2020-13

A Count Data Model with Social Interactions

Elysée Aristide Houndetoungan

Octobre / October 2020

Centre de recherche sur les risques
les enjeux économiques et les politiques publiques

www.crrep.ca

ABSTRACT

I present a model of peer effects in which the dependent variable takes integer values. I present an incomplete information game rationalizing the model, and I provide sufficient conditions under which the equilibrium of the game is unique. I estimate the model's parameters using the Nested Partial Likelihood method. I show that the counting nature of the dependent variable is important and that assuming incorrectly that it is continuous significantly underestimates the peer effects. I estimate peer effects on the the number of extracurricular activities in which students are enrolled. Increasing the number of activities in which friends are enrolled by one implies an increase of 0.295 in the number of activities in which students are enrolled, when controlling for network endogeneity. Ignoring the endogeneity of the network overestimates the peer effects.

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

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

Elysée Aristide Houndetoungan : Université Laval, Department of Economics.

I would like to thank Vincent Boucher for his helpful comments and insights. I would also like to thank Bernard Fortin, Arnaud Dufays, Luc Bissonnette, and Maripier Isabelle for helpful comments and discussions. Thank you also to the participants of the Applied Young Economists webinar at Monash University. 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 github.com/ahoundetoungan/CDatanet.
This research uses data from Add Health, a program directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by Grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is given to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain Add Health data files is available on the Add Health website (www.cpc.unc.edu/addhealth). No direct support was received for this research.

1 Introduction

There is a large and growing literature on peer effects in economics.[1] Recent contributions include, among others, models for limited dependent variables, including binary (e.g., Brock and Durlauf, 2001; Lee et al., 2014; Liu, 2019), ordered (e.g., Boucher et al., 2018), multinomial (e.g., Guerra and Mohnen, 2017), and censored (e.g., Xu and Lee, 2015b) variables. To my knowledge, however, there are no existing models for count variables with microeconomic foundations, despite these variables being prevalent in survey data (e.g., Duncan et al., 2005; Daniels and Leaper, 2006; Mays et al., 2010; Ali et al., 2011; Liu et al., 2012; Fujimoto and Valente, 2013; Liu et al., 2014; Fortin and Yazbeck, 2015; Boucher, 2016; Lee et al., 2020).

In this paper, I present a static game with incomplete information (see Harsanyi, 1967; Osborne and Rubinstein, 1994) to examine a network model in which the dependent variable takes unbounded integer values (0, 1, 2, ...). An example of variable is the number of occurrences of an event[2] in a constant period. The model generalizes the rational expectations model of Lee et al. (2014), which is used to study peer effects on binary data. I show that the model's parameters can be estimated using the Nested Partial Likelihood (NPL) method (Aguirregabiria and Mira, 2007). I show that the counting nature of the dependent variable is important and that assuming incorrectly that the dependent variable is continuous significantly underestimates the peer effects. 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). I control for network endogeneity. I find that ignoring the endogeneity of the network overestimates the peer effects. Finally, I present an easy-to-use R package—named `CDatanet`—for implementing the estimator and the network formation model.[3]

I present a microeconomic model in which individuals have private information and in which they can simultaneously choose their strategy. Individuals interact through a directed network and are influenced by their belief over the choice of their peers. As in many discrete games (e.g., Xu and Lee, 2015a; Liu, 2019), I assume that individuals do not directly choose the observed integer outcome. Instead, they choose a latent variable that can be interpreted as an intention. This latent variable determines the observed integer outcome (see also Maddala, 1986; Cameron and Trivedi, 2013).
I provide sufficient conditions under which the model game has a unique Bayesian Nash Equilibrium (BNE). When the distribution of the outcome is almost degenerated such that the outcome takes only two values, the structure of the game and the BNE are similar to Lee et al. (2014). The individuals' beliefs represent a discrete distribution over the infinite set of possible count choices $\{0, 1, 2, \dots\}$. As

[1] For recent reviews, see Boucher and Fortin (2016), De Paula (2017), and Bramoullé et al. (2019).
[2] Number of cigarettes smoked, frequency of restaurant visits, frequency of participation in activities, etc.
[3] The package is available at github.com/ahoundetoungan/CDatanet.

discussed by Reeves and Wellman (2012), the equilibrium belief computation may be cumbersome when the strategy space is infinite. I show that the BNE only depends on the average belief (expected outcome), which does not require computing the entire distribution of beliefs.

To estimate the model parameters, I rely on the Nested Pseudo Likelihood (NPL) algorithm proposed by Aguirregabiria and Mira (2007). The estimation process is straightforward and can be readily implemented. Moreover, it does not require computing the game equilibrium. I show that the estimator is consistent, and I study its limit distribution.

Using Monte Carlo simulations, I show that modeling count data using a misspecified continuous model, such as the Spatial Autoregressive Tobit (SART) model (see Xu and Lee, 2015b) or the standard Spatial Autoregressive (SAR) model (see Lee, 2004), significantly underestimates the peer effects. A similar result (in a context of no social interactions) is also discussed by Cameron and Trivedi (2013) as well as Hellerstein and Mendelsohn (1993). Ignoring the integer and the left-censoring nature of count data leads to biased estimations.

I use the Add Health data to estimate peer effects on the number of extracurricular activities in which students are enrolled. I find that increasing the number of activities in which a student's friends are enrolled by one implies an increase in the number of activities in which the student is enrolled by 0.295. As in the Monte Carlo study, I find that the SART and the SAR models underestimate peer effects at 0.141 and 0.166, respectively.

I control for the endogeneity of the network using a two-stage estimation strategy. In the first stage, I consider an dyadic linking model in which the probability of link formation between two students depends, among others, on their gregariousness (see Graham, 2017; Breza et al., 2020). Using a Markov Chain Monte Carlo (MCMC) approach, I simulate the posterior distribution of this gregariousness. In the second stage, the estimator of gregariousness is included in the count data model as a supplementary explanatory variable.[4] I find that the network is endogenous and that ignoring the endogeneity overestimates peer effects at 0.363.

This paper contributes to the literature on social interaction models for limited dependent variables. The existing models deal with binary (e.g., Brock and Durlauf, 2001; Soetevent and Kooreman, 2007; Lee et al., 2014; Xu and Lee, 2015a; Liu, 2019), censored (e.g., Xu and Lee, 2015b), and multinomial outcomes (e.g., Guerra and Mohnen, 2017). As mentioned above, the model generalizes Lee et al. (2014) and the number of values taken by the outcome is not bounded. Recently, Boucher et al. (2018) studied peer effects on the formation of beliefs regarding college participation. The outcome is polytomous ordered. However, as the number of parameters increases with the number of values taken by the outcome, the estimation strategy of Boucher et al. (2018) could raise numerical issues when the

[4] I use the posterior distribution of the estimator of gregariousness to account for the uncertainty related to first-stage estimation in the second stage.

dependent variable takes many values.

The paper contributes to the extensive literature on social interactions by being the first to deal with count variables. Existing papers studying the influence of social networks on a count outcome rely on linear-in-means models in which the outcome is assumed to be continuous (e.g., Duncan et al., 2005; Daniels and Leaper, 2006; Mays et al., 2010; Ali et al., 2011; Liu et al., 2012, 2014; Fortin and Yazbeck, 2015; Boucher, 2016; Lee et al., 2020). As shown through the Monte Carlo study (see Section 4), peer effects may be underestimated when a continuous variable model is used (see also Cameron and Trivedi, 2013; Hellerstein and Mendelsohn, 1993).

Importantly, in the literature on spatial autoregressive models for limited dependent variables, cases of count data have been studied (e.g, Karlis, 2003; Liesenfeld et al., 2016; Inouye et al., 2017; Glaser, 2017). These papers consider reduced form equations in which the dependent count variable is spatially autocorrelated. However, the models are not based on any process (game) that explains how the individuals choose their strategy, and thus how they are influenced by their peers. Therefore, the reduced form cannot be interpreted as a best-response function, and the spatial dependence parameter cannot be interpreted as peer effects.

The paper also contributes to the literature on peer effects models with endogenous networks. Goldsmith-Pinkham and Imbens (2013) as well as Hsieh and Lee (2016) consider a Bayesian hierarchical model to control for endogeneity. They use a MCMC approach to jointly simulate from the posterior distribution of the network formation model parameters and the outcome model parameters. While this method is efficient as the estimation is done in a single step, it can be cumbersome to implement with a non-linear outcome model. In this paper, the method used to control for endogeneity can be easily implemented regardless of the network formation and the outcome models. The network formation model is estimated, in a first stage, separately from the outcome model estimation. Moreover, I provide a way to properly estimate the variance of the estimator of the outcome model, which takes into account the uncertainty of the estimation at the first stage.[5]

The remainder of the paper is organized as follows. Section 2 presents the microeconomic foundation of the model based on an incomplete information network game. Section 3 addresses the identification and the estimation of the model parameters. The link between this new model and the SART model is also discussed. 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 discusses some limits, some areas for future research, and some general implications of the results. Section 7 concludes this paper.

[5]The approach used to correct the variance at the second stage is similar in spirit to that of Johnsson and Moon (2015).

2 Incomplete Information Network Game

I present a game of incomplete information with social interactions. Let $\mathcal{V} = \{1, \dots, n\}$ be a set of n players indexed by i and y_i, the observed integer outcome of player i (e.g., the number of cigarettes smoked per day or per week). As in Xu and Lee (2015a) and Liu (2019), I assume that the players do not directly choose y_i. Instead, they choose y_i^*, a latent variable that determines the observed outcome y_i. This latent variable can be interpreted as an intention that leads to the observed choice y_i (see Maddala, 1986).

I assume that y_i^* and y_i are linked as follows:

Assumption 1. *Let $(a_q)_{q\in\mathbb{N}}$ be a sequence given by $a_0 = -\infty$, $a_1 \in \mathbb{R}$, and $a_q = a_1 + \gamma(q-1)$ for $q \in \mathbb{N}^*$ and $\gamma \in \mathbb{R}_+^*$. If $y_i^* \in (a_q, a_{q+1}]$, then $y_i = q$.*

The outcome y_i is called the *count variable* or *count data*. As in a binary game (e.g., Liu, 2019), Assumption 1 sets $y_i = 0$, if y_i^* is not greater than some real value a_1. When $y_i^* > a_1$, Assumption 1 implies that there are increasing boundaries a_1, a_2, $\dots$, such that $y_i = q$, if $y_i^* \in (a_q, a_{q+1}]$. A similar assumption is also set to link a polytomous ordered variable to a latent variable (e.g., Amemiya, 1981; Baetschmann et al., 2015; Boucher et al., 2018).

Assumption 1 restricts the boundaries to be equally spaced from a_1; that is, a_1, $a_1 + \gamma$, $a_1 + 2\gamma$, and so on. This is stronger than the usual assumption for an ordered model, which allows the boundary increment to vary (see Amemiya, 1981). However, two important points motivate such simplification. First, if the increment varies, then the number of unknown parameters increases with the number of values taken by y_i. In practice, estimating the model can be cumbersome when the outcome takes many values (see Boucher et al., 2018). As the count variable y_i is unbounded, Assumption 1 fixes this curse of dimensionality. Second, it is intuitively natural to set that the boundaries increase uniformly by γ as the outcome y_i increases uniformly by 1. This is different than for an ordered polytomous variable, where the values are not numeric but can only be ranked. As a result, the increase between two consecutive values is likely to be variable.

Interestingly, Assumption 1 also generalizes the binary outcome game of Lee et al. (2014). Indeed, if $\gamma = \infty$, then $a_r = \infty$ for $r \geq 2$. In that case, y_i can only take 2 values: $y_i = 0$, if $y_i^* \leq a_1$, and $y_i = 1$ otherwise.

Individuals interact through a directed network. Let $\mathbf{G} = [g_{ij}]$ be an $n \times n$ adjacency matrix, where the (i, j)-th element is non-negative and captures the proximity of the individuals i and j in the network. I define the peers of individual i as the set of individuals $\mathcal{V}_i = \{j, g_{ij} > 0\}$. By convention, nobody interacts with himself/herself, that is $g_{ii} = 0\ \forall\ i \in \mathcal{V}$.

I assume that the individuals' preferences can be characterized by the following linear-quadratic utility function:[6]

$$\mathcal{U}_i = \underbrace{(\psi_i + \varepsilon_i)\, y_i^* - \frac{y_i^{*2}}{2}}_{\text{private sub-utility}} + \underbrace{\lambda y_i^* \sum_{j \neq i} g_{ij} y_j}_{\text{social sub-utility}}\ , \tag{1}$$

where ψ_i, $\lambda \in \mathbb{R}$, and ε_i is an idiosyncratic shock. The first two terms of the utility function (1) are the private subutility, in which $-\dfrac{1}{2}y_i^{*2}$ is the intention cost, and $\psi_i + \varepsilon_i$, an individual characteristic (own marginal benefit). The marginal benefit is composed of ψ_i, an observable part by all players, and ε_i, only observed by i. I assume that ε_i is identically and independently distributed over i with a symmetric distribution, common knowledge among all the players. The third term is a social sub-utility. It depends on the intention y_i^*, the average of the peers' outcomes $\sum_{j \neq i} g_{ij} y_j$, and the peer effects parameter λ. Importantly, each individual i chooses the intention y_i^*, but each is affected by their peers' outcomes y_j, $j \in \mathcal{V}$. As argued by Fortin and Boucher (2015), the utility function (1) describes *complementarity* in social interactions if $\lambda > 0$ and *substitutability* in social interactions if $\lambda < 0$. A similar utility function is used by Liu (2019) to model bivariate binary outcomes with social interactions.

Individuals observe neither the private information ε_j of their peers, nor do they then observe the outcome y_j of their peers. The utility function (1) characterizes a game of incomplete information (Bayesian game) in which the players form beliefs regarding their peers' outcomes. Moreover, as the players know the common distribution of their type, they form rational beliefs (see Lee et al., 2014; Liu, 2019); that is, the probability that any player i puts on the event $\{y_j = q\}$, $q \in \mathbb{N}$, and $i \neq j$ corresponds to the *true* probability of realization of the event.

Individuals simultaneously choose their strategy y_i^* as to maximize their expected utilities.

$$\mathbf{E}\left(\mathcal{U}_i | y_i^*, \varepsilon_i, \lambda, \boldsymbol{\psi}, \mathbf{G}\right) = (\psi_i + \varepsilon_i)\, y_i^* - \frac{y_i^{*2}}{2} + \lambda y_i^* \sum_{j \neq i} \sum_{r=0}^{\infty} r g_{ij} p_{jr}, \tag{2}$$

where $\forall\ i \in \mathcal{V},\ \boldsymbol{\psi} = (\psi_1, \dots \psi_n)'$, and $\forall\ i \in \mathcal{V},\ q \in \mathbb{N},\ p_{iq} = \mathcal{P}(y_i = q | \lambda, \boldsymbol{\psi}, \mathbf{G})$. From the first-order conditions (f.o.cs) of the expected utility maximization,

$$y_i^* = \lambda \mathbf{g}_i \bar{\mathbf{y}} + \psi_i + \varepsilon_i, \tag{3}$$

where $\forall\ i \in \mathcal{V}$, $\mathbf{g}_i = (g_{i1} \dots g_{in})$, $\bar{\mathbf{y}} = (\bar{y}_1\ \dots \bar{y}_n)'$, and $\bar{y}_i = \sum_{r=0}^{\infty} r p_{ir}$ is the expectation of the outcome y_i.

[6]The linear-quadratic specification of the utility function is common for network games (e.g., Ballester et al., 2006; Calvó-Armengol et al., 2009).

Let $\mathbf{y}^* = (y_1^* \ \dots y_n^*)'$, and $\boldsymbol{\varepsilon} = (\varepsilon_1 \ \dots \ \varepsilon_n)'$. The f.o.cs (3) is also equivalent to

$$\mathbf{y}^* = \lambda \mathbf{G}\bar{\mathbf{y}} + \boldsymbol{\psi} + \boldsymbol{\varepsilon}. \tag{4}$$

Let F_ε be the cumulative distribution function (cdf) of ε_i. As $p_{iq} = \mathcal{P}(y_i^* \in (a_q, a_{q+1})|\lambda, \boldsymbol{\psi}, \mathbf{G})\ \forall\ i \in \mathcal{V}$, $q \in \mathbb{N}$, the f.o.cs (3) implies

$$p_{iq} = \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \psi_i - a_q\right) - \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \psi_i - a_{q+1}\right). \tag{5}$$

It follows that any vector of beliefs $(p_{iq})_{\substack{q\in\mathbb{N}\\ i\in\mathcal{V}}}$ characterizes a BNE (see Osborne and Rubinstein, 1994) of the network game with the utility (1) if

$$\begin{cases} p_{iq} = \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \psi_i - a_q\right) - \mathrm{F}_\varepsilon\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \psi_i - a_{q+1}\right), \\ \text{where } \ \bar{\mathbf{y}} = (\bar{y}_1 \ \dots \bar{y}_n) \text{ and } \bar{y}_i = \displaystyle\sum_{r=0}^{\infty} r p_{ir}. \end{cases} \tag{6}$$

The BNE has an interesting characterization. Indeed, Equation (6) shows that there is a bijective correspondence between the vector of beliefs $(p_{iq})_{\substack{q\in\mathbb{N}\\ i\in\mathcal{V}}}$ and the expected outcome $(\bar{y}_i)_{i\in\mathcal{V}}$ at equilibrium; that is, the knowledge of the expected outcome is sufficient to compute the beliefs and vice versa. This result has a very useful implication: to prove the uniqueness of the equilibrium belief, it is sufficient to prove that the expected equilibrium outcome is unique.

In the following assumption, I define sufficient conditions that ensure the existence and the uniqueness of the BNE.

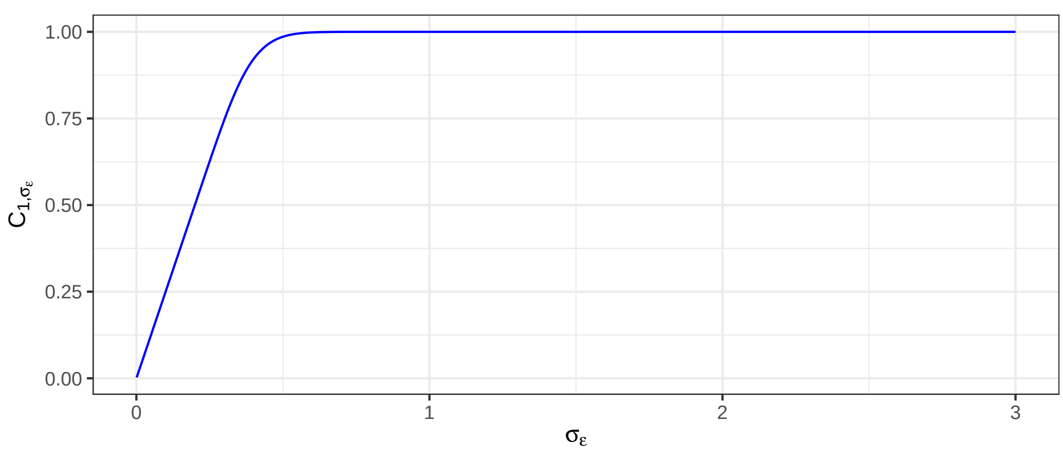
Assumption 2. *(i)* $\varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$ *and (ii)* $|\lambda| < \dfrac{C_{\gamma,\sigma_\varepsilon}}{||\mathbf{G}||_\infty}$, *where* $C_{\gamma,\sigma_\varepsilon} = \dfrac{\sigma_\varepsilon}{\phi\left(0\right) + 2\sum_{k=1}^{\infty}\phi\left(\frac{\gamma k}{\sigma_\varepsilon}\right)}$ *and* ϕ *is the probability density function (pdf) of* $\mathcal{N}\left(0,1\right)$.

