Misclassification and the Hidden Silent Rivalry

Yingyao Hu
Johns Hopkins University

Zhongjian Lin
Emory University

2019 ES Winter Meeting

January 5, 2019

Observation

- There is peer effects among students on attitude towards learning (Silent Rivalry).
- Self reported attitude suffers from misclassification error (Overreport).
- Binary Choice Model with Social Interactions (and Misclassification).

Setting

- Students $\mathcal{I} \equiv \{1, 2, \cdots, n\}$.
- Each student i is associated with a friends set $F_i = \{j \in \mathcal{I} : F_{ij} = 1\}$ where $F_{ij} = 1$ denotes that student i considers student j as best friend. Denote N_i as the number of friends.
- Each student i is associated with demographic characteristics $X_i \in \mathbb{R}^d$ and a random utility shock ε_i.
- Students make attitude decisions $\{Y_i^*\}_{\mathcal{I}}$ simultaneously.
- The latent true attitude Y_i^* suffers from misclassification error due to social desirability of “diligence”.
- Two repeated measurements of Y_i^* are observed: Y_i and Z_i.
- Public information: $W_n \equiv \{X_i, F_i\}_{\mathcal{I}}$; private information ε_i.

Peer Effects in Attitude towards Learning

Utility of being diligent:

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

and utility of null-action is normalized as 0.

Utility Components

1. Deterministic utility: $X_i^T \beta$;
2. Deterministic social utility: $\frac{\gamma}{N_i} \sum_{j \in F_i} Y_j^*$;
3. Random utility shock: ε_i.

Based on "Incomplete Information" structure in the game, we have

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

Assumption

Assumption 1

(i) The private random utility terms ε_i's are i.i.d. across students and conform to the standard Logistic distribution; (ii) The strength of peer effects is moderate, i.e. $0 < \gamma < 4$.

Assumption 1 renders the choice probability as

$$P(Y_i^* = 1|W_n) = \frac{exp\left(X_i^T\beta + \frac{\gamma}{N_i}\sum_{j\in F_i} P(Y_j^* = 1|W_n)\right)}{1 + exp\left(X_i^T\beta + \frac{\gamma}{N_i}\sum_{j\in F_i} P(Y_j^* = 1|W_n)\right)} \equiv \Gamma_i(W_n, P^*), \quad (3)$$

where $P^* \equiv (P_1^*, \cdots, P_n^*) \equiv \left(P(Y_1^* = 1|W_n), \cdots, P(Y_1^* = 1|W_n)\right)$ is the equilibrium probabilities profile.

Lemma 1

With assumption 1 hold, the Bayesian Nash game in Equation (3) has a unique equilibrium.

Misclassification on Latent Attitude

Assumption 1 leads to the identification of $P(Y^* = 1|W_n)$ from the data if Y^*s are accurately observed. While the attitude may suffers from the misclassification error due to social desirability of "diligence".

Repeated Measurements

In the *The National Longitudinal Study of Adolescent Health (Add Health)* dataset, there are two identical questions [Skipped school without an excuse] from at-home survey and in-school survey which serve as the two repeated measurement of attitude towards learning. We denote the answers from the at-home and in-school surveys as Y and Z.

Assumption on Measurement Errors

Assumption 2: Conditional Independence

(Y, Z) are jointly independent conditional on Y^* and W

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

Assumption 2 is standard in the nonlinear measurement error literature, e.g. Hu (2008), Hu and Schennach (2008), Hu (2017). Conditional independence means that the repeated measurements provides no extra useful information other than those embedded in the true latent attitude. We denote the misclassification from "negative" to positive attitude as "desired misclassification" and the reverse one as "evasive misclassification". Due to the social desirability of "diligence", we make the following assumption regarding the evasive misclassification.

Assumption 3

Students do not underreport their positive attitude, i.e.,

$$P(Y = 0|Y^* = 1, W) = P(Z = 0|Y^* = 1, W) = 0.$$

Our identification of the parameter of interest is in two step. First, we identify the conditional distribution of the latent attitude, $P(Y^*|W)$ from the observables. Second, we identify the parameter of interest, $\mu \equiv (\beta^T, \gamma)^T$, in a constructive way. Define

$$M_{Y,Z|W} \equiv \begin{pmatrix} f_{Y,Z|W}(0,0|w) & f_{Y,Z|W}(0,1|w) \\ f_{Y,Z|W}(1,0|w) & f_{Y,Z|W}(1,1|w) \end{pmatrix} \equiv [f_{Y,Z|W}(i-1,j-1|w)]_{i,j}.$$

Similarly, we define $M_{Y|Y^*,W} = [f_{Y|Y^*,W}(i-1|j-1,w)]_{i,j}$, $M_{Z|Y^*,W} = [f_{Z|Y^*,W}(i-1|j-1,w)]_{i,j}$, $M_{Y,Y^*|W} = [f_{Y,Y^*|W}(i-1,j-1|w)]_{i,j}$ and $M_{Z,Y^*|W} = [f_{Z,Y^*|W}(i-1,j-1|w)]_{i,j}$. Notice that the latter two matrices are lower triangular matrices. Denote

$$D_{Y^*|W} \equiv \begin{pmatrix} f_{Y^*|W}(0|w) & 0 \\ 0 & f_{Y^*|W}(1|w) \end{pmatrix} = \left[f_{Y^*|W}(i|w)\right]_i.$$

Theorem

Theorem

With assumption 1-3 hold, we identify the conditional distribution of the latent attitude, i.e. $D_{Y^*|W}$.

