IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15070

Heterogeneous Peer Effects under Endogenous Selection: An Application to Local and Migrant Children in Elementary Schools in Shanghai

Yuanyuan Chen
Shuaizhang Feng
Chao Yang

FEBRUARY 2022

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15070

Heterogeneous Peer Effects under Endogenous Selection: An Application to Local and Migrant Children in Elementary Schools in Shanghai

Yuanyuan Chen
Shanghai University of Finance and Economics

Shuaizhang Feng
Jinan University and IZA

Chao Yang
Shanghai University of Finance and Economics

FEBRUARY 2022

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Heterogeneous Peer Effects under Endogenous Selection: An Application to Local and Migrant Children in Elementary Schools in Shanghai*

This paper develops a model that allows for heterogenous contemporaneous peer effects among different types of agents who are endogenously selected into different peer groups. Using our framework, we characterize the reduced-form coefficient in the peer effect literature and show that it is a priori ambiguous in sign. We apply our approach to migrant and local students in Shanghai, where local students all go to public schools, but migrant students are endogenously selected into either public schools or lower-quality private schools. The results suggest large contemporaneous peer effects among all student groups. We conduct policy experiments to examine the effect of transferring migrant students from private schools to public schools. We show that peer effect can be substantially more important than the school effect in accounting for the total treatment effect of moving to better schools.

JEL Classification: C31, C34, I21

Keywords: peer effects, sample selection, education, migrant children

Corresponding author:
Shuaizhang Feng
School of Economics and Institute for Economic and Social Research
Jinan University
Guangzhou 510632
China
E-mail: shuaizhang.feng@jnu.edu.cn

* Chen acknowledges support from the National Natural Science Foundation of China (No. 71773074). Feng acknowledges support from the National Natural Science Foundation of China (No. 72073052). Yang owes thanks to the National Natural Science Foundation of China (No. 72173077, No.71703092, No.71833004) for financial support. We thank participants at the 2019 China meeting of the Econometric Society, IESR Conference on peer effects, the UCL seminar, and the seminar at National School of Development at Peking University, as well as Lars Lefgren and Kevin Lang for very helpful comments. All errors are our own.

1 Introduction

There is a large and growing literature investigating the spillover effects of one type of students, such as immigrants, females, ethnic minorities, disruptive kids, on their peers in the same classrooms (Ohinata et al., 2013, Carrell et al., 2018, Gong et al., 2021, Bifulco et al., 2011). In this literature, the most standard reduced-form specification is to regress the outcomes of the "affected" students on the proportion of "affecting" peers in a classroom, with a negative and significant coefficient being interpreted as evidence for the detrimental impact of the "affecting" group on the rest of the class. However, by focusing only on the effect of the "affecting" students on the "affected" students, this specification implicitly assumed away all the other peer interactions, including the influence of the "affected" students on the "affecting" ones, as well as interactions within each student type group. Moreover, these "reduced-form" estimates are less informative in answering policy questions of student regrouping when peer groups are endogenously formed. In this paper, we extend this literature by developing a more general framework that takes account of endogenous peer group formation and fully incorporates interactions among all students of different types. We characterize the reduced-form coefficient in our framework, and show that its sign is a priori ambiguous, as evidenced by the conflicting findings reported in the existing studies.

The specific empirical question we tackle in this paper is how domestic migrant students and local students affect each other in Shanghai's elementary school classrooms. Similar to the situation in many developed countries where schools have witnessed significant inflows of immigrant children, China's large cities have also received very substantial numbers of internal migrant students due to the country's rapid urbanization process. Because public schools are capacity-constrained, many lower-quality private schools are established to cater to migrant children's educational needs (Chen and Feng, 2013, 2017, 2019). Our data set contains students from both types of schools in Shanghai, including private schools with only migrant students and public schools with both local and migrant students. There is clear endogenous selection for migrant students in terms of their school choices, as they generally prefer public schools. Our aim is to evaluate the simultaneous interactions among local and migrant students in public schools and among migrant students in private schools, given the endogenously formed peer groups, i.e., classrooms in different types of schools.

Our framework builds on a theoretical model of interactions on the basis of a simultaneous-move game under incomplete information. Agents first select into different peer groups without considering other agents' choices. Then, conditional on knowing the types and characteristics of all its peers, an agent chooses its optimal response in a game-theoretical setting. In the equilibrium, all outcomes of the agents are jointly determined. Therefore, the model allows heterogenous peer interactions in a very flexible way, as all pre-determined characteristics included in the outcome equation also affect other agents' outcomes. Econometrically, by incorporating a social interaction model with a switch regression, we build a framework allowing for migrant students to first choose between public schools and migrant-only private schools. Once school types are chosen, students are (conditionally) randomly assigned to different classrooms, in which heterogeneous peer effects conditional on student type are fully incorporated. We propose a nested-fixed point 2-stage Nonlinear Least Square (NLS) estimator, which is shown to be consistent and asymptotically normal. We show that our model and estimation methods provide robust estimates for the effects from different types of classmates (peers), controlling for peer heterogeneity and school type selection.

We use data collected in 2012 on fifth grade students in Shanghai's elementary schools for our empirical work. In the first stage, we examine migrant students' choices between public and private schools. The excluded variable is parental residence prior to 2008 when the students in our sample started elementary school. In 2008, the Shanghai government unexpectedly enacted a policy that has substantially changed the geographical distribution of migrant-only private schools which explicitly cater to migrant children. Before 2008, these private schools were distributed in all districts of Shanghai, but they were only allowed to operate in the peripheral districts since then (Chen and Feng, 2013). Due to lack of mobility in residence, migrant children whose family resided in inner districts before the policy change were much more likely to enroll in a public school than otherwise. On the other hand, local children always go to public schools which are widely perceived as of higher quality.

A remaining identification challenge is the selection of schools and classrooms once school types are determined. We first include school fixed effects to control for quality differences among different schools. Within a school, the key concern is whether students (as well as teachers) are assigned to different classrooms exogenously, i.e., whether classmate relationships are uncorrelated with individual unobservables conditional on observed school and personal characteristics. Although Chinese elementary schools usually stick strictly to random class assignment as a general rule, there are some segregated classrooms in some public schools in our sample for historical reasons.[1] To play on the safe side, we treat segregated classrooms with only migrant students or local students as separate "pseudo" schools and only use variations among classes in the same "pseudo" schools.[2] We perform multiple tests and cannot reject the hypothesis of random assignments of students and teachers.

We then estimate the second-stage model with heterogeneous peer interactions, and derive significantly positive effects among all peer relations. We can not reject the hypothesis that all peer effect coefficients are the same. The inverse mills ratios in the regressions are statistically significant, bespeaking the importance of accounting for endogenous selection of school types. We perform placebo tests using simulated data where students are reshuffled into different classrooms or schools. Estimates from these placebo tests are not significantly different from zero, which suggest that our model does capture interactions among students in the same classrooms rather than something else.

Our results are robust when we also directly control for some key observed peer characteristics, which suggests that the estimated peer effect parameters mainly capture contemporaneous interactions among students. Although most of the empirical literature focuses on the spillover effects from pre-determined characteristics, contemporaneous peer interaction entails a "multiplier effect" and is potentially more important in affecting outcomes, as discussed in Fruehwirth (2014).

Lastly, we conduct policy experiments that transfer migrant students from a migrant-only private school to a public school in the same district. To keep the classroom size unchanged, we gradually replace students in the public school with migrant students from the private school. We find that migrant students transferred to the public school would initially experience substantial test score gains. But as more and more students from the private school are moved in, their test score gains gradually decrease. The students who are originally enrolled in the public

[1] For example, when a private school is closed by the government, typically all existing students will be transferred to a nearby public school, and these students usually form a migrant-only class. We discuss the related background information more in the next section.

[2] In our empirical work, we also try alternative ways of treating these segregated classes, and the results are qualitatively similar.

school and remain there suffer losses in test scores as qualities of their peers decline. Finally, we decompose the test score gains for transferred migrant students into three components, one due to changes in school quality, another due to changes in the quality of peers, and the third due to changes in the intensity of peer interactions. We show that the results are mainly driven by improved peer qualities. School quality effect is also positive but relatively small. On the other hand, peer intensity effects are negative as student peer influences are in general larger in private schools than in public schools.

While this paper speaks directly to the "reduced form" literature on peer effects as we noted at the very beginning, it is also closely related to the studies on social interaction models. Models by Manski (1993), Brock and Durlauf (2001), Lee et al. (2014), and Yang and Lee (2017) analyze the influence from a socially associated agent under incomplete information with the social relationship between any two agents exogenously given. We build on that literature by considering the endogenous sorting of agents into different peer groups (classes) before interactions begin. A related literature considers peer interactions in smaller social groups within classes, such as Leung (2015), Sheng (2020), and Hsieh and Lee (2016). In these studies, only endogenously formed friends interact with each other within a larger classroom. In principal, one can extend our framework to consider both interactions among friends as well as class-wide peer interactions. Finally, Hoshino (2019) studies endogenous missing outcomes in a social interaction model under incomplete information. While Hoshino models ex post selection of outcomes as a result of peer group formations and peer interactions, we model the ex ante formation of such peer groups.

The empirical approach proposed in this paper is applicable to all settings in which peer groups are formed prior to peer interactions, including many scenarios in the field of education. In tracking, students are endogenously sorted into different classrooms based on their academic abilities before peer interactions take place. Similarly, one may consider issues such as residential segregation in which parental choice of different school districts based on family background happens before peer effects kick in. Without explicitly modeling the process of peer group formation, one cannot answer important counterfactual policy questions of student regrouping, and analyze their welfare implications. Of course, similar to this paper, additional identification assumptions are required with respect to group formation.

The rest of the paper proceeds as follows. The next section sets the stage by reviewing the "reduced-form" empirical peer effect literature and introducing background information on the education of Shanghai's migrant students. Section 3 describes the model, discusses its various implications, and also characterizes the "reduced-form" estimates in our more general framework. Section 4 presents the data and performs various tests to examine whether students are randomly assigned to different classrooms within schools. The main empirical results are reported in section 5, with placebo tests and robustness checks and comparisons with other findings from the literature. In section 6, we perform policy experiments to examine the effects of transferring students from private schools to public schools, and decompose its treatment effect into three components: school effects, peer effect due to peer quality, and peer effect due to peer interaction intensity. The last section concludes. We include some technical details of the model in the Appendix.

2 Related Literature and Background

2.1 The "reduced-form" peer effects literature

There is a large and expanding literature studying the effects of a particular (and usually disadvantaged) group of students on their classmates, such as females (Hoxby, 2000; Lavy and Schlosser, 2011), minorities (Angrist and Lang, 2004; Card and Rothstein, 2007), disruptive kids (Figlio, 2007; Carrell, Hoekstra, and Kuka, 2018) or low achievers (Lavy, paserman and Schlosser, 2012). Most of these studies adopt the "reduced form" specification and use the ratio of the "affecting" group as the key right-hand-side variable, and use the outcomes of the "affected" group as the left-hand-side variables. A negative coefficient on the ratio is interpreted as the "affecting" group having a negative effect on the outcome of the "affected" group. Identification relies on some exogenous variations of the ratio of the "affecting" group. For example, Hoxby (2000) explores within-school across-cohorts variations in the composition of female students. Many other studies, such as Ohinata and Van Ours (2013), use "as if" random classroom assignments within schools.

Within this literature, some studies focus on the effects of migrant children on their local peers. Many western developed countries have witnessed large inflows of immigrants in the last several decades. Some developing countries, such as China, have also experienced substantial internal rural-to-urban and regional migration as they develop and urbanize. Generally speaking, these international and domestic migrants are of lower social and economic status (SES) compared to local residents. Thus one would expect migrant children to affect their local peers negatively. However, the empirical findings in the literature have been rather mixed.

On the one hand, many studies report negative peer effects from migrant students, i.e., as the percentage of migrant students increases, outcomes such as test scores or educational attainment of their native peers tend to deteriorate (Jensen and Rasmussen; 2011, Ballatore et al., 2018; Nordin, 2013; Hoxby 2000; Flores and Scorezafave, 2014; Wang et al. 2018). In some cases, the negative peer effects even persist years after the classmate relationship is over, such as in Gould, Lavy and Paserman (2009). In addition, migrants themselves could also be negatively affected by the increase in immigrant student population, as shown by Jensen and Rasmussen (2011).

On the other hand, many other studies do not find significant negative impacts of migrant students on local students. Ohinata and Van Ours (2013) study primary school students in Netherland, and do not find strong evidence of negative spillover effects from the presence of immigrant children on the academic performance of the native students. Schneeweis (2015) finds no significant peer effects on native students by exploring variations in the ratio of immigrant students in different cohorts of primary students in Austria. Geay et al. (2013) show that after accounting for sorting at the school level, the negative effects of non-native English speakers on native English speakers can be ruled out.

Although this linear regression approach is widely used in empirical studies, it focuses only on the effect of the “affecting” group on the “affected” group. One should also interpret this “reduced-form” coefficient with caution, especially when used in scenarios that involve regrouping students, as students are generally sorted into different types of schools based on unobservables. When the model is misspecified, estimates could induce large biases in policy simulations, as shown by Carrell et al. (2011) and Fruehwirth (2014). In this paper, we extend this approach by developing a more general framework that explicitly considers all peer interactions between and within heterogenous student subgroups, and takes consideration of endogenous formation of the peer groups. We characterize the reduced-form coefficient in our extended framework, and show that its sign is a priori ambiguous, which is consistent with the conflicting findings reported in the literature.

2.2 Migrant Children's Education in China

Along with rapid economic growth, China has experienced a historically unprecedented surge of rural-to-urban migration during the last several decades. A large number of rural peasants have migrated to work in the cities to seek better-paid jobs and living environments (Démurger et al, 2009; Messinis, 2013). However, in most cases, their official household registration statuses, or "*hukou*", stay with their places of birth. In China, *hukou* acts like a "citizenship" in different jurisdictions within the country, to which many social benefits and rights are tied. According to the most recent 2020 population Census, there are around 376 million such non-*hukou* migrants who live in a city other than where her official residence is registered. In this paper, we are going to follow the literature on China's migration and simply use "migrants" as synonym to "non-local *hukou*" people.

Because migrants do not possess local *hukou*, their children do not have the legal right to go to public schools in destination cities. In the 1990s, public schools usually charged hefty "out-of-district" fees to students without local *hukou* and most migrant children had to go to the then-newly-emerged private schools. These private migrant schools were mostly informal schools, privately-run, and usually had much worse facilities and teachers compared to public schools (Chen and Feng, 2013; Chen et al., 2019). Since 2001, the Chinese central government has required the destination city governments to take a more proactive approach towards migrant children's education. As a result, public schools have become more accessible to migrant kids, and the proportion of migrant students enrolled in private schools has declined gradually. Nevertheless, even today, there are still around 20% of migrant students in private migrant schools across the country (Xu and Zhang, 2016).

Our study is based in Shanghai, the largest Chinese city and one of the most attractive destinations for domestic migrants. Since 2008, Shanghai government has implemented a new policy to shut down all migrant-only private schools in its central districts. Migrant students from these closed private schools are allowed to enter nearby public schools. In more peripheral suburban districts, where there is limited number of public schools, selected private schools are allowed to continue to operate. These private schools also receive financial subsidies from the local governments to cover tuition for migrant students. Therefore, for a migrant child, the probability of entering a public school crucially depends on where her family lives in the city. More detailed discussions regarding policy changes for migrant-only private schools in Shanghai can be found in Chen and Feng (2013, 2017).

Given that migrant children can go to either public schools or private schools, previous studies have paid much attention to the treatment effect associated with enrolling migrant children in public schools. Chen and Feng (2013), Lai et al. (2014), and Chen and Feng (2017) all show significant positive test score gains from studying in public schools. Chen et al., (2020) find a positive long-term impact of studying in public elementary schools on high school enrollment. Because public schools are better than private migrant schools in almost all dimensions, including school and classroom facilities, quality of teachers, as well as quality of classmates, the treatment effect of public schools estimated in the literature actually reflects a total effect. So far, no studies have tried to disentangle these separate elements and identify the component of peer effects, which we will do in this paper.

There are a few studies on the issue of migrant children's peer effects in the Chinese context. Most of them focus on middle school students due to data limitations, as the publicly-available China Education Panel Survey (CEPS) only samples middle schools. However, migrant students in middle schools are a much smaller and selected sample compared to their counterparts in elementary schools. Due to China's current education system, many migrant children have to return to their hometown once they finish elementary school.[3] These studies also adopt the "reduce form" approach when estimating peer effects, and have reached somewhat contradicting conclusions. Using data on Grade 7 and 9 students in 86 schools from CEPS, Hu (2018) reports that the test scores of local students significantly decrease with the ratio of migrant students. On the other hand, Wang (2018) use Grade 9 students from 43 schools in CEPS and find that the ratio of migrant students actually improves the Chinese test scores of local students, and has no impact on Math or English test scores. One reason for the inconsistent results might be

[3] Typically, migrants face more and more obstacles as they move up the educational ladder. In China, college admission quotas are allocated to each province, with possibly different college entrance exams. This forced migrant children who want to go to college to go back to their own province in order to get prepared, usually in the middle school stage. On the other hand, middle schools and high schools in destination cities have little incentives to educate those non-*hukou* students who would not qualify for college entrance exams anyway.

the different samples used by the two studies. As we will show later, the sign of the reduced-form specification estimate is a priori uncertain, thus can be sensitive to changes in sample compositions.

3 The Model

3.1 The Model Framework

There are two types of students, local and migrant students; and two types of schools, public and private schools[4]. Each family has a predetermined residence, which belongs to a school district.[5] A school district contains one public school and may also contain a private school. Local children always go to the public school in their school district. For migrant children, they may go to the public school or the private school (if there is one) in their school district. Migrants' school type choice is determined by Eq. (3.1) below.

$$S_i^m = I(z_i^{'}\gamma + \epsilon_{s,i}^m > 0) \tag{3.1}$$

where $I(.)$ is an indicator function, $S_i^m = 1$ if migrant i goes to a public school and $S_i^m = 0$ if migrant i goes to a private school. z_i is a vector of child, household characteristics and other factors that influence student i's school type choice and $\epsilon_{s,i}^m$ is the error term in the selection equation.

We observe a migrant's outcome measure (i.e., test score) as in (3.2) below.

$$y_i^m = S_i^m y_{1,i}^{m*} + (1 - S_i^m) y_{0,i}^{m*} \tag{3.2}$$

That is, if i studies in a public school, $S_i^m = 1$, her test score is $y_i^m = y_{1,i}^{m*}$. If i is enrolled in a private school, $S_i^m = 0$, then we observe test score $y_{0,i}^{m*}$. In both types of schools, a student's performance measure depends on school and classroom characteristics, her personal and family characteristics, peer effects from their classmates, as well as a random error term.

If migrant student i studies in a public school, then her test score is:

$$y_{1,i}^{m*} = u_g^{'}\alpha^P + x_i^{'}\beta^P + \lambda_{mm}^P \sum_k W_{mm,ik}^P E[y_k|\Im_i] + \lambda_{ml}^P \sum_h W_{ml,ih}^P E[y_h|\Im_i] + \epsilon_{y,1,i}^m; \tag{3.3}$$

[4]We use subscripts "l" and "m" to represent "local" and "migrant" children respectively. Superscripts "P" and "M" are employed to distinguish public schools which have both local and migrant children and private schools with exclusively migrant students.

[5]In reality, people can change their residence and switch school districts, but the cost of moving is very high and the choice of residential locations is strongly constrained by job opportunities and housing market conditions.

where u_g represent characteristics of schools and classes, and α^P measures the marginal influence of school and class observables on the behavior of an individual child, which is called the "contextual effects" by Manski (1993). x_i contains a student's personal and family characteristics and their coefficients are denoted by β^P.

The next two terms capture peer influence which is the focus of our paper. Because there are two types of students (migrant and local) in a public school, a migrant student can receive two different types of peer effects, one from other migrant students, which is denoted by $\lambda^P_{mm} \sum_k W^P_{mm,ik} E[y_k|\mathfrak{I}_i]$, and the other from local students, denoted by $\lambda^P_{ml} \sum_h W^P_{ml,ih} E[y_h|\mathfrak{I}_i]$.

Let us take a closer look at the peer effect terms. Note that there are three terms for $\lambda^P_{mm} \sum_k W^P_{mm,ik} E[y_k|\mathfrak{I}_i]$. First, $E[y_k|\mathfrak{I}_i]$ is the expected peer test score, or peer quality, given i's information set $\mathfrak{I}_i$. $W^P_{mm,ik}$ measures peer relationship between student i and k, as W^P_{mm} is a matrix that signifies migrant-migrant peer relationships between all agents in a public school classroom. λ^P_{mm} is the coefficient on migrant-to-migrant peer effect. This term sums all influences from i's migrant peers. Similarly, $\lambda^P_{ml} \sum_h W^P_{ml,ih} E[y_h|\mathfrak{I}_i]$ captures all influences from local student peers of i, with λ^P_{ml} denoting local-to-migrant peer effects.

Note that student i's information set $\mathfrak{I}_i$ contains both public and private information. The public information part is $\mathfrak{I}_g$ which contains what is known to all students in this class and can be further decomposed into two subsets, (1) $\mathfrak{K}_g$, which includes school and classroom-level characteristics, as well as student characteristics and other policy and background variables observable to everybody (the u, x, z terms); and (2) $\mathfrak{W}_g$, which includes the classmate relationships in the classrooms, i.e., the W matrices, also observable to everybody once students are assigned to different classrooms.

