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 Fishbourne 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 an-
swering survey questions; social desirability for some answers; essential survey con-
ditions; and applicability of findings to the measurement of economic phenomena;
see Bound, Brown, and Mathiowetz (2001) for details. Decision variables possess dis-
creteness and sometimes are dichotomous. Discrete measurement error is also called
misclassification error. Typically, discrete decision variables require nonlinear tech-
niques that are different from those deployed in linear models. Econometricians have
devoted increasing attention to the magnitude and consequences of measurement er-
ror 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 excep-
tions, 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 ef-
ciciency 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, in-
formation diffusion, word of mouth, etc. Brock and Durlauf (2001a,b) pioneer the dis-
crete 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 in-
teractions and the linear social interations model, see Durlauf and Ioannides (2010);
Blume, Brock, Durlauf, and Ioannides (2011); Blume, Brock, Durlauf, and Jayaraman
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); Bramoulle, 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); Bramoulle, Kran ton, 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 Bramoulle 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 stra-
 tegy in Section 4. Eight Monte Carlo experiments are conducted in Section 5 to illus-
trate 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 Interac-
tions

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 interac-
tions 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 friend-
ships 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

\[
U_i(Y_i^*, X_i, F_i, \epsilon_i) = X_i^T \beta + \frac{Y_i^*}{N_i} \sum_{j \in F_i} Y_j^* - \epsilon_i, \tag{1}
\]

where superscript \( T \) stands for the matrix transpose, \( X_i \in \mathcal{X} \) is a \( d \times 1 \) vector repre-
senting the demographic characteristics\(^1\), \( Y_j^* \) are the decisions of others, and \( \epsilon_i \) is the

\(^1X_i \) 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), $\frac{1}{N_i} \sum_{j \in F_i} Y_j^*$, and a private utility shock, $\epsilon_i$. $\gamma$ 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 $\mathcal{W} = \{\{X_i\}_{i \in I}, \{F_i\}_{i \in I}\}$ be the public information set including all demographic characteristics and friendship information$^2$. The private utility shock $\epsilon_i$ 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:

$$Y_i^* = \begin{cases} X_i^T \beta + \frac{\gamma}{N_i} \sum_{j \in F_i} \mathbb{E}(Y_j^*|\mathcal{W}, \epsilon_i) - \epsilon_i \geq 0 \end{cases},$$

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 $\epsilon_i$’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., \( \mathbb{E}(Y_j|W, \varepsilon_i) = \mathbb{E}(Y_j|W) \).

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

\[
\sigma^*_i(W; \mu) = \Lambda \left[ X^T_i \beta + \frac{Y}{N_i} \sum_{j \in F_i} \sigma^*_j(W; \mu) \right] \equiv \Gamma_i(W, \sigma^*; \mu), i \in I,
\]

Equation (3) is a simultaneous system of equations of \( \sigma^* \). Let \( P \) be an arbitrary choice probabilities profile. The equilibrium choice probability profile \( \sigma^* \) defined in Equation (3) is then a fixed point of

\[
\Gamma(W, P; \mu) \equiv (\Gamma_1(W, P; \mu), \ldots, \Gamma_n(W, P; \mu))^T = P.
\]

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. \( N_i \leq M \) for \( i \in I \).

**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, $0 < \gamma < 1 \sup f_{z}(\cdot)$ 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 $\{Y_{i}^{*}\}_{i \in I}$ 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 \( \{Y_i \}_{i \in I} \) and \( \{Z_i \}_{i \in I} \) 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 \( \mathcal{W} \),

\[
Y \perp Z \mid (Y^*, \mathcal{W}).
\] (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 measurement. 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\textsuperscript{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\textsuperscript{8}, see Hu (2017) for review of the 3-measurement model. Assumption 5 implies $P(Y = 1 | Y^* = 1, W) = P(Z = 1 | Y^* = 1, W) = 1$.

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., $P(Y^* = 1 | W)$\textsuperscript{9}. Identification is about the recovery of $P(Y^* = 1 | W)$ uniquely from observables, i.e., $P(Y = 1 | W), P(Z = 1 | W)$ and $P(Y, Z | W)$.

We define

$$M_{Y,Z|W} \equiv \begin{pmatrix} P(Y = 0, Z = 0 | W) & P(Y = 0, Z = 1 | W) \\ P(Y = 1, Z = 0 | W) & P(Y = 1, Z = 1 | W) \end{pmatrix},$$

$$\equiv \left[ P(Y = i - 1, Z = j - 1 | W) \right]_{i,j=1}^2.$$

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

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

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

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

\textsuperscript{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.

