the same size. These nonidentification results are well known with linear-in-means models and have been studied by Manski (1993) and Moffitt et al. (2001). The theoretical counterpart of this full rank condition on $\breve{\mathbf{Z}}$ is that $\lim_{n\to\infty}\left(\dfrac{1}{n}\sum_{i=1}^{n}\mathbf{z}_i\mathbf{z}_i'\right)$ is a finite nonsingular matrix. The latter is important for establishing the limiting distribution of the parameter estimator.

Proposition 3.1. *Under Assumption 3.3, λ, $\boldsymbol{\theta}$, and $\boldsymbol{\delta}$ can be identified .*

The identification is based on the fact that y_i is unbounded. Using this condition, I show that the mapping $\mathbf{L}$ is approximately linear. As a result, the identification result is similar to that of the linear-in-means model (see Proposition 1 of Bramoullé et al. (2009)).

3.3 Estimation

The estimation strategy is based on the Nested Pseudo-Likelihood (NPL) algorithm proposed by Aguirregabiria and Mira (2007) and recently used by Lin and Xu (2017) and Liu (2019). If $\mathbf{y}^e$ were observed, estimating the model would result in a simple *probit* model estimation by the maximum likelihood (ML) method. As $\mathbf{y}^e$ is not observed, the ML estimation requires computing $\mathbf{y}^e$; that is, solve a fixed point problem in $\mathbb{R}^n$ for each value of the parameter. This may be computationally cumbersome for large samples.

In contrast, the NPL algorithm uses an iterative process and does not require solving a fixed point problem. I distinguish both the cases of exogenous networks and endogenous networks.

3.3.1 Exogenous Networks

I assume that all the regressors are strictly exogenous in the sense that $\mathbb{E}\left(\varepsilon_i|\mathbf{G},\mathbf{W},\mathbf{X}\right) = 0$ for any $i \in \mathcal{V}$. I also assume ε_i's are independent across i. The NPL algorithm is based on a pseudo-likelihood function defined as

$$\mathcal{L}(\boldsymbol{\Gamma},\mathbf{y}^e) = \sum_{i=1}^{n}\sum_{r=0}^{\infty} d_{ir}\log(p_{ir}), \tag{8}$$

where $\boldsymbol{\Gamma} = \left(\lambda,\ \boldsymbol{\theta}', \log(\boldsymbol{\delta}')\right)'$, $p_{ir} = \Phi\left(\lambda\mathbf{g}_i\mathbf{y}^e + \mathbf{z}_i'\boldsymbol{\theta} - a_r\right) - \Phi\left(\lambda\mathbf{g}_i\mathbf{y}^e + \mathbf{z}_i'\boldsymbol{\theta} - a_{r+1}\right)$, $\mathbf{g}_i$ is the i-th row of $\mathbf{G}$, and $d_{ir} = 1$ if $y_i = r$ and $d_{ir} = 0$ otherwise.[14]

The NPL algorithm is an iterative approach that consists of starting with a proposal $\mathbf{y}_0^e$ for $\mathbf{y}^e$ and constructing the sequence of estimators $(\mathcal{Q}_t)_{t\geqslant 1}$, such that $\mathcal{Q}_t = \{\boldsymbol{\Gamma}_t, \mathbf{y}_m^e\}$ for $t \geqslant 1$, where $\boldsymbol{\Gamma}_t = \arg\max_{\boldsymbol{\Gamma}} \mathcal{L}(\boldsymbol{\Gamma}, \mathbf{y}_{t-1}^e)$ is the estimator of $\boldsymbol{\Gamma}$ at the t-th stage, and $\mathbf{y}_m = \mathbf{L}\left(\mathbf{y}_{t-1}^e, \boldsymbol{\theta}_t\right)$ is the estimator

[14] The parameter $\boldsymbol{\delta}$ is estimated in logarithm to take into account the constraint $\delta_r > 0$ for any $r \geqslant 2$.

of $\mathbf{y}^e$ at the t-th stage. In other words, given the guess $\mathbf{y}_0^e$, $\boldsymbol{\theta}_1 = \arg\max_{\boldsymbol{\Gamma}} \mathcal{L}(\boldsymbol{\Gamma}, \mathbf{y}_0^e)$ and $\mathbf{y}_1 = \mathbf{L}\left(\mathbf{y}_0^e, \boldsymbol{\theta}_1\right)$, then $\boldsymbol{\theta}_2 = \arg\max_{\boldsymbol{\Gamma}} \mathcal{L}(\boldsymbol{\Gamma}, \mathbf{y}_1^e)$, $\mathbf{y}_2 = \mathbf{L}\left(\mathbf{y}_1^e, \boldsymbol{\theta}_2\right)$, and so forth.

The sequence $\mathcal{Q}_t$ is well defined for any $t > 1$. Notice that each value of $\mathcal{Q}_t$ requires evaluating the mapping $\mathbf{L}$ only once. If $(\mathcal{Q}_t)_{t\geqslant 1}$ converges, regardless of the initial guess $\mathbf{y}_0^e$, its limit $\{\hat{\boldsymbol{\Gamma}}, \hat{\mathbf{y}^e}\}$ satisfies the following two properties: $\hat{\boldsymbol{\Gamma}}$ maximizes the pseudo-likelihood $\mathcal{L}(\boldsymbol{\Gamma}, \hat{\mathbf{y}^e})$ and $\hat{\mathbf{y}}^e = \mathbf{L}(\hat{\boldsymbol{\Gamma}}, \hat{\mathbf{y}^e})$.

As shown by Kasahara and Shimotsu (2012), a key determinant of the convergence of the NPL algorithm is the contraction property of the fixed point mapping $\mathbf{L}$ guaranteed by Theorem 2.1. In practice, when $||\hat{\boldsymbol{\Gamma}}_T - \hat{\boldsymbol{\Gamma}}_{T-1}||_1$ and $||\hat{\mathbf{y}}_T^e - \hat{\mathbf{y}}_{T-1}^e||_1$ are less than some tolerance values (for example 10^{-4}), I set $\hat{\boldsymbol{\Gamma}} = \hat{\boldsymbol{\Gamma}}_T$ and $\hat{\mathbf{y}}^e = \hat{\mathbf{y}}_T^e$. Aguirregabiria and Mira (2007) prove that the NPL estimator is root-n consistent and asymptotically normal. I adapt their proof to my framework (see Appendix A.6).

Some numerical aspects about the NPL estimator must be pointed out. First, the pseudo-likelihood (8) involves an infinite sum. However, as $d_{ir} = 0$ for any $r \neq y_i$, this pseudo-likelihood can also be expressed as $\mathcal{L}(\boldsymbol{\Gamma}, \mathbf{y}^e) = \sum_{i=1}^{n} \log(p_{iy_i})$. Second, the mapping $\mathbf{L}$, which is used to compute the sequence $(\mathcal{Q}_t)_t$ and the asymptotic variance of $\hat{\boldsymbol{\Gamma}}$, also involves an infinite sum. However, note that the summed elements decrease exponentially. A very good approximation of these sums can be readily reached by only summing a few elements. My R package may be used for this purpose.

3.3.2 Endogenous Networks

Assuming the network is exogenous implies that link formation does not depend on the error term ε_i. This assumption is strong and may lead to an inconsistent estimator. I now assume only $\mathbf{X}$ and $\mathbf{W}$ are exogenous with respect to ε_i; i.e., $\mathbb{E}\left(\varepsilon_i|\mathbf{W}, \mathbf{X}\right) = 0$, but ε_i depends on the network.

The strategy to control for endogeneity relies on a common intuition that some unobservable individual-level attributes influence both link formation and the counting outcome (see Hsieh et al., 2020; Johnsson and Moon, 2015). Given that these individual-level attributes are not included as unobserved variables, they are captured by the error term ε_i, which is then correlated to the network. To consistently estimate the model parameters, the individual-level attributes must be estimated and used as additional explanatory variables. This strategy is similar to the control function approach proposed by Johnsson and Moon (2015), which was initially used by Heckman (1979) is his selection model.

Let $\mathbf{A} = [a_{ij}]$ be the network data, such that $a_{ij} = 1$ if i knows j, and $a_{ij} = 0$ otherwise. Let also a_{ij}^* be the link formation utility, such that $a_{ij} = 1$ if $a_{ij}^* > 0$, and $a_{ij} = 0$ otherwise. I consider a dyadic linking model in which the probability of link formation between two agents i and j is specified with degree heterogeneity. Following Graham (2017) and Hsieh et al. (2020), I assume a_{ij}^* is determined by observed dyad-specific variables, denoted $\ddot{\mathbf{x}}_{ij}$, and on unobserved individual-level attributes (gregariousness) that capture *degree heterogeneity*. Formally, I specify that

$$a^*_{ij} = \ddot{\mathbf{x}}'_{ij}\bar{\boldsymbol{\beta}} + \mu_i + \nu_j + \varepsilon^*_{ij}, \tag{9}$$

where $\bar{\boldsymbol{\beta}}$ is the slope of the utility with respect to $\ddot{\mathbf{x}}_{ij}$, μ_i and ν_j are unobserved attributes, and ε^*_{ij} is an error term. The term $\ddot{\mathbf{x}}'_{ij}\bar{\boldsymbol{\beta}}$ can be interpreted as the social distance between i and j. Two restrictions regarding the network formation model must be pointed out. First, ε^*_{ij} is assumed independent across pairs $(i,\ j)$ and identically distributed according to $\mathcal{N}(0,1)$. This implies a dense network that does not suit several applications. On the other hand, I treat μ_i and ν_j as random effects following, respectively, $\mathcal{N}\left(0, \sigma^2_\mu\right)$ and $\mathcal{N}\left(0, \sigma^2_\nu\right)$.[15] Moreover, $(\mu_i, \nu_i)'$ and $(\mu_j, \nu_j)'$ are independent for $i \neq j$; i.e., $\mathbb{E}\left(\mu_i\mu_j\right) = 0$, $\mathbb{E}\left(\nu_i\nu_j\right) = 0$, and $\mathbb{E}\left(\mu_i\nu_j\right) = 0$. But μ_i and ν_i may be correlated: $\mathbb{E}\left(\mu_i\nu_i\right) = \rho_{\mu,\nu}\sigma_\mu\sigma_\nu$, where $\rho_{\mu,\nu}$ is the correlation between μ_i and ν_i. The assumption of random effects deals with the incidental parameter issue as the number of unobservable attributes grows to infinity with n.

The restrictions can be released and have no direct impact on the result of this section. The result of Proposition 3.2 below only requires μ_i and ν_j to be estimated consistently and the econometrician to be able to sample from the distribution of their estimator. This is encouraging, to the extent that the network formation model I present can be replaced by a more general model, and Proposition 3.2 still holds as long as the network formation model estimator is consistent.[16]

The probability of link formation between i and j, conditional on $\ddot{\mathbf{x}}_{ij}$, $\bar{\boldsymbol{\beta}}$, μ_i, and ν_j, is defined as

$$\mathbb{P}\left(a_{ij} = 1|\ddot{\mathbf{x}}_{ij}, \bar{\boldsymbol{\beta}}, \mu_i, \nu_j\right) = P_{ij} = \Phi\left(\ddot{\mathbf{x}}'_{ij}\bar{\boldsymbol{\beta}} + \mu_i + \nu_j\right). \tag{10}$$

By convention, I set $P_{ii} = 0$ and $P_{ij} = 0$ if $s(i) \neq s(j)$. Unlike most network formation models, the specification (10) includes two unobservable factors μ_i and ν_i. This implies a nonsymmetric matrix of link probabilities. The parameter μ_i only influences the probabilities of links going from i to another agent, whereas ν_i influences the probabilities of links going from other agents to i. This feature is relevant for directed networks.

The identification of the network formation model requires the standard assumption that the observed explanatory variables are linearly independent.

Assumption 3.4. *The matrix of explanatory variables,* $\ddot{\mathbf{X}} = [\ddot{\mathbf{x}}_{ij}; i \neq j, s(i) = s(j)]'$, *is a full rank matrix.*

Although μ_i and ν_i enter Equation (10) only by their sum, the identification issue faced by Graham

[15] As μ_i and ν_j are treated as random effects, the identification of the model does not require μ_i and ν_j being bounded as set by Graham (2017).

[16] For example, Dzemski (2019) proposed a strategy to correct the systematic bias due to the incidental parameter if μ_i and ν_i are treated as fixed effects.

(2017) does not arise here, as $\mathbb{E}(\mu_i) = 0$ and $\mathbb{E}(\nu_j) = 0$. These restrictions are used in the estimation strategy to ensure the identification (see Appendix A.7).

In the counting variable model, μ_i and ν_i are potentially correlated to ε_i, which implies network endogeneity. Indeed, $\mathbb{E}\left(\mu_i\varepsilon_i\right) = \rho_{\mu,\varepsilon}\sigma_{\mu},\sigma_{\varepsilon}$ and $\mathbb{E}\left(\nu_i\varepsilon_i\right) = \rho_{\nu,\varepsilon}\sigma_{\nu}\sigma_{\varepsilon}$, where $\rho_{\mu,\varepsilon}$ is the correlation between μ_i and ε_i, and $\rho_{\nu,\varepsilon}$ is the correlation between ν_i and ε_i. The error term ε_i can be rewritten as $\varepsilon_i = \theta_\mu\mu_i + \theta_\nu\nu_i + \tilde{\varepsilon}_i$, where $\tilde{\varepsilon}_i$'s are independent and identically distributed according to $\mathcal{N}\left(0, \sigma^2_{\tilde{\varepsilon}}\right)$. To control network endogeneity, μ_i and ν_i must be included in the counting variable model as additional explanatory variables. The RE equation becomes

$$p_{ir} = \Phi\left(\lambda\bar{y}^e_i + \mathbf{z}'_i\boldsymbol{\theta} + \theta_\mu\mu_i + \theta_\nu\nu_i - a_r\right) - \Phi\left(\lambda\bar{y}^e_i + \mathbf{z}'_i\boldsymbol{\theta} + \theta_\mu\mu_i + \theta_\nu\nu_i - a_{r+1}\right), \tag{11}$$

where y^e_i is given by

$$y^e_i = \sum_{r=1}^{\infty}\Phi\left(\lambda\bar{y}^e_i + \mathbf{z}'_i\boldsymbol{\theta} + \theta_\mu\mu_i + \theta_\nu\nu_i - a_r\right). \tag{12}$$

My estimation strategy is in two stages. The first stage is based on a Bayesian approach. Using a Gibbs sampler, I simulate $\bar{\boldsymbol{\beta}}$, μ_i, ν_i, $\sigma^2_{\mu,}$, σ^2_{ν}, and $\rho_{\mu,\nu}$ from their posterior distributions (see details in Appendix A.7). At the second stage, I estimate the counting variable model using the NPL approach after replacing μ_i and ν_i by their respective estimators $\hat{\mu}_i$ and $\hat{\nu}_i$ in Equations (11) and (12). I establish a new limiting distribution for the NPL estimator at the second stage, which accounts for the uncertainty of the first stage estimation.

Proposition 3.2. *Let* $\boldsymbol{\Gamma}^* = \left(\boldsymbol{\Gamma}',\ \theta_\mu,\ \theta_\nu\right)'$*. Let also* $\hat{\boldsymbol{\Gamma}}^*$ *be its NPL estimator when* μ_i *and* ν_i *are replaced by their respective consistent estimator* $\hat{\mu}_i$ *and* $\hat{\nu}_i$*. Then* $\hat{\boldsymbol{\Gamma}}^*$ *is consistent, and*

$$\sqrt{n}(\hat{\boldsymbol{\Gamma}}^* - \boldsymbol{\Gamma}^*_0 + \boldsymbol{\zeta}^*_n) \xrightarrow{d} \mathcal{N}\left(0,\ \left(\boldsymbol{\Sigma}^*_0 + \boldsymbol{\Omega}^*_0\right)^{-1}\bar{\boldsymbol{\Sigma}}_0\left(\boldsymbol{\Sigma}^{*\prime}_0 + \boldsymbol{\Omega}^{*\prime}_0\right)^{-1}\right), \tag{13}$$

where $\operatorname{plim}\boldsymbol{\zeta}^*_n = \mathbf{0}$*,* $\boldsymbol{\Gamma}^*_0$ *is the true value of* $\boldsymbol{\Gamma}^*$*;* $\boldsymbol{\zeta}^*_n$*,* $\bar{\boldsymbol{\Sigma}}_0$*,* $\boldsymbol{\Sigma}^*_0$*, and* $\boldsymbol{\Omega}^*_0$ *are given in Appendix A.8.*

The perturbation term $\boldsymbol{\zeta}^*_n$ appears because μ_i and ν_i are replaced by their estimator. However, since $\operatorname{plim}\boldsymbol{\zeta}^*_n = 0$, the result of Proposition 3.2 is sufficient to perform statistical tests on $\hat{\boldsymbol{\Gamma}}^*$ in large samples because $\boldsymbol{\zeta}^*_n \approx 0$. Such a limiting distribution with a perturbation term is well known in the variable selection literature (e.g., Fan and Li, 2001; Fan and Peng, 2004). To compute the asymptotic variance in practice, one should be able to sample from the distribution of $(\hat{\mu}_i,\ \hat{\nu}_i)$.[17]

[17] See Appendix A.8.

4 Monte Carlo Experiments

In this section, I conduct a Monte Carlo study to assess the performance of the estimator in finite sample. I consider both cases where the network is exogenous and endogenous. I also discuss how to set the value of $\bar{\mathrm{R}}$ in practice. As pointed out above, the linear-in-means model is based on a strong restriction ($\bar{\mathrm{R}} = 2$), which leads to inconsistent estimations. I illustrate this result with simulations by comparing the model to the spatial autoregressive Tobit (SART) model. I use the SART model, as it controls for the left-censure issue.

In this simulation study, $\psi_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \gamma_1 \bar{x}_{1i} + \gamma_2 \bar{x}_{2i}$. There is only one intercept α regardless the number of subnetworks. This means that I do not account for fixed effects. In fact, note that as I assume that the number of subnetworks is finite, the fixed effects can be viewed as ordinary explanatory variables. The parameters are set as follows: $\lambda = 0.3$, $\alpha = 2.5$, $\boldsymbol{\beta} = (1.5, -1.2)'$, and $\boldsymbol{\gamma} = (0.5, -0.9)'$. The exogenous variables x_1 and x_2 are simulated from $\mathcal{N}(1, 1)$ and $\mathcal{P}oisson(2)$, respectively.

4.1 Exogenous Networks

In the case of an exogenous network, I assume only one subnetwork. Each individual i is randomly assigned to n_i friends, where n_i is randomly chosen between 0 and 30. The network matrix $\mathbf{G}$ is row normalized. I consider two specifications of the model with different values of $\bar{\mathrm{R}}$. In the first specification (model A), $\bar{\mathrm{R}} = 6$, with $\boldsymbol{\delta} = (1,\ 0.87,\ 0.75,\ 0.55,\ 0.35)'$. I recall that $\boldsymbol{\delta} = (a_2 - a_1,\ \dots,\ a_{\bar{\mathrm{R}}} - a_{\bar{\mathrm{R}}-1})'$ and that $\forall\ r > \bar{\mathrm{R}}$, $a_r - a_{r-1} = a_{\bar{\mathrm{R}}} - a_{\bar{\mathrm{R}}-1}$. In the second specification (model B), $\bar{\mathrm{R}} = 13$, with $\boldsymbol{\delta} = (1.2,\ 0.7,\ 0.55,\ 0.5,\ 0.5,\ 0.4,\ 0.4,\ 0.3,\ 0.3,\ 0.25,\ 0.25,\ 0.2)'$.