Note that Assumption 2 (i) imposes that ε_i's distribution is normal. This is not a necessary condition, but the proof of the equilibrium uniqueness requires to be specific about the distribution of ε_i. Other distributions, such as logistic, can be used by adapting the proof. The normal distribution is chosen to facilitate the comparison of this model with the SART or SAR models (see Section 3.3) and also to deal with the endogeneity of the network (see Section 5.3).

Assumption 2 (ii) gives the upper bound of $|\lambda|$. This generalizes the restriction imposed on $|\lambda|$ in other rational expectations models for binary data. Indeed, due to an identification issue in the binary model, it is assumed that $\sigma_\varepsilon = 1$. If $\gamma = \infty$, then Assumption 2 (ii) implies that $|\lambda| < \dfrac{1}{||\mathbf{G}||_\infty \phi(0)}$, which is the restriction set on $|\lambda|$ in Lee et al. (2014) and Liu (2019). In addition, when $\mathbf{G}$ is row-normalized, $||\mathbf{G}||_\infty = 1$, and the restriction on $|\lambda|$ implies that $|\lambda| < \dfrac{1}{\sqrt{2\pi}}$. However, it is generally assumed in practice that $|\lambda| < 1$; that is, individuals will not experience an increase in their intention/outcome greater than the increase in their peers' outcomes. The restriction under the binary case is not stronger than $|\lambda| < 1$.

Importantly, Assumption 2 (ii) is still somewhat weak when $\gamma < \infty$ and $||\mathbf{G}||_\infty = 1$. As shown in Section 3.1, γ and σ_ε are closely linked and cannot be identified. Thus, to analyze the upper bound $C_{\gamma,\sigma_\varepsilon}$, I set $\gamma = 1$. Figure 1 plots C_{1,σ_ε} as a function of σ_ε.[7] One can notice that $C_{1,\sigma_\varepsilon} \approx 1$, if $\sigma_\varepsilon > 0.5$. In that case, Assumption 2 (ii) is not much stronger than $|\lambda| < 1$. In contrast, when σ_ε is low, Assumption 2 (ii) implies a stronger restriction. For example, if $\sigma_\varepsilon = 0.3$, then $|\lambda| < 0.746$. To understand why the upper bound of $|\lambda|$ must decrease together with σ_ε, notice that $\dfrac{\lambda}{\sigma_\varepsilon}$ is a multiplicative term in the marginal effect of peer expected outcomes $\mathbf{g}_i\bar{\mathbf{y}}$ on the individual expected outcome $\bar{y}_i$. If σ_ε decreases when λ is fixed, the marginal effect explodes (see Appendix B.2) and may involve multiple equilibria.

Figure 1: C_{1,σ_ε} (upper bound of λ when $\gamma = 1$ and $||\mathbf{G}||_\infty = 1$) as a function of σ_ε



However, the condition $\sigma_\varepsilon < 0.5$ is likely violated in practice when $\gamma = 1$. Indeed, σ_ε is the standard deviation of y_i^*, conditional on ψ_i and $\mathbf{g}_i\bar{\mathbf{y}}$. As y_i^* takes values in disjoint intervals of range γ, the standard deviation must be sufficiently large for y_i^* to span several intervals. If σ_ε is too low, y_i will not vary given ψ_i and $\mathbf{g}_i\bar{\mathbf{y}}$.

In practice, a sufficient condition for the equilibrium uniqueness dealing with $|\lambda| < 1$ is that the econometrician observes several different choices of y_i, given the observable characteristics ψ_i and average outcomes of the peers $\mathbf{g}_i\bar{\mathbf{y}}$. This likely ensures that $\sigma_\varepsilon > 0.5$.

Under Assumption 2, $\mathrm{F}_\varepsilon(.) = \Phi\left(\dfrac{.}{\sigma_\varepsilon}\right)$, where Φ is the cdf of $\mathcal{N}(0,1)$. The following theorem establishes the existence and uniqueness of the pure strategy BNE of the incomplete information network game.

[7] I show in Appendix A.3 that $C_{\gamma,\sigma_\varepsilon}$ can be evaluated using the third Theta function (see Section 2 in Bellman, 2013) available in most software.

Theorem 1. *Let* $\mathbf{L}(\bar{\mathbf{y}}) = (\ell_1(\bar{\mathbf{y}}) \ \dots \ell_n(\bar{\mathbf{y}}))'$, *where*

$$\ell_i(\bar{\mathbf{y}}) = \sum_{r=1}^{\infty} \Phi\left(\frac{\lambda \mathbf{g}_i \bar{\mathbf{y}} + \psi_i - a_r}{\sigma_\varepsilon}\right) \quad \textit{for all } i \in \mathcal{V}. \tag{7}$$

Under Assumptions 1 and 2, the incomplete information network game with the utility (1) has a unique pure strategy BNE with the equilibrium strategy profile $\mathbf{y}^{e*}$, *given by* $\mathbf{y}^{e*} = \lambda \mathbf{G}\bar{\mathbf{y}}^e + \boldsymbol{\psi} + \boldsymbol{\varepsilon}$, *where* $\bar{\mathbf{y}}^e = (\bar{y}_1^e \dots \bar{y}_n^e)$ *is the unique solution of* $\bar{\mathbf{y}} = \mathbf{L}(\bar{\mathbf{y}})$.

Proof. See Appendix A. □

There are two important remarks concerning Theorem 1. First, the model generalizes the rational expectations model proposed by Lee et al. (2014) for discrete binary outcomes. Indeed, if $\gamma = \infty$, then $p_{ir} = 0$ for $r \geq 2$ and $i \in \mathcal{V}$. As a result, $\bar{y}_i = \sum_{r=0}^{\infty} r p_{ir} = p_{i1} \ \forall \ i \in \mathcal{V}$, and $\bar{\mathbf{y}} = \mathbf{p}_1$, where $\mathbf{p}_1 = (p_{11} \ \dots \ p_{n1})$. Under these considerations, Assumption 2 still ensures that the game has a unique BNE with the equilibrium strategy $\mathbf{y}^{e*}$, given by $\mathbf{y}^{e*} = \lambda \mathbf{G}\mathbf{p}_1^e + \boldsymbol{\psi} + \boldsymbol{\varepsilon}$, where $p_{i1}^e = \Phi\left(\dfrac{\lambda \mathbf{g}_i \mathbf{p}_1^e + \psi_i - a_1}{\sigma_\varepsilon}\right)$ for all $i \in \mathcal{V}$. For $\sigma_\varepsilon = 1$, this characterization of the equilibrium is the same as that of Lee et al. (2014).

Second, the equilibrium belief is not necessary to compute the equilibrium strategy. The knowledge of the first moment of y_i at equilibrium is sufficient to compute the equilibrium strategy and the equilibrium belief. Although the outcome y_i takes an infinite number of values, the equilibrium computation requires solving a fixed point problem in $\mathbb{R}_+^n$ as in other rational expectation discrete games with social interactions (see Lee et al., 2014; Guerra and Mohnen, 2017; Liu, 2019)

Theorem 1 guarantees that the mapping $\mathbf{L}$ has a unique fixed point, which is sufficient to compute the BNE. This also suggests using the Nested Pseudo Likelihood (NPL) algorithm proposed by Aguirregabiria and Mira (2007) to estimate the model. In the next section, I study the parameter identification and present the model estimation strategy.

3 Econometric Model

This section presents the identification and estimation of the model. It also discusses the link between the model and the SART model (Xu and Lee, 2015b) that is used when the data are left-censored at 0.

3.1 Identification

In this section, I describe restrictions on the model parameters that are necessary to ensure identifiability. Let $\boldsymbol{\psi} = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X} = (\mathbf{x}_1 \ \dots \ \mathbf{x}_n)'$ is an $n \times K$-dimensional matrix of explanatory variables, and $\boldsymbol{\beta}$ is a K-dimensional vector of unknown parameters. The matrix $\mathbf{X}$ may also include the average of the explanatory variables of the peers; that is, $\boldsymbol{\psi} = \tilde{\mathbf{X}}\boldsymbol{\beta}$, where $\tilde{\mathbf{X}} = [\mathbf{X},\ \mathbf{G}\mathbf{X}]$. The coefficients of $\mathbf{G}\mathbf{X}$ represent the contextual effects (Manski, 1993).

The BNE characterization (6) becomes

$$p_{iq} = \Phi\left(\frac{\lambda \mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} - a_q}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\lambda \mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} - a_{q+1}}{\sigma_\varepsilon}\right). \tag{8}$$

As $a_0 = -\infty$, and $a_q = a_1 + \gamma(q-1)$ for $q \in \mathbb{N}^*$,

$$p_{iq} = \begin{cases} 1 - \Phi\left(\dfrac{\lambda \mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} - a_1}{\sigma_\varepsilon}\right) & \text{if } q = 0, \\ \Phi\left(\dfrac{\lambda \mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} - a_1 - \gamma(q-1)}{\sigma_\varepsilon}\right) - \Phi\left(\dfrac{\lambda \mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} - a_1 - \gamma q}{\sigma_\varepsilon}\right) & \text{if } q \in \mathbb{N}^*. \end{cases} \tag{9}$$

Estimating the model requires additional restrictions on the parameters. Equation (9) poses two identification issues. First, Equation (9) does not change when λ, $\boldsymbol{\beta}$, a_1, γ, and σ_ε are multiplied by any positive number. To fix this identification issue, I set γ to one.[8] Second, if the explanatory variables include a constant, such that $\mathbf{x}_i'\boldsymbol{\beta} = \beta_1 + x_{2i}\beta_2 + \dots x_{Ki}\beta_K$, the parameters β_1 and a_1 cannot be identified because they enter the equation only through their difference. Therefore, I also set $a_1 = 0$. Following these restrictions, Assumption 1 can be simplified.

Assumption 1′. *Let $(a_q)_{q\in\mathbb{N}}$ be a sequence given by $a_0 = -\infty$, $a_q = q-1$ for $q \in \mathbb{N}^*$. If $y_i^* \in (a_q, a_{q+1}]$, then $y_i = q$.*

Under Assumptions 1′ and 2, the parameters $\boldsymbol{\theta} = \left(\lambda, \boldsymbol{\beta}', \sigma_\varepsilon\right)'$ are identified if $\mathbf{Z} = [\mathbf{G}\bar{\mathbf{y}},\ \mathbf{X}]$ is a full rank matrix. Indeed, given the adjacency matrix $\mathbf{G}$ and the exogenous variable $\mathbf{X}$, the parameters $\boldsymbol{\theta} = \left(\lambda, \boldsymbol{\beta}', \sigma_\varepsilon\right)'$ and the alternative parameters $\tilde{\boldsymbol{\theta}} = \left(\tilde{\lambda}, \tilde{\boldsymbol{\beta}}', \tilde{\sigma}_\varepsilon\right)'$ are equivalent if they lead to the same BNE equilibrium; that is $\bar{\mathbf{y}} = \tilde{\bar{\mathbf{y}}}$, where $\bar{\mathbf{y}}$ and $\tilde{\bar{\mathbf{y}}}$ are the expected outcomes associated with $\boldsymbol{\theta}$ and $\tilde{\boldsymbol{\theta}}$, respectively. In addition, Theorem 1 ensures that $\bar{\mathbf{y}}$ and $\tilde{\bar{\mathbf{y}}}$ are uniquely determined by the fixed point mappings. Then,

$$\begin{gathered} \bar{\mathbf{y}} = \sum_{r=1}^{\infty}\Phi\left(\frac{\lambda \mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} - a_r}{\sigma_\varepsilon}\right) = \sum_{r=1}^{\infty}\Phi\left(\frac{\tilde{\lambda} \mathbf{g}_i\tilde{\bar{\mathbf{y}}} + \mathbf{x}_i'\tilde{\boldsymbol{\beta}} - a_r}{\tilde{\sigma}_\varepsilon}\right),\ \forall\ i \in \mathcal{V}, \\ \left(\frac{\lambda}{\sigma_\varepsilon} - \frac{\tilde{\lambda}}{\tilde{\sigma}_\varepsilon}\right)\mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\left(\frac{\boldsymbol{\beta}}{\sigma_\varepsilon} - \frac{\tilde{\boldsymbol{\beta}}}{\tilde{\sigma}_\varepsilon}\right) + q\left(\frac{1}{\sigma_\varepsilon} - \frac{1}{\tilde{\sigma}_\varepsilon}\right) = 0,\ \forall\ i \in \mathcal{V}, q \in \mathbb{N}. \end{gathered} \tag{10}$$

As $\mathbf{Z}$ is a full rank matrix, it follows from Equation (10) that $\sigma_\varepsilon = \tilde{\sigma}_\varepsilon$, $\lambda = \tilde{\lambda}$, and $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$, except for

[8] Alternatively, I could also set σ_ε to one. However, this complicates the comparison of the model with the SAR and SART models. Moreover, the restriction $\gamma = 1$ excludes the binary cases for which $\gamma = \infty$. Therefore, the restrictions set in this section are only for the dependent variables defined as a counting variable (which excludes binary cases).

zero-measure cases (see Bramoullé et al., 2009). Therefore, $\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}$.

In the next section, I present the strategy used to estimate $\boldsymbol{\theta}$, and I study the limit distribution of the estimator.

3.2 Estimation

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

Let $\mathcal{L}$ be the pseudo likelihood[9] function in $(\boldsymbol{\theta}, \bar{\mathbf{y}})$, defined as

$$\mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}}) = \sum_{i=1}^{n}\sum_{r=0}^{\infty} d_{ir}\log(p_{ir}), \tag{11}$$

where $p_{iq} = \Phi\left(\dfrac{\lambda\mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} - a_q}{\sigma_\varepsilon}\right) - \Phi\left(\dfrac{\lambda\mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} - a_{q+1}}{\sigma_\varepsilon}\right)$ $\forall\ i \in \mathcal{V}$, $q \in \mathbb{N}$, and $d_{ir} = 1$, if $y_i = r$ and $d_{ir} = 0$ otherwise. As I set above that $\boldsymbol{\psi} = \mathbf{X}\boldsymbol{\beta}$, the mapping $\mathbf{L}$ can be redefined as $\mathbf{L}(\bar{\mathbf{y}}, \boldsymbol{\theta}) = (\ell_1(\bar{\mathbf{y}}, \boldsymbol{\theta})\ \dots \ell_n(\bar{\mathbf{y}}, \boldsymbol{\theta}))'$, where

$$\ell_i(\bar{\mathbf{y}}, \boldsymbol{\theta}) = \sum_{r=1}^{\infty}\Phi\left(\frac{\lambda\mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} - a_r}{\sigma_\varepsilon}\right) \quad \text{for all } i \in \mathcal{V}. \tag{12}$$

The NLP algorithm consists of starting with a proposal $\bar{\mathbf{y}}_0$ for $\bar{\mathbf{y}}$ and constructing a sequence of estimators $(\mathcal{Q}_m)_{m\geq 1}$, defined as $\mathcal{Q}_m = \{\boldsymbol{\theta}_m, \bar{\mathbf{y}}_m\}$ for $m \geq 1$, where $\boldsymbol{\theta}_m = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}}_{m-1})$ is the estimator of $\boldsymbol{\theta}$ at the m-th stage, and $\mathbf{y}_m = \mathbf{L}(\bar{\mathbf{y}}_{m-1}, \boldsymbol{\theta}_m)$ is the estimator of $\bar{\mathbf{y}}$ at the m-th stage. In other words, given the guess $\bar{\mathbf{y}}_0$, $\boldsymbol{\theta}_1 = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}}_0)$, and $\mathbf{y}_1 = \mathbf{L}(\bar{\mathbf{y}}_0, \boldsymbol{\theta}_1)$; then $\boldsymbol{\theta}_2 = \arg\max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}}_1)$, $\mathbf{y}_2 = \mathbf{L}(\bar{\mathbf{y}}_1, \boldsymbol{\theta}_2)$, ...

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

As shown by Kasahara and Shimotsu (2012), a key determinant of the convergence of the NPL algorithm is the contraction property of the fixed point mapping $\mathbf{L}$ guaranteed by Theorem 1. In practice, when $||\hat{\boldsymbol{\theta}}_M - \hat{\boldsymbol{\theta}}_{M-1}||_1$ and $||\hat{\bar{\mathbf{y}}}_M - \hat{\bar{\mathbf{y}}}_{M-1}||_1$ are less than some tolerance values (for example 10^{-6}), I set $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_M$ and $\hat{\bar{\mathbf{y}}} = \hat{\bar{\mathbf{y}}}_M$. Aguirregabiria and Mira (2007) prove that the NPL estimator is root-n

[9] This is a pseudo likelihood because it is defined for any $\boldsymbol{\theta}$ and $\bar{\mathbf{y}}$, where $\bar{\mathbf{y}}$ is not necessary for the equilibrium associated with $\boldsymbol{\theta}$ (see Aguirregabiria and Mira, 2007).

consistent and asymptotically normal. Their approach can be adapted to the current context. The convergence and the limit distribution of $\hat{\boldsymbol{\theta}}$ are given by the following proposition.

Proposition 1. *Under regularity conditions (see Proposition 2 of Aguirregabiria and Mira, 2007), the NLP estimator $\hat{\boldsymbol{\theta}}$ is consistent, and*

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(0,\ (\boldsymbol{\Sigma}_0 + \boldsymbol{\Omega}_0)^{-1}\boldsymbol{\Sigma}_0\left(\boldsymbol{\Sigma}_0' + \boldsymbol{\Omega}_0'\right)^{-1}\right), \tag{13}$$

where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$; $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Omega}_0$ are given in Appendix B.1.

Proof. See Appendix B.1. □

Some numerical aspects about the NLP estimator must be pointed out. First, the pseudo likelihood (11) involves an infinite sum. However, note that this does not pose any numerical issues because $d_{iq} = 0$ for $q \neq y_i$. Therefore, one only needs to compute p_{iy_i}; that is, one probability p_{iq} per individual at $q = y_i$. For this purpose, the pseudo likelihood can be redefined as $\mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}}) = \displaystyle\sum_{i=1}^{n}\log(p_{iy_i})$. Second, the mapping $\mathbf{L}$, which is used to compute the sequence $(\mathcal{Q}_m)$ and the asymptotic variance of $\hat{\boldsymbol{\theta}}$, also involves an infinite sum. However, note that the summed elements decrease exponentially. A very good approximation of these sums can be readily reached by only summing a few elements.

3.3 Link with the Spatial Autoregressive Tobit Model

In this section, I make the link between the count variable model and the SART model (see Xu and Lee, 2015b). Tobit models are also used for counting responses because the data are left-censored at 0 (e.g., Jones, 1989). The SART model is given as follows:

$$\begin{cases} y_i^* = \lambda \mathbf{g}_i \mathbf{y} + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \\ y_i = \max\{0;\ y_i^*\}. \end{cases} \tag{14}$$

Let us recall the f.o.cs (3).

$$y_i^* = \lambda \mathbf{g}_i \bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i. \tag{15}$$

Equations (14) and (15) differ because $\mathbf{g}_i\mathbf{y}$, the average of the peers' outcomes, is an explanatory variable in (14) whereas $\mathbf{g}_i\bar{\mathbf{y}}$, the expectation of the average of the peers' outcomes, as explanatory variable in (15). A misspecification of y_i as $\max\{0;\ y_i^*\}$ instead of Assumption 1′ is costless if the specification error is not correlated to the explanatory variable.[10] Fundamentally, the difference between

[10]In contrast, the specification of the dependent variable following the standard SAR model can involve substantial bias. Indeed, under the SAR model, $y_i = y_i^*$. Given that y_i cannot be negative, this specification can be a source of bias if the data contains many *zeros* (see Monte Carlo results in Section 4).

both models appears in $\mathbf{y}$ in (14) instead of $\bar{\mathbf{y}}$; that is, the expected outcome is approximated by the observed outcome. Although $\mathbf{y}$ and $\bar{\mathbf{y}}$ could be very similar, the approximation of a continuous variable by a discrete variable performs *poorly* in most cases when the count variable has a weak dispersion. Therefore, I expect that estimating a SART model with data generated following the count response model would involve classical bias: the smaller the dispersion of the dependent variable, the greater the bias.

