Journal of Econometrics 238 (2024) 105551

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Binary choice with misclassification and social interactions, with an application to peer effects in attitude ☆

Zhongjian Lin [a,*], Yingyao Hu [b]

[a] John Munro Godfrey, Sr. Department of Economics, University of Georgia, 620 S Lumpkin St, Athens, GA 30602, United States
[b] Department of Economics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, United States

ARTICLE INFO

JEL classification:
C25
C57
C63
I20

Keywords:
Measurement error
Misclassification
Social interactions
Nested pseudo likelihood
Large network
Peer effects
Attitude

ABSTRACT

The interaction of economic agents is one of the most important elements in economic analyses. Social interactions on subjective outcomes, behavior, or decisions, are inherently difficult to identify and estimate because these variables are prone to misclassification errors. This paper puts forth a binary choice model with misclassification and social interactions to rectify the misclassification problems in social interactions studies. We achieve the identification of the conditional choice probability of the latent dependent variable by the technique of repeated measurements and a monotonicity condition. We construct the complete likelihood function from the two repeated measurements and propose a nested pseudo likelihood algorithm for estimation. Consistency and asymptotic normality results are shown for the proposed estimation method. We illustrate the finite sample performance of the model and the estimation method by three Monte Carlo experiments and an application to the study of peer effects among students in their attitudes towards learning.

1. Introduction

Models with strategic interactions; e.g., peer effects, competitive effects, etc., have been estimated across many fields in economics, including financial economics, industrial organization, labor economics, and socioeconomics. Much of the existing empirical work has taken the behavior and decisions data as accurately measured; however, behavior and decisions data usually suffer from measurement errors when drawn from surveys. When modeling strategic interactions, peers'/competitors' decisions enter utility/profit functions, which lead to a simultaneous equation system of conditional choice probabilities. With mismeasured decision variables, the simultaneity of a strategic interactions study naturally raises problems from both the left and right (Hausman, 2001). In this paper, we propose a binary choice model with misclassification and social interactions and use the model to analyze the peer effects among students in their attitudes towards learning. We rectify the biases due to misclassification errors by using a tool with two repeated measurements and a monotonicity condition. We find both significant overreporting and underreporting in attitude and recover the hidden peer effects among students in their attitudes towards learning.

Measurement errors prevail in the survey data for economics analyses. There are four sources of measurement errors: mistakes

☆ The paper benefited greatly from comments and suggestions provided by Editor Elie Tamer, an associate editor, three referees, Yonghong An, Xiaodong Liu, Konrad Menzel, Ning Yu, as well as audiences at the Texas A&M University, 2019 ES winter meeting, 2018 ES China meeting, 2018 Kansas Econometrics Workshop. All errors are our own.

* Corresponding author.
E-mail address: zhongjian.lin@uga.edu (Z. Lin).

https://doi.org/10.1016/j.jeconom.2023.105551
Received 29 April 2020; Received in revised form 12 September 2023; Accepted 3 October 2023
Available online 9 November 2023
0304-4076/

made during the cognitive processes of answering survey questions; social desirability for some answers; essential survey conditions; and applicability of findings to the measurement of economic phenomena [see Bound et al. (2001) for details]. For continuous mismeasured variables, there are methods of decomposition (Hu and Schennach, 2008; Li and Vuong, 1998; Schennach, 2004) and auxiliary data (Carroll et al., 2010; Chen et al., 2005; Ridder and Moffitt, 2007) to deal with the measurement errors in nonlinear models. Decision variables possess discreteness and are sometimes dichotomous. A discrete measurement error is also called a misclassification error. Typically, discrete decision variables require nonlinear techniques that are different from those deployed in linear models. Econometricians have devoted increasing attention to the magnitude and consequences of measurement errors in nonlinear models; see details in Chen et al. (2011), Schennach (2016), Hu (2017) and references therein.

A major development of solutions to misclassification errors is on the right-hand side with few exceptions; for examples, see Hausman et al. (1998), Lewbel (2000) and Meyer and Mittag (2017) for binary choice models; Hsiao and Sun (1998) for multinomial models; Abrevaya and Hausman (1999) for duration models; and Li et al. (2003), Cameron et al. (2004) for count models. In the continuous setting, Lewbel (1996) and De Nadai and Lewbel (2016) investigate the measurement errors on both sides of the regression. Unlike in linear models where measurement errors occurring on the left-hand side cause only efficiency loss, there is a sizable distortion in the econometric analysis of nonlinear models with misclassification errors on the dependent variable. This paper attempts to study a case where there are misclassification errors on both sides of the regression due to the simultaneity of strategic interactions (static simultaneous game).

In the last three decades, tremendous attention was paid to social interactions and peer effects[1] among economic agents in many fields; e.g., education, production adoption, information diffusion, financial decisions, word of mouth, etc. Scholars propose many mechanisms to understand the peer effects; such as social image concern and signaling mechanism (Breza and Chandrasekhar, 2019; Bursztyn et al., 2019; Bursztyn and Jensen, 2015; 2017); mutual insurance (De Giorgi and Pellizzari, 2013); social learning (Bandura and Walters, 1977; Bursztyn et al., 2014; Cai et al., 2009; Moretti, 2011); and social utility (Brock and Durlauf, 2001a; Bursztyn et al., 2014; Gilchrist and Sands, 2016).

Brock and Durlauf (2001a,b) pioneer the discrete choice analysis with social interactions (discrete game). Brock and Durlauf (2001a, 2007) provide a novel equilibrium characterization of the discrete game and the identification strategies for unique equilibrium and multiple equilibria. For more discussion on the identification of discrete choice with social interactions and the linear social interactions model, see Blume et al. (2011); Durlauf and Ioannides (2010) and Blume et al. (2015). This paper generalizes to the case where there are misclassification errors in the binary dependent variable on the left and peers' binary decisions in the social utility on the right. We denote the model as the binary choice with misclassification and social interactions.

The binary choice with misclassification and social interactions is modeled through a simultaneous game played on an exogenously-given large social network. Exogenous network setting prevails in peer effects studies, either in the linear-in-mean model.[2] or in the discrete choice model with social interactions[3] There is a growing literature on the econometrics of dynamic network formation; e.g., Christakis et al. (2010), Graham (2015), Graham (2016), Graham (2017), Leung (2015), Menzel (2017), Chandrasekhar and Jackson (2016), Mele (2017), de Paula et al. (2018), Sheng (2020), Badev (2021). We focus on the simultaneous game played on an exogenous network and do not study the network formation issue. For more discussion of games played on networks, see Bramoullé and Kranton (2016).

We obtain the identification of the true model, which is the conditional distribution of the latent true decision variable, through the technique of two repeated measurements and a monotonicity condition. We extend the likelihood-like algorithm [the nested pseudo likelihood (NPL) estimation] from Aguirregabiria and Mira (2007) (dynamic game) and Lin and Xu (2017) (social interactions) and apply it to our model with a homogeneous misclassification condition. We establish the asymptotic properties of the NPL estimator and illustrate its finite sample performance with Monte Carlo studies and an application to peer effects in attitude.

The paper unfolds as follows. Section 2 introduces the binary choice model with misclassification and social interactions. Section 3 provides theoretical results on the identification of the conditional distribution of the latent variable and the structural parameter. We then demonstrate the nested pseudo likelihood (NPL) estimation strategy in Section 4. Two Monte Carlo experiments are conducted in Section 5 to illustrate the finite sample performance of the model and the NPL algorithm. Section 6 presents results on the peer effects among students in their attitudes towards learning. The last section concludes. Proofs are rendered in Section Appendix A.

2. Binary choice with misclassification and social interactions

There are n individuals, $\mathscr{I} = \{1, \ldots, n\}$, located (socially) in a single exogenously-given large social network. Each individual i is associated with a group of friends, F_i. Let $F_{ij} = 1$ denote that individual i considers j as a best friend and friendships are taken exogenously. The friendship is not necessarily reciprocal; i.e., $F_{ij} \neq F_{ji}$ is allowed. We denote $F_{ii} = 0$ by convention. Therefore, the friends set is $F_i = \{j \in \mathscr{I} : F_{ij} = 1\}$. Denote N_i as the number of friends of individual i, i.e., $N_i = \#F_i$.

Individuals make binary choices $\{Y_i^* \in \{0, 1\}\}_{i \in \mathscr{I}}$ simultaneously. The interactions transit through the directed link, F_{ij}, which

[1] We use "social interactions" and "peer effects" interchangeably.

[2] See Manski (1993), Manski (2000), Lee (2007), Graham (2008), Bramoullé et al. (2009), Calvó-Armengol et al. (2009), Lee et al. (2010), Lin (2010), Liu and Lee (2010), Goldsmith-Pinkham and Imbens (2013), Bramoullé et al. (2014), Dahl et al. (2014), Blume et al. (2015), Eraslan and Tang (2017), Hoshino (2019), Johnsson and Moon (2021), Lin and Tang (2021) to name only a few,

[3] See Brock and Durlauf (2001a, 2007), Card and Giuliano (2013), Lee et al. (2014), Song (2014), Menzel (2016), Li and Zhao (2016), Canen et al. (2017), Lin and Xu (2017), Yang and Lee (2017), Xu (2018), Liu (2019), to mention but a few.

means that individuals take into account the choices of their friends when they make decisions. Though individuals, rather than friends, do not directly deliver peer effects, the transitions through friendships render indirect effects over the network. For example, in the peer effects in attitude study, $Y_i^* = 1$ means that student i has a positive attitude towards learning. Here we use Y_i^* to denote the true latent choice of individual i. We will use Y_{1i}, and Y_{2i} for the reported measurements of the latent variable. Following the standard binary choice literature (McFadden, 1974; Train, 2009), we normalize the utility of choosing $Y_i^* = 0$ as 0. We specify the latent utility of $Y_i^* = 1$ as

$$U_i\left(Y_{-i}^*, X_i, F_i, \varepsilon_i\right) = X_i'\beta + \frac{\alpha}{N_i}\sum_{j\in F_i} Y_j^* - \varepsilon_i, \tag{2.1}$$

where $X_i \in \mathscr{X}$ is a $d \times 1$ vector representing the demographic characteristics[4], Y_{-i}^* are the choices of others, and ε_i is the private utility shock. The utility of individual i has three components: the deterministic part from demographics, $X_i'\beta$; the deterministic social utility from the average choice of friends (peer effects), $\frac{\alpha}{N_i}\sum_{j\in F_i} Y_j^*$; and a private utility shock, ε_i. α captures the peer effects from friends. Denote $\mu = (\beta', \alpha)'$.

To complete the setting for the model, we further specify the information structure. Let $\mathbb{I}_n \equiv (\{X_i\}_{i\in\mathscr{I}}, \{F_i\}_{i\in\mathscr{I}})$ be the public information set including all demographic characteristics and friendship information.[5] The private utility shock ε_i is only known to individual i. Therefore, we consider an incomplete information structure in the Bayesian game and individuals form beliefs on the choices of their friends.[6] The decision rule is:

$$Y_i^* = \mathbf{1}\left\{X_i'\beta + \frac{\alpha}{N_i}\sum_{j\in F_i}\mathbb{E}\left(Y_j^*\middle|\mathbb{I}_n, \varepsilon_i\right) - \varepsilon_i \geq 0\right\}, \tag{2.2}$$

where the incomplete information structure is presented by the conditional expectation (belief). Individuals make choices based on the belief of their peers' choices and not on their friends' actual choices. A similar setting can be found in Brock and Durlauf (2001a), Brock and Durlauf (2001b), Ioannides (2006), Durlauf and Ioannides (2010), Lin and Xu (2017), Xu (2018), Jackson et al. (2020) and Lin (2021).

2.1. Bayesian Nash equilibrium

With an incomplete information structure, we consider the Bayesian Nash equilibrium (BNE) of the Bayesian game. To characterize the equilibrium, we make the following assumptions on the random utility terms.

Assumption 1. The private random utility terms ε_i's are i.i.d. across individuals and conform to the standard Logistic distribution. ε_i is independent from $\mathbb{I}_n$.

Remark 1. Assumption 1 is fairly standard in the discrete game model literature (Bajari et al., 2010). As a matter of fact, Assumption 1 provides a closed-form expression for individuals' conditional choice probabilities in terms of friends' choice probabilities and streamlines the belief term; i.e., $\mathbb{E}(Y_j^*\big|\mathbb{I}_n,\varepsilon_i) = \mathbb{E}(Y_j^*\big|\mathbb{I}_n)$. This is because that given $\mathbb{I}_n$, the left random term in Y_j^*, ε_j is independent from ε_i.

Denote $\Lambda(t) = \frac{e^t}{1+e^t}$. We define $\mathbb{P}(Y_i^* = 1|\mu; \mathbb{I}_n)$ as the equilibrium choice probability of individual i. With $\mathbb{E}(Y_i^*|\mu; \mathbb{I}_n) = \mathbb{P}(Y_i^* = 1|\mu; \mathbb{I}_n)$, we have

$$\mathbb{P}\left(Y_i^* = 1\middle|\mu; \mathbb{I}_n\right) = \Lambda\left[X_i'\beta + \frac{\alpha}{N_i}\sum_{j\in F_i}\mathbb{P}\left(Y_j^* = 1\middle|\mu; \mathbb{I}_n\right)\right], i \in \mathscr{I}, \tag{2.3}$$

where $P_{[n]}^* \equiv (P_1^*, \dots, P_n^*)' \equiv [P(Y_1^* = 1|\mu; \mathbb{I}_n), \dots, P(Y_n^* = 1|\mu; \mathbb{I}_n)]'$ is the equilibrium choice probabilities profile. Eq. (2.3) is a simultaneous system of equations of $P_{[n]}^*$. Let $P_{[n]} \equiv (P_1, \dots, P_n)'$ be an arbitrary choice probabilities profile. Define

[4] X_i contains an intercept. We consider a single large network; therefore, the characteristics of the network itself are constant for all individuals and are absorbed in the intercept term. Our model can also be brought to multiple networks including network characteristics as there is variation across networks. The identification strategy is similar to the one using between-group variation in linear-in-mean models, as seen in Graham (2008).

