Count Data Models with Social Interactions under Rational Expectations

Elysée Aristide Houndetoungan*

Thema, Cy Cergy Paris Université

July 18, 2023

Abstract

This paper proposes a peer effect model for counting variables using a game of incomplete information. I show that the game has a unique equilibrium under standard conditions. I also show that the identification argument in Bramoullé et al. (2009) extends to nonlinear models, particularly to the model of this paper. The parameters are estimated using the Nested Partial Likelihood (NPL) approach, controlling for network endogeneity. I demonstrate that the linear-in-means/Tobit models with a counting outcome are particular cases of my model. However, by ignoring the counting nature of the outcome, these models can lead to inconsistent estimators. I use the model to evaluate peer effects on students' participation in extracurricular activities. I find that a one-unit increase in the expected number of activities in which a student's friends are enrolled yields an increase in the expected number of activities in which the student is enrolled by 0.08. This point estimate is three times higher with the Tobit model.

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 cover, among others, models for limited dependent variables, including binary (e.g., Brock and Durlauf, 2001; Aradillas-Lopez, 2010; 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. Identification of the model parameters relies on a general analysis, which can also be applied to other nonlinear models such as binary and ordered response models. I show that the 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 also 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 model (Xu and Lee, 2015b), can lead 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). The assumption of incomplete information is extensively considered in the

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

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

literature on peer effect models for discrete outcomes as it implies a unique game equilibrium under standard conditions (e.g., Brock and Durlauf, 2001; Bajari et al., 2010; Lee et al., 2014; Yang and Lee, 2017; Boucher and Bramoullé, 2020). Individuals in the game interact through a directed network, simultaneously choose their strategy, and receive a payoff whose expectation depends on their belief about the choice of their peers. Unlike linear models that assume a linear-quadratic payoff, the counting nature of the outcome allows for a more flexible payoff. Note that the assumption of a linear-quadratic payoff is generally made in the linear-in-means model as 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 can yield inconsistent estimates of peer effects.

I establish sufficient conditions for the game to have a unique Bayesian Nash Equilibrium (BNE). However, the econometric specification of the model raises an identification challenge. Parameter identification is generally achieved by setting a rank condition on the design matrix. In rational expectation models, the design matrix depends on the expected outcome, which is an unobserved variable to the practitioner. Consequently, it is difficult to empirically verify the rank condition (see Lee, 2004; Yang and Lee, 2017). I present an identification analysis with easily verifiable conditions. The main condition for identifying the parameters of my model is that the network includes friends of friends who are not friends. While this condition has so far been used for identifying parameters in linear-in-means models (Bramoullé et al., 2009), I demonstrate its applicability in nonlinear models as well. Importantly, this result holds true for various nonlinear specifications, including binary and ordered response models, regardless of whether information in the game is incomplete or not. It provides a practical alternative way to achieve parameter identification in nonlinear models, often based on conditions 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 approach to the case where the network is endogenous because it depends on unobserved individual attributes, which can also explain the game 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 the unobserved attributes using a dyadic linking model. In the second stage, I control for the estimated 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 SAR Tobit 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.

Related Literature

This paper contributes to the literature on social interaction models for limited dependent variables by dealing with counting outcomes. The existing models deal with binary (e.g., Brock and Durlauf, 2001; Soetevent and Kooreman, 2007; Aradillas-Lopez, 2010; 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; Aradillas-Lopez and Rosen, 2022), 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 analysis, which can also be applied to many other nonlinear models.

A relevant study in my framework is that of Aradillas-Lopez and Rosen (2022) who examined a game with complete information for ordered responses in a potentially unbounded space. The unbounded nature of the strategy space may eventually allow for dealing with count data. However, my study differs from theirs in many ways. Aradillas-Lopez and Rosen (2022) consider a finite number of players observed across a large number of environments (markets), whereas I assume a large number of players observed only once. My asymptotic analysis is based on the number of players rather than the number of environments in which players are observed. Moreover, they consider a game with complete information and emphasize the ordinal nature of the outcome. The specific values of the outcome are deemed unimportant; rather, the focus lies on the relative order of these values.

Furthermore, in the literature on spatial autoregressive models for limited dependent variables, cases of count data have been studied (for a recent review, see 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 may not be endogenous peer effects.[3]

The paper contributes to the literature on peer effect models with endogenous networks. Goldsmith-

[3]For example, one way to model the dependence between each pair in the network it to use copula function, where the marginal distribution of the outcome of each individual is a Poisson distribution. It is difficult to know whether this dependence comes from endogenous peer effects, exogenous peer effects, or correlated effects (see Manski, 1993).

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. 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. 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). Agents interact through a directed network. Let $\mathbf{G} = [g_{ij}]$ be an $n \times n$ network matrix (observable by all agents and the practitioner), where g_{ij} is non-negative and captures the proximity of the individuals i and j in the network. The element g_{ij} may depend on n. I do not use the notation $g_{n,ij}$ for simplicity, and this does not create confusion. Throughout, my notations follow this simplicity. The subscript i is used instead of n and i. I define the set of peers of individual i as $\mathcal{P}_i = \{j, g_{ij} > 0\}$. By convention, no one interacts with himself/herself, that is $g_{ii} = 0 \ \forall \ i \in \mathcal{V}$.

2.1 Incomplete Information Network Game

The model is based on a game of incomplete information (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

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

where $\lambda \geqslant 0$, $\mathbf{y}_{-i} = (y_1, y_2, \dots, y_{i-1}, y_{i+1}, \dots, y_n)$, and $\bar{y}_i = \sum_{j \in \mathcal{P}_i} g_{ij} y_j$. 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 characteristics ψ_i. I assume that $\boldsymbol{\psi} = (\psi_1, \dots, \psi_n)'$ is observed by all agents.[5] $c(y_i)$ is the cost associated with the choice of y_i, which is assumed to be finite for any integer y_i.[6] 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 $\frac{\lambda}{2}\left(y_i - \bar{y}_i\right)^2$ is a social cost that increases with the gap between the agent and peers' choices. The parameter λ, called *peer effects*, captures the influence of this gap on the payoff.[7] Such a specification of the social cost implies conformist preferences. This is different from the social subpayoff assumed by Blume et al. (2015), which implies complementary preferences.

Agent's type is described by $\boldsymbol{\tau}_i := \left(e_i(r)\right)_{r \in \mathbb{N}}$, a sequence of random variables. Each agent observes their own type, i.e., 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 $\mathbf{y}_{-i}$. Let $\mathcal{I} = \{\boldsymbol{\psi}, \mathbf{G}\}$ be the common observation set of all players and $\mathcal{I}_i = \{\boldsymbol{\tau}_i, \boldsymbol{\psi}, \mathbf{G}\}$ be the information set of agent i. As agent i does not observe $\mathbf{y}_{-i}$,

[4] The assumption of incomplete information in a discrete game is interesting as it implies a unique equilibrium under weak 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).

[5] In the econometric model, $\psi_i = \alpha + \boldsymbol{x}_i'\boldsymbol{\beta} + \bar{\boldsymbol{x}}_i'\boldsymbol{\gamma}$, where $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}_i$ are vectors of observable individual-specific characteristics (control variables) and peers' average characteristics, respectively, and $\alpha \in \mathbb{R}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ are unknown parameters.

[6] 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 which y_i is unbounded.

[7] 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.

they maximize the expectation of the random payoff (1) with respect to their beliefs conditionally on $\mathcal{I}_i$. I assume that the private information $\boldsymbol{\tau}_i$ is independently and identically distributed across i conditionally on $\mathcal{I}$, and that this distribution is common knowledge to all the agents (see Brock and Durlauf, 2001; Bajari et al., 2010; Lee, 2004; Yang and Lee, 2017).[8] Thus, agents form rational expectations, i.e., their expectation of the payoff is the *true* mathematical expectation conditionally on $\mathcal{I}_i$, and can be expressed as

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

where $\mathbb{E}_{\bar{y}_i|\mathcal{I}}$ means that the expectation is taken with respect to $\bar{y}_i = \sum_{j\in\mathcal{P}_i} g_{ij}y_j$ (average outcome among i's friends) conditionally on $\mathcal{I}$. The expectation of $\bar{y}_i$ is conditionly on $\mathcal{I}$ and not on $\mathcal{I}_i$ because for any $i \neq j$, $\boldsymbol{\tau}_i$ and $\boldsymbol{\tau}_j$ are independent. Thus, the belief of agent i about y_j (friend j's outcome, where $j \neq i$) does not depend on $\boldsymbol{\tau}_i$. This expectation depends on the true distribution of y_j conditionally on $\mathcal{I}$, which is common knowledge to any agent. 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 $\mathbb{E}_{\bar{y}_i|\mathcal{I}}\left[(y_i - \bar{y}_i)^2\right]$ involves an infinite summation 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 $i \in \mathcal{V}$, $r \in \mathbb{N}^* := \mathbb{N}\backslash\{0\}$, $e_i(r) = e_i(r-1) + \varepsilon_i$, where $\varepsilon_i|\mathcal{I}$'s are independent across i 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

[8] To be precise regarding the notations, I should also include this distribution in the information sets as it is common knowledge. This is also true for various elements in the game, such as the expression of the payoff (1), which is commonly known to all agents. However, I maintain the definitions of the information sets as mentioned earlier to simplify the notations. Nonetheless, it should be noted that each player has complete knowledge about everything in the game, except for $\boldsymbol{\tau}_i$, which is private information exclusive to agent i.

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, implying that $e_i(y_i) = \varepsilon_i y_i + e_i(0)$. Specifying $e_i(y_i)$ as a linear function in y_i simplifies the econometric model.[9] As I will show later, this condition also implies that the SAR Tobit model is a particular case of the model proposed in this paper. Assumption 2.2 also sets that conditionally on $\mathcal{I}$, 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. 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. If $\rho = 0$, then the first 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). For the usual continuous distributions whose tail decays exponentially, such as the normal or logistic distribution, Assumption 2.3 is verified for any non-negative ρ.