\textsuperscript{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.

\textsuperscript{9}Because we are considering the binary case, $P(Y^* = 1 | W)$ fully characterize the conditional distribution, i.e., $P(Y^* = 0 | W) = 1 - P(Y^* = 1 | W)$. 
probability and Assumption 4, we obtain

\[
M_{Y,Z|W} = M_{Y|Y^*,W} \times M_{Z,Y^*|W} = M_{Z|Y^*,W} \times M_{Y,Y^*|W}, \tag{5}
\]

\[
M_{Y,Z|W} = M_{Y|Y^*,W} \times D_{Y^*|W} \times M_{Z,Y^*|W}. \tag{6}
\]

With condition on the evasive misclassification probabilities (Assumption 5), \(M_{Y|Y^*,W}\) and \(M_{Z,Y^*|W}\) 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 \(M_{Y|Y^*,W}\) equals one and that the sum of all the entries in \(M_{Z,Y^*|W}\) also equals one. Thus we have identified two misclassification matrices \(M_{Y|Y^*,W}\) and \(M_{Z,Y^*|W}\). Similarly we identify \(M_{Z|Y^*,W}^T\) and \(M_{Y,Y^*|W}\). For more details on LU decomposition, see Hu and Sasaki (2017). Then the conditional distribution of the latent variable is identified through:

\[
D_{Y^*|W} = M_{Y|Y^*,W}^{-1} \cdot M_{Y,Z|W} \cdot M_{Z,Y^*|W}^{-1}. \tag{7}
\]

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

### 3.2 Identification of the Structural Parameter, \(\mu\)

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

\[
\Xi(W) \equiv \log \left[ P(Y^*_i = 1|W) \right] - \log \left[ P(Y^*_i = 0|W; \mu) \right] = X^T_i \beta + \frac{Y_i}{N_i} \sum_{j \in F_i} P(Y^*_j = 1|W), i \in I. \tag{8}
\]

We make the following rank condition assumption to achieve identification.

**Assumption 6.** \(\mathbb{E}\left[ \left( X^T_i, \frac{\sum_{j \in F_i} P(Y^*_j = 1|W)}{N_i} \right) \right] \) is with full rank \(d + 1\).

**Remark 6.** Assumption 6 requires no perfect collinearity of \(\left( X^T_i, \frac{\sum_{j \in F_i} P(Y^*_j = 1|W)}{N_i} \right) \). 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 \(W\) and the
distribution of the $\varepsilon$. 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 $\mu$ as

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

(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 Likelihood (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

$$
P(Y_i = 1 | Y_i^* = 0, \mathcal{W}) = P(Y_i = 1 | Y_i^* = 0) = \alpha \in (0, 1),
$$

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

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 relax 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

$$
P(Y_i = 1 | W; \theta) = \alpha + (1 - \alpha)P(Y^*_i = 1 | W; \mu),$$

$$
P(Z_i = 1 | W; \theta) = \delta + (1 - \delta)P(Y^*_i = 1 | W; \mu).$$

(10)

Our log likelihood function is formulated by the observed conditional distribution function $f(Y_i, Z_i | W; \theta)$. Let $P^* = (P^*_1, P^*_2, \cdots, P^*_n) = (P(Y^*_1 = 1 | W; \theta), P(Y^*_2 = 1 | W; \theta), \cdots, P(Y^*_n = 1 | W; \theta))$. With Equation (10), we have the log-likelihood function:

$$
\mathcal{L}(\theta, P^*) = \sum_{i \in I} \left\{ Y_i \log \left[ \alpha + (1 - \alpha)P^*_i \right] + (1 - Y_i) \log \left[ 1 - \alpha - (1 - \alpha)P^*_i \right] 
+ Z_i \log \left[ \delta + (1 - \delta)P^*_i \right] + (1 - Z_i) \log \left[ 1 - \delta - (1 - \delta)P^*_i \right] \right\}. 
$$

(11)

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

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

(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 $P = \Gamma(P, W; \mu)$ 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 $\theta$ 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 $\theta$ becomes a modified Logit regression, $\hat{\theta}^{(1)} = \arg \max_{\theta \in \Theta} \mathcal{L}(\theta, P^{(0)})$. The algorithm then updates the choice probabilities profile using $P^{(1)} = \Gamma(W, P^{(0)}; \hat{\theta}^{(1)})$ 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., $|\hat{\theta}^{(K+1)} - \hat{\theta}^{(K)}| < tol$. 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.