In practice, the econometrician does not know the true value of $\bar{\mathrm{R}}$ and has to set it. Let $\hat{\bar{\mathrm{R}}}$ be the value set empirically. The model is well specified if $\hat{\bar{\mathrm{R}}} \geqslant \bar{\mathrm{R}}$. In contrast, a misspecification issue could occur when $\hat{\bar{\mathrm{R}}} < \bar{\mathrm{R}}$. A very low $\hat{\bar{\mathrm{R}}}$ implies a strong restriction and would lead to an inconsistent estimator. Furthermore, the number of parameters to be estimated increases with $\hat{\bar{\mathrm{R}}}$; thus, $\hat{\bar{\mathrm{R}}}$ cannot be set as large as possible. For the sake of identification, $\hat{\bar{\mathrm{R}}}$ must be less than the empirical maximum of y; that is $\hat{\bar{\mathrm{R}}} < \max\limits_i(y_i)$.

Figure 1 presents the histogram of an example of the simulated data for $n = 1,500$. I found that setting $\hat{\bar{\mathrm{R}}}$ over the 90th percentile of y gives very satisfactory results.[18] This is because the correlation between p_{ir} and a_r decreases exponentially as r grows. The 90th percentile corresponds to $\hat{\bar{\mathrm{R}}} = 8$ in the case of the model A and $\hat{\bar{\mathrm{R}}} = 9$ in the case of the model B. Thus, $\hat{\bar{\mathrm{R}}} > \bar{\mathrm{R}}$ for model A and one expects "good" estimations. However, the distribution of y under model B has a long tail and $\hat{\bar{\mathrm{R}}} < \bar{\mathrm{R}}$. This case is interesting for assessing my strategy that suggests setting $\hat{\bar{\mathrm{R}}}$ over the 90th percentile of

[18] I confirm this result with several other data generator processes.

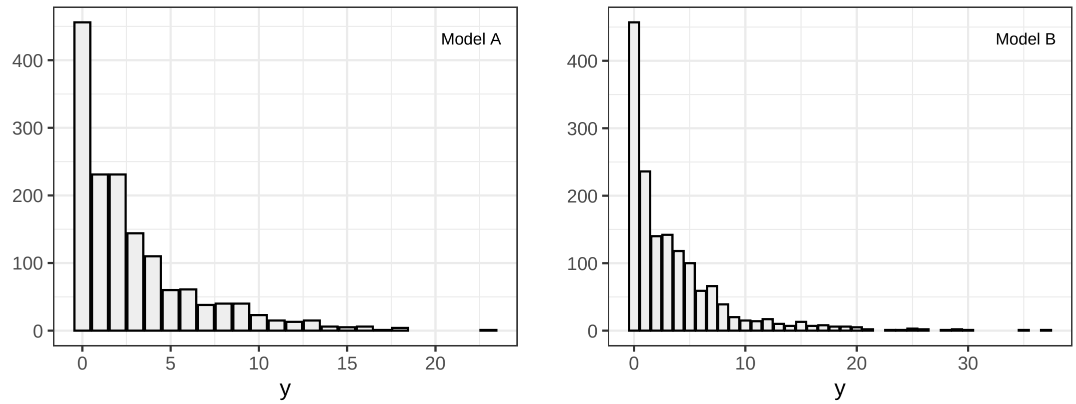

Figure 1: Simulated data using the count data model with social interactions

y. Moreover, the situation in the case of model B is frequent with survey data. One example is my empirical application. With such a distribution, the cost function is not likely quadratic. Therefore, the SART estimator would be strongly biased.

The simulation results (for 1,000 replications) are presented in Table 1. Note that one cannot directly interpret the parameters of the counting variable model, nor can one compare those parameters to those of the SART model. Table 1 reports the marginal effect (ME) of each variable.[19] Column (1) presents the true ME. The notation $\delta(.)$ denotes the ME of the variable in parentheses. Columns (2) and (3) report the estimated MEs and their corresponding standard deviation, respectively, under the counting variable model, where the empirical value of $\bar{\mathrm{R}}$ is $\hat{\bar{\mathrm{R}}}(1)$. For both models A and B, the NPL estimator performs well. As expected, the finite sample bias of the marginal peer effect is higher for model B. However, regarding the standard deviation, one can consider this bias to be negligible.

Columns (4) and (5) report the same MEs but estimated from the SART model. The latter significantly overestimates the marginal peer effects. To understand why the bias is positive, observe that by imposing $\hat{\bar{\mathrm{R}}} = 2$, the SART model overestimates the distance between the cut points a_r's as r grows.[20] Therefore, the expected outcome y_i^e is underestimated, and thus the marginal peer effect is overestimated. As expected, the bias is greater under model B. Moreover, the SART model slightly underestimates the marginal effect of the control variables.

To illustrate that the linear model indirectly imposes $\hat{\bar{\mathrm{R}}} = 2$, I conduct further simulations. I re-estimate the MEs using the same data sets generated from the models A and B and by setting $\hat{\bar{\mathrm{R}}}$ at $\hat{\bar{\mathrm{R}}}(2) = 2$. As expected, the results approximately replicate the bias of the SART model (see columns (6) and (7)). I also consider a new specification (model C), in which the true value of $\bar{\mathrm{R}}$ is 2. The

[19] I compute the ME for each individual and take the average. I present how to derive the marginal effects and the corresponding standard errors for the count data model in Appendix B.1.

[20] This is because the true distance between the cut points decreases. If the true distance were increasing, the SART model would underestimate the marginal peer effects.

results show that the SART model performs well in this case.

4.2 Endogenous Networks

I consider a population of four subnetworks, each composed of 250 nodes (agents). This allows us to bound the number of friends with whom agents interact as in the case of the exogenous network. Interactions are restricted to people in the same subnetwork. The probability for two individuals from the same network to be connected is $P_{ij} = \Phi(\bar{\beta}_0 + \bar{\beta}_1|x_{1i} - x_{1j}| + \bar{\beta}_2|x_{2i} - x_{2j}| + \mu_i + \nu_j)$. The exogenous variables x_1 and x_2 are simulated from $\mathcal{N}(1,1)$ and $\mathcal{P}oisson(2)$, respectively. The unobserved attributes $(\mu_i,\ \nu_i)$ are independent across i and follow a centered normal distribution, where $\sigma_\mu = 0.2$, $\sigma_\nu = 0.3$, and $\rho_{\mu,\nu} = 0.2$. I set $\bar{\beta}_1 = \bar{\beta}_2 = 0.1$ and $\bar{\beta}_0 \in \{-2.3,\ -1.8,\ -1.2,\ -2\}$, where one value is used for each subnetwork.

Regarding the counting variable model, I set $\psi_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \gamma_1 \bar{x}_{1i} + \gamma_2 \bar{x}_{2i} + \theta_\mu \mu_i + \theta_\nu \nu_i$, where $\theta_\mu = 0.8$ and $\theta_\nu = 0.5$. I use the same value set in the case of exogeneity for β_0, β_1, and β_2. The true value of $\bar{\mathrm{R}}$ is 6, and $\boldsymbol{\delta}$ is defined as in model A.

The results are presented in Table 2. The empirical value of $\bar{\mathrm{R}}$ is set at $\hat{\bar{\mathrm{R}}} = 8$, which corresponds to the 90th percentile. Column (1) reports the true marginal effects. Columns (2) and (3) report the estimated marginal effects and their respective standard deviations when one does not control for endogeneity. The estimated marginal peer effect is biased upward. The bias is positive because θ_μ and θ_ν are positive. Columns (4) and (5) present the estimated marginal effects and their standard deviation while controlling for endogeneity. The finite sample bias of the marginal peer effect is low and can be negligible regarding its corresponding standard deviation.

5 Effect of Social Interactions on Participation in Extracurricular Activities

In this section, I present an empirical illustration of the model using a unique and now widely used data set provided by the National Longitudinal Study of Adolescent Health (Add Health).

5.1 Data

The studied counting dependent variable is the number of extracurricular activities in which students are enrolled. Participation in extracurricular activities is associated with positive educational, social, and developmental outcomes, such as increased achievement, improved interpersonal skills, reduced levels of delinquency, reduced likelihood of dropping out, and improved self-esteem (see Holland and Andre, 1987; McNeal Jr, 1999; Darling, 2005).

Table 1: Monte Carlo simulations with under exogenous networks

(1)	(2)	(3)	(4)	(5)	(6)	(7)
	CDSI(1)		**SART**		**CDSI(2)**	
Marginal effects	**Mean**	**Sd.**	**Mean**	**Sd.**	**Mean**	**Sd.**
			Model A, $\bar{\text{R}} = 6$, $\hat{\bar{\text{R}}}(1) = 8$, $\hat{\bar{\text{R}}}(2) = 2$			
$\delta(\bar{y}^e) = 0.387$	0.385	(0.047)	0.420	(0.058)	0.422	(0.058)
$\delta(x_1) = 1.933$	1.932	(0.091)	1.902	(0.089)	1.893	(0.089)
$\delta(x_2) = -1.546$	-1.545	(0.073)	-1.483	(0.067)	-1.470	(0.067)
$\delta(\bar{x}_1) = 0.644$	0.649	(0.165)	0.600	(0.183)	0.596	(0.184)
$\delta(\bar{x}_2) = -1.160$	-1.160	(0.095)	-1.146	(0.100)	-1.144	(0.101)
			Model B, $\bar{\text{R}} = 13$, $\hat{\bar{\text{R}}}(1) = 9$, $\hat{\bar{\text{R}}}(2) = 1$			
$\delta(\bar{y}^e) = 0.451$	0.458	(0.049)	0.525	(0.070)	0.534	(0.084)
$\delta(x_1) = 2.255$	2.258	(0.137)	2.194	(0.131)	2.186	(0.131)
$\delta(x_2) = -1.804$	-1.805	(0.109)	-1.694	(0.095)	-1.682	(0.096)
$\delta(\bar{x}_1) = 0.752$	0.743	(0.186)	0.663	(0.226)	0.639	(0.250)
$\delta(\bar{x}_2) = -1.353$	-1.353	(0.122)	-1.342	(0.136)	-1.335	(0.137)
			Model C, $\bar{\text{R}} = 2$, $\hat{\bar{\text{R}}}(1) = 5$, $\hat{\bar{\text{R}}}(2) = 2$			
$\delta(\bar{y}^e) = 0.215$	0.213	(0.051)	0.210	(0.051)	0.213	(0.051)
$\delta(x_1) = 1.074$	1.074	(0.031)	1.089	(0.032)	1.074	(0.031)
$\delta(x_2) = -0.860$	-0.859	(0.025)	-0.878	(0.026)	-0.859	(0.025)
$\delta(\bar{x}_1) = 0.358$	0.361	(0.097)	0.370	(0.099)	0.361	(0.097)
$\delta(\bar{x}_2) = -0.645$	-0.645	(0.049)	-0.653	(0.050)	-0.645	(0.049)

CDSI stands for count data model with social interactions. The number of simulations performed is 1,000. The "Mean" column reports the average of the 1,000 estimations, and the "Sd." column reports the standard deviation. CDSI(1) is estimated by setting $\hat{\bar{\text{R}}}$ at $\hat{\bar{\text{R}}}(1)$, whereas CDSI(2) is estimated by setting $\hat{\bar{\text{R}}}$ at $\hat{\bar{\text{R}}}(2)$.

Table 2: Monte Carlo simulations under endogenous networks

(1)	(2)	(3)	(4)	(5)
	CDSI(1)		**CDSI(2)**	
Marginal effects	**Mean**	**Sd.**	**Mean**	**Sd.**
		$\bar{\text{R}} = 6$, $\hat{\bar{\text{R}}} = 8$		
$\delta(\bar{y}^e) = 0.381$	0.439	(0.083)	0.375	(0.068)
$\delta(x_1) = 1.904$	1.906	(0.143)	1.907	(0.139)
$\delta(x_2) = -1.524$	-1.513	(0.115)	-1.522	(0.113)
$\delta(\bar{x}_1) = 0.635$	0.627	(0.247)	0.661	(0.215)
$\delta(\bar{x}_2) = -1.143$	-1.038	(0.158)	-1.144	(0.149)

CDSI stands for count data model with social interactions. The number of simulations performed is 1,000. The "Mean" column reports the average of the 1,000 estimations, and the "Sd." column reports the standard deviation.

The Add Health data provide national representative information on 7th–12th graders in the United States (US). I use the Wave I in-school data, which were collected between September 1994 and April 1995. The surveyed sample comprises 80 high schools and 52 middle schools. In particular, the data provides information on the social and demographic characteristics of students as well as their friendship links (i.e., best friends, up to 5 females and up to 5 males), education level, occupation of parents, etc. Students were presented with a list of clubs, organizations, and teams found in many schools. The students were asked to identify any of these activities in which they participated during the current school year or in which they planned to participate later in the school year. The students do not observe the activities in which their peers plan to participate. Therefore, the studied dependent variable is a good example for illustrating the model because the outcome is suited to a Bayesian game used to address the model.

I remove self-friendships and friendships between two students from different schools. Moreover, an important number of listed friend identifiers are missing or associated with "error codes."[21] I, therefore, remove from the study sample schools having many missing links and those having less than 100 students. I end up with 72,291 students from 120 schools. The largest school has 2,156 students, and about 50% of the schools have more than 500 students. The average number of friends per student is 3.8 (1.8 male friends and 2.0 female friends).

McNeal Jr (1999) shows that school characteristics and other characteristics of the household in which student lives likely influence their participation in extracurricular activities. In this empirical study, I determine whether social interactions also play an important role. In the matrix of explanatory variables $\mathbf{X}$, I include several other potential factors, such as age, sex, race of the student, whether the student is Hispanic, the number of years spent at their current school by the student, whether the student lives with both parents, mother's education, and mother's profession. Table 6 provides the data summary and Figure 3 the histogram of the number of extracurricular activities in which the students are enrolled. The number of activities varies from 0 to 33 with an average of 2.4. The histogram looks like the histogram of model B with a long tail. I expect the SART model to overestimate the marginal peer effect. For the categorical explanatory variables, the level in italics is set as the reference level in the econometric models (see Table 6).

5.2 Empirical results

I set $\hat{\bar{\mathrm{R}}}$ at 5, which corresponds to the quantile at 90% of the number of activities in which students are enrolled. Table 3 presents the estimation results when I do not control for school heterogeneity and

[21] In the recent literature, numerous papers have developed methods for estimating peer effects using partial network data (e.g., Boucher and Houndetoungan, 2020). To focus on the main purpose of this paper, I do not address that issue here.

network endogeneity. The findings indicate that social interactions play an important role in student participation in extracurricular activities. An increase by one in the expected number of activities in which friends are enrolled implies an increase in the expected number of activities in which the students are enrolled by 0.384. As in the Monte Carlo study, the SART model overestimates this effect at 0.552. The difference between both estimates highlights how important it is to use the counting variable model instead of the linear or the Tobit model to estimate peer effects on count data.

I also find that most of the control variables influence student participation in extracurricular activities. This corroborates several results in the literature (e.g., McNeal Jr, 1999); for example, Black and Asian students participate more than White students. Students who have been in their current school for a longer time and those who live with both parents also participate more. Morevover, students whose mother's education is high school or whose mother's job is professional are more involved in recreational activities. However, I find that Hispanic students, old students, and boys are less involved in extracurricular activities. This result is in part at odds with that of McNeal Jr (1999). Several contextual factors also impact student participation in recreational activities. For instance, being friend with an old, Hispanic, Black, or Asian students impinges upon student participation; however, interacting with a student whose mother's education is high school or whose mother's job is professional likely increases participation.

As found by McNeal Jr (1999), school characteristics such as size, pupil/teacher ratio, and general school climate also determine student participation. This suggests that school heterogeneity plays an important role in student participation. I control for this heterogeneity by including school fixed effects in the model. Indeed, as argued by Lee et al. (2014) and Liu (2019), the number of schools (120) is low relative to the sample size. Therefore, this does not raise an incidental parameter issue. Those fixed effects control for any observed or unobserved school attribute (which is not taken into account in the previous estimation) that influences participation. Table 4 reports the new findings. The pseudo-log-likelihood increases by 1,384 for 119 additional explanatory variables. The likelihood ratio (LR) test confirms the importance of those school fixed effects. The marginal peer effect decreases at 0.306 but is still overestimated at 0.358 by the SART model. Some control variables are no longer significant such as being Hispanic or interacting with a male, Hispanic, or Black student.

I also control for network endogeneity. Participation in extracurricular activity may depend on personality, such as sociability degree. Indeed, evidence has been found in sociology that specific personality traits are associated with activity participation, extroverted people work more often in jobs having more social interactions and highly gregarious individuals are more likely to be a member of a group (e.g., Newton et al., 2018; Pfeiffer and Schulz, 2012; Erbe, 1962). Thus, these personality traits also are likely to increase student probability of interacting with others. This implies that network matrix $\mathbf{G}$ is potentially endogenous. One can control for this endogeneity by including the personality traits in the participation model as additional explanatory variables. However, these traits are not observed and can only be estimated. To do so, I use the dyadic linking model presented in Section 3.3.2. The posterior distribution of the parameters is presented in Appendix B.3.

The use of the estimated unobserved attributes (and not the true unobserved attributes) complicates the asymptotic of the counting variable model. Therefore, Proposition 3.2 comes into play here. From the posterior distribution of the parameters of the network formation model, I take into account the uncertainty associated with the estimation of the unobserved traits in the participation model. Table 5 presents the estimation results. The coefficients θ_μ and θ_ν are the parameters associated with the additional explanatory variables. They are significant, which confirms that the network is endogenous. The marginal peer effect now decreases at 0.256 and is still overestimated at 0.325 by the SART model. To understand the decrease in the marginal peer effects, one should note that the peer effects capture the effect of any positive shocks the sociability degree when one does not control for the network endogeneity. Thus, the marginal peer effect of Table 4 includes the influence of the degree of student sociability on participation.

6 Conclusion

In this paper, I develop a social network model for counting data using a static game of incomplete information. I find sufficient conditions under which the game has a unique equilibrium and propose a strategy to estimate the model parameters. I generalize this estimation strategy to the case where the network is endogenous. The microfoundations of the model are similar to those of the linear-in-mean model. Individuals in the game interact through a directed network, simultaneously choose their strategy, and receive a payoff that depends on their belief about the choice of their peers. However, unlike the linear model, which assumes a linear-quadratic payoff, the counting nature of the outcome allows to deal with a more flexible payoff. I show that the restriction of the linear-quadratic payoff leads to an inconsistent estimator of peer effects on counting variables. I support this result using Monte Carlo simulations.