I compare both models via Monte Carlo experiments (see Section 4). The results show that the SART model biases peer effects downward.

4 Monte Carlo Experiments

In this section, I conduct a Monte Carlo study to assess the performance of the NPL estimator in a finite sample. I also compare the model to the spatial autoregressive Tobit (SART) and the standard linear-in-mean spatial autoregressive (SAR) models.

I consider two types of data generating processes (DGP). The DGP of type A simulates many *zeros*,[11] whereas the DGP of type B simulates few *zeros*. In both cases, the latent variables y_i^* are defined as follows:

$$y_i^* = \lambda \mathbf{g}_i \bar{\mathbf{y}} + \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \gamma_1 \mathbf{g}_i \mathbf{x}_1 + \gamma_2 \mathbf{g}_i \mathbf{x}_2 + \varepsilon_i,$$

where $\bar{\mathbf{y}} = \mathbf{L}(\bar{\mathbf{y}}, \boldsymbol{\theta})$. The explanatory variables $\mathbf{g}_i \mathbf{x}_1$ and $\mathbf{g}_i \mathbf{x}_2$ are the averages x_1 and x_2, respectively, among friends. Once $\mathbf{y}^*$ is generated, I compute the count outcome $\mathbf{y}$ following Assumption $1'$.

As pointed out in Section 3.3, the estimator of $\boldsymbol{\theta}$ from the count data model may be close to that of the SART model if the dependent variable has a large dispersion. To illustrate this through the Monte Carlo study, I set two values for the parameter $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}',\ \boldsymbol{\gamma}')'$ by type of DGP. This allows simulating a count dependent variable with either low or high dispersion, depending on $\tilde{\boldsymbol{\beta}}$. The values used for $\tilde{\boldsymbol{\beta}}$ are presented in Table 1.

Table 1: Slope of the observed explanatory variables

	Low dispersion	High dispersion
Type A	(-2, -2.5, 2.1, 1.5, -1.2)	(-1, -6.8, 2.3, -2.5, 2.5)
Type B	(1, 0.4, 0.5, 0.5, 0.6)	(3, -1.8, 2.3, 2.5, 2.5)

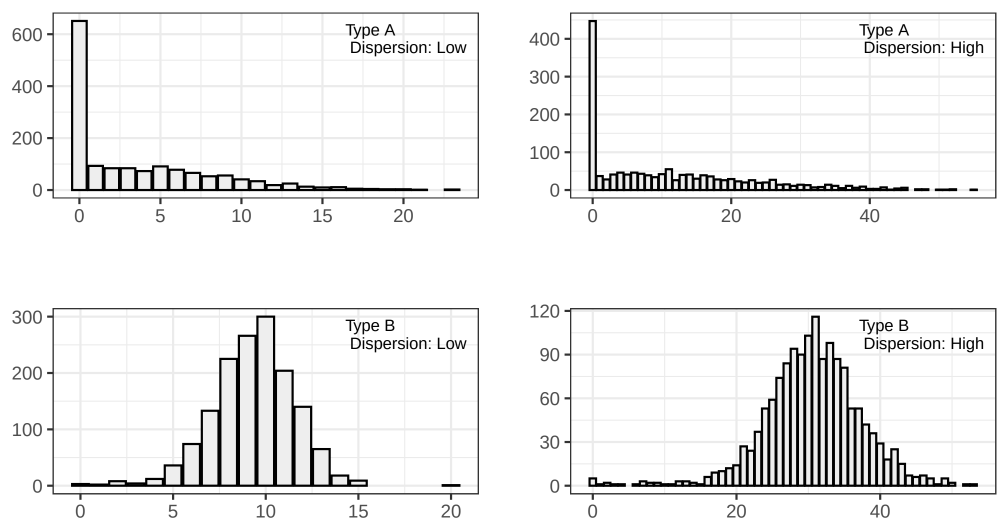
This table presents the values of $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}',\ \boldsymbol{\gamma}')'$ by type of DGP to simulate count data having either low or high dispersion. For instance, to simulate data from the DGP of type B with a low dispersion, I set $\boldsymbol{\beta} = (1, 0.4, 0.5)$ and $\boldsymbol{\gamma} = (0.5, 0.6)$.

The exogenous variables x_1 and x_2 are simulated from $\mathcal{N}(0, 4)$ and $\mathcal{P}oisson(3)$, respectively. I also

[11] When the proportion of zeros is very high, one may need zero-inflated or hurdle specifications (see Jones, 1989; Lambert, 1992). I discuss this point in Section 6.3.

consider several sample sizes, N $\in \{250, 750, 1500\}$. The adjacency matrix $\mathbf{G}$ is such that $g_{ij} = \dfrac{1}{n_i}$, if i is connected to j, and $g_{ij} = 0$ otherwise, where n_i is the degree of i randomly chosen between 0 and 20 for $N = 250$, 0 and 35 for $N = 750$, and 0 and 50 for $N = 1500$. Figure 2 presents the histogram of the simulated data for $N = 1500$. Data from a DGP of type A exhibit excess zeros (e.g., *number of cigarettes smoked daily* for low dispersion data or *weekly* for high dispersion data), whereas data from a DGP of type B concern frequent events (e.g., *number of recreational activities in which students participated in the last school year* for low dispersion data or *the last two school years* for large dispersion data).

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



I simulate each DG 1,000 times. The results can be replicated using my R package `CDatanet` and the replication code.[12]

The Monte Carlo results show that the NPL estimator of the count data model performs well in finite samples regardless of the type of DGP (see Tables 2 and 3). The estimator seems consistent. Moreover, the model performs better when the dependent variable has a higher dispersion. This implies that the data observation period may influence the results as a longer observation period would provide data having a large range (see Hakim et al., 1991). I discuss how to control for this in Section 6.2.

When comparing the count data model to the SART and SAR models, it stands out that the SART and SAR models bias the peer effects downward. The bias remains substantial in a large sample for both types of DGP when the dependent variable has a lower dispersion. This suggests that the counting nature of the data is important. In contrast, when the dependent variable has a large dispersion, the

[12]The package and the replication code are located at github.com/ahoundetoungan/CDatanet.

SART model estimator is close to that of the count data model. However, the bias of the SAR model is still large for the DGP of type A. Indeed, the SAR model does not control for the left-censoring nature of the dependent variable.

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 provides 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 is made up of 80 high schools and 52 middle schools. In particular, the data provides information on the social and demographic characteristics of students as well as their friendship links (i.e., best friends, up to 5 females and up to 5 males), education level, occupation of parents, etc.

I remove self-friendships and friendships between two students from different schools. Moreover, an important number of listed friend identifiers are missing or associated with "error codes."[13] I therefore remove from the study sample schools having many missing links and those having less than 100 students. I end up with 72,291 students from 120 schools. The largest school has 2,156 students, and about 50% of the schools have more than 500 students. The average number of friends per student is 3.8 (1.8 male friends and 2.0 female friends).

The studied counting dependent variable is the number of extracurricular activities in which students are enrolled. Students were presented with a list of clubs, organizations, and teams found in many schools. The students were asked to identify any of these activities in which they participated during the current school year or in which they planned to participate later in the school year. The students do not observe the activities in which their peers plan to participate. Therefore, the studied dependent variable is a good example for illustrating the model because the outcome is suited to a Bayesian game used to address the model. Throughout the paper, I write "*the number of extracurricular activities in which students are enrolled*" to mean the number of extracurricular activities in which the students participate during the year or in which they plan to participate.

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

Table 2: Monte Carlo simulations with low dispersion

Statistic	CDSI[1] Mean	CDSI[1] Sd.	SART Mean	SART Sd.	SAR Mean	SAR Sd.
			$N = 250$			
			Type A			
$\lambda = 0.4$	0.399	0.171	0.270	0.141	0.193	0.139
$\beta_0 = -2$	−2.009	0.441	−1.698	0.455	0.946	0.488
$\beta_1 = -2.5$	−2.500	0.075	−2.543	0.076	−1.689	0.078
$\beta_2 = 2.1$	2.100	0.072	2.133	0.073	1.534	0.084
$\gamma_1 = 1.5$	1.499	0.313	1.300	0.281	0.887	0.286
$\gamma_2 = -1.2$	−1.196	0.280	−1.016	0.247	−0.707	0.252
$\sigma_\varepsilon = 1.5$	1.469	0.085	1.546	0.087	2.013	0.106
			Type B			
$\lambda = 0.4$	0.407	0.088	0.303	0.076	0.283	0.104
$\beta_0 = 1$	0.984	0.454	1.806	0.451	1.911	0.492
$\beta_1 = 0.4$	0.400	0.049	0.400	0.049	0.399	0.049
$\beta_2 = 0.5$	0.500	0.057	0.501	0.058	0.500	0.058
$\gamma_1 = 0.5$	0.496	0.127	0.537	0.126	0.545	0.130
$\gamma_2 = 0.6$	0.588	0.164	0.738	0.148	0.754	0.178
$\sigma_\varepsilon = 1.5$	1.480	0.071	1.528	0.072	1.523	0.071
			$N = 750$			
			Type A			
$\lambda = 0.4$	0.394	0.112	0.263	0.096	0.171	0.118
$\beta_0 = -2$	−1.991	0.284	−1.685	0.298	0.945	0.334
$\beta_1 = -2.5$	−2.500	0.042	−2.543	0.043	−1.684	0.047
$\beta_2 = 2.1$	2.099	0.041	2.132	0.042	1.534	0.048
$\gamma_1 = 1.5$	1.489	0.206	1.288	0.190	0.854	0.235
$\gamma_2 = -1.2$	−1.193	0.181	−1.012	0.164	−0.679	0.201
$\sigma_\varepsilon = 1.5$	1.490	0.049	1.564	0.050	2.028	0.062
			Type B			
$\lambda = 0.4$	0.399	0.064	0.292	0.057	0.275	0.085
$\beta_0 = 1$	1.002	0.323	1.874	0.317	1.971	0.389
$\beta_1 = 0.4$	0.401	0.028	0.401	0.028	0.400	0.028
$\beta_2 = 0.5$	0.501	0.032	0.502	0.032	0.501	0.032
$\gamma_1 = 0.5$	0.500	0.088	0.543	0.088	0.550	0.091
$\gamma_2 = 0.6$	0.601	0.118	0.749	0.109	0.762	0.139
$\sigma_\varepsilon = 1.5$	1.494	0.040	1.533	0.040	1.531	0.040
			$N = 1500$			
			Type A			
$\lambda = 0.4$	0.402	0.088	0.268	0.078	0.143	0.132
$\beta_0 = -2$	−2.009	0.225	−1.705	0.234	0.930	0.271
$\beta_1 = -2.5$	−2.500	0.029	−2.543	0.029	−1.682	0.030
$\beta_2 = 2.1$	2.101	0.028	2.135	0.028	1.532	0.031
$\gamma_1 = 1.5$	1.502	0.162	1.296	0.149	0.804	0.238
$\gamma_2 = -1.2$	−1.200	0.141	−1.015	0.132	−0.632	0.217
$\sigma_\varepsilon = 1.5$	1.496	0.035	1.569	0.036	2.030	0.042
			Type B			
$\lambda = 0.4$	0.401	0.056	0.288	0.050	0.272	0.074
$\beta_0 = 1$	0.995	0.280	1.915	0.278	2.006	0.343
$\beta_1 = 0.4$	0.401	0.020	0.401	0.020	0.400	0.020
$\beta_2 = 0.5$	0.499	0.023	0.500	0.023	0.499	0.023
$\gamma_1 = 0.5$	0.503	0.072	0.549	0.072	0.555	0.076
$\gamma_2 = 0.6$	0.599	0.101	0.753	0.093	0.764	0.118
$\sigma_\varepsilon = 1.5$	1.497	0.028	1.533	0.028	1.531	0.028

(1): CDSI stands for count data model with social interactions. The count data model is estimated using the NPL method as described in Section 3.2, whereas the SART and the SAR models are estimated using the ML method. The number of simulations performed is 1,000. The "Mean" column reports the average of the 1,000 estimations, and the "Sd." column reports the standard deviation.

Table 3: Monte Carlo simulations with high dispersion

Statistic	CDSI[1] Mean	CDSI[1] Sd.	SART Mean	SART Sd.	SAR Mean	SAR Sd.
			$N = 250$			
			Type A			
$\lambda = 0.4$	0.401	0.032	0.387	0.033	0.306	0.092
$\beta_0 = -1$	-1.007	0.492	-0.406	0.498	2.812	1.558
$\beta_1 = -6.8$	-6.801	0.058	-6.807	0.058	-6.289	0.178
$\beta_2 = 2.3$	2.298	0.061	2.300	0.061	2.146	0.100
$\gamma_1 = -2.5$	-2.499	0.251	-2.591	0.256	-2.725	0.665
$\gamma_2 = 2.5$	2.497	0.185	2.559	0.188	2.395	0.376
$\sigma_\varepsilon = 1.5$	1.481	0.071	1.533	0.073	2.585	0.395
			Type B			
$\lambda = 0.4$	0.401	0.025	0.389	0.025	0.388	0.025
$\beta_0 = 3$	2.986	0.439	3.610	0.443	3.663	0.436
$\beta_1 = -1.8$	-1.800	0.048	-1.801	0.048	-1.798	0.049
$\beta_2 = 2.3$	2.300	0.056	2.301	0.056	2.299	0.056
$\gamma_1 = 2.5$	2.505	0.132	2.485	0.135	2.481	0.135
$\gamma_2 = 2.5$	2.497	0.178	2.560	0.180	2.562	0.180
$\sigma_\varepsilon = 1.5$	1.477	0.069	1.528	0.071	1.528	0.070
			$N = 750$			
			Type A			
$\lambda = 0.4$	0.400	0.024	0.384	0.023	0.299	0.078
$\beta_0 = 1$	-0.999	0.356	-0.359	0.358	2.751	1.219
$\beta_1 = -6.8$	-6.801	0.031	-6.807	0.031	-6.354	0.102
$\beta_2 = 2.3$	2.300	0.034	2.302	0.034	2.169	0.056
$\gamma_1 = -2.5$	-2.500	0.180	-2.607	0.179	-2.793	0.545
$\gamma_2 = 2.5$	2.502	0.133	2.571	0.133	2.447	0.297
$\sigma_\varepsilon = 1.5$	1.494	0.041	1.538	0.041	2.454	0.229
			Type B			
$\lambda = 0.4$	0.400	0.019	0.389	0.019	0.387	0.019
$\beta_0 = 3$	2.991	0.314	3.632	0.316	3.681	0.316
$\beta_1 = -1.8$	-1.801	0.028	-1.801	0.028	-1.800	0.028
$\beta_2 = 2.3$	2.301	0.034	2.301	0.034	2.300	0.034
$\gamma_1 = 2.5$	2.508	0.091	2.487	0.094	2.484	0.094
$\gamma_2 = 2.5$	2.499	0.133	2.560	0.134	2.563	0.134
$\sigma_\varepsilon = 1.5$	1.494	0.042	1.535	0.042	1.535	0.042
			$N = 1500$			
			Type A			
$\lambda = 0.4$	0.400	0.020	0.383	0.020	0.296	0.063
$\beta_0 = -1$	-1.006	0.298	-0.339	0.299	2.717	1.006
$\beta_1 = -6.8$	-6.801	0.023	-6.806	0.023	-6.381	0.072
$\beta_2 = 2.3$	2.301	0.023	2.302	0.023	2.180	0.038
$\gamma_1 = -2.5$	-2.501	0.148	-2.615	0.149	-2.828	0.441
$\gamma_2 = 2.5$	2.504	0.106	2.576	0.107	2.475	0.231
$\sigma_\varepsilon = 1.5$	1.496	0.029	1.536	0.029	2.391	0.158
			Type B			
$\lambda = 0.4$	0.400	0.016	0.387	0.016	0.385	0.016
$\beta_0 = 3$	3.012	0.269	3.672	0.272	3.721	0.272
$\beta_1 = -1.8$	-1.800	0.020	-1.800	0.020	-1.799	0.020
$\beta_2 = 2.3$	2.300	0.023	2.301	0.023	2.300	0.023
$\gamma_1 = 2.5$	2.500	0.074	2.477	0.075	2.474	0.076
$\gamma_2 = 2.5$	2.498	0.106	2.563	0.107	2.566	0.107
$\sigma_\varepsilon = 1.5$	1.224	0.012	1.239	0.011	1.239	0.011
$\sigma_\varepsilon = 1.5$	1.498	0.029	1.536	0.028	1.536	0.028

(1): CDSI stands for count data model with social interactions. The count data model is estimated using the NPL method as described in the Section 3.2, whereas the SART and the SAR models are estimated using the ML method. The number of simulations performed is 1,000. The "Mean" column reports the average of the 1,000 estimations, and the "Sd." column reports the standard deviation.

Table 7 provides the data summary. Figure 3 in Appendix C presents the distribution of the number of extracurricular activities in which the students are enrolled. It varies from 0 to 33 with an average of 2.4. Most students are enrolled in fewer than 10 activities. As observable characteristics, I consider age, sex, indicator of being Hispanic, race, number of years spent at their current school, living with both parents, mother's education, and mother's profession. Sex, race, mother's education and profession are discretized into subvariables.

5.2 Empirical estimation

I estimate the count data model as well as the SART and the SAR models by controlling for contextual effects and school heterogeneity as fixed effects. It is well known that controlling for fixed effects in a non-linear model leads to an inconsistent estimation because of the accidental parameter issue (see Neyman and Scott, 1948; Lancaster, 2000). However, as argued by Lee et al. (2014) and Liu (2019), school fixed effects can be included as *dummy* variables because the number of schools in the Add Health data is low relative to sample size. Moreover, I remove schools having fewer than 100 students from the data.

The estimation results without school heterogeneity are reported in Table 4, whereas those with school heterogeneity are reported in Table 5. The comparison of log-likelihoods of both estimations confirms that there is a school heterogeneity effect.[14] As in the Monte Carlo study, the SART and SAR models significantly underestimate the peer effects. Moreover, the estimation results of the SART and the SAR models are quite similar. This is because the DGP of the number of extracurricular activities in which students are enrolled is similar to the DGP of type B (see Section 4). As a result, the left-censoring nature of the dependent variable is not too important.

The coefficient of the count data model cannot be interpreted directly. Policy makers may be interested in the marginal effect of the explanatory variables on the expected number of extracurricular activities in which students are enrolled.[15] I present how to derive the marginal effects and the corresponding standard errors for the count data model in Appendix B.2.

The results confirm that an increase by one in the number of activities in which friends are enrolled implies an increased number of activities in which the students are enrolled of 0.363 (when controlling for school fixed effects). However, the SART and the SAR models underestimate this effect at 0.157 and 0.185, respectively. This result implies that the counting nature of the dependent variable is important.

Moreover, the own control variables are also significant. For instance, older students participate less

[14]This result is found using the likelihood ratio test. The test statistic is compared with the value of the Chi-squared distribution table for 119 degrees of freedom.

[15]Notice that only the estimators of the SART model's parameters can be interpreted as marginal effects.

in extracurricular activities, whereas Black and Asian students as well as students who have spent a greater number of years at their current school participate more. It is also found that many contextual effects are significant; for example, being a friend with male students increases one's participation, whereas being a friend with a student who has spent a greater number of years at their current school decreases ones's participation.

5.3 Endogeneity of the network

The estimation results above are based on the exogeneity of the network; that is, link formation does not depend on the error term ε_i in Equation (3). This assumption is strong and may imply inconsistent estimations (see Hsieh and Lee, 2016). To release this assumption, I consider a dyadic linking model in which the probability of link formation between two students i and j is specified with degree heterogeneity (e.g., Graham, 2017).

