Social Networks with Heterogeneity and Group Fixed Effects: A Likelihood Approach

Yanli Lin, Yang Yang

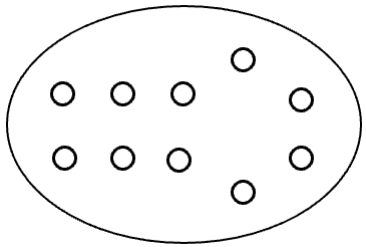
Department of Economics
The Ohio State University

EWMES

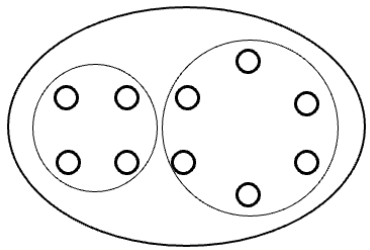
December 15, 2021

Motivation

- Empirical driven: difficult task to estimate social interaction effects
  - "reflection problem" (Manski, 1993); separation of social effects from other confounding effects; data limitation
  - SAR model

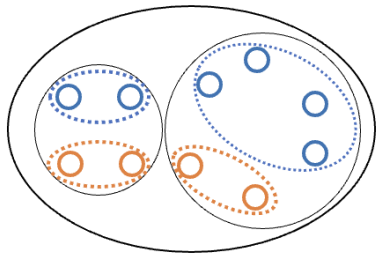


Lee (2007)



Lin (2010)

- **Research question**: How to identify heterogeneous endogenous and exogenous social interaction effects with group-level common factors?



  - application: heterogeneous peer effects in student academic achievement

Empirical Specification

$$y_i = \sum_{g\in\{F,M\}} \sum_{p\in\{F,M\}} d_{g,i}\lambda_{g,p}\bar{y}_{-i,p,c_i} + \sum_{g\in\{F,M\}} \sum_{p\in\{F,M\}} d_{g,i}\bar{x}'_{-i,p,c_i}\gamma_{g,p} + x_i'\beta_1 + t'_{c_i}\beta_2 + \alpha_{s_i} + u_i$$

- y_i: student i's academic achievement
- $d_{g,i}$: dummy variable for i's gender, $d_{F,i} + d_{M,i} = 1$
- $\bar{y}_{-i,p,c_i} = \begin{cases} \text{average achievement of female classmates} & \text{if } p = F \\ \text{average achievement of male classmates} & \text{if } p = M \end{cases}$
- $\bar{x}_{-i,p,c_i} = \begin{cases} \text{average characteristics of female classmates} & \text{if } p = F \\ \text{average characteristics of male classmates} & \text{if } p = M \end{cases}$

Empirical Specification

$$y_i = \sum_{g\in\{F,M\}} \sum_{p\in\{F,M\}} d_{g,i}\lambda_{g,p}\bar{y}_{-i,p,c_i} + \sum_{g\in\{F,M\}} \sum_{p\in\{F,M\}} d_{g,i}\bar{x}'_{-i,p,c_i}\gamma_{g,p} + x_i'\beta_1 + t'_{c_i}\beta_2 + \alpha_{s_i} + u_i$$

- $\{\lambda_{F,F}, \lambda_{M,M}\}$ and $\{\gamma_{F,F}, \gamma_{M,M}\}$: within gender peer and contextual effects; $\{\lambda_{M,F}, \lambda_{F,M}\}$ and $\{\gamma_{M,F}, \gamma_{F,M}\}$: cross gender peer and contextual effects
- x_i: student i's pre-determined characteristics
- t_{c_i}: characteristics of the head teacher of class c_i
- α_{s_i}: effects of common variables, identical for all students in the same group
- parameters of interest: $\theta_0 = (\underbrace{\lambda_{F,F,0}, \lambda_{M,F,0}, \lambda_{F,M,0}, \lambda_{M,M,0}}_{\Lambda_0}, \underbrace{\gamma_{F,F,0}, \gamma_{M,F,0}, \gamma_{F,M,0}, \gamma_{M,M,0}, \beta_{1,0}, \beta_{2,0}}_{\varphi_0}, \sigma_0^2)'$

Empirical Specification

In Matrix/Vector Form:

$$Y_n = \sum_{g\in\{F,M\}} \sum_{p\in\{F,M\}} \lambda_{g,p} H_g W_{p,n} Y_n + \sum_{g\in\{F,M\}} \sum_{p\in\{F,M\}} H_g W_{p,n} X_n \gamma_{p,g}$$
$$+ X_n\beta_1 + T_n\beta_2 + \sum_{s=1}^{S} \alpha_s h_s + u_n$$

- $Y_n = (y_1, \cdots, y_n)'$, $H_g = \text{diag}\{d_{g,1}, \ldots, d_{g,n}\}$, $X_n = (x_1, \cdots, x_n)'$, $T_n = (t_{c_1}, \cdots, t_{c_n})'$, $u_n = (u_1, \cdots, u_n)'$, h_s is the dummy variable for each school s
- $W_{F,n}$, $W_{M,n}$: row normalized $n \times n$ matrix with each element

$$w_{ij,F,n} = \begin{cases} 1/\left(|F_{c_i}| - 1\right) & c_i = c_j \\ 0 & c_i \neq c_j \end{cases} \text{ and } w_{ij,M,n} = \begin{cases} 1/\left(|M_{c_i}| - 1\right) & c_i = c_j \\ 0 & c_i \neq c_j \end{cases}$$