[5] The usage of all demographics and friendships as public information is for the tractability of the equilibrium, as we will see below.

[6] The importance of an incomplete information structure is well documented in the discrete game literature; see Brock and Durlauf (2001a,b), Bajari et al. (2010); Lin and Xu (2017) and Xu (2018) for social interactions/peer effects studies; Seim (2006) and Sweeting (2009) for competition in industrial organization; Aradillas-Lopez (2010, 2012), Tang (2010), de Paula and Tang (2012) and Xu (2014) for estimation and inference of the static games; and Aguirregabiria and Mira (2002, 2007), Pesendorfer and Schmidt-Dengler (2008) and Arcidiacono et al. (2016)for dynamic games. We would like to refer interested readers to the global game literature with an incomplete information structure; e.g., Morris and Shin (2003).

$$\Gamma_i\left(\mu; P_{[n]}, \mathbb{I}_n\right) \equiv \Lambda\left(X_i'\beta + \frac{\alpha}{N_i}\sum_{j\in F_i} P_j\right) \tag{2.4}$$

The equilibrium choice probability profile $P^*_{[n]}$ defined in Eq. (2.3) is then a fixed point of

$$\Gamma\left(\mu; P_{[n]}, \mathbb{I}_n\right) \equiv \left(\Gamma_1\left(\mu; P_{[n]}, \mathbb{I}_n\right), \ldots, \Gamma_n\left(\mu; P_{[n]}, \mathbb{I}_n\right)\right)' = P_{[n]}.$$

We make the following assumptions to achieve the uniqueness of the BNE for the identification of conditional choice probabilities from the data.

Assumption 2. There is an upper bound, $M > 0$, for the number of friends; i.e., $N_i \le M$ for $i \in \mathscr{I}$ for all $n \ge 1$.

Remark 2. Assumption 2 excludes some specific networks, such as the star network. In social networks, it is feasible to limit the number of friends as human beings do not have infinite efforts to maintain too many friendships. Furthermore, we here use the directed network, and only out-degree delivers peer effects. One individual is reasonably influenced by a limited number of influencers. Assumption 2 leads to a sparse network in our model as n increases. With a sparse network, we could identify the conditional distribution of (Y_{1i}, Y_{2i}) and then of Y^*_i. To identify the joint conditional distribution of (Y_{1i}, Y_{2i}), we look for possible many j's (increasing but slower than n) whose characteristics and network positions are the same as i's. Taking the average of all these (Y_{1j}, Y_{2j}) gives the empirical conditional distribution. The sparsity enables us to draw n subnetworks, which have limited overlap, i.e., when the distance of two subnetworks is large enough, they are nearly independent. This feature establishes the network decaying dependence condition of the model, which provides support for feasible inference. We defer the illustration of the network decaying dependence condition after the result of the unique BNE.

Assumption 3. The strength of interactions is moderate; i.e., $0 \le \alpha < 4$.

Remark 3. In the literature concerning interaction games, a similar assumption is denoted as the Moderate Social Influence (MSI) condition for uniqueness; see Glaeser and Scheinkman (2003) and Horst and Scheinkman (2006, 2009). In the literature of discrete choice with social interactions, Brock and Durlauf (2001a,b), Lin and Xu (2017), Xu (2018), Liu (2019) and Jackson et al. (2020) employ a similar condition to characterize the uniqueness of the BNE in the Bayesian game. Assumption 3 restricts the size of interactions along individuals' choices. This size restriction is similar to the stationarity condition in the autoregressive model; e.g., in an AR(1) model, the dependence parameter is within $(-1, 1)$. The time series analysis is one dimensional and our social interactions analysis is multiple-dimensional such that each friend of an individual provides one dimension. Similar to the existence of explosive time series, there are exceptions with dominant peer effects. Examples include tipping (Gladwell, 2000; Granovetter, 1978; Schelling, 1971) or rush into the market (Anderson et al., 2017; Park and Smith, 2008). In the peer effects in attitude study in Section 6, this condition is feasible so that peer effects would not be dominant in a school environment. There is also literature to work with multiple equilibria by partial identification technique; e.g., Li and Zhao (2016) construct moments inequalities based on subnetworks for partial identification analysis. For more discussion on multiple equilibria and partial identification, see Tamer (2003), Ciliberto and Tamer (2009), Tamer (2010) and de Paula (2013). The upper bound, 4, comes from the Logistic distribution of the private utility terms. For a standard normal distribution in Probit-type models, we should change the upper bound to $\sqrt{2\pi}$. In general, $0 \le \alpha < 1/\sup f_\varepsilon(\cdot)$ is required to establish uniqueness; see Horst and Scheinkman (2006) for more details.

Lemma 1. *With Assumptions 1 to 3, there exists a unique pure strategy Bayesian Nash equilibrium for the Bayesian game, represented in Equation (2.3).*

Proof. See Appendix A.□

Assumptions 3 is the key to establish the uniqueness of the BNE. The MSI condition restricts the interactions between an individual's conditional choice probability and friends' conditional choice probabilities. Combined with the parametric logistic assumption over the random utility shock $[\Lambda(\cdot)\cdot[1-\Lambda(\cdot)] \le 1/4]$, the MSI condition guarantees that the best response function is contractive–see details in Eq. (A.1). Lemma 1 establishes the uniqueness of the Bayesian Nash equilibrium. The uniqueness ensures that the conditional distributions of repeated measurements are identified from the data. Another option for the equilibrium characterization is to assume that the data comes from one single equilibrium; see Bajari et al. (2010). The uniqueness based on Assumption 3 has the advantage that we can impose the restriction in our estimation strategy to ensure that the data is from the unique equilibrium.

2.2. Misclassification

Our Bayesian game builds on the binary latent choices $\{Y^*_i\}_{i\in\mathscr{I}}$, which are prone to misclassification errors. de Paula (2017) points out the importance of the measurement error issue in network studies. It is well accepted that misclassification induces problems of analysis and interpretation. In the binary choice with misclassification and social interactions, the simultaneity of social interactions raises misclassification errors on the left and on the right. There are several ways to deal with the misclassification problem: repeated measurements, validation data, instrumental variables, etc. Mahajan (2006) resorts to using an instrumental variable for identification of a nonparametric model with the presence of misclassified regressor. Hu (2008) provides a general framework for the identification and estimation of the misclassification problem with repeated measurements. For the misclassification on the dependent variable, Lewbel (2000) establishes the identification of the model with misclassification on the left using an instrument variable (exogenous shifter). Hausman et al. (1998) propose a partial maximum likelihood estimator to handle a misclassified response variable. In this paper, we resort to repeated measurements and a monotonicity condition for identification and estimation.

3. Closed-form identification

In this paper, we adopt a technique of two repeated measurements and a monotonicity condition to identify the true conditional distribution of the latent choice variable and the structural parameter. In general, we could derive identification and estimation results with three repeated measurements. However, it is difficult to obtain three clean repeated measurements of the same latent status. To fit the empirical study, we replace the third measurement with a monotonicity condition.

3.1. Two measurements and a monotonicity condition

In this section, we provide closed-form identification of the conditional choice probability of the latent variable using two measurements, (Y_{1i}, Y_{2i}), and a monotonicity condition. In some scenarios, theory or information provides monotonicity that individuals do not overreport or underreport the latent status. Without loss of generality, we take the monotonicity that individuals do not underreport the latent status, i.e., $\mathbb{P}(Y_{1i}=0 \big| Y_i^*=1, \mu, \mathbb{I}_n) = \mathbb{P}(Y_{2i}=0 \big| Y_i^*=1, \mu, \mathbb{I}_n) = 0$. We further impose the conditional independence assumption.

Assumption 4. (i) (Y_{1i}, Y_{2i}) are jointly independent conditional on Y_i^* and $\mathbb{I}_n$,

$$Y_{1i} \perp Y_{2i} \mid \left(Y_i^*, \mathbb{I}_n\right). \tag{3.1}$$

(ii) (Y_{1i}, Y_{2i}) are cross-sectionally independent given $\mathbb{I}_n$.

Remark 4. Assumptions 4 is standard in the nonlinear measurement error literature (Hu, 2008; Hu, 2017; Hu and Schennach, 2008; Li, 2002; Li and Hsiao, 2004; Li and Vuong, 1998; Mahajan, 2006; Schennach, 2016; and references therein). Assumption 4 means that the repeated measurements provide no extra useful information other than those embedded in the true latent choices and characteristics. After controlling the true latent variables and public information, the data collection processes for the two repeated measurements are independent.

A second condition is technical: the conditional probability of (Y_{1i}, Y_{2i}) at (0,0), $\mathbb{P}(Y_{1i}=0, Y_{2i}=0 | \mu, \mathbb{I}_n)$, is positive. This term is in the denominators of our closed-form formulas for identification.

Assumption 5. $\mathbb{P}(Y_{1i}=0, Y_{2i}=0 | \mu, \mathbb{I}_n) > 0$ for all realizations of $\mathbb{I}_n$.

Further, a *monotonicity condition* precludes *underreporting*, i.e., the measurements cannot take the value of 0 if the latent binary regressand is 1.

Assumption 6. $\mathbb{P}(Y_{1i}=0 \big| Y_i^*=1, \mu, \mathbb{I}_n) = \mathbb{P}(Y_{2i}=0 \big| Y_i^*=1, \mu, \mathbb{I}_n) = 0$ for $i \in \mathscr{I}$ and any realization of $\mathbb{I}_n$.

Note that if there is no overreporting in an application, the monotonicity condition holds after we switch what 0 and 1 each represent for the latent binary regressands and measurements.

Proposition 1. *Given Assumptions 4–6, there is a closed-form identification of $\mathbb{P}(Y_i^*=0 \mid \mu, \mathbb{I}_n)$ and $\mathbb{P}(Y_i^*=1 \mid \mu, \mathbb{I}_n)[=1-\mathbb{P}(Y_i^*=0 \mid \mu, \mathbb{I}_n)]$:*

$$\mathbb{P}\left(Y_i^*=0 \big| \mu, \mathbb{I}_n\right) = \frac{\mathbb{P}_{Y_{2i}}(0) \cdot \mathbb{P}_{Y_{1i}}(0)}{\mathbb{P}_{Y_{1i}, Y_{2i}}(0,0)},$$

where $\mathbb{P}_{Y_{1i}, Y_{2i}}(j,k)$ is the jointly conditional choice probability of (Y_{1i}, Y_{2i}) at (j,k), $j,k \in \{0,1\}$. All proofs are relegated to Appendix A.

We then take $\mathbb{P}(Y_i^*=1 | \mu, \mathbb{I}_n)$ as known for the next step identification of the structural parameter.

3.2. Identification of the structural parameter, μ

For the equilibrium presented in Eq. (2.3), the identification of μ is standard in a constructive way. As shown in the first step identification, $P(Y_i^*=1 | \mu, \mathbb{I}_n), i \in \mathscr{I}$ is identified from the observables. From Eq. (2.3), we have

$$\Xi(\mathbb{I}_n) \equiv \log\left[\mathbb{P}\left(Y_i^*=1 \big| \mu, \mathbb{I}_n\right)\right] - \log\left[\mathbb{P}\left(Y_i^*=0 \big| \mu; \mathbb{I}_n\right)\right] = X_i'\beta + \frac{\alpha}{N_i}\sum_{j \in F_i}\mathbb{P}\left(Y_j^*=1 \Big| \mathbb{I}_n\right), i \in \mathscr{I} \tag{3.2}$$

We make the following rank condition assumption to achieve identification.

Assumption 7. $\mathbb{E}\left[\left(X_i', \frac{\sum_{j \in F_i}\mathbb{P}(Y_j^*=1|\mathbb{I}_n)}{N_i}\right)' \times \left(X_i', \frac{\sum_{j \in F_i}\mathbb{P}(Y_j^*=1|\mathbb{I}_n)}{N_i}\right)\right]$ is with full rank $d+1$ for all n sufficiently large.

Remark 5. Assumption 7 requires no perfect collinearity of $\left(X_i', \frac{\sum_{j \in F_i}\mathbb{P}(Y_j^*=1|\mathbb{I}_n)}{NF_i}\right)$. This assumption is essentially a full rank condition. As is pointed out in Bajari et al. (2010), it is other individuals' exclusive payoff shifters that induce independent variations in individual i's beliefs, which render the rank condition meaningful. The variation in the friends sets makes the peer effects term with enough variation and mitigates or solves the perfect collinearity issue. The BNE profile is determined through the fixed point and therefore implicitly by the $\mathbb{I}_n$ and the distribution of the ε. Furthermore, there is a variation in friends sets, which prevents the perfect collinearity/multiplicity problem between the peer effect covariate, $\frac{\sum_{j\in F_i}\mathbb{P}(Y_j^*=1|\mathbb{I}_n)}{N_i}$, and the demographic characteristics, X_i. It is worth pointing out that the expectation operator is for X_i, Y_i. The network information, $F_i, i \in \mathscr{I}$ is taken exogenously and thus does not enter the expectation operator.