Let be $\mathbf{A} = [a_{ij}]$, the network data, such that $a_{ij} = 1$, if i knows j, and $a_{ij} = 0$ otherwise. Let also the latent variable a^*_{ij}, given by $a^*_{ij} = \Delta\mathbf{x}'_{ij}\bar{\boldsymbol{\beta}} + \mu_i + \mu_j + \varepsilon^*_{ij}$, where $\Delta\mathbf{x}_{ij}$ is a vector of observed dyad-specific variables, $\bar{\boldsymbol{\beta}}$ contains the parameters associated with the dyad-specific variables, μ_i is an unobserved individual-level attribute (gregariousness) that captures the *degree heterogeneity*, and $\varepsilon^*_{ij} \overset{iid}{\sim}$ *logistic*. The latent variable a^*_{ij} can be interpreted as a link formation utility. I assume that $a_{ij} = 1$, if $a^*_{ij} > 0$. Therefore, the probability of link formation between i and j, denoted P_{ij}, is defined as

$$P_{ij} = \frac{\exp\left(\Delta\mathbf{x}'_{ij}\bar{\boldsymbol{\beta}} + \mu_i + \mu_j\right)}{1 + \exp\left(\Delta\mathbf{x}'_{ij}\bar{\boldsymbol{\beta}} + \mu_i + \mu_j\right)}. \tag{16}$$

By convention, I set $P_{ii} = 0$ and $P_{ij} = 0$, if i and j come from different schools. A similar network formation model can be found in McCormick and Zheng (2015) and Breza et al. (2020), where the term $\Delta\mathbf{x}'_{ij}\bar{\boldsymbol{\beta}}$ is replaced by the distance between the individuals on a latent space.

As dyad-specific variables, I choose the absolute value of age difference, the absolute value of the difference in the number of years spent at the current school, whether both students are of the same sex, Hispanic, White, Black, Asian, and whether the mother's job for both students is professional. Importantly, the probability of link formation (16) is symmetric ($P_{ij} = P_{ji}$ for any $i,\ j \in \mathcal{V}$), but it allows the network to be directed because $\varepsilon^*_{ij} \neq \varepsilon^*_{ji}$. This specification is different from that of Graham (2017) in which $\varepsilon^*_{ij} = \varepsilon^*_{ji}$ and $a_{ij} = a_{ji}$ for all $i,\ j \in \mathcal{V}$.

For any student i from school s, I assume that the unobserved attribute μ_i is random and distributed according to $\mathcal{N}\left(u_{\mu s}, \sigma^2_{\mu s}\right)$. It is important to notice that the mean and the variance of μ_i vary across schools. Such a specification enables the capturing of school heterogeneity (as fixed effects) in the probability of link formation.

As pointed out in Hsieh and Lee (2016), the unobserved attributes μ_i may be correlated to the error

Table 4: Application results without fixed effects

Parameters	CDSI(1) Coef.	CDSI(1) Marginal effects		SART Coef.	SART Marginal effects		SAR	
λ	0.668	0.549	(0.025)***	0.249	0.203	(0.004)***	0.237	(0.006)***
Own effects								
Intercept	1.061	0.870	(0.096)***	2.415	1.963	(0.07)***	2.597	(0.094)***
Age	−0.019	−0.016	(0.006)**	−0.077	−0.063	(0.004)***	−0.075	(0.006)***
Male	−0.237	−0.195	(0.017)***	−0.243	−0.198	(0.017)***	−0.208	(0.019)***
Hispanic	0.036	0.029	(0.027)	0.012	0.010	(0.02)	0.052	(0.029)*
Race								
Black	0.250	0.205	(0.031)***	0.210	0.170	(0.023)***	0.235	(0.034)***
Asian	0.670	0.550	(0.035)***	0.651	0.529	(0.023)***	0.639	(0.039)***
Other	0.211	0.173	(0.029)***	0.197	0.160	(0.023)***	0.192	(0.033)***
Years at school	0.122	0.100	(0.008)***	0.132	0.107	(0.005)***	0.127	(0.008)***
With both par.	0.160	0.131	(0.020)***	0.158	0.129	(0.019)***	0.150	(0.022)***
Mother Educ.								
<High	−0.065	−0.054	(0.024)**	−0.068	−0.055	(0.024)**	−0.054	(0.027)**
>High	0.376	0.309	(0.02)***	0.381	0.310	(0.021)***	0.359	(0.022)***
Missing	0.222	0.182	(0.033)***	0.206	0.167	(0.028)***	0.240	(0.037)***
Mother job								
Professional	0.211	0.174	(0.025)***	0.219	0.178	(0.026)***	0.197	(0.029)***
Other	0.058	0.047	(0.021)**	0.055	0.045	(0.021)**	0.041	(0.024)*
Missing	−0.081	−0.066	(0.03)**	−0.080	−0.065	(0.027)**	−0.061	(0.033)*
Contextual effects								
Age	−0.078	−0.064	(0.004)***	−0.035	−0.028	(0.004)***	−0.042	(0.004)***
Male	0.108	0.088	(0.029)***	0.013	0.010	(0.031)	0.051	(0.034)
Hispanic	−0.153	−0.126	(0.039)***	−0.241	−0.196	(0.042)***	−0.217	(0.046)***
Race								
Black	−0.169	−0.139	(0.037)***	−0.095	−0.077	(0.035)**	−0.102	(0.043)**
Asian	−0.589	−0.484	(0.046)***	−0.447	−0.363	(0.047)***	−0.440	(0.058)***
Other	−0.279	−0.229	(0.05)***	−0.229	−0.186	(0.061)***	−0.220	(0.061)***
Years at school	−0.028	−0.023	(0.010)**	0.021	0.017	(0.01)*	0.021	(0.011)*
With both par.	0.069	0.057	(0.037)	0.244	0.198	(0.039)***	0.226	(0.041)***
Mother Educ.								
<High	−0.222	−0.182	(0.042)***	−0.204	−0.166	(0.049)***	−0.175	(0.05)***
>High	0.019	0.016	(0.036)	0.250	0.203	(0.038)***	0.239	(0.040)***
Missing	−0.247	−0.203	(0.060)***	−0.152	−0.123	(0.064)*	−0.099	(0.071)
Mother job								
Professional	0.094	0.078	(0.045)*	0.272	0.221	(0.051)***	0.252	(0.054)***
Other	−0.006	−0.005	(0.036)	0.107	0.087	(0.041)**	0.093	(0.044)**
Missing	−0.030	−0.024	(0.053)	0.067	0.055	(0.056)	0.054	(0.064)
σ_ε	2.426			2.447			2.315	
N	72,291			72,291			72,291	
log-likelihood	$-159,923.7$			$-160,606.6$			$-163,430.3$	
Fixed effects	No			No			No	

(1): CDSI stands for count data model with social interactions. The count data model is estimated using the NPL method as described in Section 3.2, whereas the SART and the SAR models are estimated using the ML method. Under the CSDI and the SART models, the column Coef. refers to the parameter values, while both columns of marginal effects refer to the marginal effects with their corresponding standard errors reported in parentheses. The columns under SAR report the parameter values (equal to the marginal effects) of the SAR model, with their standard error reported in parentheses. The codes ***, **, * mean that the corresponding parameter is significant at 1%, 5%, and 10%, respectively.

Table 5: Application results with fixed effects

Parameters	CDSI[1] Coef.	CDSI[1] Marginal effects		SART Coef.	SART Marginal effects		SAR	
λ	0.443	0.363	(0.028)***	0.194	0.157	(0.005)***	0.185	(0.006)***
Own effects								
Age	−0.049	−0.040	(0.008)***	−0.073	−0.059	(0.006)***	−0.061	(0.009)***
Male	−0.253	−0.207	(0.017)***	−0.261	−0.212	(0.018)***	−0.225	(0.019)***
Hispanic	0.123	0.101	(0.026)***	0.128	0.104	(0.021)***	0.158	(0.03)***
Race								
Black	0.309	0.253	(0.031)***	0.308	0.250	(0.025)***	0.312	(0.035)***
Asian	0.701	0.576	(0.035)***	0.704	0.572	(0.025)***	0.689	(0.04)***
Other	0.220	0.181	(0.028)***	0.217	0.176	(0.024)***	0.209	(0.033)***
Years at school	0.120	0.099	(0.007)***	0.120	0.097	(0.006)***	0.112	(0.009)***
With both par.	0.158	0.129	(0.019)***	0.153	0.124	(0.019)***	0.149	(0.022)***
Mother Educ.								
<High	−0.044	−0.036	(0.024)	−0.045	−0.036	(0.025)	−0.033	(0.027)
>High	0.392	0.321	(0.019)***	0.389	0.316	(0.021)***	0.369	(0.022)***
Missing	0.231	0.190	(0.032)***	0.214	0.174	(0.029)***	0.246	(0.037)***
Mother job								
Professional	0.236	0.193	(0.025)***	0.238	0.193	(0.026)***	0.217	(0.029)***
Other	0.069	0.057	(0.02)***	0.069	0.056	(0.022)***	0.057	(0.024)**
Missing	−0.064	−0.052	(0.029)*	−0.063	−0.051	(0.028)*	−0.042	(0.033)
Contextual effects								
Age	−0.064	−0.052	(0.005)***	−0.032	−0.026	(0.004)***	−0.039	(0.005)***
Male	0.032	0.026	(0.030)	−0.034	−0.027	(0.032)	0.011	(0.034)
Hispanic	−0.048	−0.039	(0.042)	−0.071	−0.057	(0.046)	−0.059	(0.049)
Race								
Black	−0.085	−0.070	(0.039)*	−0.028	−0.023	(0.038)	−0.045	(0.045)
Asian	−0.331	−0.272	(0.052)***	−0.219	−0.178	(0.054)***	−0.229	(0.062)***
Other	−0.245	−0.201	(0.052)***	−0.208	−0.169	(0.063)***	−0.203	(0.061)***
Years at school	−0.015	−0.012	(0.011)	−0.002	−0.001	(0.011)	−0.004	(0.013)
With both par.	0.165	0.135	(0.037)***	0.239	0.194	(0.040)***	0.228	(0.041)***
Mother Educ.								
<High	−0.180	−0.148	(0.043)***	−0.173	−0.141	(0.050)***	−0.147	(0.051)***
>High	0.190	0.156	(0.038)***	0.299	0.243	(0.040)***	0.286	(0.041)***
Missing	−0.178	−0.146	(0.061)**	−0.145	−0.118	(0.066)*	−0.095	(0.072)
Mother job								
Professional	0.257	0.211	(0.047)***	0.341	0.277	(0.053)***	0.321	(0.055)***
Other	0.076	0.062	(0.038)*	0.133	0.108	(0.043)**	0.124	(0.045)***
Missing	0.055	0.045	(0.054)	0.105	0.085	(0.059)	0.091	(0.064)
σ_ε	2.394			2.425			2.295	
N	72,291			72,291			72,291	
log-likelihood	$-158,963.9$			$-159,881.0$			$-162,744.4$	
Fixed effects	Yes			Yes			Yes	

(1): CDSI stands for count data model with social interactions. The count data model is estimated using the NPL method as described in Section 3.2, whereas the SART and the SAR models are estimated using the ML method. Under the CSDI and the SART models, the column Coef. refers to the parameter values, while both columns of marginal effects refer to the marginal effects with their corresponding standard errors reported in parentheses. The columns under SAR report the parameter values (equal to the marginal effects) of the SAR model, with their standard error reported in parentheses. The codes ***, **, * mean that the corresponding parameter is significant at 1%, 5%, and 10%, respectively.

term ε_i. In that case, there are unobserved variables (e.g., degree of student sociability) explaining both link formation and the students' participation in extracurricular activities. This violates the exogeneity condition on $\mathbf{G}$.

For any i, let $\boldsymbol{\eta}_i = (\varepsilon_i, \mu_i)'$. The variable $\boldsymbol{\eta}_i$ is distributed according to a bivariate normal distribution. Let $\boldsymbol{\Sigma}_{\mu\varepsilon}$ be the covariance matrix of $\boldsymbol{\eta}_i$.

$$\boldsymbol{\Sigma}_{\mu\varepsilon} = \begin{pmatrix} \sigma_\varepsilon^2 & \rho\sigma_\varepsilon\sigma_{\mu s} \\ \rho\sigma_\varepsilon\sigma_{\mu s} & \sigma_{\mu s}^2 \end{pmatrix}, \tag{17}$$

where ρ is the partial correlation between μ_i and ε_i. The error term ε_i can be rewritten as $\varepsilon_i = \rho\sigma_\varepsilon \dfrac{\mu_i - u_{\mu s}}{\sigma_{\mu s}} + \nu_i$, where $\nu_i \sim \mathcal{N}\left(0, (1-\rho^2)\sigma_\varepsilon^2\right)$, and $\mathbf{Cov}(\mu_i, \nu_i) = 0$. Let $\tilde{\mu}_i = \dfrac{\mu_i - u_{\mu s}}{\sigma_{\mu s}}$. By looking for more evidence of endogeneity, one can also control for the contextual effect of $\tilde{\mu}_i$. In that case, $\varepsilon_i = \rho\sigma_\varepsilon\tilde{\mu}_i + \bar{\rho}\sigma_\varepsilon\bar{\tilde{\mu}}_i + \tilde{\nu}_i$, where $\bar{\tilde{\mu}}_i$ is the average of $\tilde{\mu}_i$ among i's friends, $\bar{\rho}$ is the partial correlation between $\bar{\tilde{\mu}}_i$ and ε_i and $\tilde{\nu}_i \sim \mathcal{N}\left(0, \bar{\sigma}_\varepsilon^2\right)$. If μ_i or μ_j is correlated to ε_i, that is $\rho \neq 0$ or $\bar{\rho} \neq 0$, then the network is endogenous. To control for endogeneity, $\tilde{\mu}_i$ and $\bar{\tilde{\mu}}_i$ may simply be included in the count data model as additional explanatory variables (see Johnsson and Moon, 2015; Boucher and Houndetoungan, 2020). In that case, the BNE characterization (8) becomes

$$p_{iq} = \Phi\left(\frac{\lambda\mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} + \rho\sigma_\varepsilon\tilde{\mu}_i + \bar{\rho}\sigma_\varepsilon\bar{\tilde{\mu}}_i - a_q}{\bar{\sigma}_\varepsilon}\right) - \Phi\left(\frac{\lambda\mathbf{g}_i\bar{\mathbf{y}} + \mathbf{x}_i'\boldsymbol{\beta} + \rho\sigma_\varepsilon\tilde{\mu}_i + \bar{\rho}\sigma_\varepsilon\bar{\tilde{\mu}}_i - a_{q+1}}{\bar{\sigma}_\varepsilon}\right).$$

The paper now deals with a two-stage estimator. The first stage is based on a Bayesian approach. Using MCMC, I simulate $\bar{\boldsymbol{\beta}}$, μ_i for any $i \in \mathcal{V}$, $u_{\mu s}$, and $\sigma_{\mu s}^2$ for $s = 1, \dots, 120$ from their posterior distributions (see details in Appendix D.1). The simulations from the posterior distribution are then used to draw $\tilde{\mu}_i$ and $\bar{\tilde{\mu}}_i$. At the second stage, the draws of $\tilde{\mu}_i$ and $\bar{\tilde{\mu}}_i$ are used as additional explanatory variables to estimate the count data model. It is important to remember that the asymptotic variance of $\hat{\boldsymbol{\theta}}$, derived in Appendix B.1, is conditional on the explanatory variables. In the current framework, the asymptotic variance is conditional on $\mathbf{G}$, $\mathbf{X}$, and $\tilde{\boldsymbol{\mu}}$, the vector of $\tilde{\mu}_i$. This variance is no longer valid because $\tilde{\boldsymbol{\mu}}$ is not observed. By replicating drawings of μ_i, $u_{\mu s}$, and $\sigma_{\mu s}^2$ from the posterior distribution, I correct the asymptotic variance. The approach I use is similar in spirit to that of Krinsky and Robb (1986). The new variance accounts for the variability of $\tilde{\mu}_i$ (see details in Appendix D.2).

The estimation results (controlling for schools' heterogeneity and network endogeneity) are presented in Table 6. The results are significantly different to those of Table 5. The parameters of the additional explanatory variables are significantly different to zero at 1%. This confirms that the network is endogenous.

Although friends incite participation in extracurricular activities, the sociability degree (gregariousness) of the students also plays an important role. Students with high μ_i are more *extroverted* (more likely to form links) and also participate in more extracurricular activities.[16] In contrast, *introverted* students participate less in extracurricular activities. Similar evidence has been found in sociology studies, which highlight that an individual's gregariousness determines their participation in activities.[17] As well, being friends of a highly gregarious student also increases one's participation in extracurricular activities.[18]

Peer effects are reduced when controlling for network endogeneity but remain significant. An increase by one in the number of activities in which friends are enrolled implies an increase in the number of activities in which students are enrolled of 0.295. The endogeneity of the network is also confirmed with the models SART and SAR. However, they still underestimate peer effects at 0.141 and 0.166, respectively.

To understand the decrease in peer effects, notice that λ could capture other effects if students' gregariousness is not included in the count data model. For example, λ can capture the effect of an exogenous shock that increases students' and peers' gregariousness because students and their friends will experiment and increase in their participation in extracurricular activities. This is similar to the correlated effects (see Manski, 1993).

6 Discussions

In this section, I discuss some limits, some areas for future research, and some general implications of the results.

6.1 Flexibly dispersed count variable model

The most commonly used models to study count data (without social interactions) are the Poisson model and related models, such as the generalized Poisson (Consul and Jain, 1973) and Negative Binomial (Hilbe, 2011).[19] The main difference between these models is in the way they fit the dispersion of the dependent variable. The count data model of this paper is flexible in terms of dispersion fitting. The fundamental feature of the Poisson model is the mean-variance equality conditional on the explanatory variables (equidispersion). However, count variables generally have values across a large range and have many low values. Therefore, their variance is sometimes greater than their mean

[16]Because $\rho\sigma_\varepsilon$, the sign of $\tilde{\mu}_i$ is positive in the count data model.

[17]For example, specific personality traits are associated with activity participation (e.g., Newton et al., 2018); extroverted people work more often in jobs having more social interactions (e.g., Pfeiffer and Schulz, 2011), and highly gregarious individuals are more likely to be a member of a group (e.g., Erbe, 1962).

[18]Because $\bar{\rho}\sigma_\varepsilon$, the sign of $\mathbf{g}_i\tilde{\mu}_i$ is positive in the count data model.

[19]See also Cameron and Trivedi (1990); Böhning et al. (1999); Chang (2005) and Chiou and Fu (2013). Models of the Poisson family that deal with spatial autocorrelation also exist (e.g., Karlis, 2003; Inouye et al., 2017). However, as pointed out in the introduction, these models are not based on any microeconomic foundation. The spatial correlation parameter cannot be interpreted as peer effects.

