Essays in the Economics of Education and Program Evaluation

by
Hongliang Zhang
B.S., Peking University (2001)
MURP, University of Minnesota (2003)

Submitted to the Department of Economics and the Department of Urban Studies and Planning in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2009

Signature of Author ....
Department of Economics
Department of Urban Studies and Planning
July 15, 2009

Certified by ....
Joshua D. Angrist
Ford Professor of Economics
Thesis Supervisor

Certified by ....
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Thesis Supervisor

Certified by ....
Karen R. Polenske
Professor of Regional Political Economy and Planning
Thesis Supervisor

Accepted by ....
Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics
Chair, Graduate Studies Committee of the Department of Economics

Accepted by ....
Eran Ben-Joseph
Associate Professor of Landscape Architecture and Urban Planning
Chair, PhD Committee of the Department of Urban Studies and Planning

Essays in the Economics of Education and Program Evaluation

by

Hongliang Zhang

Submitted to the Department of Economics and the Department of Urban Studies and Planning
on 15 August 2009, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The dissertation consists of three essays studying the economics of education and program evaluation. Chapter 1 examines the impact of attending a magnet school on student achievement using school admissions lotteries in Wuhan, China. Although lottery winners were more likely to attend magnet schools that appear better in many dimensions, including peer achievement, I find little evidence that winning a lottery improved students' performance on the High School Entrance Exam or their enrollment status at elite high schools three years later. Magnet school popularity, measured by either the competitiveness of the admission lottery or the take-up rate of lottery winners, is highly positively correlated with the average student achievement, but largely unrelated to the treatment effect on test scores that I estimate for each school. This evidence suggests that parents value peer quality beyond its effect on achievement gains, or confuse average student achievement with value added. The finding that magnet schools are sought mainly for their observed superiority in average student achievement rather than for their academic value added casts doubt on the potential of school choice to improve student achievement, at least in this context.

Chapter 2 studies peer effects on student test scores in middle school using a multi-cohort longitudinal data set from China. I base the identification on variation in peer composition across adjacent cohorts within the same school to control for student sorting across schools and the unobserved school characteristics that affect student outcomes. The existing peer effects literature pays little attention to the potential positive correlation in measurement errors between the individual- and the peer-level lagged test score variables, which I find important in my data. Such a positive correlation in measurement errors arises because the individual- and the peer-level lagged test score variables are subject to transitory common shocks due to the continuing presence of a student's former peers in her current peer group. I derive formally that the presence of transitory common shocks on lagged test scores will lead to a negative bias in the estimate of peer effects, and propose an empirical strategy to address this problem by using the lagged test score measures of new peers to instrument for the corresponding lagged test score measures of all peers. The within-school IV estimate of

the linear-in-means model shows little evidence that having peers of higher average lagged test score significantly improves a student's test score. Estimates of heterogeneous peer effects models, however, show some evidence in favor of ability tracking.

Chapter 3 investigates the nonparametric analysis of randomized program evaluation. Observational problems following randomization sometimes prevent researchers from collecting complete and error-free outcome data. The problem is more serious if sample selection varies by treatment status and interacts with data contamination (imperfect matching). In Chapter 3, I develop a trimming procedure for nonparametric analysis of average treatment effects in the presence of sample attrition and imperfect matching, as well as their interactions. I show that, if prior knowledge or a consistent estimator of data contamination rate is available and the treatment status affects sample selection in "one direction," the proposed trimming procedure can construct bounds on average treatment effects for a specific subpopulation whose outcomes would be observed irrespective of treatment assignment status.

Thesis Supervisor: Joshua D. Angrist
Title: Ford Professor of Economics

Thesis Supervisor: Esther Duflo
Title: Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics

Thesis Supervisor: Karen R. Polenske
Title: Professor of Regional Political Economy and Planning

Acknowledgement

I am deeply indebted to my advisors, Josh Angrist, Esther Duflo, and Karen Polenske for their infinite patience, encouragement, guidance, and support. I cannot thank Josh, Esther, and Karen enough for being so generous with the time and attention they devoted to my development throughout my years at MIT. I feel unbelievably fortunate to have the opportunity to work with and learn from Josh, Esther, and Karen.