With Assumption 7, we have identified μ as

$$\mu = \mathbb{E}\left[\left(X_i^{'}, \frac{\sum_{j\in F_i}\mathbb{P}\left(Y_j^*=1\middle|\mathbb{I}_n\right)}{N_i}\right)^{'} \times \left(X_i^{'}, \frac{\sum_{j\in F_i}\mathbb{P}\left(Y_j^*=1\middle|\mathbb{I}_n\right)}{N_i}\right)\right]^{-1} \times \mathbb{E}\left[\left(X_i^{'}, \frac{\sum_{j\in F_i}P\left(Y_j^*=1\middle|\mathbb{I}_n\right)}{N_i}\right)^{'} \times \Xi(\mathbb{I}_n)\right] \tag{3.3}$$

4. Estimation strategy

The identification in Section 3 is for the population and it takes $P(Y_{1i}, Y_{2i}|\mathbb{I}_n)$ as identified from the observables. However, the nonparametric estimation of the joint conditional distribution is infeasible due to the large dimension of $\mathbb{I}_n$. To avoid such a problem, we adopt a sequential algorithm, the Nested Pseudo Likelihood (NPL) method, to estimate the structural parameter. This method is first introduced by Aguirregabiria and Mira (2002, 2007) for dynamic discrete choice models and dynamic games. Lin and Xu (2017) extend the method to social interactions studies. We suppress the parameter μ in the conditional choice probabilities when there is no ambiguity. Before we proceed to the details of the NPL estimator, we make the following simplifying assumption:

Assumption 8. The misclassification probabilities satisfy

$$\begin{aligned} P\left(Y_{1i}=1\middle|Y_i^*=0,\mathbb{I}_n\right) &= P\left(Y_{1i}=1\middle|Y_i^*=0\right)=\gamma\in[0,1], \\ P\left(Y_{2i}=1\middle|Y_i^*=0,\mathbb{I}_n\right) &= P\left(Y_{2i}=1\middle|Y_i^*=0\right)=\delta\in[0,1]. \end{aligned}$$

Remark 6. Assumptions 8 reduces the number of unknowns in the misclassification probabilities. This assumption is introduced to make the empirical analysis feasible given the sample size and the complexity of social network analysis. We can relax this assumption by parameterization over some observed covariates with richer data. Similar constant misclassification probabilities assumption can be found in Copas (1988), Hausman et al. (1998), Neuhaus (2002), Ramalho (2002). See Carroll et al. (2006) for more details.

Define $P_i^* \equiv \mathbb{P}(Y_i^*=1|\mathbb{I}_n)$ and $P_{[n]}^* \equiv (P_1^*, \ldots, P_n^*)^{'}$. We now have the structural parameter, $\theta \equiv (\gamma, \delta, \mu^{'})^{'}$, and

$$\begin{aligned} P(Y_{1i}=1|\mathbb{I}_n) &= \mathbb{P}\left(Y_{1i}=1\middle|Y_i^*=1,\mathbb{I}_n\right)\mathbb{P}\left(Y_i^*=1\middle|\mathbb{I}_n\right)+\mathbb{P}\left(Y_{1i}=1\middle|Y_i^*=0,\mathbb{I}_n\right)\mathbb{P}\left(Y_i^*=0\middle|\mathbb{I}_n\right) \\ &= 1\cdot\mathbb{P}\left(Y_i^*=1\middle|\mathbb{I}_n\right)+\gamma\left[1-\mathbb{P}\left(Y_i^*=1\middle|\mathbb{I}_n\right)\right] \\ &= \gamma+(1-\gamma)\mathbb{P}\left(Y_i^*=1\middle|\mathbb{I}_n\right)=\gamma+(1-\gamma)P_i^* \\ P(Y_{2i}=1|\mathbb{I}_n) &= \mathbb{P}\left(Y_{2i}=1\middle|Y_i^*=1,\mathbb{I}_n\right)\mathbb{P}\left(Y_i^*=1\middle|\mathbb{I}_n\right)+\mathbb{P}\left(Y_{2i}=1\middle|Y_i^*=0,\mathbb{I}_n\right)\mathbb{P}\left(Y_i^*=0\middle|\mathbb{I}_n\right) \\ &= 1\cdot\mathbb{P}\left(Y_i^*=1\middle|\mathbb{I}_n\right)+\delta\left[1-\mathbb{P}\left(Y_i^*=1\middle|\mathbb{I}_n\right)\right] \\ &= \delta+(1-\delta)\mathbb{P}\left(Y_i^*=1\middle|\mathbb{I}_n\right)=\delta+(1-\delta)P_i^* \end{aligned} \tag{4.1}$$

Our log likelihood function is formulated by the observed conditional distribution function $f(Y_{1i}, Y_{2i}|\mathbb{I}_n;\theta)$. With Eq. (4.1), we have the log-likelihood function:[7]

$$\begin{aligned} \mathscr{L}\left(\theta, P_{[n]}^*\right) \equiv \frac{1}{n}\sum_{i\in\mathscr{I}}\Bigg\{ & Y_{1i}\log\left[\gamma+(1-\gamma)\Gamma_i\left(\mu;P_{[n]}^*,\mathbb{I}_n\right)\right]+(1-Y_{1i})\log\left[1-\gamma-(1-\gamma)\Gamma_i\left(\mu;P_{[n]}^*,\mathbb{I}_n\right)\right] \\ & +Y_{2i}\log\left[\delta+(1-\delta)\Gamma_i\left(\mu;P_{[n]}^*,\mathbb{I}_n\right)\right]+(1-Y_{2i})\log\left[1-\delta-(1-\delta)\Gamma_i\left(\mu;P_{[n]}^*,\mathbb{I}_n\right)\right]\Bigg\} \end{aligned} \tag{4.2}$$

We first introduce the MLE to motivate the NPL estimation method.

$$\widehat{\theta}_{MLE} = \underset{\theta\in\Theta}{\operatorname{argmax}}\ \mathscr{L}\left(\theta, P_{[n]}\right) \ \text{ s.t. } \ P_{[n]} = \Gamma\left(\mu; P_{[n]}, \mathbb{I}_n\right). \tag{4.3}$$

For a small number of players, we can implement the MLE method by the nested fixed point (NFP) algorithm (Rust, 1987), which repeatedly solves all of the fixed points of $P = \Gamma(\mu; P, \mathbb{I}_n)$ for each candidate parameter value. As n becomes large, the NFP algorithm for the MLE is computationally intensive to solve the $n-$ dimensional fixed points for each candidate value of θ and obtain the optimal $\widehat{\theta}$

[7] Here we construct a complete likelihood function based on two measurements: Y_1, and Y_2. Hausman et al. (1998) use either Y_1, or Y_2 to construct partial likelihood function.

with a maximized log-likelihood function. To address the computational burden, we adopt the Nested Pseudo Likelihood estimation method, which swaps the order of the NFP algorithm.

The equilibrium choice probabilities profile is solved through iterated steps. Therefore, in this section, we suppress the public information $\mathbb{I}_n$; i.e., $P_{[n]} = \Gamma(\theta; P_{[n]})$. We make similar assumptions as in Aguirregabiria and Mira (2007), Kasahara and Shimotsu (2012) and Lin and Xu (2017). Define the pseudo log-likelihood function as

$$\begin{aligned}\mathscr{L}\left(\theta, P_{[n]}\right) &= \frac{1}{n}\sum_{i\in\mathscr{I}}\mathscr{L}_i\left(\theta, P_{[n]}\right)\\ &\equiv \frac{1}{n}\sum_{i\in\mathscr{I}} Y_{1i}\log\left[\gamma + (1-\gamma)\Gamma_i\left(\mu; P_{[n]}, \mathbb{I}_n\right)\right] + (1-Y_{1i})\log\left[1-\gamma-(1-\gamma)\Gamma_i\left(\mu; P_{[n]}, \mathbb{I}_n\right)\right]\\ &\quad + Y_{2i}\log\left[\delta + (1-\delta)\Gamma_i\left(\mu; P_{[n]}, \mathbb{I}_n\right)\right] + (1-Y_{2i})\log\left[1-\delta-(1-\delta)\Gamma_i\left(\mu; P_{[n]}, \mathbb{I}_n\right)\right],\end{aligned} \tag{4.4}$$

where $P_{[n]} = (P_1, \ldots, P_n)$ is not necessarily the true equilibrium choice probabilities profile. We illustrate the NPL algorithm below

Initiation: Obtain an initial guess of choice probabilities profile, e.g., Logit regression without social interactions, denoted by $\widehat{P}_{[n]}^{(0)}$.

Iteration: Given $\widehat{P}_{[n]}^{(K)}$, we obtain the $K+1$th estimate, $\widehat{\theta}^{(K+1)} = \underset{\theta\in\Theta}{\operatorname{argmax}}\ \mathscr{L}(\theta, \widehat{P}_{[n]}^{(K)})$ and update the choice probabilities profile by $\widehat{P}_{[n]}^{(K+1)} = \Gamma(\widehat{\mu}^{(K+1)}; \widehat{P}_{[n]}^{(K)}, \mathbb{I}_n)$.

Termination: Iterate until the difference between two consecutive $P_{[n]}$ estimates is sufficiently small, say, when $\| \widehat{P}_{[n]}^{(K+1)} - \widehat{P}_{[n]}^{(K)} \| < tol$ with some preset tolerance. Set the NPL estimates as $\widehat{\theta} = \widehat{\theta}^{(K+1)}$ and $\widehat{P}_{[n]} = \widehat{P}_{[n]}^{(K+1)}$.

It is computationally feasible that we do not actually calculate the BNE choice probabilities profile but instead adopt a recursive method starting from reasonable guesses of probability values. The NPL method is a sequential version of the extremum estimation. It augments the maximum likelihood estimation with an equilibrium condition and mitigates the dimensionality issue when calculating the conditional choice probabilities. The computational cost of NPL is moderate with the contraction mapping property of the Bayesian game derived from the bounded degree and the moderate social influence conditions.

4.1. Consistency and asymptotic normality of the NPL estimator

Let Θ and $\mathscr{P}_n$ be the support sets of θ and $P_{[n]}$, respectively. Let

$$\begin{aligned}\mathscr{L}_0\left(\theta, P_{[n]}\right) &\equiv \mathbb{E}\left[\mathscr{L}_i\left(\theta, P_{[n]}\right)\right],\\ \widetilde{\theta}_n\left(P_{[n]}\right) \equiv arg\max_{\theta\in\Theta}\ \mathscr{L}_0\left(\theta, P_{[n]}\right);&\ \ \phi_0\left(P_{[n]}\right) \equiv \Gamma\left(\widetilde{\theta}_0\left(P_{[n]}\right), P_{[n]}\right);\\ \widetilde{\theta}_0\left(P_{[n]}\right) \equiv arg\max_{\theta\in\Theta}\ \mathscr{L}\left(\theta, P_{[n]}\right);&\ \ \phi_n\left(P_{[n]}\right) \equiv \Gamma\left(\widetilde{\theta}_n\left(P_{[n]}\right), P_{[n]}\right);\end{aligned}$$

The conditional choice probabilities profile in the Bayesian game with n players is a function of the true parameter θ_0 and the public information set $\mathbb{I}_n$. Our asymptotic analysis draws on the fact that the number of players in the large network goes to infinity. Thus, we are considering a growing single network with some stable growing mechanisms. Define a sequence of NPL fixed points sets as $\Lambda_{0n} \equiv \{(\theta, P_{[n]}) \in (\Theta, \mathscr{P}_n) : \theta = \widetilde{\theta}_0(P_{[n]}), P_{[n]} = \phi_{0n}(P_{[n]})\}$ and the NPL fixed points set of sample size n as $\Lambda_n \equiv \{(\theta, P_{[n]}) \in (\Theta, \mathscr{P}_{[n]}) : \theta = \widetilde{\theta}_n(P_{[n]}), P_{[n]} = \phi_n(P_{[n]})\}$. Let θ_{NPL} be the maximizer of $\mathscr{L}_0(\theta_0, P_{[n]}^*)$. Let $\mathscr{N}$ denote a closed neighborhood of $(\theta_0, P_{[n]}^*)$. The first order condition for the NPL estimation is

$$\frac{\partial\mathscr{L}\left(\theta, P_{[n]}\right)}{\partial\theta}\Big|_{\left(\theta, P_{[n]}\right)=(\hat{\theta}_{NPL}, \hat{P}_{NPL})} = 0. \tag{4.5}$$

This first order condition is similar to that for dynamic game (Aguirregabiria and Mira, 2007).