---

10Here 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 $W$, i.e., $P = \Gamma(P;\theta)$\textsuperscript{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:

$$L(\theta, P) = \frac{1}{n} \sum_{i \in I} L_i(\theta, P),$$

$$= \frac{1}{n} \sum_{i \in I} \left\{ Y_i \log \left[ \alpha + (1 - \alpha)P_i \right] + (1 - Y_i) \log \left[ 1 - \alpha - (1 - \alpha)P_i \right] 
+ Z_i \log \left[ \delta + (1 - \delta)P_i \right] + (1 - Z_i) \log \left[ 1 - \delta - (1 - \delta)P_i \right] \right\},$$

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

$$\tilde{\theta}_n(P) \equiv \arg\max_{\theta \in \Theta} L(\theta, P),$$

$$\phi_n(P) \equiv \Gamma(\tilde{\theta}_n(P), P),$$

$$L_0(\theta, P) \equiv \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left[ \sum_{i \in I} L_i(\theta, P) \right],$$

$$\tilde{\theta}_0(P) \equiv \arg\max_{\theta \in \Theta} L_0(\theta, P),$$

$$\phi_0(P) \equiv \Gamma(\tilde{\theta}_0(P), P).$$

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 $(\theta_0, P^*)$. The first order condition for the NPL estimation is

$$\frac{\partial L(\theta, \Gamma(P;\theta))}{\partial \theta} \bigg|_{(\theta, P) = (\theta_{\text{NPL}}, \hat{P}_{\text{NPL}})} = 0. \quad (13)$$

**Assumption 8.** (a) $\Theta$ is compact, $\theta_0$ is an interior point of $\Theta$, and $\mathcal{P}$ is a compact and convex subset of $(0,1)^n$; (b) $(\theta_0, P^*)$ 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 $\Lambda_0$; (c) $\frac{\partial^2 L_0(\theta, P_0)}{\partial \theta \partial \theta^T}$ is a nonsingular matrix in $\mathcal{N}$; (d) The operator $\phi(P) - P$ has a nonsingular

\textsuperscript{11}Here we abuse the notation to use $\theta$ instead of $\mu$ though $\alpha$ and $\delta$ 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

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

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

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

Remark 8. Since we are assuming the logistic error term, we can check that $\hat{\theta}_0(P)$ 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 $\theta$ 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 $\hat{\theta}_{NPL} \overset{p}{\to} \theta_0$ and

$$
\sqrt{n}(\hat{\theta}_{NPL} - \theta_0) \overset{d}{\to} N(0, V_{NPL}), \quad (14)
$$

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

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

$$Y^*_i = \begin{cases} 
\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \frac{Y}{N_i} \sum_{j \in F_i} \mathbb{E}(Y^*_j | \mathcal{W}) - \epsilon_i \geq 0 
\end{cases}$$  \hspace{1cm} (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**: $\theta_0 = (0.05, 0.05; -1, 1, -1, 1; 1)$

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{\alpha}_{NPL}$</th>
<th>$\hat{\delta}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
<th>$\tilde{\beta}_{NPL}$</th>
<th>$\tilde{\gamma}_{NPL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.002</td>
<td>0.001</td>
<td>-0.061</td>
<td>0.046</td>
<td>-0.033</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>(0.037)</td>
<td>(0.037)</td>
<td>(0.34)</td>
<td>(0.187)</td>
<td>(0.184)</td>
<td>(0.181)</td>
</tr>
<tr>
<td>1,000</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.002</td>
<td>0.011</td>
<td>-0.013</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.026)</td>
<td>(0.23)</td>
<td>(0.124)</td>
<td>(0.117)</td>
<td>(0.118)</td>
</tr>
<tr>
<td>2,000</td>
<td>0.001</td>
<td>0.001</td>
<td>-0.013</td>
<td>0.012</td>
<td>-0.010</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.019)</td>
<td>(0.16)</td>
<td>(0.087)</td>
<td>(0.087)</td>
<td>(0.082)</td>
</tr>
</tbody>
</table>