My heartfelt thanks also go to Bill Wheaton, who offered generous advice, support, and recommendation at various stages of my study at MIT. I am also grateful to Daron Acemoglu, Glenn Ellison, and Frank Levy for their generous recommendation for my application to the Department of Economics and to Nancy Rose for her invaluable advice and support during my job search.

Many professors at MIT have also generously offered their time and advice, especially David Autor, Abhijit Banerjee, Amy Finkelstein, Michael Greenstone, and Benjamin Olken. I have benefited enormously from my fellow graduate students. Thanks are due to my entire cohort, and especially to James Berry, Rongzhu Ke, Cynthia Kinnan, Weifeng Li, and Tao Jin. I also want to thank Robert Irwin for his patient and tireless writing consultations, which greatly improved this thesis. I am also grateful to Tosha Hairston, Peet Hoagland, Gary King, Malinda Nicolosi, Kim Scantlebury, Selene Victor, and Sandra Wellford for their generous administrative assistance.

Special thanks go to my master's advisor at University of Minnesota, Ragui Assaad, who introduced me with quantitative research methods and inspired me to conduct research on education. Part of my field work was conducted in the Institute of Advanced Studies at Wuhan University. I am grateful to Heng-fu Zou for his warm invitation. I also thank Bing

Liang, Tao Mao, Xiao Qin, and Ye Yang from Wuhan University for assistance in data entry. The George and Obie Shultz Fund and the MIT Center for International Studies Summer Study Grant provide financial support for my field trip.

Finally, my greatest thanks are to my family. I am eternally grateful to my parents, Yongquan Zhang and Lili Zhu. They have always believed in me and offered their unwavering support in every step of my life. My wife, Qin Zhao, has accompanied me for the past five years at MIT. Her patience, understanding, and love have been the source of great inspiration at many times during my PhD process. Without her, the completion of this thesis would have been impossible. I could not hope for a more supportive and understanding family. I dedicate this thesis to them.

Chapter 1

Magnet Schools and Student Achievement: Evidence from a Randomized Natural Experiment in China

1.1 Introduction

Recent decades have seen growing efforts in several countries (e.g., Chile, China, Pakistan, and the United States) to improve educational outcomes by increasing the scope of schooling alternatives available to parents and students. The popularity of school choice is based upon the belief that increasing parental choice can yield improved efficiency in education production through enhanced competition or better matches between students and schools (e.g., Friedman, 1962; Chub and Moe, 1990; Hoxby, 2000).

The ability of the choice mechanism to improve educational outcomes turns in part on the extent to which parents express their preference for achievement gains. Some previous studies have found that parents do not necessarily seem to place the highest weight on academic outcomes (Hastings et al., 2005, 2006; Jacob and Levitt, 2006; Jacob and Lefgren, 2007), and may not know which schools are likely to benefit their child the most academically (Figlio and Lucas, 2004; Mizala and Urquiola, 2007). Moreover, if peer externalities play a significant role in value added, parents may rationally prefer a poorly-run school with good peers to a well-run school with bad peers (Willms and Echols, 1992). As noted by Rothstein (2006), any of these factors will dilute the incentives for efficiency improvement that choice might otherwise create.

A growing body of empirical work has used random assignment to estimate the effect on academic outcomes of attending a school other than the local public school.[1] Some of these studies show benefits (e.g., Green et al., 1999; Hoxby and Rockoff, 2004; Lai, et al., 2008) but others show little or no effect (e.g., Howell and Peterson, 2002; Krueger and Zhu, 2004). Cullen et al. (2006), in a study closely related to this work, use lotteries to evaluate the impact of attending a sought-after school in the Chicago Public Schools. They find no evidence that winning a lottery to attend a sought-after school improves students' performance on a variety of traditional academic measures. In contrast, Angrist et al. (2002,

[1] Theoretically, it could be that school choice benefits all students through increased competition, including those who remain in attendance area schools. So the partial equilibrium effect could be an overestimate or an underestimate of the overall effect. Unfortunately, due to the extreme difficulty in the identification, little evidence exists regarding the potential for schools to respond to the enhanced competitive pressure induced by school choice (e.g., Hoxby, 2003; Figlio and Rouse, 2006; Hsieh and Urquiola, 2006; Carnoy et al., 2007). The research designs of the limited studies on the general equilibrium impact of school choice are far from ideal, and do not necessarily rule out other explanations for the improvements (Rouse and Barrow, 2008).

