Count Data Models with Social Interactions under Rational Expectations

Elysée Aristide Houndetoungan*

Thema, Cy Cergy Paris Université

October 24, 2022

Abstract

I propose a peer effect model for counting variables using a game of incomplete information. I present sufficient conditions under which the game equilibrium is unique and study the model parameter identification. My identification result is general and can be applied to other nonlinear models with social interactions, such as binary and ordered response models. I show that the parameters can be estimated using the Nested Partial Likelihood (NLP). I generalize the estimator to the case of endogenous networks and study its asymptotic properties. I show that the linear-in-means/Tobit models with a counting outcome are particular cases of my model. However, these models ignore the counting nature of the dependent variable and lead to inconsistent estimators. I use the model to evaluate peer effects on students' participation in extracurricular activities. 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 0.08. The estimate of this effect with the Tobit model is three times higher.

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 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`.

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 assumption of a linear-quadratic payoff is made because it leads to a linear reduced form that is simple to estimate. I show that this assumption 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 design matrix. In the case of rational expectation models, the design matrix depends on the expected outcome, which is an unobserved variable. Therefore, the rank condition cannot be verified empirically (e.g., Lee, 2004; Yang and Lee, 2017). I present an identification analysis that leads to verifiable conditions. In particular, I show that the parameters are identified under similar conditions to those imposed by Bramoullé et al. (2009) for linear-in-means models. This analysis can be extended to other nonlinear models with social interactions, such as binary and ordered response models, for which the identification of parameters has so far been based on conditions that are difficult to verify in practice.

I show that the parameters can be estimated using the NPL algorithm proposed by Aguirregabiria and Mira (2007). I assess the finite sample performance of the estimation strategy using Monte Carlo simulations. I generalize the estimation strategy to the case where the network is endogenous. Endogeneity is due to unobservable individual attributes, which influence both link formation and the outcome (see Johnsson and Moon, 2021; Graham, 2017). To control for the endogeneity, I use a two-stage estimation strategy. In the first stage, I estimate a dyadic linking model in which the probability of link formation between two agents depends on, among other things, the unobservable attributes. In the second stage, I control for the unobserved attributes in the counting variable model. Under regularity conditions, I establish the asymptotic normality of the estimator at the second stage.

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. 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 0.08. As in the Monte Carlo study, I also find that the SART model overestimates these marginal peer effects. The bias depends on the tail of the outcome. In my empirical study, the estimate of the marginal peer effects with the Tobit model is three times higher than that of the proposed model.

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). My model bridges the gap between the binary/ordered response models and the linear-in-means models by generalizing both models. Indeed, 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 the binary model studied by Lee et al. (2014). On the other hand, I show that the linear-in-means/Tobit models are particular cases of the proposed model. Moreover, I present a general identification condition that can be verified in practice.
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 jointly simulate the posterior distribution of the network formation model parameters and the outcome model parameters. Although this method could be more efficient as the estimation is done in a single step, it can be cumbersome to implement with a discrete data model. In addition, their model treats the unobservable individual attributes that influence the outcome and the network as random effects. In the current paper, I assume that the attributes are fixed effects. Furthermore, my approach 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. My estimation strategy is similar to the control function approach proposed by Johnsson and Moon (2021).

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), which 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. Agent's 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 subgroup of individual i (observable by all agents and the econometrician). Agents interact through a directed network. Let $\mathbf{G} = [\mathbf{g}_{ij}]$ be an $n \times n$ network matrix (observable by all agents and the econometrician), where $\mathbf{g}_{ij}$ is non-negative and captures the proximity of the individuals i and j in the network. The element $\mathbf{g}_{ij}$ may depend on n. I do not use the notation $\mathbf{g}_{n,ij}$ for simplicity, and this does not create confusion. Throughout the paper, my notations follow this simplicity. The subscript i is used instead of n, $s(i)$, i. Interactions are restricted to individuals from the same subgroup.[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 about the choice of y_i are described by an additive discrete payoff function defined by

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

[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 an 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.

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 vector's (resp. matrix's) subscript $-i$ is used to denote the vector (resp. matrix) excluding the i-th row. 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 assume that $c(y_i)$ is finite for any $y_i \in \mathbb{N}$.[7] 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 below, a quadratic cost function implies a strong restriction on the econometric model.

The term $\dfrac{\lambda}{2}(y_i - \bar{y}_i)^2$ is a social cost that increases with the gap between the agent and peers' choices. The parameter λ, called *peer effects*, captures the influence of this gap on the payoff.[8] Such a specification of the social cost implies conformist preferences (see Akerlof, 1997). This social cost is different from the social subpayoff assumed by Blume et al. (2015), which implies complementary preferences. However, as Boucher and Fortin (2016) show using a linear-in-means model, both subpayoffs lead to similar reduced forms that involve the same econometric issues.

Agent's type is described by $(e_i(r))_{r \in \mathbb{N}}$, 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 the others' type, and thus, they do not observe the others' choice $\mathbf{y}_{-i}$. Therefore, the agent maximizes the expectation of the random payoff (1), where the expectation is taken with respect to their beliefs over $\mathbf{y}_{-i}$ conditionally on their observable characteristics ψ_i, others' observable characteristics $\boldsymbol{\psi}_{-i}$, and the network matrix components $\mathbf{g}_i$ and $\mathbf{G}_{-i}$, where $\mathbf{g}_i$ is the i-th row of of $\mathbf{G}$. As common in Bayesian game literature, I assume that, conditionally on $\mathcal{I}_i = \{\psi_i,\ \boldsymbol{\psi}_{-i},\ \mathbf{g}_i,\ \mathbf{G}_{-i}\}$, the private information $e_i(r)$ is distributed identically 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

$$U_i^e(y_i) = \psi_i y_i - c(y_i) - \frac{\lambda}{2}\mathbb{E}_{\bar{y}_i|\mathcal{I}_i}\left[(y_i - \bar{y}_i)^2\right] + e_i(y_i). \tag{2}$$

I set, by convention, that $c(-1) = +\infty$, which implies $U_i^e(-1) = -\infty$. This will be helpful to simplify many equations. As the space of y_i is $\mathbb{N}$, the expectation $\mathbb{E}_{\bar{y}_i|\mathcal{I}_i}\left[(y_i - \bar{y}_i)^2\right]$ involves an infinite sum-

[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] The strategy y_i that maximizes the payoff would be bounded if $c(y_i)$ is infinite from a large y_i. I exclude this particular case from my analysis because it leads to a bounded outcome that can be analyzed using a binary or an ordered response model (see Lee et al., 2014; Liu and Zhou, 2017). The current framework is similar to that of the linear model in that y_i is unbounded, but the space of its expectation hinges on whether ψ_i is bounded or not.

[8] The case where λ is negative is excluded from this analysis because it requires stronger assumptions for the payoff function to be concave. Moreover, this case is less relevant in the literature since agents tend to conform to their peers.

mation that may not be finite. Under Assumptions 2.1–2.3 stated below, all infinite summations used in the paper are finite (see Online Appendix (OA) S.1).

Let Δ be the first difference operator; i.e., for any sequence $(b_r)_r$, $\Delta b_r = b_r - b_{r-1}$. 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.

Assumption 2.1. *$c(.)$ is a strictly convex and strictly increasing function on $\mathbb{N}$.*

Assumption 2.2. *For all $i \in \mathcal{V}$, $r \in \mathbb{N}^*$, $e_i(r) = e_i(r-1) + \varepsilon_i$, where $\varepsilon_i|\mathcal{I}_i$'s are independent and identically follow a continuous symmetric distribution with a cumulative distribution function (cdf) $F_{\varepsilon|\mathcal{I}}$ and a probability density function (pdf) $f_{\varepsilon|\mathcal{I}}$.*

Assumption 2.3. *$\lim\limits_{r\to\infty} r^{-\rho}(\Delta c(r+1) - \Delta c(r)) > 0$ and $f_{\varepsilon|\mathcal{I}}(x) = o\left(|x|^{-\kappa}\right)$ at ∞, where $\rho \geqslant 0$ and $(1+\rho)(\kappa - 1) > 2$.*

The strict convexity condition set for the cost function in Assumption 2.1 means a strictly increasing difference in the cost: $\Delta c(r+1) - \Delta c(r) > 0$, $\forall\ r \in \mathbb{N}$. This implies a strictly concave expected payoff under the conditions set in Assumption 2.2. This feature plays an important role in Proposition 2.1. It suggests that the expected payoff has a global maximum that is reached at a single point almost surely (a.s.). As discussed in Section 3.1, the assumption of a strictly 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).

Assumption 2.2 characterizes 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, the restriction is set for the distribution of $\Delta e_i(r) := e_i(r) - e_i(r-1)$ for any $r \in \mathbb{N}^*$. First, Assumption 2.2 sets that 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, Assumption 2.2 sets that ε_i's are independent and identically follow a continuous symmetric distribution. This is a classic restriction in the literature on social interactions that simplifies many equations.[9] A similar restriction is also set for binary response models and is expressed as $e_i(1) - e_i(0) = \varepsilon_i$, where ε_i's are independent and identically distributed according to a logistic distribution (e.g., Brock and Durlauf, 2001; Li and Lee, 2009; Lee et al., 2014; Lin and Xu, 2017).

Assumption 2.3 imposes the minimum rate at which the cost increases when y_i is sufficiently high. This condition is not too restrictive since ρ may take any nonnegative value. If $\rho = 0$, then the first

[9]I release this assumption later to control for network endogeneity.

condition of Assumption 2.3 is $\lim_{r\to\infty}(\Delta c(r+1)-\Delta c(r))>0$, which is slightly more constraining than the strictly increasing difference assumption set in Assumption 2.1. In addition, Assumption 2.3 sets the rate at which the tail of the density function $f_{\varepsilon|\mathcal{I}}$ must decay. The condition $(1+\rho)(\kappa-1)>2$ is a trade-off condition between ρ and κ. It ensures that the probability that y_i takes the value r converges to zero at some rate as r grows to infinity. This condition is necessary so that the infinite summations defined in the paper (e.g., the expected choice y_i^e) be finite (see OA S.1). As κ is necessarily greater than one for any density function having a limit at infinity,[10] any continuous distribution (whose density function has a limit at infinity) could suit Assumption 2.3. However, this would require setting ρ greater than $2/(\kappa-1)-1$ to meet the trade-off condition. For the usual distributions whose tail decays exponentially, such as the normal or logistic distribution, ρ can take any non-negative value and the trade-off condition would still hold. Assumptions 2.1–2.3 imply that there is a unique count choice that maximizes the payoff a.s.

Proposition 2.1. *Under Assumptions 2.1–2.3, (i) $U_i^e(.)$ has a unique maximizer, $r_0\in\mathbb{N}$, almost surely; (ii) $U_i^e(r)\geqslant\max\{U_i^e(r-1),\ U_i^e(r+1)\}$ iff $r=r_0$.*

Under Assumptions 2.1–2.3, the expected payoff has a global maximum. Since $U_i^e(.)$ is a discrete function, the global maximum could be reached at two different points. However, this has zero probability of occurrence because the distribution of ε_i is continuous (see Appendix A.1). Furthermore, Proposition 2.1 states that $y_i=r$ iff $\Delta U_i^e(r+1)\leqslant 0\leqslant\Delta U_i^e(r)$, which is equivalent 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 $a_r=\Delta c(r)+\lambda r-\frac{\lambda}{2}$, $\bar{y}_i^e=\sum_{j\in\mathcal{V}_i}g_{ij}y_j^e$, and y_i^e is the *rational* expected (*true* expectation of the) choice given $\mathcal{I}_i=\{\psi_i,\ \boldsymbol{\psi}_{-i},\ \mathbf{g}_i,\ \mathbf{G}_{-i}\}$. This characterization is useful as it allows to write the probability that $y_i=r$ given $\mathcal{I}_i$. Let $p_{ir}=\mathbb{P}(y_i=r|\mathcal{I}_i)$ be this probability. Using the symmetry of the distribution of ε_i, p_{ir} can be written as

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

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 microfoundations 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 (see Liu, 2019). Moreover, unlike a classical ordered model, y_i is unbounded, and there is then an infinite number of cut points a_r, $r\in\mathbb{N}$.

The model can also be compared to an ordered response model by interpreting $y_i^*=\lambda\bar{y}_i^e+\psi_i+\varepsilon_i$ as

[10] Because $\lim_{|x|\to\infty}f_{\varepsilon|\mathcal{I}}(x)/|x|^{-\kappa}=0$, for some $\kappa>1$. Otherwise, $\lim_{|x|\to\infty}f_{\varepsilon|\mathcal{I}}(x)/|x|^{-1}>0$, and there would exist $h_0,\ x_0>0$ such that, for any $x\geqslant x_0$, $f_{\varepsilon|\mathcal{I}}(x)>h_0x^{-1}$. This would imply that $F_{\varepsilon|\mathcal{I}}(x)\geqslant h_0\int_{x_0}^x x^{-1}dx$ and $F_{\varepsilon|\mathcal{I}}(x)$ would diverge at infinity.

an idiosyncratic marginal utility and a_r as a common marginal cost. Therefore, $y_i = r$ if and only if $y_i^* \geqslant a_r$, $y_i^* \leqslant a_{r+1}$, and $U_i^e(r) > U_i^e(r+1)$. The condition $y_i^* \geqslant a_r$ means that the agent can choose a value greater than or equal to r by increasing their utility greater than the increase in their cost. In contrast, the condition $y_i^* \leqslant a_{r+1}$ implies that the optimal choice is less or equal to $r+1$. Both conditions means that the optimal choice is either r or $r+1$. The third condition $U_i^e(r) > U_i^e(r+1)$ guarantees that r is preferred to $r+1$. These three conditions also lead to Equation (3).

