Misclassification and the Hidden Silent Rivalry

Yingyao Hu* — Johns Hopkins University
Zhongjian Lin† — Emory University

November 1, 2018

Abstract

The interaction of economic agents is one of the most important elements in economic analyses. While peer effects on subjective outcomes, behavior or decisions, are inherently difficult to identify and estimate because these variables are prone to misclassification errors. In this paper, we propose a binary choice model with misclassification and social interactions to rectify the misclassification problems in peer effects studies. We achieve identification of the model by the tool of repeated measurements and propose nested pseudo likelihood algorithm for estimation. We bring the model to estimate the peer effects among students on attitudes towards learning (silent rivalry). Peer effects on students' attitudes towards learning are believed to have a significant impact on their achievements, while we find that these peer effects are distorted by the misclassification error. Our estimates suggest that peer effects are not only significant, but also much larger than estimates ignoring the misreporting errors and a significant proportion of students overreport their attitudes towards learning.

Keywords: Misclassification, Binary Choice, Peer Effects, Nested Pseudo Likelihood, Attitude Towards Learning, Social Desirability.
JEL Classifications: C25; C57; C63; I20.

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 data as accurately measured while decisions data usually suffer from measurement error in the surveys. With mismeasured decision variables, the simultaneity of strategic interactions study naturally raises the problems from the left and the problems from the right [Hausman (2001)]. In this paper, we propose a

*Department of Economics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218. Email: yhu@jhu.edu.
†Department of Economics, Emory University, 1602 Fishburne Drive, Atlanta, GA 30322, Email: zhongjian.lin@emory.edu.

binary choice model with misclassification and social interactions and bring the model to analyze the peer effects among students on attitude towards learning. We rectify the biases due to misclassification by using tool with repeated measurements. We find significant overreporting in attitude and recover the hidden silent rivalry of students on attitude towards learning.

Measurement errors prevail in survey data for economics analyses. There are four sources of measurement errors: mistakes 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, Brown, and Mathiowetz (2001) for details. Decision variables possess discreteness and sometimes are dichotomous. Discrete measurement error is also called 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 error in the nonlinear models, see Chen, Hong, and Nekipelov (2011); Schennach (2016); Hu (2017) and reference therein for details. In the peer effects on attitude, the self-reported attitudes of students are prone to misclassification errors as some answers to attitude questions are social desirable, e.g. "hard-working".

Major development of misclassification is on the right hand side with few exceptions, e.g. see Hausman, Abrevaya, and Scott-Morton (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, Trivedi, and Guo (2003) for count models. In the continuous setting, Lewbel (1996); De Nadai and Lewbel (2016) investigate the measurement errors on both sides of regression. Unlike in the linear models where measurement errors on the left hand side cause only efficiency loss, there is sizable distortion of econometric analysis of nonlinear models with measurement errors on the dependent variable. This paper attempts to study a case where there are misclassification errors on both sides due to the simultaneity of strategic interactions.

In the last three decades, tremendous attention was paid to social interactions and peer effects among individuals in many fields, e.g. education, production adoption, information diffusion, word of mouth, etc. Brock and Durlauf (2001a,b) pioneer the discrete choice analysis with social interactions. The nonlinear model of discrete choice avoids the reflection problem raised by Manski (1993), see Brock and Durlauf (2007). Brock and Durlauf (2001a, 2007) provide 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 intearctions model, see Durlauf and Ioannides (2010); Blume, Brock, Durlauf, and Ioannides (2011); Blume, Brock, Durlauf, and Jayaraman

(2015).

The binary choice model with misclassification and social interactions is modeled through a static game played on an exogenously given large social network. Exogenous network setting prevails in peer effects study, either in the linear-in-mean model, see Manski (1993, 2000); Lee (2007); Graham (2008); Bramoullé, Djebbari, and Fortin (2009); Calvó-Armengol, Patacchini, and Zenou (2009); Lee, Liu, and Lin (2010); Lin (2010); Liu and Lee (2010); Goldsmith-Pinkham and Imbens (2013); Bramoullé, Kranton, and D'amours (2014); Dahl, Løken, and Mogstad (2014); Blume, Brock, Durlauf, and Jayaraman (2015); Eraslan and Tang (2017); Hoshino (2017) to name only a few, or in the discrete choice with social interactions, e.g. Brock and Durlauf (2001a, 2007); Card and Giuliano (2013); Lee, Li, and Lin (2014); Song (2014); Menzel (2015); Li and Zhao (2016); Canen, Schwartz, and Song (2017); Lin and Xu (2017); Liu (2017); Yang and Lee (2017); Xu (2018) to mention but a few. There is a growing literature on the econometrics of dynamic network formation, e.g. Christakis, Fowler, Imbens, and Kalyanaraman (2010); Graham (2015, 2016, 2017); Leung (2015); Menzel (2017); Chandrasekhar and Jackson (2016); Badev (2017); Mele (2017); de Paula, Richards-Shubik, and Tamer (2017); Sheng (2017). We focus on the static game played on exogenous network and and does 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, the conditional distribution of latent true decision variable through the technique of the two repeated measurements, see Hu (2008, 2017). We extend the likelihood-like algorithm (nested pseudo likelihood (NPL) estimation) from Aguirregabiria and Mira (2007) (dynamic game) and Lin and Xu (2017) (social interactions) to our model with a homogeneous misclassification condition. We establish the asymptotic properties of the NPL estimator and illustrate its finite sample performance with eight Monte Carlo experiments.

We bring our model to estimate the peer effects of students on attitudes towards learning. In a school, peer effects on students' attitudes towards learning are believed to have a significant impact on their achievements. Extant empirical studies pay much attention to the peer effects on the final achievements with implicitly assumed production function while the inquiry on students' attitudes is underdeveloped. In this paper, we aim to bridge this gap by investigating the peer effects on attitude towards learning. We provide empirical evidence on the presence of misclassification errors in students' self-reported attitudes and correct such misreporting errors for estimating the peer effects on attitude. We denote such peer effects as silent rivalry as students strive in a silent manner. Our estimates show that a significant proportion of students overreport their attitudes towards learning and that peer effects are not only significant, but also much larger than estimates ignoring the misreporting errors. These stronger peer effects elaborate the statement in the Coleman report of 1966 that it is more well-grounded to improve school performance through manipulation of peer group influence than by increased per student expenditures [Coleman et al. (1966)].

The paper unfolds as follows. Section 2 introduces the binary choice model with misclassification and social interactions. Section 3 provides the theoretical results on the identification of the conditional distribution of latent variable and the structural parameter. We then demonstrate the nested pseudo likelihood (NPL) estimation strategy in Section 4. Eight Monte Carlo experiments are conducted in Section 5 to illustrate the finite sample performance of the model and the NPL algorithm. Section 6 presents our main empirical results on the silent rivalry among high school students on attitude towards learning. The last section concludes. Proofs are rendered in the Appendix A. We also check the robustness of our estimation with different discretized definition of positive attitude towards learning in Appendix B.

2 Binary Choice Model with Misclassification and Social Interactions

There are n individuals, $\mathcal{I} = \{1, \cdots, n\}$, 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 $F_i = \{j \in \mathcal{I} : F_{ij} = 1\}$. Denote N_i as the number of friends of individual i.

Individuals make binary decisions $\{Y_i^* \in= \{0,1\}\}_{i\in\mathcal{I}}$ simultaneously. The interactions transit through the directed link, F_{ij}, which means that individuals take into account the actions 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 silent rivalry study, $Y_i^* = 1$ means that student i chooses to work hard. Here we use Y_i^* to denote the true latent decision of individual i. We will use Y_i and Z_i for the reported measurements of the latent decision. Following the standard binary choice literature, e.g. 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

$$U_i(Y_{-i}^*, X_i, F_i, \varepsilon_i) = X_i^T \beta + \frac{\gamma}{N_i} \sum_{j \in F_i} Y_j^* - \varepsilon_i, \tag{1}$$

where superscript T stands for the matrix transpose, $X_i \in \mathcal{X}$ is a $d \times 1$ vector representing the demographic characteristics[1], Y_{-i}^* are the decisions of others, and ε_i is the

[1] X_i contains intercept. We are considering a single large network, therefore the characteristics of

private utility shock. The utility of individual i has three components: the deterministic part from demographics, $X_i^T\beta$; the deterministic social utility from the average choice of friends (peer effects), $\frac{\gamma}{N_i}\sum_{j\in F_i} Y_j^*$, and a private utility shock, ε_i. γ captures the peer effects from friends. Denote $\mu = (\beta^T, \gamma)^T$.

To complete the setting for the model, we further specify the information structure: let $\mathcal{W} = (\{X_i\}_{i\in\mathcal{I}}, \{F_i\}_{i\in\mathcal{I}})$ be the public information set including all demographic characteristics and friendship information[2]. The private utility shock ε_i is only known to individual i. Therefore we consider an incomplete information structure in the Bayesian Nash game and individuals form beliefs on the actions of their friends[3].The decision rule is:

$$Y_i^* = \mathbf{1}\left\{X_i^T\beta + \frac{\gamma}{N_i}\sum_{j\in F_i}\mathbb{E}(Y_j^*|\mathcal{W},\varepsilon_i) - \varepsilon_i \geq 0\right\}, \tag{2}$$

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

2.1 Bayesian Nash Equilibrium

With an incomplete information structure, we consider the Bayesian Nash equilibrium (BNE) of the Bayesian Nash 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.*

Remark 1. *Assumption 1 is fairly standard in the discrete choice model literature, e.g. Bajari, Hong, Krainer, and Nekipelov (2010). As a matter of fact, Assumption 1 provides*

the network itself is constant for all individuals and absorbed in the intercept term. Our model can also be brought to multiple networks which could include the network characteristics as there is variation across networks. The identification strategy is similar as the one using between-group variation in linear-in-mean models, see Graham (2008).

[2]The usage of all demographics and friendship as public information is for the tractability of the equilibrium as we will see below. This can be approximated by information from subnetwork if we have weak dependence. Xu (2018) establishes such weak dependence (network decaying dependence property) of the discrete game that the conditional probabilities based on the whole network can be well approximated by the counterpart calculated based on subnetwork, e.g. the one with individual, friends and the friends of friends.