By the total law of probability and conditional independence in assumption 2, we have

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

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

With condition on the evasive misclassification probabilities (assumption 3), $M_{Y|Y^*,W}$, $M_{Z,Y^*|W}$, $M_{Z|Y^*,W}$ and $M_{Y,Y^*|W}$are lower triangular matrices. The point identification of these unknown matrices is feasible through the so-called LU decomposition.

Then the conditional distribution of the latent attitude is identified through:

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

Estimation

We use the Nested Pseudo Likelihood (NPL) algorithm to estimate the silent rivalry in attitude towards learning. Before we proceed to the details of the NPL estimator, we make the following simplification assumption

Assumption 5

The misclassification probabilities satisfy

$$P(Y_i = 1|Y_i^* = 0, W_n) = P(Y_i = 1|Y_i^* = 0) \equiv \alpha,$$
$$P(Z_i = 1|Y_i^* = 0, W_n) = P(Z_i = 1|Y_i^* = 0) \equiv \delta.$$

Assumption 5 reduces the number of unknown in the misclassification probabilities. Theoretically, this assumption is not necessary, but is introduced to make the empirical analysis feasible given the sample size. Hausman, Abrevaya and Scott-Morton (1998) make the same assumption to construct the likelihood function. Define $\theta \equiv (\alpha, \delta, \mu^T)^T$. With assumption 5, we have

$$\begin{aligned} P(Y_i = 1|W_n; \theta) &= \alpha + (1-\alpha)P(Y_i^* = 1|W_n; \mu) \\ P(Z_i = 1|W_n; \theta) &= \delta + (1-\delta)P(Y_i^* = 1|W_n; \mu) \end{aligned} \tag{8}$$

Likelihood

We construct our log likelihood function as

$$
\begin{aligned}
\mathcal{L}(\theta, P^*) &= \sum_{i=1}^{n} log\left[f_{Y,Z|W}(1,1|w) \cdot f_{Y,Z|W}(1,0|w) \cdot f_{Y,Z|W}(0,1|w) \cdot f_{Y,Z|W}(0,0|w)\right] \\
&= \sum_{i=1}^{n} \Big\{ Y_i \log\left[\alpha + (1-\alpha)P_i^*\right] + (1-Y_i)\log\left[1-\alpha-(1-\alpha)P_i^*\right] \\
&\qquad + Z_i \log\left[\delta + (1-\delta)P_i^*\right] + (1-Z_i)\log\left[1-\delta-(1-\delta)P_i^*\right] \Big\},
\end{aligned} \tag{9}
$$

and the responding pseudo log likelihood function as

$$
\begin{aligned}
\mathcal{L}(\theta, P) = \sum_{i=1}^{n} \Big\{ & Y_i \log\left[\alpha + (1-\alpha)\Gamma_i(W_n, P)\right] + (1-Y_i)\log\left[1-\alpha-(1-\alpha)\Gamma_i(W_n, P)\right] \\
& + Z_i \log\left[\delta + (1-\delta)\Gamma_i(W_n, P)\right] + (1-Z_i)\log\left[1-\delta-(1-\delta)\Gamma_i(W_n, P)\right] \Big\},
\end{aligned} \tag{10}
$$

where P can be arbitrary choice probabilities profile.

The NPL algorithm is as follows:

1. Start with a conjecture $P^{(0)}$ and the maximization of $\mathcal{L}(\theta, P)$ with respect to θ becomes the modified Logit estimation. Denote the estimate as $\hat{\theta}^{(1)}$.
2. Update the choice probabilities profile using $P^{(0)}$ and $\hat{\theta}^{(1)}$ through the best response functions $\Gamma \equiv (\Gamma_1, \cdots, \Gamma_n)$ defined in eq. (3). Denote the new choice probabilities profile as $P^{(1)}$.
3. Repeat step 1 and 2 until the distance between estimates in two consecutive iterations is less than a preset tolerance, e.g. $|\hat{\theta}^{(K+1)} - \hat{\theta}^{(K)}| < 10^{-6}$.

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. In the *Add Health* dataset, each student has nominations of at most five male friends and at most five female friends, from which we construct the school network with direct links $[\{F_{ij}\}_{i,j=1}^{n}]$. *Add Health* dataset also include questionnaires for demographic characteristics such as age, parents' education, race information, gender, etc.

Table 1: 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

When it comes to the silent rivalry, we have three options to back out the peer effects parameter. We can either take Y or Z as the true latent attitude to estimate the interaction-based model 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 2, models without misclassification correction either fails to detect significant silent rivalry ($\hat{\gamma} = 0$ in model M1) or underestimates the peer effects ($\hat{\gamma} = 0.482$ in model M2).
Our 2M model estimates a significant 1.543 peer effects parameter which is as three times bigger than the model with In-School measurement. We also provide results for simple Logit model without simultaneous peer effects in attitude towards learning. The results are very similar for demographic covariates, e.g. older students pay more attention to study as they turn more matured. Furthermore, there are a large proportion of students over report their attitude, 25.6% at home and 28.9% in school.

Main Result

Table 2: Estimation Results

	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)

significances of α, δ and γ obtain from the one–sided test.

Conclusion

1. We provide identification and estimation of the binary choice model with misclassification and social interactions.
2. We find significant silent rivalry among students in attitude towards learning which is either hidden or underestimated if omitting the misclassification errors.
3. A large proportion of students overreport their attitude due to the social desirability of “diligence”.