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 to 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, $p_{i0} = 1 - F_{\varepsilon|\mathcal{I}}\left(\lambda \bar{y}_i^e + \psi_i - a_1\right)$ and $p_{i1} = F_{\varepsilon|\mathcal{I}}\left(\lambda \bar{y}_i^e + \psi_i - a_1\right)$. The variable y_i only takes the values 0 and 1, almost surely, and y_i^e is equal to p_{i1}. The condition $p_{i1} = F_{\varepsilon|\mathcal{I}}\left(\lambda \sum_{j=1}^n g_{ij} p_{j1} + \psi_i - a_1\right)$ is the characterization of the rational beliefs in the case of the binary choice.

I now explain the reason why the quadratic cost function is too restrictive econometrically. 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 may not be verified 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 estimates. 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 a specification more flexible than that of the SAR/SART models.

2.2 Bayesian Nash Equilibrium

Proposition 2.1 states that there is a unique count choice $\mathbf{y} = (y_1, \dots, y_n)'$ that maximizes the expected payoff (2) given $\mathbf{y}^e = (y_1^e, \dots, y_n^e)'$. However, there may exist more than one expected outcome and belief system $\mathbf{p} = (p_{ir})$ that verify the RE condition (3). In the case of the binary or multinational response models (e.g., Brock and Durlauf, 2001; Lee et al., 2014; Guerra and Mohnen, 2020), Brouwer's fixed point theorem applied to Equation (3) guarantees the existence of $\mathbf{p}$. In addition, sufficient conditions are set to show the uniqueness using the contraction mapping theorem. This approach cannot be used in the current framework because $\mathbf{p}$ is an infinite dimensional vector. Indeed, as

$y_i^e = \sum_{r=1}^{\infty} rp_{ir}$, Equation (3) can also be expressed as $\mathbf{p} = \mathbb{H}(\mathbf{p})$, where $\mathbb{H}$ is some mapping defined from $[0,1]^{\infty}$ to itself. Neither Brouwer's fixed point theorem nor Schauder's fixed point theorem[11] can be used since $[0,1]^{\infty}$ is not a compact set.[12] The main condition for the application of these fixed point theorems is that the mapping is defined on a compact set. Moreover, it would be burdensome to apply the contraction mapping theorem to $\mathbb{H}$.

To circumvent this issue, let us consider Equation (3), which 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 $\ell_i(\mathbf{y}^e) = \sum_{r=1}^{\infty} F_{\varepsilon|\mathcal{I}}(\lambda \bar{y}_i^e + \psi_i - a_r)$ 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 suggests that the game can be solved in a finite-dimensional space. The uniqueness of $\mathbf{y}^e$ can be directly established if $\mathbf{L}$ is a contracting mapping. I then make the following assumption.

Assumption 2.4. *$\lambda < B_c/\lVert\mathbf{G}\rVert_{\infty}$, where $B_c = \left(\max_{u\in\mathbb{R}} \sum_{r=1}^{\infty} f_{\varepsilon|\mathcal{I}}(u - a_r)\right)^{-1}$.*

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 2.4 defines this threshold for the case of my model.[13] Assumption 2.4 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 2.4 implies that $\lambda < (\lVert\mathbf{G}\rVert_{\infty} f_{\varepsilon|\mathcal{I}}(0))^{-1}$, which is the restriction set on λ in the binary data model.

When the network matrix is row normalized ($\lVert\mathbf{G}\rVert_{\infty} = 1$), Assumption 2.4 implies that $\lambda < B_c$. This is equivalent to assuming that the maximum of the marginal peer effects is less than one. Indeed, from the expected choice expression, $y_i^e = \ell_i(\mathbf{y}^e)$, the marginal expected choice with respect to average expected peers' choice $\partial \ell_i(\mathbf{y}^e)/\partial \bar{y}_i^e$ is given by $\lambda \sum_{r=1}^{\infty} f_{\varepsilon|\mathcal{I}}(\lambda \bar{y}_i^e + \psi_i - a_r)$. Thus, $\max_i(\partial \ell_i(\mathbf{y}^e)/\partial \bar{y}_i^e) < 1$ by Assumption 2.4. That is, 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).

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

[12] If $[0,1]^{\infty}$ were a compact set, any of its sequences would converge or would have a subsequent that converges. One can consider the sequence $(b_r)_r$ defined for any $r \geqslant 1$ by $b_r = (\dots, 0, 1, 0, \dots)$, an infinite-dimensional vector, where only the r-th entry is equal to one and the rest of the entries are equal to zero. This sequence and any subsequent of this sequence do not have a limit.

[13] It is important that $\max_{u\in\mathbb{R}} \sum_{r=1}^{\infty} f_{\varepsilon|\mathcal{I}}(u - a_r)$ be finite so that there exists a nonempty convex and compact set of λ's that verify Assumption 2.4. I state and prove a general lemma in OA S.1 on the convergence of all the infinite summations used in this paper.

The following theorem established the uniqueness of the Bayesian Nash equilibrium (BNE) and that of the rational belief system.

Proposition 2.3. *Under Assumptions 2.1–2.4, the game of incomplete information associated with payoff (1) has a unique BNE given by* $\mathbf{y}^* = (y_i^*, \ldots, y_n^*)'$ *and a unique rational expected outcome* $\mathbf{y}^{e*}$*, where* y_i^* *is the maximizer of the expected payoff* $U_i^e(.)$ *and* $\mathbf{y}^{e*} = \mathbf{L}(\mathbf{y}^{e*})$*.*

Proposition 2.3 gives sufficient conditions for the uniqueness of the rational belief system.[14] 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 2.4. Moreover, $\mathbf{p}^*$ 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 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.

3.1 Specification

As pointed out above, 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 $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_M)'$, $\boldsymbol{\beta}$, 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 $\mathbf{X} = [\mathbf{x}_1 \ \ldots \ \mathbf{x}_n]'$ and $\mathbf{W}$ be an $n \times M$ matrix, where the (i, m)-th entry is one if i belongs to the m-th subgroup and zero otherwise. I have $\boldsymbol{\psi} = \mathbf{Z}\boldsymbol{\Gamma}$, where $\mathbf{Z} = [\mathbf{W} \ \mathbf{X} \ \mathbf{G}\mathbf{X}]$ and $\boldsymbol{\Gamma} = \left(\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\gamma}'\right)'$. 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 (see Lancaster, 2000). When n grows to infinity, the number of groups M is fixed but the number of individuals in each subgroup grows to infinity.[15] With the specification of ψ_i, the set of information observed by the agent i and the econometrician is $\mathcal{I}_i = \{\mathbf{x}_i, \ \mathbf{X}_{-i}, \ \mathbf{g}_i, \ \mathbf{G}_{-i}\}$. Moreover, all unknown parameters of the model are assumed to be known by the agents.

[14] Assumption 2.4 may not be a necessary condition. In the case of the binary outcomes, Brock and Durlauf (2001) give a full picture of the equilibrium multiplicity. It is not straightforward to know the implication of the violation of Assumption 2.4 in this paper because the mapping $\mathbf{L}$ does not have a closed-form. Since $\mathbf{L}$ is defined on the unbounded space $[0, \infty[^n$, there is not even any guarantee that the expected outcome exists if $\lambda \geqslant B_c/\lVert\mathbf{G}\rVert_\infty$. This is different from the case of the binary response model, where the rational belief system necessarily exists.

[15] In some nonlinear models, M and n/M are assumed to converge to infinity. In this case, a consistent estimator can be constructed using a bias reduction approach (see Fernández-Val, 2009). This goes beyond the scope of this paper.

For any $r \geqslant 2$, let $\delta_r = a_r - a_{r-1}$ and $\delta_1 = 0$. The parameter a_r can be written as $a_r = a_1 + \sum_{k=1}^r \delta_k$ for any $r \geqslant 1$. As $a_r = \Delta c(r) + \lambda r - \frac{\lambda}{2}$, I have $\delta_r = \Delta\Delta c(r) + \lambda$ for any $r \geqslant 2$. This means that $\delta_r > \lambda$ because $c(.)$ is strictly convex. Indeed, the role of the strictly convex cost assumption is to have $a_r > 0$. I can get the same result with a weaker condition than Assumption 2.1. I could assume that $c(.)$ is a strictly increasing function that verifies $\Delta\Delta c(r) + \lambda > \epsilon$ for some $\epsilon > 0$. This rule allows some concave cost functions depending on the value of λ. However, I keep the condition $\delta_r \geqslant \lambda$ in practice because the uniqueness of the RE equilibrium set in Assumption 2.4 could be violated if $\delta_r < \lambda$ for large r.

In Equation (3), 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 relies the condition $\lim_{r\to\infty} r^{-\rho}(\Delta c(r+1) - \Delta c(r)) > 0$ set in Assumption 2.3. In particular, I assume that this limit is reached for large values of r.

Assumption 3.1. *There exists a constant $R \in \mathbb{N}^*$, such that $\forall\ r > R$, $\delta_r = (r-1)^\rho \bar{\delta} + \lambda$, where $\bar{\delta} > 0$ and $\rho \geqslant 0$.*

Assumption 3.1 is equivalent to setting that $r^{-\rho}(\Delta c(r+1) - \Delta c(r)) = \bar{\delta}$ for any $r \geqslant R$, i.e., the limit set in Assumption 2.3 is reached from R. As $\rho \geqslant 0$, Assumption 3.1 includes many representations of the cost function. The SAR/SART specifications are particular cases of my model because they impose $\rho = 0$ and $R = 1$. The case $\rho > 0$ corresponds to the situation where δ_r diverges as r grows to infinity. The RE Equation (3) can now be expressed as

$$p_{ir} = F_{\varepsilon|\mathcal{I}}\left(\lambda \bar{y}_i^e + \mathbf{z}_i'\mathbf{\Gamma} - a_r\right) - F_{\varepsilon|\mathcal{I}}\left(\lambda \bar{y}_i^e + \mathbf{z}_i'\mathbf{\Gamma} - a_{r+1}\right), \tag{4}$$

where $a_0 = -\infty$, $a_r = a_1 + \sum_{k=1}^r \delta_k$ for any $r \geqslant 1$, $\delta_1 = 0$, $\delta_r = (r-1)^\rho\bar{\delta} + \lambda$ for any $r > R$, $y_i^e = \ell_i(\mathbf{y}^e) = \sum_{r=1}^\infty F_{\varepsilon|\mathcal{I}}(\lambda \bar{y}_i^e + \mathbf{z}_i'\mathbf{\Gamma} - a_r)$, and $\mathbf{z}_i'$ is the i-th row of $\mathbf{Z}$.

3.2 Identification

The free objects to be identified in Equation (4) are λ, $\mathbf{\Gamma}$, $\boldsymbol{\delta} = (\delta_2, \dots, \delta_R)'$, $\bar{\delta}$, a_1, ρ, R, and $F_{\varepsilon|\mathcal{I}}$. Note that if Assumption 3.1 holds for some $R \in \mathbb{N}^*$, in the sense that $\delta_r = (r-1)^\rho\bar{\delta} + \lambda$ for any $r > R$, it also holds for all R' greater than R. Let $\bar{R} := \min\{R :\ \forall\ r > R,\ \delta_r = (r-1)^\rho\bar{\delta} + \lambda\}$, i.e., $\bar{R}$ is the smallest R for which Assumption 3.1 holds. My identification analysis focuses on $\bar{R}$ rather than R. The following definition states the conditions for two parameters to be observationally equivalent.

Definition 3.1. *The set $\{\lambda,\ \mathbf{\Gamma},\ \boldsymbol{\delta},\ \bar{\delta},\ a_1,\ \rho,\ \bar{R},\ F_{\varepsilon|\mathcal{I}}\}$ is observationally equivalent to the alternative $\{\tilde{\lambda},\ \tilde{\mathbf{\Gamma}} = (\tilde{\boldsymbol{\alpha}}', \tilde{\boldsymbol{\beta}}', \tilde{\boldsymbol{\gamma}}')',\ \tilde{\boldsymbol{\delta}} = (\tilde{\delta}_2, \dots, \tilde{\delta}_{\tilde{\bar{R}}})',\ \tilde{\bar{\delta}},\ \tilde{a}_1,\ \tilde{\rho},\ \tilde{\bar{R}},\ \tilde{F}_{\varepsilon|\mathcal{I}}\}$ if $p_{ir} = \tilde{p}_{ir}$ for any $i \in \mathcal{V}$ and $r \in \mathbb{N}$, where p_{ir} and $\tilde{p}_{ir}$ are defined as in Equation (4) with their corresponding parameters.*

Since $\boldsymbol{\alpha}$ and a_1 enter Equation (4) only through their difference, they cannot be identified. As in an ordered model, I set $a_1 = 0$.