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

a closed-form expression for individuals' conditional choice probabilities in terms of choice probabilities and streamlines the belief term, i.e., $\mathbb{E}(Y_j|\mathcal{W},\varepsilon_i) = \mathbb{E}(Y_j|\mathcal{W})$.

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

$$\sigma_i^*(\mathcal{W};\mu) = \Lambda\left[X_i^T\beta + \frac{\gamma}{N_i}\sum_{j\in F_i}\sigma_j^*(\mathcal{W};\mu)\right] \equiv \Gamma_i(\mathcal{W},\sigma^*;\mu), i \in \mathcal{I}, \tag{3}$$

where $\sigma^* = [\sigma_1^*(\mathcal{W};\mu),...,\sigma_n^*(\mathcal{W};\mu)]^T$ is the equilibrium choice probabilities profile . Equation (3) is a simultaneous system of equations of σ^*. Let P be an arbitrary choice probabilities profile. The equilibrium choice probability profile σ^* defined in Equation (3) is then a fixed point of

$$\Gamma(\mathcal{W},P;\mu) \equiv (\Gamma_1(\mathcal{W},P;\mu),\cdots,\Gamma_n(\mathcal{W},P;\mu))^T = P.$$

To obtain the uniqueness of the BNE, we make following assumptions

Assumption 2. *There is an upper bound,* $M > 0$*, for the number of friends, i.e.* $N_i \le M$ *for* $i \in \mathcal{I}$.

Remark 2. *Assumption 2 excludes some specific networks, e.g. star network. In the social networks, it is feasible to limit the number of friends as the human beings do not have infinite efforts to maintain too many friendships. In the Add Health dataset,* $M = 10$ *by the design of the survey. Assumption 2 leads to sparse network in our model as n increases.*

Assumption 3. *The strength of interactions is moderate, i.e.* $0 < \gamma < 4$.

Remark 3. *In the literature concerning interaction games, a similar assumption is denoted as Moderate social influence (MSI) condition for uniqueness, see Glaeser and Scheinkman (2003); Horst and Scheinkman (2006, 2009). In the literature of discrete choice with social interactions, Brock and Durlauf (2001a,b); Lin and Xu (2017); Liu (2017); Xu (2018) employ a similar condition to characterize the uniqueness of the BNE in the Bayesian Nash game. Assumption 3 restricts the size of dependence along individuals' decisions. 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 dimension and our social interactions analysis is multiple-dimensional that each friend of a individual provides one dimension. Similar as the existence of explosive time series, there are exceptions with dominant peer effects, e.g. tipping (Schelling (1971); Granovetter (1978); Gladwell (2000)); rush into the market (Park and Smith (2008); Anderson, Smith, and Park (2017)). In the silent rivalry study in Section 6, this condition is feasible that peer effects would not be dominant. There is also literature to work with multiple equilibria with* *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, 2010); de Paula (2013). The upper bound, 4 is from the Logistic distribution of the private utility terms. For standard normal distribution, Probit type assumption, we should change the upper bound to $\sqrt{2\pi}$. In general, $0 < \gamma < 1/\sup f_\varepsilon(\cdot)$ 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 Nash game represented in Equation (3).*

Proof. See Appendix A. □

Lemma 1 establishes the uniqueness of the Baysian Nash equilibrium. The uniqueness ensures that the conditional distributions of repeated measurements are identified from the data[4]. Other option for the equilibrium characterization is to assume that the data comes from one single equilibrium, see Bajari, Hong, Krainer, and Nekipelov (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 Nash game builds on the binary latent decisions $\{Y_i^*\}_{i\in\mathcal{I}}$ which are prone to measurement errors. de Paula (2017) points out the importance of measurement error issue in the 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 instrumental variable for identification of nonparametric model with 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, Abrevaya, and

[4]The identification strategy is similar as the identification in time series. We identified through the observations with same demographic characteristics and network feature (position). While this is identification in infinity. Xu (2018) establishes network decaying dependence condition for feasible inference of the network analysis. The idea is to use subnetwork results to approximate the whole network results mimicking the strategy in the time series, e.g. Markovian property.

Scott-Morton (1998) propose a partial maximum likelihood estimator to handle misclassified response variable. In this paper, we resort to repeated measurements for identification and estimation.

3 Identification

In this paper, we adopt repeated measurements to identify the true conditional distribution of the latent response variable and the structural parameter. Let $\{Y_i\}_{i\in\mathcal{I}}$ and $\{Z_i\}_{i\in\mathcal{I}}$ be two observed measurements of the latent decisions. For instance, in the silent rivalry study in Section 6, there are two repeated measurements of students' attitude in the *Add Health* data regarding the question "Skipped school without an excuse" which was asked in both the in-school and at-home surveys. We adopt the two measurements technique based on the social desirability feature of the survey question[5]. We denote two misclassification types: desired misclassification that individuals overreport from a null latent attitude, i.e. $Y = 1$ or $Z = 1$ when $Y^* = 0$; and evasive misclassification that individuals underreport a positive attitude, i.e. $Y = 0$ or $Z = 0$ when $Y^* = 1$[6]. We establish the identification of the conditional distribution of the latent decisions, $P(Y^*|\mathcal{W})$, through the LU decomposition with a condition on the evasive misclassification and propose an estimator based on the complete likelihood function drawing on both Y and Z. We introduce the following two assumptions for our closed-form identification.

Assumption 4. *(Y,Z) are jointly independent conditional on Y^* and $\mathcal{W}$,*

$$Y \perp Z \mid (Y^*,\mathcal{W}). \tag{4}$$

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

Assumption 5. *Individuals do not underreport their positive attitudes, i.e.,*

$$P(Y = 0|Y^* = 1,\mathcal{W}) = P(Z = 0|Y^* = 1,\mathcal{W}) = 0.$$

[5]The social desirability leads to additional condition, accompanying with two repeated measurements for identification. While social desirability can be relaxed if we could draw a third measurements. To fit our study on silent rivalry, we take two repeated measurements here and illustrate the social desirability in following assumption.

[6]For notational simplicity, we suppress the subscription in this section.

Remark 5. *Assumption 5 states that there are zero evasive misclassification probabilities. This assumption builds on the social desirability feature of survey data*[7]*. In data collection, some socially and personally sensitive questions are often asked. It is well documented that such questions provoke patterns of underreporting (for socially undesirable behavior and attitudes) as well as overreporting (for socially desirable behaviors and attitudes), see Bound, Brown, and Mathiowetz (2001). Assumption 5 can be further dropped with a third repeated measurement*[8]*, see Hu (2017) for review of the 3-measurement model. Assumption 5 implies* $P(Y=1|Y^*=1,\mathcal{W}) = P(Z=1|Y^*=1,\mathcal{W}) = 1$.

3.1 A Closed-form Identification

In this section, we establish a closed-form identification result for the conditional probabilities of the latent decisions, i.e., $P(Y^*=1|\mathcal{W})$[9]. Identification is about the recovery of $P(Y^*=1|\mathcal{W})$ uniquely from observables, i.e., $P(Y=1|\mathcal{W})$, $P(Z=1|\mathcal{W})$ and $P(Y,Z|\mathcal{W})$.

We define

$$
\begin{aligned}
M_{Y,Z|\mathcal{W}} &\equiv \begin{pmatrix} P(Y=0,Z=0|\mathcal{W}) & P(Y=0,Z=1|\mathcal{W}) \\ P(Y=1,Z=0|\mathcal{W}) & P(Y=1,Z=1|\mathcal{W}) \end{pmatrix}, \\
&\equiv \Big[P(Y=i-1,Z=j-1|\mathcal{W})\Big]_{i,j=1}^{2}.
\end{aligned}
$$

Similarly, we define $M_{Y|Y^*,\mathcal{W}} = [P(Y=i-1|Y^*=j-1,\mathcal{W})]_{i,j}$, $M_{Z|Y^*,\mathcal{W}} = [P(Z=i-1|Y^*=j-1,\mathcal{W}]_{i,j}$, $M_{Y,Y^*|\mathcal{W}} = [P(Y=i-1,Y^*=j-1|\mathcal{W})]_{i,j}$ and $M_{Z,Y^*|\mathcal{W}} = [P(Z=i-1,Y^*=j-1|\mathcal{W})]_{i,j}$. These are lower triangular matrices by Assumption 5. Denote

$$
D_{Y^*|\mathcal{W}} \equiv \begin{pmatrix} P(Y^*=0|\mathcal{W}) & 0 \\ 0 & P(Y^*=1|\mathcal{W}) \end{pmatrix}.
$$

Theorem 1. *With Assumptions 1 to 5, we identify the conditional distribution of the latent variable, i.e.,* $D_{Y^*|\mathcal{W}}$.

Proof. With Assumptions 1 to 3, $M_{Y,Z|\mathcal{W}}$ is identified from the data. By law of total

[7] We take social desirability as "social norm" rather than outcome of peer effects as desirability is a longrun stable standard. Therefore we suppress the potential peer effects on whether overreport. Empirically, we never observe the true latent variable and therefore do not have data on overreport behavior which prevents us to conduct inference on peer effects in overreporting.

[8] In empirical analyses, there are seldom three repeated measurements of binary outcomes, e.g. same questions asked three times in different venues. Therefore, in this paper, we adopt above assumption to fit the data structure in our empirical study.