Suppose the same migrant student i ends up with a private school, then her test score would be:

$$y^{m*}_{0,i} = u'_{g'} \alpha^M + x'_i \beta^M + \lambda^M_{mm} \sum_k W^M_{mm,ik} E[y_k|\mathfrak{I}_i] + \epsilon^m_{y,0,i}. \tag{3.4}$$

In this case, only one type of peer effect is present because there are only migrant students in private schools. λ^M_{mm} is the migrant-to-migrant peer effect coefficient in a private school, while W^M_{mm} is the matrix for peer relationships among migrant students in private schools.

On the other hand, local students always go to public schools. A local student j's test score is determined as follows:

$$y_j^l = u_g'\alpha^P + x_j'\beta^P + \lambda_{lm}^P \sum_k W_{lm,jk}^P E[y_k|\Im_j] + \lambda_{ll}^P \sum_h W_{ll,jh}^P E[y_h|\Im_j] + \epsilon_{y,j}^l. \qquad (3.5)$$

where λ_{lm}^P and λ_{ll}^P denote the migrant-to-local and local-to-local peer effect coefficient in public schools, respectively.

3.2 Model assumptions, identification and estimation

For further investigation, we make the following assumptions on the distribution of the disturbances.

Assumption 3.1 *The error terms all have zero means and are identically and independently distributed (iid). Let* $\epsilon_i^m = (\epsilon_{s,i}^m, \epsilon_{y,1,i}^m, \epsilon_{y,0,i}^m)'$ *for migrant students and let* $\epsilon_j^l = \epsilon_{y,j}^l$ *for local students. Then we have:*

- $E[\epsilon_i^m] = 0$ *for any migrant child* i *and* $E[\epsilon_j^l] = 0$ *for any local child* j.
- ϵ_i^m*'s are independent across all migrants.* ϵ_j^l*'s are also independent among all local children.*
- ϵ_i^m*'s and* ϵ_j^l*'s are independent.*

Assumption 3.2 *The observable characteristics* $\mathfrak{K}_g = (u, x, z)$ *are exogenous, i.e., have no predictive powers on the error terms. To be more specific,*

- $E[\epsilon_{s,i}^m|z] = E[\epsilon_{s,i}^m]$ *for each migrant student* i.
- $E[\epsilon_{y,1,i}^m|\mathfrak{K}_g] = E[\epsilon_{y,1,i}^m]$ *and* $E[\epsilon_{y,0,i}^m|\mathfrak{K}_g] = E[\epsilon_{y,0,i}^m]$ *for each migrant student* i.
- $E[\epsilon_{y,j}^l|\mathfrak{K}_g] = E[\epsilon_{y,j}^l]$ *for each local student* j.

Assumptions 3.1 and 3.2 are conventional in the literature. Under these assumptions, an individual cannot get additional information about other classmates' behaviors out of his/her private shocks given the public information. It then follows that the expectation on others' outcomes are actually based on public information, i.e., $E[y_k|\Im_i] = E[y_k|\Im_g]$, which is proved in the Appendix.

Assumption 3.3 *Class size does not matter per se for peer effects. Suppose there are n_g students in a classroom, for a given student, if we clone his $n_g - 1$ classmates such that the class size becomes $2n_g$-1, then the test score of the student would be unchanged.*

Assumption 3.3 is intuitive, otherwise student test scores would increase solely as a result of increased class size. Under Assumption 3.3 , we can normalize the peer relationship matrices such that the row sum of $W^P_{mm} + W^P_{ml}$, of $W^P_{lm} + W^P_{ll}$, and of W^M_{mm} all equal to 1. For example, for a migrant student i in a public school, $W^P_{mm,ik} = 1/(n_g - 1)$ if student k is also a migrant student and is in the same classroom as i, otherwise $W^P_{mm,ik} = 0$. Then λ^P_{mm} measures average peer effect from a migrant peer to a migrant student (denoted as student i) in public schools, i.e., if the test score of one migrant classmate of student i increased by one unit, then holding everything else constant and not allowing for multiplier effect, student i's test score would increase by $\lambda^P_{mm}/(n_g - 1)$ units. The other coefficients, including λ^P_{ml}, λ^P_{lm}, λ^P_{ll}, and λ^M_{mm}, can be interpreted similarly. Therefore, Eq. (3.3), (3.4) and (3.5) can be rewritten as follows,

$$y^{m*}_{1,i} = u^{'}_g\alpha^P + x^{'}_i\beta^P + \lambda^P_{mm}\sum_{k\in\mathfrak{M}}\frac{1}{n_g - 1}E[y^m_k|\mathfrak{I}_g] + \lambda^P_{ml}\sum_{h\in\mathfrak{L}}\frac{1}{n_g - 1}E[y^l_h|\mathfrak{I}_g] + \epsilon^m_{y,1,i}; \tag{3.6}$$

$$y^{m*}_{0,i} = u^{'}_{g'}\alpha^M + x^{'}_i\beta^M + \lambda^M_{mm}\sum_{k\in\mathfrak{M}}\frac{1}{n_g - 1}E[y^m_k|\mathfrak{I}_{g'}] + \epsilon^m_{y,0,i}. \tag{3.7}$$

$$y^l_j = u^{'}_g\alpha^P + x^{'}_j\beta^P + \lambda^P_{lm}\sum_{k\in\mathfrak{M}}\frac{1}{n_g - 1}E[y^m_k|\mathfrak{I}_g] + \lambda^P_{ll}\sum_{h\in\mathfrak{L}}\frac{1}{n_g - 1}E[y^l_h|\mathfrak{I}_g] + \epsilon^l_{y,j}. \tag{3.8}$$

where $\mathfrak{M}$ and $\mathfrak{L}$ denote the set of migrant peers and local peers for a student in her classroom, respectively.

Assumption 3.4 *Conditional on school type selection, students are assigned to classes exogenously, i.e., class assignment rules depend only on individual and school observed characteristics.*

Assumption 3.4 is a key identification assumption, and is weaker than the "random classroom assignment" assumption typically seen in many peer effect studies. Under this assumption, we do not have to require classroom assignment to be completely random. School administrators can still take observable characteristics, typically gender, into consideration when assigning students to different classrooms.

We make additional parametric assumptions on the distributions of the idiosyncratic shocks.

Assumption 3.5 *ϵ_j^l's are i.i.d. $N(0, \sigma_{l,y}^2)$. ϵ_i^m's are i.i.d. $N(0, \Sigma_m)$, where*

$$\Sigma_m = \begin{pmatrix} \sigma_{m,s}^2 & \rho_{m,1}\sigma_{m,s}\sigma_{m,y,1} & \rho_{m,0}\sigma_{m,s}\sigma_{m,y,0} \\ \rho_{m,1}\sigma_{m,s}\sigma_{m,y,1} & \sigma_{m,y,1}^2 & \rho_{m,y}\sigma_{m,y,1}\sigma_{m,y,0} \\ \rho_{m,0}\sigma_{m,s}\sigma_{m,y,0} & \rho_{m,y}\sigma_{m,y,1}\sigma_{m,y,0} & \sigma_{m,y,0}^2 \end{pmatrix}.$$

Under the previous assumptions we can simply apply the standard Heckman sample selection results. In particular, we have:

$$\begin{aligned} & E[y_i^m | S_i^m = 1] = E[y_{1,i}^{m*} | S_i^m = 1] \\ =& u_g^{'}\alpha^P + x_i^{'}\beta^P + \lambda_{mm}^P \sum_{k\in\mathfrak{M}} \frac{1}{n_g - 1} E[y_k^m | \Im_g] + \lambda_{ml}^P \sum_{h\in\mathfrak{L}} \frac{1}{n_g - 1} E[y_h^l | \Im_g] \\ & + E[\epsilon_{y,1,i}^m | S_i^m = 1]. \end{aligned} \tag{3.9}$$

$$\begin{aligned} & E[y_i^m | S_i^m = 0] = E[y_{1,i}^{m*} | S_i^m = 0] \\ =& u_{g'}^{'}\alpha^M + x_i^{'}\beta^M + \lambda_{mm}^M \sum_{k\in\mathfrak{M}} \frac{1}{n_g - 1} E[y_k^m | \Im_{g'}] + E[\epsilon_{y,0,i}^m | S_i^m = 0]. \end{aligned} \tag{3.10}$$

where the last terms of Eq. (3.9) and Eq. (3.10) are the inverse mills ratios, which can be estimated from the selection equation (Eq. (3.1)) first.

Finally, identification also requires that the average peer effects are all less than one in absolute numbers, otherwise the system will explode[6], i.e.:

Assumption 3.6 $\max\{|\lambda_{mm}^P|, |\lambda_{ml}^P|, |\lambda_{lm}^P|, |\lambda_{ll}^P|, |\lambda_{mm}^M|\} < 1.$

It is proved in the Appendix that under certain full rank conditions, after normalizing $\sigma_{m,s} = 1$, all parameters other than $\rho_{m,y}$ can be identified.[7] Those full-rank conditions require full ranks in X and in Z and also an excluding condition that Z contains some variables which are not contained in X.

In estimating the model, we first estimate Eq. (3.1), which allows us to calculate the inverse mills ratios in Eq. (3.9) and Eq. (3.10). We then use two-stage NLS method to estimate the system of equations for the outcomes (Eqs. 3.8, 3.9 and 3.10). We first take expectations for the test scores conditional on observables on both sides of the equations, and estimate the predicted

[6] In the Appendix, we show that the same assumptions ensure a unique Bayesian Nash Equilibrium (BNE) of the incomplete information game on which our econometric model is based.

[7] For a migrant child i, $y_{1,i}^m$ and $y_{0,i}^m$ can not be observed at the same time. Thus, their correlation coefficient, $\rho_{m,y}$ can not be identified

values of the expected test scores. We then plug back the predicted expected test scores in the equations to estimate the parameters.

3.3 Comparison with the "reduced-form" peer effect models

As we have reviewed before, the vast majority of the peer effect literature takes a "reduced-form" approach. In our example, the standard practice would be to simply regress the test scores of local students on the percentage of migrant students in a public school classroom. If the coefficient on migrant student ratio is negative, then this would be interpreted as evidence that migrant students negatively affect their local classmates. This specification can be considered as a special case of our model if school type selection biases do not exist, i.e., $E[\epsilon^m_{y,1,i}|S^m_i = 1] = 0$. In reality, this condition is unlikely to hold, as migrant students in public and private schools are expected to differ in both observable and unobservable dimensions. Therefore, the "reduced-form" estimates are uninformative in policies involving regrouping students between the two different types of schools.

Even without considering endogenous school type selection, our model is still more extensive and "structural" than the "reduced-form" specification, as we explicitly model migrant-to-local, local-to-migrant, local-to-local and migrant-to-migrant peer interactions. Our specification also allows all conditioning variables to affect outcomes of peers in a nonlinear fashion due to the existence of the multiplier effect. Therefore, it makes sense to characterize the "reduced-form" coefficient in our more general framework in order to understand what the "reduce-form" coefficient actually identifies.

We first denote $\iota_m = E[y^m_k|migrant]$ and $\iota_l = E[y^l_k|local]$, i.e., ι_m and ι_l are the expected test score of all migrant students and all local students, respectively. For each class, its expected migrant test score deviates from the population mean ι_m by ι_{mg}, i.e. for class g, the expected test score of its migrant students is $\iota_m + \iota_{mg}$. We allow ι_{mg} to be correlated with class migrant student ratio $n_{g,m}/n_g$, where $n_{g,m}$ is the number of migrant students in class g and $\frac{n_{g,m}}{n_g}$ is the ratio of migrant students. To make things simple, we assume that $\iota_{mg} = \gamma_g n_{g,m}/n_g$, where γ_g is the correlation coefficient. If $\gamma_g < 0$, then classes with more migrant students also have "worse" migrant students in terms of expected test scores.

We can then replace $E[y_k^m|\Im_g]$ and $E[y_h^l|\Im_g]$ in Eq.(3.8) and have the following equation.[8]

$$y_j^l = u_g^{'}\alpha^P + x_j^{'}\beta^P + \lambda_{lm}^P \sum_{k\in\mathfrak{M}} \frac{1}{n_g-1}(\iota_m + \iota_{mg}) + \lambda_{ll}^P \sum_{h\in\mathfrak{L}} \frac{1}{n_g-1}\iota_l + \epsilon_{y,j}^l.$$

which can be rewritten as

$$\begin{aligned} y_j^l =& u_g^{'}\alpha^P + x_j^{'}\beta^P + \lambda_{ll}^P\iota_l + (\lambda_{lm}^P\iota_m + \lambda_{lm}^P\iota_{mg} - \lambda_{ll}^P\iota_l)\frac{n_{g,m}}{n_g-1} + \epsilon_{y,j}^l, \\ =& u_g^{'}\alpha^P + x_j^{'}\beta^P + \lambda_{ll}^P\iota_l + [(\lambda_{lm}^P - \lambda_{ll}^P)\iota_m + \lambda_{ll}^P(\iota_m - \iota_l) + \lambda_{lm}^P\gamma_g\frac{n_{g,m}}{n_g}]\frac{n_g}{n_g-1}\frac{n_{g,m}}{n_g} + \epsilon_{y,j}^l, \end{aligned}$$

Therefore, give that $\frac{n_g}{n_g-1} \approx 1$, for a local student, the marginal effect of migrant ratio in his/her class is approximately $(\lambda_{lm}^P - \lambda_{ll}^P)\iota_m + \lambda_{ll}^P(\iota_m - \iota_l) + 2\lambda_{lm}^P\gamma_g\frac{n_{g,m}}{n_g}$ which is roughly what we get if we regress test scores of local students on migrant ratio in a class in a typical "reduced-form" specification.

To start, if migrant test scores are uncorrelated with class migrant ratios, i.e., $\gamma_g = 0$, the coefficient on migrant ratio can be decomposed into two parts. The first term is $(\lambda_{lm}^P - \lambda_{ll}^P)\iota_m$, which is the difference between migrant-to-local peer effect and local-to-local peer effect times the average test scores of local peers. This term is positive when migrant students influence a local student more than his local peers, i.e., $\lambda_{lm}^P > \lambda_{ll}^P$, otherwise the term will be negative.[9] The second term is $\lambda_{ll}^P(\iota_m - \iota_l)$, which is the difference between migrant and local peer test scores times local-to-local peer effect. This term would be negative if the average migrant test score is lower than the average test score of local peers. Otherwise the term will be positive. Clearly, even in this simple setting, the sign of the coefficient is a priori undetermined. In a typical practical setting, it is likely that migrant students have lower average test scores than local students ($\iota_m < \iota_l$) so the second term is negative, and migrants influence locals less than (or equal to) locals do ($\lambda_{lm}^P \leq \lambda_{ll}^P$). If that is the case, then we would find a negative coefficient, as is typical in the literature. Nevertheless, in that case, we should be careful in interpreting the coefficients. It is completely possible that we get a negative coefficient in the "reduced-form" regressions only because migrants have much lower test scores than local students, which does not tell us much about the sign and magnitude of λ_{lm}^P, which is the migrant-to-local peer effect

[8]We assume that the number of local students is large enough so that for any given local students, the average test score of his local peers is approximately the same as the average test scores of all local students (including himself).

[9]We assume that test scores are all positive, i.e., $\iota_m > 0$ and $\iota_l > 0$, and all average peer effects are nonnegative, i.e., increase in peer test scores will not decrease a student's test score.

in which the researcher is supposed to be interested. It is also uninformative of other λs, the true peer effects parameters.

In more realistic scenarios, it is highly likely that migrant test scores are correlated with class migrant ratios, i.e., $\gamma_g \neq 0$, then the coefficient on migrant ratio would consist of three parts. If classes with more migrant students are also likely to have migrants that are on average worse in terms of test scores, then this third term will be negative. Otherwise it will be positive. Again, the sign of the "reduced-form" coefficient is indeterminate and cannot be interpreted in a straightforward way. Our approach provides a framework to decompose and interpret the coefficient in a more systematic fashion, and could be more useful in advising policy.

4 The Shanghai Elementary School Data

4.1 Data and Summary Statistics

We use data collected by two of the coauthors of this paper from the fifth grade students in Shanghai in November 2012. Among Shanghai's 16 districts, we first chose 5 districts, including 2 central districts: Huangpu and Yangpu, and 3 suburban districts: Pudong, Baoshan and Minhang. These 5 districts collectively cover about 53% of all public primary schools, and around half of all migrant students in Shanghai. Because the objective of the study is migrant children's education, we only sampled from schools in these districts that had a substantial proportion of migrant students, including public schools with at least 20% of migrant students, and all migrant-only private schools. Therefore, we have excluded elite public and private schools that are basically inaccessible to migrants and even ordinary local Shanghai students. In the end, our school sample includes 5 public schools in the central districts, 6 public schools and 9 migrant-only private schools in the suburban districts.[10] We then conduct interviews with all fifth grade students from these 20 selected schools.

To test the representativeness of our school sample, we compare average school characteristics of our sample schools with those of all schools in the 5 districts. Table 1 shows the results for public schools and migrant-only private schools, respectively. Compared to public schools, migrant-only private schools have slightly less teachers and substantially more students, thus a

[10]No migrant-only private schools were allowed to operate in central districts by the time we conducted the survey.

much lower teacher-student ratio. But there is no difference between our sample schools and all schools from these 5 chosen districts. In terms of teachers with middle rank and higher titles, which is a measure of teacher qualifications, the proportion in our sample public schools is lower than that of all public schools (0.479 vs. 0.568). This is not surprising as our sampling protocol precludes elite public schools with very few or no migrant students, which are presumably high-quality schools. On the other hand, there is no systematic difference between our sampled private schools and all migrant-only private schools.[11]

We gave all fifth grade students from the sample schools the same standardized Math tests. This is important because the exams given by the schools themselves were different for public schools and migrant-only private schools, leading to incomparable test scores. Our Math test was designed by experts outside the sample schools, with a prefect score of 100. The actual test lasted for 20 minutes and was proctored by both the head teacher of the class and one of our interviewers. We also collected data on individual characteristics, including age, gender, whether the student is the only child in her/his family, *hukou* status, parents' education, family income, and whether the child has attended kindergarten before entering primary school, as well as characteristics of their math teachers, such as the education, experience and tenure.

Table 2 presents the summary statistics for students in three groups, local students in public schools, migrant students in public schools, and migrant students in private schools, respectively. There are significant test score differences among the three groups. The local and migrant students in public schools on average scored 60 and 56 out of 100 in the math test, while migrant students from private schools only scored 44. In general, local students have the strongest individual and family characteristics, followed by migrant students in public schools, who in turn have significant better background than their counterparts in private schools. Thus, there is a clear pattern of positive selection into public schools for migrant students. Migrant students in public and private schools also come from different origin provinces. Migrant students in public schools are more likely to come from Jiangsu, a wealthy province geographically close to Shanghai. On the other hand, students in private schools are more likely to come from poorer migrant-sending provinces such as Anhui, Henan and Sichuan. In terms of where the families lived in 2008, right before the students entered primary school, there is also a significant

[11] Please also see discussions in Chen and Feng (2017).

difference for migrant students from the two different types of schools. Before 2008, almost 40% of families whose children study in public schools lived in central Shanghai, while almost no families (0.7%) whose children study in private schools lived in these central districts. This pattern is consistent with the abrupt policy change in 2008 that let to the shutdown of all migrant-only private schools in central Shanghai.

Finally, compared to students from public schools, students in the migrant-only private schools have much lower percentages of teachers with college degree, and much higher student-teacher ratios, showing that classes are much larger and teacher qualities are much lower in these private schools. On average, local students from public schools have 16 local student peers (classmates) and 12.6 migrant student peers. Migrant students in public schools have slightly less local peers (11.4) and more migrant peers (16.6) on average. Migrant students in private schools have no local student peers and on average 45.6 migrant peers. The effects of these classroom peers on student performance will be our focus in the following sections.

4.2 Tests of Random Assignment

In Chinese elementary schools, typically students are randomly assigned to different classrooms in the same grade, as the education authority specifically forbids tracking.[12] However, migrant students in public schools pose a special challenge as their performances are usually much worse than their local counterparts when they first migrate to the cities. When public schools started to admit large number of migrant students, they usually form migrant-only classes so that it is easy to teach them, and also to avoid complaints from parents of local students. However, as internal migration intensified, the education authority started to require mixed classrooms, and schools gradually turned to mixed classrooms with both migrant and local students. However, at the time of the survey, we still observe some segregated classrooms in some public schools. As Table 3 shows, there are about five schools with segregated classes, where three schools (5, 6, 9) assign one or two classes with only migrant students, and two other schools (10 and 11) have one class with (almost) only local students.

There is no guarantee that students from migrant-only or local-only classes are randomly

[12]Most recently published peer effect studies using data on Chinese junior high schools rely on the random classroom assignment assumption, see e.g. Hu (2018), Wang et al. (2018), Gong et al. (2021), Huang and Zhu (2020) and Xu et al. (2020). Elementary schools in China are even more likely to conform to the no-tracking regulation than junior higher schools.

assigned to these classrooms. For example, when a private school is shut down, a nearby public school is usually required to take their students and may form a migrant-only class. To play safe, in all of our follow-up analyses, we treat these migrant-only or local-only classes as independent "pseudo schools". To be specific, class 1 and 2 in school 5 is treated as a separate "school" from class 3 in school 5. Similarly, class 1 of school 6, classes 2 and 3 of school 6, class 1 of school 9, class 2 of school 9, class 1 of school 10, class 2, 3, 4 of school 10, class 1 of school 11, and class 2, 3, 4, 5 of school 11 are all individual "pseudo schools". Therefore, we will be working with 25 "pseudo-schools" in total instead of 20 schools. For brevity, we shall still call those "pseudo schools" as "schools".

We shall then perform a batch of tests to show that the actual data do not violate the "as if" random classroom assignment rule, which would in turn ensure that Assumption 3.4 holds.