Let f_ε and F_ε be the (unconditional) pdf and cdf of the error term ε_i. If ε_i is independent of $\mathbf{Z}$ and $\mathbf{G}$ and if f_ε is positive almost everywhere on $\text{supp}(\varepsilon_i)$,[16] the main condition for the model parameters and the distribution function F_ε to be identified (up to scale) is that the design matrix $\tilde{\mathbf{Z}} = [\bar{\mathbf{y}}^e \ \mathbf{Z}]$ be a full rank matrix (see Manski, 1988). The issue related to this condition is that $\tilde{\mathbf{Z}}$ contains $\bar{\mathbf{y}}^e$, which is an unobserved variable. It therefore seems difficult to verify this condition in practice. In the case of the binary and bounded ordered outcomes, $\tilde{\mathbf{Z}}$ 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; Lin et al., 2021). In the literature, this assumption is justified by the fact that $\text{supp}(y_i^e)$ is bounded independently of $\mathbf{z}_i$. Thus, $\mathbf{G}\mathbf{y}^e$ is not likely to be linearly dependent on $\mathbf{Z}$ in general (except for special cases of $\mathbf{Z}$). In contrast, in the current framework, $\text{supp}(y_i^e)$ depends on $\mathbf{z}_i$. As for linear-in-means models, the expected outcome is unbounded if there exists an explanatory variable (with a nonzero coefficient) whose density function is positive everywhere on $\mathbb{R}$, given the other regressors. It is then more difficult to argue that $\tilde{\mathbf{Z}}$ is full rank without providing formal proof. For example, if $\bar{R} = 1$ and $\rho = 0$, it can be shown that y_i^e is approximately linear for large values. Therefore, the identification may fail under certain conditions, as is the case in linear models (see Bramoullé et al., 2009).

I generalize the identification analysis provided by Bramoullé et al. (2009) to nonlinear models. My identification strategy is not only valid for the current model, but it also holds for binary and ordered outcome models (e.g., Lee et al., 2014; Liu, 2019) without assuming that $\tilde{\mathbf{Z}}$ is a full rank matrix.

Let $\boldsymbol{\omega} = (1, \ \mathbf{x}')'$ and $\tilde{\boldsymbol{\omega}} = (\boldsymbol{\omega}', \ \bar{\mathbf{x}}')'$. The subscript i is removed to denote the vectors for an arbitrary agent. In the following assumptions and elsewhere, for any vector $\mathbf{b}$, $\mathbf{b}_k$ denotes its k-th component.

Assumption 3.2. *(i) For any i, ε_i is independent of $\mathbf{Z}$ and $\mathbf{G}$. (ii) The density function f_ε is positive almost everywhere on $\mathbb{R}$.*

Assumption 3.3. *(i) $\sup_i \sum_{j=1}^n \mathbf{g}_{ij}$ is uniformly bounded in n. (ii) The elements $\mathbf{g}_{ij}$'s are at most of order h_n^{-1} uniformly in all i, j,*[17] *where the sequence $(h_n)_n$ can be bounded or divergent, such that h_n/n converges to zero as n grows to infinity.*

Assumption 3.4. *There exists at least one $l \in [2, \ \dots, \ K+1]$ such that the coefficient associated with ω_l is nonzero and such that for almost every value of the vector $\boldsymbol{\omega}_{-l}$, the conditional distribution function of ω_l given $\boldsymbol{\omega}_{-l}$ has everywhere positive density.*

Assumption 3.5. *(i) $\text{supp}(\boldsymbol{\omega})$ is not contained in a proper linear subspace of $\mathbb{R}^{K+1}$. (ii) If a subgroup*

[16] Throughout, for a random variable V, $\text{supp}(V)$ denotes the support of V.

[17] That is, there exists $g_0 \geqslant 0$, such that for any i, j, and n, $|\mathbf{g}_{ij}h_n| < g_0$, and for any sequence $(\hat{h}_n)_n$, such that $\hat{h}_n/h_n$ diverges, there exists i, j, such that $\mathbf{g}_{ij}\hat{h}_n$ diverges almost surely.

s contains a positive proportion of individuals who have friends, then $\text{supp}(\tilde{\boldsymbol{\omega}}_{|s})$ *is not contained in a proper linear subspace of* $\mathbb{R}^{2K+1}$*, where* $\tilde{\boldsymbol{\omega}}_{|s}$ *is* $\tilde{\boldsymbol{\omega}}_i$ *for an arbitrary* i *from the subgroup* s*.*

Assumption 3.6. *(i) There exists* $k_0 \in [1, K]$ *such that* $\boldsymbol{\beta}_{k_0}\boldsymbol{\gamma}_{k_0} \geqslant 0$ *and* $\boldsymbol{\gamma}_{k_0} \neq 0$*. (ii) There exists a subgroup* s_0 *in which the proportion of agents who have friends is strictly positive (also as* n *grows to infinity). (iii) The cardinality of the set* $\mathcal{T}_{s_0,n} = \{i \in s_0 : \exists\ j, l \in s_0,$ *where* $i \neq l$ *such that* $\mathbf{g}_{ij} > 0,\ \mathbf{g}_{jl} > 0,$ *and* $\mathbf{g}_{il} = 0\}$ *in proportion to* $|s_0|$ *is strictly positive (also as* n *grows to infinity).*

Interestingly, Assumptions 3.2–3.6 do not involve the expected outcome $\mathbf{y}^e$. Condition (i) of Assumption 3.2 suggests there is no omission of important regressors in $\mathbf{X}$, so that $F_{\varepsilon|\mathcal{I}} = F_{\varepsilon}$. This restriction is common in the literature (e.g., Lee et al., 2014; Blume et al., 2015; Yang and Lee, 2017; Guerra and Mohnen, 2020). I discuss in Section 3.4 how to control for network endogeneity in the case where unobserved factors are correlated to both ε_i and $\mathbf{G}$. Condition (ii) ensures that there is much variation in y to allow for the identification of the cut points. In particular, the event $\{y_i > \bar{R}\}$ has a nonzero probability of occurrence. In practice, it is important to set $\bar{R}$ smaller than $\hat{\max}(y) - 2$ to identify $\bar{\delta}$ and ρ, where $\hat{\max}(y)$ is the empirical maximum of y (see Section 3.3).

Assumption 3.3 is taken directly from Lee (2004) and rules out the cases where $\bar{y}_i^e$ or $\bar{\mathbf{x}}_i$ diverges as n grows to infinity. This assumption is verified if $\tilde{n}_i$, the number of friends i has, and $\mathbf{g}_{ij}$ are uniformly bounded in i, j, and n. It can also hold when $\tilde{n}_i$ diverges but at some rate lower than that of n (since $h_n/n \to 0$). However, the assumption is violated when agents are connected to each other in their subgroup (even if the network matrix is row normalized).

Assumptions 3.4 and 3.5 originate from Manski (1988). They impose an unbounded support for one element of $\boldsymbol{\omega}$ and linear independence among the observable regressors $\mathbf{X}$ and $\mathbf{GX}$. Assumption 3.4 is only necessary for the identification of F_{ε}. It can be released if F_{ε} is known or assumed by the practitioner (see Manski, 1988, Proposition 5). Both assumptions require sufficient variation in the regressors $\mathbf{X}$ and $\mathbf{GX}$ to allow certain nonlinearities.

Condition (i) of Assumption 3.6 is a sufficient condition for the expected choice y_i^e to be affected by contextual variables. To see why, without loss of generality, assume that $\boldsymbol{\beta}_{k_0} \geqslant 0$ and $\boldsymbol{\gamma}_{k_0} > 0$. If j is i's friend, then an increase in $\mathbf{x}_{j,k_0}$, the k_0-th component of $\mathbf{x}_j$, implies an increase in y_i^e (because $\boldsymbol{\gamma}_{k_0} > 0$) and possibly an increase in y_j^e (because $\boldsymbol{\beta}_{k_0} \geqslant 0$). There would also be an indirect positive effect on y_i^e because of the increase in y_j^e (since $\lambda \geqslant 0$). Overall, the increase in $\mathbf{x}_{j,k_0}$ certainly implies an increase in y_i^e. A similar restriction is also set by Bramoullé et al. (2009). The condition in their case is $\lambda\boldsymbol{\beta}_{k_0} + \boldsymbol{\gamma}_{k_0} \neq 0$, which is implied by Condition (i). Bramoullé et al. (2009) set the necessary condition for the outcome to be affected by the contextual variables whereas Condition (i) is sufficient. Since my model is nonlinear, the necessary condition would not be straightforward as for the linear models. With several characteristics, the condition I set is likely to be verified.

Condition (ii) of Assumption 3.6 imposes the existence of a subgroup s_0 with a positive proportion of individuals who have friends. This is necessary for the identification of $\boldsymbol{\gamma}$. Finally, Condition (iii) means that s_0 contains a positive proportion of agents whose friends' friends are not all friends. This condition implies that $\mathbf{I}_n$,[18] $\mathbf{G}$, and $\mathbf{G}^2$ are linearly independent, which is also one of the conditions for identifying the linear-in-means model parameters (see Bramoullé et al., 2009, Proposition 1). As for Condition (i), Condition (iii) is sufficient and not necessary.

Proposition 3.1. *Under Assumptions 3.2–3.6, λ, $\boldsymbol{\Gamma}$, $\boldsymbol{\delta}$, $\bar{\delta}$, and the distribution function F_ε are identified up to scale, whereas ρ and $\bar{R}$ are point identified.*

The proof of Proposition 3.1 is done in several steps. I first show that $\tilde{\mathbf{Z}} = [\bar{\mathbf{y}}^e \ \mathbf{Z}]$ is a full rank matrix. As $\mathbf{Z}$ is a full rank matrix (Assumption 3.5, Condition (ii)), it is sufficient to show that $\bar{y}_i^e$ cannot be written as $\check{\alpha}+\mathbf{x}_i'\check{\boldsymbol{\beta}}+\bar{\mathbf{x}}_i'\check{\boldsymbol{\gamma}}$ for all i, where $\check{\alpha} \in \mathbb{R}$ and $\check{\boldsymbol{\beta}}, \check{\boldsymbol{\gamma}} \in \mathbb{R}^K$. Assume that $\bar{y}_i^e = \check{\alpha}+\mathbf{x}_i'\check{\boldsymbol{\beta}}+\bar{\mathbf{x}}_i'\check{\boldsymbol{\gamma}}$ is true. By taking the total differential of $\bar{y}_i^e$ (keeping $\mathbf{G}$ and the parameters fixed), I get $d\bar{y}_i^e = (d\mathbf{x}_i)'\check{\boldsymbol{\beta}} + (d\bar{\mathbf{x}}_i)'\check{\boldsymbol{\gamma}}$. Then, assume an increase in $\mathbf{x}_l$ for some $l \neq i$ who is not i's friend, *ceteris paribus*, i.e., $d\mathbf{x}_l > 0$, $d\mathbf{x}_i = 0$, and $d\bar{\mathbf{x}}_i = 0$. This implies that $d\bar{y}_i^e = 0$. Since $\boldsymbol{\beta}_{k_0}\boldsymbol{\gamma}_{k_0} \geqslant 0$ and $\boldsymbol{\gamma}_{k_0} \neq 0$, an increase in $\mathbf{x}_l$ cannot imply both $dy^e < 0$ for some agents and $dy^e > 0$ for others. Thus, $d\bar{y}_i^e = 0$ necessarily means that $dy_j^e = 0$ for any j who is i's friend. Put differently, $\mathbf{x}_l$ cannot influence y_j^e if j is i's friend and l is not. This cannot hold under Condition (iii) of Assumption 3.6 because l could be j's friend.

For example, consider a network of three agents i, j, and l, as illustrated by Figure 1. Assume that j is i's only friend. This means that $\bar{y}_i^e = y_j^e$ and $\bar{\mathbf{x}}_i = \mathbf{x}_j$ (I normalize $\mathbf{g}_{ij}$ to one). Assume also that l is j's only friend. Thus, $dy_j^e/d\mathbf{x}_l \neq 0$ by Assumption 3.6, Condition (i). However, Equation $\bar{y}_i^e = \check{\alpha}_{s_0} + \mathbf{x}_i'\check{\boldsymbol{\beta}} + \bar{\mathbf{x}}_i'\check{\boldsymbol{\gamma}}$, also equivalent to $y_j^e = \check{\alpha}_{s_0} + \mathbf{x}_i'\check{\boldsymbol{\beta}} + \mathbf{x}_j'\check{\boldsymbol{\gamma}}$, means that $dy_j^e/d\mathbf{x}_l = 0$. As a result, if there are intransitive triads in the network, $\bar{y}_i^e$ cannot be written as $\check{\alpha}_{s_0} + \mathbf{x}_i'\check{\boldsymbol{\beta}} + \bar{\mathbf{x}}_i'\check{\boldsymbol{\gamma}}$.

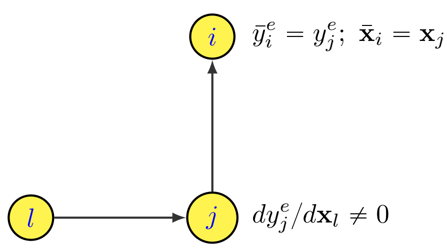

Figure 1: Illustration of the identification

I now show the identification of λ, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and F_ε. By Definition 3.1, if λ, α_{s_0}, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and F_ε are observationally equivalent to their alternative, then $p_{i0} = \tilde{p}_{i0}$ for any $i \in s_0$. By the exogeneity of $\mathbf{Z}$ and $\mathbf{G}$, this implies $F_\varepsilon(\alpha_{s_0}+\lambda y_i^e+\mathbf{x}_i'\boldsymbol{\beta}+\bar{\mathbf{x}}_i'\boldsymbol{\gamma}) = \tilde{F}_\varepsilon(\tilde{\alpha}_{s_0}+\tilde{\lambda}\tilde{y}_i^e+\mathbf{x}_i'\tilde{\boldsymbol{\beta}}+\bar{\mathbf{x}}_i'\tilde{\boldsymbol{\gamma}})$, where $\tilde{y}_i^e$ is the expected outcome in the case of the alternative parameters. Moreover, since $p_{ir} = \tilde{p}_{ir}$ for any r, I have $\sum_{r=1}^{\infty} rp_{ir} =$

[18] $\mathbf{I}_n$ is the identity matrix of dimension n.