[9] Because we are considering the binary case, $P(Y^*=1|\mathcal{W})$ fully characterize the conditional distribution, i.e., $P(Y^*=0|\mathcal{W}) = 1 - P(Y^*=1|\mathcal{W})$.

probability and Assumption 4, we obtain

$$M_{Y,Z|\mathcal{W}} = M_{Y|Y^*,\mathcal{W}} \times M^T_{Z,Y^*|\mathcal{W}} = M_{Z|Y^*,\mathcal{W}} \times M^T_{Y,Y^*|\mathcal{W}}, \tag{5}$$

$$M_{Y,Z|\mathcal{W}} = M_{Y|Y^*,\mathcal{W}} \times D_{Y^*|\mathcal{W}} \times M^T_{Z|Y^*,\mathcal{W}}. \tag{6}$$

With condition on the evasive misclassification probabilities (Assumption 5), $M_{Y|Y^*,\mathcal{W}}$ and $M^T_{Z,Y^*|\mathcal{W}}$ are lower and upper triangular matrices, respectively. The point identification of these two unknown matrices is feasible through the so-called LU decomposition. One can show that such a decomposition is unique given that each column sum of $M_{Y|Y^*,\mathcal{W}}$ equals one and that the sum of all the entries in $M^T_{Z,Y^*|\mathcal{W}}$ also equals one. Thus we have identified two misclassification matrices $M_{Y|Y^*,\mathcal{W}}$ and $M_{Z,Y^*|\mathcal{W}}$. Similarly we identify $M_{Z|Y^*,\mathcal{W}}$ and $M_{Y,Y^*|\mathcal{W}}$. For more details on LU decomposition, see Hu and Sasaki (2017). Then the conditional distribution of the latent variable is identified through:

$$D_{Y^*|\mathcal{W}} = M^{-1}_{Y|Y^*,\mathcal{W}} \cdot M_{Y,Z|\mathcal{W}} \cdot M^{T-1}_{Z|Y^*,\mathcal{W}}. \tag{7}$$

□

We then take $P(Y^* = 1|\mathcal{W})$ as known for the next step identification of the structural parameter.

3.2 Identification of the Structural Parameter, μ

For equilibrium presented in Equation (3), the identification of μ is standard in a constructive way. As shown in the first step identification, $P(Y_i^* = 1|\mathcal{W}), i \in \mathcal{I}$ is identified from the observables. From Equation (3), we have

$$\begin{aligned} \Xi(\mathcal{W}) &\equiv \log\Big[P(Y_i^* = 1|\mathcal{W})\Big] - \log\Big[P(Y_i^* = 0|\mathcal{W};\mu)\Big] \\ &= X_i^T\beta + \frac{\gamma}{N_i}\sum_{j\in F_i} P(Y_j^* = 1|\mathcal{W}), i \in \mathcal{I}. \end{aligned} \tag{8}$$

We make the following rank condition assumption to achieve identification.

Assumption 6. $\mathbb{E}\left[\left(X_i^T, \frac{\sum_{j\in F_i} P(Y_j^*=1|\mathcal{W})}{N_i}\right)^T \times \left(X_i^T, \frac{\sum_{j\in F_i} P(Y_j^*=1|\mathcal{W})}{N_i}\right)\right]$ *is with full rank $d+1$.*

Remark 6. *Assumption 6 requires no perfect collinearity of* $\left(X_i^T, \frac{\sum_{j\in F_i} P(Y_j^*=1|\mathcal{W})}{NF_i}\right)$. *This assumption is essentially a full rank condition. As is pointed out in Bajari, Hong, Krainer, and Nekipelov (2010), it is other individuals' exclusive payoff shifters that induce independent variations in individual i's beliefs, which render the rank condition meaningful. The BNE profile is determined through the fixed point and therefore implicitly by the $\mathcal{W}$ and the* *distribution of the ε. Furthermore, the variation of friends set which prevents the perfect collinearity/multiplicity problem between the peer effect covariate, i.e.* $\frac{\sum_{j\in F_i} P(Y_j^*=1|\mathcal{W})}{N_i}$, *and the demographic characteristics,* X_i.

With assumption 6, we have identified μ as

$$\begin{aligned}\mu = \mathbb{E}&\left[\left(X_i^T, \frac{\sum_{j\in F_i} P(Y_j^* = 1|\mathcal{W})}{N_i}\right)^T \times \left(X_i^T, \frac{\sum_{j\in F_i} P(Y_j^* = 1|\mathcal{W})}{N_i}\right)\right]^{-1} \\ &\times \mathbb{E}\left[\left(X_i^T, \frac{\sum_{j\in F_i} P(Y_j^* = 1|\mathcal{W})}{N_i}\right)^T \times \Xi(\mathcal{W})\right].\end{aligned} \tag{9}$$

4 Estimation Strategy

The identification in Section 3 is for the population and it takes $P(Y_i, Z_i|\mathcal{W})$ as identified from the observables. This is the identification at infinity in the nonparametric identification literature, see Matzkin (2007, 2013). However, the nonparametric estimation of the joint conditional distribution is infeasible due to the large dimension of $\mathcal{W}$. To avoid such problem, we adopt the sequential algorithm, the Nested Pseudo Likehood (NPL) estimation to estimate the structural parameter. The method is first introduced by Aguirregabiria and Mira (2002, 2007) for dynamic discrete games. Lin and Xu (2017) extend the method to social interactions studies. Before we proceed to the details of the NPL estimator, we make the following simplification assumption

Assumption 7. *The misclassification probabilities satisfy*

$$\begin{aligned}P(Y_i = 1|Y_i^* = 0, \mathcal{W}) = P(Y_i = 1|Y_i^* = 0) = \alpha \in (0,1),\\ P(Z_i = 1|Y_i^* = 0, \mathcal{W}) = P(Z_i = 1|Y_i^* = 0) = \delta \in (0,1).\end{aligned}$$

Remark 7. *Assumption 7 reduces the number of unknown 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 relex this assumption by parametrization over some observed covariates with richer data. Hausman, Abrevaya, and Scott-Morton (1998) make the same assumption when constructing the partial likelihood function. From Assumption 5, we have* $P(Y_i = 0|Y_i^* = 1, \mathcal{W}) = P(Z_i = 0|Y_i^* = 1, \mathcal{W}) = 0$, *therefore the monotone condition in Hausman, Abrevaya, and Scott-Morton (1998) is satisfied, i.e.,* $\alpha + 0 \in (0,1)$ *and* $\delta + 0 \in (0,1)$. *The exclusion of 0 or 1 probabilities is for identification since probabilities matrix in Section 3 would be singular without these exclusions. In the estimation of the misclassification parameters, we have 0 and 1 as the lower bound and upper bound and therefore the estimated misclassification probabilities can be arbitrarily close to 0 or 1.*

We now have the structural parameter, $\theta \equiv (\alpha, \delta, \mu^T)^T$ and

$$
\begin{aligned}
P(Y_i = 1|\mathcal{W};\theta) &= \alpha + (1-\alpha)P(Y_i^* = 1|\mathcal{W};\mu), \\
P(Z_i = 1|\mathcal{W};\theta) &= \delta + (1-\delta)P(Y_i^* = 1|\mathcal{W};\mu).
\end{aligned} \tag{10}
$$

Our log likelihood function is formulated by the observed conditional distribution function $f(Y_i, Z_i|\mathcal{W};\theta)$. Let $P^* = (P_1^*, P_2^*, \cdots, P_n^*) = \Big(P(Y_1^* = 1|\mathcal{W};\theta), P(Y_2^* = 1|\mathcal{W};\theta), \cdots, P(Y_n^* = 1|\mathcal{W};\theta)\Big)$. With Equation (10), we have the log-likelihood function[10]:

$$
\begin{aligned}
\mathcal{L}(\theta, P^*) = \sum_{i\in\mathcal{I}} \Bigg\{ & Y_i \log\Big[\alpha + (1-\alpha)P_i^*\Big] + (1-Y_i)\log\Big[1-\alpha-(1-\alpha)P_i^*\Big] \\
& + Z_i \log\Big[\delta + (1-\delta)P_i^*\Big] + (1-Z_i)\log\Big[1-\delta-(1-\delta)P_i^*\Big]\Bigg\}.
\end{aligned} \tag{11}
$$

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

$$
\hat{\theta}_{MLE} = \underset{\theta\in\Theta}{\arg\max}\ \mathcal{L}(\theta, P) \ \text{ s.t. } \ P = \Gamma(P, \mathcal{W};\mu). \tag{12}
$$

For small number of players, we can implement the MLE method by the nested fixed point (NFP) algorithm [Rust (1987)], which repeatedly solves all the fixed points of $P = \Gamma(P, \mathcal{W};\mu)$ 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 $\hat{\theta}$ with 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 NPL algorithm starts with an arbitrary choice probabilities profile, $P^{(0)}$, and the estimation of θ becomes a modified Logit regression, $\hat{\theta}^{(1)} = \underset{\theta\in\Theta}{\arg\max}\ \mathcal{L}(\theta, P^{(0)})$. The algorithm then updates the choice probabilities profile using $P^{(1)} = \Gamma\Big(\mathcal{W}, P^{(0)};\hat{\theta}^{(1)}\Big)$ defined in Equation (3) with the Logit estimate and the previous probabilities profile. The algorithm stops when the gap between the estimates in the consecutive two iterations is smaller than some preset tolerance value, i.e., $\Big|\hat{\theta}^{(K+1)} - \hat{\theta}^{(K)}\Big| < tol$. It is computationally feasible that we do not actually calculate the BNE choice probabilities profile but instead adopt a recursive method starting from an arbitrary probabilities value.

[10]Here we construct complete likelihood function based on both Y and Z. Hausman, Abrevaya, and Scott-Morton (1998) use either Y or Z to construct partial likelihood function.