Table 6: Application results controlling for fixed effects and network endogeneity

Parameters	CDSI[1] Coef.	CDSI[1] Marginal effects		SART Coef.	SART Marginal effects		SAR	
λ	0.359	0.294	(0.028)***	0.173	0.141	(0.005)***	0.166	(0.006)***
$\rho\sigma_\varepsilon$	0.246	0.202	(0.011)***	0.253	0.205	(0.010)***	0.240	(0.013)***
$\bar{\rho}\sigma_\varepsilon$	0.202	0.166	(0.019)***	0.240	0.195	(0.018)***	0.218	(0.020)***
Own effects								
Age	−0.049	−0.040	(0.008)***	−0.066	−0.053	(0.006)***	−0.061	(0.009)***
Male	−0.241	−0.198	(0.017)***	−0.249	−0.202	(0.018)***	−0.213	(0.019)***
Hispanic	0.179	0.147	(0.027)***	0.184	0.150	(0.022)***	0.211	(0.031)***
Race								
Black	0.557	0.457	(0.033)***	0.564	0.458	(0.027)***	0.552	(0.038)***
Asian	0.848	0.696	(0.035)***	0.847	0.687	(0.026)***	0.827	(0.041)***
Other	0.281	0.231	(0.028)***	0.281	0.228	(0.024)***	0.269	(0.033)***
Years at school	0.099	0.081	(0.007)***	0.097	0.079	(0.006)***	0.092	(0.009)***
With both par.	0.145	0.119	(0.019)***	0.142	0.115	(0.019)***	0.135	(0.022)***
Mother Educ.								
<High	−0.021	−0.017	(0.024)	−0.021	−0.017	(0.025)	−0.012	(0.027)
>High	0.377	0.309	(0.019)***	0.376	0.305	(0.021)***	0.354	(0.022)***
Missing	0.226	0.185	(0.032)***	0.210	0.170	(0.029)***	0.242	(0.036)***
Mother job								
Professional	0.209	0.171	(0.024)***	0.209	0.170	(0.026)***	0.191	(0.029)***
Other	0.054	0.044	(0.020)**	0.056	0.045	(0.022)**	0.043	(0.023)*
Missing	−0.060	−0.050	(0.029)*	−0.058	−0.047	(0.028)*	−0.041	(0.033)
Contextual effects								
Age	−0.075	−0.061	(0.005)***	−0.051	−0.041	(0.004)***	−0.056	(0.005)***
Male	−0.002	−0.002	(0.029)	−0.042	−0.034	(0.032)	0.002	(0.034)
Hispanic	0.002	0.001	(0.042)	−0.009	−0.007	(0.047)	−0.001	(0.049)
Race								
Black	0.171	0.140	(0.043)***	0.241	0.196	(0.042)***	0.205	(0.048)***
Asian	−0.114	−0.094	(0.055)*	−0.013	−0.011	(0.055)	−0.039	(0.064)
Other	−0.157	−0.129	(0.053)**	−0.122	−0.099	(0.063)	−0.127	(0.061)**
Years at school	−0.016	−0.013	(0.011)	−0.010	−0.008	(0.011)	−0.009	(0.013)
With both par.	0.153	0.126	(0.037)***	0.207	0.168	(0.04)***	0.193	(0.041)***
Mother Educ.								
<High	−0.152	−0.125	(0.043)***	−0.143	−0.116	(0.050)**	−0.122	(0.051)**
>High	0.169	0.139	(0.038)***	0.246	0.200	(0.040)***	0.236	(0.041)***
Missing	−0.147	−0.120	(0.062)*	−0.124	−0.101	(0.065)	−0.081	(0.071)
Mother job								
Professional	0.205	0.168	(0.047)***	0.269	0.218	(0.053)***	0.246	(0.055)***
Other	0.034	0.028	(0.038)	0.083	0.067	(0.043)	0.072	(0.045)
Missing	0.037	0.030	(0.055)	0.083	0.067	(0.059)	0.065	(0.064)
$\bar{\sigma}_\varepsilon$	2.377			2.412			2.283	
N	72,291			72,291			72,291	
log-likelihood	$-158,467.7$			$-159,462.2$			$-162,328.3$	
Fixed effects	Yes			Yes			Yes	

(1): CDSI stands for count data model with social interactions. The count data model is estimated using the NPL method as described in Section 3.2, whereas the SART and the SAR models are estimated using the ML method. Under the CSDI and the SART models, the column Coef. refers to the parameter values, while both columns of marginal effects refer to the marginal effects with their corresponding standard errors reported in parentheses. The columns under SAR report the parameter values (equal to the marginal effects) of the SAR model, with their standard error reported in parentheses. The codes ***, **, * mean that the corresponding parameter is significant at 1%, 5%, and 10%, respectively.

(overdispersion). In that case, Negative Binomial models offer a better fit to the data. The Negative Binomial specification has an additional parameter, which links the conditional mean to the conditional variance as $\mathbf{Var}\left(y_i\right) = \bar{y}_i + \alpha\bar{y}_i^2$, where $\bar{y}_i = \mathbf{E}(y_i)$ and $\alpha > 0$.[20] In addition to the overdispersion, the generalized Poisson allows the variance to be smaller than the mean (underdispersion) as $\mathbf{Var}\left(y_i\right) = \dfrac{\bar{y}_i}{(1-\alpha)^2}$, where $\max\left(-1,\ -\dfrac{\bar{y}_i}{4}\right) < \alpha < 1$. Under this specification, data are overdispersed if $\alpha > 0$ and underdispersed otherwise.

The specification of the relationship between the conditional mean and the conditional variance is important in this paper. The conditional mean $\bar{y}_i$ is used to compute the explanatory variable of interest of the model (see Equation 8). Misspecification of that relationship may involve classical bias, especially on the peer effect estimation. The new count variable model is flexible and allows equidispersion, underdispersion, and overdispersion. From Theorem 1, it follows that

$$\mathbf{Var}\left(y_i\right) = \bar{y}_i + \underbrace{2\sum_{r=1}^{\infty} r\Phi\left(\hat{\psi}_{ir}\right) - \bar{y}_i^2}_{\Delta(\sigma_\varepsilon)}, \tag{18}$$

where $\forall\ i \in \mathcal{V},\ q \in \mathbb{N}^*$, and $\hat{\psi}_{iq} = \dfrac{\lambda\mathbf{g}_i\bar{\mathbf{y}} + \psi_i - q + 1}{\sigma_\varepsilon}$. The equation $\Delta(\sigma_\varepsilon) = 0$ does not have a closed form, but $\Delta(\sigma_\varepsilon)$ is increasing in σ_ε. Depending on σ_ε, the term $\Delta(\sigma_\varepsilon)$ may be null, negative, or positive. The parameter σ_ε is a dispersion parameter for the new count data model as α is for the generalized Poisson and Negative Binomial models.

For too-low $\hat{\psi}_{ir}$, overdispersion is possible when σ_ε is large, but the underdispersion condition may be violated as $\bar{y}_i$ will decrease to 0. This is also the case of the generalized Poisson model in which underdispersion is less frequent when $\bar{y}_i$ decreases to 0.

6.2 Count data observation period

Data from "*How many times do you smoke a day*?" are not the same as those of "*How many times do you smoke a week*?" The data observation period, commonly known as *exposure time* (Hakim et al., 1991), is important when modeling count variables. Using daily or weekly data affects the results and thus policy implications. The Monte Carlo study highlights that the estimator of the model parameters performs better when the outcome takes values across a large range. Therefore, data from a high exposure time may lead to a more precise estimator. If one does not control for the exposure time, the interpretation of peer effects and the policy implications are only valid for the considered exposure time.

In the traditional count data models (Poisson and Negative Binomial), this issue can be fixed using an

[20]Conditioning of the explanatory variables in the expectancy and the variance is omitted to simplify the notations.

offset. This consists of adding the log of the exposure time as a supplementary explanatory variable and constraining its coefficient to one (see Winkelmann and Zimmermann, 1995). In that case, $\frac{\bar{y}_i}{e}$ does not depend on the exposure time e, as $\bar{y}_i$ is a log-linear function of explanatory variables. Since the expected outcome $\bar{y}_i$ of the new count variable model has a more complex function, the offset approach cannot be used.

To make results comparable across exposure times for the new count variable model, the sequence $(a_q)_{q\in\mathbb{N}}$ of Assumption $1'$ may be redefined as $a_0 = -\infty$ and $a_q = e(q-1)\ \forall\ q \in \mathbb{N}^*$. Under this specification, the distribution of $\frac{y_i}{e}$ does not depend on the exposure time. This way to control the exposure time is more general because it scales not only the expected outcome of $\bar{y}_i$ (as the offset approach) but also the full distribution of y_i.

6.3 Zero-inflated and Hurdle specifications

In applications with excess zeros, zero-inflated (see Lambert, 1992) or Hurdle (see Jones, 1989) specifications are suggested for modeling count data. These specifications assume that "zeros" could be generated by processes other than those of the positive values. For instance, for the question "*How many times did you smoke during the last week*?" smokers may report zero because they did not smoke during that *specific* week. However, other individuals may report zero because they are non-smokers. The first type of zeros are *sampling*, whereas the second type of zeros are *structural*. It may be important to distinguish both processes because they do not have the same policy implications (see Tüzen and Erbaş, 2018).

The zero-inflated and Hurdle specifications are given respectively by Equations (19) and (20).

$$p_{iq} = \begin{cases} w_i + (1-w_i)p_{i0} & \text{if } q = 0, \\ (1-w_i)p_{iq} & \text{if } q \geq 1, \end{cases} \tag{19}$$

$$p_{iq} = \begin{cases} w_i & \text{if } q = 0, \\ (1-w_i)p_{iq} & \text{if } q \geq 1 \end{cases}, \tag{20}$$

where w_i is the probability of generating a structural zero.

The zero-inflated model assumes that there is a mix of sampling of structural zeros, whereas the Hurdle specification allows only structural zeros. I refer the reader to Jones (1989) and (Lambert, 1992) for more details. However, these specifications are not compatible with the microeconomic foundation of the model. This could be investigated in future research.

7 Conclusion

In this paper, I study a social network model for count data using a Bayesian game. I show that the counting nature of the dependent variable is important, especially when the variable has a small range. Assuming that this variable is continuous underestimates peer effects. However, bias decreases when the range of the dependent variable is large.

I also study peer effects on the number of extracurricular activities in which a student is enrolled, controlling for the endogeneity of the network. I find that an increase by one in the number of activities in which friends are enrolled implies an increase in the number of activities in which students are enrolled by 0.295, whereas the SART and SAR models underestimate this effect at 0.141 and 0.166, respectively. I also show that ignoring the endogeneity overestimates the peer effects.

I present an easy to use R package that implements all the methods used in this paper.[21] Although this is, to my knowledge, the first paper to study count data with social interactions, there remain many important challenges that are not addressed by this paper. In particular, this paper does not consider zero-inflated specifications for data having excess zeros.

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

References

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

Ali, M. M., A. Amialchuk, and F. W. Heiland (2011): "Weight-related behavior among adolescents: the role of peer effects," *PloS one*, 6, e21179.

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

Atchadé, Y. F., J. S. Rosenthal, et al. (2005): "On adaptive markov chain monte carlo algorithms," *Bernoulli*, 11, 815–828.

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

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

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

Böhning, D., E. Dietz, P. Schlattmann, L. Mendonca, and U. Kirchner (1999): "The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162, 195–209.

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

Boucher, V., F. A. Dedewanou, and A. Dufays (2018): "Peer-Induced Beliefs Regarding College Participation," .

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

Boucher, V. and E. A. Houndetoungan (2020): "Estimating peer effects using partial network data," .

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

——— (2019): "Peer Effects in Networks: A Survey," *Annual Review of Economics, forthcoming.*

Breza, E., A. G. Chandrasekhar, T. H. McCormick, and M. Pan (2020): "Using aggregated relational data to feasibly identify network structure without network data," *American Economic Review*, 110, 2454–84.

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

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

Cameron, A. C. and P. K. Trivedi (1990): "Regression-based tests for overdispersion in the Poisson model," *Journal of econometrics*, 46, 347–364.

——— (2013): *Regression analysis of count data*, vol. 53, Cambridge university press.

Chang, L.-Y. (2005): "Analysis of freeway accident frequencies: negative binomial regression versus artificial neural network," *Safety science*, 43, 541–557.

Chiou, Y.-C. and C. Fu (2013): "Modeling crash frequency and severity using multinomial-generalized Poisson model with error components," *Accident Analysis & Prevention*, 50, 73–82.

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

Consul, P. C. and G. C. Jain (1973): "A generalization of the Poisson distribution," *Technometrics*, 15, 791–799.

Daniels, E. and C. Leaper (2006): "A longitudinal investigation of sport participation, peer acceptance, and self-esteem among adolescent girls and boys," *Sex roles*, 55, 875–880.

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

Duncan, G. J., J. Boisjoly, M. Kremer, D. M. Levy, and J. Eccles (2005): "Peer effects in drug use and sex among college students," *Journal of abnormal child psychology*, 33, 375–385.

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

Fortin, B. and V. Boucher (2015): "Some challenges in the empirics of the effects of networks," in *The Oxford Handbook of the Economics of Networks*, Oxford University Press.

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.

Gelfand, A. E. and A. F. Smith (1990): "Sampling-based approaches to calculating marginal densities," *Journal of the American statistical association*, 85, 398–409.

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 (2017): "Multinomial choice with social interactions: occupations in Victorian London," *Documento CEDE*.

Hakim, S., D. Shefer, A.-S. Hakkert, and I. Hocherman (1991): "A critical review of macro models for road accidents," *Accident Analysis & Prevention*, 23, 379–400.

Halmos, P. R. (2012): *A Hilbert space problem book*, vol. 19, Springer Science & Business Media.

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

Hellerstein, D. and R. Mendelsohn (1993): "A theoretical foundation for count data models," *American journal of agricultural economics*, 75, 604–611.

Hilbe, J. M. (2011): *Negative binomial regression*, Cambridge University Press.

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.

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

Jones, A. M. (1989): "A double-hurdle model of cigarette consumption," *Journal of applied econometrics*, 4, 23–39.

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

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

Krinsky, I. and A. L. Robb (1986): "On approximating the statistical properties of elasticities," *The Review of Economics and Statistics*, 715–719.

Lambert, D. (1992): "Zero-inflated Poisson regression, with an application to defects in manufacturing," *Technometrics*, 34, 1–14.

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

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

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

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

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

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

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

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

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

MADDALA, G. S. (1986): *Limited-dependent and qualitative variables in econometrics*, 3, Cambridge university press.

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

MAYS, D., N. THOMPSON, H. I. KUSHNER, D. F. MAYS II, D. FARMER, AND M. WINDLE (2010): "Sports-specific factors, perceived peer drinking, and alcohol-related behaviors among adolescents participating in school-based sports in Southwest Georgia," *Addictive behaviors*, 35, 235–241.

MCCORMICK, T. H. AND T. ZHENG (2015): "Latent surface models for networks using Aggregated Relational Data," *Journal of the American Statistical Association*, 110, 1684–1695.

METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, AND E. TELLER (1953): "Equation of state calculations by fast computing machines," *The journal of chemical physics*, 21, 1087–1092.

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.

NEYMAN, J. AND E. L. SCOTT (1948): "Consistent estimates based on partially consistent observations," *Econometrica: Journal of the Econometric Society*, 1–32.

OSBORNE, M. J. AND A. RUBINSTEIN (1994): *A course in game theory*, MIT press.

PFEIFFER, F. AND N. J. SCHULZ (2011): "Gregariousness, interactive jobs and wages," *ZEW-Centre for European Economic Research Discussion Paper*.

REEVES, D. AND M. P. WELLMAN (2012): "Computing best-response strategies in infinite games of incomplete information," *arXiv preprint arXiv:1207.4171*.

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.

TÜZEN, M. F. AND S. ERBAŞ (2018): "A comparison of count data models with an application to daily cigarette consumption of young persons," *Communications in Statistics-Theory and Methods*, 47, 5825–5844.

WINKELMANN, R. AND K. F. ZIMMERMANN (1995): "Recent developments in count data modelling: theory and application," *Journal of economic surveys*, 9, 1–24.

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.

Zill, D. and P. Shanahan (2009): *A first course in complex analysis with applications*, Jones & Bartlett Learning.

Appendices

A Proof of the BNE uniqueness

To simplify the notations, I assume that f and F are the pdf and the cdf of $\mathcal{N}\left(0,\sigma_\varepsilon^2\right)$, respectively; that is, $\mathrm{f}(.) = \dfrac{1}{\sigma_\varepsilon}\phi\left(\dfrac{.}{\sigma_\varepsilon}\right)$, and $\mathrm{F}(.) = \Phi\left(\dfrac{.}{\sigma_\varepsilon}\right)$. In this Appendix, I proof the Theorem 1. To do so, I first state and prove the following lemma.

Lemma 1. *Let $\left(p^e_{iq}\right)_{\substack{i\in\mathcal{V}\\ q\in\mathbb{N}}}$ be a belief system that verifies the BNE characterization (6), $\bar{y}^e_i = \sum_{r=0}^{\infty} r p^e_{ir}$ for all $i \in \mathcal{V}$, and $\bar{\mathbf{y}}^e = (\bar{y}^e_1 \;\dots\; \bar{y}^e_n)$. If $\varepsilon_i \overset{iid}{\sim} \mathcal{N}\left(0,\sigma_\varepsilon^2\right)$, then $\bar{\mathbf{y}}^e = \mathbf{L}\left(\bar{\mathbf{y}}^e\right)$.*

A.1 Proof of Lemma 1

The mapping $\mathbf{L}$ is defined for any $\bar{\mathbf{y}} \in \mathbb{R}^N_+$ by $\mathbf{L}\left(\bar{\mathbf{y}}\right) = \left(\ell_1\left(\bar{\mathbf{y}}\right) \;\dots \ell_n\left(\bar{\mathbf{y}}\right)\right)'$, where

$$\ell_i\left(\bar{\mathbf{y}}\right) = \sum_{r=1}^{\infty} \mathrm{F}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}} + \psi_i - a_r\right) \quad \text{for all } i \in \mathcal{V}. \tag{21}$$

As (p^e_{iq}) verifies (6),

$$\begin{aligned} p^e_{iq} &= \mathrm{F}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_q\right) - \mathrm{F}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_{q+1}\right), \\ \bar{y}^e_i &= \underbrace{\sum_{r=0}^{\infty} r\mathrm{F}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_r\right)}_{S_1} - \underbrace{\sum_{r=0}^{\infty} r\mathrm{F}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_{r+1}\right)}_{S_2}. \end{aligned} \tag{22}$$

Equation (22) holds by the fact that the both sums are finite. To prove this, let $x < 0$ with $|x|$ being sufficiently large. It follows that $\mathrm{F}(x) = \int_{-\infty}^{x} \mathrm{f}(t)dt < -\int_{-\infty}^{x} \dfrac{t}{\sigma_\varepsilon^2}\mathrm{f}(t)dt = \mathrm{f}(x)$. Therefore, $S_1 < \underbrace{\sum_{r=r_0}^{\infty} r\mathrm{f}\left(\mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_r\right)}_{S_3}$, for some integer r_0, is sufficiently large. In addition, $S_3 < \infty$ because f is a decreasing exponential function as r grows. Hence, $S_1 < \infty$. Analogously, $S_2 < \infty$.

$$\begin{aligned} \bar{y}^e_i &= \sum_{r=0}^{\infty} r\mathrm{F}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_r\right) - \sum_{r=0}^{\infty} (r+1)\mathrm{F}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_{r+1}\right) + \sum_{r=0}^{\infty} \mathrm{F}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_{r+1}\right), \\ \bar{y}^e_i &= \sum_{r=1}^{\infty} r\mathrm{F}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_r\right) - \sum_{r=1}^{\infty} r\mathrm{F}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_r\right) + \sum_{r=0}^{\infty} \mathrm{F}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_{r+1}\right), \\ \bar{y}^e_i &= \sum_{r=1}^{\infty} \mathrm{F}\left(\lambda \mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_r\right) = \sum_{r=1}^{\infty} \Phi\left(\frac{\lambda \mathbf{g}_i \bar{\mathbf{y}}^e + \psi_i - a_r}{\sigma_\varepsilon}\right). \end{aligned} \tag{23}$$

The proof is given by (23) because $\bar{y}^e_i = \ell_i\left(\bar{\mathbf{y}}^e\right)$ for all $i \in \mathcal{V}$. Hence, $\bar{\mathbf{y}}^e = \mathbf{L}\left(\bar{\mathbf{y}}^e\right)$.

A.2 Proof of Theorem 1

The BNE characterization (6) implies that there is a bijection correspondence between the beliefs and the expected outcome at equilibrium. Following this result, the key determinant of the proof is to establish that the vector of equilibrium beliefs exists and that the vector of expected equilibrium outcome is unique. This implies that there is a unique vector of equilibrium beliefs and a unique vector of expected equilibrium outcome.