$\sum_{r=1}^{\infty} r\tilde{p}_{ir}$, meaning that $y_i^e = \tilde{y}_i^e$. Thus, $F_\varepsilon(\alpha_{s_0} + \lambda y_i^e + \mathbf{x}_i'\boldsymbol{\beta} + \bar{\mathbf{x}}_i'\boldsymbol{\gamma}) = \tilde{F}_\varepsilon(\tilde{\alpha}_{s_0} + \tilde{\lambda} y_i^e + \mathbf{x}_i'\tilde{\boldsymbol{\beta}} + \bar{\mathbf{x}}_i'\tilde{\boldsymbol{\gamma}})$. By Proposition 2, Corollary 5 in Manski (1988), λ, α_{s_0}, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and F_ε are identified up to scale.[19] In addition, $\boldsymbol{\alpha}$ is identified up to scale by using the equation $p_{i0} = \tilde{p}_{i0}$ in each subgroup.

I now turn to the proof of the identification of $\boldsymbol{\delta}$, $\bar{\delta}$, ρ, and $\bar{R}$. The equation $p_{ir} = \tilde{p}_{ir}$ implies that $\sum_{k=1}^{r} \delta_k = \sum_{k=1}^{r} \tilde{\delta}_k$ for any $r \geqslant 1$, where $\tilde{\delta}_1 = 0$ and $\tilde{\delta}_r = (r-1)^{\tilde{\rho}}\tilde{\bar{\delta}} + \lambda$ for any $r > \tilde{\bar{R}}$. The condition $\sum_{k=1}^{r} \delta_k = \sum_{k=1}^{r} \tilde{\delta}_k$ suggests that $\delta_r = \tilde{\delta}_r$ for any $r \geqslant 1$. Especially, for $r \geqslant \max\{\bar{R}, \tilde{\bar{R}}\}$, I have $(r-1)^{\rho}\bar{\delta} = (r-1)^{\tilde{\rho}}\tilde{\bar{\delta}}$ and $r^{\rho}\bar{\delta} = r^{\tilde{\rho}}\tilde{\bar{\delta}}$, which means that $\bar{\delta} = \tilde{\bar{\delta}}$ and $\rho = \tilde{\rho}$. As a result, $\bar{\delta}$ is identified up to scale and ρ is point identified. Note that ρ is point identified because the scale of F_ε has no impact on ρ. The identification of $\bar{R}$ is trivial. Assume without loss of generality that $\bar{R} > \tilde{\bar{R}}$. This means that $\delta_{\bar{R}} = \tilde{\delta}_{\bar{R}} = (r-1)^{\rho}\bar{\delta} + \lambda$ for any $r \in (\tilde{\bar{R}}, \bar{R}]$. This representation of δ_r is not possible at $r = \bar{R}$ given the definition of $\bar{R}$. Finally $\boldsymbol{\delta}$ is identified up to scale because $\bar{R} = \tilde{\bar{R}}$ and $\delta_r = \tilde{\delta}_r$.

Showing that $\tilde{\mathbf{Z}}$ is a full rank matrix is sufficient to identify the model parameters if F_ε were known or assumed by the practitioner (see Manski, 1988, Proposition 5). I then have the following result, which is a direct implication of Proposition 3.1.

Corollary 3.1 (Proposition 3.1)**.** *Assume that Condition (i) of Assumption 3.2 and Assumptions 3.3, 3.5–3.6 hold. If F_ε is known (assumed by the practitioner), then λ, $\boldsymbol{\Gamma}$, $\boldsymbol{\delta}$, $\bar{\delta}$, ρ, and $\bar{R}$ are point identified.*

This identification strategy can be applied to a large class of models. Indeed, for any model defined by $y^e = \check{h}(\alpha_s + \check{m}(y^e) + \mathbf{x}'\boldsymbol{\beta} + \bar{\mathbf{x}}'\boldsymbol{\gamma})$, where $\check{h}$ and $\check{m}$ are strictly increasing functions, $\check{m}(y^e)$ cannot be a linear combination of the other regressors under Assumptions 3.5, and 3.6. This result is also true for models with complete information. Moreover, unlike Bramoullé et al. (2009), it does not require that there are no isolated agents in the network.

3.3 Estimation

The estimation strategy is based on a likelihood approach. Computing the likelihood of the outcome 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. 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).

Assumption 3.7. $\varepsilon_i \sim \mathcal{N}(0, 1)$.

[19] The Corollary imposes three conditions: the statistical independence (SI) assumption and Conditions X1 and X3. Assumptions 2.2 and 3.2 guarantee the SI condition. Condition X1 is exactly the linear independence between y^e and $\tilde{\boldsymbol{\omega}}$ shown above. Condition X3 is filled by Assumption 3.4. Indeed, by increasing ω_l of Assumption 3.4 to ∞ (or $-\infty$), y^e will tend either to ∞ or to zero.

The variance of ε_i is normalized to one because some parameters are identified up to scale. Under Assumption 3.7, Equation (4) can now be expressed as

$$p_{ir} = \Phi\left(\lambda \mathbf{g}_i \mathbf{y}^e + \mathbf{z}_i'\boldsymbol{\Gamma} - a_r\right) - \Phi\left(\lambda \mathbf{g}_i \mathbf{y}^e + \mathbf{z}_i'\boldsymbol{\Gamma} - a_{r+1}\right), \tag{5}$$

where Φ is the cdf of $\mathcal{N}(0,1)$, $a_0 = -\infty$, $a_r = \sum_{k=1}^{r} \delta_k$ for any $r \geqslant 1$, and $\mathbf{y}^e = \mathbf{L}(\boldsymbol{\theta}, \mathbf{y}^e)$. Given the specification $\boldsymbol{\psi} = \mathbf{Z}\boldsymbol{\Gamma}$, the mapping $\mathbf{L}$ is now defined as $\mathbf{L}(\boldsymbol{\theta}, \mathbf{y}^e) = (\ell_1(\boldsymbol{\theta}, \mathbf{y}^e), \dots, \ell_n(\boldsymbol{\theta}, \mathbf{y}^e))'$, where $\ell_i(\boldsymbol{\theta}, \mathbf{y}^e) = \sum_{r=1}^{\infty} \Phi\left(\lambda \mathbf{g}_i \mathbf{y}^e + \mathbf{z}_i'\boldsymbol{\Gamma} - a_r\right)$. A direct implication of Assumption 3.7 is that the model parameters are point identified (see Corollary 3.1).

In practice, $\bar{R}$ is unknown and has to be set by the practitioner. For any fixed $\bar{R}$, if $\mathbf{y}^e$ were observed, estimating the model would result in a classical *probit* model estimation by the maximum likelihood (ML) method. As $\mathbf{y}^e$ is not observed, the ML approach requires computing $\mathbf{y}^e$, i.e., solving a fixed point problem in $\mathbb{R}^n$ for every value of $\boldsymbol{\theta}$. This may be computationally cumbersome for large samples. I then use the NPL algorithm proposed by Aguirregabiria and Mira (2007), which is computationally more attractive than the ML method. This algorithm uses an iterative process that does not require solving a fixed point problem.

It is worth mentioning that Aguirregabiria and Mira (2007) show the consistency of the NPL estimator by assuming that the expected outcome is bounded. A sufficient condition for this assumption to hold is that supp($\mathbf{x}$) is bounded. Although this condition would be generally verified in practice, it could be incompatible with Assumption 3.4. However, this does not raise any issues because Assumption 3.4 is no longer necessary when the distribution of ε_i is assumed (see Corollary 3.1).

To deal with the constraint $\delta_r \geqslant \lambda$ for all $r \geqslant 2$, I define $\tilde{\boldsymbol{\delta}} = (\tilde{\delta}_2, \dots, \tilde{\delta}_{\bar{R}})$, where $\tilde{\delta}_r = \delta_r - \lambda$. The NPL algorithm is based on a pseudo-likelihood function defined as

$$\mathcal{L}_n(\boldsymbol{\theta}, \mathbf{y}^e) = \frac{1}{n} \sum_{i=1}^{n} \sum_{r=0}^{\infty} d_{ir} \log(p_{ir}), \tag{6}$$

where $\boldsymbol{\theta} = \Big(\log(\lambda),\ \boldsymbol{\Gamma}', \log(\tilde{\boldsymbol{\delta}}'), \log(\bar{\delta}), \log(\rho)\Big)'$, $d_{ir} = 1$ if $y_i = r$ and $d_{ir} = 0$ otherwise. By taking λ and ρ in log, I indirectly assume that they are strictly positive. However, they can be zero. In particular, $\rho = 0$ is an interesting case, meaning that the model is approximately linear for large values of y_i^e. In practice, these alternative specifications can be estimated and criterion information or a likelihood ratio test can be used to compare them to the general model.

To describe NPL iterations, it is useful to define the operators $\boldsymbol{\phi}_n(\mathbf{y}^e) \equiv \mathbf{L}(\tilde{\boldsymbol{\theta}}(\mathbf{y}^e), \mathbf{y}^e)$, where $\tilde{\boldsymbol{\theta}}_n(\mathbf{y}^e) \equiv \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathcal{L}_n(\boldsymbol{\theta}, \mathbf{y}^e)$. The NPL algorithm starts with a proposal $\mathbf{y}^e_{(0)}$ for $\mathbf{y}^e$ and constructing the sequence of estimators $(\mathcal{Q}_t)_{t \geqslant 1}$, such that $\mathcal{Q}_t = \{\boldsymbol{\theta}_{(t)}, \mathbf{y}^e_{(t)}\}$, where $\boldsymbol{\theta}_{(t)} = \tilde{\boldsymbol{\theta}}_n(\mathbf{y}^e_{(t-1)})$ and $\mathbf{y}_{(t)} = \boldsymbol{\phi}_n(\mathbf{y}^e_{(t-1)})$. In other words, given the guess $\mathbf{y}^e_0$, $\boldsymbol{\theta}_{(1)} = \tilde{\boldsymbol{\theta}}_n(\mathbf{y}^e_{(0)})$ and $\mathbf{y}^e_{(1)} = \boldsymbol{\phi}_n(\mathbf{y}^e_{(0)})$, then $\boldsymbol{\theta}_{(2)} =$

$\tilde{\theta}_n(\mathbf{y}^e_{(1)})$ and $\mathbf{y}^e_{(2)} = \boldsymbol{\phi}_n(\mathbf{y}^e_{(1)})$, and so forth. Notice that each value of $(\mathcal{Q}_t)_{t\geqslant 1}$ requires evaluating the mapping $\mathbf{L}$ only once. If $(\mathcal{Q}_t)_{t\geqslant 1}$ converges, regardless of the initial guess $\mathbf{y}^e_0$, its limit, denoted $\{\hat{\theta}_n(\bar{R}), \hat{\mathbf{y}}^e_n(\bar{R})\}$, is the NLP estimator. This limit satisfies the following two properties: $\hat{\theta}_n(\bar{R})$ maximizes the pseudo-likelihood $\mathcal{L}_n(\theta, \hat{\mathbf{y}}^e_n(\bar{R}))$ and $\hat{\mathbf{y}}^e_n(\bar{R}) = \mathbf{L}(\hat{\theta}_n(\bar{R}), \hat{\mathbf{y}}^e_n(\bar{R}))$. 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 $\mathbf{L}$. In practice, the convergence is reached when $\|\theta_{(t)} - \theta_{(t-1)}\|_1$ and $\|\mathbf{y}^e_{(t)} - \mathbf{y}^e_{(t-1)}\|_1$ are less than some tolerance values (for example 10^{-4}).[20]

Choice of $\bar{R}$ and asymptotic properties

I now discuss how to set $\bar{R}$ and study the limiting distribution of the proposed estimator.
Let $\hat{\bar{R}}$ be the value of $\bar{R}$ set empirically and $\bar{R}^0$ be the true value. In this section and elsewhere, any parameter with a superscript zero denotes its true value. As $\boldsymbol{\delta}$ is an $(\bar{R}-1)$-vector, $\dim(\hat{\theta}_n(\hat{\bar{R}}))$, the dimension of $\hat{\theta}_n(\hat{\bar{R}})$ depends on $\hat{\bar{R}}$, whereas $\dim(\theta^0)$ depends on $\bar{R}^0$. Thus, $\hat{\theta}_n(\hat{\bar{R}})$ and $\bar{R}^0$ are not in the same space if $\hat{\bar{R}} \neq \bar{R}^0$. However, they can be redefined in the same space. For example, if $\hat{\bar{R}} > \bar{R}^0$, then $\boldsymbol{\delta}^0$ can be redefined as $\boldsymbol{\delta}^0 = (\delta^0_2, \dots, \delta^0_{\bar{R}^0}, \dots, \delta^0_{\hat{\bar{R}}})'$, where $\delta^0_r = (r-1)^{\rho^0}\delta^0 + \lambda^0$ for any $r > \bar{R}^0$. I establish the following result on the consistency of the NPL parameter.