I also control for network endogeneity using a two-stage estimation strategy. In the first stage, I estimate a dyadic linking model in which the probability of link formation between two students depends, among others, on unobserved attributes. In the second stage, the estimated attributes are included in the peer effect model. I establish a new limiting distribution of the peer effect model parameters that accounts for the uncertainty associated with the first stage estimation.

I provide an empirical application. I estimate peer effects on the number of extracurricular activities in which a student is enrolled. By controlling for the endogeneity of the network, I find that an increase by one in the expected number of activities in which friends are enrolled implies an increase in the

Table 3: Application results without fixed effects

(1)	(2)	(3)	(4)	(5)	(6)	(7)
		CDSI			**SART**	
Parameters	**Coef.**	**Marginal effects**		**Coef.**	**Marginal effects**	
λ	0.200	0.384	(0.003)	0.681	0.552	(0.018)
Own effects						
Age	−0.028	−0.054	(0.004)	−0.018	−0.015	(0.004)
Male	−0.132	−0.253	(0.017)	−0.246	−0.200	(0.011)
Hispanic	−0.047	−0.089	(0.026)	0.025	0.020	(0.017)
Race						
Black	0.088	0.169	(0.030)	0.241	0.195	(0.019)
Asian	0.219	0.420	(0.036)	0.668	0.542	(0.022)
Other	0.065	0.125	(0.029)	0.211	0.171	(0.018)
Years at school	0.049	0.094	(0.007)	0.123	0.100	(0.005)
With both par.	0.079	0.152	(0.019)	0.161	0.131	(0.012)
Mother Educ.						
<High	−0.069	−0.132	(0.024)	−0.068	−0.055	(0.015)
>High	0.186	0.358	(0.019)	0.381	0.309	(0.012)
Missing	0.037	0.070	(0.032)	0.211	0.171	(0.021)
Mother job						
Professional	0.114	0.218	(0.025)	0.216	0.176	(0.016)
Other	0.037	0.072	(0.021)	0.061	0.049	(0.013)
Missing	−0.049	−0.095	(0.029)	−0.085	−0.069	(0.019)
Contextual effects						
Age	−0.017	−0.033	(0.004)	−0.077	−0.062	(0.003)
Male	0.029	0.055	(0.028)	0.106	0.086	(0.019)
Hispanic	−0.084	−0.160	(0.039)	−0.148	−0.120	(0.025)
Race						
Black	−0.044	−0.084	(0.036)	−0.161	−0.131	(0.024)
Asian	−0.174	−0.333	(0.047)	−0.588	−0.477	(0.031)
Other	−0.092	−0.176	(0.051)	−0.279	−0.227	(0.033)
Years at school	0.006	0.012	(0.009)	−0.029	−0.023	(0.006)
With both par.	0.090	0.172	(0.034)	0.070	0.056	(0.024)
Mother Educ.						
<High	−0.137	−0.263	(0.042)	−0.225	−0.183	(0.027)
>High	0.073	0.140	(0.032)	0.016	0.013	(0.024)
Missing	−0.091	−0.174	(0.060)	−0.253	−0.205	(0.039)
Mother job						
Professional	0.098	0.187	(0.044)	0.092	0.075	(0.030)
Other	0.029	0.057	(0.036)	−0.006	−0.005	(0.024)
Missing	−0.020	−0.038	(0.053)	−0.026	−0.021	(0.035)
$\boldsymbol{\delta}$	$(1.512,\ 0.511,\ 0.452,\ 0.201)'$					
σ_ε					2.475	
N		72,291			72,291	
log-likelihood		-131,130.3			-161,224.7	
Fixed Effects		No			No	
Endogeneity		No			No	

CDSI stands for count data model with social interactions, where $\hat{\bar{R}} = 5$. Columns (2)–(4) (respectively (5)–(7)) report the coefficients, the MEs, and the standard deviations of the MEs of the counting variable (respectively SART) model.

Table 4: Application results with fixed effects

(1)	(2)	(3)	(4)	(5)	(6)	(7)
		CDSI			**SART**	
Parameters	**Coef.**	**Marginal effects**		**Coef.**	**Marginal effects**	
λ	0.163	0.306	(0.020)	0.442	0.358	(0.019)
Own effects						
Age	−0.037	−0.069	(0.007)	−0.051	−0.041	(0.005)
Male	−0.143	−0.269	(0.017)	−0.262	−0.212	(0.011)
Hispanic	0.001	0.002	(0.026)	0.114	0.093	(0.017)
Race						
Black	0.136	0.255	(0.030)	0.304	0.247	(0.020)
Asian	0.241	0.453	(0.035)	0.700	0.568	(0.023)
Other	0.073	0.136	(0.028)	0.220	0.178	(0.018)
Years at school	0.048	0.090	(0.008)	0.121	0.098	(0.005)
With both par.	0.079	0.148	(0.019)	0.159	0.129	(0.012)
Mother Educ.						
<High	−0.055	−0.104	(0.023)	−0.046	−0.037	(0.015)
>High	0.195	0.367	(0.019)	0.396	0.321	(0.013)
Missing	0.043	0.080	(0.032)	0.221	0.179	(0.021)
Mother job						
Professional	0.128	0.241	(0.025)	0.241	0.195	(0.016)
Other	0.044	0.083	(0.020)	0.072	0.058	(0.013)
Missing	−0.040	−0.075	(0.029)	−0.068	−0.056	(0.019)
Contextual effects						
Age	−0.017	−0.033	(0.004)	−0.062	−0.050	(0.003)
Male	0.001	0.002	(0.030)	0.024	0.020	(0.019)
Hispanic	−0.027	−0.052	(0.041)	−0.047	−0.038	(0.027)
Race						
Black	−0.015	−0.029	(0.038)	−0.076	−0.061	(0.025)
Asian	−0.074	−0.139	(0.051)	−0.324	−0.263	(0.035)
Other	−0.087	−0.163	(0.052)	−0.242	−0.197	(0.034)
Years at school	0.005	0.008	(0.011)	−0.014	−0.011	(0.007)
With both par.	0.105	0.197	(0.036)	0.170	0.138	(0.024)
Mother Educ.						
<High	−0.117	−0.221	(0.043)	−0.184	−0.150	(0.028)
>High	0.123	0.231	(0.037)	0.194	0.158	(0.025)
Missing	−0.084	−0.157	(0.060)	−0.183	−0.148	(0.040)
Mother job						
Professional	0.151	0.284	(0.047)	0.263	0.213	(0.031)
Other	0.051	0.095	(0.038)	0.079	0.064	(0.025)
Missing	0.012	0.022	(0.054)	0.061	0.049	(0.036)
$\boldsymbol{\delta}$	$(1.542,\ 0.525,\ 0.465,\ 0.204)'$					
σ_ε					2.442	
N		72,291			72,291	
log-likelihood		-129,746.1			-160,258.4	
Fixed Effects		Yes			Yes	
Endogeneity		No			No	

CDSI stands for count data model with social interactions, where $\hat{\bar{\mathrm{R}}} = 5$. Columns (2)–(4) (respectively (5)–(7)) report the coefficients, the MEs, and the standard deviations of the MEs of the counting variable (respectively SART) model.

Table 5: Application results fixed effects and network endogeneity

(1)	(2)	(3)	(4)	(5)	(6)	(7)
		CDSI			**SART**	
Parameters	**Coef.**	**Marginal effects**		**Coef.**	**Marginal effects**	
λ	0.138	0.256	(0.018)	0.400	0.325	(0.016)
θ_μ	0.329	0.612	(0.097)	0.791	0.642	(0.060)
θ_ν	0.396	0.736	(0.078)	0.668	0.542	(0.046)
Own effects						
Age	−0.046	−0.086	(0.008)	−0.063	−0.052	(0.005)
Male	−0.143	−0.266	(0.018)	−0.258	−0.209	(0.012)
Hispanic	0.005	0.009	(0.026)	0.120	0.097	(0.017)
Race						
Black	0.200	0.371	(0.033)	0.429	0.348	(0.029)
Asian	0.225	0.417	(0.035)	0.662	0.537	(0.023)
Other	0.026	0.048	(0.028)	0.125	0.102	(0.019)
Years at school	0.036	0.067	(0.008)	0.094	0.076	(0.005)
With both par.	0.061	0.114	(0.020)	0.121	0.098	(0.014)
Mother Educ.						
<High	−0.057	−0.105	(0.026)	−0.048	−0.039	(0.018)
>High	0.187	0.348	(0.020)	0.375	0.304	(0.014)
Missing	0.031	0.058	(0.034)	0.197	0.160	(0.023)
Mother job						
Professional	0.120	0.223	(0.026)	0.221	0.179	(0.018)
Other	0.026	0.048	(0.021)	0.034	0.027	(0.015)
Missing	−0.052	−0.096	(0.031)	−0.093	−0.075	(0.021)
Contextual effects						
Age	−0.023	−0.042	(0.004)	−0.073	−0.059	(0.003)
Male	−0.028	−0.053	(0.031)	−0.033	−0.027	(0.020)
Hispanic	−0.044	−0.081	(0.043)	−0.075	−0.061	(0.029)
Race						
Black	−0.004	−0.007	(0.038)	−0.058	−0.047	(0.032)
Asian	−0.056	−0.103	(0.052)	−0.287	−0.233	(0.036)
Other	−0.093	−0.173	(0.052)	−0.249	−0.202	(0.035)
Years at school	0.010	0.019	(0.011)	−0.006	−0.005	(0.007)
With both par.	0.065	0.121	(0.036)	0.086	0.070	(0.024)
Mother Educ.						
<High	−0.126	−0.234	(0.044)	−0.213	−0.173	(0.029)
>High	0.111	0.206	(0.038)	0.143	0.116	(0.026)
Missing	−0.092	−0.171	(0.061)	−0.208	−0.169	(0.041)
Mother job						
Professional	0.104	0.194	(0.047)	0.157	0.128	(0.032)
Other	−0.001	−0.001	(0.038)	−0.031	−0.025	(0.026)
Missing	−0.036	−0.067	(0.054)	−0.036	−0.029	(0.037)
$\boldsymbol{\delta}$	$(1.561,\ 0.533,\ 0.472,\ 0.205)'$					
σ_ε					2.424	
N		72,291			72,291	
log-likelihood		-128,995.0			-159,722.9	
Fixed Effects		Yes			Yes	
Endogeneity		Yes			Yes	

CDSI stands for count data model with social interactions, where $\hat{\bar{\mathrm{R}}} = 5$. Columns (2)–(4) (respectively (5)–(7)) report the coefficients, the MEs, and the standard deviations of the MEs of the counting variable (respectively SART) model.

expected number of activities in which students are enrolled by 0.256. However, the SART model overestimates this effect at 0.325. I also find that network endogeneity is important, and that ignoring this endogeneity overestimates the peer effects. Finally, I provide an easy-to-use R package that implements all the methods used in this paper.[22]

The findings of this paper raise an important question. Because the assumption of a quadratic cost function leads to inconsistent estimations of peer effects on counting variables, it is questionable whether this restriction is not also strong for the linear model. This question would be difficult to answer and requires releasing some important parametric assumptions in the microfoundations of the linear model.

[22] The package is available at github.com/ahoundetoungan/CDatanet.

Appendices

A Proof of Propositions

A.1 Proof of Proposition 2.1

First, I state and prove the following lemma, which adapts Murota (1998) to the case of univariate concave discrete functions.

Lemma A.1. *Let $\bar{D}$ be a convex subset of $\mathbb{R}$ and let h be a discrete concave function defined on $D_h = \bar{D} \cap \mathbb{Z}$. Let also $r_0 \in D_h$ such that $r_0 - 1,\ r_0 + 1 \in D_h$. Then, $h(r_0) \geqslant \max\{h(r_0-1),\ h(r_0+1)\}$ iff $h(r_0)$ is the global maximum of h.*

A.1.1 Proof of Lemma A.1

Assume first that $h(r_0)$ is the global maximum of h. This implies that $h(r_0) \geqslant h(r_0+1)$ and $h(r_0) \geqslant h(r_0-1)$. As a result, $h(r_0) \geqslant \max\{h(r_0-1),\ h(r_0+1)\}$.

Assume now that $h(r_0) \geqslant \max\{h(r_0-1),\ h(r_0+1)\}$ and let's prove that $h(r_0)$ is the global maximum of h.

As pointed out by Murota (1998), a discrete function is concave if and only if it can be extended to a continuous concave function. As h is concave, let $\bar{h}$ be an extension of h on $\bar{D}$, where $\bar{h}$ is concave and $\bar{h}(r) = h(r),\ \forall\ r \in D_h$. In particular, one can choose $\bar{h}$ by linearly joining $h(r_0-1)$ to $h(r_0)$ and then $h(r_0)$ to $h(r_0+1)$. Thus, $\bar{h}$ is linear on $[r_0-1,\ r_0]$ and on $[r_0,\ r_0+1]$. This implies that $\bar{h}(r_0)$ is a local maximum of $\bar{h}$ on $[r_0-1,\ r_0+1]$. As $\bar{h}$ is concave, $\bar{h}(r_0)$ is the global maximum of h.

A.1.2 Proof of Proposition 2.1

The expected outcome is $U^e\left(y_i, \mathbf{y}_{-i}\right) = \psi_i y_i - c(y_i) - \dfrac{\lambda}{2}\mathbb{E}_{\mathbf{y}_{-i}}\left[\,(y_i - \bar{y}_i)^2\,\right] + e_i(y_i)$.
Under Assumptions 2.1 and 2.2, $c(.)$ is convex and $e_i(.)$ is linear, then $U^e\left(., \mathbf{y}_{-i}\right)$ is strictly concave. As a result, there is a unique $r_0 \in \mathbb{N}$ at which $U^e\left(., \mathbf{y}_{-i}\right)$ is maximized.
The second part of Proposition 2.1 is given by Lemma A.1. Given that $U\left(r_0, \mathbf{y}^e_{-i}\right)$ is the global maximum of $U\left(., \mathbf{y}^e_{-i}\right)$, then $U\left(r, \mathbf{y}^e_{-i}\right) \geqslant \max\left\{U\left(r-1, \mathbf{y}^e_{-i}\right),\ U\left(r+1, \mathbf{y}^e_{-i}\right)\right\}$ iff $r = r_0$.

A.2 Proof of the Convergence of Infinite Summations

Many infinite summations appear in the paper, and there is no reason to *a priori* believe that these quantities are finite (e.g., the expected choice, the infinite summations in Proposition 2.2, Assumption

2.3, and several others used throughout the proofs and in the limiting distribution). In this section, I state and prove a general lemma on the convergence of these infinite sums.

Lemma A.2. *Let h be a continuous function on $\mathbb{R}$ and f_γ be a function defined for any $u \in \mathbb{R}$ as $f_\gamma(u) = \sum_{r=0}^{+\infty} r^\gamma h(u - b_r)$, where $\gamma \geqslant 0$ and $(b_k)_{k\in\mathbb{N}}$ is an increasing sequence, such that $\lim_{r\to\infty} b_{r+1} - b_r > 0$. The following statements hold.*

(a) For any $u \in \mathbb{R}$, if $h(x) = o(|x|^{-\alpha})$, for some $\alpha > \gamma + 1$ at $-\infty$, then $f_\gamma(u) < \infty$.

(b) if $h(x) = o(|x|^{-\alpha})$, for some $\alpha > 1$ at both $-\infty$ and $+\infty$, then f_0 is bounded on $\mathbb{R}$.

Statement (b) and Assumptions 2.2 and 2.1 ensure that B_c defined in Assumption 2.3 is finite. On the other hand, Statement (a) and Assumptions 2.2 and 2.1 also imply that all the infinite summations in the paper are finite.

Proof of Lemma A.2

The proof is done in several steps.

Step 0: I show that if $h(x) = o\left(|x|^{-\alpha}\right)$ for some $\alpha > 1$ at both $-\infty$ and $+\infty$, then $\exists\ M \geqslant 1$, such that $|h(u - b_r)| \leqslant M\left(|u - b_r| + 1\right)^{-\alpha}$. Moreover, this is also true for large r even if $h(x) = o\left(|x|^{-\alpha}\right)$ only at $-\infty$.

$h(x) = o\left(|x|^{-\alpha}\right)$ at both $-\infty$ and $+\infty$ $\implies |h(x)| = o\left((|x| + 1)^{-\alpha}\right)$.
Thus, $\exists\ x_0 \in \mathbb{R}_+ \ /\ \forall\ x < -x_0$ or $x > x_0,\ |h(x)| < (|x| + 1)^{-\alpha}$.
As h is continuous, this implies that $\exists\ M \geqslant 1 \ /\ \forall\ x \in \mathbb{R},\ |h(x)| \leqslant M\left(|x| + 1\right)^{-\alpha}$. It follows that $\forall\ u \in \mathbb{R},\ r \in \mathbb{N}$,

$$|h(u - b_r)| \leqslant M\left(|u - b_r| + 1\right)^{-\alpha}. \tag{14}$$

Step 1: I prove Statement (a).

Let f^* be the real-valued function defined as $f^*(u) = \sum_{r=0}^{\infty} \left(|u - b_r| + 1\right)^{-\alpha},\ \forall\ u \in \mathbb{R}$.
$\lim_{r\to\infty} b_{r+1} - b_r > 0 \implies \exists\ r^* \in \mathbb{N}$ and $b > 0$ such that $\forall\ r \geqslant r^*$, $b_{r+1} - b_r \geqslant b$. As $\lim_{r\to\infty} b_r = \infty$, $\forall\ u \in \mathbb{R}$, it is possible to choose r^* sufficiently large, such that $b_{r*} > u$. It follows that $\forall\ r \geqslant r^*$,

$$\begin{aligned}
&b_r \geqslant (r - r^*)b + b_{r*},\\
&|u - b_r| = b_r - u \geqslant (r - r^*)b + b_{r*} - u \geqslant 0,\\
&\left(|u - b_r| + 1\right)^{-\alpha} \leqslant \left((r - r^*)b + b_{r*} - u\right)^{-\alpha},\\
&\left(|u - b_r| + 1\right)^{-\alpha} \leqslant O\left(r^{-\alpha}\right).
\end{aligned} \tag{15}$$

Inequation (15) implies that $f^*(u) < \infty$, $\forall\ u \in \mathbb{R}$. Using the result of the step 0, it follows that, $\forall\ u \in \mathbb{R},\ \gamma \geqslant 0,\ r^\gamma h(u - b_r) = O\left(r^{-(\alpha-\gamma)}\right)$. Hence, $f_\gamma(u) < \infty$ if $\alpha > \gamma + 1$.

Step 2: I prove Statement (b).

As Equation (14) holds for any r if $h(x) = o\left(|x|^{-\alpha}\right)$ at both $-\infty$ and $+\infty$, to prove that f_0 is bounded it is sufficient to prove that f^* is also bounded. As f^* is a continuous function, it is also sufficient to prove that $\lim_{u \to -\infty} f^*(u) < \infty$ and $\lim_{u \to +\infty} f^*(u) < \infty$.
If $u \leqslant 0$, then $\left(|u - b_r| + 1\right)^{-\alpha} = \left(b_r - u + 1\right)^{-\alpha} \leqslant \left(b_r + 1\right)^{-\alpha}$. Thus, $f^*(u) \leqslant f^*(0)$.
As f^* is a positive function, then $\lim_{u \to -\infty} f^*(u) < \infty$.