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, $y_i^* \in \mathbb{N}$, a.s.; (ii) $U_i^e(r) > \max\{U_i^e(r-1),\ U_i^e(r+1)\}$ if and only if $r = y_i^*$.*

As $c(.)$ is convex and $e_i(y_i)$ is linear in y_i, the expected payoff function $U_i^e(.)$ is strictly concave. Moreover, since $U_i^e(y_i)$ tends to $-\infty$ as y_i grows to ∞, $U_i^e(.)$ has a global maximum. However, this does not directly imply that $U_i^e(.)$ is maximized at a single point because $U_i^e(.)$ is defined in an integer space. Using the continuity of the distribution of ε_i, I show that with a probability equal to one, the maximum cannot be reached at many points (see Appendix A.1 for formal proof). On the other hand,

[9] This assumption can be generalized to $e_i(r) = e_i(r-1) + \varepsilon_{ir}$, where ε_{ir} is observed by agent i only. This would imply that $e_i(y_i) = \sum_{r=0}^{y_i} \varepsilon_{ir}$, where $\varepsilon_{i0} = e_i(0)$, which is not linear in y_i. However, one must be specific regarding the correlation between ε_{ir}'s across r. As one desires $U_i^e(y_i)$ to be concave at y_i, then ε_{ir} must not be increasing in r.

the condition $U_i^e(r) > \max\{U_i^e(r-1),\ U_i^e(r+1)\}$ is only verified when r is equal to the unique maximizer y_i^* because a strictly concave function exhibits a decreasing difference.

Proposition 2.1 plays a crucial role. As agents make their choice by maximizing their expected payoff, then the observed outcome in the data set is assumed to be equal to y_i^*. Consequently, for any $r \in \mathbb{N}$, I have $y_i = r$ if and only if $U_i^e(r) > \max\{U_i^e(r-1),\ U_i^e(r+1)\}$, which is equivalent to $-\psi_i - \lambda\bar{y}_i^e + a_r < \varepsilon_i < -\psi_i - \lambda\bar{y}_i^e + a_{r+1}$, where $a_r = \Delta c(r) + \lambda r - \frac{\lambda}{2}$, $\bar{y}_i^e = \sum_{j\in\mathcal{P}_i} g_{ij}y_j^e$, and y_j^e is the agent i's belief regarding agent j's choice, i.e., $y_j^e = \mathbb{E}(y_j|\mathcal{I})$. This characterization can be used to express the probability of the event $\{y_i = r\}$ conditionally on $\mathcal{I}$. Let $p_{ir} = \mathbb{P}(y_i = r|\mathcal{I})$ 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) expresses p_{ir}, the *true* probability of the event $\{y_i = r\}$, as a function of $\bar{y}_i^e$, agent i's belief about the average outcome among friends. The assumption of RE implies that this belief also corresponds to the *true* expected average outcome among friends, i.e., $y_i^e = \sum_{r=1}^{\infty} rp_{ir}$. Thus, Equation (3) is equivalent to

$$p_{ir} = F_{\varepsilon|\mathcal{I}}\left(\lambda\textstyle\sum_{j\in\mathcal{P}_i} g_{ij}\sum_{t=1}^{\infty} tp_{jt} + \psi_i - a_r\right) - F_{\varepsilon|\mathcal{I}}\left(\lambda\textstyle\sum_{j\in\mathcal{P}_i} g_{ij}\sum_{t=1}^{\infty} tp_{jt} + \psi_i - a_{r+1}\right). \tag{4}$$

Taking Equation (4) for all $i \in \mathcal{V}$ and $r \in \mathbb{N}$ yields a vector fixed point equation in $\mathbf{p} = (p_{ir})_{\substack{i\in\mathcal{V}\\ r\in\mathbb{N}}}$, which is the consistency condition of any rational belief system. The belief system $\mathbf{p}$ is said to be *rational* or *consistent* (with respect to the distribution of ε_i) if and only if it verifies Equation (4).

I now discuss some implications arising from Equation (4). First, imposing a quadratic cost function would be overly restrictive from an econometric perspective. As $a_r = \Delta c(r) + \lambda r - \dfrac{\lambda}{2}$, a 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 SAR Tobit 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/SAR Tobit models.

Second, Equation (4) generalizes the case of binary outcomes under RE studied by Lee et al. (2014). Indeed, if the cost function is such that $\Delta c(r) = +\infty$ for any $r \geqslant 2$, then y_i can only take the values 0 and 1 and $p_{jr} = 0$ for any j and $r \geqslant 2$. In this case, Equation (4) would imply that $p_{i0} = 1 - F_{\varepsilon|\mathcal{I}}(\lambda\sum_{j\in\mathcal{P}_i} g_{ij}p_{j1} + \psi_i - a_1)$ and $p_{i1} = F_{\varepsilon|\mathcal{I}}(\lambda\sum_{j\in\mathcal{P}_i} g_{ij}p_{j1} + \psi_i - a_1)$, which is the characterization of the rational beliefs in binary response models.

Finally, Equation (4) is similar to the specification of an ordered model (see Amemiya, 1981). One can get the same characterization by assuming a latent variable $\hat{y}_i := \lambda \sum_{j \in \mathcal{P}_i} g_{ij} \sum_{t=1}^{\infty} tp_{jt} + \psi_i + \varepsilon_i$, such that $y_i = r$ if and only if $\hat{y}_i \in (a_r,\ a_{r+1})$. However, the microfoundations behind both specifications are different. In the case of an ordered model, agents choose the latent variable $\hat{y}_i$ 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, $r \in \mathbb{N}$.

2.2 Bayesian Nash Equilibrium

Proposition 2.1 states that there is a unique count choice $\mathbf{y}^* = (y_1^*, \dots, y_n^*)'$, such that y_i^* maximizes the expected payoff (2), given the expected outcome $\mathbf{y}^e = (y_1^e, \dots, y_n^e)'$. This implies that $\mathbf{y}^*$ is the unique BNE of the game associated with payoff (1), given agent beliefs about the choice of other players. By assumption, these beliefs are rational and verify the fixed point Equation (4). However, there is no guarantee that a unique belief system verifies Equation (4). It is therefore important to establish conditions under which Equation (4) has a unique solution.

Taking Equation (4) for all $i \in \mathcal{V}$ and $r \in \mathbb{N}$ yields a vector fixed point equation in $\mathbf{p} = (p_{ir})_{\substack{i \in \mathcal{V} \\ r \in \mathbb{N}}}$, defined in the space $[0,1]^{\infty}$. This equation may not have a solution because $[0,1]^{\infty}$ is not a compact set.[10] This is different from the case of the binary or multinational response models, where Brouwer's fixed point theorem guarantees that a rational belief system exists (Brock and Durlauf, 2001; Guerra and Mohnen, 2020). Unlike binary response models, there is no guarantee that a rational belief system exists here. Moreover, as $\mathbf{p}$ is an infinite-dimensional vector, it would be burdensome to apply the contraction mapping theorem to Equation (4) to show that it has a unique solution..

I circumvent this issue using Equation (3). As $y_i^e = \sum_{r=1}^{\infty} rp_{ir}$, Equation (3) implies that the knowledge of the expected outcome $\mathbf{y}^e$ is sufficient to compute the underlying beliefs $\mathbf{p}$ and vice versa. Consequently, if there is a unique $\mathbf{y}^e$ that verifies Equation (3), such that $y_i^e = \sum_{r=1}^{\infty} rp_{ir}$, then the rational belief system is unique. Since $\mathbf{y}^e$ 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* $\ell_i(\mathbf{y}^e) = \sum_{r=1}^{\infty} F_{\varepsilon|\mathcal{I}}(\lambda \bar{y}_i^e + \psi_i - a_r)$. *Let* $\mathbf{p}^* = (p_{ir}^*)_{\substack{i \in \mathcal{V} \\ r \in \mathbb{N}}}$ *be a rational belief system, in the sense that it verifies Equation (4). The associated rational expected outcome* $\mathbf{y}^{e*} = (y_1^{e*}, \dots, y_n^{e*})'$, *such that* $y_i^{e*} = \sum_{r=1}^{\infty} rp_{ir}^*$, *verifies* $\mathbf{y}^{e*} = \mathbf{L}(\mathbf{y}^{e*})$.

[10] Neither Brouwer's fixed point theorem nor Schauder's fixed point theorem, which is a generalization of Brouwer's fixed point theorem to an infinite-dimensional space (see Smart, 1980, Chapter 2), can be applied to Equation (4) since $[0,1]^{\infty}$ is not a compact set. Indeed, If $[0,1]^{\infty}$ were a compact set, any of its sequences would converge or would have a subsequence 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 subsequence of this sequence do not have a limit.

The condition for a belief system $\mathbf{p}$ to be rational is given by Equation (4). Proposition 2.2 states that the expected outcome $\mathbf{y}^e$ associated to this belief system verifies $y_i^e = \ell_i(\mathbf{y}^e)$, where $y_i^e = \sum_{r=1}^{\infty} r p_{ir}$. The uniqueness of the rational expected outcome can be directly established if $\mathbf{L}$ is a contracting mapping. I then make the following assumption.

Assumption 2.4. *$\lambda < B_c/\|\mathbf{G}\|_\infty$, where $B_c = \left(\max_{u\in\mathbb{R}} \sum_{r=1}^{\infty} f_{\varepsilon|\mathcal{I}}(u - a_r)\right)^{-1}$.*[11]

The multiple RE equilibria issue generally arises in peer effect models when the peer effect parameter exceeds some threshold.[12] Assumption 2.4 generalizes the restriction imposed on λ in the binary model. If $\Delta c(r) = +\infty$ for $r \geqslant 2$, Assumption 2.4 implies that $\lambda < (\|\mathbf{G}\|_\infty f_{\varepsilon|\mathcal{I}}(0))^{-1}$, which is the restriction set on λ in the binary data model (see Lee et al., 2014). When the network matrix is row normalized ($\|\mathbf{G}\|_\infty = 1$), 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 a lot of cases (see Bramoullé et al., 2009).