**Mean Square Errors**

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{\alpha}_{NPL}$</th>
<th>$\hat{\delta}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
<th>$\tilde{\beta}_{NPL}$</th>
<th>$\tilde{\gamma}_{NPL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.001</td>
<td>0.001</td>
<td>0.12</td>
<td>0.037</td>
<td>0.035</td>
<td>0.034</td>
</tr>
<tr>
<td>1,000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.05</td>
<td>0.016</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>2,000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.02</td>
<td>0.008</td>
<td>0.007</td>
<td>0.007</td>
</tr>
</tbody>
</table>

### Table 2: Experiment II

**True Parameters**: $\theta_0 = (0.1, 0.1; -1, 1, -1, 1; 1)$

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{\alpha}_{NPL}$</th>
<th>$\hat{\delta}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
<th>$\tilde{\beta}_{NPL}$</th>
<th>$\tilde{\gamma}_{NPL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.060</td>
<td>0.048</td>
<td>-0.035</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.046)</td>
<td>(0.38)</td>
<td>(0.207)</td>
<td>(0.205)</td>
<td>(0.200)</td>
</tr>
<tr>
<td>1,000</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.005</td>
<td>0.012</td>
<td>-0.015</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.032)</td>
<td>(0.25)</td>
<td>(0.135)</td>
<td>(0.128)</td>
<td>(0.128)</td>
</tr>
<tr>
<td>2,000</td>
<td>0.001</td>
<td>0.001</td>
<td>-0.020</td>
<td>0.014</td>
<td>-0.013</td>
<td>0.012</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.175)</td>
<td>(0.092)</td>
<td>(0.095)</td>
<td>(0.089)</td>
</tr>
</tbody>
</table>

**Mean Square Errors**

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{\alpha}_{NPL}$</th>
<th>$\hat{\delta}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
<th>$\tilde{\beta}_{NPL}$</th>
<th>$\tilde{\gamma}_{NPL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.002</td>
<td>0.002</td>
<td>0.14</td>
<td>0.045</td>
<td>0.043</td>
<td>0.042</td>
</tr>
<tr>
<td>1,000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.06</td>
<td>0.018</td>
<td>0.017</td>
<td>0.017</td>
</tr>
<tr>
<td>2,000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.031</td>
<td>0.009</td>
<td>0.009</td>
<td>0.008</td>
</tr>
</tbody>
</table>
Table 3: Experiment III

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{a}_{NPL}$</th>
<th>$\hat{\delta}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.078</td>
<td>0.057</td>
<td>-0.047</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.052)</td>
<td>(0.418)</td>
<td>(0.232)</td>
<td>(0.227)</td>
<td>(0.229)</td>
</tr>
<tr>
<td>1,000</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.020</td>
<td>0.017</td>
<td>-0.020</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.037)</td>
<td>(0.294)</td>
<td>(0.151)</td>
<td>(0.144)</td>
<td>(0.146)</td>
</tr>
<tr>
<td>2,000</td>
<td>0.001</td>
<td>0.001</td>
<td>-0.023</td>
<td>0.017</td>
<td>-0.015</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.026)</td>
<td>(0.206)</td>
<td>(0.104)</td>
<td>(0.106)</td>
<td>(0.104)</td>
</tr>
</tbody>
</table>

Mean Square Errors

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{a}_{NPL}$</th>
<th>$\hat{\delta}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.003</td>
<td>0.003</td>
<td>0.181</td>
<td>0.057</td>
<td>0.054</td>
<td>0.055</td>
</tr>
<tr>
<td>1,000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.087</td>
<td>0.023</td>
<td>0.021</td>
<td>0.021</td>
</tr>
<tr>
<td>2,000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.043</td>
<td>0.011</td>
<td>0.012</td>
<td>0.011</td>
</tr>
</tbody>
</table>

Table 4: Experiment IV

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{a}_{NPL}$</th>
<th>$\hat{\delta}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>-0.004</td>
<td>-0.004</td>
<td>-0.101</td>
<td>0.078</td>
<td>-0.066</td>
<td>0.069</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.059)</td>
<td>(0.529)</td>
<td>(0.281)</td>
<td>(0.281)</td>
<td>(0.274)</td>
</tr>
<tr>
<td>1,000</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.033</td>
<td>0.025</td>
<td>-0.027</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.042)</td>
<td>(0.363)</td>
<td>(0.185)</td>
<td>(0.178)</td>
<td>(0.178)</td>
</tr>
<tr>
<td>2,000</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.028</td>
<td>0.020</td>
<td>-0.019</td>
<td>0.019</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.028)</td>
<td>(0.250)</td>
<td>(0.125)</td>
<td>(0.124)</td>
<td>(0.123)</td>
</tr>
</tbody>
</table>