Let $k_0 \in \mathbb{N}^*$ such that $\forall\ r \geqslant k_0$, $b_{r+1} - b_r \geqslant b$, for some $b > 0$.
For u sufficiently large, $\exists\ k^* \in \mathbb{N}$ (with k^* depending on u), where $k^* > k_0$ and $\forall\ r \leqslant k^* - 1$, $u > b_r$, and $\forall\ r \geqslant k^*$, $u \leqslant b_r$. Thus, $f^*(u)$ can be decomposed as

$$
\begin{aligned}
f^*(u) &= \sum_{r=0}^{k_0-1} \left(|u - b_r| + 1\right)^{-\alpha} + \sum_{r=k_0}^{k^*-1} \left(|u - b_r| + 1\right)^{-\alpha} + \sum_{r=k^*}^{\infty} \left(|u - b_r| + 1\right)^{-\alpha}, \\
f^*(u) &\leqslant k_0 + \sum_{r=k_0}^{k^*-1} \left(u - b_r + 1\right)^{-\alpha} + \sum_{r=k^*}^{\infty} \left(b_r - u + 1\right)^{-\alpha}, \\
f^*(u) &\leqslant k_0 + \sum_{r=k_0}^{k^*-1} \left(b_{k^*-1} - b_r + 1\right)^{-\alpha} + \sum_{r=k^*}^{\infty} \left(b_r - b_{k^*} + 1\right)^{-\alpha}.
\end{aligned}
$$

If $k_0 \leqslant r \leqslant k^* - 1$, then $b_{k^*-1} - b_r \geqslant (k^* - 1 - r)b$, because $b_{r+1} - b_r \geqslant b$.
Thus, $\left(b_{k^*-1} - b_r + 1\right)^{-\alpha} \leqslant \left((k^* - 1 - r)b + 1\right)^{-\alpha}$.
Analogously, if $k^* \leqslant r$, then $b_r - b_{k^*} \geqslant (r - k^*)b$.
Thus, $\left(b_r - b_{k^*} + 1\right)^{-\alpha} \leqslant \left((r - k^*)b + 1\right)^{-\alpha}$.
Finally,

$$
\begin{aligned}
f^*(u) &\leqslant k_0 + \sum_{r=k_0}^{k^*-1} \left((k^* - 1 - r)b + 1\right)^{-\alpha} + \sum_{r=k^*}^{\infty} \left((r - k^*)b + 1\right)^{-\alpha}, \\
f^*(u) &\leqslant k_0 + \sum_{r=0}^{k^*-k_0-1} \left(br + 1\right)^{-\alpha} + \sum_{r=0}^{\infty} \left(br + 1\right)^{-\alpha}, \\
f^*(u) &\leqslant k_0 + 2\sum_{r=0}^{\infty} \left(br + 1\right)^{-\alpha}.
\end{aligned}
$$

The quantity $k_0 + 2\sum_{r=0}^{\infty} \left(br + 1\right)^{-\alpha}$ does not depend on u and is finite. Hence, $\lim_{u \to +\infty} f^*(u) < \infty$.
As a result, f_0 is bounded.

A.3 Proof of Proposition 2.2

For any $\mathbf{y}^e \in \mathbb{R}_+^n$, $\mathbf{L}(\mathbf{y}^e) = (\ell_1(\mathbf{y}^e)\ \ldots \ell_n(\mathbf{y}^e))'$, where $\ell_i(\mathbf{y}^e) = \sum_{r=1}^{\infty} F_\varepsilon(\lambda \bar{y}_i^e + \psi_i - a_r)$ for all $i \in \mathcal{V}$ and $\bar{y}_i^e = \sum_{j=1}^{n} g_{ij} y_j^e$.

Assume that $\mathbf{p} = (p_{ir})$ is rational and let $\mathbf{y}^e$ be the associated expected outcome. $\mathbf{p}$ and $\mathbf{y}^e$ verify (3). Thus,

$$
\begin{aligned}
p_{ir} &= F_\varepsilon(\lambda \bar{y}_i^e + \psi_i - a_r) - F_\varepsilon(\lambda \bar{y}_i^e + \psi_i - a_{r+1}), \\
y_i^e &= \sum_{r=0}^{\infty} r p_{ir} = \underbrace{\sum_{r=0}^{\infty} r F_\varepsilon(\lambda \bar{y}_i^e + \psi_i - a_r)}_{S_1} - \underbrace{\sum_{r=0}^{\infty} r F_\varepsilon(\lambda \bar{y}_i^e + \psi_i - a_{r+1})}_{S_2}.
\end{aligned} \tag{16}
$$

Equation (16) holds because $S_1 < \infty$ and $S_2 < \infty$. To prove this, let $x < 0$ with $|x|$ being sufficiently large. By Assumption 2.2, $f_\varepsilon(x) = o\left(|x|^{-\kappa}\right)$ at ∞ for $\kappa > 3$. Then, $F_\varepsilon(x) = O\left(|x|^{-\kappa+1}\right)$ at $-\infty$, and $F_\varepsilon(x) = o(|x|^{-\frac{\kappa+1}{2}})$ at $-\infty$. By Lemma A.2, $S_1 < \infty$, and $S_2 < \infty$.

$$
\begin{aligned}
y_i^e &= \sum_{r=0}^{\infty} r F_\varepsilon(\lambda \bar{y}_i^e + \psi_i - a_r) - \sum_{r=0}^{\infty} (r+1) F_\varepsilon(\lambda \bar{y}_i^e + \psi_i - a_{r+1}) + \sum_{r=0}^{\infty} F_\varepsilon(\lambda \bar{y}_i^e + \psi_i - a_{r+1}), \\
y_i^e &= \sum_{r=1}^{\infty} r F_\varepsilon(\lambda \bar{y}_i^e + \psi_i - a_r) - \sum_{r=1}^{\infty} r F_\varepsilon(\lambda \bar{y}_i^e + \psi_i - a_r) + \sum_{r=0}^{\infty} F_\varepsilon(\lambda \bar{y}_i^e + \psi_i - a_{r+1}), \\
y_i^e &= \sum_{r=1}^{\infty} F_\varepsilon(\lambda \bar{y}_i^e + \psi_i - a_r) = \ell_i(\mathbf{y}^{*e}).
\end{aligned}
$$

Hence, $\mathbf{y}^e = \mathbf{L}(\mathbf{y}^e)$.

A.4 Proof of Theorem 2.1

A belief system $\mathbf{p} = (p_{ir})$ is rational if it verifies Equation (3).

Let's prove the existence of a rational belief system.

Let us denote by $\mathbf{p}_{\mathrm{r}} = (p_{1r}, \ldots, p_{nr})'$, an n-dimensional vector for any $r \in \mathbb{N}$, $\mathbf{p} = \left(\mathbf{p}_0', \mathbf{p}_1', \mathbf{p}_2', \mathbf{p}_3', \ldots\right)'$, $\mathbf{h}_1 = (a_0, a_1, a_2, a_3, \ldots)'$, $\mathbf{h}_2 = (a_1, a_2, a_3, a_4, \ldots)'$ infinite-dimensional vectors, and $\mathbf{1}_d$, the d-dimensional vector of ones for any $d \in \mathbb{N}^*$ or $d = \infty$. Let $\mathbf{J} = (0, 1, 2, 3, \ldots)$, an infinite-dimensional row-vector, and $\mathbf{B} = \mathbf{1}_\infty \otimes \mathbf{J} \otimes \mathbf{G}$. Let also $\mathbf{F}_\varepsilon$ be a mapping defined for any $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3, \ldots) \in \mathbb{R}^\infty$ as $\mathbf{F}_\varepsilon(\boldsymbol{\omega}) = (F_\varepsilon(\omega_1), F_\varepsilon(\omega_2), F_\varepsilon(\omega_3), \ldots)$. Equation (3) in matrix form is given by $\mathbf{p} = \mathbf{H}(\mathbf{p})$, where

$$
\mathbf{H}(\mathbf{p}) = \mathbf{F}_\varepsilon(\lambda \mathbf{B}\mathbf{p} + \mathbf{1}_\infty \otimes \Psi - \mathbf{h}_1 \otimes \mathbf{1}_n) - \mathbf{F}_\varepsilon(\lambda \mathbf{B}\mathbf{p} + \mathbf{1}_\infty \otimes \Psi - \mathbf{h}_2 \otimes \mathbf{1}_n). \tag{17}
$$

$\mathbf{H}$ is a continuous. By Schauder's fixed point theorem (generalization of Brouwer's fixed point theorem to an infinite-dimensional space, see Smart, 1980, Chapter 2), there exists $\mathbf{p}^* \in [0,\ 1]^\infty$, such that $\mathbf{p}^* = \mathbf{H}(\mathbf{p}^*)$. As a result, a rational belief system exists.

Let us prove the uniqueness of rational belief system.

If $\mathbf{p}^*$ is a rational belief system, Proposition 2.2 states that its associated expected outcome $\mathbf{y}^{e*} = (y_1^{e*} \dots y_n^{e*})$ verifies $\mathbf{y}^{e*} = \mathbf{L}(\mathbf{y}^{e*})$. To prove the uniqueness, it is sufficient to establish that $\mathbf{L}$ does not have more than one fixed point. By contracting mapping theorem, it is sufficient to prove that $\mathbf{L}$ is contracting; that is, $\forall\ \mathbf{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$, $\left|\left|\dfrac{\partial \mathbf{L}(\mathbf{u})}{\partial \mathbf{u}'}\right|\right|_\infty \leqslant \bar{\kappa_c}$ for some $\bar{\kappa_c} < 1$ not depending on $\mathbf{u}$. Let $\mathbf{g}_i$ be the i-th row of $\mathbf{G}$. For all i and j,

$$\frac{\partial \ell_i(\mathbf{u})}{\partial u_j} = \lambda g_{ij} \underbrace{\sum_{r=1}^{\infty} f_\varepsilon(\lambda \mathbf{g}_i \mathbf{u} + \psi_i - a_r)}_{f_i^*} = \lambda g_{ij} f_i^*. \tag{18}$$

From Equation (18), $\dfrac{\partial \mathbf{L}(\mathbf{u})}{\partial \mathbf{u}'}$ is defined by

$$\frac{\partial \mathbf{L}(\mathbf{u})}{\partial \mathbf{u}'} = \lambda \begin{pmatrix} g_{11} f_1^* & \dots & g_{1n} f_1^* \\ \vdots & \ddots & \vdots \\ g_{n1} f_N^* & \dots & g_{nn} f_N^* \end{pmatrix}.$$

It follows that

$$\begin{aligned} \left|\left|\frac{\partial \mathbf{L}(\mathbf{u})}{\partial \mathbf{u}'}\right|\right|_\infty &= \lambda \max_i \left\{ f_i^* \sum_{j=1}^{n} g_{ij} \right\} \leqslant \lambda \left(\max_i f_i^*\right) \max_i \left\{ \sum_{j=1}^{n} g_{ij} \right\}, \\ \left|\left|\frac{\partial \mathbf{L}(\mathbf{u})}{\partial \mathbf{u}'}\right|\right|_\infty &\leqslant \lambda \left(\max_i f_i^*\right) ||\mathbf{G}||_\infty. \end{aligned} \tag{19}$$

I now focus on the term f_i^*.

$$\begin{aligned} f_i^* &= \sum_{r=1}^{\infty} f_\varepsilon(\lambda \mathbf{g}_i \mathbf{u} + \psi_i - a_r), \\ f_i^* &\leqslant \max_{u \in \mathbb{R}} \sum_{k=1}^{\infty} f_\varepsilon(u - a_r) = \frac{1}{B_c}. \end{aligned} \tag{20}$$

From Equations (19) and (20),

$$\left|\left|\frac{\partial \mathbf{L}(\mathbf{u})}{\partial \mathbf{u}'}\right|\right|_\infty \leqslant \frac{\lambda ||\mathbf{G}||_\infty}{B_c} < 1 \text{ by Assumption 2.3.} \tag{21}$$

Hence, $\mathbf{L}$ is a contracting mapping and does not have more than one fixed point.

As a result, there is a unique rational belief system $\mathbf{p}^*$, where the associated expected outcome $\mathbf{y}^{e*}$ is the unique solution of $\mathbf{y}^e = \mathbf{L}(\mathbf{y}^e)$.

A.5 Proof of Proposition 3.1

According to Definition 3.1, two sets of parameters, $\boldsymbol{\Lambda}_{(1)} = \left(\lambda_{(1)}, \boldsymbol{\theta}'_{(1)}, \boldsymbol{\delta}'_{(1)}\right)'$ and $\boldsymbol{\Lambda}_{(2)} = \left(\lambda_{(2)}, \boldsymbol{\theta}'_{(2)}, \boldsymbol{\delta}'_{(2)}\right)'$, are observationally equivalent if $p_{ir}^{(1)} = \Phi\left(\lambda_{(1)}\bar{y}^e_{i(1)} + \mathbf{z}'_i\boldsymbol{\theta}_{(1)} - a_r^{(1)}\right) - \Phi\left(\lambda_{(1)}\bar{y}^e_{i(1)} + \mathbf{z}'_i\boldsymbol{\theta}_{(1)} - a_{r+1}^{(1)}\right) =$
$p_{ir}^{(2)} = \Phi\left(\lambda_{(2)}\bar{y}^e_{i(2)} + \mathbf{z}'_i\boldsymbol{\theta}_{(2)} - a_r^{(2)}\right) - \Phi\left(\lambda_{(2)}\bar{y}^e_{i(2)} + \mathbf{z}'_i\boldsymbol{\theta}_{(2)} - a_{r+1}^{(2)}\right)$, where $y^e_{i(1)} = \sum_{r=1}^{\infty} r p_{ir}^{(1)} = \sum_{r=1}^{\infty} r p_{ir}^{(2)} = y^e_{i(2)}$, and for $s \in \{1,2\}$, $a_0^{(s)} = -\infty$, $a_r^{(s)} = \sum_{k=1}^{r}\delta_k^{(1)}$ if $1 \leqslant r < \bar{\mathrm{R}}$, $\delta_1^{(s)} = 0$, and $a_r^{(s)} = (r - \bar{\mathrm{R}})\delta_{\bar{\mathrm{R}}}^{(s)} + \sum_{k=1}^{\bar{\mathrm{R}}}\delta_k^{(s)}$ if $r \geqslant \bar{\mathrm{R}}$.

For $s \in \{1,2\}$, $p_{i0}^{(s)} = 1 - \Phi\left(\lambda_{(1)}\bar{y}^e_{i(1)} + \mathbf{z}'_i\boldsymbol{\theta}_{(1)}\right)$. Thus, if $p_{i0}^{(1)} = p_{i0}^{(2)}$, then $\lambda_{(1)}\bar{y}^e_{i(1)} + \mathbf{z}'_i\boldsymbol{\theta}_{(1)} = \lambda_{(2)}\bar{y}^e_{i(2)} + \mathbf{z}'_i\boldsymbol{\theta}_{(2)}$. Since $\bar{y}^e_{i(1)} = \bar{y}^e_{i(2)}$, to establish that $\lambda_{(1)} = \lambda_{(2)}$ and $\boldsymbol{\theta}_{(1)} = \boldsymbol{\theta}_{(2)}$, it is sufficient to prove that $[\mathbf{G}\mathbf{y}^e\ \mathbf{Z}]$ is a full rank matrix. Let us recall that $\mathbf{Z} = [\mathbf{W}\ \mathbf{X}\ \mathbf{G}\mathbf{X}]$. By Assumption 3.3, $\check{\mathbf{Z}} = \left[\mathbf{W}\ \mathbf{X}\ \mathbf{G}\mathbf{X}\ \mathbf{G}^2\mathbf{X}\right]$ is a full rank matrix, then so is $\mathbf{Z}$. It follows that $[\mathbf{G}\mathbf{y}^e\ \mathbf{Z}]$ is a full rank matrix if $\mathbf{G}\mathbf{y}^e$ cannot be expressed as $\mathbf{Z}\tilde{\boldsymbol{\theta}}$, where $\tilde{\boldsymbol{\theta}}$ is an unknown parameter.

The rational expected outcome verifies $y_i^e = \sum_{r=1}^{\infty}\Phi\left(\tilde{\mathbf{z}}'_i\boldsymbol{\Lambda} - a_r\right)$, where $\boldsymbol{\Lambda} = (\lambda,\ \boldsymbol{\theta}')'$ and $\tilde{\mathbf{z}}_i = (\bar{y}^e_i,\ \mathbf{z}'_i)$. As y_i^e is unbounded, if $\mathbf{G}\mathbf{y}^e = \mathbf{Z}\tilde{\boldsymbol{\theta}}$, then $\mathbf{z}_i$, $\mathbf{x}_i$, and $\tilde{\mathbf{z}}_i\boldsymbol{\Lambda}$ are also unbounded. In particular, $\tilde{\mathbf{z}}_i\boldsymbol{\Lambda}$ is only unbounded on the top. In fact, if $\tilde{\mathbf{z}}_i\boldsymbol{\Lambda}$ grows to $-\infty$, then $y_i^e \approx 0$, and the condition $\mathbf{G}\mathbf{y}^e = \mathbf{Z}\tilde{\boldsymbol{\theta}}$ would not hold.