Proposition 2.3. *Under Assumptions 2.1–2.4, the game of incomplete information associated with payoff (1) has a unique BNE given by $\mathbf{y}^* = (y_i^*, \dots, y_n^*)'$ and a unique rational belief system $\mathbf{p}^*$, where y_i^* is the maximizer of the expected payoff $U_i^e(.)$ and $\mathbf{p}^*$ verifies Equation (4).*

The uniqueness of the BNE is a direct implication of Proposition 2.1. In addition, Assumption 2.4 implies that $\mathbf{L}$ is a contracting mapping. There is thus a unique rational expected outcome $\mathbf{y}^{e*}$, such that $\mathbf{y}^{e*} = \mathbf{L}(\mathbf{y}^{e*})$ and a unique rational belief system that verifies Equation (4).

It is worth mentioning that Assumption 2.4 may not be a necessary condition for the uniqueness of the rational belief system. 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 implications of the violation of Assumption 2.4 in the current 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 an expected outcome exists if $\lambda \geqslant B_c/\|\mathbf{G}\|_\infty$. This is also pointed out above with Equation 4, which may not have a solution because $\mathbf{p}$ is an infinite-dimensional vector. This is different from the case of the binary response model, where the rational belief system necessarily exists. The same issue also rises in the linear-in-means models. If the peer effect parameter is not sufficiently low for Assumption 2.4

[11] For any $n \times n$-matrix $\mathbf{Q} = (q_{ij})_{ij}$, $\|\mathbf{Q}\|_\infty = \max_i \sum_{i=1}^{\infty} |q_{ij}|$.

[12] 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.

to be verified, it would be challenging to characterize the game's equilibrium. This problem generally leads to corner solutions when the outcome is bounded (see Tamer, 2003; Soetevent and Kooreman, 2007; De Paula, 2013, 2017). In my case, since y_i^e is not bounded above, the solution of the fixed point equation $y_i^e = \ell_i(\mathbf{y}^e)$ may explode if Assumption 2.4 is violated. Therefore, the case of multiple equilibria goes beyond the scope of this paper. As peer effect estimate in nonlinear models is generally low in practice, I expect Assumption 2.4 to be verified in many cases.[13]

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

I assume that $\psi_i = \alpha + \boldsymbol{x}_i'\boldsymbol{\beta} + \bar{\boldsymbol{x}}_i'\boldsymbol{\gamma}$, where $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}_i$ are K-vectors of observable individual-specific characteristics and peers' average characteristics, respectively, and α, $\boldsymbol{\beta}$, 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). Let $\mathbf{X} = [\boldsymbol{x}_1 \ \dots \ \boldsymbol{x}_n]'$, $\mathbf{1}_n = (1, \dots, 1)' \in \mathbb{R}^n$, $\mathbf{Z} = [\mathbf{1}, \ \mathbf{X}, \ \mathbf{GX}]$ and $\boldsymbol{\Gamma} = \left(\alpha, \boldsymbol{\beta}', \boldsymbol{\gamma}'\right)'$. It follows that $\boldsymbol{\psi} = \mathbf{Z}\boldsymbol{\Gamma}$.

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 $\delta_r = \Delta c(r) - \Delta c(r-1) + \lambda$ for any $r \geqslant 2$. This means that $\delta_r > \lambda$ because the assumption of strictly convex cost function implies that $\Delta c(r) - \Delta c(r-1) > 0$. 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 $\Delta c(r) - \Delta c(r-1) + \lambda > \epsilon$ for some $\epsilon > 0$. This rule allows some concave cost functions depending on the value of λ but still guarantees that the expected payoff (2) is concave in y_i. 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 (4), there is an infinite number of cut points (or δ_r) to be estimated because the cost function is nonparametric. Without additional restrictions on δ_r's, the model identification would be challenging. My identification strategy relies the condition $\lim\limits_{r\to\infty} r^{-\rho}(\Delta c(r+1) - \Delta c(r)) > 0$ set in Assumption 2.3. In particular, I assume that this limit is reached for large values of r.

[13]This is also the reason why I consider a game of incomplete information. In fact, a model similar to that studied in this paper under complete information would yield multiple equilibria unless one sets highly constrained assumptions. See for example the case of binary outcomes studied by Tamer (2003) and Soetevent and Kooreman (2007). It is not clear how to generalize those frameworks to the case where the outcome is unbounded.

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/SAR Tobit 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 (4) can now be expressed as

$$p_{ir} = F_{\varepsilon|\mathcal{I}}\left(\lambda \sum_{j\in\mathcal{P}_i} g_{ij} \sum_{t=1}^{\infty} tp_{jt} + \boldsymbol{z}_i'\boldsymbol{\Gamma} - a_r\right) - F_{\varepsilon|\mathcal{I}}\left(\lambda \sum_{j\in\mathcal{P}_i} g_{ij} \sum_{t=1}^{\infty} tp_{jt} + \boldsymbol{z}_i'\boldsymbol{\Gamma} - a_{r+1}\right), \tag{5}$$

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 $r > R$, and $\boldsymbol{z}_i'$ is the i-th row of $\mathbf{Z}$. Two other specifications of the model are important for the econometric analysis. For instance, since $y_i^e = \sum_{r=1}^{\infty} rp_{jr}$, we also have

$$p_{ir} = F_{\varepsilon|\mathcal{I}}\left(\lambda \bar{y}_i^e + \boldsymbol{z}_i'\boldsymbol{\Gamma} - a_r\right) - F_{\varepsilon|\mathcal{I}}\left(\lambda \bar{y}_i^e + \boldsymbol{z}_i'\boldsymbol{\Gamma} - a_{r+1}\right). \tag{6}$$

From Proposition 2.2, another specification is

$$y_i^e = \sum_{r=1}^{\infty} F_{\varepsilon|\mathcal{I}}\left(\lambda \bar{y}_i^e + \boldsymbol{z}_i'\boldsymbol{\Gamma} - a_r\right). \tag{7}$$

Note that Equation (7) is not equivalent to the first two specifications, but rather is implied by these specifications. In practice, the observed counting variable in the data set is assumed to be equal to the BNE $\mathbf{y}^*$. Unlike players who know $\boldsymbol{\psi}$ and can perfectly compute the rational expected outcome $\mathbf{y}^{e*}$ and the rational belief system $\mathbf{p}^*$, the practitioner does not observe $\boldsymbol{\psi}$ as it depends on the unknown parameters of the model. As a result, they cannot directly compute $\mathbf{y}^{e*}$ and $\mathbf{p}^*$ by solving Equations (6) and (7). However, given λ, $\boldsymbol{\Gamma}$, a_1, a_2, $\dots$, and $F_{\varepsilon|\mathcal{I}}$, $\mathbf{y}^{e*}$ and $\mathbf{p}^*$ can be computed under Assumption 2.4. Indeed, as stated in Proposition 2.3, there is a unique $\mathbf{y}^{e*}$ that verifies Equation (7). This solution can be replaced in Equation (6) to obtain the outcome's distribution $\mathbf{p}^*$.

3.2 Identification

The free objects to be identified in Equation (5) are λ, $\boldsymbol{\Gamma}$, $\boldsymbol{\delta} = (\delta_2, \dots, \delta_R)'$, $\bar{\delta}$, 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. Given that the rational expected outcome depends on the cdf $F_{\varepsilon|\mathcal{I}}$, it is challenging to estimate the model parameters without specifying 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 3.2. *ε_i is distributed as $\mathcal{N}(0,1)$ identically across i, conditionally on $\mathbf{Z}$ and $\mathbf{G}$.*

Following Manski (1988), I conduct a more general identification analysis in OA S.3 without making any specific assumptions about the distribution of ε_i. Assumption 3.2 indirectly imposes that $\mathbf{X}$ and $\mathbf{G}$ are exogenous, so that $F_{\varepsilon|\mathcal{I}} = \Phi$, where Φ is the cdf of $\mathcal{N}(0,1)$. This suggests that there is no omission of important regressors in $\mathbf{X}$. I discuss in Section 3.4 how to control for network endogeneity in the case where unobserved factors that may explain the counting outcome and the network are missing in $\mathbf{X}$. Assumption 3.2 also sets the variance of ε_i to one. This is an important identification restriction also imposed in ordered response models.

Given λ, $\mathbf{\Gamma}$, $\boldsymbol{\delta} = (\delta_2, \dots, \delta_R)'$, $\bar{\delta}$, a_1, ρ, and $\bar{R}$, the rational expected outcome and the rational belief system can be perfectly computed from Equations (6) and (7). Moreover, the rational belief system $\mathbf{p}^*$ defines the likelihood function of the model. Consequently, the nonidentification issue raises when two different parameter collections yield the same rational belief system $\mathbf{p}^*$, i.e., the same distribution for the outcome $\mathbf{y} = (y_1, \dots, y_n)'$. This is formally stated by the following definition.

Definition 3.1. *The parameter collection $\mathcal{B} = \{\lambda,\ \mathbf{\Gamma},\ \boldsymbol{\delta},\ \bar{\delta},\ a_1,\ \rho,\ \bar{R}\}$ is observationally equivalent to the alternative $\tilde{\mathcal{B}} = \{\tilde{\lambda},\ \tilde{\mathbf{\Gamma}} = (\tilde{\alpha}, \tilde{\boldsymbol{\beta}}', \tilde{\boldsymbol{\gamma}}')',\ \tilde{\boldsymbol{\delta}} = (\tilde{\delta}_2, \dots, \tilde{\delta}_{\tilde{R}})',\ \tilde{\bar{\delta}},\ \tilde{a}_1,\ \tilde{\rho},\ \tilde{\bar{R}}\}$ if $p^*_{ir} = \tilde{p}^*_{ir}$ for any $i \in \mathcal{V}$ and $r \in \mathbb{N}$, where p^*_{ir} and $\tilde{p}^*_{ir}$ verify Equation (5), with the corresponding parameter collection, $\mathcal{B}$ or $\tilde{\mathcal{B}}$.*

Definition 3.1 is a generalization of the concept of observational equivalence in binary outcome models. See for example the definition of Brock and Durlauf (2001) on page 252 and that of Brock and Durlauf (2007) on page 54. Importantly, as $y^{e*}_i = \sum_{r=1}^{\infty} r p^*_{ir}$ and $\tilde{y}^{e*}_i = \sum_{r=1}^{\infty} r \tilde{p}^*_{ir}$, where $\tilde{y}^{e*}_i$ is the expected outcome associated with $\tilde{\mathbf{p}} = (\tilde{p}^*_{ir})_{\substack{i \in \mathcal{V} \\ r \in \mathbb{N}}}$, the definition of observational equivalence also implies that $y^{e*}_i = \tilde{y}^{e*}_i$. This remark is obvious in the case of the binary outcome models because $y^{e*}_i = p^*_{i1}$, but is important to emphasize in my framework. Even though we cannot directly observe the rational expected outcomes, the definition of observational equivalence implies that they are identical.