4.2.1 Correlation of characteristics between students and classroom peers

If students are randomly assigned to classes within schools, we should not observe that similar students are more likely to be assigned into the same classes. Therefore, the characteristics of students should not be systematically positively correlated with those of their peers in the same classes. Following Guryan et al (2009), we regress each student's predetermined characteristics on the mean of the same characteristics of his/her class peers, controlling for the school fixed effects. As shown in Table 4, none of the coefficients are statistically significantly different from zero. The largest t-stat is only 1.3 for female, but the coefficient is negative. This suggests that schools might have a tendency to balance sex ratios in a classroom. Overall, there is no evidence of positive sorting of students in classroom assignment.

4.2.2 Test for standard deviations of class-level averages

To provide further evidence of random class assignment, we test whether the variations of class-level averages of key predetermined variables are consistent with random class assignments within schools. Following Bifulco et al. (2011), we compare the standard deviation of real data and simulated data with random class assignment. We focus on the 19 schools with at least two classes in the fifth grade and exclude the 6 schools with only one class.[13] If class assignments

[13]Note those are "pseudo schools" as explained previously.

are random, then the real data and the simulated data should have similar standard deviations for the raw data and for the residuals after removing school fixed effects, as well for the ratios of the two standard deviations.

For the real data, we first compute class-level averages for each variable, and then calculate standard deviations (s.d.) across classes. Columns (1) and (2) in Table 5 show the raw s.d. and the s.d. after removing school fixed effects in the real data, which measures within-school across-class variations. Column (3) reports the ratio of the first two columns.

We then ran 500 simulations. In each simulation, we randomly assign students of each school into different classes, and compute the standard deviations of class-level averages for each variable. Columns (4) and (5) report the averages of raw s.d. and the s.d. after removing school fixed effects, while column (6) reports their ratios. The simulated data display strong similarities with their real data counterparts. For example, s.d. for the variable "father education high school and above" is 0.208 in both the actual and the simulated data. After removing school fixed effects, the two s.d. become 0.065 and 0.063 which are also very close. Finally, the ratios are 31% and 30% which show comparable levels of importance of the school fixed effects. The results for other variables are similar and suggests our data are rather consistent with patterns implied by random class assignment.

4.2.3 Resampling Tests

We also use resampling techniques to test random assignment of students in classes within schools, as done in Carrell (2010) and Lim and Meer (2017). We also exclude the 6 schools with only one class as in section 4.2.2. For each school, students are randomly assigned to different classes without replacement. Based on 1000 such simulations, we can generate class-level averages for the variables we choose in each class for 1000 times. For example, for the variable "female", we can compute female ratios for each class based on each simulated data set. We then compute an empirical p-value for each class, which is the percentage of the 1000 simulated values of female ratio lower than the actual female ratio in that class. This will produce 63 p-values for each variable (including Shanghai *hukou*, female, only child, etc..), as there are 63 classes in those 19 schools. If classes are assigned randomly within schools, then the empirical p-values will be uniformly distributed. We then use both the Kolmogorov-Smirnov test and Chi-squared test to examine whether the distribution of the empirical p-values for each variable is uniform. Based on results reported in Table 6, we cannot reject the null hypotheses of random assignment for all of the 8 variables at the 5 percent level. Again, there is no evidence indicating nonrandom assignment of students into classes within schools.

4.2.4 Assignment of Teachers

Finally, it is also important to test whether teachers are randomly assigned to each class within schools. If teacher qualifications are correlated with student characteristics, then one might erroneously attribute the effect of teachers to students themselves or their peers. Table 7 reports results on within-school correlations of teacher characteristics with student characteristics. We choose four characteristics of the math teacher (female, teaching experience, within-class tenure, college degree) and regress them separately on each of the 8 student characteristics and school fixed effects. Most of the coefficients from the 32 regressions are not statistically different from zero. Only four coefficients are marginally significant at the 10 percent level. There is no consistent pattern of sorting of teachers with students. For example, teachers with a college degree are less likely to teacher students with better-educated parents, but they are more likely to teach students from families with monthly income between 3-5 thousand, compared to those with family monthly income less than 3 thousand. Overall, we find no evidence against the null of random teacher assignment across classes.

5 Empirical Results

5.1 Peer effect estimates

We start with estimating the school selection equation for migrant students as denoted by Eq. (3.1). Table 8 reports the results. As expected, there is strong positive selection into public schools among migrant students based on individual and family background. Students who have urban-*hukou*, are the only child in their family, have attended kindergarten are more likely to enter public schools. Similarly, families with higher levels of income are also more likely to send their kids to public schools. In terms of origin province, we see that migrant students from Anhui, Henan and Sichuan are strongly less likely to enter public schools. Migrants from these provinces are more likely to be rural farmers and have relatively low social-economic statuses in Shanghai. Also, if one's family lived in central Shanghai before 2008, then the probability of entering a public school is much higher. This is consistent with the policy change in 2008 that have substantially altered the geographic distribution of public and migrant-only private schools in Shanghai. Because residential mobility for migrants is strictly constrained by job opportunities and cost of living concerns, students whose families lived in central districts prior to 2008 benefited from this policy change because they are now entitled to public schools. On the other hand, families who have lived in a peripheral district still face the fierce competition for local public schools, and their kids are more likely to enter a migrant-only private school.

We then proceed to estimate the main equations on math exam test scores (Eq. (3.6), (3.7), and (3.8)). The results for public schools are reported in Table 9. Column (1) of Table 9 reports results for the specification without accounting for migrant children's school selection. In column (2), we consider the sample selection issue by adding the inverse mills ratio calculated from the Probit regression reported in Table 8. In column (3), we further include school dummies to exploit variations across classes within schools. This also eliminates common contextual effects, which has been the standard practice in the literature (e.g. Ohinata and Van Ours, 2013; Schneeweis, 2015).

We focus on discussing the results based on our preferred specification as reported in column (3) of Table 9. First, note that females on average score 1.7 points less than male students. Coefficients of father's education and whether the family monthly income is above 5 thousand are both positive and statistically significant. Also, the inverse mills ratio is positive and significant, which suggests that it is important to account for sample selection of migrants. Our main interest is, however, the various peer effect coefficients. The results show that λ^P_{mm}, λ^P_{ml}, λ^P_{lm}, and λ^P_{ll} are 0.802, 0.764, 0.822 and 0.815, respectively. These estimates are all highly significant statistically and are relative large in magnitude. It is normal in the literature to find that contemporaneous peer effects to be significantly larger than peer effects based on predetermined variables (see e.g. Fruehwirth, 2013).

One interesting hypothesis is whether peer effects are heterogeneous according to the types of peers. For example, migrant students are typically viewed as having detrimental effects on local students because of their low test scores and perceived bad behaviors by some local parents.

Since the estimates of λ_{mm}^{P}, λ_{lm}^{P} are all positive, we can be sure that at least an increase in a migrant's test score will increase rather than decrease their peer's test score. Nevertheless, it is still important to formally test whether peer effects are the same in magnitude, as heterogeneity of peer effects may have important implications on optimal class assignments. In panel B of Table 9, we formally test the hypothesis and cannot reject the null that all four peer effects are the same. We have also tested other potentially interesting hypotheses, including that the within-group peer effects (migrant-to-migrant, local-to-local) are stronger than those between different groups (migrant-to-local, local-to-migrant). We also tested symmetry, in that the effect of migrant students on local students are the same as the effect of local students on migrants. In all cases, we cannot reject the null hypotheses.

In Table 10, we report the results for migrant-only private schools. Again, the preferred specification is column (3), which accounts for sample selection of migrant students and includes school dummies. The basic results on the key individual and family characteristics are qualitatively similar to those in Table 9. The estimated peer effect parameter is 0.82, indicating that for a one point increase in average peer math test scores, a migrant student's math test score would increase by 0.82. The magnitude of the peer effect in private schools are comparable to those in public schools.

5.2 Robustness checks

To examine the robustness of peer effects results, we try different ways of treating the segregated classes in public schools. In the baseline specification reported in Table 9 and Table 10, we treat these as separate public schools (An example would be to treat classes 1 and 2 from school 5 as a separate public school). We think it is proper to do so as these classes could be different from other classes in the same school, thus violate the random class assignment assumption. Nevertheless, we try three alternative ways of treating these classes. First, we ignore the issue of segregated classes, which means they are still treated as part of the original school, rather than a separate school (e.g., classes 1 and 2 from school 5 are still part of school 5). Second, we completely drop these classes from our sample (e.g., drop classes 1 and 2 from school 5). Lastly, we treat them as separate schools, and label those migrant-only classes as private schools rather than public schools (e.g., classes 1 and 2 from school 5 would be treated as a separate private school).

Table 11 report the results, for both peer effects in public and private schools. In all of these regressions, we control for school fixed effects (as appropriate in each case) and include an inverse mills ratio which controls for sample selection of migrants. The estimated peer effects are all positive and highly significant. In most cases, their magnitudes are also close to the estimates from our baseline results. When segregated classed are deleted or treated as private schools, sample sizes in terms of the number of classes for public schools become smaller, thus the coefficient estimates become less precise.

We also try to directly control for peer's characteristics. In Table 12, we try to add two important average peer variables, family income and father's education, which have been proved to have significant impact on test scores. The estimated peer effects in our model remain robust. This shows that our model mainly captures the more interesting endogenous contemporaneous peer effects, which cannot be fully captured by these pre-determined variables (Fruehwirth, 2013, 2014).

5.3 Placebo tests

To make sure that our estimates do reflect peer interactions among classmates, rather than something else from model misspecification, we do some placebo tests using simulated data. Specifically, we reassign students randomly to different classes and re-estimate the peer effects using the false peer relationships in the simulated data. In the first case, we randomly reassign students to different classes in each school. Thus in the simulated data the "classmates" now could come from different classes although they are indeed schoolmates. In the second case, we first pool all students in all public schools and reassign them randomly to different schools and different classes. We do the same for all private school students. In this case, the "classmates" in the simulated data could come from completely different schools. If our peer effects parameters are meant to capture true interactions among classmates, we expect the estimates based on the simulated data to be significantly smaller, if not completely zero.

The simulation results are summarized in Table 13. Compared with the estimates for the true classmate relations, the estimated peer effects for the simulated data are in general much closer to zero. When we randomly assign students within each school across 500 simulations, the sample mean of those estimates are between -0.08 and 0.02. When we reallocate students across public (private) schools, the average estimates become negative and very small in magnitude. We also report percentages of significant estimates among the 500 simulations. In the first case, for the estimates of the five peer effects parameters, the percentages that they are significantly different from zero at the 5% level are 23%, 22%, 18%, 17%, 35%, respectively. In the second case, the percentages are even smaller, at 13%, 15%, 10%, 10%, and 25%, respectively. This is intuitive, because when students are reassigned within schools, the likelihood that they are true classmates are higher than when they are reassigned across different schools.

5.4 The marginal effects of predetermined characteristics

As noted before, an interesting feature of our model is its multiplier effect. When a student's characteristic change, it will not only affect her own test score, but also affect other students' test scores through peer interactions, which would then affect her own scores, and so on. This process will go on until a new equilibrium is reached. Therefore, the coefficients in the regression results as reported in Table 9 and 10 would underestimate the full marginal impact of a change in a predetermined characteristic. Because the marginal effects are heterogenous for each individual, we shall take the averages on the respective samples and present the average marginal effects.

In Table 14, we report the average marginal effects for a change in one's father's education and family monthly income, respectively. Again, we shall focus on discussing the results based on our preferred specification reported in columns (3) and (6) of Table 14. In terms of father's education, when a migrant student's father's education increases from below high school to high school and above, her test score would increase by 3.128 points on average, which is larger than 2.8, the corresponding coefficient reported in Table 9, suggesting a multiplier effect. Her migrant peers and local peers would also enjoy a test score gain of 0.406 an 0.343, respectively. Likewise, if a local student's father's education change similarly, his test score would increase by 3.125, and his local and migrant peers will also enjoy test score gains as a result. Similarly, in private schools, when a student's father's education changes from less than high school to high school and above, the student would enjoy a test score gain of 1.43. This marginal effect is smaller than in public schools, as the coefficient on father's education is lower in private schools compared with those in public schools.

Results for family's monthly income are similar. When the monthly family income of a migrant student in a public school increases from less than 3 thousand RMB per month to more than 5 thousand RMB, his test score would increase by 5.307 points. The average marginal effects for local students in public schools and migrant students in private schools are 5.302 and 2.066, respectively.

5.5 Comparison to alternative models

We first compare our estimation results with the reduced-form specification widely used in the literature. In this specification, the scores of local students are regressed on class-level migrant ratio, controlling for individual and family characteristics. Of course, this specification only applies to the public school sample with both local and migrant students. The results are reported in the first two columns of Table 15. In the first column, we do not control for school fixed effects. The coefficient of migrant ratio in class is -11.6. Taking at the face value, it says that a 10% increase in migrant ratio of a classroom would decrease local students' test scores on average by 1.16 points. However, given that schools with more migrant students are likely to be worse off in other respects, this estimates are likely contaminated by contextual effects. Therefore, in the second column of Table 15, we control for school fixed effects and only use variations across different classes within the same schools. Here we see that the coefficient becomes positive at 8.87 but not statistically significantly different from zero. As we have discussed in section 3.3, the "reduced-form" peer effect coefficient estimate is a composite and its sign depends on the relative magnitudes of its three components.

For our public school sample, we can also estimate a more restricted model compared to our baseline specification, assuming that the intensities of peer effects do not vary with peer types, i.e., $\lambda_{mm}^P = \lambda_{ml}^P = \lambda_{lm}^P = \lambda_{ll}^P = \lambda_0^P$. Estimates are reported in column (3) of Table 15. The estimate of λ_0^P is 0.797 and is significant at the 1% level. Note that the estimate here is in between the estimates reported in column (3) in Table 9, where all the four peer effects are allowed to be different. The estimated peer effect coefficient in public schools is still slightly less than in private schools, which is reported as 0.82 in Table 10.

6 Policy Simulations on School Desegregation

6.1 Simulation setup

Although the Chinese government has forbidden segregated classes for migrant students in public schools, somewhat ironically, segregated migrant-only private schools are still prevalent. To date, around 20% of migrant students in China are still enrolled in these low-quality private schools. Therefore, one of the most important policy questions regarding migrant children's education is whether they should all be entitled to public schools. Given that the literature has shown that migrant children perform better in public schools in terms of test scores (Chen and Feng, 2013; Lai et al. 2014; Chen and Feng 2017), this seems like a no brainer at a first glance. Nevertheless, even if the government is able to build enough public schools, there are still two remaining questions. First, what are the impacts of mixed schools/classes on local students? Many local parents do worry that the inflow of migrant students in public schools would adversely affect academic performances of their children (Chen and Feng, 2019), and in some cases, local parents have transferred their children out when migrant students moved in. Therefore, it is important to understand effects on all students when desegregation happens. Second, although the test scores of migrant students increase when they move into a public school from a private school, it is less clear what has contributed to this positive "treatment effect". If higher scores just come from better school quality, i.e., better school facilities and more qualified teachers in public schools, then it might be enough to just invest more in private schools. On the other hand, if peer effects are important, then it might be necessary for actions of desegregation at the school level. In this section, we try to answer these questions based on estimates from our model framework.

As an illustrative example, we perform a desegregation experiment of transferring fifth grade students from a migrant-only private school to a public school in our sample. Both schools are from the Pudong district. The sample summary statistics are reported in Table 16. There are 82 students in the private school, and 92 students in the public school, of which 63 are local students and 29 are migrants. As we can see, the overall pattern is similar to the whole sample as reported in Table 2. Students in the public school performed much better than those in the private school. But the test score difference between local and migrant students within the public school is small. In terms of individual and family characteristics, migrant students in the public school are much better than those in the private school. On the other hand, compared to migrant students in the public school, local students have higher parental education levels but lower family incomes on average.

In the experiment, each time we would randomly pick a local student (scenario 1) or a migrant student (scenario 2) from the public school and drop him or her from the sample. We then randomly pick one migrant student from the private school to replace that dropped student in the public school. This practice keeps the class sizes in the public school unchanged.[14] We continue this process until all local students (scenario 1) or migrant students (scenario 2) in the public school are replaced by the migrant students from the private school. We repeat this process for 500 times and calculate the averages.

6.2 Changes in test scores for different student groups

The results are reported in Table 17. We first focus on scenario 1, in which local students in the public school are replaced by migrant students from the private school. Simulation results based on our baseline estimates are reported in columns (1), (3), (5), and (8). First, column (1) shows that migrant students moved in enjoyed a huge test score gain. The test scores of those who have initially moved into the public school would increase by almost 19 points. However, this test score gain decreases as more and more migrants moved into the public school. When all local students are replaced, the test score gain for moved-in migrants almost halved to only 9.5 points. This suggests that peer effects might be important, as the newly moved-in migrants are in general worse than those moved-out local students in terms of observed and unobserved qualities. On the other hand, as column (3) suggests, migrant students who are originally in the public school suffer some test score loss, and the loss becomes quite large when all local students are replaced. Again, this should be due to the quality differences of their peers. For the remaining local students, the story is similar and their test scores would also decrease, as shown in column (5). Finally, column (8) says that average test scores of all students in the public school would initially decease slightly, but then increase as more and more migrant students are

[14] We have also tried a different scenario in which migrant students from the private school are moved into the public school, while all the original public school students stay. In that case, class sizes in the public school become larger. The results are similar and are available upon request.

moved in.[15]

The main takeaway from the above results is that there are important heterogeneities for different groups of students. Because of the importance of peer effects, when the local students are replaced by migrant students from private schools, the moved-in students would benefit but the original students in the school would suffer. Therefore, the decision of desegregation depends on the social welfare function. If we care more about the welfare (test score) of the low-performing migrants in the private schools, or we care more about equalizing test scores for different groups, then we would be more likely to favor desegregation with mixed schools and classes. If this is not the case, then policy makers may prefer segregation. It is not a straightforward decision.

Column (2), (4), (6), and (9) report simulation results when we force the peer effects in the public school to be the same, as shown in the last column of Table 15. Although the basic patterns are similar, results are somewhat different from the baseline. As more and more migrant students moved in, the homogenous peer effect specification would first overestimate and then underestimate benefits for the moved-in migrants, and consistently overestimate costs for the remaining migrant and local students. Finally, column (7) reports results based on the reduced-form regression, in which test scores of local students are regressed on migrant ratios at the class level. Based on the positive reduced-form estimates (refer to the column (2) of Table 15), we would conclude that the remaining local students would benefit from the inflow of migrant students. In addition, the reduced-form specification results cannot be used to examine test score changes of other student groups.

Table 17 also reports results when migrant students from the public school are gradually replaced by migrants from the private school (scenario 2). All local students would remain in the public school. Again, the general pattern is the same as in scenario 1. The moved-in migrants would benefit substantially in test scores, and the remaining students would suffer. However, in scenario 2, the change in peer quality is quite modest in the public school, as migrants from private schools are replacing migrants originally in the public school, not the more advantaged local students. Therefore, in all the columns we see the change to be relatively small when more and more students are replaced.

[15]Note that this only includes students currently in the public school. Students who were replaced and moved out of the school are not included.

6.3 Decomposing the test score change for the moved-in migrants

As shown in the previous subsection, migrants who are fortunate to switch from a private school to a public school could enjoy large test score gains. This is consistent with findings from the existing literature on a substantial "public-school premium" (Chen and Feng, 2013; Lai et al. 2014; Chen and Feng 2017). It is therefore important to understand sources of such gains. On the one hand, this "treatment effect" could be a result of studying in a better school, with better physical facilities, more qualified teachers and maybe better pedagogy. On the other hand, it could be that students benefit more from their stronger peers in public schools. Answers to this question has important implications on how to better educate migrant students in China's school system.

Using our framework, we can decompose the total change in test scores into three components. For a migrant student initially in the private school, his test score is $y_{0,i}^{m*}$ as defined in Equation (3.7). Suppose he is moved to a public school, his test score would then be $y_{1,i}^{m*}$ as defined in Equation (3.6). In addition, we may define the following hypothetical test score:

$$y_{1,i}^{m*1} = u_g^{'}\alpha^P + x_i^{'}\beta^P + \lambda_{mm}^M \sum_{k\in\mathfrak{M}} \frac{1}{n_g - 1} E[y_k^m|\mathfrak{I}_g] + \lambda_{mm}^M \sum_{h\in\mathfrak{L}} \frac{1}{n_g - 1} E[y_h^l|\mathfrak{I}_g] + \epsilon_{y,1,i}^m; \qquad (6.1)$$

where $y_{1,i}^{m*1}$ is person i's test score in the public school, with exactly the same peer relationships. The only difference with $y_{1,i}^{m*}$ in Equation (3.6) is that the peer effects are homogenous and equal to the private school peer effect λ_{mm}^M.

We can also define the hypothetical test score when student i is in the public school, but all his classmates are the same as in the private school, and the only difference is the school effect (i.e., including school facilities, teachers and so on).

$$y_{0,i}^{m*2} = u_g^{'}\alpha^P + x_i^{'}\beta^P + \lambda_{mm}^M \sum_{k\in\mathfrak{M}} \frac{1}{n_{g'} - 1} E[y_k^m|\mathfrak{I}_{g'}] + \epsilon_{y,0,i}^m. \qquad (6.2)$$

Please note that the expected test score difference between public school and private school $E(y_{1,i}^{m*} - y_{0,i}^{m*})$ can be then decomposed into three components. First, $E(y_{1,i}^{m*} - y_{1,i}^{m*1})$ denotes the effect of peer effect intensity, which captures the differences in peer effects (λs) in the two types of schools. Second, $E(y_{1,i}^{m*1} - y_{1,i}^{m*2})$ denotes the effect of peer quality. Finally, $E(y_{1,i}^{m*2} - y_{0,i}^{m*})$

denotes the school effect, holding the effects of peers unchanged.