2006) find consistently positive effects of private school vouchers on grade completion and test scores in Colombia. One possible explanation of this difference in findings is that there might be more potential for school choice programs to generate large efficiency gains in developing countries that have less efficient public schools and less competitive schooling environments.

This chapter presents new evidence on the impact of choice on student achievement and the question of how parents choose schools. The evidence comes from a particular form of school choice in China that allows students access to alternative magnet (*gaizhi*) schools outside their assigned local public schools by paying an additional charge. This chapter answers two questions: First, what is the direct impact of attending a magnet school on student achievement? Second, is there any evidence that parents chose the most academically beneficial school for their child?

Students in China are assigned to primary schools (grades 1-6) and neighborhood middle schools (grades 7-9) based on their residence. Middle school graduates take a citywide uniform High School Entrance Exam (HSEE) and are tracked into different types of secondary schools (grades 10-12) based on their HSEE test scores. Magnet schools exist only at the middle school level and open their enrollment to all interested students within the city willing to pay the tuition. Magnet schools in China share many common features with their U.S. counterparts:[2] they are located mainly in large cities,[3] have better qualified teachers and higher per pupil spending than neighborhood public schools, are highly sought after,

[2]See Steel and Levine (1994) for an in-depth study of magnet schools in the U.S.

[3]One in six middle school students in large Chinese cities are enrolled in magnet schools. While in the U.S., though magnet schools only account for 4 percent of the overall enrollment nationwide, they enroll 15 percent students in districts where magnet programs are available. The majority of these magnet districts are urban districts in large metropolitan areas.

and enroll a disproportionate number of students from families with high socioeconomic status (SES) and high primary school test scores. Lotteries are often used by oversubscribed magnet schools to determine admission eligibility, providing identification for magnet school attendance for students with a variety of academic backgrounds.

A number of features of Chinese magnet schools make the Chinese context especially interesting and informative. First, in contrast to the United States, magnet schools in China have the same curriculum as other neighborhood schools. Achievement is measured by test scores on the HSEE, which are almost the sole determinant of their secondary school admission status. This unidimensional, high-stakes exam provides a good benchmark for evaluating the impact of magnet school attendance. Second, because choice is costly for parents, the perceived difference between magnet schools and neighborhood public schools must be substantial, making it an ideal context to compare parental perception of what constitutes a "good school" to evidence of effective value added.

The data set used in this chapter consists of over 13,000 students who participated in the admissions lotteries of magnet schools during 2002-2004 in three school districts in Wuhan, China. I match applicants in the lottery data to administrative records of the city's HSEE database to obtain student information on middle school enrollment status, HSEE test scores, and secondary school admission status. Although not all students complied with their lottery assignment and about half of the lottery losers managed to get into the magnet school they chose through the "back door," winning a lottery still increased the probability of enrolling in that magnet school by 33 percentage points. Despite the fact that lottery winners had access to a better peer group, better qualified teachers, and a school they chose (and that hence might better suit their learning needs), I find little evidence that winning a lottery is associated with any academic benefit to students. The point estimates of the two-stage least squares (2SLS) regressions of students' HSEE test scores on their magnet school enrollment status using lottery outcomes as an instrument are all insignificant, and allow me to reject a modest gain of 0.1 standard deviation (SD) at the five percent level.