Since α and a_1 enter Equation (5) only through their difference, they cannot be identified. As in an ordered response model, I impose that $a_1 = 0$. If $\mathcal{B} \neq \tilde{\mathcal{B}}$ are observationally equivalent, then specifically, $p^*_{i0} = \tilde{p}^*_{i0}$, which means that $\Phi(\lambda \bar{y}^{e*}_i + \boldsymbol{z}_i'\mathbf{\Gamma}) = \Phi(\tilde{\lambda} \bar{y}^{e*}_i + \boldsymbol{z}_i'\tilde{\mathbf{\Gamma}})$. By assuming that $\mathbf{Z}$ is a full rank matrix, the previous equation implies that $\bar{y}^{e*}_i = \boldsymbol{z}_i' \frac{\mathbf{\Gamma} - \tilde{\mathbf{\Gamma}}}{\tilde{\lambda} - \lambda}$, where λ is necessarily different from $\tilde{\lambda}$. In other words, a crucial condition for the identification is that $\bar{y}^{e*}_i$ cannot be written as a linear combination of the components in $\boldsymbol{z}_i$ for all i.

As $y^{e*}_i = p^*_{i1}$ in binary outcome models, the necessary condition for the identification is that $p^*_{i1} = \boldsymbol{z}_i' \frac{\mathbf{\Gamma} - \tilde{\mathbf{\Gamma}}}{\tilde{\lambda} - \lambda}$ is not possible. This condition is generally assumed to be true because the support of

p^*_{i1} is bounded independently of $\boldsymbol{z}_i$ (e.g., see Brock and Durlauf, 2001; Lee et al., 2014; Liu, 2019; Yang and Lee, 2017; Guerra and Mohnen, 2020).[14] Another way to justify this assumption is to use Equation (7). Taking this equation for the rational expected outcome for all agents would express $\mathbf{y}^{e*}$ as a nonlinear function of $\mathbf{G}\mathbf{y}^{e*}$ and $\mathbf{Z}$. Therefore, the relationship that links $\mathbf{G}\mathbf{y}^{e*}$ to $\mathbf{Z}$ would not be a linear function. This argument is for example used by Lin et al. (2021).

Although those arguments can also be employed in my framework, it is worth mentioning that they may be restrictive. First, y^{e*}_i is bounded only at the bottom by zero. As in the Tobit model, the upper bound depends on the support of $\boldsymbol{z}_i$. The expected outcome may be unbounded above if there is an unbounded explanatory regressor in $\mathbf{X}$. The second argument seems more realistic. However, as pointed out above, my model is closer to the Tobit model when $\rho = 0$ and $\bar{R} = 1$. Indeed, if $\rho = 0$ and $\bar{R} = 1$, then Equation (7) can be approximated using a linear equation for large $y^e_i + \boldsymbol{z}'_i\boldsymbol{\Gamma}$ (see OA S.2). Therefore, the argument that the nonlinearity of Equation (7) can allow for identifying the model parameters may seem unpersuasive. This argument would be valid only when $\bar{y}^{e*}_i$ is close to zero with minimal variations for some i's whereas $\boldsymbol{z}'_i$ exhibits significant variations.

I show that the parameters can be identified under conditions similar to the identification conditions of linear-in-means models (see Bramoullé et al., 2009). My identification analysis is general as it can also be applied to many other nonlinear models, such as binary and ordered outcome models, irrespective of whether the game information is complete or incomplete. Let $\text{supp}(\boldsymbol{z})$ be the common support of $\boldsymbol{z}_i$'s. For any vector $\mathbf{b}$, b_k denotes its k-th component. I set the following assumption.

Assumption 3.3. $\sup_i \sum_{j=1}^n g_{ij}$ *is uniformly bounded in* n.

Assumption 3.4. $\text{supp}(\boldsymbol{z})$ *is not contained in a proper linear subspace of* $\mathbb{R}^{2K+1}$.

Assumption 3.5. *(i) There exists an integer* $k_0 \in [1, K]$ *such that* $\beta_{k_0}\gamma_{k_0} \geqslant 0$ *and* $\gamma_{k_0} \neq 0$. *(ii) There is a strictly positive proportion of agents whose friends' friends are not friends, as* n *grows to infinity.*

Assumption 3.3 rules out the cases where $\bar{y}^{e*}_i$ or $\bar{\boldsymbol{x}}_i$ diverges as n grows to infinity. This assumption is verified if $\tilde{n}_i$, the number of friends i has, and g_{ij} are uniformly bounded in i, j, and n. It can also hold when $\tilde{n}_i$ diverges and $\mathbf{G}$ is row-normalized. Assumption 3.4 originates from Manski (1988) and imposes that $\mathbf{Z}$ is a full rank matrix.

Condition (i) of Assumption 3.5 is a sufficient condition for the rational expected choice y^{e*}_i to be affected by at least one contextual variable. Indeed, without loss of generality, assume that $\beta_{k_0} \geqslant 0$ and $\gamma_{k_0} > 0$. If j is i's friend, then an increase in $\boldsymbol{x}_{j,k_0}$, the k_0-th component of $\boldsymbol{x}_j$, implies an increase

[14] For example, Brock and Durlauf (2001) highlight that "*If the characteristics and behaviours of the neighbourhoods always move in proportion as one moves across neighbourhoods, then clearly one could not determine the respective roles of the characteristics as opposed to the behaviour of the group in determining individual outcomes. This can never happen in the logistic binary choice case given that the expected average choice must be bounded between* -1 *and 1. So, for example, if one moves across a sequence of richer and richer communities, the percentage of high school graduates cannot always increase proportionately with income.*"

in y_i^{e*} (because $\gamma_{k_0} > 0$) and possibly an increase in y_j^{e*} (because $\beta_{k_0} \geqslant 0$). 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 $\boldsymbol{x}_{j,k_0}$ 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 $\lambda\beta_{k_0} + \gamma_{k_0} \neq 0$, 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) sets that there is a positive proportion of agents whose friends' friends are not all friends. This condition is satisfied in many social network settings. However, it is violated if agents are connected to each other. Condition (ii) implies that $\mathbf{I}_n$,[15] $\mathbf{G}$, 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).

Proposition 3.1. *Under Assumptions 2.1–3.5, λ, $\mathbf{\Gamma}$, $\boldsymbol{\delta}$, $\bar{\delta}$, ρ and $\bar{R}$ are point identified.*

The key point of the proof of Proposition 3.1 is to show that $\bar{y}_i^{e*}$ is not linearly dependent on $\boldsymbol{z}_i$ (see proof in Appendix A.4). For example, consider a network of three agents i, j, and l, as illustrated by Figure 1. The network structure follows Condition (ii) of Assumption 3.5. Indeed, j is i's friend and l is j's friend, but l is not i's friend. If $\bar{y}_i^{e*}$ is linearly dependent on $\boldsymbol{z}_i$, I can write that

$$\bar{y}_i^{e*} = \breve{\alpha} + \boldsymbol{x}_i'\breve{\boldsymbol{\beta}} + \bar{\boldsymbol{x}}_i'\breve{\boldsymbol{\gamma}}, \tag{8}$$

for all i, where $\breve{\alpha} \in \mathbb{R}$ and $\breve{\boldsymbol{\beta}}$, $\breve{\boldsymbol{\gamma}} \in \mathbb{R}^K$. As j is i's only friend, thus $\bar{y}_i^{e*} = y_j^{e*}$ and $\bar{\boldsymbol{x}}_i = \boldsymbol{x}_j$ (I normalize g_{ij} to one). Therefore, Equation (8) becomes $y_j^{e*} = \breve{\alpha} + \boldsymbol{x}_i'\breve{\boldsymbol{\beta}} + \boldsymbol{x}_j'\breve{\boldsymbol{\gamma}}$, which means that y_j^{e*} does not depend on $\boldsymbol{x}_l$. This is not possible because l is a direct friend of j, i.e., $\boldsymbol{x}_l$ is the vector of contextual variables for y_j^{e*}. Condition (i) of Assumption 3.5 implies the rational expected choice should be affected by at least one contextual variable.

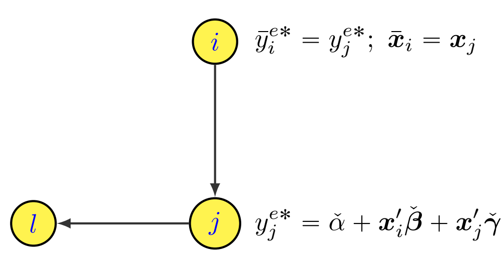

Figure 1: Illustration of the identification

My identification analysis is general as it can also be applied to many other nonlinear models, irrespec-

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

tive of whether the game information is complete or incomplete. It can be applied to any game, whose best response function implies $y_i^e = \check{h}(\check{m}(\bar{y}_i^e) + \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{x}_j'\boldsymbol{\gamma})$ or $y_i = \check{h}(\check{m}(\bar{y}_i) + \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{x}_j'\boldsymbol{\gamma})$, where $\check{h}$ and $\check{m}$ are strictly increasing functions. A multicollinearity issue cannot raise in such models if the network includes agents whose friends' friends are not friends. The intuition is that the linear relationship that would link $m(\bar{y}_i^e)$ to $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}_i$ only involves agent i (through the term $\boldsymbol{x}_i$) and i's friends (through the terms $m(\bar{y}_i^e)$ and $\bar{\boldsymbol{x}}_i$), but not i's friends' friends who are not friends. Therefore, this relationship cannot capture changes in the characteristics of i's friends' friends who are not friends. Indeed, these changes would influence $m(\bar{y}_i^e)$ through the impact of the contextual variables of i's friends but would have no influence on $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}_i$. In the relationship, only $m(\bar{y}_i^e)$ cannot vary whereas $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}_i$ are constant. This means that having a relationship linking only $m(\bar{y}_i^e)$ to $\boldsymbol{x}_i$ and $\bar{\boldsymbol{x}}_i$ is not compatible with the condition that the network includes agents whose friends' friends are not all friends.