We perform 500 simulations based on our baseline estimates and report average changes in Figure 1. The left panel illustrate scenario 1 when local students are replaced in the public school. For the total treatment effect, initially it was nearly 20 points, but as more and more local students are replaced, it declines steadily to slightly less than 10 points, as originally reported in Table 17. Peer quality is the most important component, initially almost the same as the total effect, but declines more rapidly to about 7 points. The school effect is about 6.7 point and remains constant. Finally, the peer intensity effect is negative, as the peer effect coefficient in private schools are larger than those in public schools. The intensity effect becomes less negative slightly as more and more migrant students enter the public school.

In the right panel of Figure 1 we report results for scenario 2 when migrant students from the public school are replaced. The overall pattern is similar to that in scenario 1. The school effect remains the same and unchanged as more and more students are replaced. The effect from peer quality was the same as in scenario 1 but declines much slowly, which explains why the total treatment effect also declines slowly compared to scenario 1. This is understandable as quality of the remaining peers are quite different in these two scenarios. Overall, the decomposition exercises show that peer quality could be very important in accounting for test score differences for migrant students in public schools and private schools.

7 Conclusion

In this paper, we study peer interactions among heterogenous agents under endogenous selection, using a game-theory based model. In our model, agents are first endogenously sorted into different peer groups. They then choose their best responses given their expectations of their peers' choices. In the equilibrium, all outcomes are jointly determined. The corresponding econometrics model provides a parsimonious framework to deal with endogenous social relations formed before interactions happen, and can fully incorporate heterogeneities in peer effects.

We apply the model to study peer effects among classmates in primary schools of Shanghai. We focus on the education of migrant students, who can either go to public schools, which have both migrant and local students, or go to low-quality private schools, which only enroll migrant students. We use a policy change in 2008 to identify the non-random school choices of migrant students. We then obtain significantly positive and large estimates for the four types of peer effects in public schools: migrant-to-migrant, local-to-migrant, migrant-to-local, and local-to-local. We also estimate the migrant-to-migrant peer effect coefficient in private schools. Our results are robust when peer contextual characteristics are controlled for, which suggests that the peer effect coefficients mainly capture contemporaneous peer interactions among students.

Finally, we perform policy experiments of desegregation using one public school and one private school in Pudong district. We randomly drop students from the public school and fill the vacancies left using migrant students from the private school. We show that migrant students who are moved to the public school enjoy substantial test score gains, but the remaining students from the public school would experience a decline in test score. We further decompose the test score gains for the moved-in migrant students into three components: school effect, which comes from different school facilities and qualities of teachers, etc.; peer quality effect, which comes from different qualities (expected test scores) of their peers; and peer intensity effect, which accounts for differences in the magnitudes of peer effect coefficients. We show that peer quality effect is the most important among the three, although its importance declines as more and more migrants are moved in. School effect is positive and constant, while the peer intensity effect is negative as peer effect coefficients are larger in private schools. Overall, our results suggest that peer effects should be an important consideration in the education of migrant students.

References

Angrist, J. and Lang, K. 2004. Does school integration generate peer effects? Evidence from Boston's Metco Program. *American Economic Review*, 94(5):1613–1634.

Ballatore, R. M., Fort, M., and Ichino, A. 2018. Tower of Babel in the classroom: Immigrants and natives in italian schools. *Journal of Labor Economics*, 36(4):885–921.

Bifulco, R., Fletcher, J. M., and Ross, S. L. 2011. The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy*, 3(1):25–53.

Brock, W. and Durlauf, S. 2001. Discrete choice with social interactions. *The Review of Economic Studies*, 68(2):235–260.

Card, D. and Rothstein, J. 2007. Racial segregation and the black-white test score gap. *Journal of Public Economics*, 91(11-12):2158–2184.

Carrell, S. E., Hoekstra, M., and Kuka, E. 2018. The long-run effects of disruptive peers. *American Economic Review*, 108(11):3377–3415.

Carrell, S. E., Page, M. E., and West, J. E. 2010. Sex and science: How professor gender perpetuates the gender gap. *The Quarterly Journal of Economics*, 125(3):1101–1144.

Carrell, S. E., Sacerdote, B. I., and West, J. E. 2011. From natural variation to optimal policy? The Lucas critique meets peer effects. *NBER Working Papers No 16865.*

Chen, Y. and Feng, S. 2013. Access to public schools and the education of migrant children in China. *China Economic Review*, 26:75–88.

Chen, Y. and Feng, S. 2017. Quality of migrant schools in China: Evidence from a longitudinal study in Shanghai. *Journal of Population Economics*, 30(3):1007–1034.

Chen, Y. and Feng, S. 2019. The education of migrant children in China's urban public elementary schools: Evidence from Shanghai. *China Economic Review*, 54:390–402.

Chen, Y., Feng, S., and Han, Y. 2019. Research on the education of migrant children in China: A review of the literature. *Frontiers of Economics in China*, 14(2):168–202.

Chen, Y., Feng, S., and Han, Y. 2020. The effect of primary school type on the high school opportunities of migrant children in China. *Journal of Comparative Economics*, 48(2):325–338.

Démurger, S., Gurgand, M., Li, S., and Yue, X. 2009. Migrants as second-class workers in urban China? A decomposition analysis. *Journal of Comparative Economics*, 37(4):610–628.

Figlio, D. N. 2007. Boys named sue: Disruptive children and their peers. *Education Finance and Policy*, 2(4):376–394.

Flores, R. and Scorezafave, L. 2014. Effect of racial segregation on proficiency of Brazilian elementary school students. *Economia*, 15(1):20–29.

Fruehwirth, J. C. 2013. Identifying peer achievement spillovers: Implications for desegregation and the achievement gap. *Quantitative Economics*, 4(1):85–124.

Fruehwirth, J. C. 2014. Can achievement peer effect estimates inform policy? A view from inside the black box. *The Review of Economics and Statistics*, 96(3):514–523.

Geay, C., McNally, S., and Telhaj, S. 2013. Non-Native Speakers of English in the Classroom: What are the Effects on Pupil Performance? *The Economic Journal*, 123(570):F281–F307.

Gong, J., Lu, Y., and Song, H. 2021. Gender peer effects on student's academic and noncognitive outcomes: Evidence and mechanisms. *Journal of Human Resources*, 56(3):686–710.

Gould, E., Lavy, V., and Paserman, M. 2009. Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence. *The Economic Journal*, 119(540):1243–69.

Guryan, J., Kroft, K., and Notowidigd, M. J. 2009. Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, 1(4):34–68.

Hoshino, T. 2019. Two-step estimation of incomplete information social interaction models with sample selection. *Journal of Business and Economic Statistics*, 37(4):598–612.

Hoxby, C. 2000. Peer effects in classroom: Learning from gender and race variation. *NBER working paper*, (No. 7867).

Hsieh, C. S. and Lee, L. 2016. A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*, 31(2):301–319.

Hu, F. 2018. Migrant peers in the classroom: Is the academic performance of local students negatively affected? *Journal of Comparative Economics*, 46(2):582–597.

Huang, B. and Zhu, R. 2020. Peer effects of low-ability students in the classroom: evidence from China's middle schools. *Journal of Population Economics*, 33(4):1343–1380.

Jensen, P. and Rasmussen, A. W. 2011. The effect of immigrant concentration in schools on native and immigrant children's reading and math skills. *Economics of Education Review*, 30 (6):1503–1515.

Lai, F., Liu, C., Luo, R., Zhang, L., Ma, X., Bai, B., Y.and Sharbono, and Rozelle, S. 2014. The education of China 's migrant children : The missing link in China's education system. *International Journal of Educational Development*, 37:68–77.

Lavy, V. and Schlosser, A. 2011. Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3(2):1–33.

Lavy, V., Paserman, M. D., and Schlosser, A. 2012. Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal*, 122(559):208–237.

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

Leung, M. 2015. Two-step estimation of network formation models with incomplete information. *Journal of Econometrics*, 188(1):182–195.

Lim, J. and Meer, J. 2017. The impact of teacher-student gender matches random assignment: Evidence from South Korea. *Journal of Human Resources*, 52(4):979–997.

Manski, C. 1993. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

Messinis, G. 2013. Returns to education and urban-migrant wage differentials in China: IV quantile treatment effects. *China Economic Review*, 26:39–55.

Nordin, M. 2013. Immigrant school segregation in Sweden. *Population Research and Policy Review*, 32(3):415–435.

Ohinata, A. and Van Ours, J. C. 2013. How immigrant children affect the academic achievement of native Dutch children. *The Economic Journal*, 123(570):F308–F331.

Schneeweis, N. 2015. Immigrant concentration in schools: Consequences for native and migrant students. *Labour Economics*, 35:63–76.

Sheng, S. 2020. A structural econometric analysis of network formation games through subnetworks. *Econometrica*, 88(5):1829–1858.

Wang, H., Cheng, Z., and Smyth, R. 2018. Do migrant students affect local students' academic achievements in urban china? *Economics of Education Review*, 63:64–77.

Xu, D., Zhang, Q., and Zhou, X. 2020. The impact of low-ability peers on cognitive and non-cognitive outcomes: Random assignment evidence on the effects and operating channels. *Journal of Human Resources*. doi: 10.3368/jhr.57.2.0718-9637R2. forthcoming.

Xu, X. and Zhang, X. 2016. Incorporating the family angle into public policy: An analysis based on the evolution of the policy of compulsory education for migrant children. *Social Sciences in China (in Chinese)*, 6:151–169.

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

Table 1: Comparison of school-level characteristics: sample schools vs. all schools

	Public schools		Private schools	
	(1)	(2)	(3)	(4)
	All schools	Sample schools	All schools	Sample schools
Number of teachers	61.4	59.2	49	55.3
	(31.9)	(22.9)	(29.8)	(11.3)
Number of students	797.3	712.2	1157.5	1315.1
	(531.6)	(340.1)	(485.5)	(361.3)
Teacher-student ratio	0.083	0.087	0.050	0.043
	(0.021)	(0.014)	(0.013)	(0.006)
Proportion of teachers with	0.568*	0.479	0.262	0.276
at least a middle rank title	(0.151)	(0.216)	(0.194)	(0.213)
Number of schools	355	11	32	9

Note: "Private schools" refers to migrant-only private schools. The numbers in parentheses are standard deviations. Significance of the difference between the sample schools and all schools at the 10%, 5% and 1% levels are represented by $*$, $**$, and $***$, respectively.

Table 2: Summary Statistics

	(1)	(2)	(3)
School Type	Public schools	Public schools	Private schools
Student Type	Local students	Migrant students	Migrant students
Math score	60.45	55.96***	44.31^^^
	(18.77)	(19.16)	(16.89)
Female	0.484	0.449	0.408^
	(0.500)	(0.498)	(0.492)
Rural *Hukou*	0.065	0.705***	0.948^^^
	(0.247)	(0.456)	(0.221)
Only Child in family	0.993	0.459***	0.286^^^
	(0.086)	(0.498)	(0.452)
Father's education HS and above	0.806	0.437***	0.241^^^
	(0.394)	(0.496)	(0.428)
Mother's education HS and above	0.770	0.302***	0.158^^^
	(0.421)	(0.459)	(0.365)
Age in months	136.3	138.8***	140.4^^^
	(4.41)	(7.70)	(11.28)
Attended Kindergarten	0.996	0.966***	0.880^^^
	(0.061)	(0.181)	(0.325)
Family monthly income between 3-5K	0.303	0.363**	0.433^^^
	(0.460)	(0.481)	(0.496)
Family monthly income above 5K	0.492	0.457	0.207^^^
	(0.500)	(0.498)	(0.405)
Origin Province of migrant students:			
Jiangsu		0.155	0.053^^^
		(0.362)	(0.224)
Anhui		0.278	0.488^^^
		(0.448)	(0.500)
Jiangxi		0.076	0.052^^
		(0.265)	(0.221)
Henan		0.071	0.113^^^
		(0.257)	(0.317)
Sichuan		0.064	0.100^^
		(0.245)	(0.300)
Lived in Central Shanghai before 2008		0.398	0.007^^^
		(0.490)	(0.082)
Lived in Peripheral Shanghai before 2008		0.427	0.767^^^
		(0.495)	(0.423)
Proportion of college degree teachers in school	0.649	0.645	0.186^^^
	(0.104)	(0.119)	(0.071)
Student-teacher ratio in school	12.56	11.43***	23.07^^^
	(2.747)	(2.592)	(1.425)
# of Local Peers	15.99	11.40***	
	(5.506)	(6.711)	
# of Migrant Peers	12.63	16.61***	45.60^^^
	(5.923)	(6.054)	(4.812)
No. Obs.	535	593	1337

Note: The numbers in parentheses are standard deviations. We denote statistical significance levels of the difference between local and migrant students in the public schools in column (2), where ∗∗∗, ∗∗, and ∗ stands for significance levels at the 1%, 5% and 10%, respectively. We denote statistical significance levels of the difference between migrant students in the public schools and private schools in column (3), where ^^^, ^^, and ^ stands for significance levels at the 1%, 5% and 10%, respectively.

Table 3: Migrant Ratio for Classes in Public Schools

(1) School	(2) Class	(3) Migrant Ratio	(4) School	(5) Class	(6) Migrant Ratio	(7) School	(8) Class	(9) Migrant Ratio
1	1	63.9%	5	1	94.1%	9	1	34.6%
	2	60.5%		2	100.0%		2	100.0%
	3	64.1%		3	27.8%			
2	1	83.9%	6	1	43.3%	10	1	9.1%
	2	65.6%		2	100.0%		2	66.7%
	3	63.3%		3	100.0%		3	52.4%
	4	50.0%					4	30.4%
3	1	55.6%	7	1	42.9%	11	1	3.1%
	2	71.4%		2	37.9%		2	50.0%
	3	50.0%		3	29.0%		3	45.5%
	4	50.0%		4	32.4%		4	41.2%
							5	60.6%
4	1	47.8%	8	1	28.1%			
	2	45.5%		2	30.0%			
	3	60.0%		3	36.7%			
	4	60.0%						
	5	50.0%						

Table 4: Correlations of predetermined characteristics between students and their classroom peers

Variables	(1) Coefficient	(2) s.e.	(3) T-statistics
Shanghai *Hukou*	0.020	(0.117)	0.168
Female	-0.134	(0.104)	-1.288
Only child	0.046	(0.095)	0.485
Father above high school	0.061	(0.063)	0.975
Mother above high school	0.025	(0.122)	0.204
Kindergarten	0.117	(0.164)	0.713
Family monthly income 3-5K	0.001	(0.079)	0.018
Family monthly income >5K	0.062	(0.063)	0.987

Note: Each row represents a regression in which the y variable is the student characteristic and the x variables include average peer characteristic and a school fixed effect.

Table 5: Test for standard deviations

	Real Data			Simulated Data		
	(1)	(2)	(3)=(2)/(1)	(4)	(5)	(6)=(5)/(4)
Shanghai *hukou*	0.253	0.051	20%	0.253	0.047	19%
Female	0.080	0.062	78%	0.084	0.068	81%
Only Child	0.236	0.071	30%	0.233	0.061	26%
Father high school & above	0.208	0.065	31%	0.208	0.063	30%
Mother high school & above	0.211	0.063	30%	0.209	0.059	28%
Kindergarten	0.078	0.040	52%	0.074	0.032	43%
Family monthly income 3-5K	0.101	0.072	72%	0.096	0.066	68%
Family monthly income >5K	0.171	0.074	44%	0.166	0.061	37%

Note: Authors' calculation based on real and simulated data.

Table 6: Resampling Tests

	(1) Kolmogorov-Smirnov Test	(2) Chi-squared Goodness-of-fit Test
Shanghai *Hukou*	0.612	0.585
Female	0.628	0.533
Only child	0.614	0.646
Father high school & above	0.336	0.379
Mother high school & above	0.721	0.968
Kindergarten	0.348	0.948
Family monthly income 3-5K	0.654	0.915
Family monthly income >5K	0.610	0.374

Note: Based on 1000 simulations.

Table 7: Correlations of Teachers' Characteristics with Students' Characteristics

Variables	(1) Female	(2) Experience	(3) Tenure in class	(4) College Degree
Shanghai *hukou*	0.390	34.95	0.359	-1.412
	(0.886)	(22.23)	(4.761)	(0.870)
Female	0.860	14.21	8.500**	-0.381
	(0.773)	(20.07)	(3.999)	(0.789)
Only child	0.115	-0.105	0.922	-0.430
	(0.682)	(17.57)	(3.657)	(0.686)
Father high school & above	-0.305	28.09	1.574	-1.488**
	(0.748)	(18.80)	(4.009)	(0.721)
Mother high school & above	-0.520	26.91	3.677	-1.279*
	(0.750)	(18.98)	(4.005)	(0.376)
Kindergarten	-1.011	33.96	-3.760	0.486
	(1.308)	(33.51)	(7.039)	(1.327)
Family monthly income 3-5K	0.984	23.28	-5.390	1.148*
	(0.660)	(17.06)	(3.536)	(0.661)
Family monthly income >5K	-0.776	23.61	3.103	-0.176
	(0.605)	(15.46)	(3.272)	(0.622)

Note: Each coefficient is from a separate regression in which the y variable is the teacher's characteristic (one in each column) and the x variable is the student characteristic (one in each row), controlling for school fixed effects. The numbers in parentheses are standard errors. $***$, $**$, and $*$ stands for significance levels at the 1%, 5% and 10%, respectively.

Table 8: Estimations for School-type Selection of Migrant Students

Variables	(1) Parameter	(2) Standard Error
Female	0.138*	(0.081)
Rural *hukou*	-1.129***	(0.110)
One child	0.319***	(0.082)
Father high school & above	0.131	(0.101)
Mother high school & above	0.151	(0.108)
Age in month	-0.006	(0.005)
Kindergarten	0.558***	(0.171)
Family monthly income between 3k-5k	0.199*	(0.104)
Family monthly income above 5k	0.648***	(0.111)
Origin Provinces:		
Jiangsu	0.095	(0.140)
Auhui	-0.428***	(0.096)
Jiangxi	-0.171	(0.180)
Henan	-0.302*	(0.163)
Sichuan	-0.426***	(0.155)
Parental Residence before 2008 (a):		
In Central Shanghai	2.589***	(0.175)
In Peripheral Shanghai	-0.285***	(0.087)

Note: Total number of observations is 1,930. $***$, $**$, and $*$ stands for significance levels at the 1%, 5% and 10%, respectively. (a) The omitted group are those who moved to Shanghai after 2008.

Table 9: Peer effects on Math Test Scores in Public Schools

Variables	(1)	(2)	(3)
	Panel A: Regression results		
Female	-0.600	-1.994**	-1.710*
	(0.618)	(0.886)	(0.913)
One child	2.642***	2.2290*	0.639
	(0.886)	(1.147)	(1.267)
Father education above high school	0.966	3.554***	2.802**
	(0.738)	(1.136)	(1.194)
Mother education above high school	1.769**	1.155	0.081
	(0.740)	(1.098)	(1.097)
Age in month	0.002	0.038	-0.027
	(0.033)	(0.045)	(0.075)
Kindergarten	3.155	0.322	3.420
	(2.116)	(2.851)	(3.240)
Family monthly income between 3k-5k	1.966**	2.390*	1.735
	(0.924)	(1.247)	(1.293)
Family monthly income above 5k	3.213***	5.091***	4.754***
	(1.013)	(1.3040)	(1.363)
Inverse Mills ratio		2.897***	3.157**
		(0.774)	(1.290)
λ^P_{mm}	0.905***	0.905***	0.802***
	(0.032)	(0.038)	(0.068)
λ^P_{ml}	0.868***	0.835***	0.764***
	(0.039)	(0.047)	(0.080)
λ^P_{lm}	0.894***	0.589***	0.822***
	(0.049)	(0.094)	(0.090)
λ^P_{ll}	0.906***	0.685***	0.815***
	(0.035)	(0.072)	(0.083)
School Dummies	No	No	Yes
Origin Provinces	Yes	Yes	Yes
No. Observed Scores	1105	1105	1105
No. Obs.	1128	1128	1128
	Panel B: Hypotheses tests		
Overall homogeneity			
$H_0 : \lambda^P_{mm} = \lambda^P_{ml} = \lambda^P_{lm} = \lambda^P_{ll}$ vs $H_1 : \lambda^P_{mm} = \lambda^P_{ml} = \lambda^P_{lm} = \lambda^P_{ll}$ *does not hold*	0.060*	0.003***	0.837
Stronger within			
$H_0 : \lambda^P_{mm} = \lambda^P_{ml}$ vs $H_1 : \lambda^P_{mm} \neq \lambda^P_{ml}$	0.104	0.009***	0.400
$H_0 : \lambda^P_{mm} \geq \lambda^P_{ml}$ vs $H_1 : \lambda^P_{mm} < \lambda^P_{ml}$	0.948	0.996	0.800
$H_0 : \lambda^P_{lm} = \lambda^P_{ll}$ vs $H_1 : \lambda^P_{lm} \neq \lambda^P_{ll}$	0.600	0.008***	0.913
$H_0 : \lambda^P_{lm} \leq \lambda^P_{ll}$ vs $H_1 : \lambda^P_{lm} > \lambda^P_{ll}$	0.700	0.996	0.457
Symmetry			
$H_0 : \lambda^P_{ml} = \lambda^P_{lm}$ vs $H_1 : \lambda^P_{ml} \neq \lambda^P_{lm}$	0.697	0.016	0.621

Note: λ^P_{mm}, λ^P_{ml}, λ^P_{lm}, and λ^P_{ll} measure the migrant-to-migrant, local-to-migrant, migrant-to-local, and local-to-local peer effects in public schools. The numbers in parentheses are standard errors. $***$, $**$, and $*$ stands for significance levels at the 1%, 5% and 10%, respectively.