Assumption 9. (i) Θ is compact, θ_0 is an interior point of Θ, and $\mathscr{P}_n$ is a compact and convex subset of $(0,1)^n$; (ii) $(\theta_0, P_{[n]}^*)$ is an isolated population NPL fixed point; i.e., there is an open ball around it that does not contain any other element of Λ_{0n}; (iii) the operator $\phi_0(P) - P$ has a nonsingular Jacobian matrix at $P_{[n]}^*$; (iv) there exist non-singular matrices $V_1(\theta_0)$ and $V_2(\theta_0)$ such that

$$\mathbb{E}\left[\frac{\partial^2\mathscr{L}\left(\theta_0, P_{[n]}^*\right)}{\partial\theta\partial\theta'} + \frac{\partial^2\mathscr{L}\left(\theta_0, P_{[n]}^*\right)}{\partial\theta\partial P'}\cdot\left[I - \left(\frac{\partial\Gamma\left(P_{[n]}^*; \theta_0\right)}{\partial P}\right)'\right]^{-1}\cdot\frac{\partial\Gamma\left(P_{[n]}^*; \theta_0\right)}{\partial\theta'}\Bigg|\mathbb{I}_n\right] \xrightarrow{p} V_1(\theta_0),$$

$$\mathbb{E}\left[\frac{\partial\mathscr{L}_i\left(\theta_0, P_{[n]}^*\right)}{\partial\theta}\frac{\partial\mathscr{L}_i\left(\theta_0, P_{[n]}^*\right)}{\partial\theta'}\Bigg|\mathbb{I}_n\right] \xrightarrow{p} V_2(\theta_0)$$

Moreover, $V_1(\theta_0)$ is negative definite.

Remark 7. Assumption 9(i) is standard for asymptotic analysis in the literature (see Newey and McFadden, 1994). Assumption 9(ii) is an identification assumption to use the NPL algorithm. It is straightforward to show that $\theta_0 = \widetilde{\theta}_0(P^*_{[n]})$ solves $\underset{\theta\in\Theta}{\operatorname{argmax}}\ \mathscr{L}_0(\theta, P^*_{[n]})$. Without Assumption 9(ii), $\underset{\theta\in\Theta}{\operatorname{argmax}}\ \mathscr{L}_0(\theta, P^*_{[n]})$ might admit multiple solutions and each of these represents a fixed point of the NPL algorithm. It is worth noting here $(\theta, P_{[n]})$ is the fixed point of the NPL algorithm. It differs from the fixed point solution of the best response functions, i.e., $P^*_{[n]}$ is the fixed point solution of $P_{[n]} = \Gamma(\mu; P_{[n]}, \mathbb{I}_n)$ and it varies on different realizations of the conditional set $\mathbb{I}_n$. We are not able to find more primitive condition to assure this point. Assumption 9(iii) assures that the matrix $I - \left(\frac{\partial\Gamma(P^*_{[n]};\theta_0)}{\partial P}\right)'$ is invertible and thus there exists a sample nested pseudo likelihood fixed point (estimate) close to $(\theta_0, P^*_{[n]})$. Assumption 9 (iv) is a high-level condition for non-singular limiting matrices as n goes to infinity. This condition requires that adding more individuals (growing network) would not change the interactions pattern in the network. The sparsity of the network originated from Assumption 2 implies that the CCPs $P^*_{[n]}$ enter the social interactions term of individual i through $\frac{1}{N_i}\sum_{j\in F_i} P^*_j$ which has at most M elements of $P^*_{[n]}$. The social interactions term, $\frac{1}{N_i}\sum_{j\in F_i} P^*_j$ is bounded in [0,1] and behaves as a standard term, though solved from the fixed point of the Bayesian game. Moreover, the non-degeneracy of $V_1(\theta)$ and $V_2(\theta)$ requires that all of the determinants of the finite counterparts are outside an open ball of zero for all n, which is a rank condition. It is worth pointing out here $\frac{\partial^2\mathscr{L}(\theta_0,P^*)}{\partial\theta\partial P'} \cdot \left[I - \left(\frac{\partial\Gamma(P^*;\theta_0)}{\partial P}\right)'\right]^{-1} \cdot \frac{\partial\Gamma(P^*;\theta_0)}{\partial\theta'}$ is a $(d_x+3)\times(d_x+3)$ matrix.

Assumption 10. (i) The family $\{\mathscr{L}_i(\theta, P_{[n]}) : \theta\in\Theta\}$ is a Vapnik-Cernonenkis class of functions; (ii)the class $\{\mathscr{L}_i(theta, P_{[n]}, y)\}$ is Donsker with respect to the distribution of X_i for $y = 1$ or 0 with a square-integrable function; (iii) $E[|\mathscr{L}_i(\theta, P_{[n]}, y)|\mathbb{I}]$ and $|E[\mathscr{L}_i(\theta, P_{[n]}, 1) - \mathscr{L}_i(\theta, P_{[n]}, 0)|\mathbb{I}]|$ are bounded by a constant.

Theorem 1. *Suppose Assumptions 1–10 hold, $\widehat{\theta}_{NPL}$ is consistent and*

$$\sqrt{n}(\widehat{\theta}_{NPL} - \theta_0) \xrightarrow{d} \mathscr{N}(0, V_{NPL}), \tag{4.6}$$

where $V_{NPL} = V_1^{-1}(\theta_0)V_2(\theta_0)V_1^{'-1}(\theta_0)$.

Proof. See Appendix A□

The NPL estimator has the same convergent rate as the MLE (implemented by the NFXP algorithm). It has a different asymptotic variance than the MLE that an additional term $\mathbb{E}\left[\frac{\partial^2\mathscr{L}(\theta_0,P^*_{[n]})}{\partial\theta\partial P'} \cdot \left[I - \left(\frac{\partial\Gamma(P^*_{[n]};\theta_0)}{\partial P}\right)'\right]^{-1} \cdot \frac{\partial\Gamma(P^*_{[n]};\theta_0)}{\partial\theta'} | \mathbb{I}_n\right]$ coming from the equilibrium condition shows in V_1. A similar result can be seen in the seminal work of Menzel (2016) for games with many players where he shows that the limiting distribution of players' choices and characteristics is equivalent to a single-agent discrete choice problem that is augmented by an aggregate equilibrium condition. Menzel (2016) establishes a conditional central limit theorem for a moment-based estimator in games with many players where both the number of games and number of players in each game go to infinity.

The local convergence of the NPL algorithm is ensured by the local contraction condition established in Aguirregabiria and Mira (2002) for a single agent dynamic discrete choice problem and in Kasahara and Shimotsu (2012) for dynamic games. Our Lemma 1 under Assumption 1, 2, 3 establishes the contraction mapping condition of the BNE. Another condition for the convergence of NPL is that the initial starting point needs to be in the neighborhood of the true parameter. Thus, NPL algorithm converges to a consistent estimator in our binary choice with misclassification and social interactions. In our Monte Carlo experiments and the empirical application, we choose the initial starting value of the parameter as the combination of the Logit regression on the covariate parameter (β) and different points for those misclassification parameters (γ,δ) and the peer effects parameter (α). We notice similar convergence results as those seen in the literature (Aguirregabiria and Mira, 2007; Kasahara and Shimotsu, 2012; Lin and Xu, 2017). An analog estimator for the variance is

$$\widehat{V}_{NPL} = \widehat{V}_1^{-1}(\widehat{\theta}_{NPL})\widehat{V}_2(\widehat{\theta}_{NPL})\widehat{V}_1^{'-1}(\widehat{\theta}_{NPL}),$$

where

$$\widehat{V}_1(\widehat{\theta}_{NPL}) = \frac{\partial^2\mathscr{L}(\widehat{\theta}_{NPL}, \widehat{P}_{NPL})}{\partial\theta\partial\theta'} + \frac{\partial^2\mathscr{L}(\widehat{\theta}_{NPL}, \widehat{P}_{NPL})}{\partial\theta\partial P'}\left[I - \left(\frac{\partial\Gamma(\widehat{P}_{NPL}, \widehat{\theta}_{NPL})}{\partial P}\right)'\right]^{-1}\frac{\partial\Gamma(\widehat{P}_{NPL}, \widehat{\theta}_{NPL})}{\partial\theta'}$$

and

$$\widehat{V}_2(\widehat{\theta}_{NPL}) = \frac{1}{n}\sum_{i\in\mathscr{I}}\frac{\partial\mathscr{L}_i(\widehat{\theta}_{NPL}, \widehat{P}_{NPL})}{\partial\theta}\frac{\partial\mathscr{L}_i(\widehat{\theta}_{NPL}, \widehat{P}_{NPL})}{\partial\theta'}.$$

5. Monte Carlo experiments

The Monte Carlo experiments are designed to mimic the peer effects in attitude study in Section 6. We conduct three Monte Carlo experiments to investigate the finite sample performance of the model and the NPL algorithm. The Monte Carlo designs have three covariates: X_1 is drawn from a standard normal distribution, X_2 is drawn from a uniform distribution $U[-\sqrt{3},\sqrt{3}]$, and X_3 is drawn from a discrete distribution taking values from $\{-1,1\}$ with equal probability $\frac{1}{2}$. X_1,X_2, and X_3 have a mean of 0 and a variance of 1. We generate a random network with a maximum number of friends at 10 (the same as in the *Add Health* dataset). Each individual i has a degree independently drawn from $N_i \in \{0,1,\ldots,10\}$ with equal probabilities. Then, we randomly choose N_i of the other $n-1$ individuals for this individual to have as friends. The network is thus directed, in that j can influence i without requiring (but not precluding) that i influence j in return. The example can also be done with any other network, but this simplifies the code. We then solve fixed points with known parameter values for $\mathbb{E}(Y_i^*|\mathbb{I}_n), i \in \mathscr{I}$. The latent dependent variable is given by

$$Y_i^* = \mathbf{1}\left\{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \frac{\alpha}{N_i}\sum_{j\in F_i}\mathbb{E}\left(Y_j^*\middle|\mathbb{I}_n\right) - \varepsilon_i \geq 0\right\}, \tag{5.1}$$

where ε_i is drawn from a standard logistic distribution. β is set to be $(-1,1,-1,1)$ for all three experiments. We let $\alpha = 1$ in all experiments. Two observed measurements, Y_{1i}, and Y_{2i} are generated with misclassification probabilities, $(\gamma,\delta) = (0.1,0.1),(0.2,0.2)$ and (0.4,0.4) in three experiments, respectively.

We generate 1,000 samples of pseudo-random numbers with $n \in \{500,\ 1,000,\ 2,000\}$. We denote $\widehat{\theta} = (\widehat{\gamma},\widehat{\delta},\widehat{\beta},\widehat{\alpha})$ as the NPL estimates with misclassification correction (2M model) and $\widetilde{\mu}_1 = (\widetilde{\beta}_1,\widetilde{\alpha}_1), \widetilde{\mu}_2 = (\widetilde{\beta}_2,\widetilde{\alpha}_2)$ as the NPL estimates without misclassification correction; i.e., taking Y_1 or Y_2 as the truly observed binary choice (proxy method, denoted as M1 and M2 models), respectively. We report the average biases and the mean square errors of our misclassification corrected estimates in Tables 1, 3 and 5 and of non-corrected estimates in Tables 2, 4, and 6. We report the average correlations between the true latent variable Y_i^* and the conditional choice probabilities from the 2M, M1, and M2 models in Table 7.

The NPL estimators with a misclassification correction converge to the true parameter at the $\sqrt{n}$ rate, while those without a misclassification correction (proxy method) do not converge even when studying such a large sample size. The results demonstrate the good finite sample performance of the NPL algorithm for the binary choice model with misclassification and social interactions.

6. Peer effects in attitude

Students live in two distinct social worlds: the hierarchical world with adults and the egalitarian world with peers. The former introduces students to the society as new members and the latter helps students develop skills like negotiation, cooperation, and so on. Students interact with peers in many different activities; e.g., studying together, attending sports clubs, conducting delinquent behaviors, etc. Among these spillovers, the peer effects in education have received considerable attention in the literature; see more details in Epple and Romano (2011), Sacerdote (2011) and Bursztyn and Jensen (2017). When it comes to the learning spillover, scholars emphasize the achievements of students, e.g. Hoxby (2000), Zimmerman (2003) and Calvó-Armengol et al. (2009) to name only a few. However, in the context of education, students have partial control over the outcomes and the simple production function is difficult to illustrate the process from inputs to the outcomes.

There are two main factors determining students' achievements: ability and attitude[8]. Ability is the physical or mental power to do something and is usually unobserved. The unobserved ability causes endogeneity problems in many studies; e.g., return to schooling. A proxy or IV approach is adopted to handle the unobserved ability in a cross-sectional setting. Arcidiacono et al. (2012) treat ability as the unobserved heterogeneity in the panel data model and remove this unobserved heterogeneity by standard approaches in panel data models with fixed effects. Fruehwirth (2014) deploys a specific relationship between achievement and the ability to investigate the "black box". Generally, genetics and learning shape ability, and people do not make conscious choices on ability.

Attitude towards learning is the way of thinking or feeling about studying and educational aspirations. Typically, attitude is reflected in a student's behavior and originates from the student's choices. Peer effects demonstrate the interconnection among students on choices; e.g., whether one makes tremendous efforts in the study, exercises, smokes, drinks, etc. For learning spillover, peer effects play a role in the chosen attitude rather than in one's final achievements. Thus, the investigation on peer effects in attitude is interesting, however, attitude is subjective and difficult to measure. In the *National Longitudinal Study of Adolescent Health (Add Health)* dataset, we obtain several measurements[9] in the survey regarding the attitudes of students. Attitudes regarding questions are socially and personally sensitive and students tend to misreport. This feature raises the issue of misclassification errors.

Using both in-school and at-home surveys, we obtain two repeated measurements for attitude from the question, "Skipped school without an excuse" in the in-school and at-home surveys. We estimate the peer effects in attitude with our method to rectify potential misclassification errors and compare the results to those obtained when each of the two measurements is used as a proxy of the latent attitude.