Let $\mathbb{R}^{\infty}$ be the vector space of infinite sequences $(\alpha_1, \alpha_2, \alpha_3, \dots)$ of real numbers.[22] To express the characterization of the BNE (6) as a matrix, let us denote by $\mathbf{p}_{\text{q}} = (p_{1q}, \dots, p_{nq})'$, an n-dimensional vector for any $q \in \mathbb{N}$, $\mathbf{p} = (\mathbf{p}_0', \mathbf{p}_1', \mathbf{p}_2', \mathbf{p}_3', \dots)'$, $\mathbf{h}_1 = (a_0, a_1, a_2, a_3, \dots)'$, $\mathbf{h}_2 = (a_1, a_2, a_3, a_4, \dots)'$ infinite-dimensional vectors, and $\mathbf{1}_d$, the d-dimensional vector of ones for any $d \in \mathbb{N}^*$ or $d = \infty$. Let also $\mathbf{J} = (0, 1, 2, 3, \dots)$, an infinite-dimensional row-vector, and $\mathbf{B} = \mathbf{1}_{\infty} \otimes \mathbf{J} \otimes \mathbf{G}$. Equation (6) in matrix form is given by

$$\mathbf{p} = \mathbf{\Phi}\left(\frac{\lambda \mathbf{B}\mathbf{p} + \mathbf{1}_{\infty} \otimes \Psi - \mathbf{h}_1 \otimes \mathbf{1}_n}{\sigma_{\varepsilon}}\right) - \mathbf{\Phi}\left(\frac{\lambda \mathbf{B}\mathbf{p} + \mathbf{1}_{\infty} \otimes \Psi - \mathbf{h}_2 \otimes \mathbf{1}_n}{\sigma_{\varepsilon}}\right), \tag{24}$$

where $\mathbf{\Phi}$ is defined for any $\mathbf{q} = (q_1, q_2, q_3, \dots) \in \mathbb{R}^{\infty}$ as $\mathbf{\Phi}(\mathbf{q}) = (\Phi(q_1), \Phi(q_2), \Phi(q_3), \dots)$.

Let $\mathbf{C} \subset \mathbb{R}^{\infty}$ defined by $\mathbf{C} := \left\{\mathbf{p} \ / \forall\ i \in \mathcal{V} \text{ and } q \in \mathbb{N},\ p_{iq} \in [0,1] \ \text{ and } \forall\ i \in \mathcal{V},\ \sum_{r=0}^{\infty} p_{ir} = 1\right\}$. Note that any belief system $\mathbf{p} \in \mathbf{C}$. Let also $\mathbf{H}$ be a mapping from $\mathbf{C}$ to itself, such that $\forall\ \mathbf{p} \in \mathbf{C}$,

$$\mathbf{H}(\mathbf{p}) = \mathbf{\Phi}\left(\frac{\lambda \mathbf{B}\mathbf{p} + \mathbf{1}_{\infty} \otimes \Psi - \mathbf{h}_1 \otimes \mathbf{1}_n}{\sigma_{\varepsilon}}\right) - \mathbf{\Phi}\left(\frac{\lambda \mathbf{B}\mathbf{p} + \mathbf{1}_{\infty} \otimes \Psi - \mathbf{h}_2 \otimes \mathbf{1}_n}{\sigma_{\varepsilon}}\right). \tag{25}$$

Any $\mathbf{p} \in \mathbf{C}$ is an equilibrium belief of the incomplete information network game with the utility (1) if $\mathbf{p} = \mathbf{H}(\mathbf{p})$. $\mathbf{H}$ is a continuous mapping from $\mathbf{C}$ to itself. The set $\mathbf{C}$ is a compact and convex nonempty subset of the infinite dimensional normed space $\mathbb{R}^{\infty}$. Schauder's fixed point Theorem (generalization of Brouwer's fixed point Theorem to an infinite dimensional space) gives the existence of $\mathbf{p}$ (see Smart, 1980, Chapter 2). Thus, there exists $\mathbf{p}^e \in \mathbf{C}$, such that $\mathbf{p}^e = \mathbf{H}(\mathbf{p}^e)$. By Lemma 1, there is also $\bar{\mathbf{y}}^e = (\bar{y}_1^e \dots \bar{y}_n^e)$, where $\bar{y}_i^e = \sum_{r=0}^{\infty} r p_{ir}^e$, such that $\bar{\mathbf{y}}^e = \mathbf{L}(\bar{\mathbf{y}}^e)$.

I will show that $\mathbf{u} = \mathbf{L}(\mathbf{u})$ has a unique solution. By the contraction mapping Theorem, it is sufficient to prove that $\left|\left|\dfrac{\partial \mathbf{L}(\mathbf{u})}{\partial \mathbf{u}'}\right|\right|_{\infty} < 1$ for any $\mathbf{u} = (u_1, \dots, u_n) \in \mathbb{R}^n$.

For all i and j,

$$\frac{\partial \ell_i(\mathbf{u})}{\partial u_j} = \lambda g_{ij} \underbrace{\sum_{r=1}^{\infty} \mathrm{f}(\lambda \mathbf{g}_i \mathbf{u} + \psi_i - a_r)}_{f_i^*},$$

[22] A natural generalization of $\mathbb{R}^k$, $k \in \mathbb{N}^*$ (see Halmos, 2012).

$$\frac{\partial \ell_i\left(\mathbf{u}\right)}{\partial u_j} = \lambda g_{ij} f_i^*. \tag{26}$$

From Equation (26), $\dfrac{\partial \mathbf{L}\left(\mathbf{u}\right)}{\partial \mathbf{u}'}$ is defined by

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

It follows that

$$\begin{aligned}
\left|\left|\frac{\partial \mathbf{L}\left(\mathbf{u}\right)}{\partial \mathbf{u}'}\right|\right|_\infty &= |\lambda| \max_i \left\{ |f_i^*| \sum_{j=1}^{N} g_{ij} \right\}, \\
\left|\left|\frac{\partial \mathbf{L}\left(\mathbf{u}\right)}{\partial \mathbf{u}'}\right|\right|_\infty &= |\lambda| \max_i |f_i^*| \max_i \left\{ \sum_{j=1}^{N} g_{ij} \right\}, \\
\left|\left|\frac{\partial \mathbf{L}\left(\mathbf{u}\right)}{\partial \mathbf{u}'}\right|\right|_\infty &\leq |\lambda| \max_i |f_i^*| ||\mathbf{G}||_\infty. \qquad (27)
\end{aligned}$$

I will now focus on the term f_i^*.

$$\begin{aligned}
f_i^* &= \sum_{r=1}^{\infty} \mathrm{f}\left(\lambda \mathbf{g}_i \mathbf{u} + \psi_i - a_r\right), \\
f_i^* &= \sum_{r=1}^{\infty} \mathrm{f}\left(m_i + a_1 - a_r\right),
\end{aligned}$$

where $m_i^* = \lambda \mathbf{g}_i \mathbf{u} + \psi_i - a_1$. As $a_q = a_1 + \gamma(q-1)$ for any $q \in \mathbb{N}^*$,

$$\begin{aligned}
f_i^* &= \sum_{r=1}^{\infty} \mathrm{f}\left(m_i - \gamma(r-1)\right), \qquad (28) \\
f_i^* &< \sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right). \qquad (29)
\end{aligned}$$

By the Poisson summation formula (see Bellman, 2013, Section 6)

$$\sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right) = \sum_{k=-\infty}^{\infty} \hat{\mathrm{f}}\left(m_i^* + \gamma k\right), \tag{30}$$

where $\hat{\mathrm{f}}(.)$ is the Fourier transform of $\mathrm{f}(.)$ given by

$$\hat{\mathrm{f}}\left(m_i^* + \gamma k\right) = \int_{-\infty}^{\infty} \mathrm{f}(\gamma x + m_i^*) e^{-2\pi i k x} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{1}{2\sigma_\varepsilon^2}(\gamma x + m_i^*)^2 - 2\pi i k x} dx. \tag{31}$$

In Equation (31), i as subscript refers to the individual i, whereas i in the calculations refers to the pure imaginary complex number, such that $i^2 = -1$.

$$\begin{aligned}
\hat{\mathrm{f}}\left(m_i^* + \gamma k\right) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{1}{2\sigma_\varepsilon^2}\left(\gamma^2 x^2 + 2m_i^*\gamma x + (m_i^*)^2 + 4\pi i \sigma_\varepsilon^2 k x\right)} dx, \\
\hat{\mathrm{f}}\left(m_i^* + \gamma k\right) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{1}{2\sigma_\varepsilon^2}\left(\gamma^2 x^2 + 2\left(m_i^*\gamma + 2\pi i \sigma_\varepsilon^2 k\right)x + (m_i^*)^2\right)} dx, \\
\hat{\mathrm{f}}\left(m_i^* + \gamma k\right) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{\gamma^2}{2\sigma_\varepsilon^2}\left(x + \frac{m_i^*\gamma + 2\pi i \sigma_\varepsilon^2 k}{\gamma^2}\right)^2 + \frac{1}{2\sigma_\varepsilon^2\gamma^2}\left(m_i^*\gamma + 2\pi i \sigma_\varepsilon^2 k\right)^2 - \frac{1}{2\sigma_\varepsilon^2}(m_i^*)^2} dx, \\
\hat{\mathrm{f}}\left(m_i^* + \gamma k\right) &= \frac{1}{\gamma} e^{\frac{1}{2\sigma_\varepsilon^2\gamma^2}\left(m_i^*\gamma + 2\pi i \sigma_\varepsilon^2 k\right)^2 - \frac{1}{2\sigma_\varepsilon^2}(m_i^*)^2} \underbrace{\int_{-\infty}^{\infty} \frac{\gamma}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{\gamma^2}{2\sigma_\varepsilon^2}\left(x + \frac{m_i^*\gamma + 2\pi i \sigma_\varepsilon^2 k}{\gamma^2}\right)^2} dx}_{=1}, \\
\hat{\mathrm{f}}\left(m_i^* + \gamma k\right) &= \frac{1}{\gamma} e^{\frac{1}{2\sigma_\varepsilon^2\gamma^2}\left(m_i^*\gamma + 2\pi i \sigma_\varepsilon^2 k\right)^2 - \frac{1}{2\sigma_\varepsilon^2}(m_i^*)^2}, \\
\hat{\mathrm{f}}\left(m_i^* + \gamma k\right) &= \frac{1}{\gamma} e^{-2\pi^2 k^2 \left(\frac{\sigma_\varepsilon}{\gamma}\right)^2 + 2\pi i k \frac{m_i^*}{\gamma}}.
\end{aligned} \tag{32}$$

By replacing the Fourier transform (32) in Equation (30), I have

$$\begin{aligned}
\sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right) &= \frac{1}{\gamma} \sum_{k=-\infty}^{\infty} e^{-2\pi^2 k^2 \left(\frac{\sigma_\varepsilon}{\gamma}\right)^2 + 2\pi i k \frac{m_i^*}{\gamma}} = \frac{1}{\gamma} \sum_{k=-\infty}^{\infty} e^{-2\pi^2 k^2 \left(\frac{\sigma_\varepsilon}{\gamma}\right)^2} e^{2\pi i k \frac{m_i^*}{\gamma}}, \\
\sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right) &= \frac{1}{\gamma} \left(1 + \sum_{k=-\infty}^{-1} e^{-2\pi^2 k^2 \left(\frac{\sigma_\varepsilon}{\gamma}\right)^2} e^{2\pi i k \frac{m_i^*}{\gamma}} + \sum_{k=1}^{\infty} e^{-2\pi^2 k^2 \left(\frac{\sigma_\varepsilon}{\gamma}\right)^2} e^{2\pi i k \frac{m_i^*}{\gamma}}\right), \\
\sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right) &= \frac{1}{\gamma} \left(1 + \sum_{k=1}^{\infty} e^{-2\pi^2 (-k)^2 \left(\frac{\sigma_\varepsilon}{\gamma}\right)^2} e^{-2\pi i k \frac{m_i^*}{\gamma}} + \sum_{k=1}^{\infty} e^{-2\pi^2 k^2 \left(\frac{\sigma_\varepsilon}{\gamma}\right)^2} e^{2\pi i k \frac{m_i^*}{\gamma}}\right), \\
\sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right) &= \frac{1}{\gamma} \left(1 + \sum_{k=1}^{\infty} e^{-2\pi^2 k^2 \left(\frac{\sigma_\varepsilon}{\gamma}\right)^2} \left(e^{-2\pi i k \frac{m_i^*}{\gamma}} + e^{2\pi i k \frac{m_i^*}{\gamma}}\right)\right).
\end{aligned} \tag{33}$$

Let us focus on the term $e^{-2\pi i k \frac{m_i^*}{\gamma}} + e^{2\pi i k \frac{m_i^*}{\gamma}}$. By Euler's formula (Zill and Shanahan, 2009),

$$\begin{aligned}
e^{-2\pi i k \frac{m_i^*}{\gamma}} + e^{2\pi i k \frac{m_i^*}{\gamma}} &= \cos\left(-2\pi k \frac{m_i^*}{\gamma}\right) + i \sin\left(-2\pi k \frac{m_i^*}{\gamma}\right) + \cos\left(2\pi k \frac{m_i^*}{\gamma}\right) + i \sin\left(2\pi k \frac{m_i^*}{\gamma}\right), \\
e^{-2\pi i k \frac{m_i^*}{\gamma}} + e^{2\pi i k \frac{m_i^*}{\gamma}} &= \cos\left(2\pi k \frac{m_i^*}{\gamma}\right) - i \sin\left(2\pi k \frac{m_i^*}{\gamma}\right) + \cos\left(2\pi k \frac{m_i^*}{\gamma}\right) + i \sin\left(2\pi k \frac{m_i^*}{\gamma}\right), \\
e^{-2\pi i k \frac{m_i^*}{\gamma}} + e^{2\pi i k \frac{m_i^*}{\gamma}} &= 2\cos\left(2\pi k \frac{m_i^*}{\gamma}\right).
\end{aligned} \tag{34}$$

By replacing (34) in (33), I get

$$
\begin{aligned}
\sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right) &= \frac{1}{\gamma}\left(1 + 2\sum_{k=1}^{\infty} e^{-2\pi^2 k^2 \left(\frac{\sigma_\varepsilon}{\gamma}\right)^2} \cos\left(2\pi k \frac{m_i^*}{\gamma}\right)\right), \\
\sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right) &\leq \frac{1}{\gamma}\left(1 + 2\sum_{k=1}^{\infty} e^{-2\pi^2 k^2 \left(\frac{\sigma_\varepsilon}{\gamma}\right)^2}\right) = \sum_{k=-\infty}^{\infty} \mathrm{f}\left(\gamma k\right), \qquad (35) \\
\sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right) &\leq \mathrm{f}\left(0\right) + \sum_{k=-\infty}^{-1} \mathrm{f}\left(\gamma k\right) + \sum_{k=1}^{\infty} \mathrm{f}\left(\gamma k\right), \\
\sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right) &\leq \mathrm{f}\left(0\right) + 2\sum_{k=1}^{\infty} \mathrm{f}\left(\gamma k\right), \quad \text{as f is symmetric.} \\
\sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right) &\leq \frac{\phi\left(0\right) + 2\sum_{k=1}^{\infty} \phi\left(\frac{\gamma k}{\sigma_\varepsilon}\right)}{\sigma_\varepsilon}, \\
\sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right) &\leq \frac{1}{C_{\gamma,\sigma_\varepsilon}}. \qquad (36)
\end{aligned}
$$

From Equations (29) and (36),

$$
f_i^* < \sum_{k=-\infty}^{\infty} \mathrm{f}\left(m_i^* + \gamma k\right) \leq \frac{1}{C_{\gamma,\sigma_\varepsilon}}. \qquad (37)
$$

From Equations (27) and (37),

$$
\begin{aligned}
\left|\left|\frac{\partial \mathbf{L}\left(\mathbf{u}\right)}{\partial \mathbf{u}'}\right|\right|_\infty &\leq |\lambda| \max_i |f_i^*| ||\mathbf{G}||_\infty, \\
\left|\left|\frac{\partial \mathbf{L}\left(\mathbf{u}\right)}{\partial \mathbf{u}'}\right|\right|_\infty &< \frac{|\lambda| ||\mathbf{G}||_\infty}{C_{\gamma,\sigma_\varepsilon}}, \\
\left|\left|\frac{\partial \mathbf{L}\left(\mathbf{u}\right)}{\partial \mathbf{u}'}\right|\right|_\infty &< 1 \text{ by Assumption 2 (ii).} \qquad (38)
\end{aligned}
$$

Hence, $\mathbf{u} = \mathbf{L}\left(\mathbf{u}\right)$ has a unique solution $\bar{\mathbf{y}}^e$.

By Equation (6) and Lemma 1, it follows that $\mathbf{p} = \mathbf{H}(\mathbf{p})$ also has a unique solution $\mathbf{p}^e$, such that $p_{iq}^e = \Phi\left(\frac{\lambda \mathbf{g}_i \bar{\mathbf{y}}_j^e + \psi_i - a_q}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\lambda \mathbf{g}_i \bar{\mathbf{y}}_j^e + \psi_i - a_{q+1}}{\sigma_\varepsilon}\right)$ due to the bijective correspondence between the equilibrium belief and the equilibrium expected outcome associated with that belief.

As a result, the incomplete information network game with the utility (1) has a unique pure strategy BNE with the equilibrium strategy profile $\mathbf{y}^{e*}$ given by $\mathbf{y}^{e*} = \lambda \mathbf{G} \bar{\mathbf{y}}^e + \boldsymbol{\psi} + \boldsymbol{\varepsilon}$, where the equilibrium belief system $\left(p_{iq}^e\right)_{\substack{i \in \mathcal{V} \\ q \in \mathbb{N}}}$ is such that $\bar{y}_i^e = \sum_{r=0}^{\infty} r p_{ir}^e$ is the unique solution of $\mathbf{u} = \mathbf{L}\left(\mathbf{u}\right)$.

A.3 Exact computation of the upper bound in Assumption 2 (ii)

It is clear that $C_{\gamma,\sigma_\varepsilon} = \dfrac{\sigma_\varepsilon}{\phi(0) + 2\sum_{k=1}^{\infty}\phi\left(\frac{\gamma k}{\sigma_\varepsilon}\right)} = \dfrac{1}{\sum_{k=-\infty}^{\infty}\mathrm{f}(\gamma k)}$.

The quantity $\displaystyle\sum_{k=-\infty}^{\infty}\mathrm{f}(\gamma k)$ can be computed exactly using the third Theta function (see Bellman, 2013, Section 2). From (35), it follows that

$$\sum_{k=-\infty}^{\infty}\mathrm{f}(\gamma k) = \frac{1}{\gamma}\left(1 + 2\sum_{k=1}^{\infty}e^{-2\pi^2k^2\left(\frac{\sigma_\varepsilon}{\gamma}\right)^2}\right),$$
$$\sum_{k=-\infty}^{\infty}\mathrm{f}(\gamma k) = \frac{1}{\gamma}\theta_3\left(0, e^{-2\pi^2\left(\frac{\sigma_\varepsilon}{\gamma}\right)^2}\right), \tag{39}$$

where for any complex z and $q \in \mathbb{R}_+$, $\theta_3(z,q)$ is the third Theta function evaluated at (z,q). Therefore, $C_{\gamma,\sigma_\varepsilon} = \dfrac{\gamma}{\theta_3\left(0, e^{-2\pi^2\left(\frac{\sigma_\varepsilon}{\gamma}\right)^2}\right)}$.