Table 10: Peer effects on Math Test Scores in Private Schools

Variables	(1)	(2)	(3)
Female	-0.702	-0.649	-1.864**
	(0.459)	(0.441)	(0.803)
One child	-0.518	-0.445	0.398
	(0.345)	(0.314)	(0.857)
Father education above high school	0.486	0.524	1.325
	(0.346)	(0.378)	(1.025)
Mother education above high school	-0.251	-0.295	3.113**
	(0.242)	(0.258)	(1.245)
Age in month	-0.000	0.000	0.032
	(0.003)	(0.003)	(0.027)
Kindergarten	-0.143	-0.107	2.016*
	(0.184)	(0.164)	(1.205)
Family monthly income between 3k-5k	-0.262	-0.264	0.786
	(0.195)	(0.198)	(0.844)
Family monthly income above 5k	0.805	0.786	1.911
	(0.533)	(0.539)	(1.227)
Inverse Mills ratio		0.0970	2.093*
		(0.102)	(1.136)
λ_{mm}^M	0.985***	0.986***	0.820***
	(0.010)	(0.010)	(0.069)
School Dummies	No	No	Yes
Origin Provinces	Yes	Yes	Yes
No. Observed Scores	1299	1299	1299
No. Obs.	1337	1337	1337

Note: λ_{mm}^M denotes the peer effects in private schools where there are only migrant students. The numbers in parentheses are standard errors. $***$, $**$, and $*$ stands for significance levels at the 1%, 5% and 10%, respectively.

Table 11: Robustness checks: Different Ways of Treating Segregated Classes in Public Schools

	(1) Ignore Segregated Classes		(2) Delete Segregated Classes		(3) Treat Migrant-dominated Classes as Private Schools	
	(1) Public Schools	(2) Private Schools	(3) Public Schools	(4) Private Schools	(5) Public Schools	(6) Private Schools
Female	-1.732**	-1.864**	-2.728**	-1.858**	-2.396**	-0.850
	(0.878)	(0.803)	(1.144)	(0.804)	(1.034)	(0.640)
One child	0.485	0.398	0.762	0.396	0.756	0.580
	(1.217)	(0.857)	(1.653)	(0.859)	(1.480)	(0.714)
Father education above high school	2.917**	1.325	4.574***	1.325	2.879**	1.206
	(1.184)	(1.025)	(1.457)	(1.029)	(1.306)	(0.878)
Mother education above high school	-0.013	3.113**	-0.285	3.149**	0.281	2.509**
	(1.083)	(1.245)	(1.344)	(1.250)	(1.193)	(1.110)
Age in month	-0.021	0.032	-0.040	0.029	-0.014	0.036
	(0.059)	(0.027)	(0.093)	(0.027)	(0.087)	(0.023)
Kindergarten	2.559	2.016*	7.414	2.125*	11.226*	2.002*
	(2.984)	(1.205)	(7.996)	(1.235)	(6.317)	(1.059)
Family monthly income between 3k-5k	2.129*	0.786	1.877	0.794	1.858	0.381
	(1.237)	(0.844)	(1.643)	(0.849)	(1.483)	(0.691)
Family monthly income above 5k	4.563***	1.911	5.078***	1.900	4.462***	2.305**
	(1.335)	(1.227)	(1.646)	(1.233)	(1.481)	(1.059)
Inverse Mills ratio	2.572**	2.093*	3.677**	2.107*	1.051	1.538**
	(1.223)	(1.136)	(1.704)	(1.200)	(1.063)	(0.770)
λ_{mm}^P	0.850***		0.582***		0.770***	
	(0.051)		(0.142)		(0.116)	
λ_{ml}^P	0.771***		0.580***		0.774***	
	(0.062)		(0.136)		(0.113)	
λ_{lm}^P	0.753***		0.848***		0.805***	
	(0.088)		(0.114)		(0.103)	
λ_{ll}^P	0.826***		0.811***		0.754***	
	(0.063)		(0.104)		(0.092)	
λ_{mm}^M		0.820***		0.817***		0.878***
		(0.069)		(0.071)		(0.041)
School Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Origin Provinces	Yes	Yes	Yes	Yes	Yes	Yes
No. Observed Scores	1105	1299	873	1299	997	1407
No. Obs.	1128	1337	890	1337	1018	1447

Note: λ_{mm}^P, λ_{ml}^P, λ_{lm}^P, and λ_{ll}^P measure the migrant-to-migrant, local-to-migrant, migrant-to-local, and local-to-local peer effects in public schools, respectively. λ_{mm}^M measures migrant-to-migrant peer effects in migrant-only private schools. The numbers in parentheses are standard errors. $***$, $**$, and $*$ stands for significance levels at the 1%, 5% and 10%, respectively.

Table 12: Robustness Checks: Adding average peer characteristics

Variables	(1)	(2)	(3)
Public Schools			
(a) Control Peer Family Income			
λ^P_{mm}	0.905***	0.905***	0.821***
	(0.035)	(0.036)	(0.076)
λ^P_{ml}	0.869***	0.849***	0.786***
	(0.042)	(0.049)	(0.091)
λ^P_{lm}	0.893***	0.887***	0.847***
	(0.054)	(0.057)	(0.089)
λ^P_{ll}	0.906***	0.907***	0.847***
	(0.039)	(0.040)	(0.086)
(b) Control Peer Fathers' Education			
λ^P_{mm}	0.905***	0.905***	0.807***
	(0.033)	(0.036)	(0.078)
λ^P_{ml}	0.874***	0.855***	0.771***
	(0.044)	(0.055)	(0.101)
λ^P_{lm}	0.885***	0.883***	0.829***
	(0.051)	(0.055)	(0.103)
λ^P_{ll}	0.906***	0.907***	0.825***
	(0.039)	(0.042)	(0.104)
Private Schools			
(a) Control Peer Family Income			
λ^M_{mm}	0.990***	0.990***	0.837***
	(0.012)	(0.011)	(0.063)
(b) Control Peer Fathers' Education			
λ^M_{mm}	0.990***	0.990***	0.961***
	(0.010)	(0.010)	(0.021)
Selection	No	Yes	Yes
School Dummies	No	No	Yes
Origin Provinces	Yes	Yes	Yes

Note: λ^P_{mm}, λ^P_{ml}, λ^P_{lm}, and λ^P_{ll} measure the migrant-to-migrant, local-to-migrant, migrant-to-local, and local-to-local peer effects in public schools. λ^M_{mm} measures migrant-to-migrant peer effects in migrant-only private schools. The numbers in parentheses are standard errors. $***$, $**$, and $*$ stands for significance levels at the 1%, 5% and 10%, respectively.

Table 13: Estimated Peer Effects for Randomly Reassigned Classes

		(1) Actual Data	(2) Simulated Data (a): Random assignment within Schools	(3): Simulated Data (b): Random assignment across Schools
λ_{mm}^P		0.802	-0.001	-0.029
		(0.068)	(1.454)	(0.832)
			[0.645]	[0.700]
	Percentage significant at 5%		23%	13%
λ_{ml}^P		0.764	0.018	-0.020
		(0.080)	(1.426)	(0.771)
			[0.621]	[0.662]
	Percentage significant at 5%		22%	15%
λ_{lm}^P		0.822	-0.076	-0.207
		(0.090)	(1.460)	(0.870)
			[0.655]	[0.702]
	Percentage significant at 5%		18%	10%
λ_{ll}^P		0.815	-0.067	-0.162
		(0.083)	(1.430)	(0.808)
			[0.625]	[0.614]
	Percentage significant at 5%		17%	10%
λ_{mm}^M		0.820	-0.071	-0.118
		(0.069)	(2.421)	(1.643)
			[0.781]	[0.658]
	Percentage significant at 5%		35%	25%

Note: The numbers in the parentheses are the (mean) of theoretical standard deviations calculated based on the estimates' large sample properties and the figures in the brackets are the sample standard deviations of the parameter estimates over 500 simulations. The percentage of the estimates which are 5% significant across 500 simulated samples are also reported. $***$, $**$, and $*$ stands for significance levels at the 1%, 5% and 10%, respectively.

Table 14: Average Marginal Effects of Personal Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
	Father's Education			Family Income		
			Public Schools			
Migrant Students						
Average effect on oneself	1.238	4.082	3.128	4.114	5.848	5.307
Average effect on migrant peers	0.302	0.644	0.406	1.005	0.923	0.690
Average effec on local peers	0.243	0.290	0.343	0.807	0.416	0.582
Local Students						
Average effect on oneself	1.224	3.788	3.125	4.068	5.426	5.302
Average effect on migrant peers	0.282	0.421	0.383	0.936	0.603	0.650
Average effect on local peers	0.287	0.317	0.402	0.954	0.455	0.682
			Private Schools			
Migrant Students						
Average effect on oneself	1.150	1.309	1.433	1.905	1.965	2.066
Average effect on peers	0.674	0.796	0.131	1.117	1.195	0.189
Selection	No	Yes	Yes	No	Yes	Yes
School Dummies	No	No	Yes	No	No	Yes
Origin Provinces	Yes	Yes	Yes	Yes	Yes	Yes

Note: Results show the effects of one's father's education increasing from below high school to high school and above (columns 1 to 3), or one's monthly family income increasing from less than 3 thousand RMB to above 5 thousand RMB. The specifications reported are the same as in Table 9 and Table 10.

Table 15: Results from alternative models for public schools

Samples	(1) Local students	(2) Local students	(3) Local and migrant students
Female	-2.122	-2.513	-1.849**
	(1.803)	(1.713)	(0.896)
One child			0.447
			(1.195)
Father education above high school	6.443**	6.864***	2.846**
	(2.521)	(2.399)	(1.193)
Mother education above high school	-0.203	-1.931	0.158
	(2.431)	(2.315)	(1.095)
Age in month	-0.036	-0.007	-0.025
	(0.204)	(0.194)	(0.075)
Kindergarten			3.520
			(3.128)
Family monthly income between 3k-5k	2.948	1.345	1.787
	(2.554)	(2.429)	(1.287)
Family monthly income above 5k	7.963***	6.149***	4.729***
	(2.462)	(2.348)	(1.357)
Migrant ratio in class	-11.589*	8.872	
	(6.865)	(11.264)	
Inverse Mills Ratio			2.993**
			(1.211)
λ_0^P			0.797***
			(0.055)
School Dummies	No	Yes	Yes
Origin Provinces	Yes	Yes	Yes
No. Observed Scores	426	426	1105
No. Obs.	436	436	1128

Note: Columns (1) and (2) report results from the linear reduced-form regressions of local students' test scores on class-level migrant ratios. Column (1) does not control for school fixed effects while column (2) does. Column (3) uses the same specification as the baseline reported in column (3) of Table 9, except that the peer effects are restricted to be the same, λ_0^P. Significance at 10%, 5% and 1% is marked with $*$, $**$, and $***$ respectively.

Table 16: Summary statistics for a public and a private school in Pudong district

	(1)	(2)	(3)
School Type	Public schools	Public schools	Private schools
Student Type	Local students	Migrant students	Migrant students
Math test scores	64.9	62.9	29.6
	(16.7)	(14.8)	(14.1)
Female	0.571	0.379	0.415
	(0.495)	(0.485)	(0.493)
Only child in family	1.00	0.517	0.268
	(0)	(0.500)	(0.443)
Father education high school & above	0.730	0.552	0.122
	(0.444)	(0.497)	(0.327)
Mother education high school & above	0.794	0.310	0.110
	(0.405)	(0.463)	(0.313)
Age in month	135.7	136.7	145.0
	(4.55)	(8.15)	(11.6)
Kindergarten	1.00	1.00	0.902
	(0)	(0)	(0.297)
Family monthly income between 3-5K	0.365	0.207	0.366
	(0.481)	(0.405)	(0.482)
Family monthly income above 5K	0.524	0.724	0.171
	(0.499)	(0.447)	(0.376)
Origin Provinces of Migrants:			
Jiangsu		0.172	0.024
		(0.378)	(0.154)
Anhui		0.448	0.305
		(0.497)	(0.460)
Jiangxi		0.034	0.024
		(0.182)	(0.154)
Henan			0.171
			(0.376)
Sichuan		0.034	0.122
		(0.182)	(0.327)
No. Obs	63	29	82

Note: Authors' calculations based on two schools (one private and one public) from Pudong district in our sample. Numbers in parentheses are standard deviations.

Table 17: Average changes in test score for different student groups in the public school in Pudong

	New Migrants		Remaining Migrants		Remaining Locals			All	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Hetero	Homo	Hetero	Homo	Hetero	Homo	RF	Hetero	Homo
Scenario 1: % of Locals Replaced									
10%	18.906	19.839	-1.353	-1.509	-1.526	-1.488	0.563	-0.139	-0.104
20%	17.687	18.378	-2.629	-3.017	-2.945	-2.950	1.123	-0.154	-0.189
30%	16.543	16.934	-3.799	-4.483	-4.270	-4.390	1.684	-0.049	-0.247
40%	15.293	15.263	-5.085	-6.193	-5.733	-6.074	2.342	0.185	-0.313
50%	14.263	13.812	-6.158	-7.694	-6.934	-7.531	2.906	0.453	-0.391
60%	13.236	12.307	-7.209	-9.224	-8.129	-9.039	3.467	0.753	-0.513
70%	12.112	10.571	-8.330	-10.969	-9.437	-10.790	4.134	1.218	-0.630
80%	11.240	9.120	-9.225	-12.451	-10.438	-12.224	4.693	1.726	-0.696
90%	10.414	7.675	-10.072	-13.927	-11.403	-13.672	5.261	2.296	-0.759
100%	9.469	5.954	-11.032	-15.674				3.007	-0.864
Scenario 2: % of Migrants Replaced									
10%	19.408	20.544	-0.399	-0.393	-0.459	-0.432		-0.010	0.035
20%	18.924	20.069	-0.983	-0.970	-1.129	-1.062		-0.001	0.110
30%	18.255	19.405	-1.585	-1.566	-1.816	-1.710		-0.018	0.159
40%	17.713	18.870	-2.142	-2.115	-2.468	-2.324		0.009	0.251
50%	17.116	18.282	-2.739	-2.702	-3.150	-2.966		0.001	0.310
60%	16.517	17.694	-3.304	-3.262	-3.847	-3.621		-0.013	0.364
70%	16.029	17.218	-3.855	-3.803	-4.487	-4.222		0.034	0.480
80%	15.451	16.647	-4.490	-4.428	-5.171	-4.866		0.029	0.541
90%	14.883	16.088	-5.095	-5.025	-5.826	-5.482		0.050	0.629
100%	14.322	15.536			-6.507	-6.122		0.059	0.705

Note: "hetero" refers to the baseline model in our paper with heterogenous peer effects as reported in the last column of Table 9. "homo" refers to the model with homogenous peer effects in public schools as reported in the last column of Table 15. "RF" refers to the reduced-form regression model as reported in column (2) of Table 15.

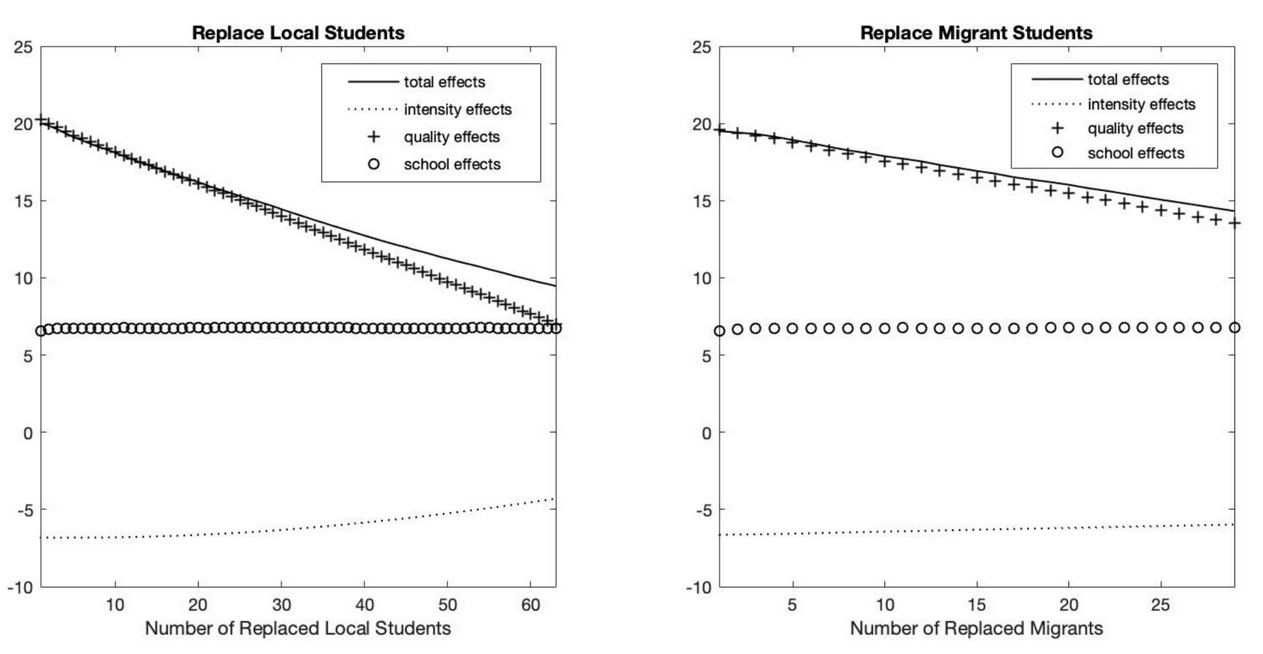

Figure 1: Decomposing the average test score changes of migrant students moving from a private school to a public school in Pudong

Note: The left panel corresponds to scenario 1 where local students are replaced in the public school. In the right panel, migrant students are replaced in the public school. Results based on 500 simulations.

Appendices

A Micro-foundations

Our econometric framework is built on the following behavioral model. There are two time periods, $t = 0, 1$. At $t = 0$, a migrant child is selected to either a public school or a private one. Whether a migrant child i is enrolled in a public or a private school is denoted by a binary variable, S_i^m. $S_i^m = 1$ if i is affiliated to a public school; and $S_i^m = 0$ if i enrolls in a private one. In our paper, $S_i^m = I(z_i^{m'}\gamma + \epsilon_{s,i}^m > 0)$, where z_i^m and $\epsilon_{s,i}^m$ represent respectively observed and unobserved individual characteristics. At $t = 0$, parents do not know about the detail characteristics of students in a school, and only have prior information about schools, such as public schools on average excel in teachers' quality and facility, compared with the private schools. They would try to enter better schools given their own qualification. Hence, we do not consider peer effects in the stage of school type selection.

After enrolled in a school of either type, a student will be randomly assigned to a class. We focus on interactions among students within a class, which is called a "group", g. Different groups are independent of each other. A student's utility depends on his/her own performances, such as math score, y_i, efforts e_i and peer's performance y_{-i}. Specifically,

$$U(y_i, y_{-i}, e_i) = a_i + by_i - \frac{c}{2}y_i^2 + \sum_{j \neq i} \rho_{ij} y_i y_j - \frac{d}{2}e_i^2, \tag{A.1}$$

where a_i include personal and school features. b, c, and d are parameters. When $d > 0$, i gets disutility by making more efforts. In addition to his/her own performances y_i, i's utility may also depend on the performances of his/her classmates due to externalities effects (See Calvó-Armengol et al. (2009), Epple and Romano (1998), and Davezies et al (2009) for example.), which is captured by the term $\sum_{j \neq i} \rho_{ij} y_i y_j$. The marginal utility for y_i is $b - cy_i + \sum_{j \neq i} \rho_{ij} y_j$. That is, the marginal utility varies with one's own score, y_i, and may also be affected by others' performances, y_{-i}. If $\rho_{ij} > 0$, i is happier with own score increases when his/her classmate j's score also increases; and if $\rho_{ij} < 0$, i's marginal utility is higher when his/her own scores rises while his/her classmate j's score decreases.

The performance y_i, depends on i's own efforts, e_i, as well as some random shocks.

$$y_i = f_i + e_i + \eta_{u,i}. \tag{A.2}$$

$f_i = f(\widetilde{u}_g, x_i, \eta_{o,i})$, where $\widetilde{u}_g$ and x_i respectively represent group and personal characteristics, $\eta_{o,i}$ includes some personal traits observed only by i himself/herself and $\eta_{u,i}$ represents the unobservable idiosyncratic shocks. Individual i chooses efforts e_i to maximize the expected utility, $E[U(y_i, y_{-i}, e_i)|\Im_i]$, based on his/her information set, $\Im_i$, which contains the public information $\Im_g$ (the school-class features, $\widetilde{u}_g$, and the predetermined personal features for all students in the class, X) and personal information $\eta_{o,i}$. Assume that $(\eta_{o,i}, \eta_{u,i})$'s are independent of $\widetilde{u}_g$ and X, have zero means, and are independent across all the students. In addition, $\eta_{o,i}$ is independent of $\eta_{u,i}$ for every i, and $Var(\eta_{u,i}) = \sigma_u^2$. In this case, conditional on $\Im_i$, y_i and y_j are independent. Namely, $E[y_i y_j|\Im_i] = E[y_i|\Im_i]E[y_j|\Im_i]$, if $i \neq j$. Thus,

$$E[U(y_i, y_{-i}, e_i)|\Im_i] = a_i + bf_i + be_i - \frac{c}{2}f_i^2 - cf_ie_i - \frac{c}{2}e_i^2 - \frac{c}{2}\sigma_u^2 + (f_i + e_i)\sum_{j\neq i}\rho_{ij}E[y_j|\Im_i] - \frac{d}{2}e_i^2.$$

With $c > 0$ and $d > 0$, this is a quadratic function of efforts. The peer effect is realized by choosing the best effort level based peer's performance. The first-order condition follows,

$$e_i^* = \frac{1}{c+d}(b - cf_i + \sum_{j\neq i}\rho_{ij}E[y_j|\Im_i]). \tag{A.3}$$

As all the students will choose their best response, from (A.3), we would get a Bayesian Nash Equilibrium (BNE) for studying efforts, e^*, which satisfies the following equation system,

$$e_i^* = \frac{1}{c+d}(b - cf_i + \sum_{j\neq i}\rho_{ij}E[f_j + e_j^* + \eta_{u,j}|\Im_i]).$$

Although efforts are usually unobservable from the data set, the equilibrium outcomes (y_i^*, y_{-i}^*) are observed. From the relationship between performance and efforts, we can derive the interactions between observed outcomes,

$$y_i^* = \frac{b}{c+d} + \frac{d}{c+d}f_i + \frac{1}{c+d}\sum_{j\neq i}\rho_{ij}E[y_j^*|\Im_i] + \eta_{u,i}. \tag{A.4}$$

Specifically, assume that f_i is a linear function of group and individual features, i.e., $f_i = \widetilde{u}_g'\widetilde{\alpha} + x_i'\widetilde{\beta} + \eta_{o,i}$, we get

$$y_i^* = \frac{b}{c+d} + \frac{d}{c+d}\widetilde{u}_g'\widetilde{\alpha} + \frac{d}{c+d}x_i'\widetilde{\beta} + \frac{1}{c+d}\sum_{j\neq i}\rho_{ij}E[y_j^*|\Im_i] + \frac{d}{c+d}\eta_{o,i} + \eta_{u,i}.$$

Reorganizing terms on group and individual features and let $\epsilon_{y,i} \equiv \frac{d}{c+d}\eta_{o,i} + \eta_{u,i}$, we derive the following equation for observed equilibrium outcomes,

$$y_i^* = u_g'\alpha + x_i'\beta + \frac{1}{c+d}\sum_{j\neq i}\rho_{ij}E[y_j^*|\Im_i] + \epsilon_{y,i}. \tag{A.5}$$

In addition, like Nakajima (2007), we assume that after the exogenous characteristics are controlled, the marginal effects of peer's outcomes depend on the peer's "type", i.e., "l" for "local" and "m" for "migrant" in our case.