4.1 Consistency and Asymptotic Normality of the NPL Estimator

Because the equilibrium choice probabilities profile is solved through iterated steps, in this section, we suppress the public information $\mathcal{W}$, i.e., $P = \Gamma(P;\theta)$[11]. We make similar assumptions as in Aguirregabiria and Mira (2007); Kasahara and Shimotsu (2012); Lin and Xu (2017). Define the pseudo log-likelihood function:

$$\begin{aligned}
\mathcal{L}(\theta,P) &= \frac{1}{n}\sum_{i\in\mathcal{I}}\mathcal{L}_i(\theta,P),\\
&\equiv \frac{1}{n}\sum_{i\in\mathcal{I}}\Bigg\{Y_i\log\Big[\alpha+(1-\alpha)P_i\Big]+(1-Y_i)\log\Big[1-\alpha-(1-\alpha)P_i\Big]\\
&\qquad + Z_i\log\Big[\delta+(1-\delta)P_i\Big]+(1-Z_i)\log\Big[1-\delta-(1-\delta)P_i\Big]\Bigg\},
\end{aligned}$$

where $P=(P_1,\cdots,P_n)$ is not necessarily the true equilibrium choice probabilities profile. Let

$$\begin{aligned}
\tilde{\theta}_n(P) &\equiv \arg\max_{\theta\in\Theta}\ \mathcal{L}(\theta,P),\\
\phi_n(P) &\equiv \Gamma(\tilde{\theta}_n(P),P),\\
\mathcal{L}_0(\theta,P) &\equiv \lim_{n\to\infty}\frac{1}{n}\mathbb{E}\Big[\sum_{i\in\mathcal{I}}\mathcal{L}_i(\theta,P)\Big],\\
\tilde{\theta}_0(P) &\equiv \arg\max_{\theta\in\Theta}\ \mathcal{L}_0(\theta,P),\\
\phi_0(P) &\equiv \Gamma(\tilde{\theta}_0(P),P).
\end{aligned}$$

Define the population NPL fixed points set as $\Lambda_0 \equiv \{(\theta,P)\in(\Theta,\mathcal{P}) : \theta=\tilde{\theta}(P), P=\phi_0(P)\}$ and the NPL fixed points set as $\Lambda_n \equiv \{(\theta,P)\in(\Theta,\mathcal{P}) : \theta=\tilde{\theta}_n(P), P=\phi_n(P)\}$. Let $\mathcal{N}$ denote a closed neighborhood of (θ_0,P^*). The first order condition for the NPL estimation is

$$\frac{\partial\mathcal{L}\Big(\theta,\Gamma(P;\theta)\Big)}{\partial\theta}\Bigg|_{(\theta,P)=(\hat{\theta}_{NPL},\hat{P}_{NPL})}=0. \tag{13}$$

Assumption 8. *(a) Θ is compact, θ_0 is an interior point of Θ, and $\mathcal{P}$ is a compact and convex subset of $(0,1)^n$; (b) (θ_0,P^*) is an isolated population NPL fixed point, i.e., it is unique, or else there is an open ball around it that does not contain any other element of Λ_0; (c) $\frac{\partial^2\mathcal{L}_0(\theta,P^*)}{\partial\theta\partial\theta^T}$ is a nonsingular matrix in $\mathcal{N}$; (d) The operator $\phi(P)-P$ has a nonsingular*

[11] Here we abuse the notation to use θ instead of μ though α and δ do not enter the conditional choice probabilities.

Jacobian matrix at P^*; (e) There exist non-singular matrices $V_1(\theta_0)$ and $V_2(\theta_0)$ such that

$$\mathbb{E}\left[\frac{\partial^2\mathcal{L}(\theta_0,P^*)}{\partial\theta\partial\theta^T}+\frac{\partial^2\mathcal{L}(\theta_0,P^*)}{\partial\theta\partial P^T}\cdot\left[I-\left(\frac{\partial\Gamma(P^*;\theta_0)}{\partial P}\right)^T\right]^{-1}\cdot\frac{\partial\Gamma(P^*;\theta_0)}{\partial\theta^T}\right]\xrightarrow{p}V_1(\theta_0),$$

$$\mathbb{E}\left[\frac{\partial\mathcal{L}(\theta_0,P^*)}{\partial\theta}\frac{\partial\mathcal{L}(\theta_0,P^*)}{\partial\theta^T}\right]\xrightarrow{p}V_2(\theta_0).$$

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

Remark 8. *Since we are assuming the logistic error term, we can check that $\tilde{\theta}_0(P)$ is a single-valued and continuous function of P in a neighborhood of P^*. Further, with the simultaneous equation system in Equation (3), we can easily verify that $\mathcal{L}_0(\theta,P)$ is globally concave in θ for $P\in\mathcal{N}$. Assumption 8 (e) is a high-level condition for non-singular limiting matrices as n goes to infinity. Such a condition could be derived by specifying a network growing mechanism. Moreover, the non-degeneracy of $V_1(\theta)$ and $V_2(\theta)$ requires that all the determinants of the finite counterparts are outside an open ball of zero for all n, which is essentially a rank condition.*

Theorem 2. *Suppose Assumptions 1-8 hold, we have $\hat{\theta}_{NPL}\xrightarrow{p}\theta_0$ and*

$$\sqrt{n}(\hat{\theta}_{NPL}-\theta_0)\xrightarrow{d}\mathcal{N}(0,V_{NPL}),\tag{14}$$

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

Proof. See Appendix A □

The local convergence of the NPL algorithm is ensured by the local contraction condition established in Kasahara and Shimotsu (2012) which is satisfied by our Lemma 1 with Assumptions 1 to 3. While in empirics, it is difficult to verify their conditions for the convergence of the NPL algorithm. It has been noticed in the literature, and we have experienced in our Monte Carlo experiments and empirical application, that the NPL algorithm typically converges to the same fixed point, regardless of the initial values; see Aguirregabiria and Mira (2007); Lin and Xu (2017).

5 Monte Carlo Experiments

The Monte Carlo experiments are designed to mimic the silent rivalry study in Section 6. We conduct eight 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 discrete distribution taking values from $\{-1,1\}$ with equal probability $\frac{1}{2}$. X_1,X_2 and X_3 have mean 0 and variance

1. We generate a random network with maximum number of friends as 10 (the same as in the *Add Health* dataset). The latent dependent variable is given by

$$Y_i^* = \mathbf{1}\left\{\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \frac{\gamma}{N_i}\sum_{j \in F_i} \mathbb{E}(Y_j^*|\mathcal{W}) - \varepsilon_i \geq 0\right\} \tag{15}$$

Observed measurements Y_i and Z_i are generated with misclassification probabilities, $(\alpha, \delta) = (0.05, 0.05), (0.1, 0.1), (0.2, 0.2), (0.4, 0.4)$. We generate 1,000 samples of pseudo-random numbers with $n \in \{500,\ 1,000,\ 2,000\}$. We denote $\hat{\theta}_{NPL}$ as the NPL estimates with misclassification correction and $\tilde{\theta}_{NPL}$ as the NPL estimates without misclassification correction, i.e. taking Y or Z as the truly observed binary decision. We report the average biases, standard deviations and the mean square errors in Tables 1 to 8. The NPL estimators with misclassification correction converge to the true parameter at the very nice rate, $\sqrt{n}$, while those without misclassification correction 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 The Hidden Silent Rivalry

Students live in two distinct social worlds: the hierarchical world with adults and the egalitarian world with peers. The former leads 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 sport clubs, conducting delinquent behaviors, etc. Among these spillovers, the peer effects in education has received considerable attention in the literature, see more details in Epple and Romano (2011) and Sacerdote (2011). When it comes to the learning spillover, scholars emphasize the achievements of students, e.g. Hoxby (2000); Zimmerman (2003); Calvó-Armengol, Patacchini, and Zenou (2009) to name only a few. However, in the context of the 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. 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. Proxy or IV approach is adopted to handle the unobserved ability in cross sectional setting. Arcidiacono, Foster, Goodpaster, and Kinsler (2012) treat ability as the unobserved heterogeneity in panel data model and remove this unobserved heterogeneity by standard approaches in panel data models with fixed effects.

Table 1: Experiment I

True Parameters: $\theta_0 = (0.05, 0.05; -1, 1, -1, 1; 1)$												
n	$\hat{\alpha}_{NPL}$	$\hat{\delta}_{NPL}$	$\hat{\beta}_{NPL}$				$\hat{\gamma}_{NPL}$	$\tilde{\beta}_{NPL}$				$\tilde{\gamma}_{NPL}$
500	0.002	0.001	-0.061	0.046	-0.033	0.041	0.049	0.229	-0.118	0.125	-0.117	-0.109
	(0.037)	(0.037)	(0.341)	(0.187)	(0.184)	(0.181)	(0.646)	(0.247)	(0.124)	(0.124)	(0.118)	(0.561)
1,000	-0.001	-0.001	-0.002	0.011	-0.013	0.007	-0.013	0.249	-0.128	0.123	-0.126	-0.144
	(0.027)	(0.026)	(0.228)	(0.124)	(0.117)	(0.118)	(0.448)	(0.179)	(0.090)	(0.083)	(0.081)	(0.411)
2,000	0.001	0.001	-0.013	0.012	-0.010	0.010	0.015	0.245	-0.128	0.128	-0.127	-0.126
	(0.020)	(0.019)	(0.160)	(0.087)	(0.087)	(0.082)	(0.314)	(0.123)	(0.062)	(0.061)	(0.058)	(0.282)
Mean Square Errors												
500	0.001	0.001	0.120	0.037	0.035	0.034	0.419	0.113	0.029	0.031	0.028	0.326
1,000	0.001	0.001	0.052	0.016	0.014	0.014	0.201	0.094	0.025	0.022	0.022	0.190
2,000	0.000	0.000	0.026	0.008	0.008	0.007	0.098	0.075	0.020	0.020	0.020	0.095

Table 2: Experiment II

True Parameters: $\theta_0 = (0.1, 0.1; -1, 1, -1, 1; 1)$												
n	$\hat{\alpha}_{NPL}$	$\hat{\delta}_{NPL}$	$\hat{\beta}_{NPL}$				$\hat{\gamma}_{NPL}$	$\tilde{\beta}_{NPL}$				$\tilde{\gamma}_{NPL}$
500	0.000	-0.001	-0.060	0.048	-0.035	0.041	0.046	0.441	-0.220	0.223	-0.217	-0.212
	(0.046)	(0.046)	(0.378)	(0.207)	(0.205)	(0.200)	(0.671)	(0.249)	(0.118)	(0.118)	(0.114)	(0.536)
1,000	-0.002	-0.002	-0.005	0.012	-0.015	0.007	-0.015	0.457	-0.228	0.221	-0.223	-0.248
	(0.032)	(0.032)	(0.251)	(0.135)	(0.128)	(0.128)	(0.459)	(0.180)	(0.087)	(0.080)	(0.078)	(0.397)
2,000	0.001	0.001	-0.020	0.014	-0.013	0.012	0.016	0.453	-0.229	0.227	-0.223	-0.229
	(0.023)	(0.023)	(0.175)	(0.092)	(0.095)	(0.089)	(0.326)	(0.125)	(0.059)	(0.056)	(0.055)	(0.272)
Mean Square Errors												
500	0.002	0.002	0.146	0.045	0.043	0.042	0.452	0.256	0.063	0.063	0.060	0.332
1,000	0.001	0.001	0.063	0.018	0.017	0.016	0.211	0.241	0.060	0.055	0.056	0.219
2,000	0.001	0.001	0.031	0.009	0.009	0.008	0.106	0.221	0.056	0.055	0.053	0.126