[8] Scholars also work on the "effort", which to some extent is the "realization" of attitude. In this paper, we use attitude as the "choice" variable. For investigation on "effort" in field experiment studies, see Bursztyn and Jensen (2015) and Bursztyn et al. (2019).

[9] Measurement means it contains the majority of the information of the latent variable, but there is an error.

Table 1
Experiment I.

True Parameters: $\theta_0 = (0.1, 0.1; -1, 1, -1, 1; 1)$							
Average Bias							
n	$\widehat{\gamma}_{NPL}$	$\widehat{\delta}_{NPL}$	$\widehat{\beta}_{NPL}$				$\widehat{\alpha}_{NPL}$
500	-0.002	-0.002	-0.048	0.039	-0.031	0.035	0.032
1,000	-0.001	-0.002	-0.015	0.018	-0.010	0.009	0.014
2,000	0.000	0.001	-0.009	0.010	-0.009	0.010	-0.007
Mean Square Errors							
500	0.002	0.002	0.141	0.043	0.043	0.040	0.432
1,000	0.001	0.001	0.067	0.017	0.018	0.017	0.234
2,000	0.000	0.001	0.030	0.009	0.008	0.008	0.108

Table 2
Experiment I: Non-Correction.

True Parameters: $\theta_0 = (0.1, 0.1; -1, 1, -1, 1; 1)$										
Average Bias										
n	$\widetilde{\beta}_1$				$\widetilde{\alpha}_1$	$\widetilde{\beta}_2$				$\widetilde{\alpha}_2$
500	0.441	-0.219	0.222	-0.215	-0.223	0.437	-0.219	0.221	-0.217	-0.207
1,000	0.451	-0.224	0.225	-0.224	-0.228	0.449	-0.222	0.225	-0.223	-0.222
2,000	0.455	-0.229	0.226	-0.221	-0.242	0.457	-0.231	0.228	-0.222	-0.243
Mean Square Errors										
500	0.255	0.062	0.063	0.059	0.323	0.252	0.063	0.063	0.059	0.332
1,000	0.236	0.057	0.058	0.056	0.212	0.236	0.056	0.057	0.056	0.213
2,000	0.223	0.056	0.054	0.052	0.134	0.225	0.057	0.055	0.052	0.137

Table 3
Experiment II.

True Parameters: $\theta_0 = (0.2, 0.2; -1, 1, -1, 1; 1)$							
Average Bias							
n	$\widehat{\gamma}_{NPL}$	$\widehat{\delta}_{NPL}$	$\widehat{\beta}_{NPL}$				$\widehat{\alpha}_{NPL}$
500	-0.005	-0.004	-0.061	0.037	-0.037	0.033	0.073
1,000	-0.002	-0.002	-0.019	0.019	-0.021	0.017	-0.001
2,000	-0.002	-0.001	-0.006	0.007	-0.007	0.010	-0.005
Mean Square Errors							
500	0.003	0.003	0.182	0.048	0.050	0.048	0.531
1,000	0.001	0.001	0.083	0.023	0.021	0.020	0.264
2,000	0.001	0.001	0.039	0.010	0.010	0.010	0.128

Table 5
Experiment III.

True Parameters: $\theta_0 = (0.4, 0.4; -1, 1, -1, 1; 1)$							
Average Bias							
n	$\widehat{\gamma}_{NPL}$	$\widehat{\delta}_{NPL}$	$\widehat{\beta}_{NPL}$				$\widehat{\alpha}_{NPL}$
500	-0.007	-0.006	-0.075	0.062	-0.052	0.053	0.067
1,000	-0.002	-0.002	-0.026	0.036	-0.031	0.030	-0.011
2,000	0.000	-0.001	-0.017	0.017	-0.015	0.016	-0.006
Mean Square Errors							
500	0.004	0.004	0.306	0.085	0.080	0.075	0.725
1,000	0.002	0.002	0.130	0.034	0.033	0.031	0.362
2,000	0.001	0.001	0.067	0.015	0.015	0.015	0.202

Table 4
Experiment II: Non-Correction.

	True Parameters: $\theta_0 = (0.2, 0.2; -1, 1, -1, 1; 1)$									
	Average Bias									
n	$\widetilde{\beta}_1$				$\widetilde{\alpha}_1$	$\widetilde{\beta}_2$				$\widetilde{\alpha}_2$
500	0.787	-0.363	0.359	-0.354	-0.355	0.792	-0.365	0.358	-0.356	-0.362
1,000	0.814	-0.369	0.361	-0.355	-0.407	0.807	-0.365	0.360	-0.357	-0.392
2,000	0.810	-0.369	0.361	-0.355	-0.399	0.819	-0.369	0.364	-0.355	-0.414
	Mean Square Errors									
500	0.677	0.144	0.141	0.137	0.347	0.693	0.145	0.139	0.138	0.376
1,000	0.695	0.142	0.137	0.131	0.291	0.685	0.140	0.135	0.133	0.283
2,000	0.673	0.139	0.133	0.128	0.229	0.688	0.139	0.135	0.129	0.240

Table 6
Experiment III: Non-Correction.

	True Parameters: $\theta_0 = (0.4, 0.4; -1, 1, -1, 1; 1)$									
	Average Bias									
n	$\widetilde{\beta}_1$				$\widetilde{\alpha}_1$	$\widetilde{\beta}_2$				$\widetilde{\alpha}_2$
500	1.408	-0.532	0.520	-0.515	-0.565	1.419	-0.530	0.522	-0.511	-0.577
1,000	1.443	-0.532	0.521	-0.515	-0.630	1.446	-0.531	0.527	-0.514	-0.633
2,000	1.446	-0.533	0.527	-0.517	-0.633	1.438	-0.535	0.523	-0.516	-0.623
	Mean Square Errors									
500	2.045	0.295	0.281	0.277	0.475	2.080	0.291	0.284	0.272	0.494
1,000	2.116	0.289	0.277	0.270	0.483	2.123	0.287	0.283	0.269	0.483
2,000	2.110	0.286	0.281	0.270	0.454	2.089	0.289	0.277	0.269	0.441

Table 7
Prediction Power in 2M, M1 and M2 Models.

	Experiment I			Experiment II			Experiment III		
n	2M	M1	M2	2M	M1	M2	2M	M1	M2
500	0.595	0.590	0.590	0.593	0.581	0.580	0.590	0.559	0.558
1,000	0.592	0.587	0.587	0.593	0.582	0.581	0.592	0.561	0.561
2,000	0.592	0.588	0.588	0.592	0.581	0.581	0.591	0.562	0.562

6.1. *The add health data*

The National Longitudinal Study of Adolescent Health (*Add Health*) is a longitudinal study of a nationally representative sample of adolescents in grades 7-12 in the United States during the 1994-95 school year for the first wave. The study also contains Wave II, III, and IV data, which are collected in 1995-1996, 2001-2002, and 2008 (Harris et al., 2009). *Add Health* combines longitudinal survey data on respondents' social and economic features with contextual data on family, friendships, and peer groups. In this paper, we use the data from Wave I.

In the *Add Health* dataset, each student can nominate at most five male friends and at most five female friends, from which we construct the network with direct links $[\{F_{ij}\}_{i,j=1}^{n}]$. Note that although students have at most 10 outlinks, they may have more than 10 inlinks. In Wave I, there are both in-school and at-home questionnaires, which generate multiple measurements for the attitude variable. The *Add Health* dataset also includes questionnaires for demographic characteristics such as age, parents' education, race information, gender, etc.

As students, not only their achievements but also their attitudes toward learning are important. Attitude is a vague and subjective concept. Thus the study of attitude exhibits misclassification problems. Fortunately, the *Add Health* dataset contains repeated measurements of students' attitudes. There is a question, "During the past twelve months, how often did you skip school without an excuse?" in the in-school survey. In the at-home survey, there is a question "During this school year how many times {HAVE YOU SKIPPED/DID YOU SKIP} school for a full day without an excuse?". We take the answer for the at-home question as Y_1 and the answer for the in-school question as Y_2. We take the answer "never" as a "positive" attitude and all other answers as "negative" attitudes. Here, Y_1 and Y_2 are obvious measurements for the same question related to the student's attitude. This provides enough data for the identification of the conditional distribution of the latent attitude in our first step identification. In the Add Health dataset, there is also a question regarding excused absences which is not a good measurement of the latent attitude since excuses generate a lot of noise.

It is worth pointing out that the answers in the survey are ordered discrete and it is feasible to generalize the model to allow for ordered discrete choices (Greene and Hensher, 2010). We can obtain similar identification results by extending the arguments used in Hu (2008) for the ordered discrete covariates. In this paper, we focus on binary outcomes and ordered discrete choice is deferred to future research.

We consider the sister schools No. 77 and No. 177 for our analysis[10]. These two schools contain the largest single connected school network in the *Add Health* dataset. After data management, we obtain 1,173 students in the sample. For these students, the number that $Y_1 = Y_2$ is 801. This implies that these two measurements capture the main information of the latent attitude, but there are noises in each of them. Table 8 summarizes the statistics of the demographic characteristics and the attitude variables. More than half of the students leave school without an excuse at least once in the year. The attitude measurement from the at-home questionnaire regarding skipping school without excuse has a little less "positive" than the measurement obtained from the in-school questionnaire; 45.0 percent versus 47.1 percent, respectively.

6.2. *The hidden peer effects in attitude*

When it comes to the peer effects in attitude, we have three options to back out the interaction parameter. We can either take Y_1 or Y_2 as the true latent attitude to estimate the binary choice with social interactions without misclassification correction (Proxy method, Model M1, and M2), or we adopt the full information from two repeated measurements to rectify the misclassification errors (2M model).

We report our estimation results in Table 9. The overreporting probabilities are 25.7 percent, and 29.2 percent, respectively. Roughly, one-quarter to one-third of students overreport their attitudes in these two measurements when their latent attitudes are negative. Students are more likely to be honest at home where there are no peers.

In Table 9, models without misclassification correction either fail to detect a significant silent rivalry ($\widehat{\alpha} = 0$ in model M1) or underestimate the peer effects ($\widehat{\alpha} = 0.477$ in model M2). Our 2M model estimates a significant 1.658[11] peer effects parameter which is three times bigger than the model with the in-school measurement. We also provide results for simple Logit models without simultaneous peer effects on attitudes towards learning. The results are very similar for demographic covariates, e.g. older students pay more attention to their studies as they mature.

To summarize, we find that the peer effects in attitude are underestimated if we directly use these two measurements as attitudes. Our findings confirm our insight into the prevalence of peer effects in attitude among students and support the conclusion in the Coleman Report 1966 that "academic achievement was less related to the quality of a student's school and more related to the social composition of the school, the student's sense of control of his environment and future, the verbal skills of teachers, and the student's family background". The evidence of overreporting in subjective and sensitive attitude questions suggests that we should take into account the potential misclassification issues when investigating the peer effects in attitude. Our investigation has important policy implications, such as manipulating peer group influence.

Besides attitude, there are many other choices of adolescents or/and adults which prone to measurement errors due to a variety of reasons. The remedy from repeated measurements helps to pin down the bias generated by measurement errors. Our binary choice model with social interactions and misclassification is applicable to any scenario where there are potential misclassification issues and strategic interactions. For instance, students may overreport sex behavior or romantic status if they consider it to be cool. Therefore, investigation on the peer effects in sex behavior or in romantic status would bear potential misclassification issues and the resulting bias in mind.

7. Conclusion

In this paper, we propose to correct the potential misclassification error problem of the dependent variables in social interactions studies. We provide a closed-form identification result to our model primitives by adopting a technique of two measurements and a monotonicity approach. Taking into account the full information embedded in the two measurements and the monotonicity condition, we construct a complete likelihood function for the estimation of the structural parameter using the nested pseudo likelihood algorithm. We establish asymptotic results for the NPL estimator and illustrate the finite sample performance with Monte Carlo experiments and an application to peer effects in attitude towards learning. We find that a significant proportion of students overreport their attitudes towards learning. The rectified peer effects is larger than those obtained from the estimates when misclassified attitudes are taken as the correctly observed ones. The peer effects in attitude triggers multiplier effects which help improve the performance of schools and are meaningful for policy implications.

[10] There was sister school roster for friend nomination in *Add Health* dataset. This design generated friendships across the sister schools.

[11] Standard errors obtain from the last step MLE with convergence tolerance of NPL algorithm satisfied. As the last step MLE is calculated using the near equilibrium choice probabilities, the MLE standard error is very close to the NPL standard error, which is consistent with our simulation results. This is a limitation of the paper. We are working on a project to derive bootstrapping standard error for the network generated dependent data.

Table 8
Summary of Statistics of Key Variables.

Variable	Mean	Std. Dev.
Age	15.882	1.187
Female	0.497	0.500
Parents' Education	5.257	2.459
White	0.092	0.289
American Indian	0.049	0.215
Asian	0.348	0.476
African American	0.265	0.442
Hispanic	0.384	0.487
Others*	0.130	0.336
Attitude(Y_1)	0.450	0.498
Attitude(Y_2)	0.471	0.499

*Some students are associated with more than one race.

Table 9
Estimation Results on Silent Rivalry.