The identification proof is based on the fact that $\sum_{r=1}^{\infty}\Phi\left(\tilde{\mathbf{z}}'_i\boldsymbol{\Lambda} - a_r\right)$ can be approximated by a linear equation in $\mathbf{z}'_i\boldsymbol{\Lambda}$ with a bounded rest.

$$\frac{\partial y_i^e}{\partial\left(\tilde{\mathbf{z}}'_i\boldsymbol{\Lambda}\right)} = \sum_{r=1}^{\infty}\phi\left(\tilde{\mathbf{z}}'_i\boldsymbol{\Lambda} - a_r\right) = \sum_{r=1}^{\bar{\mathrm{R}}-1}\phi\left(\tilde{\mathbf{z}}'_i\boldsymbol{\Lambda} - a_r\right) + \sum_{r=\bar{\mathrm{R}}}^{\infty}\phi\left(\tilde{\mathbf{z}}'_i\boldsymbol{\Lambda} - a_{\bar{\mathrm{R}}} - \delta_{\bar{\mathrm{R}}}(r - \bar{\mathrm{R}})\right).$$

For $\tilde{\mathbf{z}}'_i\boldsymbol{\Lambda}$ positive and sufficiently large, $\sum_{r=\bar{\mathrm{R}}}^{\infty}\phi\left(\tilde{\mathbf{z}}'_i\boldsymbol{\Lambda} - a_{\bar{\mathrm{R}}} - \delta_{\bar{\mathrm{R}}}(r - \bar{\mathrm{R}})\right) \approx \sum_{r=-\infty}^{\infty}\phi\left(\tilde{\mathbf{z}}'_i\boldsymbol{\Lambda} + \delta_{\bar{\mathrm{R}}}r\right)$ and $\sum_{r=1}^{\bar{\mathrm{R}}-1}\phi\left(\tilde{\mathbf{z}}'_i\boldsymbol{\Lambda} - a_r\right) \approx 0$. Let $u = \tilde{\mathbf{z}}'_i\boldsymbol{\Lambda}$. Thus,

$$\frac{\partial y_i^e}{\partial\left(\tilde{\mathbf{z}}'_i\boldsymbol{\Lambda}\right)} \approx \sum_{r=-\infty}^{\infty}\phi\left(u + \delta_{\bar{\mathrm{R}}}r\right). \tag{22}$$

Let us focus on the quantity $\sum_{r=-\infty}^{\infty}\phi\left(u + \delta_{\bar{\mathrm{R}}}r\right)$. I simplify this expression using the Poisson summation formula (see Bellman, 2013, Section 6).

$$\sum_{r=-\infty}^{\infty} \phi\left(u + \delta_{\bar{\mathrm{R}}} r\right) = \sum_{r=-\infty}^{\infty} \tilde{f}\left(u + \delta_{\bar{\mathrm{R}}} r\right), \tag{23}$$

where $\tilde{f}$ is the Fourier transform of ϕ given by

$$\tilde{f}\left(u + \delta_{\bar{\mathrm{R}}} r\right) = \int_{-\infty}^{+\infty} \phi(u + \delta_{\bar{\mathrm{R}}} x) e^{-2\pi i r x} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\delta_{\bar{\mathrm{R}}} x + u)^2 - 2\pi i r x} dx. \tag{24}$$

In Equation (24), i is the pure imaginary complex number ($i^2 = -1$).

$$\tilde{f}\left(u + \delta_{\bar{\mathrm{R}}} r\right) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\delta_{\bar{\mathrm{R}}}^2 x^2 + 2u\delta_{\bar{\mathrm{R}}} x + u^2 + 4\pi i r x\right)} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\delta_{\bar{\mathrm{R}}}^2}{2}\left(x^2 + 2\frac{u}{\delta_{\bar{\mathrm{R}}}} x + \frac{4\pi i r}{\delta_{\bar{\mathrm{R}}}^2} x + \frac{u^2}{\delta_{\bar{\mathrm{R}}}^2}\right)} dx,$$

$$\tilde{f}\left(u + \delta_{\bar{\mathrm{R}}} r\right) = \frac{1}{\delta_{\bar{\mathrm{R}}}} e^{\frac{\delta_{\bar{\mathrm{R}}}^2}{2}\left(\frac{u}{\delta_{\bar{\mathrm{R}}}} + \frac{2\pi i r}{\delta_{\bar{\mathrm{R}}}^2}\right)^2 - \frac{u^2}{2}} \underbrace{\int_{-\infty}^{+\infty} \frac{\delta_{\bar{\mathrm{R}}}}{\sqrt{2\pi}} e^{-\frac{\delta_{\bar{\mathrm{R}}}^2}{2}\left(x + \frac{u}{\delta_{\bar{\mathrm{R}}}} + \frac{2\pi i r}{\delta_{\bar{\mathrm{R}}}^2}\right)^2} dx}_{=1},$$

$$\tilde{f}\left(u + \delta_{\bar{\mathrm{R}}} r\right) = \frac{1}{\delta_{\bar{\mathrm{R}}}} e^{-\frac{2\pi^2 r^2}{\delta_{\bar{\mathrm{R}}}^2} + \frac{2\pi i r u}{\delta_{\bar{\mathrm{R}}}}}. \tag{25}$$

By replacing the Fourier transform (25) in Equation (23) and in Equation (22),

$$\frac{\partial y_i^e}{\partial\left(\tilde{\mathbf{z}}_i' \mathbf{\Lambda}\right)} \approx \frac{1}{\delta_{\bar{\mathrm{R}}}} \sum_{r=-\infty}^{\infty} e^{-\frac{2\pi^2 r^2}{\delta_{\bar{\mathrm{R}}}^2} + \frac{2\pi i r u}{\delta_{\bar{\mathrm{R}}}}},$$

$$\frac{\partial y_i^e}{\partial\left(\tilde{\mathbf{z}}_i' \mathbf{\Lambda}\right)} \approx \frac{1}{\delta_{\bar{\mathrm{R}}}} + \frac{1}{\delta_{\bar{\mathrm{R}}}} \sum_{r=1}^{\infty} e^{-\frac{2\pi^2 (-r)^2}{\delta_{\bar{\mathrm{R}}}^2}} e^{-\frac{2\pi i r u}{\delta_{\bar{\mathrm{R}}}}} + \frac{1}{\delta_{\bar{\mathrm{R}}}} \sum_{r=1}^{\infty} e^{-\frac{2\pi^2 r^2}{\delta_{\bar{\mathrm{R}}}^2}} e^{\frac{2\pi i r u}{\delta_{\bar{\mathrm{R}}}}},$$

$$\frac{\partial y_i^e}{\partial\left(\tilde{\mathbf{z}}_i' \mathbf{\Lambda}\right)} \approx \frac{1}{\delta_{\bar{\mathrm{R}}}} + \frac{1}{\delta_{\bar{\mathrm{R}}}} \sum_{r=1}^{\infty} e^{-\frac{2\pi^2 r^2}{\delta_{\bar{\mathrm{R}}}^2}} \left(e^{-\frac{2\pi i r u}{\delta_{\bar{\mathrm{R}}}}} + e^{\frac{2\pi i r u}{\delta_{\bar{\mathrm{R}}}}}\right). \tag{26}$$

By Euler's formula,

$$e^{-\frac{2\pi i r u}{\delta_{\bar{\mathrm{R}}}}} + e^{\frac{2\pi r u}{\delta_{\bar{\mathrm{R}}}}} = \cos\left(-\frac{2\pi r u}{\delta_{\bar{\mathrm{R}}}}\right) + i \sin\left(-\frac{2\pi r u}{\delta_{\bar{\mathrm{R}}}}\right) + \cos\left(\frac{2\pi r u}{\delta_{\bar{\mathrm{R}}}}\right) + i \sin\left(\frac{2\pi r u}{\delta_{\bar{\mathrm{R}}}}\right),$$

$$e^{-\frac{2\pi r u}{\delta_{\bar{\mathrm{R}}}}} + e^{\frac{2\pi r u}{\delta_{\bar{\mathrm{R}}}}} = 2\cos\left(\frac{2\pi r u}{\delta_{\bar{\mathrm{R}}}}\right). \tag{27}$$

By replacing (27) in (26),

$$\frac{\partial y_i^e}{\partial\left(\tilde{\mathbf{z}}_i' \mathbf{\Lambda}\right)} \approx \frac{1}{\delta_{\bar{\mathrm{R}}}} + \frac{2}{\delta_{\bar{\mathrm{R}}}} \sum_{r=1}^{\infty} e^{-\frac{2\pi^2 r^2}{\delta_{\bar{\mathrm{R}}}^2}} \cos\left(\frac{2\pi r \tilde{\mathbf{z}}_i' \mathbf{\Lambda}}{\delta_{\bar{\mathrm{R}}}}\right).$$

Then, for large $\tilde{\mathbf{z}}_i'\mathbf{\Lambda}$ it follows that

$$y_i^e \approx \frac{1}{\delta_{\bar{\mathrm{R}}}}\tilde{\mathbf{z}}_i'\mathbf{\Lambda} + 2\sum_{r=1}^{\infty}\frac{1}{2\pi r}e^{-\frac{2\pi^2 r^2}{\delta_{\bar{\mathrm{R}}}^2}}\sin\left(\frac{2\pi r\tilde{\mathbf{z}}_i'\mathbf{\Lambda}}{\delta_{\bar{\mathrm{R}}}}\right) + C, \tag{28}$$

where C is a constant (not depending on $\tilde{\mathbf{z}}_i'\mathbf{\Lambda}$). As $\sum_{r=1}^{\infty}\frac{1}{2\pi r}e^{-\frac{2\pi^2 r^2}{\delta_{\bar{\mathrm{R}}}^2}}\sin\left(\frac{2\pi r\tilde{\mathbf{z}}_i'\mathbf{\Lambda}}{\delta_{\bar{\mathrm{R}}}}\right) + C$ is bounded, I can write $y_i^e - \frac{1}{\delta_{\bar{\mathrm{R}}}}\tilde{\mathbf{z}}_i'\mathbf{\Lambda} = Re_i$, where Re_i is the linear approximation error and Re_i is bounded. Figure 2 plots y_i^e as function of $\tilde{\mathbf{z}}_i'\mathbf{\Lambda}$.

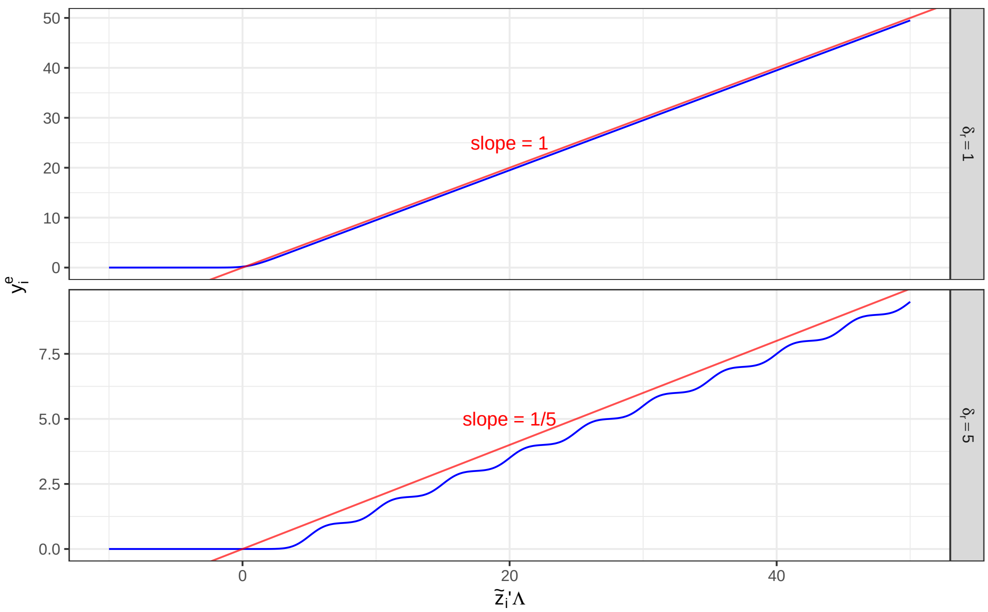

Figure 2: Expected outcome

This figure presents y_i^e as function of $\tilde{\mathbf{z}}_i'\mathbf{\Lambda}$ (*blue line*) and a straight line of null intercept and slope $\frac{1}{\delta_{\bar{\mathrm{R}}}}$ (*red line*). In this illustration, the cost function is quadratic (i.e., δ_r is constant for any r). As suggested by Equation (28), y_i^e can be approximated with a straight line of slope $\frac{1}{\delta_{\bar{\mathrm{R}}}}$ when $\tilde{\mathbf{z}}_i'\mathbf{\Lambda}$ is large.

It then follows that $\mathbf{G}\mathbf{y}^e - \frac{1}{\delta_{\bar{\mathrm{R}}}}\mathbf{G}\tilde{\mathbf{Z}}'\mathbf{\Lambda}$ is bounded. I now show that this cannot hold in general (except for special cases of $\mathbf{Z}$).

As $\mathbf{G}\mathbf{y}^e = \mathbf{Z}\tilde{\boldsymbol{\theta}}$, then $\tilde{\mathbf{Z}}'\mathbf{\Lambda} = \left[\mathbf{Z}\tilde{\boldsymbol{\theta}},\ \mathbf{Z}\right](\lambda,\ \boldsymbol{\theta}')' = \mathbf{Z}(\lambda\tilde{\boldsymbol{\theta}} + \boldsymbol{\theta})$. Thus, $\mathbf{G}\mathbf{y}^e - \frac{1}{\delta_{\bar{\mathrm{R}}}}\mathbf{G}\tilde{\mathbf{Z}}'\mathbf{\Lambda} = \mathbf{Z}\tilde{\boldsymbol{\theta}} - \frac{1}{\delta_{\bar{\mathrm{R}}}}\mathbf{G}\mathbf{Z}(\lambda\tilde{\boldsymbol{\theta}} + \boldsymbol{\theta})$. If there is no isolated individuals, I have

$$\mathbf{G}\mathbf{y}^e - \frac{1}{\delta_{\bar{\mathrm{R}}}}\mathbf{G}\tilde{\mathbf{Z}}'\mathbf{\Lambda} = [\mathbf{W}\ \mathbf{X}\ \mathbf{G}\mathbf{X}]\tilde{\boldsymbol{\theta}} - \frac{1}{\delta_{\bar{\mathrm{R}}}}[\mathbf{W}\ \mathbf{G}\mathbf{X}\ \mathbf{G}^2\mathbf{X}](\lambda\tilde{\boldsymbol{\theta}} + \boldsymbol{\theta}). \tag{29}$$

Equation (29) implies that $[\mathbf{W}\ \mathbf{X}\ \mathbf{G}\mathbf{X}]\tilde{\boldsymbol{\theta}} - \frac{1}{\delta_{\bar{\mathrm{R}}}}[\mathbf{W}\ \mathbf{G}\mathbf{X}\ \mathbf{G}^2\mathbf{X}](\lambda\tilde{\boldsymbol{\theta}} + \boldsymbol{\theta})$ is bounded. As $\mathbf{x}_i$ is unbounded, this is only possible in general if $[\mathbf{W}\ \mathbf{X}\ \mathbf{GX}]\tilde{\boldsymbol{\theta}} - \dfrac{1}{\delta_{\bar{\mathrm{R}}}}[\mathbf{W}\ \mathbf{GX}\ \mathbf{G}^2\mathbf{X}](\lambda\tilde{\boldsymbol{\theta}} + \boldsymbol{\theta}) = \mathbf{0}$; i.e., only in the case where the linear approximation error is null. In fact, the intuition of this is that, a linear combination of unbounded variables is supposed to be unbounded in general, except for special cases of variables. In particular, by Assumption 3.3, the coefficient multiplying $\mathbf{X}$ is null. Note that the coefficient multiplying $\mathbf{X}$ in Equation (29), comes from $\mathbf{G}\mathbf{y}^e = \mathbf{Z}\tilde{\boldsymbol{\theta}}$. Therefore, if this coefficient is null, then $\mathbf{G}\mathbf{y}^e$ does not depend on $\mathbf{X}$; ie $\mathbf{G}\mathbf{y}^e = \mathbf{W}\tilde{\boldsymbol{\alpha}} + \mathbf{GX}\tilde{\boldsymbol{\gamma}}$ for some parameters $\tilde{\boldsymbol{\alpha}}$ and $\tilde{\boldsymbol{\gamma}}$. As $\mathbf{y}^e$ is uniquely determined (by Theorem 2.1), this implies that $\mathbf{y}^e = \mathbf{W}\tilde{\boldsymbol{\alpha}} + \mathbf{X}\tilde{\boldsymbol{\gamma}}$. Thus, $\mathbf{y}^e$ does not depend on friends' average observable characteristics. This is not possible because $\boldsymbol{\gamma} \neq \mathbf{0}$ by Assumption 3.3.

As a result, $\lambda_{(1)} = \lambda_{(2)}$, and $\boldsymbol{\theta}_{(1)} = \boldsymbol{\theta}_{(2)}$. Moreover, given that $p_{i1}^{(1)} = p_{i1}^{(2)}, \ \dots, \ p_{i\bar{\mathrm{R}}}^{(1)} = p_{i\bar{\mathrm{R}}}^{(2)}$, it follows that $\boldsymbol{\delta}_{(1)} = \boldsymbol{\delta}_{(2)}$. Hence λ, $\boldsymbol{\theta}$, and $\boldsymbol{\delta}$ are identified.

A.6 Limiting Distribution Under the Assumption of an Exogenous Network

For simplification, I assume that the explanatory variable $\mathbf{z}_i$ and the network matrix $\mathbf{G}$ are nonstochastic in the counting variable model. This implies that the expected outcome y_i^e also is nonstochastic, as it only depends on $\mathbf{Z}$, $\mathbf{G}$, $\boldsymbol{\Gamma}$, and on the cdf F_ε.

The pseudo-log-likelihood is given by

$$\mathcal{L}(\boldsymbol{\Gamma}, \mathbf{y}^e) = \sum_{i=1}^{n} \mathcal{L}_i(\boldsymbol{\Gamma}, \mathbf{y}^e),$$

where $\mathcal{L}_i(\boldsymbol{\Gamma}, \mathbf{y}^e) = \sum_{r=0}^{\infty} d_{ir} \log\left(\Phi\left(\tilde{\mathbf{z}}_i'\boldsymbol{\Lambda} - a_r\right) - \Phi\left(\tilde{\mathbf{z}}_i'\boldsymbol{\Lambda} - a_{r+1}\right)\right)$, $\tilde{\mathbf{z}}_i' = (\bar{y}_i^e, \ \mathbf{z}_i')$, $\boldsymbol{\Lambda} = (\lambda, \ \boldsymbol{\theta}')'$, and $\boldsymbol{\Gamma} = \left(\boldsymbol{\Lambda}', \ \log\left(\boldsymbol{\delta}'\right)\right)'$, $a_0 = -\infty$, $a_r = \sum_{k=1}^{r} \delta_k$ if $1 \leqslant r < \bar{\mathrm{R}}$, $a_r = (r - \bar{\mathrm{R}})\delta_{\bar{\mathrm{R}}} + \sum_{k=1}^{\bar{\mathrm{R}}} \delta_k$ if $r \geqslant \bar{\mathrm{R}}$, and $\delta_1 = 0$.

Let $\boldsymbol{\Gamma}_0$ be the true value of $\boldsymbol{\Gamma}$ and $\mathbf{y}_0^e$ be the expected outcome associated with $\boldsymbol{\Gamma}_0$. The first-order conditions (f.o.c) of the pseudo-likelihood maximization give

$$\begin{cases} \dfrac{\partial \mathcal{L}(\boldsymbol{\Gamma}, \mathbf{y}^e)}{\partial \boldsymbol{\Lambda}} = \displaystyle\sum_{i=1}^{n} \sum_{r=0}^{\infty} d_{ir} \frac{\phi_{i,r} - \phi_{i,r+1}}{\Phi_{i,r} - \Phi_{i,r+1}} \tilde{\mathbf{z}}_i = 0, & \\ \dfrac{\partial \mathcal{L}(\boldsymbol{\Gamma}, \mathbf{y}^e)}{\partial \log(\delta_k)} = \delta_k \displaystyle\sum_{i=1}^{n} \sum_{r=k}^{\infty} \left(\frac{\phi_r}{\Phi_{i,r-1} - \Phi_{i,r}} d_{i(r-1)} - \frac{\phi_r}{\Phi_{i,r} - \Phi_{i,r+1}} d_{ir} \right) = 0, & \text{if } 2 \leqslant k < \bar{\mathrm{R}}, \\ \dfrac{\partial \mathcal{L}(\boldsymbol{\Gamma}, \mathbf{y}^e)}{\partial \log(\delta_{\bar{\mathrm{R}}})} = \delta_{\bar{\mathrm{R}}} \displaystyle\sum_{i=1}^{n} \sum_{r=\bar{\mathrm{R}}}^{\infty} \left(\frac{(r - \bar{\mathrm{R}} + 1)\phi_r}{\Phi_{i,r-1} - \Phi_{i,r}} d_{i(r-1)} - \frac{(r - \bar{\mathrm{R}} + 1)\phi_r}{\Phi_{i,r} - \Phi_{i,r+1}} d_{ir} \right) = 0, & \end{cases} \tag{30}$$

where $\phi_{i,r} = \phi\left(\tilde{\mathbf{z}}_i'\boldsymbol{\Lambda} - a_r\right)$ and $\Phi_{i,r} = \Phi\left(\tilde{\mathbf{z}}_i'\boldsymbol{\Lambda} - a_r\right)$.