Table 3: Experiment III

n	$\hat{\alpha}_{NPL}$	$\hat{\delta}_{NPL}$		$\hat{\beta}_{NPL}$			$\hat{\gamma}_{NPL}$		$\tilde{\beta}_{NPL}$			$\tilde{\gamma}_{NPL}$
True Parameters: $\theta_0 = (0.2, 0.2; -1, 1, -1, 1; 1)$												
500	-0.001	-0.001	-0.078	0.057	-0.047	0.050	0.056	0.792	-0.361	0.357	-0.354	-0.363
	(0.052)	(0.052)	(0.418)	(0.232)	(0.227)	(0.229)	(0.711)	(0.246)	(0.113)	(0.111)	(0.105)	(0.476)
1,000	-0.001	-0.002	-0.020	0.017	-0.020	0.013	-0.001	0.810	-0.371	0.360	-0.359	-0.401
	(0.038)	(0.037)	(0.294)	(0.151)	(0.144)	(0.146)	(0.511)	(0.192)	(0.079)	(0.075)	(0.072)	(0.375)
2,000	0.001	0.001	-0.023	0.017	-0.015	0.016	0.009	0.816	-0.369	0.364	-0.357	-0.408
	(0.026)	(0.026)	(0.206)	(0.104)	(0.106)	(0.104)	(0.359)	(0.132)	(0.055)	(0.052)	(0.051)	(0.261)
Mean Square Errors												
500	0.003	0.003	0.181	0.057	0.054	0.055	0.509	0.687	0.143	0.140	0.136	0.358
1,000	0.001	0.001	0.087	0.023	0.021	0.021	0.261	0.692	0.144	0.135	0.134	0.301
2,000	0.001	0.001	0.043	0.011	0.012	0.011	0.129	0.683	0.139	0.135	0.130	0.234

Table 4: Experiment IV

n	$\hat{\alpha}_{NPL}$	$\hat{\delta}_{NPL}$		$\hat{\beta}_{NPL}$			$\hat{\gamma}_{NPL}$		$\tilde{\beta}_{NPL}$			$\tilde{\gamma}_{NPL}$
True Parameters: $\theta_0 = (0.4, 0.4; -1, 1, -1, 1; 1)$												
500	-0.004	-0.004	-0.101	0.078	-0.066	0.069	0.085	1.410	-0.526	0.521	-0.512	-0.565
	(0.059)	(0.059)	(0.529)	(0.281)	(0.281)	(0.274)	(0.807)	(0.250)	(0.107)	(0.106)	(0.099)	(0.394)
1,000	-0.003	-0.002	-0.033	0.025	-0.027	0.020	0.012	1.429	-0.536	0.523	-0.518	-0.608
	(0.041)	(0.042)	(0.363)	(0.185)	(0.178)	(0.178)	(0.593)	(0.197)	(0.074)	(0.070)	(0.070)	(0.319)
2,000	0.000	0.000	-0.028	0.020	-0.019	0.019	0.016	1.443	-0.535	0.525	-0.515	-0.628
	(0.029)	(0.028)	(0.250)	(0.125)	(0.124)	(0.123)	(0.412)	(0.137)	(0.051)	(0.051)	(0.050)	(0.223)
Mean Square Errors												
500	0.003	0.003	0.289	0.085	0.083	0.080	0.657	2.050	0.288	0.283	0.272	0.475
1,000	0.002	0.002	0.133	0.035	0.032	0.032	0.352	2.080	0.293	0.278	0.274	0.471
2,000	0.001	0.001	0.063	0.016	0.016	0.016	0.170	2.102	0.289	0.278	0.267	0.444

Table 5: Experiment V

n	$\hat{\alpha}_{NPL}$	$\hat{\delta}_{NPL}$		$\hat{\beta}_{NPL}$			$\hat{\gamma}_{NPL}$		$\tilde{\beta}_{NPL}$			$\tilde{\gamma}_{NPL}$
True Parameters: $\theta_0 = (0.05, 0.05; -1, 1, -1, 1; 2)$												
500	0.003	0.004	-0.042	0.036	-0.033	0.040	0.039	0.206	-0.094	0.095	-0.084	-0.212
	(0.042)	(0.041)	(0.331)	(0.176)	(0.168)	(0.163)	(0.606)	(0.272)	(0.136)	(0.124)	(0.116)	(0.538)
1,000	-0.003	-0.003	0.009	0.010	-0.013	0.017	-0.017	0.219	-0.097	0.093	-0.088	-0.228
	(0.031)	(0.031)	(0.240)	(0.118)	(0.117)	(0.116)	(0.436)	(0.193)	(0.089)	(0.086)	(0.085)	(0.384)
2,000	0.000	0.000	-0.006	0.007	-0.008	0.008	0.000	0.211	-0.105	0.101	-0.098	-0.216
	(0.023)	(0.023)	(0.165)	(0.084)	(0.080)	(0.078)	(0.297)	(0.129)	(0.062)	(0.060)	(0.057)	(0.264)
Mean Square Errors												
500	0.002	0.002	0.111	0.032	0.029	0.028	0.368	0.117	0.027	0.024	0.021	0.334
1,000	0.001	0.001	0.057	0.014	0.014	0.014	0.190	0.085	0.017	0.016	0.015	0.200
2,000	0.001	0.001	0.027	0.007	0.007	0.006	0.088	0.061	0.015	0.014	0.013	0.117

Table 6: Experiment VI

n	$\hat{\alpha}_{NPL}$	$\hat{\delta}_{NPL}$		$\hat{\beta}_{NPL}$			$\hat{\gamma}_{NPL}$		$\tilde{\beta}_{NPL}$			$\tilde{\gamma}_{NPL}$
True Parameters: $\theta_0 = (0.1, 0.1; -1, 1, -1, 1; 2)$												
500	-0.002	-0.002	-0.032	0.034	-0.031	0.038	0.029	0.396	-0.177	0.177	-0.166	-0.387
	(0.053)	(0.053)	(0.367)	(0.193)	(0.184)	(0.180)	(0.638)	(0.280)	(0.127)	(0.118)	(0.112)	(0.526)
1,000	-0.006	-0.005	0.012	0.009	-0.013	0.017	-0.020	0.406	-0.181	0.173	-0.167	-0.402
	(0.038)	(0.037)	(0.256)	(0.126)	(0.124)	(0.123)	(0.457)	(0.197)	(0.087)	(0.085)	(0.082)	(0.377)
2,000	-0.001	0.000	-0.009	0.007	-0.010	0.008	0.001	0.400	-0.190	0.182	-0.177	-0.394
	(0.026)	(0.027)	(0.176)	(0.088)	(0.086)	(0.083)	(0.319)	(0.133)	(0.061)	(0.058)	(0.055)	(0.263)
Mean Square Errors												
500	0.003	0.003	0.135	0.038	0.035	0.034	0.407	0.235	0.048	0.045	0.040	0.427
1,000	0.001	0.001	0.066	0.016	0.016	0.015	0.209	0.203	0.040	0.037	0.034	0.304
2,000	0.001	0.001	0.031	0.008	0.008	0.007	0.102	0.177	0.040	0.037	0.034	0.224

Table 7: Experiment VII

True Parameters: $\theta_0 = (0.2, 0.2; -1, 1, -1, 1; 2)$												
n	$\hat{\alpha}_{NPL}$	$\hat{\delta}_{NPL}$	$\hat{\beta}_{NPL}$				$\hat{\gamma}_{NPL}$	$\tilde{\beta}_{NPL}$				$\tilde{\gamma}_{NPL}$
500	-0.005	-0.004	-0.043	0.039	-0.037	0.046	0.048	0.738	-0.307	0.300	-0.286	-0.676
	(0.064)	(0.064)	(0.424)	(0.216)	(0.211)	(0.210)	(0.703)	(0.299)	(0.119)	(0.112)	(0.107)	(0.521)
1,000	-0.005	-0.006	0.014	0.011	-0.017	0.021	-0.024	0.759	-0.310	0.298	-0.289	-0.709
	(0.042)	(0.043)	(0.285)	(0.140)	(0.134)	(0.134)	(0.497)	(0.201)	(0.083)	(0.080)	(0.077)	(0.354)
2,000	0.000	0.000	-0.011	0.010	-0.014	0.011	0.012	0.748	-0.318	0.306	-0.299	-0.696
	(0.030)	(0.030)	(0.196)	(0.095)	(0.095)	(0.092)	(0.352)	(0.140)	(0.057)	(0.055)	(0.053)	(0.253)
Mean Square Errors												
500	0.004	0.004	0.181	0.048	0.046	0.046	0.496	0.634	0.108	0.103	0.094	0.728
1,000	0.002	0.002	0.081	0.020	0.018	0.018	0.247	0.617	0.103	0.095	0.090	0.628
2,000	0.001	0.001	0.038	0.009	0.009	0.009	0.124	0.579	0.104	0.097	0.093	0.548

Table 8: Experiment VIII

True Parameters: $\theta_0 = (0.4, 0.4; -1, 1, -1, 1; 2)$												
n	$\hat{\alpha}_{NPL}$	$\hat{\delta}_{NPL}$	$\hat{\beta}_{NPL}$				$\hat{\gamma}_{NPL}$	$\tilde{\beta}_{NPL}$				$\tilde{\gamma}_{NPL}$
500	-0.007	-0.006	-0.061	0.057	-0.049	0.065	0.075	1.400	-0.470	0.459	-0.439	-1.126
	(0.073)	(0.073)	(0.527)	(0.257)	(0.250)	(0.247)	(0.829)	(0.317)	(0.112)	(0.108)	(0.109)	(0.466)
1,000	-0.004	-0.006	0.007	0.026	-0.024	0.031	-0.017	1.420	-0.471	0.457	-0.447	-1.158
	(0.046)	(0.047)	(0.362)	(0.170)	(0.167)	(0.167)	(0.601)	(0.225)	(0.080)	(0.077)	(0.076)	(0.345)
2,000	-0.002	-0.001	-0.011	0.010	-0.014	0.011	0.018	1.398	-0.478	0.460	-0.449	-1.134
	(0.031)	(0.031)	(0.246)	(0.114)	(0.111)	(0.109)	(0.416)	(0.154)	(0.054)	(0.051)	(0.054)	(0.237)
Mean Square Errors												
500	0.005	0.005	0.282	0.069	0.065	0.065	0.692	2.062	0.233	0.223	0.204	1.484
1,000	0.002	0.002	0.131	0.030	0.029	0.029	0.362	2.067	0.229	0.215	0.206	1.460
2,000	0.001	0.001	0.061	0.013	0.013	0.012	0.173	1.977	0.232	0.215	0.205	1.343

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 decisions on ability.