Proposition 3.2. *Assume that Condition (i) of Assumption 3.2, Assumptions 3.3, 3.5–3.7, and the regulatory conditions of the NPL estimator hold (see Appendix A.4). For any $\hat{\bar{R}}$, (i) $\hat{\theta}_n(\hat{\bar{R}})$ converges in probability to a random variable $\check{\theta}_n(\hat{\bar{R}})$, where $\check{\theta}_n(\hat{\bar{R}}) = \theta^0$ if $\hat{\bar{R}} \geqslant \bar{R}^0$; (ii) $\sqrt{n}(\hat{\theta}_n(\hat{\bar{R}}) - \check{\theta}_n(\hat{\bar{R}})) \xrightarrow{d} \mathcal{N}\left(0, (\mathbf{H}_{1,0} + \mathbf{H}_{2,0})^{-1}\boldsymbol{\Sigma}_0(\mathbf{H}'_{1,0} + \mathbf{H}'_{2,0})^{-1}\right)$, where $\boldsymbol{\Sigma}_0$, $\mathbf{H}_{1,0}$, and $\mathbf{H}_{2,0}$ are given in Appendix A.4.*

Proposition 3.2 adapts Proposition 2 in Aguirregabiria and Mira (2007) to my framework (see proof in Appendix A.4). The NPL estimator may not converge to θ^0 if $\hat{\bar{R}} < \bar{R}^0$. In this case, I establish the convergence to a random variable because my asymptotic is conditional on $\mathbf{X}$ and $\mathbf{G}$.[21] Indeed, the expected choices y^e_i's are not independent if I do not condition on $\mathbf{X}$ and $\mathbf{G}$. My asymptotic is conditional on $\mathbf{X}$ and $\mathbf{G}$ as in Lin and Xu (2017). This is different from that of Aguirregabiria and Mira (2007) where M independent markets are observed and the asymptotic assumes that M grows to infinity.
To choose a suitable $\hat{\bar{R}}$, Proposition 3.2 suggests that $\hat{\theta}_n(\hat{\bar{R}})$ can be computed for several values $\hat{\bar{R}}$ from one to $\hat{\max}(y) - 2$. An appropriate $\hat{\bar{R}}$ is reached if an increase in $\hat{\bar{R}}$ does not significantly change

[20] The pseudo-likelihood (6) involves an infinite sum. As $d_{ir} = 0$ for any $r \neq y_i$, this pseudo-likelihood can also be expressed as $\mathcal{L}_n(\theta, \mathbf{y}^e) = \frac{1}{n}\sum_{i=1}^n \log(p_{iy_i})$. Moreover, the mapping $\mathbf{L}$ and the asymptotic variance of $\hat{\theta}$ also involve an infinite sum. Note that the summed elements decrease exponentially. A very good approximation of these sums can be readily reached by summing only a few elements. My R package may be used for this purpose.

[21] Even if $\hat{\bar{R}} < \bar{R}^0$, $\check{\theta}_n(\hat{\bar{R}})$ would converge to a constant (different to θ^0) as n grows to infinity. Thus, I could simply write "$\hat{\theta}_n(\hat{\bar{R}})$ converges to a nonstochastic quantity." However, I keep the convergence to the random variable because the asymptotic normality is shown with respect to this random variable.

the estimator. If $\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}})$ varies until $\hat{\bar{R}}$ equals $\hat{\max}(y) - 2$, then the highest value can be used. Even then, I found using simulations that the finite sample bias would be marginal compared with that of the linear-in-mean model (see Section 4). As the estimation strategy is straightforward, setting $\hat{\bar{R}}$ very large raises no computational issues. Moreover, since $\bar{R}^0$ is a constant, the model does not suffer from an incidental parameter issue.

3.4 Endogenous Networks

In the recent literature, many studies point out that the network matrix $\mathbf{G}$ and the outcome $\mathbf{y} = (y_1, \dots, y_n)'$ can be linked through individual-level attributes that are unobserved by the econometrician (see Hsieh et al., 2020; Johnsson and Moon, 2021). An example related to my application is students' sociability degree (attribute), which may increase both student participation in extracurricular activities (outcome) and student likelihood of interacting with others. Not controlling for the attributes renders $\mathbf{G}$ dependent on ε_i.

As a classical argument in the literature (see Hsieh and Lee, 2016; Graham, 2017; Dzemski, 2019; Yan et al., 2019), the latent random utility associated with j being i's friend is $\mathbf{g}_{ij}^* = \ddot{\mathbf{x}}_{ij}'\bar{\boldsymbol{\beta}} + \mu_i + \nu_j + \eta_{ij}$, where $\ddot{\mathbf{x}}_{ij}$ is a $\bar{K}$-vector of observed dyad-specific variables (e.g., $\ddot{\mathbf{x}}_{ij}$ may contain the distance between agents i and j's characteristics), $\bar{\boldsymbol{\beta}} \in \mathbb{R}^{\bar{K}}$ is the slope of the utility with respect to $\ddot{\mathbf{x}}_{ij}$, μ_i and ν_j are unobserved attributes, and η_{ij} is an idiosyncratic error term. I assume that η_{ij} follows a symmetric distribution characterized by a distribution function F_η. The term $\ddot{\mathbf{x}}_{ij}'\bar{\boldsymbol{\beta}}$ is a measure of social distance between agents i and j that drives homophily of linking decisions.

Agent j is i's friend if the random utility $\mathbf{g}_{ij}^*$ exceeds a threshold. For the parameters to be identified, this threshold is set to zero. This implies that the probability for j to be i's friend conditionally on $\ddot{\mathbf{x}}_{ij}$, $\bar{\boldsymbol{\beta}}$, μ_i, and ν_j, is defined as

$$P_{ij} = \mathbb{P}\left(\mathbf{g}_{ij} > 0 | \ddot{\mathbf{x}}_{ij}, \bar{\boldsymbol{\beta}}, \mu_i, \nu_j\right) = F_\eta\left(\ddot{\mathbf{x}}_{ij}'\bar{\boldsymbol{\beta}} + \mu_i + \nu_j\right). \tag{7}$$

By convention, I set $P_{ii} = 0$ and $P_{ij} = 0$ if $s(i) \neq s(j)$. Unlike most network formation models, the specification (7) 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.

On the other hand, I assume that ε_i is independent of $\mathbf{G}$ and $\mathbf{X}$ after controlling for the unobserved attributes. The intuition behind this assumption is similar to that of the control function approach developed by Johnsson and Moon (2021) and the partially linear model proposed by Auerbach (2022).

Assumption 3.8. *For some continuous function h_ε, $\varepsilon_i = h_\varepsilon(\mu_i, \nu_i, \bar{\mu}_i, \bar{\nu}_i) + \varepsilon_i^*$, where ε_i^* is independent of $\mathbf{Z}$ and $\mathbf{G}$, $\bar{\mu}_i = \sum_{j=1}^n \mathbf{g}_{ij}\mu_j$, and $\bar{\nu}_i = \sum_{j=1}^n \mathbf{g}_{ij}\nu_j$.*

Including $\bar{\mu}_i$ and $\bar{\nu}_i$ in the function h_ε controls for situations where friend attributes influence y_i. Referring to my application, this means that friends' degree of sociability may increase one's participation in extracurricular activities.

The attributes μ_i and ν_i are unobserved by the practitioner, but I assume that they are observed by the players in the game; that is, the information set $\mathcal{I}$ includes the attributes.[22] This assumption plays an important role. Indeed, if the agents do not observe the attributes, then they should be integrated out in Equation (2.3). In contrast, if the attributes are observed, then ε_i in Assumption 2.2 can be replaced by $h_\varepsilon(\mu_i, \nu_i, \bar{\mu}_i, \bar{\nu}_i) + \varepsilon_i^*$. Given Assumption 3.8, for any $x \in \mathbb{R}$, $F_{\varepsilon|\mathcal{I}}(x) = F_\varepsilon^*(x - h_\varepsilon(\mu_i, \nu_i, \bar{\mu}_i, \bar{\nu}_i))$, where F_ε^* is the distribution function of ε_i^*. Thus, the BNE defined in Proposition 2.3 is still valid if one adapts Assumptions 2.2–2.3 to the new error term ε_i^*. I also adapt Assumption 3.7: ε_i^*'s are independent and identically distributed by $\mathcal{N}(0, 1)$.

My estimation strategy is in two stages. In the first stage, I estimate the unobserved attributes from Equation (7). There is a growing literature on network formation models with degree heterogeneity (see De Paula, 2020). Let $\hat{\boldsymbol{\mu}}_n = (\hat{\mu}_1, \dots, \hat{\mu}_n)'$ and $\hat{\boldsymbol{\nu}}_n = (\hat{\nu}_1, \dots, \hat{\nu}_n)'$, where $\hat{\mu}_i$ and $\hat{\nu}_i$ are the standard logit estimators of μ_i and ν_i, respectively. Let also $\hat{\boldsymbol{\chi}}_n = (\hat{\boldsymbol{\mu}}_n', \ \hat{\boldsymbol{\nu}}_n')'$ and $\boldsymbol{\chi}_n^0$ be the true value. Yan et al. (2019) show that $\|\hat{\boldsymbol{\chi}}_n - \boldsymbol{\chi}_n^0\|_\infty = O_p\left((\log(n)/n)^{1/2}\right)$. I refer the interested reader to Yan et al. (2019) and Dzemski (2019) for a formal discussion of the model, including its identification and consistent estimation. Alternatively, a Bayesian probit approach can be used to simulate the posterior distribution of $\hat{\boldsymbol{\chi}}_n$. However, this approach treats the attributes as random effects independent from $\mathbf{X}$ (see Hsieh et al., 2020; Albert and Chib, 1993).[23] I implement both approaches in my R package. At the second stage, I replace μ_i and ν_i with their estimator. The function h_ε can be approximated using a sieve method, such as series approximation (see Ackerberg et al., 2012). That is, $h_\varepsilon(\mu_i, \nu_i, \bar{\mu}_i, \bar{\nu}_i) = \sum_{k=1}^T (\theta_{1,k}\mu_i^k + \theta_{2,k}\nu_i^k + \theta_{3,k}\bar{\mu}_i^k + \theta_{4,k}\bar{\nu}_i^k)$, where T is a fixed integer and $\theta_{1,k}$, $\theta_{2,k}$, $\theta_{3,k}$, and $\theta_{4,k}$ are parameters to be estimated. The degree T can be set such that attributes to a power greater than T are not significant.[24]

As in Johnsson and Moon (2021), controlling for the endogeneity is substantially equivalent to including additional regressors in Equation (5). Accordingly, the parameters of the counting variable model are identified if the new set of explanatory variables verifies the condition of Corollary 3.1. Therefore,

[22] For example, agents know their sociability degree and that of their friends.

[23] The probit or logit approach assumes that η_{ij} follows a normal or logistic distribution, which implies a dense network. However, the model also performs well with sparse networks (see Dzemski, 2019).

[24] This approximation is used because the specification is nonlinear. In the case of a linear model, the estimation can be performed without using h_ε (see Johnsson and Moon, 2021). Indeed, including $h_\varepsilon(\mu_i, \nu_i, \bar{\mu}_i, \bar{\nu}_i)$ as a regressor is equivalent to regressing the expectation of the dependent variable on the expectation of the regressors, where the expectations are conditional on μ_i, ν_i, $\bar{\mu}_i$, and $\bar{\nu}_i$. General nonparametric methods can be used to estimate the conditional expectations.

a nonidentification issue does not arise here as in Auerbach (2022). Under regulatory conditions similar to those imposed by Johnsson and Moon (2021), I establish the asymptotic normality of the new estimator (see OA S.2).

4 Monte Carlo Experiments

In this section, I conduct a Monte Carlo study to assess the performance of the estimator in finite samples. I also compare the model to the SART model to illustrate that the latter is a particular case. I use the SART model because it controls for the left-censure issue.

I consider only one subnetwork of $n = 1,500$ individuals, and I assume that $\psi_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \gamma_1 \bar{x}_{1i} + \gamma_2 \bar{x}_{2i}$. My simulation study does not access the performance of the fixed effect estimator. In fact, as I assume that the number of subnetworks is finite, the fixed effects can be viewed as ordinary explanatory variables.

The exogenous variables x_1 and x_2 are simulated from $\mathcal{N}(1,1)$ and $\mathcal{P}oisson(2)$, respectively. Each individual i is randomly assigned to $\tilde{n}_i$ friends, where $\tilde{n}_i$ is randomly chosen between 0 and 30. The network matrix $\mathbf{G}$ used is the row-normalized adjacency matrix. The parameters are set as follows: $\lambda^0 = 0.25$, $\alpha^0 = 2.5$, $\boldsymbol{\beta}^0 = (1.5, -1.2)'$, and $\boldsymbol{\gamma}^0 = (0.5, -0.9)'$. I consider three data generating processes (DGP), A–C, associated with different values of $\bar{R}^0$. For the DGP A, $\bar{R}^0 = 5$, where $\boldsymbol{\delta}^0 = (1,\ 0.87,\ 0.75,\ 0.55)'$, $\bar{\delta} = 0.05$, and $\rho^0 = 0.3$. For the DGP B, $\bar{R}^0 = 13$, where $\boldsymbol{\delta}^0 = (1.2,\ 0.7,\ 0.55,\ 0.5,\ 0.5,\ 0.4,\ 0.4,\ 0.3,\ 0.3,\ 0.27,\ 0.27,\ 0.25)'$, $\bar{\delta}^0 = 0.005$, and $\rho^0 = 0$. For the DGP C, I set $\bar{R}^0 = 1$, where $\bar{\delta}^0 = 0.4$ and $\rho^0 = 0$.