3.3 Estimation

This section presents a simple likelihood-based approach for estimating the model parameters. The distribution of the observed counting variable $\mathbf{y}$ is defined by $\mathbf{p}^* = (p_{ir}^*)_{\substack{i\in\mathcal{V}\\ r\in\mathbb{N}}}$. By assigning values to the parameters in $\mathcal{B}$, the expected outcome $\mathbf{y}^{e*}$ can be determined by solving the fixed point Equation (7). By replacing the solution in Equation (6), one can also compute $\mathbf{p}^*$. Having the distribution $\mathbf{y}$ given the values assigned to the parameters in $\mathcal{B}$ suggests using the maximum likelihood (ML) approach to estimate these parameters.

The likelihood function is increasing in $\bar{R}$. For two values of $\bar{R}$, the model associated with the lower $\bar{R}$ is nested within the one associated with the higher $\bar{R}$. Consequently, estimating $\bar{R}$ by maximizing the likelihood is not a good practice as $\bar{R}$ would explode. Moreover, since $\boldsymbol{\delta}$ has $\bar{R}-1$ elements, increasing $\bar{R}$ implies more parameters to be estimated and will increase the estimate variance. To address this issue, I propose estimating the model for several values of $\bar{R}$ (e.g., using a grid of integers) and selecting the value at which the estimates stabilize or the value that minimizes the Bayesian information criterion (BIC). Further discussion on how to choose the appropriate $\bar{R}$ will follow. For now, I assume that the practitioner selects an arbitrary value for $\bar{R}$.

To deal with the constraint $\delta_r \geqslant \lambda$ for all $r \geqslant 2$ in Assumption 3.1, I define $\tilde{\boldsymbol{\delta}} = (\tilde{\delta}_2, \dots, \tilde{\delta}_{\bar{R}})$, where $\tilde{\delta}_r = \delta_r - \lambda$. Let $\boldsymbol{\theta} = \left(\log(\lambda),\ \boldsymbol{\Gamma}', \log(\tilde{\boldsymbol{\delta}}'), \log(\bar{\delta}), \log(\rho)\right)'$.[16] As I set that $\boldsymbol{\psi} = \mathbf{Z}\boldsymbol{\Gamma}$, I redefine the mapping $\mathbf{L}$ as $\mathbf{L}(\boldsymbol{\theta}, \mathbf{y}^e) = (\ell_1(\boldsymbol{\theta}, \mathbf{y}^e), \dots, \ell_n(\boldsymbol{\theta}, \mathbf{y}^e))'$, where $\ell_i(\boldsymbol{\theta}, \mathbf{y}^e) = \sum_{r=1}^{\infty} \Phi\left(\lambda \mathbf{g}_i \mathbf{y}^e + \boldsymbol{z}_i'\boldsymbol{\Gamma} - a_r\right)$,

[16] 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.

$a_0 = -\infty$, $\delta_1 = 0$, and $a_r = \sum_{k=1}^{r} \delta_k$ for any $r \geqslant 1$. Equation (6) is equivalent to

$$p_{ir} = \Phi\left(\lambda \mathbf{g}_i \mathbf{y}^e + \boldsymbol{z}_i' \boldsymbol{\Gamma} - a_r\right) - \Phi\left(\lambda \mathbf{g}_i \mathbf{y}^e + \boldsymbol{z}_i' \boldsymbol{\Gamma} - a_{r+1}\right). \tag{9}$$

Given that $\mathbf{y}^{e*}$ 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 avoids the need of solving a fixed point problem. The NPL algorithm is based on a *pseudo*-likelihood function defined as

$$\mathcal{L}_n(\boldsymbol{\theta}, \mathbf{y}^e) = \frac{1}{n} \sum_{i=1}^{n} \sum_{r=0}^{\infty} d_{ir} \log(p_{ir}), \tag{10}$$

where $d_{ir} = 1$ if $y_i = r$ and $d_{ir} = 0$ otherwise. This is a pseudo-likelihood function because $\mathbf{y}^e$ in Equation (10) is an arbitrary vector and not the true rational expected outcome. In other words, in Equation (10), p_{ir} verifies Equation (9) for arbitrary $\mathbf{y}^e$ and $\boldsymbol{\theta}$.

To describe the NPL algorithm, it is useful to define the operators $\tilde{\boldsymbol{\theta}}_n(\mathbf{y}^e) = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \mathcal{L}_n(\boldsymbol{\theta}, \mathbf{y}^e)$ and $\boldsymbol{\phi}_n(\mathbf{y}^e) = \mathbf{L}(\tilde{\boldsymbol{\theta}}(\mathbf{y}^e), \mathbf{y}^e)$. 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 $\mathbf{y}_{(t)} = \boldsymbol{\phi}_n(\mathbf{y}^e_{(t-1)})$. Specifically, 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 $\boldsymbol{\theta}_{(2)} = \tilde{\boldsymbol{\theta}}_n(\mathbf{y}^e_{(1)})$ 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 $\{\hat{\boldsymbol{\theta}}_n(\bar{R}), \hat{\mathbf{y}}^e_n(\bar{R})\}$, is the NLP estimator. This limit satisfies the following two properties: $\hat{\boldsymbol{\theta}}_n(\bar{R})$ maximizes the pseudo-likelihood $\mathcal{L}_n(\boldsymbol{\theta}, \hat{\mathbf{y}}^e_n(\bar{R}))$ and $\hat{\mathbf{y}}^e_n(\bar{R}) = \mathbf{L}(\hat{\boldsymbol{\theta}}_n(\bar{R}), \hat{\mathbf{y}}^e_n(\bar{R}))$. 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 $\|\boldsymbol{\theta}_{(t)} - \boldsymbol{\theta}_{(t-1)}\|_2$ and $\|\mathbf{y}^e_{(t)} - \mathbf{y}^e_{(t-1)}\|_2$ are less than some tolerance values (for example 10^{-4}), where $\|.\|_2$ is the euclidean norm.[17]

Asymptotic properties and choice of $\bar{R}$

I now study the limiting distribution of the proposed estimator and discuss how to set $\bar{R}$.

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, $\dim(\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}}))$, the dimension of $\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}})$ depends on $\hat{\bar{R}}$, whereas $\dim(\boldsymbol{\theta}^0)$ depends on $\bar{R}^0$. Thus, $\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}})$ and $\boldsymbol{\theta}^0$ 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 $\hat{\bar{R}} > \bar{R}^0$,

[17]The pseudo-likelihood (10) involves an infinite sum. As $d_{ir} = 0$ for any $r \neq y_i$, this pseudo-likelihood can also be expressed as $\mathcal{L}_n(\boldsymbol{\theta}, \mathbf{y}^e) = \frac{1}{n}\sum_{i=1}^{n} \log(p_{iy_i})$. Moreover, the mapping $\mathbf{L}$ and the asymptotic variance of $\hat{\boldsymbol{\theta}}$ 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.

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 $\delta^0_r = (r-1)^{\rho^0}\bar{\delta}^0 + \lambda^0$ for any $r > \bar{R}^0$. I establish the following result on the consistency of the NPL parameter.

Proposition 3.2. *Assume that Assumptions 2.1–3.5 and the regulatory conditions of the NPL estimator hold (see Appendix A.5). For any $\hat{\bar{R}}$, (i) $\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}}) - \check{\boldsymbol{\theta}}_n(\hat{\bar{R}})$ converges in probability to zero, where $\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}) = \boldsymbol{\theta}^0$ if $\hat{\bar{R}} \geqslant \bar{R}^0$; (ii) $\sqrt{n}(\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}}) - \check{\boldsymbol{\theta}}_n(\hat{\bar{R}})) \xrightarrow{d} \mathcal{N}\left(0, (\mathbf{H}_{1,0} + \mathbf{H}_{2,0})^{-1}\boldsymbol{\Sigma}_0(\mathbf{H}'_{1,0} + \mathbf{H}'_{2,0})^{-1}\right)$, where $\boldsymbol{\Sigma}_0$, $\mathbf{H}_{1,0}$, and $\mathbf{H}_{2,0}$ are given in Appendix A.5.*

Proposition 3.2 adapts Proposition 2 in Aguirregabiria and Mira (2007) to my framework (see proof in Appendix A.5). The NPL estimator may not converge to $\boldsymbol{\theta}^0$ if $\hat{\bar{R}} < \bar{R}^0$. In this case, I establish that it converges to the same quantity as the random variable $\check{\boldsymbol{\theta}}_n(\hat{\bar{R}})$.[18]

To choose a suitable $\hat{\bar{R}}$, the practitioner can compute $\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}})$ for several values $\hat{\bar{R}}$. I propose to use a grid of integer values from one to $\hat{\max}(y) - 2$. As $\bar{\delta}$ and ρ only appear in the p_{ir} when $r \geqslant \hat{\bar{R}}$, then the condition $\hat{\bar{R}} \leqslant \hat{\max}(y) - 2$ is important for the identifying $\bar{\delta}$ and ρ. An appropriate $\hat{\bar{R}}$ is reached if an increase in $\hat{\bar{R}}$ does not significantly change the estimator. In fact, as stated in Proposition 3.2, if $\hat{\bar{R}}$ reaches $\bar{R}^0$, then $\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}})$ will converge to $\boldsymbol{\theta}^0$ as n grows to infinity. As a result, $\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}})$ would not vary significantly as $\hat{\bar{R}}$ increases. The best $\hat{\bar{R}}$ can also be the one that minimizes the BIC. If $\hat{\boldsymbol{\theta}}_n(\hat{\bar{R}})$ varies significantly 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 this section, I discuss how the model parameter can be estimated when the network is endogenous. I study the case where network endogeneity is due to the omission of important variables that can explain the network and the counting outcome in $\mathbf{X}$. This issue generally occurs when these variables are not observed by the practitioner. An example in my empirical application is students' degree of sociability, which can explain students' likelihood to make friends and their participation in extracurricular activities. Students' degree of sociability is not available in Add Health data and is not included in $\mathbf{X}$. If ψ_i is misspecified in the expected payoff (2), the resulting bias will be captured by ε_i, which would be correlated to the network matrix $\mathbf{G}$, and possibly to the matrix $\mathbf{X}$ of observed characteristics. As a result, $\mathbf{G}$ and $\mathbf{X}$ would be endogenous.