Attitude towards learning is the way of thinking or feeling about study and educational aspirations. Typically, attitude is reflected in a student's behavior and originates from the student's decisions. Peer effects demonstrate the interconnection among students on choices, e.g. work hard, take exercise, smoke, drink, etc. For learning spillover, peer effects play role in the chosen attitude rather than in the final achievements. Thus investigation on peer effects on attitude is legitimate, however, attitude is subjective and difficult to measure. In the *National Longitudinal Study of Adolescent Health (Add Health)* dataset, we obtain several measurements in the survey regarding the attitudes of students. Attitude regarding questions are socially and personally sensitive and students tend to overreport. This feature raises the issue of misclassification errors due to social desirability. Fortunately, repeated measurements in the survey provide a remedy to such a misclassification error problem.

We denote the peer effects on attitude towards learning as "silent rivalry" that students strive in a silent manner. Using both in-school and at-home surveys, we obtain repeated measurements for attitude from the question "Skipped school without an excuse". Interestingly, we find that the silent rivalry either disappears or is underestimated if we directly use these two measurements as attitudes. The silent rivalry based on the binary choice with misclassification and social interactions in this paper is roughly three times larger than the direct application of the original attitude measurement (1.543 vs 0.482). Our findings confirm our insight into the prevalence of silent rivalry 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". We also find that significant proportions of students overreport their attitudes (28.9% in the in-school survey and 25.6% in the at-home survey, respectively). A corresponding policy implication of this documented silent rivalry suggests the importance of an initiation of a "diligent" atmosphere in the school. The multiplier effects from silent rivalry would help to obtain a desired result.

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, Halpern, Whitsel, Hussey, Tabor, Entzel, and Udry (2009)]. Wave V data collection began in 2016 and is still in progress. *Add Health* combines longitudinal survey data on respondents' social and economic features with contextual data on the 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 $\left[\{F_{ij}\}_{i,j=1}^{n}\right]$. Note that although students have at most 10 out-links, they may have more than 10 in-links. 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 include 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 for 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 and the answer for the in-school question as Z. We take the answer "never" as a "positive" attitude and all other answers as "negative" attitudes. Here, Y and Z 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.

We consider the sister schools No. 77 and No. 177 for our analysis. These two schools contain the largest single connected school network in the *Add Health* dataset. There are friendships across the sister schools. After data management, we obtain 1,173 students in the sample. Table 9 summarizes the statistics of the demographic characteristics and the attitude variables. More than half of students escape school without excuse at least once in the year. The attitude measurement from at-home questionnaire has a little bit less "positive" than that coming from the in-school questionnaire, 45% v.s. 47.1%.

6.2 The Hidden Silent Rivalry and the Misclassification Probabilities

We report our estimation results in Table 10. The misclassification probabilities are 28.9% and 25.6% for the in-school answer and at-home answer, respectively. Both estimates are statistically significant. Roughly, one quarter of students overreport their

Table 9: Summary of Statistics of Key Variables from the Data

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	0.385	0.487
Others*	0.130	0.336
Attitude(Y)	0.450	0.498
Attitude(Z)	0.471	0.499

†5 means "went to a business, trade, or vocational school after high school" and 6 means "went to college but did not graduate".

*Some students are associated with more than one race.

attitudes. Students are more likely to be honest at home where there are no peers. Our finding confirms the desire to be "positive" for students. When it comes to the silent rivalry, we have three options to back out the interaction parameter. We can either take Y or Z as the true latent attitude to estimate the binary choice with social interactions without misclassification correction (Model M1 and M2), or we adopt the full information from two repeated measurements to rectify the misclassification errors (2M model). In Table 10, models without misclassification correction either fail to detect a significant silent rivalry ($\hat{\gamma} = 0$ in model M1) or underestimate the peer effects ($\hat{\gamma} = 0.482$ in model M2). Our 2M model estimates a significant 1.543[12] 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 significant misclassification problems when students self-report their attitudes towards learning, either in-school or at-home. The peer effects analysis in attitude is contaminated by this misclassification error. Treatment on misclassification is needed to restore the conjectured silent rivalry among students. We also conduct robustness check of the discretized definition of "positive" attitude to include both "never" and "once" answers for the survey question in Appendix B. Results

[12]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 closed to the NPL standard error, which is consistent with our simulation results. We are working on a project to derive bootstrapping standard error for the network generated dependent data.

Table 10: Estimation Results on Silent Rivalry

	2M	M1	M2	Logit models	
				Y	Z
Age	-0.449*	-0.347*	-0.200*	-0.350*	-0.199*
	(0.123)	(0.056)	(0.053)	(0.056)	(0.053)
Female	-0.062	0.111	-0.107	0.111	-0.083
	(0.162)	(0.122)	(0.119)	(0.121)	(0.119)
Parents' Education	0.050	0.009	0.029	0.009	0.032
	(0.040)	(0.027)	(0.026)	(0.027)	(0.026)
Hispanic	-0.630*	-0.496*	-0.241	-0.499*	-0.239
	(0.285)	(0.197)	(0.192)	(0.197)	(0.192)
Asian	-0.257	-0.097	-0.161	-0.099	-0.121
	(0.255)	(0.201)	(0.197)	(0.199)	(0.196)
African American	-0.173	-0.086	-0.124	-0.090	-0.141
	(0.253)	(0.207)	(0.204)	(0.207)	(0.204)
Native American	-0.680	-0.089	-0.375	-0.091	-0.389
	(0.518)	(0.288)	(0.286)	(0.288)	(0.286)
Other	0.245	0.326*	-0.051	0.327	-0.033
	(0.259)	(0.197)	(0.194)	(0.197)	(0.193)
α	0.256*	—-	—-	—-	—-
	(0.067)	—-	—-	—-	—-
δ	0.289*	—-	—-	—-	—-
	(0.064)	—-	—-	—-	—-
Peer Effects (γ)	1.543*	0.000	0.482*	—-	—-
	(0.712)	(0.289)	(0.279)	—-	—-
Constant	5.848*	5.413*	3.007*	5.464*	3.110*
	(1.549)	(0.947)	(0.902)	(0.936)	(0.896)

a. * for 5% significance.

b. significances of α, δ and γ obtained from the one-sided test.

c. White students are left for comparison.

are similar for covariates effects and peer effects. Our rectification with two repeated measurements helps to detect a much stronger peer effects on attitude towards learning and justifies the importance of the manipulation of the peer group influence. Our investigation has important policy implications.

7 Conclusion

In this paper, we propose a binary choice model with misclassification and social interactions and bring the model to study the silent rivalry among students on attitude towards learning. We provide a closed-form identification result to our model primitives by adopting a two-measurement approach. Taking into account the full information embedded in the two measurements, we construct complete likelihood function for estimation of the structural parameter in the silent rivalry study using nested pseudo likelihood algorithm. We find peer effects on attitude towards learning are either hidden or underestimated if omitting the misclassification problem. This finding provides insights of how to improve school performance rather than monetary tools. We also find that significant proportions of students overreport their attitudes towards learning. This documents the source of the misclassification since a specific answer is socially-desired regarding attitude. The silent rivalry triggers multiplier effects which help improve the performance of schools and are meaningful for policy implications.

References

Abrevaya, J., and J. A. Hausman, 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., and P. Mira, 2002. Swapping the nested fixed point algorithm: A class of estimators for discrete Markov decision models, *Econometrica*, 70(4), 1519–1543.

———, 2007. Sequential estimation of dynamic discrete games, *Econometrica*, 75(1), 1–53.

Anderson, A., L. Smith, and A. Park, 2017. Rushes in large timing games, *Econometrica*, 85(3), 871–913.

Aradillas-Lopez, A., 2010. Semiparametric estimation of a simultaneous game with incomplete information, *Journal of Econometrics*, 157(2), 409–431.

———, 2012. Pairwise-difference estimation of incomplete information games, *Journal of Econometrics*, 168(1), 120–140.

Arcidiacono, P., P. Bayer, J. R. Blevins, and P. B. Ellickson, 2016. Estimation of dynamic discrete choice models in continuous time with an application to retail competition, *The Review of Economic Studies*, 83(3), 889–931.

Arcidiacono, P., G. Foster, N. Goodpaster, and J. Kinsler, 2012. Estimating spillovers using panel data, with an application to the classrooms, *Quantitative Economics*, 3(3), 421–470.

Badev, A., 2017. Discrete games in endogenous networks: equilibria and policy, *Working paper*.

Bajari, P., H. Hong, J. Krainer, and D. Nekipelov, 2010. Estimating static models of strategic interactions, *Journal of Business & Economic Statistics*, 28(4), 469–482.

Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides, 2011. Identification of social interactions, *Handbook of Social Economics*, 1, 853–964.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman, 2015. Linear social interactions models, *Journal of Political Economy*, 123(2), 444–496.

Bound, J., C. Brown, and N. Mathiowetz, 2001. Measurement error in survey data, *Handbook of Econometrics*, 5, 3705–3843.

Bramoullé, Y., H. Djebbari, and B. Fortin, 2009. Identification of peer effects through social networks, *Journal of Econometrics*, 150(1), 41–55.

Bramoullé, Y., and R. Kranton, 2016. Games played on networks, *The Oxford Handbook of the Economics of Networks*, pp. 1–33.

Bramoullé, Y., R. Kranton, and M. D'amours, 2014. Strategic interaction and networks, *The American Economic Review*, 104(3), 898–930.

Brock, W. A., and S. N. Durlauf, 2001a. Discrete choice with social interactions, *The Review of Economic Studies*, 5(2), 235–260.

———, 2001b. Interaction-based models, *Handbook of Econometrics*, 5, 3297–3380.

———, 2007. Identification of binary choice models with social interactions, *Journal of Econometrics*, 140(1), 52–75.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou, 2009. Peer effects and social networks in education, *The Review of Economic Studies*, 76(4), 1239–1267.

Canen, N., J. Schwartz, and K. Song, 2017. Estimating Local Interactions Among Many Agents Who Observe Their Neighbors, *Working paper*.