As discussed in Section 3.3, a suitable $\bar{R}$ can be set by using a grid of values. A good $\hat{\bar{R}}$ will be reached if an increase in $\bar{R}$ does not significantly change the estimates. To see how this approach performs, I set $\bar{R}$ to the 90th percentile of y, then to the integer after the 90th percentile. If the estimates do not change, this would suggest that the practitioner could reach a good $\bar{R}$ in practice.

Moreover, I also compute the NPL estimator for $\hat{\bar{R}} = 1$ and $\rho = 0$. In this case, I expect the counting variable model to approximately replicate the bias of the SART estimator.

Figure 2 presents the histogram of an example of the simulated data. Because I defined $(\delta_r)_r$ to be a decreasing sequence in r, the higher the $\bar{R}^0$, the longer the tail of the counting variable. This decrease in δ_r is also observed in my application on real data. This suggests that the bias of the SART model would be higher on variables having longer tails, such as those obtained from survey data sets.

My assessment in conservative by setting $\hat{\bar{R}}$ to the 90th percentile of y. In practice, $\hat{\bar{R}}$ can be set higher. The 90th percentile corresponds to 8 in the case of DGP A and B and to 7 in the case of DGP C. The NPL estimator is expected to perform very well for DGP A and C because $\hat{\bar{R}} > \bar{R}^0$. In the case of DGP B, $\hat{\bar{R}} < \bar{R}^0$. This leads to a misspecification issue. However, the resulting bias would be negligible because a small proportion of agents are affected by this issue.

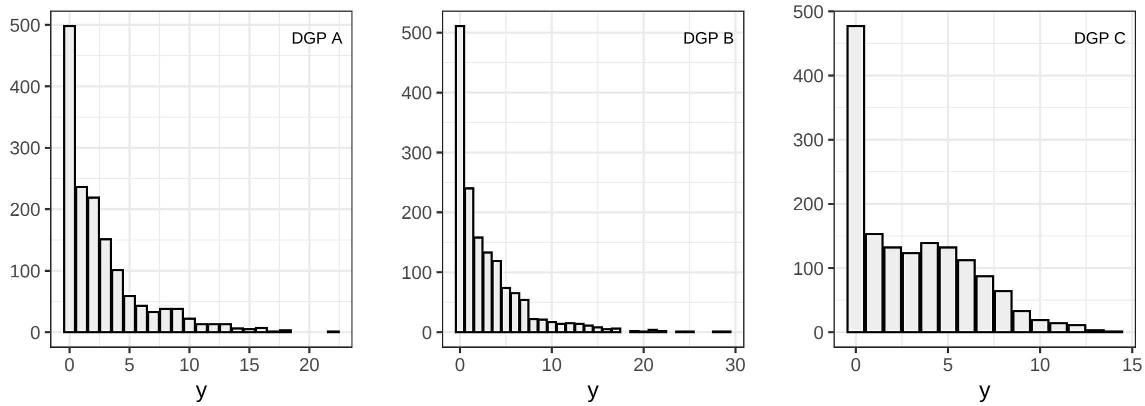

Figure 2: Simulated data using the count data model with social interactions

The simulation results (for 1,000 replications) are presented in Table 1.[25] Note that one cannot directly interpret the parameters of the counting variable model, nor can one compare these parameters to those of the SART model. Table 1 reports the marginal effect (ME) of each variable.[26] The first column presents the true ME. The notation $\delta(.)$ denotes the ME of the variable in parentheses. Model 1 corresponds to $\hat{\bar{R}} = 1$. Models 2 and 3 are cases where $\hat{\bar{R}}$ is set to the 90th percentile of y and the integer after that 90th percentile. For each model, I consider the case where ρ is flexible and the case where $\rho = 0$. Model 4 is the Tobit model.

The results for DGP A and C show that the NPL estimator performs well when $\hat{\bar{R}} \geqslant \bar{R}^0$. There is no substantial difference between Models 2 and 3. This is consistent with the result that the NPL estimator converges to θ^0 as soon as $\hat{\bar{R}} > \bar{R}^0$. The parameter ρ seems not to play an important role when $\hat{\bar{R}}$ is set high. Indeed, the estimator is slightly underestimated when ρ is set to zero and the true ρ^0 is nonzero. However, the bias is very small because this misspecification only affects a small proportion of the individuals if $\hat{\bar{R}}$ is set over the 90th percentile of y. In the case of DGP B, although $\hat{\bar{R}}$ is less than the true value, the bias of the NPL estimator is negligible. This result is interesting because $\hat{\bar{R}}$ can be set higher in practice and the bias would be smaller.

Furthermore, the Tobit model overestimates peer influence in the case of DGP A and B. As expected, the resulting bias is approximately replicated by the NPL estimator when $\hat{\bar{R}} = 1$ and $\rho = 0$. In the case of DGP C, the Tobit model performs as well as the counting variable model. However, this case is too constraining and is not likely to occur with real data.

[25] These results can be replicated using my R package – CDatanet – and the Monte Carlo replication file available at https://github.com/ahoundetoungan/CDatanet.

[26] 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 OA S.4.

Table 1: Monte Carlo simulations

True marginal	**Model 1**				**Model 2**				**Model 3**				**Model 4**	
	flexible ρ		$\rho = 0$		flexible ρ		$\rho = 0$		flexible ρ		$\rho = 0$		Tobit	
effects	Mean	Sd.	Mean	Sd.	Mean	Sd.	Mean	Sd.	Mean	Sd.	Mean	Sd.	Mean	Sd.
DGP A, $\bar{R}^0 = 4$	$\hat{\hat{R}} = 1$				$\hat{\hat{R}} = 8$				$\hat{\hat{R}} = 9$					
$\delta(\bar{y}^e) = 0.316$	0.358	0.080	0.347	0.058	0.311	0.042	0.308	0.041	0.310	0.042	0.308	0.040	0.345	0.056
$\delta(x_1) = 1.899$	1.861	0.085	1.861	0.085	1.898	0.087	1.898	0.087	1.898	0.087	1.898	0.087	1.870	0.085
$\delta(x_2) = -1.519$	−1.445	0.064	−1.445	0.064	−1.518	0.070	−1.518	0.070	−1.517	0.070	−1.518	0.070	−1.458	0.064
$\delta(\bar{x}_1) = 0.633$	0.561	0.225	0.587	0.179	0.649	0.150	0.653	0.148	0.651	0.149	0.655	0.147	0.592	0.177
$\delta(\bar{x}_2) = -1.139$	−1.113	0.107	−1.121	0.098	−1.143	0.091	−1.144	0.091	−1.144	0.091	−1.145	0.090	−1.124	0.098
Data B, $\bar{R}^0 = 13$	$\hat{\hat{R}} = 1$				$\hat{\hat{R}} = 8$				$\hat{\hat{R}} = 9$					
$\delta(\bar{y}^e) = 0.329$	0.373	0.065	0.373	0.067	0.321	0.038	0.321	0.039	0.317	0.037	0.317	0.037	0.371	0.064
$\delta(x_1) = 1.972$	1.926	0.095	1.926	0.095	1.971	0.097	1.971	0.097	1.970	0.097	1.970	0.097	1.934	0.094
$\delta(x_2) = -1.577$	−1.490	0.071	−1.490	0.071	−1.576	0.078	−1.576	0.078	−1.576	0.078	−1.576	0.078	−1.502	0.070
$\delta(\bar{x}_1) = 0.657$	0.596	0.195	0.597	0.198	0.680	0.149	0.679	0.150	0.689	0.146	0.687	0.146	0.600	0.193
$\delta(\bar{x}_2) = -1.183$	−1.165	0.107	−1.165	0.108	−1.191	0.096	−1.191	0.096	−1.194	0.095	−1.193	0.095	−1.167	0.106
Data C, $\bar{R}^0 = 1$	$\hat{\hat{R}} = 1$				$\hat{\hat{R}} = 7$				$\hat{\hat{R}} = 8$					
$\delta(\bar{y}^e) = 0.274$	0.272	0.049	0.272	0.049	0.272	0.049	0.272	0.049	0.272	0.049	0.272	0.049	0.270	0.049
$\delta(x_1) = 1.642$	1.640	0.049	1.641	0.049	1.641	0.049	1.641	0.049	1.641	0.049	1.641	0.049	1.656	0.050
$\delta(x_2) = -1.313$	−1.311	0.039	−1.312	0.039	−1.312	0.039	−1.312	0.039	−1.312	0.039	−1.312	0.039	−1.331	0.040
$\delta(\bar{x}_1) = 0.547$	0.551	0.145	0.551	0.145	0.551	0.145	0.551	0.145	0.551	0.145	0.551	0.145	0.560	0.147
$\delta(\bar{x}_2) = -0.985$	−0.985	0.073	−0.985	0.073	−0.986	0.074	−0.985	0.074	−0.985	0.074	−0.985	0.074	−0.992	0.074

This table presents the Monte Carlo simulation results. Several specifications are used for each DGP. Model 1 is estimated by setting $\hat{\hat{R}}$ to one. Model 2 is estimated by setting $\hat{\hat{R}}$ to the 90th percentile, and Model 3 has $\hat{\hat{R}}$ set to the integer after the 90th percentile. For each of those models, I consider the specification in which ρ is left flexible and another specification in which ρ to set to 0. Model 4 is the Tobit model, which indirectly sets $\hat{\hat{R}}$ to one. I perform 1,000 simulations and report the mean of the estimates and the standard deviation (Sd.)

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 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.

I remove self-friendships and friendships between two students from different schools. Furthermore, an important number of listed friend identifiers are missing or associated with "error codes."[27] 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). The network matrix $\mathbf{G}$ used is the row-normalized adjacency matrix.

The studied counting variable is the number of extracurricular activities in which students are enrolled. 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 study whether social interactions influence student participation in extracurricular activities. I control for 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. I also control for contextual variables. Table B.1 provides the data summary and Figure B.1 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 distribution has a long tail. I expect the SART model to overestimate the marginal peer effect.

[27] Numerous papers have developed methods for estimating peer effects using partial network data (e.g., Boucher and Houndetoungan, 2022). To focus on the main purpose of this paper, I do not address that issue here.

5.2 Empirical results

I first determine the suitable value of the parameter $\bar{R}$. The counting variable model is estimated using a grid of values of $\bar{R}$ starting from one. I find that the marginal peer effects and the log-likelihood of the model are stable when $\bar{R}$ reaches 12. The estimates of the peer effects and the log-likelihood are presented in Table 2. There is no significant difference between the results for $\hat{\bar{R}} = 12$ and $\hat{\bar{R}} = 13$.[28] I then set $\hat{\bar{R}} = 12$ for the rest of the empirical analysis.[29] I also compare the specification with flexible ρ to that in which ρ is set to zero and do not see a significant difference. The results of the empirical study correspond to $\rho = 0$.

I present the estimate of the marginal peer influence for $\hat{\bar{R}} = 1$ in the case where I do not control for school fixed effects nor for network endogeneity. As in my Monte Carlo experiment, the marginal peer effect corresponds to the estimate with the Tobit model and is more than five times higher than the result for $\hat{\bar{R}} = 12$. This result also confirms that the Tobit model is equivalent to the proposed counting variable model when $\hat{\bar{R}} = 1$ and $\rho = 0$.

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. The pseudo-log-likelihood increases by 1,515 for 119 additional explanatory variables. The likelihood ratio (LR) test confirms the importance of these school-fixed effects. However, including these school-fixed effects does not have a significant impact on the peer influence in the case of the counting variable model. For the Tobit model, the marginal peer effects decrease from 0.552 to 0.358.

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). Moreover, these personality traits are likely to increase student probability of interacting with others. This implies that network matrix $\mathbf{G}$ is potentially endogenous. I control for the endogeneity by using the approach presented in Section 3.4. The unobserved attributes are estimated using the fixed effect logit model studied by Yan et al. (2019) and a random effect probit model in which one can simulate the posterior distribution of the parameters using a Gibbs sampler (see Casella and George, 1992; Albert and Chib, 1993). I include a polynomial function of the at-

[28]Full results for general models that control for the network endogeneity are available in OA S.5.

[29]The suitable $\hat{\bar{R}}$ in this application corresponds to the 99th percentile.

tributes in the models as additional explanatory variables. The sufficient degree of the polynomial that captures the network endogeneity is five for the counting variable model and seven for the Tobit model, in the case where the attributes are estimated using the fixed effect logit approach. In the case of the random effect probit model, the degree is four for both models. Full results of the estimations are presented in OA S.5.

The increase in the log-likelihood suggests that the network is endogenous. However, this endogeneity does not influence the marginal peer effect in the case of the counting variable model. The estimate decreases from 0.086 to 0.084. The decrease is more important for the Tobit model, as the marginal peer effect is estimated at 0.246 after controlling for the network endogeneity, which still is three times higher than the marginal effects estimated using the counting variable model.