	2M	M1	M2	Logit models	
				Y_1	Y_2
Age	− 0.445**	− 0.348**	− 0.201**	− 0.350**	− 0.199**
	(0.128)	(0.056)	(0.053)	(0.056)	(0.053)
Female	− 0.061	0.113	− 0.108	0.111	− 0.083
	(0.164)	(0.122)	(0.119)	(0.121)	(0.119)
Parents' Education	0.046	0.009	0.029	0.009	0.032
	(0.041)	(0.027)	(0.026)	(0.027)	(0.026)
Hispanic	− 0.668**	− 0.495**	− 0.239	− 0.499**	− 0.239
	(0.303)	(0.197)	(0.192)	(0.197)	(0.192)
Asian	− 0.265	− 0.097	− 0.160	− 0.099	− 0.121
	(0.259)	(0.201)	(0.197)	(0.199)	(0.196)
African American	− 0.154	− 0.085	− 0.123	− 0.090	− 0.141
	(0.254)	(0.207)	(0.204)	(0.207)	(0.204)
Native American	− 0.764	− 0.090	− 0.374	− 0.091	− 0.389
	(0.581)	(0.288)	(0.286)	(0.288)	(0.286)
Other	0.230	0.327*	− 0.044	0.327	− 0.033
	(0.265)	(0.197)	(0.193)	(0.197)	(0.193)
γ	0.257**	–	–	–	–
	(0.071)	–	–	–	–
δ	0.292**	–	–	–	–
	(0.067)	–	–	–	–
Peer Effects (α)	1.658**	0.000	0.477*	–	–
	(0.772)	(0.289)	(0.276)	–	–
Constant	5.776**	5.416**	3.018**	5.464**	3.110**
	(1.639)	(0.947)	(0.902)	(0.936)	(0.896)

a. ** and * for 5% and 10% significance, respectively. c. White students are left for comparison.

Appendix A. Proofs

Proof of Lemma 1.. The existence of the BNE is guaranteed by Schauder's fixed-point theorem and the continuity of $\Gamma(\cdot)$. Consider that there are two distinct BNEs: $P^1 = (P_1^1, P_2^1, \ldots, P_n^1) \neq (P_1^2, P_2^2, \ldots, P_n^2) = P^2$. We have:

$$
\begin{aligned}
\left|P_i^1 - P_i^2\right| = \left|\Gamma_i(\mathbb{I}_n, P^1; \mu) - \Gamma_i(\mathbb{I}_n, P^2; \mu)\right| &= \left|\Lambda\left(X_i'\beta + \frac{\alpha}{N_i}\sum_{j\in F_i} P_j^1\right) - \Lambda\left(X_i'\beta + \frac{\alpha}{N_i}\sum_{j\in F_i} P_j^2\right)\right| \\
&= \Lambda\left(X_i'\beta + \frac{\alpha}{N_i}\sum_{j\in F_i} P_j^\dagger\right)\left[1 - \Lambda\left(X_i'\beta + \frac{\alpha}{N_i}\sum_{j\in F_i} P_j^\dagger\right)\right]\left|\frac{\alpha}{\mathrm{N_i}}\sum_{\mathrm{j}\in \mathrm{F_i}}\left(\mathrm{P_j^1} - \mathrm{P_j^2}\right)\right| \\
&\leq \frac{1}{4}\cdot|\alpha|\cdot \max_{\mathrm{j}\in\mathscr{I}}\left|\mathrm{P_j^1} - \mathrm{P_j^2}\right| < 4\cdot\frac{1}{4}\max_{\mathrm{j}\in\mathscr{I}}\left|\mathrm{P_j^1} - \mathrm{P_j^2}\right| = \max_{\mathrm{j}\in\mathscr{I}}\left|\mathrm{P_j^1} - \mathrm{P_j^2}\right|
\end{aligned}
\tag{A.1}
$$

where $P_j^\dagger$ is the probability between P_j^1 and P_j^2. The third line comes from the Mean Value theorem and the inequality is based on $\Lambda(\cdot)[1 - \Lambda(\cdot)] \leq \frac{1}{4}$. Taking maximization over $i \in \mathscr{I}$ on the left-hand side of Eq. (A.1), we have:

$$\max_{i\in\mathscr{I}}|P_i^1 - P_i^2| < \max_{j\in\mathscr{I}}\left|P_j^1 - P_j^2\right|,$$

which is a contradiction. Therefore, we have a unique BNE for the Bayesian game in Eq. (2.3).□

Proof of Proposition 1.. Consider a super population with infinite number of individuals, the uniqueness of the Bayesian game implies that, for all $y_1,\ y_2 \in \{0,\ 1\}$, we know $\mathbb{P}(Y_{1i} = y_1, Y_{2i} = y_2 \mid \mathbb{I}_n), i \in \mathscr{I}$ are identified from the observables. Define $\mathbb{P}_{Y_{1i},Y_{2i}}(y_1,y_2) \equiv \mathbb{P}(Y_{1i} = y_1, Y_{2i} = y_2|\mu, \mathbb{I}_n), \mathbb{P}_{Y_{1i}|Y_i^*}(y_1|y) \equiv \mathbb{P}(Y_{1i} = y_1\left|Y_i^* = y, \mu, \mathbb{I}_n\right), \mathbb{P}_{Y_{2i}|Y_i^*}(y_2|y) \equiv \mathbb{P}(Y_{2i} = y_2\left|Y_i^* = y, \mu, \mathbb{I}_n\right)$ and $\mathbb{P}_{Y_i^*}(y) \equiv \mathbb{P}(Y_i^* = y\left|\mu, \mathbb{I}\right)$ for $y_1, y_2, y \in \{0,1\}$. By the law of total probability, we have

$$\begin{aligned}
\mathbb{P}_{Y_{1i},Y_{2i}}(1,0) &= \sum_{y\in\{0,1\}} \mathbb{P}_{Y_{1i}|Y_i^*}(1|y)\cdot\mathbb{P}_{Y_{2i}|Y_i^*}(0|y)\cdot\mathbb{P}_{Y_i^*}(y);\\
\mathbb{P}_{Y_{1i},Y_{2i}}(0,1) &= \sum_{y\in\{0,1\}} \mathbb{P}_{Y_{1i}|Y_i^*}(0|y)\cdot\mathbb{P}_{Y_{2i}|Y_i^*}(1|y)\cdot\mathbb{P}_{Y_i^*}(y);\\
\mathbb{P}_{Y_{1i},Y_{2i}}(0,0) &= \sum_{y\in\{0,1\}} \mathbb{P}_{Y_{1i}|Y_i^*}(0|y)\cdot\mathbb{P}_{Y_{2i}|Y_i^*}(0|y)\cdot\mathbb{P}_{Y_i^*}(y).
\end{aligned}$$

With the monotonicity condition in Assumption 6, we have

$$\begin{aligned}
\mathbb{P}_{Y_{1i},Y_{2i}}(1,0) &= \mathbb{P}_{Y_{1i}|Y_i^*}(1|0)\cdot\mathbb{P}_{Y_{2i}|Y_i^*}(0|0)\cdot\mathbb{P}_{Y_i^*}(0);\\
\mathbb{P}_{Y_{1i},Y_{2i}}(0,1) &= \mathbb{P}_{Y_{1i}|Y_i^*}(0|0)\cdot\mathbb{P}_{Y_{2i}|Y_i^*}(1|0)\cdot\mathbb{P}_{Y_i^*}(0);\\
\mathbb{P}_{Y_{1i},Y_{2i}}(0,0) &= \mathbb{P}_{Y_{1i}|Y_i^*}(0|0)\cdot\mathbb{P}_{Y_{2i}|Y_i^*}(0|0)\cdot\mathbb{P}_{Y_i^*}(0).
\end{aligned}$$

Simple calculations lead to

$$\begin{aligned}
\mathbb{P}_{Y_{1i}|Y_i^*}(1|0) &= \frac{\mathbb{P}_{Y_{1i},Y_{2i}}(1,0)}{\mathbb{P}_{Y_{1i},Y_{2i}}(0,0)+\mathbb{P}_{Y_{1i},Y_{2i}}(1,0)} = \frac{\mathbb{P}_{Y_{1i},Y_{2i}}(1,0)}{\mathbb{P}_{Y_{2i}}(0)} \quad ;\\
\mathbb{P}_{Y_{2i}|Y_i^*}(1|0) &= \frac{\mathbb{P}_{Y_{1i},Y_{2i}}(0,1)}{\mathbb{P}_{Y_{1i},Y_{2i}}(0,0)+\mathbb{P}_{Y_{1i},Y_{2i}}(0,1)} = \frac{\mathbb{P}_{Y_{1i},Y_{2i}}(0,1)}{\mathbb{P}_{Y_{1i}}(0)} \quad ;\\
\mathbb{P}_{Y_i^*}(0) &= \frac{\mathbb{P}_{Y_{2i}}(0)\cdot\mathbb{P}_{Y_{1i}}(0)}{\mathbb{P}_{Y_{1i},Y_{2i}}(0,0)},
\end{aligned}$$

provided that the Assumption 5 is satisfied. The last equation and $\mathbb{P}_{Y_i^*}(1) = 1 - \mathbb{P}_{Y_i^*}(0)$ give us the closed-form identification.□

Proof of Theorem 1.. The proof is similar as that in Aguirregabiria and Mira (2007), Newey and McFadden (1994). With Assumption 9(ii), we see that $\theta_{NPL} = \theta_0$. Recall that the pseudo likelihood function is $\mathscr{L}(\theta, P_{[n]})$ in the NPL estimation. Define the function

$$T\left(\theta, P_{[n]}\right) \equiv \max_{c\in\Theta}\left\{\mathscr{L}_0\left(c, P_{[n]}\right)\right\} - \mathscr{L}_0\left(\theta, P_{[n]}\right).$$

Because $\mathscr{L}_0(\theta, P_{[n]})$ is continuous and $\Theta\times\mathscr{P}_n$ is compact, Berge's maximum theorem establishes that $T(\theta, P_{[n]})$ is a continuous function. By construction, $T(\theta, P_{[n]}) \ge 0$ for any $(\theta, P_{[n]})$. Let $\mathscr{E}$ be the set of vectors $(\theta, P_{[n]})$ that are fixed points of the equilibrium mapping Γ, i.e.,

$$\mathscr{E} \equiv \left\{\left(\theta, P_{[n]}\right) \in \Theta\times\mathscr{P}_n : P_{[n]} = \Gamma\left(\theta; P_{[n]}\right)\right\}$$

Given that $\Theta\times\mathscr{P}_n$ is compact and Γ is continuous, $\mathscr{E}$ then is a compact set. By definition, the set Λ_{0n} is included in $\mathscr{E}$. Let $B_\epsilon(\theta_0, P^*_{[n]}) = \{\theta \in \mathbb{R}^{d+3} : \|\theta - \theta_0\| + \|P_{[n]} - P^*_{[n]}\| < \epsilon\}$ be an arbitrarily small open ball that contains $(\theta_0, P^*_{[n]})$. We then see that $B^C_\epsilon(\theta_0, P^*_{[n]}) \cap \mathscr{E}$ is also compact. Define the constant

$$\tau \equiv \min_{\left(\theta, P_{[n]}\right)\in B^C_\epsilon\left(\theta_0, P^*_{[n]}\right)\cap\mathscr{E}} T\left(\theta, P_{[n]}\right). \tag{A.2}$$

When $(\theta, P_{[n]}) \in B^C_\epsilon(\theta_0, P^*_{[n]})$, Assumption 9(ii) implies that $T(\theta, P_{[n]}) \ge 0$ for all $\theta \ne \theta_0$. By construction, $\tau > 0$. Define the event

$$A \equiv \left\{\left(\theta, P_{[n]}\right) \in \Theta\times\mathscr{P}_n : \left|\mathscr{L}\left(\theta, P_{[n]}\right) - \mathscr{L}_0\left(\theta, P_{[n]}\right)\right| < \frac{\tau}{2} \text{ for all } \left(\theta, P_{[n]}\right) \in \Theta\times\mathscr{P}_n\right\}$$

Let $(\theta^{(n)}, P^{(n)})$ be an element of Λ_n. A implies

$$\mathscr{L}_0\left(\theta^{(n)}, P^{(n)}\right) > \mathscr{L}\left(\theta^{(n)}, P^{(n)}\right) - \frac{\tau}{2},$$

and

$$\mathscr{L}\left(\theta, P^{(n)}\right) > \mathscr{L}_0\left(\theta, P^{(n)}\right) - \frac{\tau}{2}.$$

Furthermore, we have $\mathscr{L}(\theta^{(n)}, P^{(n)}) \geq \mathscr{L}(\theta, P^{(n)})$ from the NPL fixed point definition. This is different from the standard argument in the literature (e.g. Newey and McFadden, 1994). Therefore, we see that $\mathscr{L}_0(\theta^{(n)}, P^{(n)}) > \mathscr{L}_0(\theta, P^{(n)}) - \tau$. We then have the following derivation:

$$\begin{aligned} A \quad &\Rightarrow \quad \left\{\mathscr{L}_0\left(\theta^{(n)}, P^{(n)}\right) > \mathscr{L}_0\left(\theta, P^{(n)}\right) - \tau \text{ for any } \theta \in \Theta\right\} \\ &\Rightarrow \quad \left\{\mathscr{L}_0\left(\theta^{(n)}, P^{(n)}\right) > \max_{\theta\in\Theta}\mathscr{L}_0\left(\theta, P^{(n)}\right) - \tau\right\} \Rightarrow \left\{\tau > T\left(\theta^{(n)}, P^{(n)}\right)\right\} \\ &\Rightarrow \quad \left\{\min_{\left(\theta, P_{[n]}\right)\in B_\epsilon^C\left(\theta_0, P^*_{[n]}\right)\cap\mathscr{E}} T\left(\theta, P_{[n]}\right)\rangle T\left(\theta^{(n)}, P^{(n)}\right)\right\} \text{ by Equation (A.2)} \\ &\Rightarrow \quad \left\{\left(\theta^{(n)}, P^{(n)}\right) \in B_\epsilon\left(\theta_0, P^*_{[n]}\right)\right\} \end{aligned}$$

The last induction uses the fact that $(\theta^{(n)}, P^{(n)}) \in \mathscr{E}$. Therefore, $\Pr(A) \leq \Pr((\theta^{(n)}, P^{(n)}) \in B_\epsilon(\theta_0, P^*_{[n]}))$.