Card, D., and L. Giuliano, 2013. Peer effects and multiple equilibria in the risky behavior of friends, *Review of Economics and Statistics*, 95(4), 1130–1149.

Chandrasekhar, A. G., and M. O. Jackson, 2016. A network formation model based on subgraphs, *Working paper*.

Chen, X., H. Hong, and D. Nekipelov, 2011. Nonlinear models of measurement errors, *Journal of Economic Literature*, 49(4), 901–937.

Christakis, N. A., J. H. Fowler, G. W. Imbens, and K. Kalyanaraman, 2010. An empirical model for strategic network formation, Discussion paper, National Bureau of Economic Research.

Coleman, J. S., et al., 1966. Equality of educational opportunity, *US Government Printing Office, Washington, DC*.

Dahl, G. B., K. V. Løken, and M. Mogstad, 2014. Peer effects in program participation, *The American Economic Review*, 104(7), 2049–2074.

De Nadai, M., and A. Lewbel, 2016. Nonparametric errors in variables models with measurement errors on both sides of the equation, *Journal of Econometrics*, 191, 19–32.

de Paula, Á., 2013. Econometric Analysis of Games with Multiple Equilibria, *Annual Review of Economics*, 5, 107–131.

———, 2017. Econometrics of Network Models, *Advances in Economics and Econometrics*, 1, 268–323.

de Paula, Á., S. Richards-Shubik, and E. T. Tamer, 2017. Identification of preferences in network formation games, *Forthcoming in Econometrica*.

de Paula, Á., and X. Tang, 2012. Inference of signs of interaction effects in simultaneous games with incomplete information, *Econometrica*, 80(1), 143–172.

Durlauf, S. N., and Y. M. Ioannides, 2010. Social interactions, *Annual Review of Economics*, 2(1), 451–478.

Epple, D., and R. Romano, 2011. Peer effects in education: A survey of the theory and evidence, *Handbook of Social Economics*, 1(11), 1053–1163.

Eraslan, H., and X. Tang, 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, *Review of Economics and Statistics*, 96(3), 514–523.

Gladwell, M., 2000. *The tipping point: How little things can make a big difference*. Little, Brown.

Glaeser, E., and J. A. Scheinkman, 2003. Nonmarket interactions, *Advances in Economics and Econometrics*, 1, 339–369.

Goldsmith-Pinkham, P., and G. W. Imbens, 2013. Social networks and the identification of peer effects, *Journal of Business & Economic Statistics*, 31(3), 253–264.

Graham, B. S., 2008. Identifying social interactions through conditional variance restrictions, *Econometrica*, 76(3), 643–660.

———, 2015. Methods of identification in social networks, *Annual Review of Economics*, 7(1), 465–485.

———, 2016. Homophily and transitivity in dynamic network formation, Discussion paper, National Bureau of Economic Research.

———, 2017. An econometric model of network formation with degree heterogeneity, *Econometrica*, 85(4), 1033–1063.

Granovetter, M., 1978. Threshold models of collective behavior, *American Journal of Sociology*, 83(6), 1420–1443.

Harris, K. M., C. T. Halpern, E. A. Whitsel, J. M. Hussey, J. Tabor, P. Entzel, and J. R. Udry, 2009. The National Longitudinal Study of Adolescent to Adult Health: Research Design, *URL: http://www.cpc.unc.edu/projects/addhealth/design*.

Hausman, J., J. Abrevaya, and F. M. Scott-Morton, 1998. Misclassification of the dependent variable in a discrete-response setting, *Journal of Econometrics*, 87(2), 239–269.

Hausman, J. A., 2001. Mismeasured variables in econometric analysis: problems from the right and problems from the left, *The Journal of Economic Perspectives*, 15(4), 57–67.

Horst, U., and J. A. Scheinkman, 2006. Equilibria in systems of social interactions, *Journal of Economic Theory*, 130, 44–77.

———, 2009. A limit theorem for systems of social interactions, *Journal of Mathematical Economics*, 45(9), 609–623.

Hoshino, T., 2017. Two-Step Estimation of Incomplete Information Social Interaction Models with Sample Selection, *Accepted by Journal of Business & Economic Statistics*.

Hoxby, C., 2000. Peer effects in the classroom: Learning from gender and race variation, Discussion paper, National Bureau of Economic Research.

Hsiao, C., and B.-H. Sun, 1998. Modeling survey response bias–with an analysis of the demand for an advanced electronic device, *Journal of Econometrics*, 89(1), 15–39.

Hu, Y., 2008. Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution, *Journal of Econometrics*, 144(1), 27–61.

———, 2017. The Econometrics of Unobservables: Applications of Measurement Error Models in Empirical Industrial Organization and Labor Economics, *Journal of Econometrics*, 200(2), 154–168.

Hu, Y., and Y. Sasaki, 2017. Identification of paired nonseparable measurement error models, *Econometric Theory*, 33(4), 955–979.

Hu, Y., and S. M. Schennach, 2008. Instrumental variable treatment of nonclassical measurement error models, *Econometrica*, 76(1), 195–216.

Ioannides, Y., 2006. Topologies of social interactions, *Economic Theory*, 28, 559–584.

Kasahara, H., and K. Shimotsu, 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, *Journal of Econometrics*, 140(2), 333–374.

Lee, L., J. Li, and X. Lin, 2014. Binary choice models with social network under heterogeneous rational expectations, *Review of Economics and Statistics*, 96(3), 402–417.

Lee, L., X. Liu, and X. Lin, 2010. Specification and estimation of social interaction models with network structures, *The Econometrics Journal*, 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, *Review of Economics and Statistics*, 78, 718–725.

———, 2000. Identification of the binary choice model with misclassification, *Econometric Theory*, 16(04), 603–609.

Li, T., 2002. Robust and consistent estimation of nonlinear errors-in-variables models, *Journal of Econometrics*, 110(1), 1–26.

Li, T., and C. Hsiao, 2004. Robust estimation of generalized linear models with measurement errors, *Journal of Econometrics*, 118(1), 51–65.

Li, T., P. K. Trivedi, and J. Guo, 2003. Modeling response bias in count: a structural approach with an application to the national crime victimization survey data, *Sociological Methods & Research*, 31(4), 514–544.

Li, T., and L. Zhao, 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, *Journal of Labor Economics*, 28(4), 825–860.

Lin, Z., and H. Xu, 2017. Estimation of social-influence-dependent peer pressures in a large network game, *Econometrics Journal*, 20(3), 86–102.

Liu, X., 2017. Simultaneous Equations with Binary Outcomes and Social Interactions, *Working paper*.

Liu, X., and L. Lee, 2010. GMM estimation of social interaction models with centrality, *Journal of Econometrics*, 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, *The Review of Economic Studies*, 60(3), 531–542.

———, 2000. Economic Analysis of Social Interactions, *The Journal of Economic Perspectives*, 14(3), 115–136.

Matzkin, R. L., 2007. Nonparametric identification, *Handbook of Econometrics*, 6, 5307–5368.

———, 2013. Nonparametric identification in structural economic models, *Annual Review of Economics*, 5(1), 457–486.

Mcfadden, D., 1974. Conditional logit analysis of qualitative choice behavior, *Frontiers in Econometrics*, pp. 105–142.

Mele, A., 2017. A structural model of dense network formation, *Econometrica*, 85(3), 825–850.

Menzel, K., 2015. Inference for games with many players, *The Review of Economic Studies*, 83(1), 306–337.

———, 2017. Strategic network formation with many agents, *Working paper*.

Meyer, B. D., and N. Mittag, 2017. Misclassification in binary choice models, *Journal of Econometrics*, 200(2), 295–311.

Morris, S., and H. S. Shin, 2003. Global games: theory and applications, *Advances in Economics and Econometrics*, 1, 56–114.

Newey, W., and D. McFadden, 1994. Large sample estimation and hypothesis testing, *Handbook of Econometrics*, 4, 2111–2245.

Park, A., and L. Smith, 2008. Caller number five and related timing games, *Theoretical Economics*, 3(2), 231–256.

Pesendorfer, M., and P. Schmidt-Dengler, 2008. Asymptotic least squares estimators for dynamic games, *The Review of Economic Studies*, 75(3), 901–928.

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?, *Handbook of the Economics of Education*, 3(3), 249–277.

Schelling, T. C., 1971. Dynamic models of segregation, *Journal of Mathematical Sociology*, 1(2), 143–186.

Schennach, S. M., 2016. Recent advances in the measurement error literature, *Annual Review of Economics*, 8(1), 341–377.

Seim, K., 2006. An empirical model of firm entry with endogenous product-type choices, *The RAND Journal of Economics*, 37(3), 619–640.

Sheng, S., 2017. Identification and estimation of network formation games, *Forthcoming in Econometrica*.

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, *The RAND Journal of Economics*, 40(4), 710–742.

Tamer, E., 2003. Incomplete simultaneous discrete response model with multiple equilibria, *The Review of Economic Studies*, 70(1), 147–165.

———, 2010. Partial identification in econometrics, *Annual Review of Economics*, 2(1), 167–195.

Tang, X., 2010. Estimating simultaneous games with incomplete information under median restrictions, *Economics Letters*, 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, *The Econometrics Journal*, 17(3), 241–270.

———, 2018. Social interactions on large networks: a game theoretic approach, *International Economic Review*, 59(1), 257–284.

Yang, C., and L. Lee, 2017. Social interactions under incomplete information with heterogeneous expectations, *Journal of Econometrics*, 198(1), 65–83.

Zimmerman, D. J., 2003. Peer effects in academic outcomes: Evidence from a natural experiment, *Review of Economics and Statistics*, 85(1), 9–23.