Mean Square Errors

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{a}_{NPL}$</th>
<th>$\hat{\delta}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.003</td>
<td>0.003</td>
<td>0.289</td>
<td>0.085</td>
<td>0.083</td>
<td>0.080</td>
</tr>
<tr>
<td>1,000</td>
<td>0.002</td>
<td>0.002</td>
<td>0.133</td>
<td>0.035</td>
<td>0.032</td>
<td>0.032</td>
</tr>
<tr>
<td>2,000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.063</td>
<td>0.016</td>
<td>0.016</td>
<td>0.016</td>
</tr>
</tbody>
</table>
Table 5: Experiment V

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{\alpha}_{NPL}$</th>
<th>$\hat{\delta}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
<th>$\bar{\beta}_{NPL}$</th>
<th>$\bar{\gamma}_{NPL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.003</td>
<td>0.004</td>
<td>-0.042</td>
<td>0.036</td>
<td>-0.033</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.041)</td>
<td>(0.331)</td>
<td>(0.176)</td>
<td>(0.168)</td>
<td>(0.163)</td>
</tr>
<tr>
<td>1,000</td>
<td>-0.003</td>
<td>-0.003</td>
<td>0.009</td>
<td>0.010</td>
<td>-0.013</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.031)</td>
<td>(0.240)</td>
<td>(0.118)</td>
<td>(0.117)</td>
<td>(0.116)</td>
</tr>
<tr>
<td>2,000</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.006</td>
<td>0.007</td>
<td>-0.008</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.023)</td>
<td>(0.165)</td>
<td>(0.084)</td>
<td>(0.080)</td>
<td>(0.078)</td>
</tr>
</tbody>
</table>

Mean Square Errors

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{\alpha}_{NPL}$</th>
<th>$\hat{\delta}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
<th>$\bar{\beta}_{NPL}$</th>
<th>$\bar{\gamma}_{NPL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.002</td>
<td>0.002</td>
<td>0.111</td>
<td>0.032</td>
<td>0.029</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.053)</td>
<td>(0.367)</td>
<td>(0.193)</td>
<td>(0.184)</td>
<td>(0.180)</td>
</tr>
<tr>
<td>1,000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.057</td>
<td>0.014</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.037)</td>
<td>(0.256)</td>
<td>(0.126)</td>
<td>(0.124)</td>
<td>(0.123)</td>
</tr>
<tr>
<td>2,000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.027</td>
<td>0.007</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.027)</td>
<td>(0.176)</td>
<td>(0.088)</td>
<td>(0.086)</td>
<td>(0.083)</td>
</tr>
</tbody>
</table>

Table 6: Experiment VI

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{\alpha}_{NPL}$</th>
<th>$\hat{\delta}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
<th>$\bar{\beta}_{NPL}$</th>
<th>$\bar{\gamma}_{NPL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>-0.002</td>
<td>-0.002</td>
<td>-0.032</td>
<td>0.034</td>
<td>-0.031</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.053)</td>
<td>(0.367)</td>
<td>(0.193)</td>
<td>(0.184)</td>
<td>(0.180)</td>
</tr>
<tr>
<td>1,000</td>
<td>-0.006</td>
<td>-0.005</td>
<td>0.012</td>
<td>0.009</td>
<td>-0.013</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.037)</td>
<td>(0.256)</td>
<td>(0.126)</td>
<td>(0.124)</td>
<td>(0.123)</td>
</tr>
<tr>
<td>2,000</td>
<td>-0.001</td>
<td>0.000</td>
<td>-0.009</td>
<td>0.007</td>
<td>-0.010</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.027)</td>
<td>(0.176)</td>
<td>(0.088)</td>
<td>(0.086)</td>
<td>(0.083)</td>
</tr>
</tbody>
</table>