If magnet schools do not improve academic outcomes, why are parents willing to pay for them? One possibility is that parents value things other than academic success when choosing a school. In a context of heterogeneous parental preferences, parents who value other school attributes may be willing to trade academic gains for utility gains in other dimensions. Although magnet schools in Wuhan are better in nonacademic attributes – such as physical environment, classroom discipline, and peer behavior – than neighborhood public schools, this is unlikely to be the main explanation here, given the importance of students' HSEE test scores in high school admission. Another possibility is that parents place high value on average student achievement when choosing a school, either because they value peer quality beyond its effect on value added, or because they confuse average achievement with value added. Even though the empirical evidence cannot distinguish these two underlying reasons, either of them being true would lead schools to be sought for their observed student achievement. Indeed, I find that magnet school popularity, measured by either the competitiveness of the admission lottery or the take-up rate of lottery winners, is highly positively correlated with the average student achievement, but largely unrelated to the treatment effect on student achievement that I estimate for each school. The finding that magnet schools are sought mainly for their observed superiority in student achievement rather than for their academic value added casts doubt on the potential of school choice to improve student achievement, at least in this context.

The rest of this chapter is organized as follows. Section 1.2 provides background on the middle school system and the admission procedure of magnet schools in Wuhan. Section 1.3 introduces the data and presents some descriptive statistics. Section 1.4 presents the impact of winning a lottery on magnet school enrollment and peer achievement. Section 1.5 discusses the reduced-form effect of winning a lottery and the causal effect of attending a magnet school – identified by using lottery status as an instrument – on students' academic outcomes. Section 1.6 examines the magnitude and the likely sign of the potential biases in the main results due to differential sample attrition between lottery winners and losers. Section 1.7 discusses possible explanations for parental school choice that are consistent with the empirical evidence here. Section 8 provides concluding remarks.

1.2 Background

With a population of over 4.5 million, Wuhan is the fourth most populous city in China. Situated at the confluence of the Yangzi River and its longest tributary, the Hanshui River, the city comprises three parts: North Bank (*Hankou*), West Bank (*Hanyang*), and South Bank (*Wuchang*) (see Figure 1-1). The three parts can be considered as independent enrollment areas as students rarely commute across river for schooling. In this chapter, I focus on all three school districts (Districts 1-3) in the North Bank for which data are available. The middle school system in the North Bank includes 74 middle schools and enrolls approximately 60,000 students, accounting for about 45 percent of the city's enrollment in middle school.

Enrollment at primary school is based on residence. Graduates from primary school are assigned to a neighborhood middle school through an assignment mechanism that works at the neighborhood level. But all eight of the magnet schools in the North Bank – two to three from each district – have opted out of the assignment mechanism and opened their enrollment to all interested students from the entire school district or even beyond district boundaries. Before 2002, these magnet schools required applicants to take entrance exams and admitted almost all the top-scoring students in each school district. To ease the exam pressure on children and reduce the across-school inequality in student ability, starting in 2002, the city education council banned the use of any form of entrance exam in middle school admissions and required admissions to all public schools to be based on residence.. In order to retain their open enrollment policy, all these magnet schools transformed into privately sponsored public schools.[4] Public middle schools are tuition-free under China's Nine-Year Compulsory Schooling Law. By switching to a semi-private financing structure, these magnet schools could continue to receive funding from the local government but also charge tuition to students.[5] All magnet schools set their tuition at US $400 per year,[6] the price ceiling allowed by the city education council. (For reference, the average annual disposable income of a three-person family was roughly $3,000 in the city during the period under study.)

[4]The semi-privatization had started even before 2002 for a few magnet schools, some of which had by then been transformed into privately sponsored public schools in order to charge discriminatory tuitions to students who scored below the admission cutoff score on the entrance exams. The tuition varied between $400 and $800 per year, depending on how far a student's test score was away from the admission cutoff. The 2002 reform, however, caused the semi-privatization movement to sweep through all magnet schools.

[5]Revenue from tuition was divided among the magnet school, the school district, and the city education council in a 5:4:1 ratio. School districts and the city education council used their share of revenue to fund capital expenditures on neighborhood middle schools with poor facilities. Basic salaries of magnet school teachers were paid from the local government budget, which was why these magnet schools were not purely private schools. But magnet schools provided much higher overall compensation to their teachers via school-funded benefits and performance pay.

[6]All monetary amounts in the paper refer to U.S. dollars, converted by the exchange rate at the time of the study.