Table 2: Empirical results

(1) Models	(2) Coef.	(3) Marginal effects	(4)	(5) log (Like)
School fixed effects: No; Network endogeneity: No				
Count data ($\hat{\bar{R}} = 1$)	0.276	0.550	(0.024)	−159924
Count data ($\hat{\bar{R}} = 11$)	0.051	0.097	(0.025)	−127659
Count data ($\hat{\bar{R}} = 12$)	0.047	0.089	(0.026)	−127641
Count data ($\hat{\bar{R}} = 13$)	0.048	0.090	(0.026)	−127640
Tobit	0.681	0.552	(0.018)	−161225
School fixed effects: Yes; Network endogeneity: No				
Count data ($\hat{\bar{R}} = 12$)	0.046	0.086	(0.022)	−126126
Tobit	0.441	0.358	(0.019)	−160259
School fixed effects: Yes; Network endogeneity: Yes (Random effects)				
Count data ($\hat{\bar{R}} = 12$)	0.046	0.084	(0.020)	−125140
Tobit	0.307	0.249	(0.019)	−159627
School fixed effects: Yes; Network endogeneity: Yes (Fixed effects)				
Count data ($\hat{\bar{R}} = 12$)	0.046	0.084	(0.024)	−125332
Tobit	0.304	0.246	(0.025)	−159735

This table presents the empirical estimate of the peer effects for different models. Column (1) indicates the estimated model. Column (2) reports the peer effect coefficient. Columns (3)–(4) report the marginal peer effects on the expected outcome and their corresponding standard error in parentheses. Column (5) is the maximized log-likelihood. The sample size is 72,291 and there are 120 schools.

6 Conclusion

In this paper, I develop a social network model for counting data using a static game of incomplete information. The model bridges the gap between binary/ordered outcome models and linear-in-means models. 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 for dealing 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. Under regularity conditions, I establish the asymptotic normality of the estimator at the second stage.

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 expected number of activities in which students are enrolled by 0.08. The estimate of this effect with the Tobit model is three times higher. Finally, I provide an easy-to-use R package that implements all the methods used in this paper.[30]

The findings of this paper raise an important question. Because the assumption of a quadratic cost function leads to inconsistent estimates 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.

A Appendix: Proofs

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.*

Proof. 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, $h(r_0) \geqslant \max\{h(r_0 - 1),\ h(r_0 + 1)\}$.

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

Assume now that $h(r_0) \geqslant \max\{h(r_0-1),\ h(r_0+1)\}$.

As pointed out by Murota (1998), a discrete function is concave iff 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 construct $\bar{h}$ 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. □

Proof of Proposition 2.1

The expected outcome is $U_i^e\,(y_i) = \psi_i y_i - c(y_i) - \dfrac{\lambda}{2}\,\mathbb{E}_{\bar{y}_i}\left[\,(y_i - \bar{y}_i)^2\,\right] + e_i(y_i)$.

Under Assumptions 2.1–2.3, $U_i^e(.)$ is strictly concave. In addition, since $U_i^e\,(y_i)$ tends to $-\infty$ as y_i grows to ∞, $U_i^e(.)$ has a global maximum, which is necessarily reached at a single point almost surely (a.s.).[31] Indeed, if the global maximum were reached at two points r and r', where $r \neq r'$, then $U_i^e\,(r) = U_i^e\,(r')$, which implies $\varepsilon_i = \dfrac{c(r)-c(r')}{r-r'} + \dfrac{\lambda}{2}(r+r') - \psi_i - \lambda\,\mathbb{E}(\bar{y}_i)$. This condition has zero probability because ε_i is continuous and the quantity on the right side of the equation is deterministic. As a result, $U_i^e(.)$ has one maximizer a.s.

The second part of Proposition 2.1 is given by Lemma A.1. As $U_i^e(.)$ is also concave, the global maximum is reached at r_0 iff $U_i(r_0) \geqslant \max\{U_i(r_0-1),\ U_i(r_0+1)\}$.

A.2 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 $\ell_i\,(\mathbf{y}^e) = \sum_{r=1}^{\infty} F_{\varepsilon|\mathcal{I}}\,(\lambda\bar{y}_i^e + \psi_i - a_r)$ 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, $p_{ir} = F_{\varepsilon|\mathcal{I}}\,(\lambda\bar{y}_i^e + \psi_i - a_r) - F_{\varepsilon|\mathcal{I}}\,(\lambda\bar{y}_i^e + \psi_i - a_{r+1})$.

Let $S_1 = \sum_{r=0}^{\infty} rF_{\varepsilon|\mathcal{I}}\,(\lambda\bar{y}_i^e + \psi_i - a_r)$ and $S_2 = \sum_{r=0}^{\infty} rF_{\varepsilon|\mathcal{I}}\,(\lambda\bar{y}_i^e + \psi_i - a_{r+1})$. Let $x < 0$ with $|x|$ being sufficiently large. By Assumption 2.3, $f_{\varepsilon|\mathcal{I}}(x) = o\left(|x|^{-\kappa}\right)$ at ∞ for $\kappa > 3$. Then, $F_{\varepsilon|\mathcal{I}}(x) = O(|x|^{-(\kappa-1)})$ at $-\infty$ and $F_{\varepsilon|\mathcal{I}}(x) = o(|x|^{-(\kappa-\upsilon-1)})$ at $-\infty$ for some υ positive and sufficiently small. By Lemma S.1 (see OA S.1), $S_1 < \infty$ and $S_2 < \infty$. Thus, $y_i^e = S_1 - S_2 < \infty$. In addition,

$y_i^e = \sum_{r=0}^{\infty} rF_{\varepsilon|\mathcal{I}}\,(\lambda\bar{y}_i^e + \psi_i - a_r) - \sum_{r=0}^{\infty}(r+1)F_{\varepsilon|\mathcal{I}}\,(\lambda\bar{y}_i^e + \psi_i - a_{r+1}) + \sum_{r=0}^{\infty} F_{\varepsilon|\mathcal{I}}\,(\lambda\bar{y}_i^e + \psi_i - a_{r+1})\,,$

$y_i^e = \sum_{r=1}^{\infty} rF_{\varepsilon|\mathcal{I}}\,(\lambda\bar{y}_i^e + \psi_i - a_r) - \sum_{r=1}^{\infty} rF_{\varepsilon|\mathcal{I}}\,(\lambda\bar{y}_i^e + \psi_i - a_r) + \sum_{r=0}^{\infty} F_{\varepsilon|\mathcal{I}}\,(\lambda\bar{y}_i^e + \psi_i - a_{r+1})\,,$

$y_i^e = \sum_{r=1}^{\infty} F_{\varepsilon|\mathcal{I}}\,(\lambda\bar{y}_i^e + \psi_i - a_r) = \ell_i\,(\mathbf{y}^{*e})$. Hence, $\mathbf{y}^e = \mathbf{L}\,(\mathbf{y}^e)$.

A.3 Proof of Proposition 2.3

If $\mathbf{p}^*$ is a rational belief system, Proposition 2.2 states that its associated expected outcome $\mathbf{y}^{e*}$ verifies $\mathbf{y}^{e*} = \mathbf{L}\,(\mathbf{y}^{e*})$. To prove the uniqueness, it is sufficient to prove that $\mathbf{L}$ is contracting; that is,

[31] The strict concavity does not imply a unique maximizer for a discrete function. There may exist two maximizers.

$\forall\ \mathbf{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$, $\|\partial \mathbf{L}(\mathbf{u})/\partial \mathbf{u}'\|_\infty \leqslant \bar{\kappa}_c$ for some $\bar{\kappa}_c < 1$ not depending on $\mathbf{u}$.
For all i and j, $\dfrac{\partial \ell_i(\mathbf{u})}{\partial u_j} = \lambda g_{ij} f_i^*$, where $f_i^* = \sum_{r=1}^{\infty} f_{\varepsilon|\mathcal{I}}(\lambda \mathbf{g}_i \mathbf{u} + \psi_i - a_r)$. Thus, the (i,j)-th element of the matrix $\partial \mathbf{L}(\mathbf{u})/\partial \mathbf{u}'$ is $g_{ij} f_j^*$. It follows that $\|\partial \mathbf{L}(\mathbf{u})/\partial \mathbf{u}'\|_\infty = \lambda \max_i \{f_i^* \sum_{j=1}^{n} g_{ij}\} \leqslant \lambda (\max_i f_i^*) \|\mathbf{G}\|_\infty$.
Besides, $f_i^* = \sum_{r=1}^{\infty} f_{\varepsilon|\mathcal{I}}(\lambda \mathbf{g}_i \mathbf{u} + \psi_i - a_r) \leqslant \max_{u \in \mathbb{R}} \sum_{k=1}^{\infty} f_{\varepsilon|\mathcal{I}}(u - a_r) = \dfrac{1}{B_c}$. Thus, $\|\partial \mathbf{L}(\mathbf{u})/\partial \mathbf{u}'\|_\infty \leqslant \dfrac{\lambda \|\mathbf{G}\|_\infty}{B_c} < 1$ by Assumption 2.4. 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.4 Proof of Proposition 3.2

Let n_m be the number of individuals in the m-th subgroup and $\mathbf{G}_m$ be the $n_m \times n_m$ network matrix of the m-th subgroup. Let $\chi_i = \{\mathbf{x}_i, \tilde{\mathbf{g}}_i\}$, where $\tilde{\mathbf{g}}_i$ is the i-th row of $\mathbf{G}_{s(i)}$ excluding $\mathbf{g}_{ii}$. Let $\boldsymbol{\Theta}(\bar{R})$ be the space of $\boldsymbol{\theta}$ for $\bar{R} \geqslant 1$ and $\boldsymbol{\theta}^0$ be the true value of $\boldsymbol{\theta}$. Let also $\mathbf{y}_{\boldsymbol{\chi}}^{e0} \in \mathbb{R}^n$, such that $\mathbf{y}_{\boldsymbol{\chi}}^{e0} = \mathbf{L}(\boldsymbol{\theta}^0, \mathbf{y}_{\boldsymbol{\chi}}^{e0})$. I adapt the regulatory conditions of the NPL estimator set by Aguirregabiria and Mira (2007), henceforth AM07, to my framework.

Assumption A.1. supp$(\mathbf{x}_i) \subset \mathbb{X}$ *for all i, where $\mathbb{X}$ is a compact subset of* $\mathbb{R}^K$.

As discussed in Section 3.3, AM07 show the consistency of the NPL estimator by assuming that the expected outcome is bounded. A sufficient condition for this assumption to hold is that supp$(\mathbf{x}_i)$ is bounded. This assumption is generally verified in empirical applications. Given Assumption A.1, y_i^e is bounded. Let $\overset{o}{y} = \sup_{n \in \mathbb{N}} \max_{i \leqslant n} y_i^e$.
For any $\bar{R} \geqslant 1$, I define the following notations:
$\mathcal{L}_0(\boldsymbol{\theta}, \mathbf{y}^e) \equiv \mathbb{E}(\mathcal{L}_n(\boldsymbol{\theta}, \mathbf{y}^e)|\boldsymbol{\chi}_n)$, $\tilde{\boldsymbol{\theta}}(\mathbf{y}^e, \bar{R}) \equiv \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}(\bar{R})} \mathcal{L}_0(\boldsymbol{\theta}, \mathbf{y}^e)$, $\boldsymbol{\phi}_0(\mathbf{y}^e, \bar{R}) \equiv \mathbf{L}(\tilde{\boldsymbol{\theta}}(\mathbf{y}^e, \bar{R}), \mathbf{y}^e)$, and $\mathcal{A}_0(\bar{R}) \equiv \{(\boldsymbol{\theta}, \mathbf{y}^e) \in \boldsymbol{\Theta}(\bar{R}) \times [0, \overset{o}{y}]^n$, such that $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_0(\mathbf{y}^e, \bar{R})$ and $\mathbf{y}^e = \boldsymbol{\phi}_0(\mathbf{y}^e, \bar{R})\}$. Note that $\mathbf{L}$, $\mathcal{L}_0(\boldsymbol{\theta}, \mathbf{y}^e)$, $\boldsymbol{\phi}_0(\mathbf{y}^e, \bar{R})$, and $\mathcal{A}_0(\bar{R})$ depend on $\boldsymbol{\chi}_n$. As $\mathbb{L}$ is continuous, $\mathcal{A}_0(\bar{R})$ is a compact subset of $\boldsymbol{\Theta}(\bar{R}) \times [0, \overset{o}{y}]^n$.

Assumption A.2. *For any finite $\bar{R} \geqslant 1$, $\boldsymbol{\Theta}(\bar{R})$ is a compact subset of* $\mathbb{R}^{\dim(\boldsymbol{\theta})}$.

Assumptions A.1–A.2 imply the following two results.

Result A.1 (Uniform convergence of $\mathcal{L}_n$ to $\mathcal{L}_0$)**.** *By the law of large numbers, for any $\bar{R} \geqslant 1$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}(\bar{R})$, $\mathcal{L}_n(\boldsymbol{\theta}, \mathbf{y}^e) - \mathcal{L}_0(\boldsymbol{\theta}, \mathbf{y}^e)$ converges in probability to zero.*[32] *Moreover, stochastic equicontinuity in $(\boldsymbol{\theta}, \mathbf{y}^e)$ follows, as $\mathcal{L}_n(\boldsymbol{\theta}, \mathbf{y}^e) - \mathcal{L}_0(\boldsymbol{\theta}, \mathbf{y}^e)$ is continuously differentiable in $(\boldsymbol{\theta}, \mathbf{y}^e)$ on the compact $\boldsymbol{\Theta}(\bar{R}) \times [0, \overset{o}{y}]^n$, and the derivative is bounded. Finally, $\mathcal{L}_n(\boldsymbol{\theta}, \mathbf{y}^e)$ uniformly converges to $\mathcal{L}_0(\boldsymbol{\theta}, \mathbf{y}^e)$ in $(\boldsymbol{\theta}, \mathbf{y}^e)$ (see Newey and McFadden, 1994, Lemma 2.8)*

[32] This can be shown using Chebyshev's inequality. Indeed, $\mathbb{V}(\mathcal{L}_n(\boldsymbol{\theta}, \mathbf{y}^e)|\boldsymbol{\chi}_n)$ converges to zero as n grows to infinity.