Mean Square Errors

<table>
<thead>
<tr>
<th>n</th>
<th>$\hat{\alpha}_{NPL}$</th>
<th>$\hat{\delta}_{NPL}$</th>
<th>$\hat{\beta}_{NPL}$</th>
<th>$\hat{\gamma}_{NPL}$</th>
<th>$\bar{\beta}_{NPL}$</th>
<th>$\bar{\gamma}_{NPL}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.003</td>
<td>0.003</td>
<td>0.135</td>
<td>0.038</td>
<td>0.035</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.047)</td>
<td>(0.367)</td>
<td>(0.193)</td>
<td>(0.184)</td>
<td>(0.180)</td>
</tr>
<tr>
<td>1,000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.066</td>
<td>0.016</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.037)</td>
<td>(0.256)</td>
<td>(0.126)</td>
<td>(0.124)</td>
<td>(0.123)</td>
</tr>
<tr>
<td>2,000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.031</td>
<td>0.008</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.027)</td>
<td>(0.176)</td>
<td>(0.088)</td>
<td>(0.086)</td>
<td>(0.083)</td>
</tr>
</tbody>
</table>
Table 7: Experiment VII

True Parameters: \( \theta_0 = (0.2, 0.2; -1,1,-1,1; 2) \)

<table>
<thead>
<tr>
<th>n</th>
<th>( \hat{\alpha}_{NPL} )</th>
<th>( \hat{\delta}_{NPL} )</th>
<th>( \hat{\beta}_{NPL} )</th>
<th>( \hat{\gamma}_{NPL} )</th>
<th>( \hat{\beta}_{NPL} )</th>
<th>( \hat{\gamma}_{NPL} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>-0.005</td>
<td>-0.004</td>
<td>-0.043</td>
<td>0.039</td>
<td>-0.037</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.064)</td>
<td>(0.424)</td>
<td>(0.216)</td>
<td>(0.211)</td>
<td>(0.210)</td>
</tr>
<tr>
<td>1,000</td>
<td>-0.005</td>
<td>-0.006</td>
<td>0.014</td>
<td>0.011</td>
<td>-0.017</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.043)</td>
<td>(0.285)</td>
<td>(0.140)</td>
<td>(0.134)</td>
<td>(0.134)</td>
</tr>
<tr>
<td>2,000</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.011</td>
<td>0.010</td>
<td>-0.014</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.030)</td>
<td>(0.196)</td>
<td>(0.095)</td>
<td>(0.095)</td>
<td>(0.092)</td>
</tr>
</tbody>
</table>

Mean Square Errors

<table>
<thead>
<tr>
<th>n</th>
<th>( \hat{\alpha}_{NPL} )</th>
<th>( \hat{\delta}_{NPL} )</th>
<th>( \hat{\beta}_{NPL} )</th>
<th>( \hat{\gamma}_{NPL} )</th>
<th>( \hat{\beta}_{NPL} )</th>
<th>( \hat{\gamma}_{NPL} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.004</td>
<td>0.004</td>
<td>0.181</td>
<td>0.048</td>
<td>0.046</td>
<td>0.046</td>
</tr>
<tr>
<td>1,000</td>
<td>0.002</td>
<td>0.002</td>
<td>0.081</td>
<td>0.020</td>
<td>0.018</td>
<td>0.018</td>
</tr>
<tr>
<td>2,000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.038</td>
<td>0.009</td>
<td>0.009</td>
<td>0.124</td>
</tr>
</tbody>
</table>

Table 8: Experiment VIII

True Parameters: \( \theta_0 = (0.4, 0.4; -1,1,-1,1; 2) \)

<table>
<thead>
<tr>
<th>n</th>
<th>( \hat{\alpha}_{NPL} )</th>
<th>( \hat{\delta}_{NPL} )</th>
<th>( \hat{\beta}_{NPL} )</th>
<th>( \hat{\gamma}_{NPL} )</th>
<th>( \hat{\beta}_{NPL} )</th>
<th>( \hat{\gamma}_{NPL} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>-0.007</td>
<td>-0.006</td>
<td>-0.061</td>
<td>0.057</td>
<td>-0.049</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.073)</td>
<td>(0.527)</td>
<td>(0.257)</td>
<td>(0.250)</td>
<td>(0.247)</td>
</tr>
<tr>
<td>1,000</td>
<td>-0.004</td>
<td>-0.006</td>
<td>0.007</td>
<td>0.026</td>
<td>-0.024</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.047)</td>
<td>(0.362)</td>
<td>(0.170)</td>
<td>(0.167)</td>
<td>(0.167)</td>
</tr>
<tr>
<td>2,000</td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.011</td>
<td>0.010</td>
<td>-0.014</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
<td>(0.031)</td>
<td>(0.246)</td>
<td>(0.114)</td>
<td>(0.111)</td>
<td>(0.109)</td>
</tr>
</tbody>
</table>