Each student could apply to only one magnet school in the city and would be disqualified from enrolling in any magnet school if caught submitting multiple applications. Districts 2 and 3 allowed magnet schools in their district to set aside a fraction (up to 50 percent) of their admission quota for advance admission exclusively for qualified applicants: applicants with award records in city- or district-level academic, artistic, and athletic contests. In order to attract the most talented students, all magnet schools guaranteed admission (through formal or informal channels) and offered full or half tuition waivers to students with extraordinary award records. As magnet schools differed in the selectivity of their criteria for advance admission and tuition waivers as well as the competitiveness of their admissions lotteries, students might be strategic in selecting which magnet school to apply to and did not necessarily choose their most preferred school.

All magnet schools were oversubscribed for the period studied here as demand at the regulated tuition far exceeded the admissions quotas. In a few cases, because the number of qualified applicants of a magnet school exceeded its advance admission quota, an advance lottery was conducted to select among qualified applicants. Qualified applicants who had lost out in the advance lottery, however, could get a second chance to gain admission (to the same magnet school) through the main lottery. Every year, a main lottery was conducted for each magnet school among its general applicants, plus its qualified applicants who did not win admission in the advance lottery (if any), to assign the rest of the admission quota randomly. In each lottery, a computer program randomly assigned a lottery number to each student. The magnet school enrolled students with the lowest numbers first until it filled its admission quota for that lottery. All admissions lotteries were certified by notaries public to verify their randomness to prevent tampering. Lottery winners were required to make an upfront payment of their three-year tuition (net of the waiver provision, if any) by the deadline announced by the school, and those who did not pay their tuition by the deadline were considered to have declined their admission offer. A significant portion of applicants who had lost the main lottery, however, still managed to get into the magnet school they applied to through the "back door."[7] A typical magnet school in the city admitted about one-third of their students by advance admission, one-third through the main lottery, and one-third through the "back door."

For students from District 2 (where pre-lottery test scores are available), Figure 1-2-A shows the Kernel density curves of their 6th grade combined math and Chinese test scores (standardized to have zero mean and unit variance for each cohort) by their magnet school enrollment status. There is clear evidence that magnet schools removed a disproportionate share of high-achieving students from their assigned neighborhood middle schools. Students enrolled in magnet schools had a mean 6th grade test score that was 0.47 SD above the district average. In contrast, students enrolled in neighborhood middle schools had a mean 6th grade test score 0.10 SD below the district average. Figure 1-2-B shows two sources of student sorting: advance admission recipients and main lottery participants had mean 6th grade test scores that were about 0.68 SD and 0.31 SD above the district average, respectively, leaving the nonparticipants a mean test score 0.21 below the district average.

[7]The tuition paid by students who entered through the "back door" varied between $400 and $800 per year, with the latter being set by the city education council as the maximum allowable charge. But being willing to pay the maximum allowable charge was not sufficient for an applicant to gain admittance to a magnet school. The most important factor determining whether a lottery loser could be admitted through the "back door" was whether she had a referee who was important enough to influence the decision of the principal (e.g., a government official her parents found through their personal social network). The student's academic performance at primary school was another factor, but of secondary importance. The final tuition charged for a student admitted through the "back door" was determined by taking into account the importance of her referee, the closeness of her relationship with her referee, and her academic performance at primary school.

Another schooling option for students is to attend a private middle school, which receives no public funding and is financed entirely by student tuition. Most private middle schools are boarding schools located in neighboring counties outside the city's urban boundary. Private middle schools are generally considered as inferior to magnet schools and only account for a very small share (less than two percent) of middle school enrollment. Nonetheless, private middle schools were still likely to be the option some lottery participants might choose, especially lottery losers who could not get into the magnet school they applied to through the "back door."

1.3 Data Description and Summary Statistics

With the cooperation of the notary public office in the North Bank, I have obtained the administrative data of three cohorts of students who participated in the main lotteries of all the eight magnet schools in the North Bank during 2002-2004. The lottery data include each applicant's name, gender, qualification status, lottery outcome, and primary school graduated from. I exclude a small fraction (3.5 percent) of applicants who were enrolled in primary schools outside the North Bank at the time of the application. (Note that excluding these students does not affect the validity of the randomization because their primary school enrollment status was predetermined at the time of the lottery.) The final sample consists of 13,769 applicants who participated in 21 lotteries of all the eight magnet schools in the North Bank during 2002-2004,[8] including 997 qualified applicants in eight of the lotteries

[8]Three lotteries conducted in the year 2003 in District 2 are excluded because the notary public office's records for these lotteries only contain information on the lottery winners, but not the losers.

of four of these magnet schools.[9] Students in the full sample constitute approximately 23 percent of all students who transitioned from primary school to middle school in the North Bank during this three-year period. For applicants from District 2, I have obtained their 6th grade test scores from the school district. But no pre-lottery test scores are available for applicants from the other two districts.