Result A.2 ($\mathcal{L}_0$ has a unique maximizer in $\mathcal{A}_0(\bar{R})$)**.** *As $\mathcal{L}_0$ is bounded on the compact $\mathcal{A}_0(\bar{R})$, there exits $(\check{\boldsymbol{\theta}}_n(\bar{R}), \check{\mathbf{y}}_n^e(\bar{R})) \in \mathcal{A}_0(\bar{R})$, depending on $\boldsymbol{\chi}_n$, such that $\mathcal{L}_0(\check{\boldsymbol{\theta}}_n(\bar{R}), \check{\mathbf{y}}_n^e(\bar{R}))$ is the global maximum of $\mathcal{L}_0$. Importantly, $(\check{\boldsymbol{\theta}}_n(\bar{R}), \check{\mathbf{y}}_n^e(\bar{R}))$ uniquely maximizes $\mathcal{L}_0$ on $\mathcal{A}_0(\bar{R})$. This is ensured by the parameter identification (Proposition 3.1).*

The following assumptions also adapt the regulatory conditions in AM07.

Assumption A.3. *For any $\bar{R} \geqslant 1$, $\check{\boldsymbol{\theta}}_n(\bar{R}) \in \operatorname{int}(\boldsymbol{\Theta}(\bar{R}))$.*

Assumption A.4. *Either $\mathcal{A}_0(\bar{R}) = \{(\check{\boldsymbol{\theta}}_n(\bar{R}), \check{\mathbf{y}}_n^e(\bar{R}))\}$ or there is an open ball around $(\check{\boldsymbol{\theta}}_n(\bar{R}), \check{\mathbf{y}}_n^e(\bar{R}))$ that does not contain any other element of $\mathcal{A}_0(\bar{R})$.*

Assumption A.5. *$\phi_0(\mathbf{y}^e, \bar{R}) - \mathbf{y}^e$ has a nonsingular Jacobian matrix at $\check{\mathbf{y}}_n^e(\bar{R})$.*

Assumptions A.3–A.5 are set with respect to the maximizer $(\check{\boldsymbol{\theta}}_n(\bar{R}), \check{\mathbf{y}}_n^e(\bar{R}))$ and not to the true value $(\boldsymbol{\theta}^0, \mathbf{y}_{\boldsymbol{\chi}}^{e0})$ as in AM07. Indeed, the NPL estimator may not converge to $\boldsymbol{\theta}^0$ when $\hat{\bar{R}} < \bar{R}^0$. Moreover, I set the assumptions conditionally on $\boldsymbol{\chi}_n$ because y_i^e are not independent when one does not condition on $\boldsymbol{\chi}_n$. My asymptotic analysis is then conditional on the data as in Lin and Xu (2017). As argued by AM07, Assumptions A.4–A.5 are not necessary when $\mathcal{A}_0(\bar{R})$ contains a single element. This seems to be the case, in general, when $\mathbf{L}(\mathbf{y}^e, \bar{R})$ is a contracting mapping in $\mathbf{y}^e$. Although there is no formal proof for this statement, simulations suggest that the NPL algorithm converges to the same solution regardless of the initial candidate for $\mathbf{y}^e$ (see also Lin and Xu, 2017; Liu, 2019).
Under Assumptions A.1–A.5, $\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}})$ converges in probability to $\check{\boldsymbol{\theta}}_n(\hat{\bar{R}})$. This is a direct implication of Proposition 2 in AM07. Furthermore, by Gibbs' inequality (see Isihara, 2013, p. 37), $\mathcal{L}_0(\boldsymbol{\theta}^0, \mathbf{y}_{\boldsymbol{\chi}}^{e0})$ is the global maximum of $\mathcal{L}_0$. Since $\boldsymbol{\theta}^0 \in \boldsymbol{\Theta}(\hat{\bar{R}})$ for any $\hat{\bar{R}} \geqslant \bar{R}^0$ (by redefinition of $\boldsymbol{\theta}^0$ in higher dimensional space), this implies that $(\boldsymbol{\theta}^0, \mathbf{y}_{\boldsymbol{\chi}}^{e0}) \in \mathcal{A}(\hat{\bar{R}})$. Hence, $(\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \check{\mathbf{y}}_n^e(\hat{\bar{R}})) = (\boldsymbol{\theta}^0, \mathbf{y}_{\boldsymbol{\chi}}^{e0})$ by Result A.2. Therefore $\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}) = \boldsymbol{\theta}^0$ if $\hat{\bar{R}} \geqslant \bar{R}^0$.

I now show the asymptotic normality.
Let $\mathcal{L}_{n,i}(\boldsymbol{\theta}, \mathbf{y}^e) = \sum_{r=0}^{\infty} d_{ir} \log\left(\Phi\left(\lambda \bar{y}_i^e + \mathbf{z}_i'\boldsymbol{\Gamma} - a_r\right) - \Phi\left(\lambda \bar{y}_i^e + \mathbf{z}_i'\boldsymbol{\Gamma} - a_{r+1}\right)\right)$. The pseudo-likelihood is $\mathcal{L}_n(\boldsymbol{\theta}, \mathbf{y}^e) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}_{n,i}$. The NPL estimator verifies $\nabla_{\boldsymbol{\theta}} \mathcal{L}_n(\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}})) = 0$ and $\hat{\mathbf{y}}_n^e(\hat{\bar{R}}) = \mathbf{L}(\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}}))$, where $\nabla_u f$ denotes the derivative function of f with respect to u.
Let us apply the mean value theorem to $\nabla_{\boldsymbol{\theta}} \mathcal{L}_n(\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}}))$ between $\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}})$ and $\check{\boldsymbol{\theta}}_n(\hat{\bar{R}})$. For some point $\dot{\boldsymbol{\theta}}_n(\hat{\bar{R}})$ between $\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}})$ and $\check{\boldsymbol{\theta}}_n(\hat{\bar{R}})$, I have

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}}) - \check{\boldsymbol{\theta}}_n(\hat{\bar{R}})) = -\left(\mathbf{H}_{1,n} + \mathbf{H}_{2,n}\right)^{-1} \sqrt{n}\nabla_{\boldsymbol{\theta}} \mathcal{L}_n(\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}})),$$

where $\mathbf{H}_{1,n} := \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} \mathcal{L}_n(\dot{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \dot{\mathbf{y}}_n^e(\hat{\bar{R}}))$,
$\mathbf{H}_{2,n} := \nabla_{\boldsymbol{\theta}\mathbf{y}^{e\prime}} \mathcal{L}_n(\dot{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \dot{\mathbf{y}}_n^e(\hat{\bar{R}})) \nabla_{\boldsymbol{\theta}'} \dot{\mathbf{y}}^{e\prime}(\hat{\bar{R}})$, $\dot{\mathbf{y}}_n^e(\hat{\bar{R}}) = \mathbf{L}(\dot{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \dot{\mathbf{y}}_n^e(\hat{\bar{R}}))$, and $\nabla_{uv} f$ denotes the derivative function of $\nabla_u f$ with respect to v. I now focus on the term $\sqrt{n}\nabla_{\boldsymbol{\theta}} \mathcal{L}_n(\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}}))$.

I have $\sqrt{n}\nabla_{\boldsymbol{\theta}}\mathcal{L}_n(\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}})) = \dfrac{\sum_{i=1}^n \nabla_{\boldsymbol{\theta}}\mathcal{L}_{n,i}(\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}}))}{\sqrt{n}}$. The terms $\nabla_{\boldsymbol{\theta}}\mathcal{L}_{n,i}(\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}}))$ are independent across i conditionally on $\boldsymbol{\chi}_n$. Moreover, $\mathbb{V}(\nabla_{\boldsymbol{\theta}}\mathcal{L}_{n,i}(\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}}))|\boldsymbol{\chi}_n)$ is bounded, and $\sum_{i=1}^n \mathbb{E}(\nabla_{\boldsymbol{\theta}}\mathcal{L}_{n,i}(\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}}))|\boldsymbol{\chi}_n) = 0$. The latter holds because $\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}(\hat{\bar{R}})} \mathcal{L}_0(\boldsymbol{\theta}, \hat{\mathbf{y}}_n^e(\hat{\bar{R}}))$. I can then apply the central limit theorem (CLT) to $\sqrt{n}\nabla_{\boldsymbol{\theta}}\mathcal{L}_n(\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}}))$ conditionally on $\boldsymbol{\chi}_n$.[33] Lindeberg's condition holds because $\nabla_{\boldsymbol{\theta}}\mathcal{L}_{n,i}(\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}}))$ is bounded.

Let $\boldsymbol{\Sigma}_n := \dfrac{\sum_{i=1}^n \mathbb{V}(\nabla_{\boldsymbol{\theta}}\mathcal{L}_{n,i}(\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \hat{\mathbf{y}}_n^e(\hat{\bar{R}}))|\boldsymbol{\chi}_n)}{n}$. Assuming that $\boldsymbol{\Sigma}_n \to \boldsymbol{\Sigma}_0$, $\mathbf{H}_{1,n} \to \mathbf{H}_{1,0}$, and $\mathbf{H}_{2,n} \to \mathbf{H}_{2,0}$ as n grows to infinity, I get

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}}) - \check{\boldsymbol{\theta}}_n(\hat{\bar{R}})) \xrightarrow{d} \mathcal{N}\left(0, (\mathbf{H}_{1,0} + \mathbf{H}_{2,0})^{-1}\boldsymbol{\Sigma}_0(\mathbf{H}_{1,0}' + \mathbf{H}_{2,0}')^{-1}\right).$$

I give the analytical expressions of $\boldsymbol{\Sigma}_0$, $\mathbf{H}_{1,0}$, and $\mathbf{H}_{2,0}$ in the Supplementary Material S.2 for $\hat{\bar{R}} \geqslant \bar{R}^0$.

B Data Summary

This section summarizes the data (see Table B.1). 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 B.1: 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

[33] See an example in Theorem 23.4 of Van der Vaart (2000).

varies from 0 to 33. However, most students declare that they participate in fewer than 10 extracurricular activities (see Figure B.1).

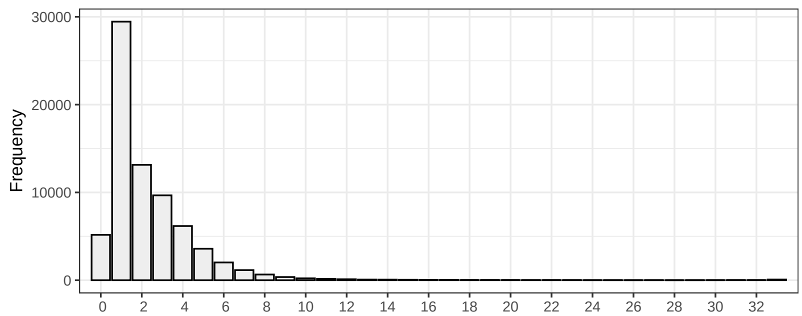

Figure B.1: Distribution of the number of extracurricular activities

C Supplementary Data

Supplementary material related to this paper can be found online at https://ahoundetoungan.com/files/Papers/CDMSIRE_SM.pdf

Acknowledgments

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

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.

References

Ackerberg, D., X. Chen, and J. Hahn (2012): "A practical asymptotic variance estimator for two-step semiparametric estimators," *Review of Economics and Statistics*, 94, 481–498.

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.

Auerbach, E. (2022): “Identification and estimation of a partially linear regression model using network data,” *Econometrica*, 90, 347–365.

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.

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 B. Fortin (2016): “Some challenges in the empirics of the effects of networks,” *The Oxford Handbook on the Economics of Networks*, 277–302.

Boucher, V. and A. Houndetoungan (2022): *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.

Casella, G. and E. I. George (1992): “Explaining the Gibbs sampler,” *The American Statistician*, 46, 167–174.

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.

De Paula, Á. (2020): "Econometric models of network formation," *Annual Review of Economics*, 12, 775–799.

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.

Fernández-Val, I. (2009): "Fixed effects estimation of structural parameters and marginal effects in panel probit models," *Journal of Econometrics*, 150, 71–85.

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.

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.

Isihara, A. (2013): *Statistical physics*, Academic Press.

Johnsson, I. and H. R. Moon (2021): "Estimation of peer effects in endogenous social networks: control function approach," *Review of Economics and Statistics*, 103, 328–345.

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.

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.

Li, J. and L.-f. Lee (2009): "Binary choice under social interactions: an empirical study with and without subjective data on expectations," *Journal of Applied Econometrics*, 24, 257–281.

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., X. Tang, and N. N. Yu (2021): "Uncovering heterogeneous social effects in binary choices," *Journal of Econometrics*, 222, 959–973.

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. (1988): "Identification of binary response models," *Journal of the American statistical Association*, 83, 729–738.

——— (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.

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

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

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.

Van der Vaart, A. W. (2000): *Asymptotic statistics*, vol. 3, Cambridge university press.

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.

Yan, T., B. Jiang, S. E. Fienberg, and C. Leng (2019): "Statistical inference in a directed network model with covariates," *Journal of the American Statistical Association*, 114, 857–868.

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