As $\mathcal{L}$ is continuous, the consistency of the NPL estimator is ensured by the fact that $\operatorname{plim}\left(\dfrac{1}{n}\mathcal{L}(\boldsymbol{\Gamma}, \mathbf{y}^e)\right)$

is maximized at $\mathbf{\Gamma} = \mathbf{\Gamma}_0$ and $\mathbf{y} = \mathbf{y}_0$, where plim stands for the probability limit.

Let us focus on the limiting distribution. The Taylor expansion of $\dfrac{\partial \mathcal{L}(\mathbf{\Gamma}, \mathbf{y}^e)}{\partial \boldsymbol{\theta}}$ around $\mathbf{\Gamma}_0$ gives

$$\frac{\partial \mathcal{L}(\mathbf{\Gamma}, \mathbf{y}^e)}{\partial \mathbf{\Gamma}} = \left.\frac{\partial \mathcal{L}(\mathbf{\Gamma}, \mathbf{y}^e)}{\partial \mathbf{\Gamma}}\right|_{\mathbf{\Gamma}_0} + \left(\left.\frac{\partial^2 \mathcal{L}(\mathbf{\Gamma}, \mathbf{y}^e)}{\partial \mathbf{\Gamma} \partial \mathbf{\Gamma}'}\right|_{\mathbf{\Gamma}_0} + \left.\frac{\partial^2 \mathcal{L}(\mathbf{\Gamma}, \mathbf{y}^e)}{\partial \mathbf{\Gamma} \partial \mathbf{y}^{e\prime}}\right|_{\mathbf{\Gamma}_0} \left.\frac{\partial \mathbf{y}^e}{\partial \mathbf{\Gamma}'}\right|_{\mathbf{\Gamma}_0} \right) (\mathbf{\Gamma} - \mathbf{\Gamma}_0) + O_p(1).$$

To simplify the notations of the partial derivatives, I will use $\dfrac{\partial \mathcal{L}(\mathbf{\Gamma}_0, \mathbf{y}_0^e)}{\partial \mathbf{\Gamma}}$ to mean $\left.\dfrac{\partial \mathcal{L}(\mathbf{\Gamma}, \mathbf{y}^e)}{\partial \mathbf{\Gamma}}\right|_{\mathbf{\Gamma}_0}$ (this notation is also applied to the second partial derivatives) and $\dfrac{\partial \mathbf{y}_0^e}{\partial \mathbf{\Gamma}'}$ to mean $\left.\dfrac{\partial \mathbf{y}^e}{\partial \mathbf{\Gamma}'}\right|_{\mathbf{\Gamma}_0}$. It follows that

$$\sqrt{n}(\mathbf{\Gamma} - \mathbf{\Gamma}_0) = -\left(\frac{1}{n} \frac{\partial^2 \mathcal{L}(\mathbf{\Gamma}_0, \mathbf{y}_0^e)}{\partial \mathbf{\Gamma} \partial \mathbf{\Gamma}'} + \frac{1}{n} \frac{\partial^2 \mathcal{L}(\mathbf{\Gamma}_0, \mathbf{y}_0^e)}{\partial \mathbf{\Gamma} \partial \mathbf{y}^{e\prime}} \frac{\partial \mathbf{y}_0^e}{\partial \mathbf{\Gamma}'} \right)^{-1} \left(\frac{1}{\sqrt{n}} \frac{\partial \mathcal{L}(\mathbf{\Gamma}_0, \mathbf{y}_0^e)}{\partial \mathbf{\Gamma}} + O_p\left(\frac{1}{\sqrt{n}}\right) \right). \tag{31}$$

Let us first apply the central limit theorem to the term $\dfrac{1}{\sqrt{n}} \dfrac{\partial \mathcal{L}(\mathbf{\Gamma}_0, \mathbf{y}_0^e)}{\partial \mathbf{\Gamma}}$.

$$\frac{1}{\sqrt{n}} \frac{\partial \mathcal{L}(\mathbf{\Gamma}_0, \mathbf{y}_0^e)}{\partial \mathbf{\Gamma}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \underbrace{\begin{pmatrix} \partial \mathcal{L}_i(\mathbf{\Gamma}_0, \mathbf{y}_0^e)/\partial \mathbf{\Lambda} \\ \partial \mathcal{L}_i(\mathbf{\Gamma}_0, \mathbf{y}_0^e)/\partial \log(\delta_2) \\ \vdots \\ \partial \mathcal{L}_i(\mathbf{\Gamma}_0, \mathbf{y}_0^e)/\partial \log(\delta_{\bar{\mathrm{R}}}) \end{pmatrix}}_{\mathbf{v}_i^0} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{v}_i^0.$$

It is obvious that $\mathbb{E}\left(\mathbf{v}_i^0 | \mathbf{Z}, \mathbf{G}\right) = 0$. Thus $\mathbb{E}\left(\mathbf{v}_i^0\right) = 0$.

Let m_{ir}^0, ϕ_{ir}^0, and Φ_{ir}^0 be defined as in (30) but with $\mathbf{\Gamma} = \mathbf{\Gamma}_0$.

Let denote by $A_i = \displaystyle\sum_{r=0}^{\infty} \frac{\left(\phi_{i,r}^0 - \phi_{i,r+1}^0\right)^2}{\Phi_{i,r}^0 - \Phi_{i,r+1}^0}$,

$B_{i,k} = \delta_k \displaystyle\sum_{r=k}^{\infty} \phi_{i,r}^0 \left(\frac{\phi_{i,r-1}^0 - \phi_{i,r}^0}{\Phi_{i,r-1}^0 - \Phi_{i,r}^0} - \frac{\phi_{i,r}^0 - \phi_{i,r+1}^0}{\Phi_{i,r}^0 - \Phi_{i,r+1}^0} \right)$ if $2 \leqslant k < \bar{\mathrm{R}}$, and

$B_{i,\bar{\mathrm{R}}} = \delta_{\bar{\mathrm{R}}} \displaystyle\sum_{r=\bar{\mathrm{R}}}^{\infty} \zeta(r) \phi_{i,r} \left(\frac{\phi_{i,r-1}^0 - \phi_{i,r}^0}{\Phi_{i,r-1}^0 - \Phi_{i,r}^0} - \frac{\phi_{i,r}^0 - \phi_{i,r+1}^0}{\Phi_{i,r}^0 - \Phi_{i,r+1}^0} \right)$, where $\zeta(r) = r - \bar{\mathrm{R}} + 1$,

$C_{i,k,k'} = -\delta_k B_{i,k'}$, if $2 \leqslant k \leqslant k' < \bar{\mathrm{R}}$ and $k \neq \bar{\mathrm{R}}$,

$C_{i,\bar{\mathrm{R}},\bar{\mathrm{R}}} = \delta_{\bar{\mathrm{R}}}^2 \displaystyle\sum_{r=\bar{\mathrm{R}}}^{\infty} \left(\frac{(\zeta(r)\phi_{i,r}^0)^2 - \zeta(r)\zeta(r-1)\phi_{i,r}^0\phi_{i,r-1}^0}{\Phi_{i,r-1}^0 - \Phi_{i,r}^0} + \frac{(\zeta(r)\phi_{i,r}^0)^2 - \zeta(r)\zeta(r+1)\phi_{i,r}^0\phi_{i,r+1}^0}{\Phi_{i,r}^0 - \Phi_{i,r+1}^0} \right)$.

$$\mathbb{V}\mathbf{ar}\left(\mathbf{v}_i^0 | \mathbf{X}, \mathbf{G}\right) = \mathbb{E}\left(\mathbf{v}_i^0 \mathbf{v}_i^{0\prime} | \mathbf{X}, \mathbf{G}\right) = \underbrace{\begin{pmatrix} A_i \tilde{\mathbf{z}}_i \tilde{\mathbf{z}}_i' & B_{i,2} \tilde{\mathbf{z}}_i & \dots & B_{i,\bar{\mathrm{R}}} \tilde{\mathbf{z}}_i \\ B_{i,2} \tilde{\mathbf{z}}_i' & C_{i,2,2} & \dots & C_{i,2,\bar{\mathrm{R}}} \\ \vdots & \vdots & \ddots & \vdots \\ B_{i,\bar{\mathrm{R}}} \tilde{\mathbf{z}}_i' & C_{i,2,\bar{\mathrm{R}}} & \dots & C_{i,\bar{\mathrm{R}},\bar{\mathrm{R}}} \end{pmatrix}}_{\mathbf{\Sigma}_i} = \mathbf{\Sigma}_i. \tag{32}$$

By the law of large numbers (LLN) applied to independent and nonidentical variables (see Chow and Teicher, 2003, p. 124), assume that plim $\left(\frac{1}{n}\sum_{i}^{n}\boldsymbol{\Sigma}_i\right)$ exists and is equal to $\boldsymbol{\Sigma}_0$. It follows by the Lindeberg–Feller Central Limit Theorem (CLT) that[23]

$$\frac{1}{\sqrt{n}}\frac{\partial\mathcal{L}(\boldsymbol{\Gamma}_0,\mathbf{y}_0^e)}{\partial\boldsymbol{\Gamma}}\xrightarrow{d}\mathcal{N}\left(0,\boldsymbol{\Sigma}_0\right). \tag{33}$$

Let us now focus on plim $\left(\frac{1}{n}\frac{\partial^2\mathcal{L}(\boldsymbol{\Gamma}_0,\mathbf{y}^e)}{\partial\boldsymbol{\Gamma}\partial\boldsymbol{\Gamma}'}\right)$ and plim $\left(\frac{1}{n}\frac{\partial^2\mathcal{L}(\boldsymbol{\Gamma}_0,\mathbf{y}^e)}{\partial\boldsymbol{\Gamma}\partial\mathbf{y}^{e\prime}}\frac{\partial\mathbf{y}_0^e}{\partial\boldsymbol{\Gamma}'}\right)$.
By the LLN, plim $\left(\frac{1}{n}\frac{\partial^2\mathcal{L}(\boldsymbol{\Gamma}_0,\mathbf{y}^e)}{\partial\boldsymbol{\Gamma}\partial\boldsymbol{\Gamma}'}\right)=$ plim $\left(\frac{1}{n}\,\mathbb{E}_d\left(\frac{\partial^2\mathcal{L}(\boldsymbol{\Gamma}_0,\mathbf{y}^e)}{\partial\boldsymbol{\Gamma}\partial\boldsymbol{\Gamma}'}\right)\right)$, where $\mathbb{E}_d$ is the expectation with respect to d_{ir}'s.

$$\mathbb{E}_d\left(\frac{\partial^2\mathcal{L}(\boldsymbol{\Gamma}_0,\mathbf{y}^e)}{\partial\boldsymbol{\Gamma}\partial\boldsymbol{\Gamma}'}\right)=-\sum_{i=1}^{n}\boldsymbol{\Sigma}_i\implies\operatorname{plim}\left(\frac{1}{n}\frac{\partial^2\mathcal{L}(\boldsymbol{\Gamma}_0,\mathbf{y}^e)}{\partial\boldsymbol{\Gamma}\partial\boldsymbol{\Gamma}'}\right)=-\operatorname{plim}\left(\frac{1}{n}\sum_{i}^{n}\boldsymbol{\Sigma}_i\right)=-\boldsymbol{\Sigma}_0. \tag{34}$$

Analogously, plim $\left(\frac{1}{n}\frac{\partial^2\mathcal{L}(\boldsymbol{\Gamma}_0,\mathbf{y}^e)}{\partial\boldsymbol{\Gamma}\partial\mathbf{y}^{e\prime}}\frac{\partial\mathbf{y}_0^e}{\partial\boldsymbol{\Gamma}'}\right)=$ plim $\left(\frac{1}{n}\,\mathbb{E}_d\left(\frac{\partial^2\mathcal{L}(\boldsymbol{\Gamma}_0,\mathbf{y}^e)}{\partial\boldsymbol{\Gamma}\partial\mathbf{y}^{e\prime}}\right)\frac{\partial\mathbf{y}_0^e}{\partial\boldsymbol{\Gamma}'}\right)$.

$$\mathbb{E}_d\left(\frac{\partial^2\mathcal{L}(\boldsymbol{\Gamma}_0,\mathbf{y}^e)}{\partial\boldsymbol{\Gamma}\partial\mathbf{y}^{e\prime}}\right)=-\lambda\sum_{i=1}^{n}\begin{pmatrix}A_i\mathbf{z}_i\mathbf{g}_i\\ B_{i,2}\mathbf{g}_i\\ \vdots\\ B_{i,\bar{\mathrm{R}}}\mathbf{g}_i\end{pmatrix}\quad\text{and}\quad\frac{\partial\mathbf{y}_0^e}{\partial\boldsymbol{\Gamma}'}=\mathbf{S}^{-1}\mathbf{W}, \tag{35}$$

where $\mathbf{S}=\mathbf{I}_n-\lambda\mathbf{D}\mathbf{G}$, $\mathbf{I}_n$ is the identity matrix of dimension n, $\mathbf{D}=\operatorname{diag}\left(\sum_{r=1}^{\infty}\phi_{1,r}^0,\ \ldots,\ \sum_{r=1}^{\infty}\phi_{n,r}^0\right)$, $\mathbf{W}=(\mathbf{D}\mathbf{Z},\mathbf{Q})$, and $\mathbf{Q}$ is an $n\times(\bar{\mathrm{R}}-1)$-matrix whose i-th row is
$\boldsymbol{q}_i=\left(-\delta_2\sum_{r=2}^{\infty}\phi_{ir}^0,\ \ldots,\ -\delta_{\bar{\mathrm{R}}-1}\sum_{r=\bar{\mathrm{R}}-1}^{\infty}\phi_{ir}^0,\ -\delta_{\bar{\mathrm{R}}}\sum_{r=\bar{\mathrm{R}}}^{\infty}(r-\bar{\mathrm{R}}+1)\phi_{ir}^0\right)$.
The partial derivative $\frac{\partial\mathbf{y}_0^e}{\partial\boldsymbol{\Gamma}'}$ is computed using the implicit definition of $\mathbf{y}^e$; that is $\mathbf{y}^e=\mathbf{L}(\mathbf{y}^e,\boldsymbol{\Gamma})$.
Assuming that plim $\left(\frac{\lambda}{n}\sum_{i=1}^{n}\begin{pmatrix}A_i\mathbf{z}_i\mathbf{g}_i\mathbf{S}^{-1}\mathbf{W}\\ B_{i,2}\mathbf{g}_i\mathbf{S}^{-1}\mathbf{W}\\ \vdots\\ B_{i,\bar{\mathrm{R}}}\mathbf{g}_i\mathbf{S}^{-1}\mathbf{W}\end{pmatrix}\right)$ exists and is equal to $\boldsymbol{\Omega}_0$,

$$\operatorname{plim}\left(\frac{1}{n}\frac{\partial^2\mathcal{L}(\boldsymbol{\Gamma}_0,\mathbf{y}^e)}{\partial\boldsymbol{\Gamma}\partial\mathbf{y}^{e\prime}}\frac{\partial\mathbf{y}_0^e}{\partial\boldsymbol{\Gamma}'}\right)=-\boldsymbol{\Omega}_0. \tag{36}$$

Proposition A.1. *From Equations (31), (33), (34), and (36), the NPL estimator $\hat{\boldsymbol{\Gamma}}$ is consistent,*

[23] See Chow and Teicher (2003).

and

$$\sqrt{n}(\hat{\mathbf{\Gamma}} - \mathbf{\Gamma}_0) \xrightarrow{d} \mathcal{N}\left(0,\ (\mathbf{\Sigma}_0 + \mathbf{\Omega}_0)^{-1}\,\mathbf{\Sigma}_0\left(\mathbf{\Sigma}_0' + \mathbf{\Omega}_0'\right)^{-1}\right). \tag{37}$$

A.7 Posterior Distribution of the Dyadic Linking Model Parameters

To estimate the dyadic linking model, I used the data augmentation approach (see Albert and Chib, 1993). This approach also simulates the latent variable a^*_{ij}. Let $\mathbf{a} = (a_{ij}; i \neq j, s(i) = s(j))'$ and $\mathbf{a}^* = \left(a^*_{ij}; i \neq j, s(i) = s(j)\right)'$. The distribution of $\mathbf{a}^*$, conditional on $\mathbf{a}$, $\ddot{\mathbf{X}}$, $\bar{\boldsymbol{\beta}}$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$, and $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)'$ can be written (proportionally) as

$$\pi(\mathbf{a}^*|\mathbf{a}, \ddot{\mathbf{X}}, \bar{\boldsymbol{\beta}}, \boldsymbol{\mu}, \boldsymbol{\nu}) \propto \prod_{\substack{i\neq j\\ s(i)=s(j)}} \left\{\mathrm{I}\left(a^*_{ij} \geqslant 0\right)\mathrm{I}\left(a_{ij} = 1\right) + \mathrm{I}\left(a^*_{ij} < 0\right)\mathrm{I}\left(a_{ij} = 0\right)\right\}\frac{1}{2}\left(a^*_{ij} - \ddot{\mathbf{x}}'_{ij}\bar{\boldsymbol{\beta}} - \mu_i - \nu_j\right)^2.$$

Thus, the distribution of $a^*_{ij}|\mathbf{a}, \ddot{\mathbf{X}}, \bar{\boldsymbol{\beta}}, \boldsymbol{\mu}, \boldsymbol{\nu}$ is $\mathcal{N}(\ddot{\mathbf{x}}'_{ij}\bar{\boldsymbol{\beta}} + \mu_i + \nu_j, 1)$, truncated at the left by 0 if $a_{ij} = 1$. The distribution is $\mathcal{N}(\ddot{\mathbf{x}}'_{ij}\bar{\boldsymbol{\beta}} + \mu_i + \nu_j, 1)$, truncated at the right by 0 if $a_{ij} = 0$.

The number of observations of the network formation model is $\sum_{m=1}^{M}(n_m^2 - n_m)$, where n_m is the number of agents in the m-th group. Given the large number of observations in the network formation model, I set a flat prior distribution for $\bar{\boldsymbol{\beta}}$, $\sigma^2_{\mu,}$, σ^2_{ν}, and $\rho_{\mu,\nu}$. It follows that

$$\bar{\boldsymbol{\beta}}|\mathbf{a}, \mathbf{a}^*, \ddot{\mathbf{X}}, \boldsymbol{\mu}, \boldsymbol{\nu} \sim \mathcal{N}\left(\left(\ddot{\mathbf{X}}'\ddot{\mathbf{X}}\right)^{-1}\ddot{\mathbf{X}}'\ddot{\mathbf{a}}^*,\ \left(\ddot{\mathbf{X}}'\ddot{\mathbf{X}}\right)^{-1}\right), \tag{38}$$

where $\ddot{\mathbf{a}}^* = (a^*_{ij} - \mu_i - \nu_j : i \neq j, s(i) = s(j))'$.