Overall, these admissions lotteries were competitive: on average, only three out of ten participants won their lottery. There is little evidence of any association between win/loss status and students' predetermined individual characteristics at the time of the lottery. This can be seen in Panel A of Table 1-1. Columns 1 and 2 present lottery losers' means for each dependent variable and the coefficients from separate regressions of each dependent variable on an indicator variable for winning a lottery and a full set of lottery fixed effects, respectively. The full sample of all lottery participants who applied from the North Bank is used in column 2, providing a test of the randomness of the lotteries among applicants from the North Bank. Columns 3 and 4 report lottery losers' means for each predetermined individual characteristic and the win/loss difference for a subsample of applicants from District 2, where students' 6th grade test scores are available. For 87 percent of the applicants from District 2, I find their 6th grade test score by matching their name, gender, and primary school to the district's test score records.[10] In both samples, lottery winners and losers were balanced on all the predetermined characteristics. There is no evidence that the lotteries favored qualified applicants or applicants with higher pre-lottery test scores. As a further check of

[9] No qualified applicants participated in the other 13 lotteries.

[10] Nonmatching is largely due to name misspelling or gender misidentification in either the lottery records or the 6th grade test score records. Students' 6th grade test scores, which were coded from the school district's handwritten records, might be subject to some degree of data entry error.

the randomness, I compare the winning rates by primary schools for schools with more than 20 applicants participating in a lottery to the winning rate of this lottery for the full sample. In only 9 (4.3 percent) out of the 207 comparisons does the winning rate of the primary school differ from the lottery winning rate at the five percent level.

Individual information in the lottery data is used to match lottery participants to administrative records of the HSEE database, which includes student information on middle school enrollment, HSEE test scores, and secondary school admission status. All middle school graduates are required to take the HSEE as it also serves as the middle school graduation exam. Unfortunately, the lottery records do not contain perfect identification information that can guarantee unique tracking of students in the HSEE database. Specifically, I can only use the combination of an applicant's name and gender to search for matches in the HSEE records, which sometimes leads to multiple matches due to common names. As dropping out and repeating a grade were almost nonexistent in middle school and students rarely commuted across river to attend middle school, I limit the matching search to the HSEE records of the corresponding cohort that graduated from middle school in the North Bank three years after each lottery.[11]

Panel B of Table 1-1 reports the matching statistics by lottery status within the universe of all middle school graduates from the North Bank. The overall match rate is high, though not perfect: about 90 percent of the lottery participants are matched to HSEE records, including 68 percent uniquely matched and 85 percent matched to no more than five records.

[11]Expanding the source data to middle school graduates from the entire city only reduces the nonmatch rate by about two percentage points, suggesting that very few applicants opt for middle school in the West Bank and South Bank. However, expanding the source data significantly increases the probability of multiple matches. In this paper, I only report the results using matched records in the North Bank. But the main results remain similar if I instead use matched records in the entire city.

I estimate the empirical results of this chapter using the single-matched sample and the combined single- and multi-matched sample consisting of applicants with up to five matches in the HSEE records, respectively. When an applicant has two to five matches in the HSEE database, I assign the applicant the mean value of the multi-matched HSEE records and a weight that is equal to the number of matches. However, I exclude applicants matched to more than five HSEE records out of concern that noise due to matching errors may outweigh information for those individuals.