Appendix A Proofs

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

$$
\begin{aligned}
|P_i^1 - P_i^2| &= |\Gamma_i(\mathcal{W}, P^1; \mu) - \Gamma_i(\mathcal{W}, P^2; \mu)|, \\
&= \left|\Lambda(X_i^T\beta + \frac{\gamma}{N_i}\sum_{j\in F_i} P_j^1) - \Lambda(X_i^T\beta + \frac{\gamma}{N_i}\sum_{j\in F_i} P_j^2)\right|, \\
&= \Lambda\Big(X_i^T\beta + \frac{\gamma}{N_i}\sum_{j\in F_i} P_j^\dagger\Big)\Big[1 - \Lambda\Big(X_i^T\beta + \frac{\gamma}{N_i}\sum_{j\in F_i} P_j^\dagger\Big)\Big]\left|\frac{\gamma}{N_i}\sum_{j\in F_i}(P_j^1 - P_j^2)\right|, \qquad (16) \\
&\leq \frac{1}{4}\cdot\gamma\cdot\max_{j\in\mathcal{I}}\left|P_j^1 - P_j^2\right| < 4\cdot\frac{1}{4}\max_{j\in\mathcal{I}}\left|P_j^1 - P_j^2\right|, \\
&= \max_{j\in\mathcal{I}}\left|P_j^1 - P_j^2\right|,
\end{aligned}
$$

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 $i \in \mathcal{I}$ on the left-hand side of Equation (16), we have

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

which is a contradiction. Therefore we have a unique BNE for the Bayesian Nash game in Equation (3). □

Proof of Theorem 2. The proof is similar as that in Aguirregabiria and Mira (2007); Newey and McFadden (1994). With Assumption 8(a), we have that $\theta_{NPL} = \theta_0$. Recall that the pseudo likelihood function is $\mathcal{L}(\theta, P)$ in the NPL estimation. Define the function

$$
T(\theta, P) \equiv \max_{c\in\Theta}\Big\{\mathcal{L}_0(c, P)\Big\} - \mathcal{L}_0(\theta, P).
$$

Because $\mathcal{L}_0(\theta, P)$ is continuous and $\Theta \times \mathcal{P}$ is compact, Berge's maximum theorem establishes that $T(\theta, P)$ is a continuous function. By construction, $T(\theta, P) \geq 0$ for any (θ, P).

Let $\mathcal{E}$ be the set of vectors (θ, P) that are fixed points of the equilibrium mapping Γ, i.e.,

$$\mathcal{E} \equiv \Big\{(\theta, P) \in \Theta \times \mathcal{P} : P = \Gamma(\theta, P)\Big\}.$$

Given that $\Theta \times \mathcal{P}$ is compact and Γ is continuous, $\mathcal{E}$ then is a compact set. By definition, the set Λ_0 is included in $\mathcal{E}$. Let $B_\epsilon(\theta_0) = \{\theta \in \mathbb{R}^{d+3} : \|\theta - \theta_0\| < \epsilon, \forall \epsilon > 0\}$ be an arbitrarily small open ball that contains θ_0. We then see that $B^c_\epsilon(\theta_0) \cap \mathcal{E}$ is also compact. Define the constant

$$\tau \equiv \min_{(\theta,P)\in B^c_\epsilon(\theta_0)\cap\mathcal{E}} T(\theta, P) > 0. \tag{17}$$

Define the event

$$A \equiv \Big\{(\theta, P) \ \in \Theta \times \mathcal{P} : \Big|\mathcal{L}(\theta, P) - \mathcal{L}_0(\theta, P)\Big| < \frac{\tau}{2} \text{ for all } (\theta, P) \in \Theta \times \mathcal{P}\Big\}.$$

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

$$\mathcal{L}_0\Big(\theta^{(n)}, P^{(n)}\Big) > \mathcal{L}\Big(\theta^{(n)}, P^{(n)}\Big) - \frac{\tau}{2},$$

and

$$\mathcal{L}\Big(\theta, P^{(n)}\Big) > \mathcal{L}_0\Big(\theta, P^{(n)}\Big) - \frac{\tau}{2}.$$

Furthermore, we have $\mathcal{L}\Big(\theta^{(n)}, P^{(n)}\Big) \geq \mathcal{L}\Big(\theta, P^{(n)}\Big)$ from the NPL fixed point definition. Therefore, we have that $\mathcal{L}_0\Big(\theta^{(n)}, P^{(n)}\Big) > \mathcal{L}_0\Big(\theta, P^{(n)}\Big) - \tau$. We then have the following derivation:

$$\begin{aligned}
A \ &\Rightarrow \ \Big\{\mathcal{L}_0\Big(\theta^{(n)}, P^{(n)}\Big) > \mathcal{L}_0\Big(\theta, P^{(n)}\Big) - \tau \text{ for any } \theta \in \Theta\Big\}, \\
&\Rightarrow \ \Big\{\mathcal{L}_0\Big(\theta^{(n)}, P^{(n)}\Big) > \max_{\theta\in\Theta} \mathcal{L}_0\Big(\theta, P^{(n)}\Big) - \tau\Big\}, \\
&\Rightarrow \ \Big\{\tau > T\Big(\theta^{(n)}, P^{(n)}\Big)\Big\}, \\
&\Rightarrow \ \Big\{\min_{(\theta,P)\in B^c_\epsilon(\theta_0)\cap\mathcal{E}} T(\theta, P) > T\Big(\theta^{(n)}, P^{(n)}\Big)\Big\} \text{ by Equation (17)}, \\
&\Rightarrow \ \Big\{(\theta^{(n)}, P^{(n)}) \in B_\epsilon(\theta_0)\Big\}.
\end{aligned}$$

The last induction uses the fact that $(\theta^{(n)}, P^{(n)}) \in \mathcal{E}$. Therefore, $\Pr(A) \leq \Pr\Big((\theta^{(n)}, P^{(n)}) \in B_\epsilon(\theta_0)\Big)$. Because $\Pr(A) \to 1$ as $n \to \infty$, $\Pr\Big((\theta^{(n)}, P^{(n)}) \in B_\epsilon(\theta_0)\Big) \to 1$. Because ϵ in $B_\epsilon(\theta_0)$ is an arbitrarily small constant, we have

$$\Big(\theta^{(n)}, P^{(n)}\Big) \xrightarrow{p} (\theta_0, P^*).$$

From the definition of Λ_n, we have that $\hat{\theta}_{NPL} \xrightarrow{p} \theta_0$. Now we establish the asymptotic normality of the NPL estimator. Taking Taylor expansion over the first order condition in Equation (13) around the true parameter (θ_0, P^*), we have

$$\frac{\partial \mathcal{L}(\theta_0, P^*)}{\partial \theta} + \frac{\partial^2 \mathcal{L}(\theta^+, P^+)}{\partial \theta \partial \theta^T}(\hat{\theta}_{NPL} - \theta_0) + \frac{\partial^2 \mathcal{L}(\theta^+, P^+)}{\partial \theta \partial P}\left[I - \left(\frac{\partial \Gamma(P^+, \mathcal{W}; \theta^+)}{\partial P}\right)^T\right]^{-1} \frac{\partial \Gamma(P^+, \mathcal{W}; \theta^+)}{\partial \theta}(\hat{\theta}_{NPL} - \theta_0) = 0, \tag{18}$$

where θ^+ is between $\hat{\theta}_{NPL}$ and θ_0 and P^+ is between $\hat{P}_{NPL}$ and P^* respectively. From Equation (18), we have that

$$\left[\frac{\partial^2 \mathcal{L}(\theta^+, P^+)}{\partial \theta \partial \theta^T} + \frac{\partial^2 \mathcal{L}(\theta^+, P^+)}{\partial \theta \partial P}\left[I - \left(\frac{\partial \Gamma(P^+, \mathcal{W}; \theta^+)}{\partial P}\right)^T\right]^{-1} \frac{\partial \Gamma(P^+, \mathcal{W}; \theta^+)}{\partial \theta}\right]\sqrt{n}(\hat{\theta}_{NPL} - \theta_0) = -\frac{1}{\sqrt{n}} \sum_{i \in \mathcal{I}} \frac{\partial \mathcal{L}_i(\theta_0, P^*)}{\partial \theta}. \tag{19}$$

Because Y_i is conditionally independent (conditional on $\mathcal{W}$), by Lindeberg-Feller theorem, Mann-Wald theorem and Assumption 8(c-e), we have

$$\sqrt{n}(\hat{\theta}_{NPL} - \theta_0) \xrightarrow{d} N(0, V_{NPL}), \tag{20}$$

where

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

□

Appendix B Robustness Check

In this section, we check the robustness of the discretization definition of attitude. Here we take "never" and "once or twice" as positive. We estimate the corrected 2M model, the M1 and M2 model as well as logit models. The results are presented in Table 11. The results are similar as those in Table 10. The peer effects is 1.553 compared to 1.543 in the 2M model for the two discretization definition of "positive". The peer effects coefficients are 0.003 and 0.466 compared to 0 and 0.482 in the M1 and M2 models respectively. The overreport proportions increase as there are more students categorized as "positive" in the new definition.

Table 11: Robustness Check

	2M	M1	M2	Logit models	
				Y	Z
Age	-0.447*	-0.335*	-0.105	-0.337*	-0.101
	(0.172)	(0.056)	(0.057)	(0.056)	(0.057)
Female	0.241	0.187	0.287*	0.186	0.332*
	(0.176)	(0.122)	(0.131)	(0.121)	(0.130)
Parents' Education	0.013	-0.003	-0.002	-0.003	0.001
	(0.042)	(0.027)	(0.028)	(0.027)	(0.028)
Hispanic	-0.672*	-0.474*	-0.249	-0.478*	-0.252
	(0.331)	(0.196)	(0.207)	(0.196)	(0.207)
Asian	0.039	0.066	0.070	0.065	0.128
	(0.278)	(0.200)	(0.215)	(0.198)	(0.213)
African American	-0.034	0.171	-0.248	0.168	-0.291
	(0.282)	(0.207)	(0.221)	(0.207)	(0.220)
Native American	-0.423	-0.258	-0.084	-0.259	-0.108
	(0.474)	(0.290)	(0.297)	(0.290)	(0.296)
Other	0.560	0.500*	-0.017	0.501*	0.005
	(0.345)	(0.198)	(0.208)	(0.197)	(0.207)
α	0.289*	—-	—-	—-	—-
	(0.105)	—-	—-	—-	—-
δ	0.573*	—-	—-	—-	—-
	(0.065)	—-	—-	—-	—-
Peer Effects (γ)	1.553*	0.003	0.466*	—-	—-
	(0.823)	(0.256)	(0.198)	—-	—-
Constant	5.879*	5.358*	2.343*	5.400*	2.457*
	(2.122)	(0.950)	(0.966)	(0.939)	(0.968)