For any $i \in \mathcal{V}$,

$$\mu_i|\bar{\boldsymbol{\beta}}, \mathbf{a}, \mathbf{a}^*, \ddot{\mathbf{X}}, \boldsymbol{\mu}_{-i}, \boldsymbol{\nu} \sim \mathcal{N}\left(\hat{u}_{\mu,s(i)},\ \hat{\sigma}^2_{\mu,s(i)}\right), \tag{39}$$

where $\hat{u}_{\mu,s(i)} = \hat{\sigma}^2_{\mu,s(i)} \sum_{\substack{i\neq j\\ s(i)=s(j)}} (a^*_{ij} - \ddot{\mathbf{x}}'_{ij}\bar{\boldsymbol{\beta}} - \nu_j)$, $\hat{\sigma}^2_{\mu,s(i)} = \dfrac{\sigma^2_{\mu}}{1 + \left(n_{s(i)} - 1\right)\sigma^2_{\mu}}$, and $n_{s(i)}$ is the number of agent in the group $s(i)$. Analogously,

$$\nu_i|\bar{\boldsymbol{\beta}}, \mathbf{a}, \mathbf{a}^*, \ddot{\mathbf{X}}, \boldsymbol{\mu}, \boldsymbol{\nu}_{-i} \sim \mathcal{N}\left(\hat{u}_{\nu,s(i)},\ \hat{\sigma}^2_{\nu,s(i)}\right), \tag{40}$$

where $\hat{u}_{\nu,s(i)} = \hat{\sigma}^2_{\nu,s(i)} \sum_{\substack{i\neq j\\ s(i)=s(j)}} (a^*_{ji} - \ddot{\mathbf{x}}'_{ji}\bar{\boldsymbol{\beta}} - \mu_j)$, and $\hat{\sigma}^2_{\nu,s(i)} = \dfrac{\sigma^2_{\nu}}{1 + \left(n_{s(i)} - 1\right)\sigma^2_{\nu}}$.

For the sake of identification, I normalize $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ to zero mean in each subnetwork for each step in the Gibbs sampling. The means of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ before this normalization are added to the intercept of the subnetwork for the posterior likelihood not to change.

Finally, let $\mathbf{\Sigma}_{\mu,\nu} = \begin{pmatrix} \sigma^2_{\mu} & \rho_{\mu,\nu}\sigma_{\mu}\sigma_{\nu} \\ \rho_{\mu,\nu}\sigma_{\mu}\sigma_{\nu} & \sigma^2_{\nu} \end{pmatrix}$,

$$\mathbf{\Sigma}_{\mu,\nu}|\bar{\boldsymbol{\beta}}, \mathbf{a}, \mathbf{a}^*, \ddot{\mathbf{X}}, \boldsymbol{\mu}, \boldsymbol{\nu} \sim \text{Inverse-Wishart}\left(n, \hat{\mathbf{V}}_{\mathbf{\Sigma}_{\mu,\nu}}\right), \tag{41}$$

where $\hat{\mathbf{V}}_{\mathbf{\Sigma}_{\mu,\nu}} = \sum_{i=1}^{n} \mathbf{d}_i\mathbf{d}_i'$ and $\mathbf{d}_i = (\mu_i, \nu_i)'$.

I use a Gibbs sampling to simulate the posterior distribution. Given a previous value of the parameters, one cycle of the Gibbs algorithm would sample $\mathbf{a}^*$, $\bar{\boldsymbol{\beta}}$, $\boldsymbol{\mu}$, $\boldsymbol{\nu}$, and $\mathbf{\Sigma}_{\mu,\nu}$ from their conditional distribution. The starting value of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$, $\boldsymbol{\mu}^{(0)}$ and $\boldsymbol{\nu}^{(0)}$, can be set to zero, whereas that of $\bar{\boldsymbol{\beta}}$, $\bar{\boldsymbol{\beta}}^{(0)}$ could be the standard ordinary least squares estimator at $\boldsymbol{\mu}^{(0)}$ and $\boldsymbol{\nu}^{(0)}$.

A.8 Limiting Distribution Under the Assumption of an Endogenous Network

Let $\mathbf{\Gamma}^* = \left(\mathbf{\Gamma}',\ \theta_{\mu},\ \theta_{\nu}\right)'$ and $\hat{\mathbf{\Gamma}}^*$ its NPL estimator after replacing μ_i and ν_i by their respective estimator $\hat{\mu}_i$ and $\hat{\nu}_i$ in Equations (11) and (12). Given that the estimation is done in two steps, I establish the limiting distribution of $\hat{\mathbf{\Gamma}}^*$.

The new pseudo-log-likelihood of the individual i is

$$\mathcal{L}_i(\mathbf{\Gamma}, \mathbf{y}^e) = \sum_{r=0}^{\infty} d_{ir} \log\left(\Phi\left(\tilde{\mathbf{z}}_i'\mathbf{\Lambda} + \theta_{\mu}\hat{\mu}_i + \theta_{\nu}\hat{\nu}_i - a_r\right) - \Phi\left(\tilde{\mathbf{z}}_i'\mathbf{\Lambda} + \theta_{\mu}\hat{\mu}_i + \theta_{\nu}\hat{\nu}_i - a_{r+1}\right)\right).$$

The f.o.c are

$$\begin{cases} \dfrac{\partial \mathcal{L}(\mathbf{\Gamma}, \mathbf{y}^e)}{\partial \mathbf{\Lambda}} = \displaystyle\sum_{i=1}^{n}\sum_{r=0}^{\infty} d_{ir}\frac{\hat{\phi}_{i,r} - \hat{\phi}_{i,r+1}}{\hat{\Phi}_{i,r} - \hat{\Phi}_{i,r+1}}\tilde{\mathbf{z}}_i = 0, \\ \dfrac{\partial \mathcal{L}(\mathbf{\Gamma}, \mathbf{y}^e)}{\partial \log(\delta_k)} = \delta_k \displaystyle\sum_{i=1}^{n}\sum_{r=k}^{\infty}\left(\frac{\hat{\phi}_r}{\hat{\Phi}_{i,r-1} - \hat{\Phi}_{i,r}}d_{i(r-1)} - \frac{\hat{\phi}_r}{\hat{\Phi}_{i,r} - \hat{\Phi}_{i,r+1}}d_{ir}\right) = 0, & \text{if } 2 \leqslant k < \bar{\mathrm{R}}, \\ \dfrac{\partial \mathcal{L}(\mathbf{\Gamma}, \mathbf{y}^e)}{\partial \log(\delta_{\bar{\mathrm{R}}})} = \delta_{\bar{\mathrm{R}}} \displaystyle\sum_{i=1}^{n}\sum_{r=\bar{\mathrm{R}}}^{\infty}\left(\frac{(r - \bar{\mathrm{R}} + 1)\hat{\phi}_r}{\hat{\Phi}_{i,r-1} - \hat{\Phi}_{i,r}}d_{i(r-1)} - \frac{(r - \bar{\mathrm{R}} + 1)\hat{\phi}_r}{\hat{\Phi}_{i,r} - \hat{\Phi}_{i,r+1}}d_{ir}\right) = 0, \end{cases} \tag{42}$$

where $\hat{\phi}_{i,r} = \phi\left(\tilde{\mathbf{z}}_i'\mathbf{\Lambda} + \theta_{\mu}\hat{\mu}_i + \theta_{\nu}\hat{\nu}_i - a_r\right)$ and $\Phi_{i,r} = \hat{\Phi}\left(\tilde{\mathbf{z}}_i'\mathbf{\Lambda} + \theta_{\mu}\hat{\mu}_i + \theta_{\nu}\hat{\nu}_i - a_r\right)$.

By taking the Taylor expansion of the f.o.c, I get an Equation similar to Equation (31).

$$\sqrt{n}(\mathbf{\Gamma} - \mathbf{\Gamma}_0) = -\left(\frac{1}{n}\frac{\partial^2 \mathcal{L}(\mathbf{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \mathbf{\Gamma}\partial \mathbf{\Gamma}'} + \frac{1}{n}\frac{\partial^2 \mathcal{L}(\mathbf{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \mathbf{\Gamma}\partial \mathbf{y}^{e\prime}}\frac{\partial \hat{\mathbf{y}}_0^e}{\partial \mathbf{\Gamma}'}\right)^{-1}\left(\frac{1}{\sqrt{n}}\frac{\partial \mathcal{L}(\mathbf{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \mathbf{\Gamma}} + O_p\left(\frac{1}{\sqrt{n}}\right)\right),$$

where $\hat{\mathbf{y}}_0^e$ is the expected outcome at the equilibrium with $\mathbf{\Gamma} = \mathbf{\Gamma}_0$, $\mu_i = \hat{\mu}_i$, and $\nu_i = \hat{\nu}_i$. Let $\mathbf{\Sigma}_n^* = \dfrac{1}{n}\dfrac{\partial^2 \mathcal{L}(\mathbf{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \mathbf{\Gamma}\partial \mathbf{\Gamma}'}$, $\mathbf{\Omega}_n^* = \dfrac{1}{n}\dfrac{\partial^2 \mathcal{L}(\mathbf{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \mathbf{\Gamma}\partial \mathbf{y}^{e\prime}}\dfrac{\partial \hat{\mathbf{y}}_0^e}{\partial \mathbf{\Gamma}'}$, and $\boldsymbol{\zeta}_n^* = \left(\mathbf{\Sigma}_n^* + \mathbf{\Omega}_n^*\right)^{-1}\left(\dfrac{1}{n}\mathbb{E}\left(\dfrac{\partial \mathcal{L}(\mathbf{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \mathbf{\Gamma}}\right)\right)$. Note

that $\operatorname{plim} \boldsymbol{\Sigma}_n^* = \boldsymbol{\Sigma}_0^*$ and $\operatorname{plim} \boldsymbol{\Omega}_n^* = \boldsymbol{\Omega}_0^*$, where $\boldsymbol{\Sigma}_0^*$ and $\boldsymbol{\Omega}_0^*$ are defined as $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Omega}_0$, respectively. The only difference is that the new explanatory variables μ_i and ν_i should be considered. Moreover, as the pseudo-log-likelihood is continuously derivable, and $\hat{\mu}_i$ and $\hat{\nu}_i$ are consistent estimators, then $\operatorname{plim}\left(\frac{1}{n}\mathbb{E}\left(\frac{\partial \mathcal{L}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}\right)\right) = \operatorname{plim}\left(\frac{1}{n}\mathbb{E}\left(\frac{\partial \mathcal{L}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}\right)\right) = \mathbf{0}$. Therefore, $\operatorname{plim} \boldsymbol{\zeta}_n^* = \mathbf{0}$.
It follows that

$$\sqrt{n}(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0 + \boldsymbol{\zeta}_n^*) = -\left(\boldsymbol{\Sigma}_n^* + \boldsymbol{\Omega}_n^*\right)^{-1}\left(\frac{1}{\sqrt{n}}\frac{\partial \bar{\mathcal{L}}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}} + O_p\left(\frac{1}{\sqrt{n}}\right)\right), \tag{43}$$

where $\frac{\partial \bar{\mathcal{L}}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}} = \frac{\partial \mathcal{L}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}} - \mathbb{E}\left(\frac{\partial \mathcal{L}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}\right)$.
As $\frac{\partial \bar{\mathcal{L}}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}$ is a centered variable, I can apply the CLT to $\frac{1}{\sqrt{n}}\frac{\partial \bar{\mathcal{L}}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}$. Note that even if $\frac{1}{\sqrt{n}}\frac{\partial \bar{\mathcal{L}}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}$ is a sum of dependent variables, because $(\hat{\mu}_i,\ \hat{\nu}_i)$ are dependent, the CLT can be applied because $(\hat{\mu}_i,\ \hat{\nu}_i)$ are asymptotically independent. I use CLT for ℓ-*mixing* array (see Withers, 1981). Under regular conditions,[24]

$$\frac{1}{\sqrt{n}}\frac{\partial \bar{\mathcal{L}}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}} \xrightarrow{d} \mathcal{N}\left(0, \bar{\boldsymbol{\Sigma}}_0\right). \tag{44}$$

Let $\bar{\boldsymbol{\Sigma}}_n = \mathbb{V}\mathbf{ar}\left(\frac{1}{\sqrt{n}}\frac{\partial \bar{\mathcal{L}}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}\right)$; that is $\operatorname{plim} \bar{\boldsymbol{\Sigma}}_n = \bar{\boldsymbol{\Sigma}}_0$. Let also $\hat{\mathbf{d}} = \{\hat{\mu}_1,\ \hat{\nu}_1,\ \dots,\ \hat{\mu}_n, \hat{\nu}_n\}$.

$$\begin{aligned}
\bar{\boldsymbol{\Sigma}}_n &= \frac{1}{n}\mathbb{V}\mathbf{ar}\left(\frac{\partial \mathcal{L}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}\right) = \frac{1}{n}\mathbb{E}_{\hat{\mathbf{d}}}\left(\mathbb{V}\mathbf{ar}\left(\frac{\partial \mathcal{L}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}|\hat{\mathbf{d}}\right)\right) + \frac{1}{n}\mathbb{V}\mathbf{ar}_{\hat{\mathbf{d}}}\left(\mathbb{E}\left(\frac{\partial \mathcal{L}(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}|\hat{\mathbf{d}}\right)\right),\\
\bar{\boldsymbol{\Sigma}}_n &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\hat{\mathbf{d}}}\left(\mathbb{V}\mathbf{ar}\left(\frac{\partial \mathcal{L}_i(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}|\hat{\mathbf{d}}\right)\right) + \frac{1}{n}\mathbb{V}\mathbf{ar}_{\hat{\mathbf{d}}}\left(\sum_{i=1}^n \mathbb{E}\left(\frac{\partial \mathcal{L}_i(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}|\hat{\mathbf{d}}\right)\right),\\
\operatorname{plim} \bar{\boldsymbol{\Sigma}}_n &= \boldsymbol{\Sigma}_0^* + \operatorname{plim}\frac{1}{n}\mathbb{V}\mathbf{ar}_{\hat{\mathbf{d}}}(\sum_{i=1}^n \hat{\mathbf{c}}_i),
\end{aligned}$$

where $\hat{\mathbf{c}}_i = \mathbb{E}\left(\frac{\partial \mathcal{L}_i(\boldsymbol{\Gamma}_0, \hat{\mathbf{y}}_0^e)}{\partial \boldsymbol{\Gamma}}|\hat{\mathbf{d}}\right) = \left(\hat{\mathbf{c}}_{i,\boldsymbol{\Lambda}}',\ \hat{\mathbf{c}}_{i,\delta_2},\ \dots,\ \hat{\mathbf{c}}_{i,\delta_{\bar{\mathrm{R}}}}\right)'$, with

$$\hat{\mathbf{c}}_{i,\boldsymbol{\Lambda}}' = \sum_{r=0}^{\infty}(\Phi_{i,r}^0 - \Phi_{i,r+1}^0)\frac{\hat{\phi}_{i,r}^0 - \hat{\phi}_{i,r+1}^0}{\hat{\Phi}_{i,r}^0 - \hat{\Phi}_{i,r+1}^0}(\tilde{\mathbf{z}}_i',\ \hat{\mu}_i,\ \hat{\nu}_i),$$

$$\hat{\mathbf{c}}_{i,\delta_k} = \delta_k\sum_{r=k}^{\infty}\left(\frac{\Phi_{i,r-1}^0 - \Phi_{i,r}^0}{\hat{\Phi}_{i,r-1}^0 - \hat{\Phi}_{i,r}^0} - \frac{\Phi_{i,r}^0 - \Phi_{i,r+1}^0}{\hat{\Phi}_{i,r}^0 - \hat{\Phi}_{i,r+1}^0}\right)\hat{\phi}_r^0 \text{ if } 2 \leqslant k < \bar{\mathrm{R}}, \text{ and}$$

$$\hat{\mathbf{c}}_{i,\delta_{\bar{\mathrm{R}}}} = \delta_{\bar{\mathrm{R}}}\sum_{r=\bar{\mathrm{R}}}^{\infty}\left(\frac{\Phi_{i,r-1}^0 - \Phi_{i,r}^0}{\hat{\Phi}_{i,r-1}^0 - \hat{\Phi}_{i,r}^0} - \frac{\Phi_{i,r}^0 - \Phi_{i,r+1}^0}{\hat{\Phi}_{i,r}^0 - \hat{\Phi}_{i,r+1}^0}\right)(r - \bar{\mathrm{R}} + 1)\hat{\phi}_r^0.$$

In practice, $\mathbb{V}\mathbf{ar}_{\hat{\mathbf{d}}}(\sum_{i=1}^n \hat{\mathbf{c}}_i)$ can be estimated using simulations of $\hat{\mu}_i$ and $\hat{\nu}_i$ from their posterior distribution. This term in the variance is due to the first-step estimation.

[24] See Withers (1981).

From (43) and (44), it follows that

$$\sqrt{n}(\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_0 + \boldsymbol{\zeta}_n^*) \xrightarrow{d} \mathcal{N}\left(0,\ \left(\boldsymbol{\Sigma}_0^* + \boldsymbol{\Omega}_0^*\right)^{-1} \bar{\boldsymbol{\Sigma}}_0 \left(\boldsymbol{\Sigma}_0^{*\prime} + \boldsymbol{\Omega}_0^{*\prime}\right)^{-1}\right).$$

B Supplementary Note on the Application

B.1 Marginal Effects and Corresponding Standard Errors

The parameters of the counting variable model cannot be interpreted directly. Policymakers are interested in the marginal effect of the explanatory variables on the expected outcome.

Let us recall the following notations: $\tilde{\mathbf{z}}_i' = (\mathbf{g}_i\mathbf{y}^e,\ \mathbf{z}_i')$, $\boldsymbol{\Lambda} = (\lambda,\ \boldsymbol{\theta}')'$, and $\boldsymbol{\Gamma} = \left(\boldsymbol{\Lambda}',\ \log\left(\boldsymbol{\delta}'\right)\right)'$, $a_0 = -\infty$, $a_r = \sum_{k=1}^{r} \delta_k$ if $1 \leqslant r < \bar{\mathrm{R}}$, $a_r = (r - \bar{\mathrm{R}})\delta_{\bar{\mathrm{R}}} + \sum_{k=1}^{\bar{\mathrm{R}}} \delta_k$ if $r \geqslant \bar{\mathrm{R}}$, and $\delta_1 = 0$. For any $k = 1, \dots, \dim(\boldsymbol{\Lambda})$, let λ_k and $\tilde{z}_{ik}$ be the k-th component in $\boldsymbol{\Lambda}$ and $\tilde{\mathbf{z}}_i$, respectively. The marginal effect of the explanatory variable $\tilde{z}_{ik}$ on $\bar{y}_i$ is given by

$$\delta_{ik}(\boldsymbol{\Gamma}) = \frac{\partial \bar{y}_i}{\partial \tilde{z}_{ik}} = \lambda_k \sum_{r=1}^{\infty} \phi\left(\tilde{\mathbf{z}}_i'\boldsymbol{\Lambda} - a_r\right). \tag{45}$$

The standard error of $\delta_{ik}(\boldsymbol{\Gamma})$ can be computed using the Delta method.