The missing outcomes can be due to either name misspelling, gender misidentification, families moving out of the city, or students opting for middle schools outside the North Bank (i.e., middle schools in the West Bank and South Bank or private middle schools outside the city). Though sample attrition due to the first three reasons is likely to be exogenous to the randomly determined lottery status, whether a student would opt for schooling outside the North Bank may depend on her lottery outcome. Column 2 shows that winning a lottery increased the likelihood of being matched by two percentage points, suggesting that a small fraction of applicants – who would have opted for schooling outside the North Bank had they lost their lottery – were induced to enroll in the North Bank after winning their lottery. As the degree of differential attrition was very small in practical terms, I first present the main results in Sections 1.4 and 1.5 by comparing the matched winners and losers under the assumption that the differential sample attrition was nonselective, i.e., the characteristics of the small group of applicants whose enrollment status in the North Bank depended on their lottery status did not differ from those who would enroll in the North Bank, irrespective of their lottery status. I then discuss in Section 1.6 the magnitude and the likely sign of the potential biases due to the differential sample attrition between the winners and losers.

1.4 Impact on Magnet School Enrollment and Peer Achievement

1.4.1 The First-stage Effect on Magnet School Enrollment

Let $i = 1, ..., N$ index students, $j = 1, ..., J$ index magnet schools, and $t = 1, .., T$ index years (cohorts). For each of the $J \times T$ independently conducted lotteries, I can estimate a lottery-specific first-stage effect δ_{jt} that measures the impact of winning the admission lottery of magnet school j in year t on the probability of enrolling in that magnet school:

$$\delta_{jt} = E[S_i|D_i = 1, A_i = jt] - E[S_i|D_i = 0, A_i = jt] \tag{1.1}$$

where S_i is a dummy variable denoting applicant i's enrollment status in the magnet school she applied to, D_i is a binary variable denoting whether applicant i won her lottery, and A_i is a categorical variable denoting which lottery applicant i participated in. Note that each applicant in the sample participated in one and only one of the $J \times T$ lotteries.

For ease of interpretation, I estimate a regression-adjusted single-parameter first-stage effect of winning a lottery on enrolling in the magnet school of one applied to by the following model:

$$S_i = \delta D_i + \sum_{j=1}^{J}\sum_{t=1}^{T}\tau_{jt}\mathbf{1}(A_i = jt) + \epsilon_i \tag{1.2}$$

where $\mathbf{1}(A_i = jt)$ is an indicator for student i having participated in the admission lottery of magnet school j in year t, τ_{jt} is a lottery fixed effect, and ϵ_i is a stochastic error term.

As shown in Cullen et al. (2006), the coefficient δ can be expressed as a weighted average of δ_{jt}'s with the weight $\omega_{jt} = \frac{N_{jt}P_{jt}(1-P_{jt})}{\sum_j \sum_t N_{jt}P_{jt}(1-P_{jt})}$ where N_{jt} is the number of participants in lottery jt and P_{jt} is the winning rate of lottery jt.

Columns 1 and 5 of Table 1-2 present the ordinary least squares regression estimates of Equation (1.2) using the single-matched applicants of the full sample and the District 2 sample, respectively. The top row reports the "back door" entry rates among the lottery losers ($P[S_i = 1|D_i = 0]$). Over one-half (53 percent) of the lottery losers in the full sample managed to get into the magnet school they applied to through the "back door." Despite the high baseline enrollment rate among the lottery losers, winning a lottery still had a substantial first-stage effect in the full sample: increasing the probability of enrolling in the magnet school for which the lottery was held by 33 percentage points. Compared to the full sample, the District 2 sample had a lower "back door" admission rate (only two-fifths of the lottery losers were enrolled in their choice magnet school), and a correspondingly higher first-stage effect of winning a lottery on enrolling in the magnet school one applied to (about 49 percentage points).

Selection in "Back Door" Entry

Columns 2 and 6 of Table 1-2 report the OLS regression estimates for specifications that include covariates – such as gender, qualification status, and pre-lottery test scores, if available – and their interactions with the lottery status for the single-matched applicants of the full sample and the District 2 sample, respectively. Results using the full sample show that being a qualified applicant increased the "back door" entry rate by about seven percentage points, suggesting evidence of positive selection on student ability via "back door" admission.