The conditional pseudo likelihood function, $\mathscr{L}(\theta, P_{[n]})$ is a likelihood function of θ and $P_{[n]}$. It is defined with an arbitrary choice probabilities profile $P_{[n]} = (P_1, \ldots, P_n)$. $P_{[n]}$ is not necessarily the equilibrium conditional choice probabilities profile, $P^*_{[n]}$ and can be any arbitrary choice probabilities profile in $[0,1]^n$, e.g., $P_{[n]} = (0.5, \ldots, 0.5)$. This is to say that $P_{[n]}$ is not the fixed point solution of $P_{[n]} = \Gamma(\mu; P_{[n]}, \mathbb{I}_n)$. For each observation, i, $P_{[n]}$ boils down to the pseudo social interactions term, $\frac{1}{N_i}\sum_{j\in F_i} P_j$ in $\Gamma_i(\mu; P_{[n]}, \mathbb{I}_n)$ defined in Eq. (2.4). By the nature of $P_j \in [0,1]$, we have that $\frac{1}{N_i}\sum_{j\in F_i} P_j \in [0,1]$. This term could be taken as a standard bounded covariate. Furthermore, the conditional pseudo likelihood function, $\mathscr{L}(\theta, P_{[n]})$, is based on the cumulative distribution function of the standard Logistic distribution and thus is continuous and differentiable in its parameter. The public information set, $\mathbb{I}_n$, is taken as an argument in the conditional pseudo likelihood function. By the nature of the Logistic CDF, we have that $E[\mathscr{L}_i(\theta, P_{[n]}) | \mathbb{I}_n]$ and $\left|E[\mathscr{L}_i(\theta, P_{[n]}, Y_i = 1) - L_i(\theta, P_{[n]}, Y_i = 0) | \mathbb{I}_n]\right|$ continuous in $\frac{1}{N_i}\sum_{j\in F_i} P_j$. Combining with the invertibility in Assumption 9(iii), boundedness from Assumption 10 (iii) and VC class from Assumption 10(i), we have all conditions of the conditional law of large number (Theorem 4.1) in Menzel (2016) hold. Thus, by the conditional law of large number theorem in Menzel (2016), we have uniform convergence of $\mathscr{L}(\theta, P_{[n]})$ to its population conditional mean, $\mathscr{L}_0(\theta, P_{[n]})$. The uniform convergence of $\mathscr{L}(\theta, P_{[n]})$ to $\mathscr{L}_0(\theta, P_{[n]})$ implies that $\Pr(A) \to 1$ as $n \to \infty$. Thus, $\Pr((\theta^{(n)}, P^{(n)}) \in B_\epsilon(\theta_0, P^*_{[n]})) \to 1$. The ϵ in $B_\epsilon(\theta_0, P^*_{[n]})$ is an arbitrarily small constant, so we have:

$$\sup_{\left(\theta^{(n)}, P_{[n]}\right)\in A} \|\theta^{(n)} - \theta\| + \|P^{(n)} - P^*_{[n]}\| = o_p(1).$$

where $\|\cdot\|$ is the Euclidean norm. From the definition of Λ_n, we see that $\widehat{\theta}_{NPL} \xrightarrow{p} \theta_0$. Now, we establish the asymptotic normality of the NPL estimator. Taking a Taylor expansion over the first order condition in Eq. (4.5) around the true parameter $(\theta_0, P^*_{[n]})$, we have:

$$\frac{\partial\mathscr{L}\left(\theta_0, P^*_{[n]}\right)}{\partial\theta} + \frac{\partial^2\mathscr{L}(\theta^+, P^+)}{\partial\theta\partial\theta'}(\widehat{\theta}_{NPL} - \theta_0) + \frac{\partial^2\mathscr{L}(\theta^+, P^+)}{\partial\theta\partial P'}\left(\widehat{P}_{NPL} - P^*_{[n]}\right) = 0 \tag{A.3}$$

where θ^+ is between $\widehat{\theta}_{NPL}$ and θ_0 and P^+ are between $\widehat{P}_{NPL}$ and $P^*_{[n]}$, respectively. Applying the same stochastic mean value theorem between $(\theta_0, P^*_{[n]})$ and $(\widehat{\theta}_{NPL}, \widehat{P}_{NPL})$ to $\widehat{P}_{NPL} = \Gamma(\widehat{\theta}_{NPL}, \widehat{P}_{NPL})$ leads to

$$\left[I - \left(\frac{\partial\Gamma(\theta^-; P^-)}{\partial P}\right)'\right]\left(\widehat{P}_{NPL} - P^*_{[n]}\right) - \frac{\partial\Gamma(\theta^-; P^-)}{\partial\theta}\left(\widehat{\theta}_{NPL} - \theta_0\right) = 0, \tag{A.4}$$

where θ^- is between $\widehat{\theta}_{NPL}$ and θ_0 and P^- are between $\widehat{P}_{NPL}$ and $P^*_{[n]}$, respectively. Solving Eq. (A.4) into Eq. (A.3) gives

$$\frac{\partial\mathscr{L}\left(\theta_0, P^*_{[n]}\right)}{\partial\theta} + \frac{\partial^2\mathscr{L}(\theta^+, P^+)}{\partial\theta\partial\theta'}(\widehat{\theta}_{NPL} - \theta_0) + \frac{\partial^2\mathscr{L}(\theta^+, P^+)}{\partial\theta\partial P'}\left[I - \left(\frac{\partial\Gamma(\theta^-; P^-)}{\partial P}\right)'\right]^{-1}\frac{\partial\Gamma(\theta^-; P^-)}{\partial\theta}(\widehat{\theta}_{NPL} - \theta_0) = 0, \tag{A.5}$$

From Eq. (A.5), we see that

$$\begin{aligned} &\left[\frac{\partial^2\mathscr{L}(\theta^+, P^+)}{\partial\theta\partial\theta'} + \frac{\partial^2\mathscr{L}(\theta^+, P^+)}{\partial\theta\partial P'}\left[I - \left(\frac{\partial\Gamma(\theta^-; P^-)}{\partial P}\right)'\right]^{-1}\frac{\partial\Gamma(\theta^-; P^-)}{\partial\theta}\right]\sqrt{n}(\widehat{\theta}_{NPL} - \theta_0) \\ &\qquad = -\frac{1}{\sqrt{n}}\sum_{i\in\mathscr{I}}\frac{\partial\mathscr{L}_i\left(\theta_0, P^*_{[n]}\right)}{\partial\theta}. \end{aligned} \tag{A.6}$$

By the consistency result of $\widehat{\theta}$ and $\widehat{P}$, we have $\frac{\partial^2 \mathscr{L}(\theta^+,P^+)}{\partial\theta\partial\theta'} + \frac{\partial^2 \mathscr{L}(\theta^+,P^+)}{\partial\theta\partial P'}\left[I - \left(\frac{\partial\Gamma(\theta-;P^-)}{\partial P}\right)'\right]^{-1}\frac{\partial\Gamma(\theta^-;P^-)}{\partial\theta} = \frac{\partial^2 \mathscr{L}(\theta_0,P^*_{[n]})}{\partial\theta\partial\theta'} + \frac{\partial^2 \mathscr{L}(\theta_0,P^*_{[n]})}{\partial\theta\partial P'}\left[I - \left(\frac{\partial\Gamma(\theta_0,P^*_{[n]})}{\partial P}\right)'\right]^{-1}\frac{\partial\Gamma(\theta_0,P^*_{[n]})}{\partial\theta} + o_p(1)$. Conditional on $\mathbb{I}_n$, the elements of $\frac{\partial^2 \mathscr{L}(\theta_0,P^*_{[n]})}{\partial\theta\partial\theta'} + \frac{\partial^2 \mathscr{L}(\theta_0,P^*_{[n]})}{\partial\theta\partial P'}\left[I - \left(\frac{\partial\Gamma(\theta_0,P^*_{[n]})}{\partial P}\right)'\right]^{-1}\frac{\partial\Gamma(\theta_0,P^*_{[n]})}{\partial\theta}$ for each observation are independent and thus this term converges to its expectation. Conditional on $\mathbb{I}_n$, (Y_{1i}, Y_{2i}) in $\frac{\partial \mathscr{L}_i(\theta_0,P^*_{[n]})}{\partial\theta}$ are independent. Thus $\frac{\partial \mathscr{L}_i(\theta_0,P^*_{[n]})}{\partial\theta}$'s are conditionally independent. With the Donsker class assumption in Assumption 10(ii), we have all conditions of the martingale central limit theorem in Menzel (2016) satisfied, we thus have the asymptotic normality result as below

$$\sqrt{n}(\widehat{\theta}_{NPL} - \theta_0) \xrightarrow{d} N(0, V_{NPL}), \tag{A.7}$$

where

$$V_{NPL} = V_1^{-1}(\theta_0) \cdot V_2(\theta_0) \cdot V_1^{'-1}(\theta_0).$$

□

References

Abrevaya, J., Hausman, J.A., 1999. Semiparametric estimation with mismeasured dependent variables: an application to duration models for unemployment spells. Annales d'Economie et de Statistique 55/56, 243–275.
Aguirregabiria, V., Mira, P., 2002. Swapping the nested fixed point algorithm: a class of estimators for discrete Markov decision models. Econometrica 70 (4), 1519–1543.
Aguirregabiria, V., Mira, P., 2007. Sequential estimation of dynamic discrete games. Econometrica 75 (1), 1–53.
Anderson, A., Smith, L., Park, A., 2017. Rushes in large timing games. Econometrica 85 (3), 871–913.
Aradillas-Lopez, A., 2010. Semiparametric estimation of a simultaneous game with incomplete information. J. Econom. 157 (2), 409–431.
Aradillas-Lopez, A., 2012. Pairwise-difference estimation of incomplete information games. J. Econom. 168 (1), 120–140.
Arcidiacono, P., Bayer, P., Blevins, J.R., Ellickson, P.B., 2016. Estimation of dynamic discrete choice models in continuous time with an application to retail competition. Rev. Econ. Stud. 83 (3), 889–931.
Arcidiacono, P., Foster, G., Goodpaster, N., Kinsler, J., 2012. Estimating spillovers using panel data, with an application to the classrooms. Quant. Econ. 3 (3), 421–470.
Badev, A., 2021. Nash equilibria on (un) stable networks. Econometrica 89 (3), 1179–1206.
Bajari, P., Hong, H., Krainer, J., Nekipelov, D., 2010. Estimating static models of strategic interactions. J. Bus. Econ. Stat. 28 (4), 469–482.
Bandura, A., Walters, R.H., 1977. Social Learning Theory, Vol. 1. Prentice-hall Englewood Cliffs, NJ.
Blume, L.E., Brock, W.A., Durlauf, S.N., Ioannides, Y.M., 2011. Identification of social interactions. Handb. Soc. Econ. 1, 853–964.
Blume, L.E., Brock, W.A., Durlauf, S.N., Jayaraman, R., 2015. Linear social interactions models. J. Polit. Economy 123 (2), 444–496.
Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement error in survey data. Handb. Econ. 5, 3705–3843.
Bramoullé, Y., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. J. Econom. 150 (1), 41–55.
Bramoullé, Y., Kranton, R., 2016. Games played on networks. Oxf. Handb. Econ. Netw. 1–33.
Bramoullé, Y., Kranton, R., D'amours, M., 2014. Strategic interaction and networks. Am. Econ. Rev. 104 (3), 898–930.
Breza, E., Chandrasekhar, A.G., 2019. Social networks, reputation, and commitment: Evidence from a savings monitors experiment. Econometrica 87 (1), 175–216.
Brock, W.A., Durlauf, S.N., 2001. Discrete choice with social interactions. Rev. Econ. Stud. 5 (2), 235–260.
Brock, W.A., Durlauf, S.N., 2001. Interaction-based models. Handb. Econ. 5, 3297–3380.
Brock, W.A., Durlauf, S.N., 2007. Identification of binary choice models with social interactions. J. Econom. 140 (1), 52–75.
Bursztyn, L., Ederer, F., Ferman, B., Yuchtman, N., 2014. Understanding mechanisms underlying peer effects: evidence from a field experiment on financial decisions. Econometrica 82 (4), 1273–1301.
Bursztyn, L., Egorov, G., Jensen, R., 2019. Cool to be smart or smart to be cool? understanding peer pressure in education. Rev. Econ. Stud. 86 (4), 1487–1526.
Bursztyn, L., Jensen, R., 2015. How does peer pressure affect educational investments? Q. J. Econ. 130 (3), 1329–1367.
Bursztyn, L., Jensen, R., 2017. Social image and economic behavior in the field: identifying, understanding, and shaping social pressure. Annu. Rev. Econ. 9, 131–153.
Cai, H., Chen, Y., Fang, H., 2009. Observational learning: evidence from a randomized natural field experiment. Am. Econ. Rev. 99 (3), 864–882.
Calvó-Armengol, A., Patacchini, E., Zenou, Y., 2009. Peer effects and social networks in education. Rev. Econ. Stud. 76 (4), 1239–1267.
Cameron, A.C., Li, T., Trivedi, P.K., Zimmer, D.M., 2004. Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts. Econom. J. 7 (2), 566–584.
Canen, N., Schwartz, J., Song, K., 2017. Estimating local interactions among many agents who observe their neighbors. Working paper.
Card, D., Giuliano, L., 2013. Peer effects and multiple equilibria in the risky behavior of friends. Rev. Econ. Stat. 95 (4), 1130–1149.
Carroll, R.J., Chen, X., Hu, Y., 2010. Identification and estimation of nonlinear models using two samples with nonclassical measurement errors. J. Nonparam. Stat. 22 (4), 379–399.
Carroll, R.J., Ruppert, D., Stefanski, L.A., Crainiceanu, C.M., 2006. Measurement Error in Nonlinear Models: A Modern Perspective. CRC Press.
Chandrasekhar, A. G., Jackson, M. O., 2016. A network formation model based on subgraphs. Working paper.
Chen, X., Hong, H., Nekipelov, D., 2011. Nonlinear models of measurement errors. J. Econ. Lit. 49 (4), 901–937.
Chen, X., Hong, H., Tamer, E., 2005. Measurement error models with auxiliary data. Rev. Econ. Stud. 72 (2), 343–366.
Christakis, N.A., Fowler, J.H., Imbens, G.W., Kalyanaraman, K., 2010. An Empirical Model for Strategic Network Formation. Technical Report. National Bureau of Economic Research.
Ciliberto, F., Tamer, E., 2009. Market structure and multiple equilibria in airline markets. Econometrica 77 (6), 1791–1828.
Copas, J.B., 1988. Binary regression models for contaminated data. J. R. Stat. Soc. Ser. B (Methodological) 50 (2), 225–253.
Dahl, G.B., Løken, K.V., Mogstad, M., 2014. Peer effects in program participation. Am. Econ. Rev. 104 (7), 2049–2074.
De Giorgi, G., Pellizzari, M., 2013. Understanding social interactions: Evidence from the classroom. Econ. J. 124 (579), 917–953.
De Nadai, M., Lewbel, A., 2016. Nonparametric errors in variables models with measurement errors on both sides of the equation. J. Econom. 191, 19–32.
de Paula, A., 2013. Econometric analysis of games with multiple equilibria. Annu. Rev. Econ. 5, 107–131.
de Paula, A., 2017. Econometrics of network models. Adv. Econ. Econom. 1, 268–323.
de Paula, A., Richards-Shubik, S., Tamer, E., 2018. Identifying preferences in networks with bounded degree. Econometrica 86 (1), 263–288.
de Paula, A., Tang, X., 2012. Inference of signs of interaction effects in simultaneous games with incomplete information. Econometrica 80 (1), 143–172.