  - $|F_{c_i}|$: # of female students in class c_i; $|M_{c_i}|$: # of male students in class c_i

Maximum likelihood estimation

- CML (Lee, 2007) and "de-group-mean" (Lin, 2010) approaches can not be applied w/ heterogeneity
- direct estimation approach (joint estimation of θ_0 and group fixed effects) can be used
- concentrate out as many parameters, e.g. group fixed effects $\alpha = (\alpha_1, \cdots, \alpha_S)'$ and linear parameters φ, as possible for numerical search
  - concentrated log-likelihood function

$$Q_n(\Lambda) = -\frac{n}{2}[1 + \ln(2\pi)] - \frac{n}{2}\ln\left[\frac{1}{n}Y_n' C_n(\Lambda) Y_n\right] + \ln|S_n(\Lambda)|$$

  - asymptotic bias exists due to $E\left[\frac{\partial Q_n(\Lambda_0)}{\partial \lambda_{g,p}}\right] \neq 0$, need bias corrections

Data

China Education Panel Survey (CEPS)

- The first large-scale, nationwide, and longitudinal survey dataset
  - different questionnaires for students, parents, teachers, and school administrators
  - Wave I, surveys 19,487 students from Grade 7 and Grade 9 in 438 classrooms of 112 schools in 28 county-level units in mainland China in 2013-2014 academic year
  - Wave II, follow-up survey for 10,279 students in Grade 8 in 2014-2015 academic year
- Select students who appear in both Waves, random assigned classrooms
  - "group": a grade level (Grade 8) in the same school
  - construct networks based on gender subgroups within a classroom
- Final sample: 3,944 students across 97 classrooms in 56 schools
  - on average, 41 students in each class with a minimum of 14, a maximum of 74 and a standard deviation 13.02

Variables

Heterogeneity	students' gender	"0" for "male student"; "1" for "female student"
Dependent variables	Chinese score Mathematics score English score Total score	standardized for all classes within the same school with mean 70 and standard deviation 10
Students' characteristics	only child in family	"0" for "no"; "1" for "yes"
	relative age	± month, compared with sample median value (February, 2001)
	minority	"0" for "no"; "1" for "yes"
	local resident	"0" for "no"; "1" for "yes"
	attended kindergarten	"0" for "no"; "1" for "yes"
	parents' education	"1" for "none"; "2" for "finished elementary school"; "3" for "junior high school degree"; "4" for "technical secondary school or technical school degree"; "5" for "vocational high school degree"; "6" for "senior high school degree"; "7" for "junior college degree"; "8" for "bachelor's degree"; "9" for "master's degree or higher"
Head teachers' information	female head teacher	"0" for "no"; "1" for "yes"
	teaching experience	year
	whether teach Chinese/ Mathematics/English	"0" for "no"; "1" for "yes"

Summary statistics

Table 11: Summary Statistics for Female and Male Students Subgroups

	Female		Male	
	Mean	SD	Mean	SD
A. Outcome Variables:				
Chinese Score	73.44	7.68	67.93	10.24
Math Score	71.43	9.35	70.00	10.11
English Score	73.17	8.75	68.22	9.95
Total Score	76.72	16.31	71.68	16.97
B. Student's characteristics:				
Only child in family	.48	.50	.55	.50
Relative age (month)	.60	7.10	1.88	7.41
Minority	.10	.30	.09	.29
Local resident (Hukou System)	.78	.41	.76	.43
Attend kindergarten	.85	.36	.84	.36
Father's Education	4.48	2.08	4.35	2.04
Mother's Education	4.11	2.03	4.09	2.06
C. Head teacher's characteristics:				
Female head teacher	.72	.45	.70	.46
Teaching experience of head teacher (year)	14.55	7.86	14.23	7.93
Chinese head teacher	.30	.46	.29	.45
Math head teacher	.32	.47	.32	.47
English head teacher	.26	.44	.25	.43
Sample Size: 3893, including 1890 females and 2003 males[7]				

Peer effects

Chinese				Mathematics			
Female Classmates		Male Classmates		Female Classmates		Male Classmates	
Female	Male	Female	Male	Female	Male	Female	Male
.7734***	.5315***	.5125***	.4112***	.3989***	.2482***	.5652***	.4683***
(.0854)	(.0666)	(.1088)	(.1008)	(.0730)	(.0839)	(.0988)	(.0967)
English				Total			
Female Classmates		Male Classmates		Female Classmates		Male Classmates	
Female	Male	Female	Male	Female	Male	Female	Male
.6140***	.4489**	.7130***	.4762	.6645***	.5055***	.4367***	.3449***
(.1197)	(.1781)	(.1606)	(.3031)	(.0854)	(.0731)	(.1107)	(.0949)