For simplification, I assume that this issue only occurs in the econometric model, but not in

[18] My asymptotic analysis is conditional on $\mathbf{X}$ and $\mathbf{G}$ as in Lin and Xu (2017). Without conditioning, on $\mathbf{X}$ and $\mathbf{G}$, the rational expected choices y_i^{e*}'s are not independent across i, which complicates the asymptotic analysis.

the game. The players perfectly observe $\boldsymbol{\psi}$, can compute the expected outcome, and make rational decisions. Therefore, all the results regarding the game in Section 2 are not affected by the endogeneity issue. In contrast, the practitioner cannot accurately compute $\mathbf{y}^{e*}$ and $\mathbf{p}^*$ because they misspecify $\boldsymbol{\psi}$.

A natural approach to dealing with the endogeneity is to estimate the unobserved missing factors in $\mathbf{X}$ and add them as additional regressors. This two-stage estimation approach is interesting as it preserves the identification of the parameters. Indeed, the unobserved factors are identified and estimated in the first stage separately from the model in the second stage. If the approach used in the first stage properly identifies the unobserved factors, then the model in the second stage consists in adding new regressors estimated in the first stage in $\mathbf{X}$. Consequently, the identification of the model in the second stage follows from Proposition 3.1, after one includes the new regressors in $\mathbf{X}$.

To estimate the unobserved missing factors in $\mathbf{X}$, I rely on a network formation model with degree heterogeneity. Let $\ddot{\boldsymbol{x}}_{ij}$ be a $\bar{K}$-vector of observed dyad-specific variables (e.g., $\ddot{\boldsymbol{x}}_{ij}$ may contain the distance between agents i and j's characteristics) and μ_i and ν_i be unobserved attributes of agent i that may influence the network and the outcome y_i. Following Graham (2017); Dzemski (2019), I assume that the probability for j to be i's friend, conditionally on $\ddot{\boldsymbol{x}}_{ij}$, $\bar{\boldsymbol{\beta}}$, μ_i, and ν_j, is defined as

$$P_{ij} = \mathbb{P}\left(g_{ij} > 0 | \ddot{\boldsymbol{x}}_{ij}, \bar{\boldsymbol{\beta}}, \mu_i, \nu_j\right) = F_g\left(\ddot{\boldsymbol{x}}_{ij}^{\prime}\bar{\boldsymbol{\beta}} + \mu_i + \nu_j\right), \tag{11}$$

where $\bar{\boldsymbol{\beta}}$ is a $\bar{K}$-vector to be estimated and function F_g is a logistic of normal distribution function depending on whether the specification is a logit or a probit model. The term $\ddot{\boldsymbol{x}}_{ij}^{\prime}\bar{\boldsymbol{\beta}}$ is a measure of social distance between agents i and j that drives homophily of linking decisions.

By convention, I set $P_{ii} = 0$. Unlike most network formation models, the specification (11) includes two unobservable factors μ_i and ν_j. This implies a nonsymmetric matrix of link probabilities. For any $i \in \mathcal{V}$, the parameter μ_i only influences the probabilities of links going from i to other agents, whereas ν_i influences the probabilities of links going from another agent to i. This feature is relevant for directed networks.

Although Equation (11) raises an incidental parameter issue due to the number of parameters μ_i's and ν_i's which increases with n, Yan et al. (2019) show that the standard logit estimator of μ_i's and ν_i's are consistent if the network is dense. Using simulations, they also show that the logit estimator performs well even when the network is sparse. 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 μ_i's and ν_i's. However, this approach treats the attributes as random effects independent from $\mathbf{X}$ (see Hsieh and Lee, 2016; Albert and Chib, 1993). I implement both approaches in my R package.

Let $\hat{\mu}_i$ and $\hat{\nu}_i$ be the estimators of μ_i and ν_i respectively obtained in the first stage. As μ_i and ν_i

are likely to explain $\mathbf{G}$ and y_i are missing in $\mathbf{X}$, I set the following assumption.

Assumption 3.6. *$\psi_i = \boldsymbol{z}_i'\boldsymbol{\Gamma} + h_\psi(\mu_i, \nu_i, \bar{\mu_i}, \bar{\nu}_i)$, where h_ψ is a continuous function, $\bar{\mu_i} = \sum_{j=1}^n g_{ij}\mu_j$ and $\bar{\nu}_i = \sum_{j=1}^n g_{ij}\nu_j$.*

Including $\bar{\mu_i}$ and $\bar{\nu}_i$ in the function h_ψ allows to deal with situations where friend attributes influence y_i. In practice, the function h_ψ can be approximated using a sieve method, such as series approximation, where μ_i and ν_i are replaced with their estimator (see Ackerberg et al., 2012). For example, $h_\psi(\mu_i, \nu_i, \bar{\mu_i}, \bar{\nu}_i)$ can be approximated by $\sum_{k=1}^T(\theta_{1,k}\hat{\mu}_i^k + \theta_{2,k}\hat{\nu}_i^k + \theta_{3,k}\hat{\bar{\mu}}_i^k + \theta_{4,k}\hat{\bar{\nu}}_i^k)$, where T is a fixed integer, $\theta_{1,k}$, $\theta_{2,k}$, $\theta_{3,k}$, and $\theta_{4,k}$ are parameters to be estimated, $\hat{\bar{\mu}}_i = \sum_{j=1}^n g_{ij}\hat{\mu}_j$, and $\hat{\bar{\nu}}_i = \sum_{j=1}^n g_{ij}\hat{\nu}_j$. The degree T can be set such that attributes to a power greater than T are not significant. Controlling for the endogeneity simply consists in adding μ_i^k, ν_i^k, $\bar{\mu_i}^k$, and $\bar{\nu}_i^k$, for $k = 1, \dots, T$ as additional explanatory variables in $\mathbf{X}$. Accordingly, the identification result in Proposition 3.1 still holds if the new set of explanatory variables verifies the identification conditions. Under regulatory conditions similar to those imposed by Johnsson and Moon (2021), I establish the asymptotic normality of the new estimator (see OA S.5).

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 SAR Tobit model to illustrate that the latter is a particular case. I use the SAR Tobit model because it controls for the left-censure issue.

In this simulation study, $n = 1,500$ and I define $\psi_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \gamma_1 \bar{x}_{1i} + \gamma_2 \bar{x}_{2i}$. 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 $\rho^0 = 0$. Note that $\bar{R}^0$ is supposed to be unknown to the practitioner, who instead uses an empirical value denoted $\hat{\bar{R}}$.

As discussed in Section 3.3, the model can be estimated using a grid of integer values for $\hat{\bar{R}}$. A good value will be reached if an increase in $\hat{\bar{R}}$ does not significantly change the estimates. To see how this approach performs, I set $\hat{\bar{R}}$ to the 90th percentile of y, then to the integer after the 90th percentile. If the estimates associated with the two values are not significantly different, this would suggest that the practitioner could reach a good $\hat{\bar{R}}$ in practice. By setting $\hat{\bar{R}}$ to the 90th percentile of y, I rely on a conservative assessment. In practice, $\hat{\bar{R}}$ can be set higher. For example, it is set to the 99th percentile of y in my application. 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 SAR Tobit 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. The decrease in δ_r is also observed in my application on real data and suggests that the bias of the Tobit model would be higher on variables having longer tails, such as those obtained from survey data sets. 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.

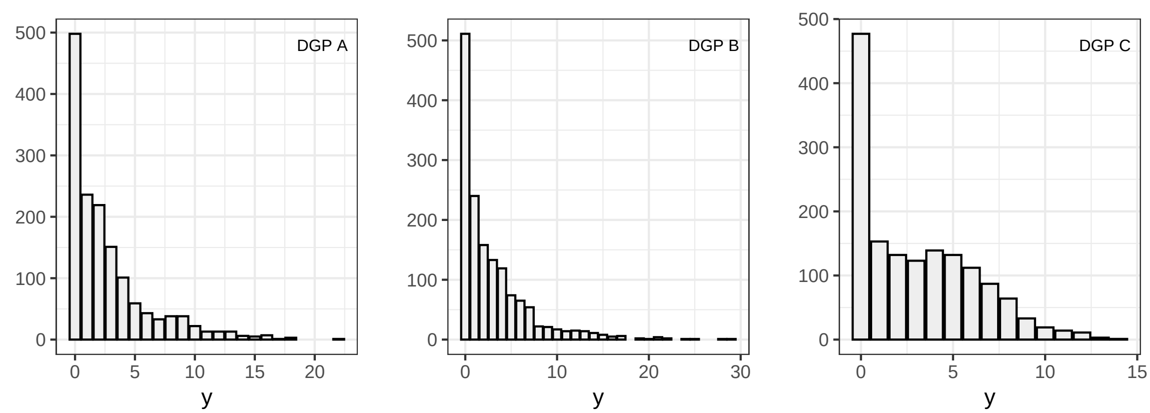

Figure 2: Simulated data using the count data model with social interactions

The simulation results (for 1,000 replications) are presented in Table 1. Note that one cannot directly interpret the parameters of the counting variable model, nor can one compare these parameters to those of the SAR Tobit model. Table 1 reports the marginal effect (ME) of each variable.[19] 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 $\boldsymbol{\theta}^0$ as soon as $\hat{\bar{R}} > \bar{R}^0$. The parameter ρ seems not to play an important role

[19] I compute the ME for each individual and take the average. I present how to derive the marginal effects and the corresponding standard errors for the count data model in OA S.6.

when $\hat{\bar{R}}$ is set high. Indeed, the peer effects are 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.