Assumption A.1 *For any $i \neq j$ and $i' \neq j'$, $\rho_{ij} = \rho_{i'j'}$ whenever i and i' are of the same type and j and j' are of the same type.*

It follows that in public schools, $\frac{\rho_{ij}}{c+d}$ captures the peer effect for two students in a public school class, and equals to $\lambda^P_{mm} W^P_{mm,ij}$, $\lambda^P_{ml} W^P_{ml,ij}$, $\lambda^P_{lm} W^P_{lm,ij}$, $\lambda^P_{ll} W^P_{ll,ij}$, when i or j are either local or migrant. Similarly, for private schools, $\frac{\rho_{ij}}{c+d} = \lambda^M_{mm} W^M_{mm,ij}$ for any two migrants who study in the same class. Then we derive outcome equations for migrant and local students in public schools, Eq. (3.3), and (3.5), and also migrants in private schools, Eq. (3.4).

Existence and uniqueness of the Bayesian Nash Equilibrium (BNE) is vital for model completeness and parameter identification. Yang and Lee (2017) provide sufficient conditions for equilibrium existence and uniqueness when there is only one type of peers. In our model, there are four types of peer relations. Hence, we need to derive sufficient conditions for our framework as follows.

Corollary A.1 *With Assumptions (3.1) and (3.2), for any two agents $i \neq j$, the conditional expected performance just depends on public information, that it, $E[y_i|\mathfrak{I}_j] = E[y_i|\mathfrak{I}_g]$.*

For example, i is a migrant student in a public school. Taking expectation on Eq. (3.3), based on $\mathfrak{I}_j = (\mathfrak{I}^P_g, \epsilon_j)$, for $i \neq j$,

$$\begin{aligned} E[y^m_i|\mathfrak{I}^P_j] =& E\Big[u'_g\alpha^P + x'_i\beta^P + \epsilon^m_{y,1,i}\Big|\mathfrak{I}^P_g, \epsilon_j\Big] \\ &+ E\Big[\lambda^P_{mm}\sum_k W^P_{mm,ik} E[y_k|\mathfrak{I}^P_g, \epsilon_i] + \lambda^P_{ml}\sum_h W^P_{ml,ih} E[y_h|\mathfrak{I}^P_g, \epsilon_i]\Big|\mathfrak{I}^P_g, \epsilon_j\Big]. \end{aligned}$$

Since u_g and x_i's of all class members are included in $\mathfrak{I}^P_g$ and exogenous, $E[u'_g\alpha^P + x'_i\beta^P|\mathfrak{I}^P_g, \epsilon_j] = u'_g\alpha^P + x'_i\beta^P$. As the idiosyncratic shocks are independent across agents, $E[\epsilon^m_{y,1,i}|\mathfrak{I}^P_g, \epsilon_j] = E[\epsilon^m_{y,1,i}|\mathfrak{I}^P_g]$. Because the peer relations, such as $W^P_{mm,ik}$'s, are included in $\mathfrak{I}^P_g$ and ϵ_i is independent of ϵ_j,

$$\begin{aligned} &E[\lambda^P_{mm}\sum_k W^P_{mm,ik} E[y_k|\mathfrak{I}^P_g, \epsilon_i] + \lambda^P_{ml}\sum_h W^P_{ml,ih} E[y_h|\mathfrak{I}^P_g, \epsilon_i]\Big|\mathfrak{I}^P_g, \epsilon_j] \\ &= E[\lambda^P_{mm}\sum_k W^P_{mm,ik} E[y_k|\mathfrak{I}^P_g, \epsilon_i] + \lambda^P_{ml}\sum_h W^P_{ml,ih} E[y_h|\mathfrak{I}^P_g, \epsilon_i]\Big|\mathfrak{I}^P_g]. \end{aligned}$$

The cases for local students in public schools and migrant children in private schools can be proved in a similar way. Following from this corollary, the conditional expectations are functions of public information. Then we can replace the information sets in the equations for equilibrium outcomes by the public information in a class, i.e., replacing $\mathfrak{I}_i$ in Eq. (3.3) and $\mathfrak{I}_j$ in Eq. (3.5) by $\mathfrak{I}_g$ and changing $\mathfrak{I}_i$ in Eq. (3.4) into $\mathfrak{I}_{g'}$.

Additionally, given public information which is observed from the data, we can represent the expected performances of students in a class by a vector. For example, in a public school class g where there are n_g students, given $\mathfrak{I}_g^P$, we define an $n_g \times 1$ vector, ψ_g^P, whose i-th element $\psi_{g,i}^P$, is equal to the expected performance of i, $E[y_i|\mathfrak{I}_g^P]$. We can use superscript to signify students' types, i.e., $E[y_i^m|\mathfrak{I}_g^P]$ if i is migrant and $E[y_i^l|\mathfrak{I}_g^P]$ if i is local. Using ψ_g^P, from Eq. (3.3), the BNE expected outcome of a migrant can be derived from the following equation,

$$\psi_{g,i}^P = u_g^{'}\alpha^P + x_i^{'}\beta^P + E[\epsilon_{y,1,i}^m|\mathfrak{I}_g^P] + \lambda_{mm}^P \sum_k W_{mm,ik}^P \psi_{g,k}^P + \lambda_{ml}^P \sum_h W_{ml,ih}^P \psi_{g,h}^P. \quad \text{(A.6)}$$

According to Eq. (3.5) the BNE expected outcome of a local student is determined analogously,

$$\psi_{g,j}^P = u_g^{'}\alpha^P + x_j^{'}\beta^P + \lambda_{lm}^P \sum_k W_{lm,jk}^P \psi_{g,k}^P + \lambda_{ml}^P \sum_h W_{ll,jh}^P \psi_{g,h}^P, \quad \text{(A.7)}$$

as $E[\epsilon_{y,,j}^l|\mathfrak{I}_g^P] = 0$. For class $g^{'}$ in private schools, we can define $\psi_{g^{'}}^M$ in a similar way.

$$\psi_{g^{'},i}^M = u_{g^{'}}^{'}\alpha^M + x_j^{'}\beta^M + E[\epsilon_{y,0,i}^m|\mathfrak{I}_{g^{'}}^M] + \lambda_{mm}^M \sum_k W_{mm,ik}^M \psi_{g^{'},k}^M. \quad \text{(A.8)}$$

Therefore, a BNE would be equivalently represented by a vector ψ_g^P satisfying (A.6) and (A.7) for a class in public schools or a vector $\psi_{g^{'}}^M$ satisfying (A.8) for a private school class. Employing the contraction mapping theorem for vectors, we can derive sufficient conditions ensuring the existence and uniqueness of a BNE, which is presented and proved in Proposition A.1 below.

Proposition A.1 *Under Assumptions (3.1) and (3.2), if, in addition,* $\max\{|\lambda_{mm}^P|\|W_{mm}^P\|_\infty + |\lambda_{ml}^P|\|W_{ml}^P\|_\infty, |\lambda_{lm}^P|\|W_{lm}^P\|_\infty + |\lambda_{ll}^P|\|W_{ll}^P\|_\infty\} < 1$*, there is a unique vector of expectations satisfying Eq. (A.6) and Eq. (A.7) for a class in public schools; and if* $|\lambda_{mm}^M|\|W_{mm}^M\|_\infty < 1$*, there is a unique vector of expectations satisfying Eq. (A.8) for private school classes*[16].

Proof. For a public school class g, it is easy to check that the space composed of $n_g \times 1$ real-valued vectors ψ with a norm $\|\psi\| = \max_{1\le i\le n_g} |\psi_i|$ is complete. Define an operator T on

[16]For an $a \times b$ matrix Q, the row-norm is defined as $\|Q\|_\infty = \max_{1\le i\le a} \sum_{j=1}^b |Q_{ij}|$.

this space, such that for a migrant child i,

$$(T(\psi))_i = u_g^{'}\alpha^P + x_i^{'}\beta^P + E[\epsilon_{y,1,i}^m|\Im_g^P] + \lambda_{mm}^P \sum_k W_{mm,ik}^P \psi_k + \lambda_{ml}^P \sum_h W_{ml,ih}^P \psi_{g,h}$$

and for a local student j,

$$(T(\psi))_j = u_g^{'}\alpha^P + x_i^{'}\beta^P + \lambda_{lm}^P \sum_k W_{lm,jk}^P \psi_k + \lambda_{ml}^P \sum_h W_{ll,jh}^P \psi_h,$$

The BNE expectation vector ψ_g^P is a fixed point of T, i.e., $(T(\psi_g^P))_i = \psi_{g,i}^P$. Without loss of generality, suppose that the first $n_{g,m}$ students are migrant and the rest $n_{g,l}$ students are local. For any two $n_g \times 1$ vectors ψ and $\psi^{'}$,

$$\begin{aligned}
&\|T(\psi) - T(\psi^{'})\| \\
=&\max\{\max_{1\leq i\leq n_{g,m}} |\lambda_{mm}^P \sum_k W_{mm,ik}^P(\psi_k - \psi_k^{'}) + \lambda_{ml}^P \sum_h W_{ml,ih}^P(\psi_h - \psi_h^{'})|, \\
&\max_{n_{g,m}+1\leq j\leq n_g} |\lambda_{lm}^P \sum_k W_{lm,jk}^P(\psi_k - \psi_k^{'}) + \lambda_{ll}^P \sum_h W_{ll,jh}^P(\psi_h - \psi_h^{'})|\} \\
\leq&\max\{\max_{1\leq i\leq n_{g,m}} \Big[|\lambda_{mm}^P| \sum_k W_{mm,ik}^P + |\lambda_{ml}^P| \sum_h W_{ml,ih}^P\Big], \\
&\max_{n_{g,m}+1\leq j\leq n_g} \Big[|\lambda_{lm}^P| \sum_k W_{lm,jk}^P + |\lambda_{ll}^P| \sum_h W_{ll,jh}^P\Big]\}\|\psi - \psi^{'}\| \\
\leq&\max\{|\lambda_{mm}^P|\|W_{mm}^P\|_\infty + |\lambda_{ml}^P|\|W_{ml}^P\|_\infty, |\lambda_{lm}^P|\|W_{lm}^P\|_\infty + |\lambda_{ml}^P|\|W_{ll}^P\|_\infty\}\|\psi - \psi^{'}\|.
\end{aligned}$$

If $\max\{|\lambda_{mm}^P|\|W_{mm}^P\|_\infty + |\lambda_{ml}^P|\|W_{ml}^P\|_\infty, |\lambda_{lm}^P|\|W_{lm}^P\|_\infty + |\lambda_{ll}^P|\|W_{ll}^P\|_\infty\} < 1$, T is a contraction mapping. The result follows from the contraction mapping theorem in a complete space. In a similar way, we can prove the sufficient condition for a private school class with exclusive migrant students. ■

With normalization, $\|W_{mm}^P\|_\infty + \|W_{ml}^P\|_\infty = 1$ and $\|W_{lm}^P\|_\infty + \|W_{ll}^P\|_\infty = 1$. Thus, Assumption (3.6) ensures that there is a unique BNE expected outcome vector for each class. From Proposition A.1, the conditional expectations consistent with BNE can be solved by iterations. That is, we begin with an arbitrary guess of $\psi_g^{P,0}$, and calculate $\psi_g^{P,1} = T(\psi_g^{P,0})$, $\psi_g^{P,2} = T(\psi_g^{P,1}), \cdots$, until the distance between two adjacent $\psi_g^{P,\tau+1}$ and $\psi_g^{P,\tau}$ is smaller than some cutoff value of precision. This result can also apply to more general cases when the outcome is not necessarily a linear function of peer's expected outcomes.

B Identification at Infinity

This section provides sufficient conditions for identification at infinity. First, parameters of the selection equation can be identified up-to-scale under conventional full-rank assumptions.

Assumption B.1 $\liminf_{n_m \to \infty} \max_{1 \le i \le n_m} |det\big(E[Var(z_i)]\big)| > 0$, *where n_m denotes the number of immigrants in the whole sample.*

Proposition B.1 *With Assumption (3.1), (3.2), and (3.5), if in addition, Assumption B.1 holds, $\frac{\gamma}{\sigma_{m,s}}$ can be identified when the number of migrant children, n_m, is sufficiently large.*

Proof. For an migrant child, $P(S_i^m = 1|z_i) = \Phi(z_i' \frac{\gamma}{\sigma_{m,s}})$. Therefore,

$$\Phi^{-1}(P(S_i^m = 1|z_i)) - E[\Phi^{-1}(P(S_i^m = 1|z_i))] = (z_i - E[z_i])' \frac{\gamma}{\sigma_{m,s}}.$$

Therefore, $\frac{\gamma}{\sigma_{m,s}}$ can be identified under Assumption B.1. ■

Second, we identify parameters in the outcome equations. In the previous section, it is shown that under Assumptions (3.1) and (3.2), given public information, private information does not provide additional information, i.e., $E[y_k^m|\Im_i] = E[y_k^m|\Im_g]$ and $E[y_h^m|\Im_i] = E[y_h^m|\Im_g]$ for any $i \neq k$, $i \neq h$. Moreover, with Assumption (3.4), conditional on school type selection, students are assigned to classes exogenously. Thus, $E[\epsilon_{s,i}^m|\Im_g] = E[\epsilon_{s,i}^m|x_i, z_i, S_i^m]$, which is written as $E[\epsilon_i^m|S_i^m]$ for simplicity. In addition, we derive the following moment conditions, which are basis for identification and estimation.

$$\begin{aligned} E[\epsilon_{y,1,i}^m - \delta^P E[\epsilon_{s,i}^m|S_i^m = 1]|\Im_g] = 0; \\ E[\epsilon_{y,j}^l|\Im_g] = 0; \end{aligned} \tag{B.1}$$

$$E[\epsilon_{y,0,i}^m - \delta^M E[\epsilon_{s,i}^m|S_i^m = 0]|\Im_g] = 0, \tag{B.2}$$

where $\delta^P = \rho_{m,1}\frac{\sigma_{m,y,1}}{\sigma_{m,s}}$ and $\delta^M = \rho_{m,0}\frac{\sigma_{m,y,0}}{\sigma_{m,s}}$.

Since the interactions in private schools can be viewed as a special case that the interactions do not vary across different types in public schools, we can focus on identifying parameters for outcomes in public schools, $\theta^P = (\alpha^{P'}, \beta^{P'}, \delta^P, \lambda_{mm}^P, \lambda_{ml}^P, \lambda_{lm}^P, \lambda_{ll}^P)'$. Identification conditions for parameters in the outcome equation for private schools, $\theta^M = (\alpha^{M'}, \beta^{M'}, \delta^M, \lambda_{mm}^M)'$, then follows straightforwardly. In addition, we can view all the public schools in the sample as a large group and classes as independent subgroups[17]. Therefore, we only need to show parameter identification within one large group.

To simplify notations, let Y_m^P be the vector of all migrants' outcomes, y_i^m's; and Y_l^P is the vector formed by all local students' outcomes, y_j^l's. $\widetilde{X}_m^P$ is the matrix formed by $\widetilde{X}_i$'s of all migrants. Similarly, $\widetilde{X}_l^P$ is the matrix of all locals' features. A row of $\widetilde{U}_m^P$ is the characteristics of the school and class that a migrant belongs to. For migrant i studying in g, the corresponding row in $\widetilde{U}_m^P$ is u_g'. $\widetilde{U}_l^P$ is formed for local students in a similar way. $E[Y_m^P|\mathfrak{I}^P]$ is the vector of all migrants' expected outcome, i.e., $E[y_i^m|\mathfrak{I}_g]$; and $E[Y_l^P|\mathfrak{I}^P]$ is the vector of all local students' expected outcome, i.e., $E[y_j^l|\mathfrak{I}_g]$. $E[\epsilon_{m,s}^P|S_m^P]$ is the vector of migrants' selection biases, i.e., $E[\epsilon_{s,i}^m|S_i^m = 1]$. The vectors $\epsilon_{y,m}^P$ and $\epsilon_{y,l}^P$ are composed of $\epsilon_{y,1,i}^m$'s and $\epsilon_{y,j}^l$'s, respectively. Note that

$$\begin{pmatrix} \epsilon_{y,m}^P \\ \epsilon_{y,l}^P \end{pmatrix} - \begin{pmatrix} E[\epsilon_{m,s}^P|S_m^P] \\ 0 \end{pmatrix} \delta^P = \begin{pmatrix} Y_m^P \\ Y_l^P \end{pmatrix} - \begin{pmatrix} \widetilde{U}_m^P \\ \widetilde{U}_l^P \end{pmatrix} \alpha^P - \begin{pmatrix} \widetilde{X}_m^P \\ \widetilde{X}_l^P \end{pmatrix} \beta^P - \begin{pmatrix} E[\epsilon_{m,s}^P|S_m^P] \\ 0 \end{pmatrix} \delta^P$$
$$- (\lambda_{mm}^P W_{mm}^P + \lambda_{ml}^P W_{ml}^P + \lambda_{lm}^P W_{lm}^P + \lambda_{ll}^P W_{ll}^P) \begin{pmatrix} E[Y_m^P|\mathfrak{I}^P] \\ E[Y_l^P|\mathfrak{I}^P] \end{pmatrix}$$

The moment condition Eq. (B.1) is equivalent to the following equation system:

$$\begin{pmatrix} E[Y_m^P|\mathfrak{I}^P] \\ E[Y_l^P|\mathfrak{I}^P] \end{pmatrix} = \begin{pmatrix} \widetilde{U}_m^P \\ \widetilde{U}_l^P \end{pmatrix} \alpha^P + \begin{pmatrix} \widetilde{X}_m^P \\ \widetilde{X}_l^P \end{pmatrix} \beta^P + \begin{pmatrix} E[\epsilon_{m,s}^P|S_m^P] \\ 0 \end{pmatrix} \delta^P$$
$$+ (\lambda_{mm}^P W_{mm}^P + \lambda_{ml}^P W_{ml}^P + \lambda_{lm}^P W_{lm}^P + \lambda_{ll}^P W_{ll}^P) \begin{pmatrix} E[Y_m^P|\mathfrak{I}^P] \\ E[Y_l^P|\mathfrak{I}^P] \end{pmatrix} \tag{B.3}$$

It is proved in the previous section that under Assumption B.1, the equation system Eq. (B.3) has a unique solution.

Define a matrix

$$Q^P(\mathfrak{I}^P) = \begin{pmatrix} \widetilde{U}_m^P & \widetilde{X}_m^P & E[\epsilon_{m,s}^P|S_m^P] & Q_{mm}^P(\mathfrak{I}^P) & Q_{ml}^P(\mathfrak{I}^P) & 0 & 0 \\ \widetilde{U}_l^P & \widetilde{X}_l^P & 0 & 0 & 0 & Q_{lm}^P(\mathfrak{I}^P) & Q_{ll}^P(\mathfrak{I}^P) \end{pmatrix},$$

where for migrants, a row of $Q_{mm}^P(\mathfrak{I}^P)$ corresponds to $Q_{mm,i}^P(\mathfrak{I}^P) = \sum_k W_{mm,ik}^P E[y_k|\mathfrak{I}_g]$ and the counterpart row for $Q_{ml}^P(\mathfrak{I}^P)$ is $Q_{ml,i}^P(\mathfrak{I}^P) = \sum_h W_{ml,ih}^P E[y_h|\mathfrak{I}_g])$. Similarly, for local students, a row of $Q_{lm,j}^P(\mathfrak{I}^P)$ refers to $Q_{lm,j}^P(\mathfrak{I}^P) = \sum_k W_{lm,jk}^P E[y_k|\mathfrak{I}_g]$ and the corresponding row of

[17] In this case, if we sort sampled students according to their affiliated schools and classes, the social matrices of the big group are block-diagonal, with each diagonal block representing the social relations within one class.