Durlauf, S.N., Ioannides, Y.M., 2010. Social interactions. Annu. Rev. Econ. 2 (1), 451–478.
Epple, D., Romano, R., 2011. Peer effects in education: a survey of the theory and evidence. Handb. Soc. Econ. 1 (11), 1053–1163.
Eraslan, H., Tang, X., 2017. Identification and estimation of large network games with private link information. Working paper.
Fruehwirth, J.C., 2014. Can achievement peer effect estimates inform policy? A view from inside the black box. Rev. Econ. Stat. 96 (3), 514–523.
Gilchrist, D.S., Sands, E.G., 2016. Something to talk about: Social spillovers in movie consumption. J. Polit. Economy 124 (5), 1339–1382.
Gladwell, M., 2000. The Tipping Point: How Little Things Can Make a Big Difference. Little, Brown.
Glaeser, E., Scheinkman, J.A., 2003. Nonmarket interactions. Adv. Econ. Econom. 1, 339–369.
Goldsmith-Pinkham, P., Imbens, G.W., 2013. Social networks and the identification of peer effects. J. Bus. Econ. Stat. 31 (3), 253–264.
Graham, B.S., 2008. Identifying social interactions through conditional variance restrictions. Econometrica 76 (3), 643–660.
Graham, B.S., 2015. Methods of identification in social networks. Annu. Rev. Econ. 7 (1), 465–485.
Graham, B.S., 2016. Homophily and Transitivity in Dynamic Network Formation. Technical Report. National Bureau of Economic Research.
Graham, B.S., 2017. An econometric model of network formation with degree heterogeneity. Econometrica 85 (4), 1033–1063.
Granovetter, M., 1978. Threshold models of collective behavior. Am. J. Sociol. 83 (6), 1420–1443.
Greene, W.H., Hensher, D.A., 2010. Modeling Ordered Choices: A Primer. Cambridge University Press.
Harris, K. M., Halpern, C. T., Whitsel, E. A., Hussey, J. M., Tabor, J., Entzel, P., Udry, J. R., 2009. The national longitudinal study of adolescent to adult health: research design. http://www.cpc.unc.edu/projects/addhealth/design.
Hausman, J., Abrevaya, J., Scott-Morton, F.M., 1998. Misclassification of the dependent variable in a discrete-response setting. J. Econom. 87 (2), 239–269.
Hausman, J.A., 2001. Mismeasured variables in econometric analysis: problems from the right and problems from the left. J. Econ. Perspect. 15 (4), 57–67.
Horst, U., Scheinkman, J.A., 2006. Equilibria in systems of social interactions. J. Econ. Theory 130, 44–77.
Horst, U., Scheinkman, J.A., 2009. A limit theorem for systems of social interactions. J. Math. Econ. 45 (9), 609–623.
Hoshino, T., 2019. Two-step estimation of incomplete information social interaction models with sample selection. J. Bus. Econ. Stat. 37 (4), 598–612.
Hoxby, C., 2000. Peer Effects in the Classroom: Learning from Gender and Race Variation. Technical Report. National Bureau of Economic Research.
Hsiao, C., Sun, B.-H., 1998. Modeling survey response bias–with an analysis of the demand for an advanced electronic device. J. Econom. 89 (1), 15–39.
Hu, Y., 2008. Identification and estimation of nonlinear models with misclassification error using instrumental variables: a general solution. J. Econom. 144 (1), 27–61.
Hu, Y., 2017. The econometrics of unobservables: applications of measurement error models in empirical industrial organization and labor economics. J. Econom. 200 (2), 154–168.
Hu, Y., Schennach, S.M., 2008. Instrumental variable treatment of nonclassical measurement error models. Econometrica 76 (1), 195–216.
Ioannides, Y., 2006. Topologies of social interactions. Econ. Theory 28, 559–584.
Jackson, M. O., Lin, Z., Yu, N., 2020. Adjusting for peer-influence in propensity scoring when estimating treatment effects. Working paper.
Johnsson, I., Moon, H.R., 2021. Estimation of peer effects in endogenous social networks: Control function approach. Rev. Econ. Stat. 103 (2), 328–345.
Kasahara, H., Shimotsu, K., 2012. Sequential estimation of structural models with a fixed point constraint. Econometrica 80 (5), 2303–2319.
Lee, L., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. J. Econom. 140 (2), 333–374.
Lee, L., Li, J., Lin, X., 2014. Binary choice models with social network under heterogeneous rational expectations. Rev. Econ. Stat. 96 (3), 402–417.
Lee, L., Liu, X., Lin, X., 2010. Specification and estimation of social interaction models with network structures. Econom. J. 13 (2), 145–176.
Leung, M., 2015. A random-field approach to inference in large models of network formation. Working paper.
Lewbel, A., 1996. Demand estimation with expenditure measurement errors on the left and right hand side. Rev. Econ. Stat. 78, 718–725.
Lewbel, A., 2000. Identification of the binary choice model with misclassification. Econom. Theory 16 (04), 603–609.
Li, T., 2002. Robust and consistent estimation of nonlinear errors-in-variables models. J. Econom. 110 (1), 1–26.
Li, T., Hsiao, C., 2004. Robust estimation of generalized linear models with measurement errors. J. Econom. 118 (1), 51–65.
Li, T., Trivedi, P.K., Guo, J., 2003. Modeling response bias in count: a structural approach with an application to the national crime victimization survey data. Sociol. Methods Res. 31 (4), 514–544.
Li, T., Vuong, Q., 1998. Nonparametric estimation of the measurement error model using multiple indicators. J. Multivar. Anal. 65 (2), 139–165.
Li, T., Zhao, L., 2016. A partial identification subnetwork approach to discrete games in large networks: an application to quantifying peer effects. Working Paper.
Lin, X., 2010. Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. J. Lab. Econ. 28 (4), 825–860.
Lin, Z., 2021. Quantile effects in discrete choice with social interactions. Working paper.
Lin, Z., Tang, X., 2021. Solving the reflection problem in social interaction models with endogeneity. Working Paper.
Lin, Z., Xu, H., 2017. Estimation of social-influence-dependent peer pressures in a large network game. Econom. J. 20 (3), 86–102.
Liu, X., 2019. Simultaneous equations with binary outcomes and social interactions. Econom. Rev. 38 (8), 921–937.
Liu, X., Lee, L., 2010. GMM estimation of social interaction models with centrality. J. Econom. 159 (1), 99–115.
Mahajan, A., 2006. Identification and estimation of regression models with misclassification. Econometrica 74 (3), 631–665.
Manski, C.F., 1993. Identification of endogenous social effects: the reflection problem. Rev. Econ. Stud. 60 (3), 531–542.
Manski, C.F., 2000. Economic analysis of social interactions. J. Econ. Perspect. 14 (3), 115–136.
McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. Front. Econom. 105–142.
Mele, A., 2017. A structural model of dense network formation. Econometrica 85 (3), 825–850.
Menzel, K., 2016. Inference for games with many players. Rev. Econ. Stud. 83 (1), 306–337.
Menzel, K., 2017. Strategic network formation with many agents. Working paper.
Meyer, B.D., Mittag, N., 2017. Misclassification in binary choice models. J. Econom. 200 (2), 295–311.
Moretti, E., 2011. Social learning and peer effects in consumption: evidence from movie sales. Rev. Econ. Stud. 78 (1), 356–393.
Morris, S., Shin, H.S., 2003. Global games: theory and applications. Adv. Econ. Econom. 1, 56–114.
Neuhaus, J.M., 2002. Analysis of clustered and longitudinal binary data subject to response misclassification. Biometrics 58 (3), 675–683.
Newey, W.K., McFadden, D., 1994. Large sample estimation and hypothesis testing. Handb. Econ. 4, 2111–2245.
Park, A., Smith, L., 2008. Caller number five and related timing games. Theor. Econ. 3 (2), 231–256.
Pesendorfer, M., Schmidt-Dengler, P., 2008. Asymptotic least squares estimators for dynamic games. Rev. Econ. Stud. 75 (3), 901–928.
Ramalho, E.A., 2002. Regression models for choice-based samples with misclassification in the response variable. J. Econom. 106 (1), 171–201.
Ridder, G., Moffitt, R., 2007. The econometrics of data combination. Handb. Econ. 6, 5469–5547.
Rust, J., 1987. Optimal replacement of GMC bus engines: An empirical model of harold zurcher. Econometrica 55 (5), 999–1033.
Sacerdote, B., 2011. Peer effects in education: How might they work, how big are they and how much do we know thus far? Handb. Econ. Educ. 3 (3), 249–277.
Schelling, T.C., 1971. Dynamic models of segregation. J. Math. Sociol. 1 (2), 143–186.
Schennach, S.M., 2004. Estimation of nonlinear models with measurement error. Econometrica 72 (1), 33–75.
Schennach, S.M., 2016. Recent advances in the measurement error literature. Annu. Rev. Econ. 8 (1), 341–377.
Seim, K., 2006. An empirical model of firm entry with endogenous product-type choices. RAND J. Econ. 37 (3), 619–640.
Sheng, S., 2020. A structural econometric analysis of network formation games through subnetworks. Econometrica 88 (5), 1829–1858.
Song, K., 2014. Econometric inference on a large Bayesian game. Working Paper.
Sweeting, A., 2009. The strategic timing incentives of commercial radio stations: An empirical analysis using multiple equilibria. RAND J. Econ. 40 (4), 710–742.
Tamer, E., 2003. Incomplete simultaneous discrete response model with multiple equilibria. Rev. Econ. Stud. 70 (1), 147–165.
Tamer, E., 2010. Partial identification in econometrics. Annu. Rev. Econ. 2 (1), 167–195.
Tang, X., 2010. Estimating simultaneous games with incomplete information under median restrictions. Econ. Lett. 108 (3), 273–276.
Train, K.E., 2009. Discrete Choice Methods with Simulation. Cambridge University Press.

Xu, H., 2014. Estimation of discrete games with correlated types. Econom. J. 17 (3), 241–270.
Xu, H., 2018. Social interactions on large networks: a game theoretic approach. Int. Econ. Rev. 59 (1), 257–284.
Yang, C., Lee, L., 2017. Social interactions under incomplete information with heterogeneous expectations. J. Econom. 198 (1), 65–83.
Zimmerman, D.J., 2003. Peer effects in academic outcomes: evidence from a natural experiment. Rev. Econ. Stat. 85 (1), 9–23.