Table 1: Monte Carlo simulations

	Model 1				**Model 2**				**Model 3**				**Model 4**	
True marginal	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, $\bar{R}^0 = 4$	$\hat{\bar{R}} = 1$				$\hat{\bar{R}} = 8$				$\hat{\bar{R}} = 9$					
$\delta(\bar{y}^e) = 0.316$	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, $\bar{R}^0 = 13$	$\hat{\bar{R}} = 1$				$\hat{\bar{R}} = 8$				$\hat{\bar{R}} = 9$					
$\delta(\bar{y}^e) = 0.329$	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, $\bar{R}^0 = 1$	$\hat{\bar{R}} = 1$				$\hat{\bar{R}} = 7$				$\hat{\bar{R}} = 8$					
$\delta(\bar{y}^e) = 0.274$	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 $\hat{\bar{R}}$ to one. Model 2 is estimated by setting $\hat{\bar{R}}$ to the 90th percentile, and Model 3 has $\hat{\bar{R}}$ 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 $\hat{\bar{R}}$ 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. 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. The network is restricted at the school level. Two students from different schools are not friends.

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."[20] 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. 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 Tobit model to overestimate the marginal peer effect.

5.2 Empirical results

I first determine the suitable value of the parameter $\hat{\bar{R}}$. The counting variable model is estimated using a grid of integer values for $\hat{\bar{R}}$ 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 99th percentile of y), which also corresponds to the lowest BIC. The estimates of the peer effects and the log-likelihoods are presented in Table 2.[21] There is no significant difference between the results for $\hat{\bar{R}} = 12$ and $\hat{\bar{R}} = 13$. I also compare the specification with flexible ρ to that in which ρ is set to zero and do not see a significant difference. I then set $\hat{\bar{R}} = 12$ and $\rho = 0$ for the rest of the empirical analysis. The results indicate that a one-unit increase in the expected number of activities in which a student's friends are enrolled yields an increase in the expected number of activities in which the student is enrolled by 0.09.

I also present the estimate of the marginal peer influence for $\hat{\bar{R}} = 1$. As in my Monte Carlo

[20] 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.

[21] Full results for general models that control for the network endogeneity are available in OA S.7.

experiment, the marginal peer effect corresponds to the estimate with the Tobit model and is six 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 as school dummy variables. 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 traits, such as sociability degree (Newton et al., 2018; Pfeiffer and Schulz, 2012). Moreover, these personality traits are likely to increase student probability of interacting with others. As they are unobserved and not included in $\mathbf{X}$, then $\mathbf{G}$ 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). As this approach requires a dense network, I also use a Bayesian random effect probit model.[22] In this model, one can simulate the posterior distribution of the unobserved attributes using a Gibbs sampler (see Casella and George, 1992; Albert and Chib, 1993). I include a polynomial function of the attributes in $\mathbf{X}$ as additional explanatory variables. The sufficient degree of the polynomial 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.7.

The increase in the log-likelihood suggests that the network is endogenous. However, this endogeneity does not significantly 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.

[22]Note that Yan et al. (2019) found using simulations that their estimator also performs well even in the case of sparse networks.

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.

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, $h(r_0) \geqslant \max\{h(r_0-1),\ h(r_0+1)\}$ if and only if $h(r_0)$ 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, $h(r_0) \geqslant \max\{h(r_0-1),\ h(r_0+1)\}$.
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 if and only if it can be extended to a continuous concave function. As h is concave, let $\bar{h}$ be an extension of h on $\bar{D}$, where $\bar{h}$ is concave and $\bar{h}(r) = h(r)$, $\forall\ r \in D_h$. In particular, one can 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 also the global maximum of h. □

Proof of Proposition 2.1
The expected outcome is $U_i^e(y_i) = \psi_i y_i - c(y_i) - \dfrac{\lambda}{2}\mathbb{E}_{\bar{y}_i|\mathcal{I}}\left[(y_i - \bar{y}_i)^2\right] + e_i(y_i)$.
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.).[23] Indeed, if the global maximum were reached at two points y_i^* and $y_i^{*\prime}$, where $y_i^* \neq y_i^{*\prime}$,

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

then $U_i^e\left(y_i^*\right) = U_i^e\left(y_i^{*\prime}\right)$, which implies $\varepsilon_i = \dfrac{c(y_i^*) - c(y_i^{*\prime})}{y_i^* - y_i^{*\prime}} + \dfrac{\lambda}{2}(y_i^* + y_i^{*\prime}) - \psi_i - \lambda\,\mathbb{E}(\bar{y}_i|\mathcal{I})$. 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 $y_i^{*\prime}$ if and only if $U_i(y_i^{*\prime}) \geqslant \max\left\{U_i(y_i^{*\prime} - 1),\ U_i(y_i^{*\prime} + 1)\right\}$.

A.2 Proof of Proposition 2.2

For any $\mathbf{y}^e \in \mathbb{R}_+^n$, $\mathbf{L}\left(\mathbf{y}^e\right) = \left(\ell_1\left(\mathbf{y}^e\right)\ \ldots \ell_n\left(\mathbf{y}^e\right)\right)'$, where $\ell_i\left(\mathbf{y}^e\right) = \sum_{r=1}^{\infty} F_{\varepsilon|\mathcal{I}}\left(\lambda\bar{y}_i^e + \psi_i - a_r\right)$ for all $i \in \mathcal{V}$ and $\bar{y}_i^e = \sum_{j=1}^n g_{ij}y_j^e$. Let $\mathbf{p}^* = (p_{ir}^*)_{\substack{i\in\mathcal{V}\\ r\in\mathbb{N}}}$ be a rational belief system and $\mathbf{y}^{e*} = (y_1^{e*}, \ldots, y_n^{e*})'$ be its associated rational expected outcome, i.e., $y_i^{e*} = \sum_{r=1}^{\infty} rp_{ir}^*$. $\mathbf{p}^*$ and $\mathbf{y}^{e*}$ verify (3). Thus, $p_{ir}^* = F_{\varepsilon|\mathcal{I}}\left(\lambda\bar{y}_i^{e*} + \psi_i - a_r\right) - F_{\varepsilon|\mathcal{I}}\left(\lambda\bar{y}_i^{e*} + \psi_i - a_{r+1}\right)$.

Let $S_1 = \sum_{r=0}^{\infty} rF_{\varepsilon|\mathcal{I}}\left(\lambda\bar{y}_i^{e*} + \psi_i - a_r\right)$ and $S_2 = \sum_{r=0}^{\infty} rF_{\varepsilon|\mathcal{I}}\left(\lambda\bar{y}_i^{e*} + \psi_i - a_{r+1}\right)$. 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,

$y_i^{e*} = S_1 - \sum_{r=0}^{\infty}(r+1)F_{\varepsilon|\mathcal{I}}\left(\lambda\bar{y}_i^{e*} + \psi_i - a_{r+1}\right) + \sum_{r=0}^{\infty} F_{\varepsilon|\mathcal{I}}\left(\lambda\bar{y}_i^{e*} + \psi_i - a_{r+1}\right),$

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

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

A.3 Proof of Proposition 2.3

Proposition 2.1 guarantees the uniqueness of the BNE since the expected outcome has a unique maximizer. I now show that there is a unique rational belief system.

If $\mathbf{p}^*$ is a rational belief system, Proposition 2.2 states that its associated expected outcome $\mathbf{y}^{e*}$ verifies $\mathbf{y}^{e*} = \mathbf{L}\left(\mathbf{y}^{e*}\right)$. To prove the uniqueness, it is sufficient to prove that $\mathbf{L}$ is contracting, i.e., $\forall\ \mathbf{u} = (u_1, \ldots, u_n) \in \mathbb{R}^n$, $\|\partial\mathbf{L}\left(\mathbf{u}\right)/\partial\mathbf{u}'\|_{\infty} \leqslant \bar{\kappa}_c$ for some $\bar{\kappa}_c < 1$ not depending on $\mathbf{u}$.

For all i and j, $\dfrac{\partial\ell_i\left(\mathbf{u}\right)}{\partial u_j} = \lambda g_{ij}f_i^*$, where $f_i^* = \sum_{r=1}^{\infty} f_{\varepsilon|\mathcal{I}}\left(\lambda\mathbf{g}_i\mathbf{u} + \psi_i - a_r\right)$. Thus, the (i,j)-th element of the matrix $\partial\mathbf{L}\left(\mathbf{u}\right)/\partial\mathbf{u}'$ is $g_{ij}f_j^*$. It follows that $\|\partial\mathbf{L}\left(\mathbf{u}\right)/\partial\mathbf{u}'\|_{\infty} = \lambda\max_i\{f_i^*\sum_{j=1}^n g_{ij}\} \leqslant \lambda\left(\max_i f_i^*\right)\|\mathbf{G}\|_{\infty}$.

On the other hand, $f_i^* = \sum_{r=1}^{\infty} f_{\varepsilon|\mathcal{I}}\left(\lambda\mathbf{g}_i\mathbf{u} + \psi_i - a_r\right) \leqslant \max_{u\in\mathbb{R}}\sum_{k=1}^{\infty} f_{\varepsilon|\mathcal{I}}\left(u - a_r\right) = \dfrac{1}{B_c}$. Thus, $\|\partial\mathbf{L}\left(\mathbf{u}\right)/\partial\mathbf{u}'\|_{\infty} \leqslant \dfrac{\lambda\|\mathbf{G}\|_{\infty}}{B_c} < 1$ by Assumption 2.4. As a result, there is a unique rational expected outcome $\mathbf{y}^{e*}$ is the unique solution of $\mathbf{y}^e = \mathbf{L}(\mathbf{y}^e)$. The unique rational expected outcome yields a unique belief system $\mathbf{p}^*$ using Equation (3).