B Supplementary note on the econometric model

B.1 Consistency and limit distribution of the NLP estimator

The pseudo likelihood is given by

$$\mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}}) = \sum_{i=1}^{n}\sum_{r=0}^{\infty} d_{ir}\log\left\{\Phi\left(\frac{\mathbf{z}_i'\boldsymbol{\Lambda} - a_r}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\mathbf{z}_i'\boldsymbol{\Lambda} - a_{r+1}}{\sigma_\varepsilon}\right)\right\}, \tag{40}$$

where $\mathbf{z}_i' = (\mathbf{g}_i\bar{\mathbf{y}},\ \mathbf{x}_i')$, $\boldsymbol{\Lambda} = (\lambda,\ \boldsymbol{\beta}')'$, and $\boldsymbol{\theta} = (\boldsymbol{\Lambda},\ \sigma_\varepsilon)'$. Let $\boldsymbol{\theta}_0$ be the true value of $\boldsymbol{\theta}$, and $\bar{\mathbf{y}}_0$ be the expected outcome associated with $\boldsymbol{\theta}$. The first-order conditions of the pseudo likelihood maximization give

$$\begin{cases} \dfrac{\partial\mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}})}{\partial\boldsymbol{\Lambda}} = \displaystyle\sum_{i=1}^{n}\sum_{r=0}^{\infty} d_{ir}\frac{f_{ir} - f_{i(r+1)}}{F_{ir} - F_{i(r+1)}}\mathbf{z}_i = 0, \\ \dfrac{\partial\mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}})}{\partial\sigma_\varepsilon} = -\displaystyle\sum_{i=1}^{n}\sum_{r=0}^{\infty} d_{ir}\frac{m_{ir}f_{ir} - m_{i(r+1)}f_{i(r+1)}}{\sigma_\varepsilon\left(F_{ir} - F_{i(r+1)}\right)} = 0, \end{cases} \tag{41}$$

where $\forall\ i \in \mathcal{V}$, $q \in \mathbb{N}$, $m_{iq} = \mathbf{z}_i'\boldsymbol{\Lambda} - a_q$, $f_{iq} = \dfrac{1}{\sigma_\varepsilon}\phi\left(\dfrac{m_{iq}}{\sigma_\varepsilon}\right)$, and $F_{iq} = \Phi\left(\dfrac{m_{iq}}{\sigma_\varepsilon}\right)$. As $\mathcal{L}$ is continuous, the consistency of the NPL estimator is ensured by the fact that plim $\left(\dfrac{1}{n}\mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}})\right)$ is maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $\bar{\mathbf{y}} = \bar{\mathbf{y}}_0$, where plim stands for the probability limit.

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

$$\frac{\partial\mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}})}{\partial\boldsymbol{\theta}} = \left.\frac{\partial\mathcal{L}(\boldsymbol{\theta}, \bar{\mathbf{y}})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}_0} + \left(\left.\frac{\partial^2\mathcal{L}(\boldsymbol{\theta}_0, \bar{\mathbf{y}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_0} + \left.\frac{\partial^2\mathcal{L}(\boldsymbol{\theta}_0, \bar{\mathbf{y}})}{\partial\boldsymbol{\theta}\partial\bar{\mathbf{y}}'}\right|_{\boldsymbol{\theta}_0}\left.\frac{\partial\bar{\mathbf{y}}}{\partial\boldsymbol{\theta}'}\right|_{\boldsymbol{\theta}_0}\right)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + O_p(1). \tag{42}$$

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

$$\sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) = -\left(\frac{1}{n}\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}_0, \bar{\mathbf{y}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} + \frac{1}{n}\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}_0, \bar{\mathbf{y}})}{\partial \boldsymbol{\theta} \partial \bar{\mathbf{y}}'}\frac{\partial \bar{\mathbf{y}}_0}{\partial \boldsymbol{\theta}'}\right)^{-1}\left(\frac{1}{\sqrt{n}}\frac{\partial \mathcal{L}(\boldsymbol{\theta}_0, \bar{\mathbf{y}})}{\partial \boldsymbol{\theta}} + O_p\left(\frac{1}{\sqrt{n}}\right)\right). \tag{43}$$

Let us first apply the central Theorem limit to the term $\dfrac{1}{\sqrt{n}}\dfrac{\partial \mathcal{L}(\boldsymbol{\theta}_0, \bar{\mathbf{y}})}{\partial \boldsymbol{\theta}}$.

$$\frac{1}{\sqrt{n}}\frac{\partial \mathcal{L}(\boldsymbol{\theta}_0, \bar{\mathbf{y}})}{\partial \boldsymbol{\theta}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underbrace{\begin{pmatrix} \displaystyle\sum_{r=0}^{\infty} d_{ir}\frac{f_{ir}^0 - f_{i(r+1)}^0}{F_{ir}^0 - F_{i(r+1)}^0}\mathbf{z}_i \\ \displaystyle -\sum_{r=0}^{\infty} d_{ir}\frac{m_{ir}^0 f_{ir}^0 - m_{i(r+1)}^0 f_{i(r+1)}^0}{\sigma_\varepsilon\left(F_{ir}^0 - F_{i(r+1)}^0\right)} \end{pmatrix}}_{\mathbf{v}_i^0} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{v}_i^0,$$

where $\forall\ i \in \mathcal{V}$, $q \in \mathbb{N}$, m_{iq}^0, f_{iq}^0, and F_{iq}^0 are defined as in (41) but with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

$\mathbf{E}\left(\mathbf{v}_i^0 | \mathbf{X}, \mathbf{G}\right) = \begin{pmatrix} \displaystyle\sum_{r=0}^{\infty}\left(f_{ir}^0 - f_{i(r+1)}^0\right)\mathbf{z}_i \\ \displaystyle -\frac{1}{\sigma_\varepsilon}\sum_{r=0}^{\infty}\left(m_{ir}^0 f_{ir}^0 - m_{i(r+1)}^0 f_{i(r+1)}^0\right) \end{pmatrix} = 0$, thus $\mathbf{E}\left(\mathbf{v}_i^0\right) = 0$.

Let denote by $A_i = \displaystyle\sum_{r=0}^{\infty}\frac{\left(f_{ir}^0 - f_{i(r+1)}^0\right)^2}{F_{ir}^0 - F_{i(r+1)}^0}$, $B_i = \displaystyle\sum_{r=0}^{\infty}\frac{\left(m_{ir}^0 f_{ir}^0 - m_{i(r+1)}^0 f_{i(r+1)}^0\right)^2}{\sigma_\varepsilon^2\left(F_{ir}^0 - F_{i(r+1)}^0\right)}$, and

$C_i = -\displaystyle\sum_{r=0}^{\infty}\frac{\left(f_{ir}^0 - f_{i(r+1)}^0\right)\left(m_{ir}^0 f_{ir}^0 - m_{i(r+1)}^0 f_{i(r+1)}^0\right)}{\sigma_\varepsilon\left(F_{ir}^0 - F_{i(r+1)}^0\right)}$.

$$\mathbf{Var}\left(\mathbf{v}_i^0 | \mathbf{X}, \mathbf{G}\right) = \mathbf{E}\left(\mathbf{v}_i^0 \mathbf{v}_i^{0\prime} | \mathbf{X}, \mathbf{G}\right) = \underbrace{\begin{pmatrix} A_i \mathbf{z}_i \mathbf{z}_i' & C_i \mathbf{z}_i \\ C_i \mathbf{z}_i' & B_i \end{pmatrix}}_{\boldsymbol{\Sigma}_i} = \boldsymbol{\Sigma}_i. \tag{44}$$

By assuming that $\operatorname{plim}\left(\dfrac{1}{n}\displaystyle\sum_{i}^{n}\boldsymbol{\Sigma}_i\right)$ exists and is equal to $\boldsymbol{\Sigma}_0$, it follows by the Lindeberg–Feller central Theorem limit (see Chow and Teicher, 2003, p. 314) that

$$\frac{1}{\sqrt{n}}\frac{\partial \mathcal{L}(\boldsymbol{\theta}_0, \bar{\mathbf{y}})}{\partial \boldsymbol{\theta}} \xrightarrow{d} \mathcal{N}\left(0, \boldsymbol{\Sigma}_0\right). \tag{45}$$

Let us now focus on $\operatorname{plim}\left(\dfrac{1}{n}\dfrac{\partial^2 \mathcal{L}(\boldsymbol{\theta}_0, \bar{\mathbf{y}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right)$ and $\operatorname{plim}\left(\dfrac{1}{n}\dfrac{\partial^2 \mathcal{L}(\boldsymbol{\theta}_0, \bar{\mathbf{y}})}{\partial \boldsymbol{\theta} \partial \bar{\mathbf{y}}'}\dfrac{\partial \bar{\mathbf{y}}_0}{\partial \boldsymbol{\theta}'}\right)$.
By the law of large numbers applied to independent and non-identical variables (see Chow and Teicher, 2003, p. 124), $\operatorname{plim}\left(\dfrac{1}{n}\dfrac{\partial^2 \mathcal{L}(\boldsymbol{\theta}_0, \bar{\mathbf{y}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right) = \operatorname{plim}\left(\dfrac{1}{n}\mathbf{E}_d\left(\dfrac{\partial^2 \mathcal{L}(\boldsymbol{\theta}_0, \bar{\mathbf{y}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right)\right)$, where $\mathbf{E}_d$ is the expectation with

respect to d_{ir}.

$$\mathbf{E}_d\left(\frac{\partial^2\mathcal{L}(\boldsymbol{\theta}_0,\bar{\mathbf{y}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) = -\sum_{i=1}^{n}\boldsymbol{\Sigma}_i \implies \operatorname{plim}\left(\frac{1}{n}\frac{\partial^2\mathcal{L}(\boldsymbol{\theta}_0,\bar{\mathbf{y}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right) = -\operatorname{plim}\left(\frac{1}{n}\sum_{i}^{n}\boldsymbol{\Sigma}_i\right) = -\boldsymbol{\Sigma}_0. \tag{46}$$

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

$$\mathbf{E}_d\left(\frac{\partial^2\mathcal{L}(\boldsymbol{\theta}_0,\bar{\mathbf{y}})}{\partial\boldsymbol{\theta}\partial\bar{\mathbf{y}}'}\right) = -\lambda\sum_{i=1}^{n}\begin{pmatrix}A_i\mathbf{z}_i\mathbf{g}_i\\ B_i\mathbf{g}_i\end{pmatrix} \quad \text{and} \quad \frac{\partial\bar{\mathbf{y}}_0}{\partial\boldsymbol{\theta}'} = \mathbf{S}^{-1}\mathbf{M}, \tag{47}$$

where $\mathbf{S} = \mathbf{I}_n - \lambda\mathbf{DG}$, $\mathbf{I}_n$ is the identity matrix of dimension n, $\mathbf{D} = \operatorname{diag}\left(\sum_{r=1}^{\infty}f_{1r}^0,\ \dots,\ \sum_{r=1}^{\infty}f_{nr}^0\right)$, $\mathbf{M} = (\mathbf{DZ},\mathbf{b})$, $\mathbf{Z} = (\mathbf{G}\bar{\mathbf{y}},\mathbf{X})$, and $\mathbf{b} = \left(-\sum_{r=1}^{\infty}\frac{f_{1r}^0 m_{1r}^0}{\sigma_\varepsilon},\ \dots,\ -\sum_{r=1}^{\infty}\frac{f_{nr}^0 m_{nr}^0}{\sigma_\varepsilon}\right)'$. The partial derivative $\frac{\partial\bar{\mathbf{y}}_0}{\partial\boldsymbol{\theta}'}$ is computed using the implicit definition of $\bar{\mathbf{y}}$; that is, $\bar{\mathbf{y}} = \mathbf{L}(\bar{\mathbf{y}},\boldsymbol{\theta})$.

Assuming that $\operatorname{plim}\left(\frac{\lambda}{n}\sum_{i=1}^{n}\begin{pmatrix}A_i\mathbf{z}_i\mathbf{g}_i\mathbf{S}^{-1}\mathbf{M}\\ B_i\mathbf{g}_i\mathbf{S}^{-1}\mathbf{M}\end{pmatrix}\right)$ exists and is equal to $\boldsymbol{\Omega}_0$,

$$\operatorname{plim}\left(\frac{1}{n}\frac{\partial^2\mathcal{L}(\boldsymbol{\theta}_0,\bar{\mathbf{y}})}{\partial\boldsymbol{\theta}\partial\bar{\mathbf{y}}'}\frac{\partial\bar{\mathbf{y}}_0}{\partial\boldsymbol{\theta}'}\right) = -\boldsymbol{\Omega}_0. \tag{48}$$

From Equations (43), (45), (46), and (48), it follows that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N}\left(0,\ (\boldsymbol{\Sigma}_0 + \boldsymbol{\Omega}_0)^{-1}\boldsymbol{\Sigma}_0\left(\boldsymbol{\Sigma}_0' + \boldsymbol{\Omega}_0'\right)^{-1}\right). \tag{49}$$

In a finite sample, an estimator of the asymptotic variance of $\hat{\boldsymbol{\theta}}$ can be computed by

$$\widehat{AsyVar}\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{n}\left(\hat{\boldsymbol{\Sigma}} + \hat{\boldsymbol{\Omega}}\right)^{-1}\hat{\boldsymbol{\Sigma}}\left(\hat{\boldsymbol{\Sigma}}' + \hat{\boldsymbol{\Omega}}'\right)^{-1}, \tag{50}$$

where $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i}^{n}\hat{\boldsymbol{\Sigma}}_i$, $\hat{\boldsymbol{\Omega}} = \frac{\hat{\lambda}}{n}\sum_{i=1}^{n}\begin{pmatrix}\hat{A}_i\mathbf{z}_i\mathbf{g}_i\hat{\mathbf{S}}^{-1}\hat{\mathbf{M}}\\ \hat{B}_i\mathbf{g}_i\hat{\mathbf{S}}^{-1}\hat{\mathbf{M}}\end{pmatrix}$, and $\hat{\boldsymbol{\Sigma}}_i$, $\hat{A}_i$, $\hat{B}_i$, $\hat{\mathbf{S}}$, $\hat{\mathbf{M}}$ are defined as above by replacing $\boldsymbol{\theta}_0$ by $\hat{\boldsymbol{\theta}}$.

B.2 Marginal effects and corresponding standard errors

The parameters $\boldsymbol{\theta}$ cannot be interpreted directly. Policy makers may be interested in the marginal effect of the explanatory variables on the expected outcome.

Let us recall the following notations: $\mathbf{z}_i' = (\mathbf{g}_i\bar{\mathbf{y}},\ \mathbf{x}_i')$ and $\boldsymbol{\Lambda} = (\lambda,\ \boldsymbol{\beta}')'$. For any $k = 1,\dots,K+1$, let

λ_k and z_{ik} be the k-th component in $\mathbf{\Lambda}$ and $\mathbf{z}_i$, respectively. The marginal effect of the explanatory variable z_{ik} on $\bar{y}_i$, the expected outcome of the individual i is given by

$$\delta_{ik}(\boldsymbol{\theta}) = \frac{\partial \bar{y}_i}{\partial z_{ik}} = \frac{\lambda_k}{\sigma_\varepsilon} \sum_{r=1}^{\infty} \phi\left(\frac{\mathbf{z}_i'\mathbf{\Lambda} - a_r}{\sigma_\varepsilon}\right). \tag{51}$$

As mentioned in Section 2, if for some $q \in \mathbb{N}^*$, $\mathbf{z}_i'\mathbf{\Lambda} - a_q = 0$, then $\delta_{i1}(\boldsymbol{\theta}) > \dfrac{\lambda}{\sqrt{2\pi}\sigma_\varepsilon}$, and $\delta_{i1}(\boldsymbol{\theta})$ explodes if $\dfrac{\lambda}{\sigma_\varepsilon}$ is not constrained. Assumption 2 prevents such cases.

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

The Taylor expansion of Equation (51) around $\boldsymbol{\theta}_0$ is

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

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

As $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$ for large n,

$$\begin{aligned}
\delta_{ik}(\hat{\boldsymbol{\theta}}) &\approx \delta_{ik}(\boldsymbol{\theta}_0) + \frac{\partial \delta_{ik}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \\
\delta_{ik}(\hat{\boldsymbol{\theta}}) &\approx \delta_{ik}(\boldsymbol{\theta}_0) + \left(\frac{\partial \delta_{ik}(\boldsymbol{\theta}_0)}{\partial \mathbf{\Lambda}'}, \quad \frac{\partial \delta_{ik}(\boldsymbol{\theta}_0)}{\partial \sigma_\varepsilon}\right)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).
\end{aligned} \tag{52}$$

It follows that a consistent estimator of the standard error of $\delta_{ik}(\hat{\boldsymbol{\theta}})$ is

$$Se\left(\delta_{ik}(\hat{\boldsymbol{\theta}})\right) = \sqrt{\left(\frac{\partial \delta_{ik}(\hat{\boldsymbol{\theta}})}{\partial \mathbf{\Lambda}'}, \quad \frac{\partial \delta_{ik}(\hat{\boldsymbol{\theta}})}{\partial \sigma_\varepsilon}\right) \widehat{AsyVar}\left(\hat{\boldsymbol{\theta}}\right) \left(\frac{\partial \delta_{ik}(\hat{\boldsymbol{\theta}})}{\partial \mathbf{\Lambda}'}, \quad \frac{\partial \delta_{ik}(\hat{\boldsymbol{\theta}})}{\partial \sigma_\varepsilon}\right)'}, \tag{53}$$

where

$$\frac{\partial \delta_{ik}(\hat{\boldsymbol{\theta}})}{\partial \mathbf{\Lambda}'} = \frac{\mathbf{e}_k}{\sigma_\varepsilon} \sum_{r=1}^{\infty} \phi\left(\frac{\mathbf{z}_i'\hat{\mathbf{\Lambda}} - a_r}{\sigma_\varepsilon}\right) - \frac{\lambda_k}{\sigma_\varepsilon^3}\mathbf{z}_i' \sum_{r=1}^{\infty} \left(\mathbf{z}_i'\hat{\mathbf{\Lambda}} - a_r\right) \phi\left(\frac{\mathbf{z}_i'\hat{\mathbf{\Lambda}} - a_r}{\sigma_\varepsilon}\right), \tag{54}$$

$$\frac{\partial \delta_{ik}(\hat{\boldsymbol{\theta}})}{\partial \sigma_\varepsilon} = \frac{\lambda_k}{\sigma_\varepsilon^4} \sum_{r=1}^{\infty} \left(\mathbf{z}_i'\hat{\mathbf{\Lambda}} - a_q\right)^2 \phi\left(\frac{\mathbf{z}_i'\hat{\mathbf{\Lambda}} - a_r}{\sigma_\varepsilon}\right) - \frac{\lambda_k}{\sigma_\varepsilon^2} \sum_{r=1}^{\infty} \phi\left(\frac{\mathbf{z}_i'\hat{\mathbf{\Lambda}} - a_r}{\sigma_\varepsilon}\right), \tag{55}$$

where $\mathbf{e}_k$ is a row vector of dimension K + 1 with the k-th term equal to one and the other terms equal to 0.

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

$$Se\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{ik}(\hat{\boldsymbol{\theta}})\right) = \sqrt{Q_{\boldsymbol{\theta}} * \widehat{AsyVar} * Q'_{\boldsymbol{\theta}}}, \tag{56}$$

and

$$Q_{\boldsymbol{\theta}} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\delta_{ik}(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\Lambda}'}, \quad \frac{1}{n}\sum_{i=1}^{n}\frac{\partial\delta_{ik}(\hat{\boldsymbol{\theta}})}{\partial\sigma_{\varepsilon}}\right). \tag{57}$$

C Data summary

This section summarizes the data (see Table 7). The categorical explanatory variables are discretized into several binary subvariables. For identification, the subvariables in italics are the omitted categories in the econometric models.

Table 7: Data summary

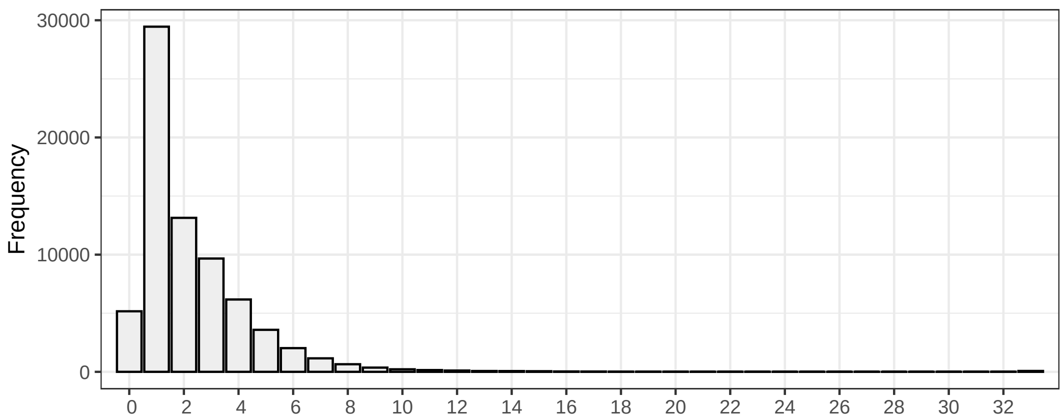
Variable	Mean	Sd.	Min	1st Qu.	Median	3rd Qu.	Max
Age	15.010	1.709	10	14	15	16	19
Sex							
Female	0.503	0.500	0	0	1	1	1
Male	0.497	0.500	0	0	0	1	1
Hispanic	0.168	0.374	0	0	0	0	1
Race							
White	0.625	0.484	0	0	1	1	1
Black	0.185	0.388	0	0	0	0	1
Asian	0.071	0.256	0	0	0	0	1
Other	0.097	0.296	0	0	0	0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

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

D Supplementary note on network endogeneity

In this section, I present the posterior distribution of the dyadic linking model parameters and show how to simulate from this posterior distribution. I also present the new asymptotic variance of $\hat{\boldsymbol{\theta}}$,

Figure 3: Distribution of the number of extracurricular activities



which includes the variability of $\tilde{\mu}_i$.