$Q^P_{ll}(\Im^P)$ is $Q^P_{ll,j}(\Im^P) = \sum_h W^P_{ll,jh} E[y_l|\Im_g])$.

Assumption B.2 $\liminf_{n_P\to\infty} \max_{i,j} |det\big(E[Var(Q^P(\Im^P))]\big)| > 0$, *where n_P denote the number of agents in this (large) group.*

Proposition B.2 *Under Assumptions (3.1) to (3.6), Assumptions B.1 and B.2,*

$\theta^P = (\alpha^{P'}, \beta_{P'}, \delta^P, \lambda^P_{mm}, \lambda^P_{ml}, \lambda^P_{lm}, \lambda^P_{ll})'$ is identified when n_P is sufficiently large.

Proof.

$$\begin{pmatrix} E[Y^P_m|\Im^P] - E[y^P_m] \\ E[Y^P_l|\Im^P] - E[y^P_l] \end{pmatrix} = (Q^P(\Im^P) - E[Q^P(\Im^P)])(\alpha^{P'}, \beta^{P'}, \delta^{P'}, \lambda^P_{mm}, \lambda^P_{ml}, \lambda^P_{lm}, \lambda^P_{ll})'$$

■

Within one class, u_g does not vary. Therefore, variations in school and class features, i.e., differences in u_g's across subgroups, is needed for the full rank condition B.2 to be satisfied. Additionally, as $E[\epsilon^m_{s,,i}|S^m_i = 1] = \sigma_{m,s}\frac{\phi(\Phi^{-1}(P(S^m_i=1)))}{P(S^m_i=1)} = \sigma_{m,s}\frac{\phi(z'\gamma/\sigma_{m,s})}{\Phi(z'\gamma/\sigma_{m,s})}$, a sufficient condition that ensures Assumption B.2 is that some variables affecting selection, Z's, are not included in the outcome equations, X's. From Proposition B.1, we normalize $\sigma_{m,s} = 1$. As $\delta^P = \rho_{m,1}\frac{\sigma_{m,y,1}}{\sigma_{m,s}}$, following from Proposition B.2, we can identify the product $\rho_{m,1}\sigma_{m,y,1}$. With $E[y^2_i|\Im_g]$, we can further identify $\rho_{m,1}$ and $\sigma_{m,y,1}$.

In analogue, the moment condition Eq. (B.2) can be equivalently represented by matrices:

$$E[Y^M_m|\Im^M] = \widetilde{U}^M_m \alpha^M + \widetilde{X}^P_m \beta^M + E[\epsilon^P_{m,s}|S^M_m]\delta^M + \lambda^M_{mm} W^M_{mm} E[Y^M_m|\Im^M] \tag{B.4}$$

For a migrant student i in a private schools, we define

$$Q^M(\Im^M) = \begin{pmatrix} \widetilde{U}^M & \widetilde{X}^M & E[\epsilon^M_{m,s}|S^M_m] & Q^P_{mm}(\Im^M) \end{pmatrix}.$$

Then with the following Assumption B.3, we can prove parameter identification in Proposition B.3.

Assumption B.3 $\liminf_{n_M\to\infty} \max_i |det\big(E[Var(Q^M(\Im^M))]\big)| > 0$, *where n_M denote the population of private schools.*

Proposition B.3 *Under Assumptions (3.1) to (3.6), and in addition to Assumptions B.1 and B.3, $\theta^M = (\alpha^{M'}, \beta^{M'}, \delta^M, \lambda^M_{mm})'$ can be identified when n_M is sufficiently large.*

As $y^{m*}_{1,i}$ and $y^{m*}_{0,i}$ cannot be observed at the same time, $corr(\epsilon^m_{y,I,1}, \epsilon^m_{y,I,2}) = \rho_{m,y}$ cannot be identified.

C Estimation Methods

C.1 Estimation Algorithms

Due to selection and peer interactions, the sample log likelihood function is very complicated. Therefore, we propose a two-step estimation algorithm. Normalize $\sigma_{m,s} = 1$, $E[\epsilon^m_{s,i}|S^m_i = 1] = \frac{\phi(z_i^{'}\gamma)}{\Phi(z_i^{'}\gamma)}$ and $E[\epsilon^m_{s,i}|S^m_i = 0] = -\frac{\phi(z_i^{'}\gamma)}{1-\Phi(z_i^{'}\gamma)}$. For migrant student i in class g of a public school, define

$$f^P_{m,i}(\Im_g, \gamma, \theta^P) = u_g^{'}\alpha^P + x_i^{'}\beta^P + \lambda^P_{mm}\sum_k W^P_{mm,ik}E[y_k|\Im_g] + \lambda^P_{ml}\sum_h W^P_{ml,ih}E[y_h|\Im_g] + \delta^P\frac{\phi(z_i^{'}\gamma)}{\Phi(z_i^{'}\gamma)};$$

Similarly, for a local student j in this class, define

$$f^P_{l,j}(\Im_g, \gamma, \theta^P) = u_g^{'}\alpha^P + x_j^{'}\beta^P + \lambda^P_{mm}\sum_k W^P_{mm,jk}E[y_k|\Im_g] + \lambda^P_{ml}\sum_h W^P_{ml,jh}E[y_h|\Im_g],$$

where $E[y_k|\Im_g]$'s are determined by equation system Eq. (B.3). Analogously, for a migrant child in class $g^{'}$ of a private school, we define

$$f^M_{m,i}(\Im_{g^{'}}, \gamma, \theta^M) = u_{g^{'}}^{'}\alpha^M + x_i^{'}\beta^M + \lambda^M_{mm}\sum_k W^M_{mm,ik}E[y_k|\Im_{g^{'}}] - \delta^M\frac{\phi(z_i^{'}\gamma)}{1-\Phi(z_i^{'}\gamma)},$$

where $E[y_k|\Im_{g^{'}}]$'s are the unique solution to Eq. (B.4). The moment conditions, Eq.(B.1) and (B.2), can be written equivalently as

$$E[y^m_i - f^P_{m,i}(\Im_g, \gamma, \theta^P)|\Im_g] = 0;$$

$$E[y^l_j - f^P_{l,j}(\Im_g, \gamma, \theta^P)|\Im_g] = 0;$$

$$E[y^m_i - f^M_{m,i}(\Im_{g^{'}}, \gamma, \theta^M)|\Im_{g^{'}}] = 0.$$

Taking unconditional expectations on both sides and applying the law of iterated expectations, the unconditional expectations of the above moment conditions are equal to 0. That is,

$$\begin{aligned} E[y^m_i - f^P_{m,i}(\Im_g, \gamma, \theta^P)] = 0; \\ E[y^l_j - f^P_{l,j}(\Im_g, \gamma, \theta^P)] = 0; \end{aligned} \tag{C.1}$$

$$E[y^m_i - f^M_{m,i}(\Im_{g^{'}}, \gamma, \theta^M)] = 0. \tag{C.2}$$

These orthogonal conditions imply the following quadratic-form criterion functions:

$$G^P(\theta^P, \gamma; Y^P, \widetilde{u}^P, \widetilde{X}^P, \widetilde{Z}_m, W^P) = \frac{1}{n_P}\Big[\sum_g\sum_{i\in m}(y^m_i - f^P_{m,i}(\Im_g, \gamma, \theta^P))^2 + \sum_g\sum_{j\in l}(y^l_j - f^P_{l,j}(\Im_g, \gamma, \theta^P))^2\Big], \tag{C.3}$$

$$G^M(\theta^M,\gamma;Y^M,\widetilde{u}^M,\widetilde{X}^M,\widetilde{Z}_m,W^M)=\frac{1}{n_M}\sum_{g'}\sum_{i\in m}(y_i^m-f^M_{m,i}(\Im_{g'},\gamma,\theta^M))^2, \tag{C.4}$$

where Y^P, $\widetilde{u}^P$, $\widetilde{X}^P$, and W^P (Y^M, $\widetilde{u}^M$, $\widetilde{X}^M$, and W^M) refer to the outcomes, group features, individual characteristics and social relations in public (private) schools and $\widetilde{Z}_m$ represent the covariates in the selection equations for migrants.

Explicitly, the estimation process can be summarized as below:

1. Estimate the selection equation, Eq. (3.1), for all the immigrants in the sample by MLE,

$$\widehat{\gamma}=argmax_{\gamma}\sum_{i\in m}\Big(S_i^m\log\Phi(z_i^{'}\gamma)+(1-S_i^m)\log[1-\Phi(z_i^{'}\gamma)]\Big). \tag{C.5}$$

2. Plug $\widehat{\gamma}$ into the quadratic-form criterion conditions for δ^P and δ^M and derive estimates by NLS ,

$$\begin{aligned}\widehat{\theta}^P_{2SNLS}&=argmax_{\theta^P}G^P(\theta^P,\widehat{\gamma};Y^P,\widetilde{u}^P,\widetilde{X}^P,\widetilde{Z}_m,W^P);\\ \widehat{\theta}^M_{2SNLS}&=argmax_{\theta^M}G^M(\theta^M,\widehat{\gamma};Y^M,\widetilde{u}^M,\widetilde{X}^M,\widetilde{Z}_m,W^M).\end{aligned} \tag{C.6}$$

In the second stage, with instruments, we can also construct more moment conditions and apply the GMM approach. To be specific, for migrant and local children in public schools, we have the population moment condition,

$$E[v_{1,i}^{'}(\epsilon^m_{y,1,i}-\delta^P E[\epsilon^m_{s,i}|S_i^m])|\Im_g]=0$$
$$E[v_{1,j}^{'}\epsilon^l_{y,j}|\Im_g]=0,$$

where $v_{1,i}$ is a $\overline{q}_1\times 1$ vector, representing some exogenous (IV) variables. which applies for any child in a public school. The corresponding sample analogue is as below:

$$\begin{aligned}&\overline{H}^P_{1,q}(\theta^P,\gamma)\\ \equiv&\overline{H}^P_{1,q}(\theta^P,\gamma;Y^P,\widetilde{u}^P,\widetilde{X}^P,\widetilde{Z}_m,W^P)\\ =&\frac{1}{n_P}\sum_g\Big[\sum_{i\in m}v_{1,i,p}(y_i^m-u_g^{'}\alpha^P-x_i^{'}\beta^P-\lambda^P_{mm}\sum_k W^P_{mm,ik}E[y_k|\Im_g]-\lambda^P_{ml}\sum_h W^P_{ml,ih}E[y_h|\Im_g]-\delta^P\frac{\phi(z_i^{'}\gamma)}{\Phi(z_i^{'}\gamma)})\\ &+\sum_{j\in l}v_{1,j,p}(y_j^l-u_g^{'}\alpha^P-x_j^{'}\beta^P-\lambda^P_{lm}\sum_k W^P_{lm,jk}E[y_k|\Im_g]-\lambda^P_{ll}\sum_h W^P_{ll,jh}E[y_h|\Im_g])\Big],\end{aligned} \tag{C.7}$$

for $q=1,\cdots,\overline{q}_1$. We can include X and Z in V_1. For private schools, the population moment condition takes a similar form:

$$E[v_{0,i}^{'}(\epsilon^m_{y,0,i}-\delta^M E[\epsilon^m_{s,i}|S_i^m])|\Im_{g'}]=0,$$

whose sample analogue is

$$\begin{aligned}
&\overline{H}^M_{1,q'}(\theta^M,\gamma)\\
\equiv&\overline{H}^M_{1,q'}(\theta^M,\gamma;Y^M,\widetilde{u}^M,\widetilde{X}^M,\widetilde{Z}_m,W^M)\\
=&\frac{1}{n_M}\sum_{g'}\sum_{i\in m}v_{0,i,p}(y_i^m-u_{g'}^{'}\alpha^M-x_i^{'}\beta^M-\lambda^M_{mm}\sum_k W^M_{mm,ik}E[y_k|\Im_{g'}]+\delta^M\frac{\phi(z_i^{'}\gamma)}{1-\Phi(z_i^{'}\gamma)}),
\end{aligned}\tag{C.8}$$

for $q^{'}=1,\cdots,\overline{q}_0$, where $V_{0,i}$ is a $\overline{q}_0\times 1$ vector. Pick a $\overline{q}_1\times\overline{q}_1$ positive definite matrix, A_1, and a $\overline{q}_0\times\overline{q}_0$ positive definite matrix A_0, we can derive quadratic forms used to estimate θ^P and θ^M in the second stage, i.e.,

$$\begin{aligned}
\widehat{\delta}^P_{2SGMM}&=argmax_{\theta^P}\overline{H}_1^{P'}(\theta^P,\widehat{\gamma})A_1\overline{H}_1^P(\theta^P,\widehat{\gamma})\\
\widehat{\delta}^M_{2SGMM}&=argmax_{\theta^M}\overline{H}_0^{M'}(\theta^M,\widehat{\gamma})A_0\overline{H}_0^M(\theta^M,\widehat{\gamma});
\end{aligned}\tag{C.9}$$

For this GMM estimator, the additional assumption for identification is that $E[V_{1,i}Q_i^{P'}]$ and $E[V_{0,i}Q_i^{M'}]$ both have full column ranks.

C.2 Large Sample Properties

Because $\epsilon^m_{s,i}$'s are i.i.d., the MLE in the first stage, $\widehat{\gamma}$, is consistent and asymptotically normal. As $\epsilon^m_{y,1,i}-\delta^P E[\epsilon^m_{s,i}|S_i^m=1]=\epsilon^m_{y,1,i}-E[\epsilon^m_{y,1,i}|S_i^m=1]$'s and ϵ^l_j's (also $\epsilon^m_{y,0,i}-\delta^M E[\epsilon^m_{s,i}|S_i^m=0]=\epsilon^m_{y,0,i}-E[\epsilon^m_{s,i}|S_i^m=0]$'s) are independent across different agents, the 2SNLS and 2SGMM estimators are consistent by applying large sample results for extreme estimators. The first-stage estimates, $\widehat{\gamma}$, however, can affect the asymptotic variance of the estimates. To be specific, in addition to the assumptions we have already made, we further impose the following assumption on the convergence rates when the population of (1) all migrants, n_m, (2) the students in public schools, n_P, (3) the students in private schools, n_M, (4) the migrants in public schools, $n_{P,m}$, and (5) the local children in public schools, $n_{P,l}$, go to infinity.

Assumption C.1 $\lim_{n_m,n_P\to\infty}\frac{n_P}{n_m}=c^2_{Pm}>0$, $\lim_{n_m,n_M\to\infty}\frac{n_M}{n_m}=c^2_{Mm}>0$. *In addition,* $\lim_{n_m,n_P\to\infty}\frac{n_{P,m}}{n_P}=r_{P,m}>0$ *and* $\lim_{n_m,n_P\to\infty}\frac{n_{P,l}}{n_P}=r_{P,l}>0$

Proposition C.1 *Under the assumptions in Proposition B.2, the two-stage NLS estimator*

$(\widehat{\alpha}', \widehat{\delta}^{P'}, \widehat{\delta}^{M'})'$ *is consistent. Moreover, with Assumption C.1,*

$$\begin{pmatrix} \sqrt{n_m}(\widehat{\gamma}-\gamma_0) \\ \sqrt{n_P}(\widehat{\delta}^P-\delta_0^P) \\ \sqrt{n_M}(\widehat{\delta}^M-\delta_0^M) \end{pmatrix} \xrightarrow{d} N\Big(0, \begin{pmatrix} \Sigma_{11} & 0 & \\ \Sigma_{21} & \Sigma_{22} & 0 \\ \Sigma_{31} & 0 & \Sigma_{33} \end{pmatrix}^{-1} \begin{pmatrix} \Omega_{11} & 0 & 0 \\ 0 & \Omega_{22} & 0 \\ 0 & 0 & \Omega_{33} \end{pmatrix} \begin{pmatrix} \Sigma_{11} & 0 & \\ \Sigma_{21} & \Sigma_{22} & 0 \\ \Sigma_{31} & 0 & \Sigma_{33} \end{pmatrix}^{-1'} \Big), \tag{C.10}$$

where $\Sigma_{11} = -\operatorname{plim}_{n_m\to\infty} \frac{1}{n_m}\sum_i \frac{\partial^2 l_i(\gamma_0;S_i^m)}{\partial\gamma\partial\gamma'}$,

$\Sigma_{21} = -2c_{Pm}\operatorname{plim}_{n_P\to\infty}\frac{1}{n_P}\sum_g\Big[\sum_{i\in m}\frac{\partial f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}\frac{\partial f^P_{m,i}(\Im_g,\gamma_0,\gamma_0^P)}{\partial\gamma'}+\sum_{j\in l}\frac{\partial f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}\frac{\partial f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\gamma'}\Big]$,

$\Sigma_{22} = -2c_{Pm}\operatorname{plim}_{n_P\to\infty}\frac{1}{n_P}\sum_g\Big[\sum_{i\in m}\frac{\partial f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}\frac{\partial f^P_{m,i}(\Im_g,\gamma_0,\gamma_0^P)}{\partial\theta^{P'}}+\sum_{j\in l}\frac{\partial f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}\frac{\partial f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^{P'}}\Big]$,

$\Sigma_{31} = -2c_{Mm}\operatorname{plim}_{n_M\to\infty}\frac{1}{n_M}\sum_{g'}\sum_i\frac{\partial f^M_{m,i}(\Im_{g'},\gamma_0,\theta_0^M)}{\partial\theta^M}\frac{\partial f^M_{m,i}(\Im_{g'},\gamma_0,\gamma_0^M)}{\partial\gamma'}$,

$\Sigma_{33} = -2c_{Mm}\operatorname{plim}_{n_M\to\infty}\frac{1}{n_M}\sum_{g'}\sum_i\frac{\partial f^M_{m,i}(\Im_{g'},\gamma_0,\theta_0^M)}{\partial\theta^M}\frac{\partial f^M_{m,i}(\Im_{g'},\gamma_0,\gamma_0^M)}{\partial\theta^{M'}}$,

$\Omega_{11} = \lim_{n_m\to}\frac{1}{n_m}\sum_{1\le i\le n_m} Var(\frac{\partial l_i(\gamma_0;S_i^m)}{\partial\gamma})$;

$\Omega_{22} = 4\Big[r_{P,m}\operatorname{plim}_{n_P\to\infty}\frac{1}{n_{P,m}}\sum_g\sum_{i\in m}\frac{\partial f^P_{m,i}}{\partial\theta^P}\frac{\partial f^P_{m,i}}{\partial\theta^{P'}}\frac{1}{n_{P,m}}\sum_g\sum_{i\in m}Var(\epsilon_{y,1,i}-E[\epsilon^m_{y,1,i}|s_i^m=1])+$
$r_{P,l}\operatorname{plim}_{n_P\to\infty}\frac{1}{n_{P,l}}\sum_g\sum_{j\in l}\frac{\partial f^P_{l,j}}{\partial\theta^P}\frac{\partial f^P_{l,ij}}{\partial\theta^{P'}}\frac{1}{n_{P,l}}\sum_g\sum_{j\in l}Var(\epsilon^l_{y,j})\Big]$;

$\Omega_{33} = 4\operatorname{plim}_{n_M\to\infty}\frac{1}{n_M}\sum_i\frac{\partial f^M_{m,i}}{\partial\theta^M}\frac{\partial f^M_{m,i}}{\partial\theta^{M'}}\frac{1}{n_M}\sum_i Var(\epsilon^m_{y,0,i}-E[\epsilon^m_{y,0,i}|s_i^m=0])$,

and $l_i(\gamma_0;S_i^m)$ *denote individual log likelihood for school type selection.*

Proof.

$$\frac{\partial^2 G^P(\theta_0^P,\gamma_0;Y^P,\widetilde{u}^P,\widetilde{X}^P,\widetilde{Z}_m,W^P)}{\partial\theta^P\partial\theta^{P'}}(\widehat{\theta}^P_{2SNLS}-\theta_0^P)+\frac{\partial^2 G^P(\theta_0^P,\gamma_0;Y^P,\widetilde{u}^P,\widetilde{X}^P,\widetilde{Z}_m,W^P)}{\partial\theta^P\partial\gamma}(\widehat{\gamma}-\gamma_0)$$
$$=-\frac{\partial G^P(\theta_0^P,\gamma_0;Y^P,\widetilde{u}^P,\widetilde{X}^P,\widetilde{Z}_m,W^P)}{\partial\theta^P}+o_p(1)$$

where

$$\frac{\partial G^P(\theta_0^P,\gamma_0;Y^P,\widetilde{u}^P,\widetilde{X}^P,\widetilde{Z}_m,W^P)}{\partial\theta^P}$$
$$=-\frac{2}{n^P}\sum_g\Big[\sum_{i\in m}(y_i^m-f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P))\frac{\partial f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}+\sum_{j\in l}(y_i^j-f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P))\frac{\partial f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}\Big],$$

$$\frac{\partial^2 G^P(\theta_0^P,\gamma_0;Y^P,\widetilde{u}^P,\widetilde{X}^P,\widetilde{Z}_m,W^P)}{\partial\theta^P\partial\theta^{P'}}$$
$$=\frac{2}{n^P}\sum_g\Big[\sum_{i\in m}\frac{\partial f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}\frac{\partial f^P_{m,i}(\Im_g,\gamma_0,\gamma_0^P)}{\partial\theta^{P'}}+\sum_{j\in l}\frac{\partial f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}\frac{\partial f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^{P'}}\Big]$$
$$-\frac{2}{n^P}\sum_g\Big[\sum_{i\in m}(y_i^m-f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P))\frac{\partial^2 f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P\partial\theta^{P'}}+\sum_{j\in l}(y_i^j-f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P))\frac{\partial^2 f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P\partial\theta^{P'}}\Big],$$

$$
\begin{aligned}
&\frac{\partial^2 G^P(\theta_0^P,\gamma_0;Y^P,\widetilde{u}^P,\widetilde{X}^P,\widetilde{Z}_m,W^P)}{\partial\theta^P\partial\gamma'}\\
=&\frac{2}{n^P}\sum_g\Big[\sum_{i\in m}\frac{\partial f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}\frac{\partial f^P_{m,i}(\Im_g,\gamma_0,\gamma_0^P)}{\partial\gamma'}+\sum_{j\in l}\frac{\partial f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}\frac{\partial f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\gamma'}\Big]\\
&-\frac{2}{n^P}\sum_g\Big[\sum_{i\in m}(y_i^m-f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P))\frac{\partial^2 f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P\partial\gamma'}+\sum_{j\in l}(y_i^j-f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P))\frac{\partial^2 f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P\partial\gamma'}\Big].
\end{aligned}
$$