The Taylor expansion of Equation (45) around $\boldsymbol{\Gamma}_0$ is

$$\delta_{ik}(\hat{\boldsymbol{\Gamma}}) = \delta_{ik}(\boldsymbol{\Gamma}_0) + \frac{\partial \delta_{ik}(\boldsymbol{\Gamma}_0)}{\partial \boldsymbol{\Gamma}'}(\hat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}_0) + O_p(\hat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}_0),$$

where $\frac{\partial \delta_{ik}(\boldsymbol{\Gamma}_0)}{\partial \boldsymbol{\Gamma}'}$ stands for the derivative of $\delta_{ik}(\boldsymbol{\Gamma})$ with respect to $\boldsymbol{\Gamma}$ applied to $\boldsymbol{\Gamma}_0$.

When n is sufficiently large,

$$\delta_{ik}(\hat{\boldsymbol{\Gamma}}) \approx \delta_{ik}(\boldsymbol{\Gamma}_0) + \frac{\partial \delta_{ik}(\boldsymbol{\Gamma}_0)}{\partial \boldsymbol{\Gamma}'}(\hat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}_0). \tag{46}$$

It follows that an estimator of the standard error of $\delta_{ik}(\hat{\boldsymbol{\Gamma}})$ is

$$Se\left(\delta_{ik}(\hat{\boldsymbol{\Gamma}})\right) = \sqrt{\frac{\partial \delta_{ik}(\hat{\boldsymbol{\Gamma}})}{\partial \boldsymbol{\Gamma}'} \widehat{AsyVar}\left(\hat{\boldsymbol{\Gamma}}\right) \frac{\partial \delta_{ik}(\hat{\boldsymbol{\Gamma}})}{\partial \boldsymbol{\Gamma}}}, \tag{47}$$

where

$$\frac{\partial \delta_{ik}(\hat{\boldsymbol{\Gamma}})}{\partial \boldsymbol{\Gamma}'} = \left(\frac{\partial \delta_{ik}(\hat{\boldsymbol{\Gamma}})}{\partial \boldsymbol{\Lambda}'},\ \frac{\partial \delta_{ik}(\hat{\boldsymbol{\Gamma}})}{\partial \log(\delta_2)},\ \dots,\ \frac{\partial \delta_{ik}(\hat{\boldsymbol{\Gamma}})}{\partial \log(\delta_{\bar{\mathrm{R}}-1})},\ \frac{\partial \delta_{ik}(\hat{\boldsymbol{\Gamma}})}{\partial \log(\delta_{\bar{\mathrm{R}}})}\right), \tag{48}$$

$$\frac{\partial \delta_{ik}(\hat{\boldsymbol{\Gamma}})}{\partial \boldsymbol{\Lambda}'} = \mathbf{e}_k \sum_{r=1}^{\infty} \phi\left(\tilde{\mathbf{z}}_i'\hat{\boldsymbol{\Lambda}} - a_r\right) - \lambda_k \tilde{\mathbf{z}}_i' \sum_{r=1}^{\infty}\left(\tilde{\mathbf{z}}_i'\hat{\boldsymbol{\Lambda}} - a_r\right)\phi\left(\tilde{\mathbf{z}}_i'\hat{\boldsymbol{\Lambda}} - a_r\right), \tag{49}$$

$$\frac{\partial \delta_{ik}(\hat{\boldsymbol{\Gamma}})}{\partial \log(\delta_l)} = \delta_l \lambda_k \sum_{r=l}^{\infty}\left(\tilde{\mathbf{z}}_i'\hat{\boldsymbol{\Lambda}} - a_r\right)\phi\left(\tilde{\mathbf{z}}_i'\hat{\boldsymbol{\Lambda}} - a_r\right), \quad \text{for } 2 \leqslant l < \bar{\mathrm{R}}, \tag{50}$$

$$\frac{\partial \delta_{ik}(\hat{\boldsymbol{\Gamma}})}{\partial \log(\delta_{\bar{\mathrm{R}}})} = \delta_{\bar{\mathrm{R}}} \lambda_k \sum_{r=\bar{\mathrm{R}}}^{\infty}(r - \bar{\mathrm{R}} + 1)\left(\tilde{\mathbf{z}}_i'\hat{\boldsymbol{\Lambda}} - a_r\right)\phi\left(\tilde{\mathbf{z}}_i'\hat{\boldsymbol{\Lambda}} - a_r\right), \tag{51}$$

where $\mathbf{e}_k$ is a row vector of dimension $\dim(\boldsymbol{\Lambda})$ with the k-th term equal to one and the other terms equal to zero.

As in any nonlinear model, the marginal effect depends on $\mathbf{z}_i$. I then report their average, $\frac{1}{n}\sum_{i=1}^{n}\delta_{ik}(\hat{\boldsymbol{\theta}})$, where

$$Se\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{ik}(\hat{\boldsymbol{\theta}})\right) = \sqrt{Q_{\boldsymbol{\theta}}\ \widehat{AsyVar}\ Q_{\boldsymbol{\theta}}'}, \quad \text{and} \quad Q_{\boldsymbol{\theta}} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \delta_{ik}(\hat{\boldsymbol{\Gamma}})}{\partial \boldsymbol{\Gamma}'}\right).$$

B.2 Data summary

This section summarizes the data (see Table 6). The categorical explanatory variables are discretized into several binary categorical variables. For the categorical explanatory variables, the level in italics is set as the reference level in the econometric models.

Table 6: Data Summary

Variable	Mean	Sd.	Min	1st Qu.	Median	3rd Qu.	Max
Age	15.010	1.709	10	14	15	16	19
Sex							
Female	0.503	0.500	0	0	1	1	1
Male	0.497	0.500	0	0	0	1	1
Hispanic	0.168	0.374	0	0	0	0	1
Race							
White	0.625	0.484	0	0	1	1	1
Black	0.185	0.388	0	0	0	0	1
Asian	0.071	0.256	0	0	0	0	1
Other	0.097	0.296	0	0	0	0	1
Years at school	2.490	1.413	1	1	2	3	6
With both parents	0.727	0.445	0	0	1	1	1
Mother Educ.							
High	0.175	0.380	0	0	0	0	1
<High	0.302	0.459	0	0	0	1	1
>High	0.406	0.491	0	0	0	1	1
Missing	0.117	0.322	0	0	0	0	1
Mother job							
Stay at home	0.204	0.403	0	0	0	0	1
Professional	0.199	0.400	0	0	0	0	1
Other	0.425	0.494	0	0	0	1	1
Missing	0.172	0.377	0	0	0	0	1
Number of activities	2.353	2.406	0	1	2	3	33

The dependent variable is the number of extracurricular activities in which students are enrolled. It varies from 0 to 33. However, most students declare that they participate in fewer than 10 extracurricular activities (see Figure 3).

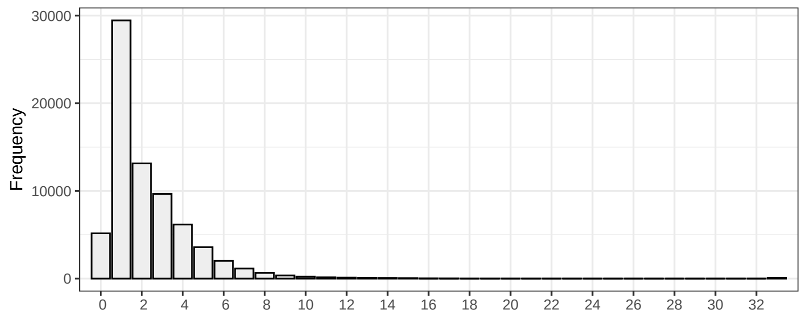

Figure 3: Distribution of the Number of Extracurricular Activities

B.3 Posterior Distribution of the Dyadic Linking Model Parameters

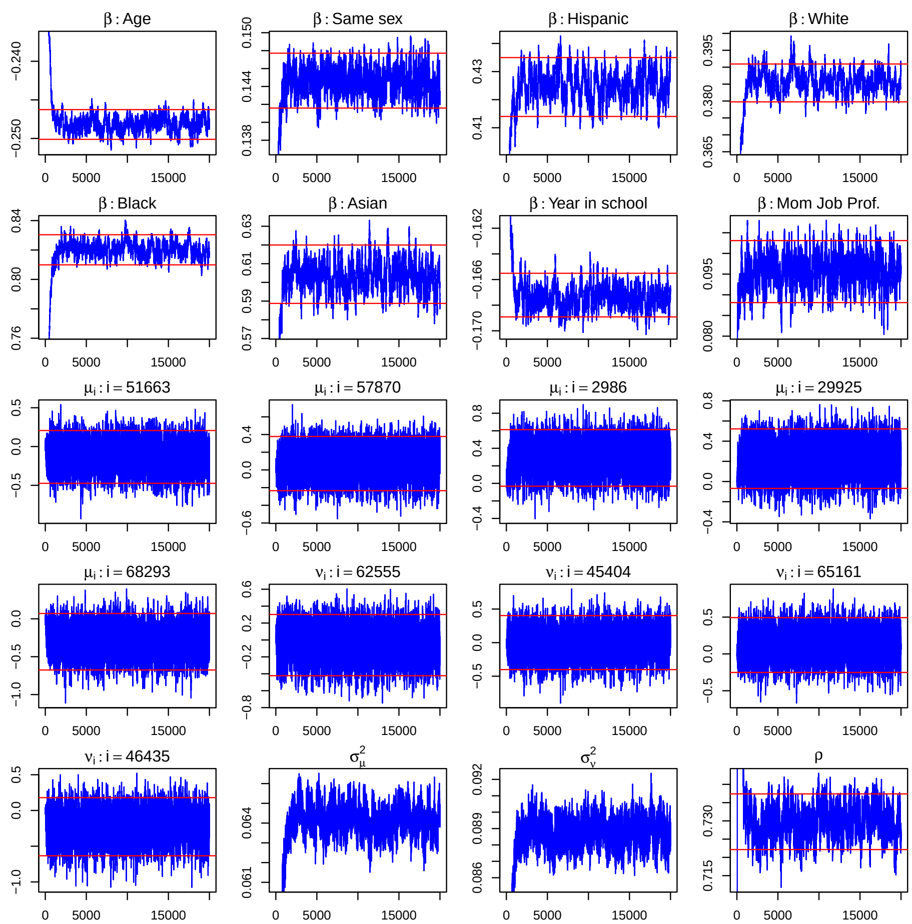

Figure 4: Posterior Distribution of the Network Formation Model Parameters

This figure presents the posterior distribution of the coefficients of the observed dyad-specific variables and some other parameters chosen at random. Students of similar age, Hispanic, Black, and Asian students, and students who have spent a similar number of years at their current school are likely to form links. In contrast, students of the same sex and White students are not likely to form links. Unobserved factors that increase the probability to give links are positively correlated with those that increase the probability to receive links

References

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

Akerlof, G. A. (1997): "Social distance and social decisions," *Econometrica: Journal of the Econometric Society*, 1005–1027.

Albert, J. H. and S. Chib (1993): "Bayesian analysis of binary and polychotomous response data," *Journal of the American statistical Association*, 88, 669–679.

Amemiya, T. (1981): "Qualitative response models: A survey," *Journal of economic literature*, 19, 1483–1536.

Baetschmann, G., K. E. Staub, and R. Winkelmann (2015): "Consistent estimation of the fixed effects ordered logit model," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 685–703.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov (2010): "Estimating static models of strategic interactions," *Journal of Business & Economic Statistics*, 28, 469–482.

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

Bellman, R. (2013): *A brief introduction to theta functions*, Courier Corporation.

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

Boucher, V. (2016): "Conformism and self-selection in social networks," *Journal of Public Economics*, 136, 30–44.

Boucher, V. and Y. Bramoullé (2020): "Binary Outcomes and Linear Interactions," Tech. rep., Aix-Marseille School of Economics, France.

Boucher, V. and A. Houndetoungan (2020): *Estimating peer effects using partial network data*, Centre de recherche sur les risques les enjeux économiques et les politiques.

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

——— (2020): "Peer effects in networks: A survey," *Annual Review of Economics*, 12, 603–629.

Brock, W. A. and S. N. Durlauf (2001): "Discrete choice with social interactions," *The Review of Economic Studies*, 68, 235–260.

——— (2002): "A multinomial-choice model of neighborhood effects," *American Economic Review*, 92, 298–303.

——— (2007): "Identification of binary choice models with social interactions," *Journal of Econometrics*, 140, 52–75.

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

Chow, Y. S. and H. Teicher (2003): *Probability theory: independence, interchangeability, martingales*, Springer Science & Business Media.

Darling, N. (2005): "Participation in extracurricular activities and adolescent adjustment: Cross-sectional and longitudinal findings," *Journal of youth and adolescence*, 34, 493–505.

De Paula, A. (2013): "Econometric analysis of games with multiple equilibria," *Annu. Rev. Econ.*, 5, 107–131.

——— (2017): "Econometrics of network models," in *Advances in economics and econometrics: Theory and applications, eleventh world congress*, Cambridge University Press Cambridge, 268–323.

Dzemski, A. (2019): "An empirical model of dyadic link formation in a network with unobserved heterogeneity," *Review of Economics and Statistics*, 101, 763–776.

Erbe, W. (1962): "Gregariousness, group membership, and the flow of information," *American Journal of Sociology*, 67, 502–516.

Fan, J. and R. Li (2001): "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American statistical Association*, 96, 1348–1360.

Fan, J. and H. Peng (2004): "Nonconcave penalized likelihood with a diverging number of parameters," *The annals of statistics*, 32, 928–961.

Fortin, B. and M. Yazbeck (2015): "Peer effects, fast food consumption and adolescent weight gain," *Journal of health economics*, 42, 125–138.

Fujimoto, K. and T. W. Valente (2013): "Alcohol peer influence of participating in organized school activities: a network approach." *Health Psychology*, 32, 1084.

Glaser, S. (2017): "A review of spatial econometric models for count data," Tech. rep., Hohenheim Discussion Papers in Business, Economics and Social Sciences.

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

Graham, B. S. (2017): "An econometric model of network formation with degree heterogeneity," *Econometrica*, 85, 1033–1063.

Guerra, J.-A. and M. Mohnen (2020): "Multinomial choice with social interactions: occupations in Victorian London," *Review of Economics and Statistics*, 1–44.

Harsanyi, J. C. (1967): "Games with incomplete information played by "Bayesian" players, I–III Part I. The basic model," *Management science*, 14, 159–182.

Heckman, J. J. (1979): "Sample selection bias as a specification error," *Econometrica: Journal of the econometric society*, 153–161.

Holland, A. and T. Andre (1987): "Participation in extracurricular activities in secondary school: What is known, what needs to be known?" *Review of educational research*, 57, 437–466.

Hsieh, C.-S. and L. F. Lee (2016): "A social interactions model with endogenous friendship formation and selectivity," *Journal of Applied Econometrics*, 31, 301–319.

Hsieh, C.-S., L.-F. Lee, and V. Boucher (2020): "Specification and estimation of network formation and network interaction models with the exponential probability distribution," *Quantitative economics*, 11, 1349–1390.

Inouye, D. I., E. Yang, G. I. Allen, and P. Ravikumar (2017): "A review of multivariate distributions for count data derived from the Poisson distribution," *Wiley Interdisciplinary Reviews: Computational Statistics*, 9, e1398.

Johnsson, I. and H. R. Moon (2015): "Estimation of peer effects in endogenous social networks: control function approach," *Review of Economics and Statistics*, 1–51.

Karlis, D. (2003): "An EM algorithm for multivariate Poisson distribution and related models," *Journal of Applied Statistics*, 30, 63–77.

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

Kelejian, H. H. and I. R. Prucha (1998): "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances," *The Journal of Real Estate Finance and Economics*, 17, 99–121.

Lancaster, T. (2000): "The incidental parameter problem since 1948," *Journal of econometrics*, 95, 391–413.

Lee, C. G., J. Kwon, H. Sung, I. Oh, O. Kim, J. Kang, and J.-W. Park (2020): "The effect of physically or non-physically forced sexual assault on trajectories of sport participation from adolescence through young adulthood," *Archives of Public Health*, 78, 1–10.

Lee, L.-F. (2004): "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models," *Econometrica*, 72, 1899–1925.

——— (2007): "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics*, 140, 333–374.

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

Liesenfeld, R., J.-F. Richard, and J. Vogler (2016): "Likelihood Evaluation of High-Dimensional Spatial Latent Gaussian Models with Non-Gaussian Response Variables', Spatial Econometrics: Qualitative and Limited Dependent Variables (Advances in Econometrics, Volume 37)," .

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

Liu, X. (2019): "Simultaneous equations with binary outcomes and social interactions," *Econometric Reviews*, 38, 921–937.

Liu, X., E. Patacchini, and Y. Zenou (2014): "Endogenous peer effects: local aggregate or local average?" *Journal of Economic Behavior & Organization*, 103, 39–59.

Liu, X., E. Patacchini, Y. Zenou, and L.-F. Lee (2012): "Criminal networks: Who is the key player?" *Unpublished manuscript, NOTA DI. LAVORO. [39.2012]*.

Liu, X. and J. Zhou (2017): "A social interaction model with ordered choices," *Economics Letters*, 161, 86–89.

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

McFadden, D. (1973): "Conditional logit analysis of qualitative choice behavior," *Frontiers in Econometrics*, 105–142.

McNeal Jr, R. B. (1999): "Participation in high school extracurricular activities: Investigating school effects," *Social Science Quarterly*, 291–309.

Moffitt, R. A. et al. (2001): "Policy interventions, low-level equilibria, and social interactions," *Social dynamics*, 4, 6–17.

Murota, K. (1998): "Discrete convex analysis," *Mathematical Programming*, 83, 313–371.

Newton, N. J., J. Pladevall-Guyer, R. Gonzalez, and J. Smith (2018): "Activity engagement and activity-related experiences: The role of personality," *The Journals of Gerontology: Series B*, 73, 1480–1490.

Osborne, M. J. and A. Rubinstein (1994): *A course in game theory*, MIT press.

Patacchini, E. and Y. Zenou (2012): "Juvenile delinquency and conformism," *The Journal of Law, Economics, & Organization*, 28, 1–31.

Pfeiffer, F. and N. J. Schulz (2012): "Gregariousness, interactive jobs and wages," *Journal for Labour Market Research*, 45, 147–159.

Smart, D. R. (1980): *Fixed point theorems*, vol. 66, CUP Archive.

Soetevent, A. R. and P. Kooreman (2007): "A discrete-choice model with social interactions: with an application to high school teen behavior," *Journal of Applied Econometrics*, 22, 599–624.

Withers, C. S. (1981): "Central limit theorems for dependent variables. I," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57, 509–534.

Xu, X. and L.-f. Lee (2015a): "Estimation of a binary choice game model with network links," *Submitted to Quantitative Economics.*

——— (2015b): "Maximum likelihood estimation of a spatial autoregressive Tobit model," *Journal of Econometrics*, 188, 264–280.

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