D.1 Posterior distribution of the dyadic linking model parameters

The likelihood of the model (16) is given by

$$\mathcal{L}(\mathbf{A}|\Delta\mathbf{X}, \bar{\boldsymbol{\beta}}, \boldsymbol{\mu}) = \prod_{s=1}^{S}\prod_{i\neq j}\frac{\exp\left(a_{ijs}(\Delta\mathbf{x}'_{ijs}\bar{\boldsymbol{\beta}} + \mu_{is} + \mu_{js})\right)}{1 + \exp\left(\Delta\mathbf{x}'_{ijs}\bar{\boldsymbol{\beta}} + \mu_{is} + \mu_{js}\right)},$$

where $\mathbf{X}$ is the matrix of dyad-specific variables, $\boldsymbol{\mu}$ is the vector of unobserved individual-level attributes, and the subscript s is used to denote the school s. The number of schools is S.
The joint distribution of $(\mathbf{A}, \boldsymbol{\mu})$ conditionally on $\boldsymbol{\Theta} = \left(\Delta\mathbf{X}, \bar{\boldsymbol{\beta}}, u_{\mu 1}, \sigma^2_{\mu 1}, \dots, u_{\mu S}, \sigma^2_{\mu S}\right)$ can be defined by

$$\pi(\mathbf{A}, \boldsymbol{\mu}|\boldsymbol{\Theta}) \propto \prod_{s=1}^{S}\left(\prod_{i\neq j}\frac{\exp\left(a_{ijs}(\Delta\mathbf{x}'_{ijs}\bar{\boldsymbol{\beta}} + \mu_{is} + \mu_{js})\right)}{1 + \exp\left(\Delta\mathbf{x}'_{ijs}\bar{\boldsymbol{\beta}} + \mu_{is} + \mu_{js}\right)}\prod_{i=1}^{n_s}\frac{1}{\sigma_{\mu s}}\exp\left(-\frac{1}{\sigma^2_{\mu s}}(\mu_{is} - u_{\mu s})^2\right)\right),$$

where n_s is the number of students in the school s.
I set a non-informative prior distribution on $\bar{\boldsymbol{\beta}}$ and conjugate prior on $\left(u_{\mu s}, \sigma^2_{\mu s}\right)$; that is, $\pi\left(\bar{\boldsymbol{\beta}}\right) \propto 1$ and $\pi\left(u_{\mu s}, \sigma^2_{\mu s}\right) \propto \dfrac{1}{\sigma_{\mu s}}$. Let Ξ be the vector containing $\bar{\boldsymbol{\beta}}$, $\boldsymbol{\mu}$, $u_{\mu 1}, \sigma^2_{\mu 1}, \dots, u_{\mu S}, \sigma^2_{\mu S}$. The posterior distribution of Ξ is

$$\pi(\Xi|\mathbf{A}, \Delta\mathbf{X}) \propto \prod_{s=1}^{S}\left(\frac{1}{\sigma^{n_s+1}_{\mu s}}\prod_{i\neq j}\frac{\exp\left(a_{ijs}(\Delta\mathbf{x}'_{ijs}\bar{\boldsymbol{\beta}} + \mu_{is} + \mu_{js})\right)}{1 + \exp\left(\Delta\mathbf{x}'_{ijs}\bar{\boldsymbol{\beta}} + \mu_{is} + \mu_{js}\right)}\prod_{i=1}^{n_s}\exp\left(-\frac{1}{\sigma^2_{\mu s}}(\mu_{is} - u_{\mu s})^2\right)\right).$$

To simulate from this posterior distribution, I use a MCMC approach that combines a Metropolis–Hasting (Metropolis et al., 1953) and a Gibbs sampler (Gelfand and Smith, 1990).

Algorithm 1. *The MCMC goes as follows:*

1. *Initialize* $\bar{\boldsymbol{\beta}}$, $\boldsymbol{\mu}$, $u_{\mu 1}, \sigma^2_{\mu 1}, \dots, u_{\mu S}, \sigma^2_{\mu S}$ *to* $\bar{\boldsymbol{\beta}}^{(0)}$, $\boldsymbol{\mu}^{(0)}$, $u^{(0)}_{\mu 1}, \sigma^{2(0)}_{\mu 1}, \dots, u^{(0)}_{\mu S}, \sigma^{2(0)}_{\mu S}$, *respectively.*

2. *For* t*, from 1 to* T*, where* T *is the number of simulations,*

   (a) *Draw the proposal* $\bar{\boldsymbol{\beta}}^*$ *from* $\mathcal{N}\left(\bar{\boldsymbol{\beta}}^{(t-1)}, \textit{jumping scale}\right)$. *Update* $\bar{\boldsymbol{\beta}}^{(t)}$ *by accepting* $\bar{\boldsymbol{\beta}}^*$ *with the probability* $\min\left\{1, \alpha_{\bar{\boldsymbol{\beta}}}\right\}$*, where*

$$\alpha_{\bar{\boldsymbol{\beta}}} = \prod_{s=1}^{S}\prod_{i\neq j} \frac{\exp\left(a_{ijs}\Delta\mathbf{x}'_{ijs}\bar{\boldsymbol{\beta}}^*\right)\left(1+\exp\left(\Delta\mathbf{x}'_{ijs}\bar{\boldsymbol{\beta}}^{(t-1)} + \mu^{(t-1)}_{is} + \mu^{(t-1)}_{js}\right)\right)}{\exp\left(a_{ijs}\Delta\mathbf{x}'_{ijs}\bar{\boldsymbol{\beta}}^{(t-1)}\right)\left(1+\exp\left(\Delta\mathbf{x}'_{ijs}\bar{\boldsymbol{\beta}}^* + \mu^{(t-1)}_{is} + \mu^{(t-1)}_{js}\right)\right)}.$$

   (b) *For* $s = 1, \dots, S$ *and* $i = 1, \dots, n_s$*, draw the proposal* μ^*_{is} *from* $\mathcal{N}\left(\mu^{(t-1)}_{is}, \textit{jumping scale}\right)$. *Update* $\mu^{(t)}_{is}$ *by accepting* μ^*_{is} *with the probability* $\min\left\{1, \alpha_{\mu_{is}}\right\}$*, where*

$$\alpha_{\mu_{is}} = \exp\left(\frac{1}{\sigma^{2(t-1)}_{\mu s}}\left(\mu^{(t-1)}_{is} - u^{(t-1)}_{\mu s}\right)^2 - \frac{1}{\sigma^{2(t-1)}_{\mu s}}\left(\mu^*_{is} - u^{(t-1)}_{\mu s}\right)^2\right)\times$$
$$\prod_{j\neq i} \frac{\exp\left(a_{ijs}\mu^*_{is}\right)\left(1+\exp\left(\Delta\mathbf{x}'_{ijs}\bar{\boldsymbol{\beta}}^{(t)} + \mu^{(t-1)}_{is} + \mu^*_{js}\right)\right)}{\exp\left(a_{ijs}\mu^{(t-1)}_{is}\right)\left(1+\exp\left(\Delta\mathbf{x}'_{ijs}\bar{\boldsymbol{\beta}}^{(t)} + \mu^*_{is} + \mu^*_{js}\right)\right)},$$

   and $\mu^*_{js} = \mu^{(t-1)}_{js}$*, if* $i < j$*, and* $\mu^*_{js} = \mu^{(t)}_{js}$*, if* $i > j$*.*

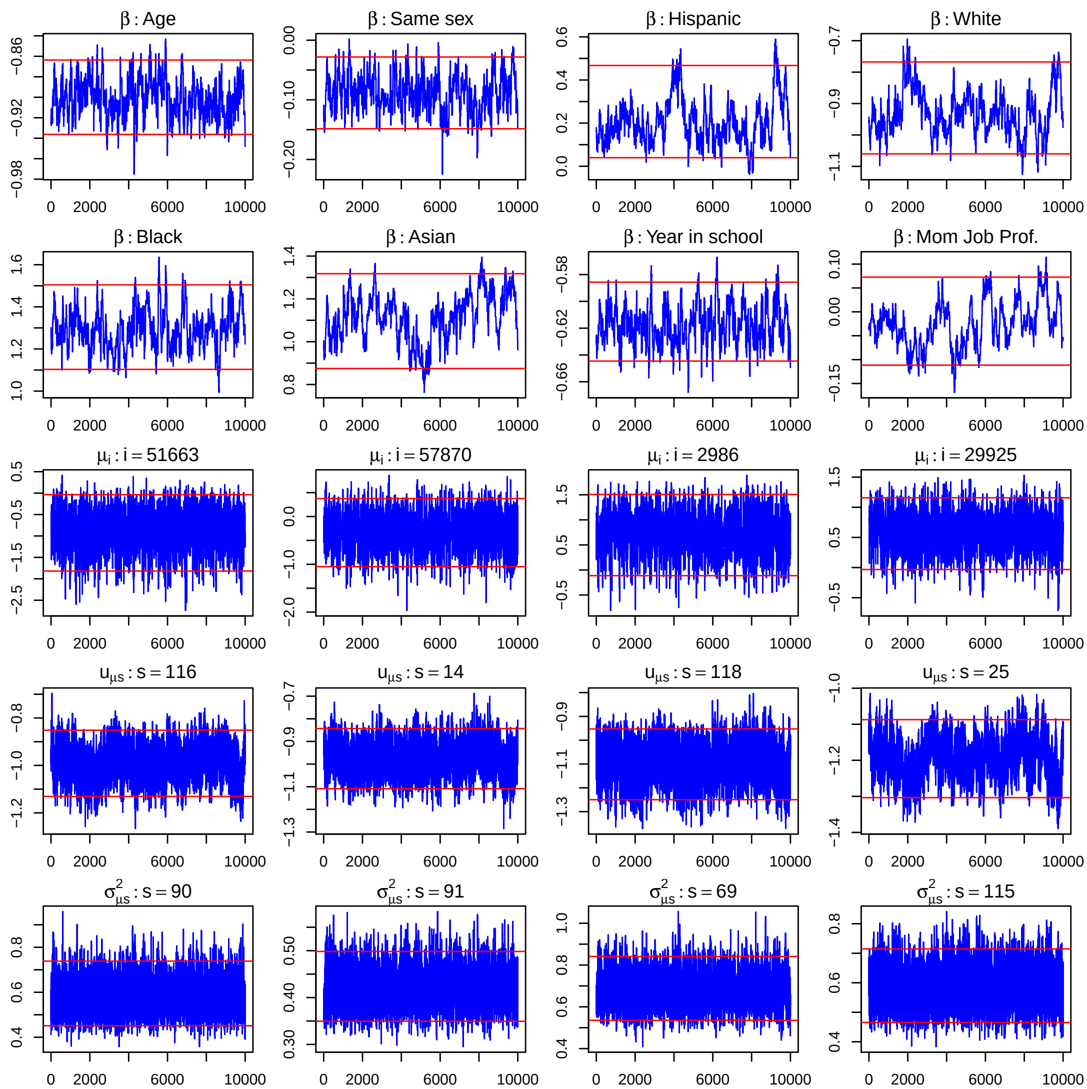
   (c) *For* $s = 1, \dots, S$*, use a Gibbs to update* $u^{(t)}_{\mu s}$ *from* $\mathcal{N}\left(\frac{\sum_{i=1}^{n_s}\mu^{(t)}_{is}}{n_s}, \frac{\sigma^{2(t-1)}_{\mu s}}{n_s}\right)$.

   (d) *For* $s = 1, \dots, S$*, use a Gibbs to update* $\sigma^{2(t)}_{\mu s}$ *from* $Inv-\chi^2\left(n_s - 1, \sum_{i=1}^{n_s}\left(\mu^{(t)}_{is} - u^{(t)}_{\mu s}\right)^2\right)$.

   (e) *Update the jumping scales following Atchadé et al. (2005) to reach an acceptance rate equal to 0.27.*

In practice the MCMC converges very quickly. I perform T = 20,000 simulations and keep the last 10,000. As the number of parameters in the model is large (72,291 parameters μ_i, 120 parameters $u_{\mu s}$, 120 parameters $\sigma^2_{\mu s}$ and, an eight-dimensional vector $\bar{\boldsymbol{\beta}}$), I randomly choose some parameters and present their posterior distribution in Figure 4.

Figure 4: Posterior distribution of the network formation model parameters



This figure presents the posterior distribution of the coefficients of the observed dyad-specific variables as well as some other parameters chosen at random. Students of similar age, Hispanic, Black, and Asian students, as well as students who have spent a similar number of years at their current school are likely to form links. In contrast, students of the same sex and white students are not likely to form links.

D.2 Correction of the asymptotic variance

As the estimation in done in two steps, the uncertainty related to $\tilde{\boldsymbol{\mu}}$ should be taken into account to correct the variance of the estimator at the second stage. The asymptotic variance, derived in Appendix B.1, is conditional on the explanatory variables, which include estimations of $\tilde{\boldsymbol{\mu}}$. In other words, the covariance of the estimator of $\hat{\theta}$ resulting from the NPL approach is given by $\mathbf{Var}\left(\hat{\theta}|\mathbf{G}, \mathbf{X}, \tilde{\boldsymbol{\mu}}\right)$ and not $\mathbf{Var}\left(\hat{\theta}|\mathbf{G}, \mathbf{X}\right)$.

To simplify the notations, I omit conditioning on $\mathbf{G}$ and $\mathbf{X}$ in this section; that is, I write $\mathbf{Var}\left(\hat{\theta}|\tilde{\boldsymbol{\mu}}\right)$ to mean $\mathbf{Var}\left(\hat{\theta}|\mathbf{G}, \mathbf{X}, \tilde{\boldsymbol{\mu}}\right)$ and $\mathbf{Var}\left(\hat{\theta}\right)$ to mean $\mathbf{Var}\left(\hat{\theta}|\mathbf{G}, \mathbf{X}\right)$. Moreover, $\mathbf{E}_u$ (respectively $\mathbf{Var}_u$) means that the expectation (respectively variance) is taken with respect to $\tilde{\boldsymbol{\mu}}$. It follows that

$$\begin{aligned}
\mathbf{Var}(\hat{\theta}) &= \mathbf{E}(\hat{\theta}\hat{\theta}^{\prime}) - \mathbf{E}(\hat{\theta})\,\mathbf{E}(\hat{\theta})^{\prime}, \\
\mathbf{Var}(\hat{\theta}) &= \mathbf{E}_u\left(\mathbf{E}(\hat{\theta}\hat{\theta}^{\prime}|\tilde{\boldsymbol{\mu}})\right) - \mathbf{E}(\hat{\theta})\,\mathbf{E}(\hat{\theta})^{\prime}, \\
\mathbf{Var}(\hat{\theta}) &= \mathbf{E}_u\left(\mathbf{E}(\hat{\theta}\hat{\theta}^{\prime}|\tilde{\boldsymbol{\mu}})\right) + \mathbf{E}_u\left(\mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})\,\mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})^{\prime}\right) - \mathbf{E}_u\left(\mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})\,\mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})^{\prime}\right) - \mathbf{E}(\hat{\theta})\,\mathbf{E}(\hat{\theta})^{\prime}, \\
\mathbf{Var}(\hat{\theta}) &= \mathbf{E}_u\Big(\underbrace{\mathbf{E}(\hat{\theta}\hat{\theta}^{\prime}|\tilde{\boldsymbol{\mu}}) - \mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})\,\mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})^{\prime}}_{\mathbf{Var}(\hat{\theta}|\tilde{\boldsymbol{\mu}})}\Big) + \underbrace{\mathbf{E}_u\left(\mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})\,\mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})^{\prime}\right) - \mathbf{E}(\hat{\theta})\,\mathbf{E}(\hat{\theta})^{\prime}}_{\mathbf{Var}_u\left(\mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})\right)}, \\
\mathbf{Var}(\hat{\theta}) &= \mathbf{E}_u\left(\mathbf{Var}(\hat{\theta}|\tilde{\boldsymbol{\mu}})\right) + \mathbf{Var}_u\left(\mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})\right). \qquad (58)
\end{aligned}$$

In Equation (58), the first component of the variance, $\mathbf{E}_u\left(\mathbf{Var}(\hat{\theta}|\tilde{\boldsymbol{\mu}})\right)$ is the variance of $\hat{\theta}$ due to the NPL algorithm. This component does not include the uncertainty of $\tilde{\boldsymbol{\mu}}$. The second component of the variance $\mathbf{Var}_u\left(\mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})\right)$ is the variance due to the estimation of $\tilde{\boldsymbol{\mu}}$ at the first stage. To compute the second component of the variance, I make the following Assumption.

Assumption 3. *Let $\tilde{\boldsymbol{\mu}}_s$ be a draw of $\tilde{\boldsymbol{\mu}}$ from its posterior distribution and $\hat{\theta}_s$ be the estimator of θ_0 associated with $\tilde{\boldsymbol{\mu}}_s$. $\hat{\theta}_s$ is a consistent estimator of $\mathbf{E}(\hat{\theta}_s|\tilde{\boldsymbol{\mu}}_s)$.*

Assumption 3 means that every estimator $\hat{\theta}_s$ associated with a draw $\tilde{\boldsymbol{\mu}}_s$ is a good estimator of $\mathbf{E}(\hat{\theta}_s|\tilde{\boldsymbol{\mu}}_s)$. This is useful because with many draws $\tilde{\boldsymbol{\mu}}_s$ the sample variance of $\hat{\theta}_s$ will be a good estimator of $\mathbf{Var}_u\left(\mathbf{E}(\hat{\theta}|\tilde{\boldsymbol{\mu}})\right)$. I also assume that the last 10,000 simulations from the posterior distribution at the first stage are sufficient to summarize well the posterior distribution of $\tilde{\boldsymbol{\mu}}_s$. Under these considerations, the variance of $\hat{\theta}_s$ is

$$\widehat{AsyVar}\left(\hat{\theta}_s\right) = \frac{1}{S}\sum_{s=1}^{S}\mathbf{Var}(\hat{\theta}_s|\tilde{\boldsymbol{\mu}}_s) + \frac{1}{S-1}\sum_{s=1}^{T}\left(\hat{\theta}_s - \hat{\bar{\theta}}\right)\left(\hat{\theta}_s - \hat{\bar{\theta}}\right)^{\prime}, \qquad (59)$$

where $\tilde{\boldsymbol{\mu}}_1, \ldots, \tilde{\boldsymbol{\mu}}_S$ are S draws of $\tilde{\boldsymbol{\mu}}$ with replacement from the population of the 10,000 simulations kept at the first stage, and $\hat{\hat{\theta}} = \frac{1}{S}\sum_{s=1}^{S}\hat{\theta}_s$. In practice, I set $S = 5,000$.

In Table 6, I present the average $\hat{\hat{\theta}}$ and the variance $\widehat{AsyVar}\left(\hat{\theta}_s\right)$ to summarize the distribution of $\hat{\theta}_s$. The same approach is used to compute the standard error of the marginal effects.