For i in public school class g, $y_i^m - f^P_{m,i}(\Im_g,\gamma,\theta^P) = \epsilon^m_{y,1,i} - \delta^P E[\epsilon^m_{s,i}|S_i^m = 1] = \epsilon^m_{y,1,i} - E[\epsilon^m_{y,1,i}|S_i^m = 1]$. Under Assumption (3.4), for all $k \neq i$,

$$
\begin{aligned}
&E[W^P_{mm,ik}(\epsilon^m_{y,1,i}-E[\epsilon^m_{y,1,i}|S_i^m=1])]\\
=&E\Big[E[W^P_{mm,ik}(\epsilon^m_{y,1,i}-E[\epsilon^m_{y,1,i}|S_i^m=1])|S_i^m=1,S^m_{-i},u_g,\widetilde{X}^P_g]\Big]\\
=&E\Big[E[W^P_{mm,ik}|S_i^m=1,S^m_{-i},u_g,\widetilde{X}^P_g]E[\epsilon^m_{y,1}-E[\epsilon^m_{y,1,i}|S_i^m=1]|S_i^m=1,S^m_{-i},u_g,\widetilde{X}^P_g]\Big]\\
=&0.
\end{aligned}
$$

Similarly, $E[W^P_{ml,ih}(\epsilon^m_{y,1,i} - E[\epsilon^m_{y,1,i}|S_i^m = 1])] = 0$ for all $h \neq i$. According to Eq. (B.3), the conditional expectations such as $E[y_k|\Im_g]$ are functions $\widetilde{u}^P$, $\widetilde{X}^P$, $\widetilde{Z}^m$ and W^P. Thus, they are also orthogonal to $\epsilon^m_{y,1,i} - E[\epsilon^m_{y,1,i}|S_i^m = 1]$'s and $\epsilon^l_{y,j}$'s. Therefore, $E[(y_i^m - f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P))\frac{\partial f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}] = E[(y_i^j - f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P))\frac{\partial f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}] = 0$. It then follows from the independence of shocks that $(y_i^m - f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P))\frac{\partial f^P_{m,i}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}$'s and $(y_j^l - f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P))\frac{\partial f^P_{l,j}(\Im_g,\gamma_0,\theta_0^P)}{\partial\theta^P}$'s are martingale difference sequences. Thus, we can apply corresponding CLT to get asymptotic normality. To be specific, define a diagonal matrix,

$$
\Xi = \begin{pmatrix} \sqrt{n_m}I_{p_0} & 0 & 0\\ 0 & \sqrt{n_P}I_{P_1} & 0\\ 0 & 0 & \sqrt{n_M}I_{P_2}\end{pmatrix},
$$

where I_{K_s}, I_{K_1}, and I_{K_0} are identity matrix of corresponding dimensions. By Taylor expansion, we have that

$$
\begin{aligned}
&\Big\{\Xi\begin{pmatrix} -\frac{\partial^2\overline{L}(\gamma_0;S^m)}{\partial\gamma\partial\gamma'} & 0 & 0\\ -\frac{\partial^2 G^P(\theta_0^P,\gamma_0)}{\partial\theta^P\partial\gamma} & -\frac{\partial^2 G^P(\theta_0^P,\gamma_0)}{\partial\theta^P\partial\theta^{P\prime}} & 0\\ -\frac{\partial^2 G^M(\theta_0^M,\gamma_0)}{\partial\theta^M\partial\gamma} & 0 & -\frac{\partial^2 G^M(\theta_0^M,\gamma_0)}{\partial\theta^M\partial\theta^{M\prime}}\end{pmatrix}\Big|_{\gamma_0,\theta_0^P,\theta_0^M}\Xi^{-1}\Big\}\Xi\begin{pmatrix}\widehat{\gamma}-\gamma_0\\ \widehat{\theta}^P-\theta_0^P\\ \widehat{\theta}^M-\theta_0^M\end{pmatrix}\\
=&\Xi\begin{pmatrix}\frac{\partial\overline{L}(\gamma_0;S^m)}{\partial\gamma}\\ \frac{\partial G^P(\theta_0^P,\gamma_0)}{\partial\theta^P}\\ \frac{\partial G^M(\theta_0^M,\gamma_0)}{\partial\theta^P}\end{pmatrix}\Big|_{\alpha_0,\delta_0^P,\delta_0^M}+\begin{pmatrix}o_p(\frac{1}{\sqrt{n_m}})\\ o_p(\frac{1}{\sqrt{n_P}})\\ o_p(\frac{1}{\sqrt{n_M}})\end{pmatrix},
\end{aligned}
$$

Applying LLN and CLT, we derive the asymptotic variance-covariance matrix. ■

D Monte Carlo Simulations

To see how the two-stage NLS method performs for small sample, we do Monte Carlo simulations. In a sample, there are two types of agents, m-type and l-type, corresponding to migrants and local children. Their populations are denoted by N_m and N_l. All "l" agents go to P-type schools. An "m" agent however, may either goes to a P-type school, if $S_i^m = 1$; or an M-type one, if $S_i^m = 0$, where

$$S_i^m = I(\gamma_0 + z_i\gamma_1 + \epsilon_{s,i}^m > 0).$$

For simplicity, suppose that there are two schools in each type. After the types of schools are determined, students are then randomly assigned to schools and classes in each type[18]. For class g of a public school, the performances of migrant children $y_i^m = y_{1,i}^{m*}$'s and local students y_j^l's are determined as follows,

$$y_{1,i}^{m*} = \alpha^P + x_i\beta^P + \lambda_{mm}^P \sum_{k\in m} \frac{1}{n_g - 1} E[y_k|\Im_i] + \lambda_{ml}^P \sum_{h\in l} \frac{1}{n_g - 1} E[y_h|\Im_i] + \epsilon_{y,1,i}^m;$$

$$y_j^l = \alpha^P + x_j\beta^P + \lambda_{lm}^P \sum_{k\in m} \frac{1}{n_g - 1} E[y_k|\Im_j] + \lambda_{ml}^P \sum_{h\in l} \frac{1}{n_g - 1} E[y_h|\Im_j] + \epsilon_{y,j}^l.$$

Similarly, in a class g' of a private school, we have that

$$y_{0,i}^{m*} = \alpha^M + x_i\beta^M + \lambda_{mm}^M \sum_{k\in m} \frac{1}{n_{g'} - 1} E[y_k|\Im_i] + \epsilon_{y,0,i}^m.$$

For simplicity, we take the true values as $\gamma_0 = 0$, $\gamma_1 = 1$, $\alpha^P = 0$, $\beta^P = 1$, $\alpha^M = 0$ and $\beta^M = 0$. We generate z_i's from i.i.d. standard normal distribution and x_i's as i.i.d. $N(1,4)$ random variables. Additionally, we set ϵ_j^l's are i.i.d. $N(0,1)$. $\epsilon_i^m = (\epsilon_{s,i}^m, \epsilon_{y,1,i}^m, \epsilon_{y,0,i}^m)'$'s are i.i.d. normal with $Var(\epsilon_{s,i}^m) = Var(\epsilon_{y,1,i}^m) = Var(\epsilon_{y,0,i}^m) = 1$, $cov(\epsilon_{s,i}^m, \epsilon_{y,1,i}^m) = cov(\epsilon_{s,i}^m, \epsilon_{y,0,i}^m) = 0.4$. To check the estimation with both positive and negative peer effects, we do simulations for

[18]With normalization, the model can correct the influence of class sizes on peer effect intensities. Therefore, we focus on the case where class sizes are similar in each type, which is the case of the real data. The number of population in each type vary with "m" agents' selection. Denote the number of "m" agents in "P" schools and "M" schools by $N_{P,m}$ and $N_{M,m}$. We set lower bounds of class sizes $\underline{c}_P$ and $\underline{c}_M$. The number of classes in "P" ("M") type schools $n_{c,P}$ ($n_{c,M}$) is then the maximal integer which satisfies $0 \leq (N_{P,m} + N_l)/n_{c,P} < \underline{c}_P$ ($0 \leq (N_{P,m} + N_l)/n_{c,M} < \underline{c}_M$). In each type, those classes are equally likely to be assigned to two schools and class populations are randomly determined. Students in each type are assigned to those schools and classes in a random way. In our simulations, the class size lower bounds are set as $\underline{c}_P = \underline{c}_M = 30$

three settings, (A) $\lambda^P_{mm} = \lambda^P_{ll} = 0.5$, $\lambda^P_{ml} = \lambda^P_{lm} = 0.3$, and $\lambda^M_{mm} = 0.5$; (B) $\lambda^P_{mm} = \lambda^P_{ll} = 0.5$, $\lambda^P_{ml} = 0.3$, $\lambda^P_{lm} = -0.1$, and $\lambda^M_{mm} = 0.5$; and (C) $\lambda^P_{mm} = \lambda^P_{ll} = 0.9$, $\lambda^P_{ml} = \lambda^P_{lm} = 0.8$, and $\lambda^M_{mm} = 0.9$. The total populations of migrants and locals, N_m and N_l, can take one of two pairs of values, (1) $N_m = 200$, $N_l = 100$; and (2) $N_m = 1000$, $N_l = 500$. In these two cases, the total population of migrants is two times of that of locals. In our setting, under the true parameter values, the ex ante proportion of migrants going to public schools is about 0.5. Then the average proportion of migrants in a public school class is about 0.5, similar to the ratio in our empirical data set. By comparing estimates in (1) and (2) when the data are generated from the same profile of true parameter values, we can see how the performances of the estimators improve as the sample size increases.

In Table E.1, E.2, and E.3, we present the mean, bias, and mean square errors of the parameters over $L = 400$ simulations under those settings with three estimation methods, (1) Peer effects models with selection assumed away; and (2) Peer effects models with selection assumed away but add school fixed effects in estimation, and (3) the 2SNLS estimation approach for peer effects with selection. As school dummies are added in (2) but not (1) or (3), we do not compare the estimates for the intercepts. Instead, we focus on the estimates for the peer effects. For λ^P_{mm}, λ^P_{ml}, λ^P_{lm}, λ^P_{ll}, and λ^M_{mm}, in most cases, compared with (1) and (2), method (3) can bring in either smaller biases or mean square errors, or both. The performances of (3) improve as the sample size increases. By adding school fixed effects, method (2) may reduce the estimation bias and improve on method (1).

E References

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

Davezies, L. and d'Haultfoeuille, X. and Fougère, D.,(2009), "Identification of peer effects using group size variation", *The Econometrics Journal*, 12(3), 397-413.

Epple, Dennis and Romano, Richard E., (1998), "Competition between private and public schools, vouchers, and peer-group effects", *American Economic Review*, 88 (1), 33-62.

Nakajima, R., (2007), ""Measuring peer effects on youth smoking behavior",*The Review of Economic Studies*, 74, 897-935.

Table E.1: Monte Carlo Simulations I

Sample Size		$N_m = 200,\ N_l = 100$			$N_m = 1000,\ N_l = 500$		
Variables		(1)	(2)	(3)	(1)	(2)	(3)
β^P	1	0.9971	0.9970	0.9977	0.9985	0.9985	0.9989
		(-0.0029)	(-0.0030)	(-0.0023)	(-0.0015)	(-0.0015)	(-0.0011)
		[0.0012]	[0.0012]	[0.0012]	[0.0003]	[0.0003]	[0.0002]
δ^P	0.4			0.4125			0.3945
				(0.0125)			(-0.0055)
				[0.0610]			[0.0099]
λ^P_{mm}	0.5	0.4812	0.4891	0.4322	0.5407	0.5373	0.4995
		(-0.0188)	(-0.0109)	(-0.0678)	(0.0407)	(0.0373)	(-0.0005)
		[0.1045]	[0.1109]	[0.1193]	[0.0167]	[0.0186]	[0.0106]
λ^P_{ml}	0.3	0.3449	0.3214	0.2842	0.3422	0.3411	0.2741
		(0.0449)	(0.0214)	(-0.0158)	(0.0422)	(0.0411)	(-0.0259)
		[0.1297]	[0.1417]	[0.1250]	[0.0211]	[0.0210]	[0.0120]
λ^P_{lm}	0.3	0.2780	0.2685	0.3194	0.2604	0.2618	0.3048
		(-0.0220)	(-0.0315)	(0.0194)	(-0.0396)	(-0.0382)	(0.0048)
		[0.1421]	[0.1508]	[0.1263]	[0.0225]	[0.0238]	[0.0114]
λ^P_{ll}	0.5	0.2747	0.2728	0.3694	0.3540	0.3498	0.4373
		(-0.2253)	(-0.2272)	(-0.1306)	(-0.1460)	(-0.1502)	(-0.0627)
		[0.2631]	[0.2748]	[0.1912]	[0.0652]	[0.0707]	[0.0245]
β^M	1	0.9958	0.9942	0.9952	1.0004	1.0004	1.0007
		(-0.0042)	(-0.0058)	(-0.0048)	(0.0004)	(0.0004)	(0.0007)
		[0.0029]	[0.0030]	[0.0028]	[0.0005]	[0.0005]	[0.0005]
δ^M	0.4			0.3804			0.4012.
				(-0.0196)			(0.0012)
				[0.0643]			[0.0114]
λ^M_{mm}	0.5	0.4450	0.3886	0.4487	0.4974	0.4959	0.4981
		(-0.0550)	(-0.1114)	(-0.0513)	(-0.0026)	(-0.0041)	(-0.0019)
		[0.0786]	[0.2069]	[0.0732]	[0.0015]	[0.0017]	[0.0014]
N^P_m		100.4575	100.4575	100.4575	500.9050	500.9050	500.9050
$\overline{r}^P_{m,c}$		0.5006	0.5006	0.5006	0.5003	0.5003	0.5003

Note: λ^P_{mm}, λ^P_{ml}, λ^P_{lm}, λ^P_{ll} are the intensities of peer effects in public schools and λ^M_{mm} represents the interaction intensities between "m" agents in private schools. N_m, N_l, and N^P_m respectively denote the population of all "m" type, all "l" type, and "m" type in public schools. $\overline{r}^P_{m,c}$ represents the average ratio of "m" types in a class of public schools. For each coefficient, we report the average estimates over 400 simulations. The figures in the parentheses and brackets are respectively biases and mean square errors.

Table E.2: Monte Carlo Simulations II

Sample Size		$N_m = 200,\ N_l = 100$			$N_m = 1000,\ N_l = 500$		
Variables		(1)	(2)	(3)	(1)	(2)	(3)
β^P	1	0.9974	0.9972	0.9976	0.9992	0.9991	0.9990
		(-0.0026)	(-0.0028)	(-0.0024)	(-0.0008)	(-0.0009)	(-0.0010)
		[0.0012]	[0.0012]	[0.0012]	[0.0002]	[0.0002]	[0.0002]
δ^P	0.4			0.4162			0.3961
				(0.0162)			(-0.0039)
				[0.0595]			[0.0095]
λ^P_{mm}	0.5	0.5087	0.5051	0.4368	0.5701	0.5686	0.5010
		(0.0087)	(0.0051)	(-0.0632)	(0.0701)	(0.0686)	(0.0010)
		[0.0899]	[0.1016]	[0.1075]	[0.0133]	[0.0135]	[0.0096]
λ^P_{ml}	0.3	0.3314	0.3143	0.2731	0.3267	0.3258	0.2727
		(0.0314)	(0.0143)	(-0.0269)	(0.0267)	(0.0258)	(-0.0273)
		[0.1215]	[0.1457]	[0.1246]	[0.0123]	[0.0125]	[0.0133]
λ^P_{lm}	-0.1	-0.1787	-0.1832	-0.0955	-0.1775	-0.1787	-0.0990
		(-0.0787)	(-0.0832)	(0.0045)	(-0.0775)	(-0.0787)	(0.0010)
		[0.1374]	[0.1592]	[0.1338]	[0.0139]	[0.0142]	[0.0076]
λ^P_{ll}	0.5	0.2990	0.2862	0.3671	0.3768	0.3756	0.4424
		(-0.2010)	(-0.2138)	(-0.1329)	(-0.1232)	(-0.1244)	(-0.0576)
		[0.2507]	[0.2716]	[0.2097]	[0.0316]	[0.0322]	[0.0176]
β^M	1	0.9958	0.9942	0.9952	1.0004	1.0004	1.0007
		(-0.0042)	(-0.0058)	(-0.0048)	(0.0004)	(0.0004)	(0.0007)
		[0.0029]	[0.0030]	[0.0028]	[0.0005]	[0.0005]	[0.0005]
δ^M	0.4			0.3804			0.4012
				(-0.0196)			(0.0012)
				[0.0643]			[0.0114]
λ^M_{mm}	0.5	0.4450	0.3886	0.4487	0.4974	0.4959	0.4981
		(-0.0550)	(-0.1114)	(-0.0513)	(-0.0026)	(-0.0041)	(-0.0019)
		[0.0786]	[0.2069]	[0.0732]	[0.0015]	[0.0017]	[0.0014]
N^P_m		100.4575	100.4575	100.4575	500.9050	500.9050	500.9050
$\bar{r}^P_{m,c}$		0.5006	0.5006	0.5006	0.5003	0.5003	0.5003

Note: λ^P_{mm}, λ^P_{ml}, λ^P_{lm}, λ^P_{ll} are the intensities of peer effects in public schools and λ^M_{mm} represents the interaction intensities between "m" agents in private schools. N_m, N_l, and N^P_m respectively denote the population of all "m" type, all "l" type, and "m" type in public schools. $\bar{r}^P_{m,c}$ represents the average ratio of "m" types in a class of public schools. For each coefficient, we report the average estimates over 400 simulations. The figures in the parentheses and brackets are respectively biases and mean square errors.

Table E.3: Monte Carlo Simulations III

Sample Size		$N_m = 200,\ N_l = 100$			$N_m = 1000,\ N_l = 500$		
Variables		(1)	(2)	(3)	(1)	(2)	(3)
β^P	1	1.0056	1.0048	1.0007	1.0021	1.0013	0.9954
		(0.0056)	(0.0048)	(0.0007)	(0.0021)	(0.0013)	(-0.0046)
		[0.0025]	[0.0020]	[0.0015]	[0.0011]	[0.0009]	[0.0003]
δ^P	0.4	0	0	0.3964	0	0	0.3901
				(-0.0036)			(-0.0099)
		[0.1600]	[0.1600]	[0.0927]	[0.1600]	[0.1600]	[0.0121]
λ^P_{mm}	0.9	0.8080	0.7660	0.8160	0.8363	0.8401	0.8717
		(-0.0920)	(-0.1340)	(-0.0840)	(-0.0637)	(-0.0599)	(-0.0283)
		[0.1055]	[0.1879]	[0.1126]	[0.0573]	[0.0493]	[0.0179]
λ^P_{ml}	0.8	0.5754	0.5132	0.5624	0.6526	0.6549	0.6573
		(-0.2246)	(-0.2868)	(-0.2376)	(-0.1474)	(-0.1451)	(-0.1427)
		[0.2638]	[0.3784]	[0.2503]	[0.1016]	[0.0981]	[0.0518]
λ^P_{lm}	0.8	0.7892	0.7479	0.8055	0.8458	0.8451	0.8343
		(-0.0108)	(-0.0521)	(0.0055)	(0.0458)	(0.0451)	(0.0343)
		[0.1124]	[0.1924]	[0.0797]	[0.0242]	[0.0260]	[0.0188]
λ^P_{ll}	0.9	0.4426	0.3899	0.5361	0.4808	0.4898	0.6580
		(-0.4574)	(-0.5101)	(-0.3639)	(-0.4192)	(-0.4102)	(-0.2420)
		[0.5390]	[0.6682]	[0.4109]	[0.4080]	[0.3846]	[0.1040]
β^M	1	0.9972	0.9951	0.9968	1.0005	1.0005	1.0008
		(-0.0028)	(-0.0049)	(-0.0032)	(0.0005)	(0.0005)	(0.0008)
		[0.0029]	[0.0031]	[0.0029]	[0.0005]	[0.0005]	[0.0005]
δ^M	0.4			0.3832			0.4001
				(-0.0168)			(0.0001)
				[0.0457]			[0.0037]
λ^M_{mm}	0.9	0.8943	0.8169	0.8953	0.8998	0.8998	0.8999
		(-0.0057)	(-0.0831)	(-0.0047)	(-0.0002)	(-0.0002)	(-0.0001)
		[0.0093]	[0.1549]	[0.0092]	[< 0.0001]	[< 0.0001]	[< 0.0001]
N^P_m		100.4575	100.4575	100.4575	500.9050	500.9050	500.9050
$\bar{r}^P_{m,c}$		0.5006	0.5006	0.5006	0.5003	0.5003	0.5003

Note: λ^P_{mm}, λ^P_{ml}, λ^P_{lm}, λ^P_{ll} are the intensities of peer effects in public schools and λ^M_{mm} represents the interaction intensities between "m" agents in private schools. N_m, N_l, and N^P_m respectively denote the population of all "m" type, all "l" type, and "m" type in public schools. $\bar{r}^P_{m,c}$ represents the average ratio of "m" types in a class of public schools. For each coefficient, we report the average estimates over 400 simulations. The figures in the parentheses and brackets are respectively biases and mean square errors.