Mean Square Errors

<table>
<thead>
<tr>
<th>n</th>
<th>( \hat{\alpha}_{NPL} )</th>
<th>( \hat{\delta}_{NPL} )</th>
<th>( \hat{\beta}_{NPL} )</th>
<th>( \hat{\gamma}_{NPL} )</th>
<th>( \hat{\beta}_{NPL} )</th>
<th>( \hat{\gamma}_{NPL} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>0.005</td>
<td>0.005</td>
<td>0.282</td>
<td>0.069</td>
<td>0.065</td>
<td>0.692</td>
</tr>
<tr>
<td>1,000</td>
<td>0.002</td>
<td>0.002</td>
<td>0.131</td>
<td>0.030</td>
<td>0.029</td>
<td>0.362</td>
</tr>
<tr>
<td>2,000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.061</td>
<td>0.013</td>
<td>0.013</td>
<td>0.173</td>
</tr>
</tbody>
</table>
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 \( \{F_{ij}\}_{i,j=1}^{n} \). 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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>15.882</td>
<td>1.187</td>
</tr>
<tr>
<td>Female</td>
<td>0.497</td>
<td>0.500</td>
</tr>
<tr>
<td>Parents’ Education†</td>
<td>5.257</td>
<td>2.459</td>
</tr>
<tr>
<td>White</td>
<td>0.092</td>
<td>0.289</td>
</tr>
<tr>
<td>American Indian</td>
<td>0.049</td>
<td>0.215</td>
</tr>
<tr>
<td>Asian</td>
<td>0.348</td>
<td>0.476</td>
</tr>
<tr>
<td>African American</td>
<td>0.265</td>
<td>0.442</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.385</td>
<td>0.487</td>
</tr>
<tr>
<td>Others*</td>
<td>0.130</td>
<td>0.336</td>
</tr>
<tr>
<td>Attitude(Y)</td>
<td>0.450</td>
<td>0.498</td>
</tr>
<tr>
<td>Attitude(Z)</td>
<td>0.471</td>
<td>0.499</td>
</tr>
</tbody>
</table>

†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 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

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

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 infor-
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


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: $P^1 = (P^1_1, P^1_2, \ldots, P^1_n) \neq (P^2_1, P^2_2, \ldots, P^2_n) = P^2$. We have

$$ |P^1_i - P^2_i| = |\Gamma_i(W, P^1; \mu) - \Gamma_i(W, P^2; \mu)|,$$

$$ = \left| \Lambda(X^T \beta + \frac{\gamma}{N} \sum_{j \in F_i} P^1_j) - \Lambda(X^T \beta + \frac{\gamma}{N} \sum_{j \in F_i} P^2_j) \right|,$$

$$ = \Lambda \left( X^T \beta + \frac{\gamma}{N} \sum_{j \in F_i} P^\dagger_j \right) \left[ 1 - \Lambda \left( X^T \beta + \frac{\gamma}{N} \sum_{j \in F_i} P^\dagger_j \right) \right] \left| \frac{\gamma}{N} \sum_{j \in F_i} (P^1_i - P^2_i) \right|,$$

$$ \leq \frac{1}{4} \cdot \gamma \cdot \max_{j \in I} |P^1_j - P^2_j| < 4 \cdot \frac{1}{4} \max_{j \in I} |P^1_j - P^2_j|,$$

$$ = \max_{j \in I} |P^1_j - P^2_j|,$$

where $P^\dagger_j$ is the probability between $P^1_j$ and $P^2_j$. 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 $i \in I$ on the left-hand side of Equation (16), we have

$$ \max_{i \in I} |P^1_i - P^2_i| < \max_{j \in I} |P^1_j - P^2_j|,$$

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 $L(\theta, P)$ in the NPL estimation. Define the function

$$ T(\theta, P) \equiv \max_{c \in \Theta} \left\{ L_0(c, P) \right\} - L_0(\theta, P).$$

Because $L_0(\theta, P)$ is continuous and $\Theta \times P$ is compact, Berge’s maximum theorem establishes that $T(\theta, P)$ is a continuous function. By construction, $T(\theta, P) \geq 0$ for any $(\theta, P)$.
Let $\mathcal{E}$ be the set of vectors $(\theta, P)$ that are fixed points of the equilibrium mapping $\Gamma$, i.e.,