A.4 Proof of Proposition 3.1

The proof of Proposition 3.1 is done in several steps. If $\mathcal{B}$ and $\tilde{\mathcal{B}}$ are observationally equivalent, then $p^*_{i0} = \tilde{p}^*_{i0}$, which implies that $\Phi(\lambda\bar{y}^{e*}_i + \boldsymbol{z}'_i\boldsymbol{\Gamma}) = \Phi(\tilde{\lambda}\bar{y}^{e*}_i + \boldsymbol{z}'_i\tilde{\boldsymbol{\Gamma}})$. As $\mathbf{Z}$ is a full rank matrix (Assumption 3.4), this means that $\bar{y}^{e*}_i = \boldsymbol{z}'_i\frac{\boldsymbol{\Gamma}-\tilde{\boldsymbol{\Gamma}}}{\tilde{\lambda}-\lambda}$, where λ is necessarily different from $\tilde{\lambda}$

I first show that the previous equation is not possible. I employ a reductio ad absurdum reasoning approach. For any $i \in \mathcal{V}$, the condition $\bar{y}^{e*}_i = \boldsymbol{z}'_i\frac{\boldsymbol{\Gamma}-\tilde{\boldsymbol{\Gamma}}}{\tilde{\lambda}-\lambda}$ is equivalent to

$$\bar{y}^{e*}_i = \check{\alpha} + \boldsymbol{x}'_i\check{\boldsymbol{\beta}} + \bar{\boldsymbol{x}}'_i\check{\boldsymbol{\gamma}}, \tag{A.1}$$

where $\check{\alpha} = \frac{\alpha-\tilde{\alpha}}{\tilde{\lambda}-\lambda}$, $\check{\boldsymbol{\beta}} = \frac{\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}}{\tilde{\lambda}-\lambda}$, and $\check{\boldsymbol{\gamma}} = \frac{\boldsymbol{\gamma}-\tilde{\boldsymbol{\gamma}}}{\tilde{\lambda}-\lambda}$. Importantly, Equation (A.1) must be true for any $i \in \mathcal{V}$, regardless of the values taken by $\bar{y}^{e*}_i$, $\boldsymbol{x}_i$, and $\bar{\boldsymbol{x}}_i$. The parameters $\check{\alpha}$, $\check{\boldsymbol{\beta}}$, and $\check{\boldsymbol{\gamma}}$ do not depend on the values taken by $\bar{y}^{e*}_i$, $\boldsymbol{x}_i$, and $\bar{\boldsymbol{x}}_i$ since $\mathcal{B}$ and $\tilde{\mathcal{B}}$ are two arbitrary parameter collections. Equation (A.1) is not possible because it only involves agent i (through the term $\boldsymbol{x}_i$) and i's friends (through the terms $\bar{y}^{e*}_i$ and $\bar{\boldsymbol{x}}_i$), but not i's friends' friends who are not friends. Therefore, it does not directly capture changes in the characteristics of i's friends' friends who are not friends.

To be specific, let us take the total differential of $\bar{y}^{e*}_i$ (keeping $\mathbf{G}$ and the parameters fixed), Equation (A.1) implies that $d\bar{y}^{e*}_i = (d\boldsymbol{x}_i)'\check{\boldsymbol{\beta}} + (d\bar{\boldsymbol{x}}_i)'\check{\boldsymbol{\gamma}}$. Now, assume an increase in $\boldsymbol{x}_l$ for some $l \neq i$ who is not i's friend, *ceteris paribus*, i.e., $d\boldsymbol{x}_l > 0$, $d\boldsymbol{x}_i = 0$, and $d\bar{\boldsymbol{x}}_i = 0$. This implies that $d\bar{y}^{e*}_i = 0$. Since $\beta_{k_0}\gamma_{k_0} \geqslant 0$ and $\gamma_{k_0} \neq 0$, an increase in $\boldsymbol{x}_l$ cannot imply both $dy^e < 0$ for some agents and $dy^{e*} > 0$ for others. Thus, $d\bar{y}^{e*}_i = 0$ necessarily means that $dy^{e*}_j = 0$ for any j who is i's friend. Put differently, $\boldsymbol{x}_l$ cannot influence y^{e*}_j if j is i's friend and l is not. This cannot hold under Condition (ii) of Assumption 3.5 because l could be j's friend.

I now show the identification of λ, α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$. The condition $p^*_{i0} = \tilde{p}^*_{i0}$ for any $i \in \mathcal{V}$ implies that $\alpha + \lambda y^{e*}_i + \boldsymbol{x}'_i\boldsymbol{\beta} + \bar{\boldsymbol{x}}'_i\boldsymbol{\gamma} = \tilde{\alpha} + \tilde{\lambda}y^{e*}_i + \boldsymbol{x}'_i\tilde{\boldsymbol{\beta}} + \bar{\boldsymbol{x}}'_i\tilde{\boldsymbol{\gamma}}$. As $\tilde{\mathbf{Z}}^* = [\bar{\mathbf{y}}^{e*}, \mathbf{Z}]$ is a full rank matrix, then $\lambda = \tilde{\lambda}$, $\alpha = \tilde{\alpha}$, $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$, and $\boldsymbol{\gamma} = \tilde{\boldsymbol{\gamma}}$. Therefore, λ, α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are identified.

I now turn to the proof of the identification of $\boldsymbol{\delta}$, $\bar{\delta}$, ρ, and $\bar{R}$. Assume that $\boldsymbol{\delta}$, $\bar{\delta}$, ρ, and $\bar{R}$ are observationally equivalent to $\tilde{\boldsymbol{\delta}} = (\tilde{\delta}_2, \dots, \tilde{\delta}_{\bar{R}})'$, $\tilde{\bar{\delta}}$, $\tilde{\rho}$, and $\tilde{\bar{R}}$. As λ, α, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ are identified, 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 $\sum_{k=1}^{r}\delta_k = \sum_{k=1}^{r}\tilde{\delta}_k$ for any $r \geqslant 1$ suggests that $\delta_r = \tilde{\delta}_r$. 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, $\bar{\delta}$ and ρ are identified. 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 $r = \bar{R}$ because $\bar{R}$ is the smallest R under which Assumption 3.1 holds. Thus, $\bar{R} = \tilde{\bar{R}}$. Finally $\boldsymbol{\delta}$ is identified because $\bar{R} = \tilde{\bar{R}}$ and $\delta_r = \tilde{\delta}_r$.

A.5 Proof of Proposition 3.2

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 $\boldsymbol{\theta}$. I have $\mathbf{y}^{e*} = \mathbf{L}(\boldsymbol{\theta}^0, \mathbf{y}^{e*})$. I adapt the regulatory conditions of the NPL estimator set by Aguirregabiria and Mira (2007), henceforth AM07, to my framework.

Assumption A.1. $\mathrm{supp}(\boldsymbol{x}_i) \subset \mathbb{X}$ *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 $\mathrm{supp}(\boldsymbol{x}_i)$ 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:
$\mathcal{L}_0(\boldsymbol{\theta}, \mathbf{y}^e) = \mathbb{E}(\mathcal{L}_n(\boldsymbol{\theta}, \mathbf{y}^e)|\boldsymbol{\chi}_n)$, $\tilde{\boldsymbol{\theta}}_0(\mathbf{y}^e, \bar{R}) = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}(\bar{R})} \mathcal{L}_0(\boldsymbol{\theta}, \mathbf{y}^e)$, $\boldsymbol{\phi}_0(\mathbf{y}^e, \bar{R}) = \mathbf{L}(\tilde{\boldsymbol{\theta}}_0(\mathbf{y}^e, \bar{R}), \mathbf{y}^e)$, and $\mathcal{A}_0(\bar{R}) = \{(\boldsymbol{\theta}, \mathbf{y}^e) \in \boldsymbol{\Theta}(\bar{R}) \times [0, \overset{o}{y}]^n$, 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 $\mathbf{L}$ 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.*[24] *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)*

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* $\bar{R} \geqslant 1$*,* $\check{\boldsymbol{\theta}}_n(\bar{R}) \in \mathrm{int}(\boldsymbol{\Theta}(\bar{R}))$*.*

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* $\mathcal{A}_0(\bar{R})$*.*

[24]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.

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 $(\boldsymbol{\theta}^0, \mathbf{y}^{e*})$ 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), $\mathcal{L}_0(\boldsymbol{\theta}^0, \mathbf{y}^{e*})$ 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 $(\boldsymbol{\theta}^0, \mathbf{y}^{e*}) \in \mathcal{A}(\hat{\bar{R}})$. Hence, $(\check{\boldsymbol{\theta}}_n(\hat{\bar{R}}), \check{\mathbf{y}}_n^e(\hat{\bar{R}})) = (\boldsymbol{\theta}^0, \mathbf{y}^{e*})$ 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 $\mathcal{L}_{n,i}(\boldsymbol{\theta}, \mathbf{y}^e) = \sum_{r=0}^{\infty} d_{ir} \log\left(\Phi\left(\lambda \bar{y}_i^e + \boldsymbol{z}_i'\boldsymbol{\Gamma} - a_r\right) - \Phi\left(\lambda \bar{y}_i^e + \boldsymbol{z}_i'\boldsymbol{\Gamma} - a_{r+1}\right)\right)$. 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 $\boldsymbol{\chi}_n$.[25] 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 $\mathbf{H}_{2,n} \to$

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

$\mathbf{H}_{2,0}$ 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.4 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 varies from 0 to 33. However, most students declare that they participate in fewer than 10 extracurricular activities (see Figure B.1).

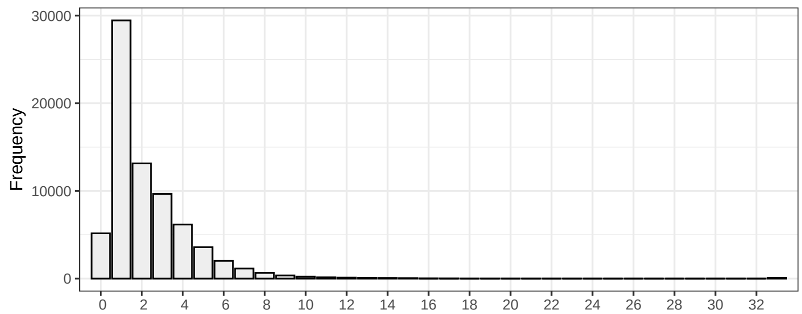

Figure B.1: Distribution of the number of extracurricular activities

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, Elie Tamer, 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.

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.

Aradillas-Lopez, A. (2010): "Semiparametric estimation of a simultaneous game with incomplete information," *Journal of Econometrics*, 157, 409–431.

Aradillas-Lopez, A. and A. M. Rosen (2022): "Inference in ordered response games with complete information," *Journal of Econometrics*, 226, 451–476.

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 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.

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

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.

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.

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.

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.

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.

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.

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

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.