- significant heterogeneous gender peer effects from female/male classmates
- female students are more subject to female/male peers' average achievement
- female peers higher impact in total and Chinese scores, male peers higher in Math
- within-gender effects > cross-gender effects for female students (total score)

Contextual effects

		Female Classmates		Male Classmates	
		Female	Male	Female	Male
Chinese	Relative age	.1823 (.1702)	-.0467 (.1675)	-.3581* (.1962)	.1224 (.1901)
	Attend kindergarten	-3.0196 (3.6818)	3.9821 (3.5969)	-6.3429* (3.6933)	1.3857 (3.6009)
Mathematics	Relative age	.2740 (.1806)	-.0730 (.1790)	-.4849** (.2123)	.0683 (.2015)
	Local resident	4.1093 (2.8141)	-2.2392 (2.8532)	-5.5803* (3.0723)	5.1577* (2.8911)
	Mother's Education	-.4686 (1.0204)	-.6195 (1.0204)	2.2530* (1.0587)	-.9135 (1.0423)
English	Relative age	.3245* (.1732)	-.0599 (.1688)	-.2713 (.2212)	.2511 (.2531)
	Local resident	3.8016 (2.6681)	-1.8771 (2.8193)	-8.1032** (2.9436)	3.7163 (3.0092)
	Mother's Education	-.0599 (.9700)	-.8382 (1.0413)	1.2656 (1.0477)	-.4033 (1.1543)
Total	Relative age	.4468** (.1872)	-.2151 (.1848)	-.3330 (.2096)	.0388 (.2001)
	Local resident	4.6416 (2.9250)	-5.5168* (2.9339)	-5.4366* (3.0552)	7.4535** (2.8492)
	Father's Education	-.7261 (.8840)	.7579 (.9119)	-1.6946* (.9415)	.3593 (.9040)
	Mother's Education	-.3528 (1.0699)	-.5139 (1.1102)	1.8081* (1.0876)	-.5879 (1.0458)

- relative age: competitive effects (Chinese and Math) vs. complementary effects (English and Total)
- mother's education: Chung and Zou (2020) find that higher classmates' maternal education raises students test score, we detect the specific channel (male classmates → females)

Characteristics of students and head teachers

		Chinese	Mathematics	English	Total
Student	Only child in family	.5946*	.7728***	.5065	.8019**
		(.3486)	(.3702)	(.3534)	(.3762)
	Relative age	-.0627***	-.1142**	-.0993***	-.1201***
		(.0233)	(.0247)	(.0236)	(.0251)
	Attend kindergarten	1.6937***	1.6219***	1.6241***	1.8566***
		(.4192)	(.4425)	(.4256)	(.4499)
	Father's Education	.2367**	.2782**	.4416***	.3761***
		(.1008)	(.1071)	(.1034)	(.1088)
	Mother's Education	.2983***	.1610	.2304**	.2504**
		(.1041)	(.1105)	(.1042)	(.1124)
Head teacher	Female	-0.1180	.4767	-.3119	.2077
		(.6354)	(.7766)	(1.4381)	(.6936)
	Teaching experience	.0435	.0477	.0232	.0411
		(.0414)	(.0439)	(.0484)	(.0451)
	Teach relative course	.0803	.3344	-.4607	.0652
		(.5618)	(.5857)	(1.8734)	(.8861)

- all students' characteristics show positive impact, except relative age
- not significant influence of head teachers' characteristics

Estimation Results

Students' characteristics-relative age

Table 13a: Correlation Coefficients Between Grades and Relative Ages

	All Range	Delayed Range (Relative Age>5)	Regular Range (-6≤Relative Age≤5)	Earlier Range (Relative Age<-6)
Chinese	-.0885	-.0224	.0013	-.0603
Mathematics	-.0886	-.0200	-.0333	-.1298
English	-.1124	-.0826	-.0150	-.0764
Total	-.1990	-.0576	-.0275	-.0766
# observations	3893	746	2768	379

Table 13b: Average Grades for Different Age Ranges

	All Range	Delayed Range (Relative Age>5)	Regular Range (-6≤Relative Age≤5)	Earlier Range (Relative Age<-6)
Chinese	70.6049	68.8624	70.8442	72.2879
Mathematics	70.6940	69.3231	70.8073	72.5650
English	70.6258	68.7782	70.8314	72.7606
Total	74.1286	66.0787	75.9426	76.7250
# observations	3893	746	2768	379

- relative age is a good approximate for whether a student has repeated or skipped a grade in primary school (noisy with lots of missing values)
  - a cutoff date regulating the precise age for entry into primary school in China
- regular range group has low correlation, other two groups have higher negative correlations
- delayed range group has lower average grades, while earlier range group is the opposite

Conclusion

- Consider higher-order spatial autoregressive models with group effects to confront some conceptual problems in social interaction estimation
  - separately identify heterogeneous peer and contextual effects
  - disentangle peer effects from other confounding effects
- Significant gender disparities in peer effects from subgroups in a classroom, provide justification for related policy intervention
- Limitation: the group fixed effect model cannot deal with possible unobservable factors in common within groups