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

Given that $\Theta \times P$ is compact and $\Gamma$ is continuous, $\mathcal{E}$ then is a compact set. By definition, the set $\Lambda_0$ 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 $\theta_0$. We then see that $B_{\epsilon}(\theta_0) \cap \mathcal{E}$ is also compact. Define the constant

$$
\tau \equiv \min_{(\theta, P) \in B_{\epsilon}(\theta_0) \cap \mathcal{E}} T(\theta, P) > 0.
$$

Define the event

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

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

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

and

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

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

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

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

$$
\left( \theta^{(n)}, P^{(n)} \right) \overset{p}{\rightarrow} (\theta_0, P^*).
$$

From the definition of $\Lambda_n$, we have that $\hat{\theta}_{NPL} \overset{p}{\rightarrow} \theta_0$. 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 \((\theta_0, P^*)\), we have

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

(18)

where \(\theta^+\) is between \(\hat{\theta}_{NPL}\) and \(\theta_0\) and \(P^+\) is between \(\hat{P}_{NPL}\) and \(P^*\) respectively. From Equation (18), we have that

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

(19)

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

\[
\sqrt{n} (\hat{\theta}_{NPL} - \theta_0) \overset{d}{\rightarrow} N(0, V_{NPL}),
\]

(20)

where

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

\[
\square
\]

**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

<table>
<thead>
<tr>
<th></th>
<th>2M</th>
<th>M1</th>
<th>M2</th>
<th>Logit models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Y$</td>
<td></td>
<td></td>
<td>$Z$</td>
</tr>
<tr>
<td>Age</td>
<td>-0.447*</td>
<td>-0.335*</td>
<td>-0.105</td>
<td>-0.337*</td>
</tr>
<tr>
<td></td>
<td>(0.172)</td>
<td>(0.056)</td>
<td>(0.057)</td>
<td>(0.056)</td>
</tr>
<tr>
<td>Female</td>
<td>0.241</td>
<td>0.187</td>
<td>0.287*</td>
<td>0.186</td>
</tr>
<tr>
<td></td>
<td>(0.176)</td>
<td>(0.122)</td>
<td>(0.131)</td>
<td>(0.121)</td>
</tr>
<tr>
<td>Parents’ Education</td>
<td>0.013</td>
<td>-0.003</td>
<td>-0.002</td>
<td>-0.003</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.027)</td>
<td>(0.028)</td>
<td>(0.027)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.672*</td>
<td>-0.474*</td>
<td>-0.249</td>
<td>-0.478*</td>
</tr>
<tr>
<td></td>
<td>(0.331)</td>
<td>(0.196)</td>
<td>(0.207)</td>
<td>(0.196)</td>
</tr>
<tr>
<td>Asian</td>
<td>0.039</td>
<td>0.066</td>
<td>0.070</td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>(0.278)</td>
<td>(0.200)</td>
<td>(0.215)</td>
<td>(0.198)</td>
</tr>
<tr>
<td>African American</td>
<td>-0.034</td>
<td>0.171</td>
<td>-0.248</td>
<td>0.168</td>
</tr>
<tr>
<td></td>
<td>(0.282)</td>
<td>(0.207)</td>
<td>(0.221)</td>
<td>(0.207)</td>
</tr>
<tr>
<td>Native American</td>
<td>-0.423</td>
<td>-0.258</td>
<td>-0.084</td>
<td>-0.259</td>
</tr>
<tr>
<td></td>
<td>(0.474)</td>
<td>(0.290)</td>
<td>(0.297)</td>
<td>(0.290)</td>
</tr>
<tr>
<td>Other</td>
<td>0.560</td>
<td>0.500*</td>
<td>-0.017</td>
<td>0.501*</td>
</tr>
<tr>
<td></td>
<td>(0.345)</td>
<td>(0.198)</td>
<td>(0.208)</td>
<td>(0.197)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.289*</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.105)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.573*</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Peer Effects ($\gamma$)</td>
<td>1.553*</td>
<td>0.003</td>
<td>0.466*</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.823)</td>
<td>(0.256)</td>
<td>(0.198)</td>
<td>—</td>
</tr>
<tr>
<td>Constant</td>
<td>5.879*</td>
<td>5.358*</td>
<td>2.343*</td>
<td>5.400*</td>
</tr>
<tr>
<td></td>
<td>(2.122)</td>
<td>(0.950)</td>
<td>(0.966)</td>
<td>(0.939)</td>
</tr>
</tbody>
</table>