Results of the District 2 sample confirm the positive "back door" selection: a one standard deviation increase in the pre-lottery test score raised the probability that a lottery loser would attend the magnet school she applied to by five percentage points. After controlling for students' pre-lottery test scores, the marginal impact of being a qualified applicant on the "back door" entry rate remained positive, though statistically insignificant. Figure 1-3-A plots the Kernel density curves of the pre-lottery test scores for the single-matched lottery losers in District 2 by their enrollment status at the magnet school of their choice. Lottery losers who attended the magnet school they applied to through the "back door" had higher pre-lottery test scores (with a mean of 0.42 SD) than those who did not (with a mean of 0.25 SD).

The take-up rate among lottery winners was about 90 percent, indicating that ten percent of the winners gave up their admission offer. Next, I check whether lottery winners who gave up their option to attend a magnet school differed from those who exercised their option in their pre-lottery achievement. Figure 1-3-B plots the Kernel density curves of the pre-lottery test scores for lottery winners in District 2 by their take-up status. The two-sample Kolmogorov-Smirnov test (with a p-value of 0.948) cannot reject the equality of the two distributions.

Columns 3-4 and 7-8 of Table 1-2 report the weighted (by number of matches) least squares regression estimates using applicants with one to five matches of the full sample and the District 2 sample, respectively. The results are qualitatively the same and quantitatively very similar to those obtained using the single-matched applicants only.

1.4.2 Impact on Peer Achievement

According to the evidence from District 2 (the only district where primary school test scores are available), there were several channels through which winning a lottery could improve peers' test scores: (1) qualified applicants admitted by advance admission had very high 6th-grade test scores (with a mean of 0.68 SD); (2) participants in the main admission lottery also had above-average 6th grade test scores (with a mean of 0.31 SD); and (3) lottery losers who attended magnet schools through the "back door" had a higher average 6th-grade test score than lottery losers who did not attend (0.42 SD vs. 0.25 SD).

In the following, I examine the impact of winning a lottery on the achievement of a student's peers measured by their test scores on the HSEE. Let $\overline{Y}_{(-i)gt}$ stand for the peer mean HSEE test score of student i, i.e., the average HSEE test score of the cohort t of the middle school g that student i attended excluding her own test score.[12] The effect of winning a lottery on peer achievement $Y_{(-i)jt}$ is estimated using the following regression model:

$$\overline{Y}_{(-i)gt} = \alpha D_i + X_i'\beta_1 + \sum_{j=1}^{J}\sum_{t=1}^{T}\rho_{jt}\mathbf{1}(A_i = jt) + \eta_{igt} \tag{1.3}$$

where X_i' is a vector of the predetermined characteristics of the applicant, such as gender, qualification status in application, and pre-lottery test scores, if available; ρ_{jt} is a lottery fixed effect; and η_{igt} is a stochastic error. For all the HSEE takers from the North Bank every year, their test scores are standardized to have zero mean and unit variance.

[12] In practice, I do not have the HSEE test scores for all students in the North Bank, but instead obtained a 3.5 percent random sample of all HSEE takers from the North Bank. For most of the lottery participants not in the random sample, their mean peer achievement is estimated as the average among all students in the random sample who belong to the cohort of the middle school of their enrollment; while for a small fraction of lottery participants who happen to be in the random sample, their own HSEE test scores are excluded when calculating their mean peer achievement.

Table 1-3 shows the impact of winning a lottery on the school-grade-level average peer achievement on the HSEE. In the full sample, winning a lottery increased the school-grade-level average peer achievement on the HSEE by about 0.11 SD (from the losers' mean of 0.19 SD); while for the District 2 sample, winning a lottery increased the average peer achievement on the HSEE by about 0.19 SD (from the losers' mean of 0.10 SD). Lottery losers from District 2 had a lower mean peer achievement than those from other districts for two reasons. First, District 2 is a relatively disadvantaged inner-city district in terms of students' average HSEE test scores. Second, the "back door" entry rate among lottery losers was lower in District 2 than in the other two districts in the North Bank. However, because of the relatively larger first-stage effect of winning a lottery on magnet school enrollment in District 2, the win/loss difference in the peer mean HSEE test scores was larger in the District 2 sample than in the full sample.

1.5 Impact on Student Outcomes

1.5.1 The Intent-to-Treat Effect on Student Outcomes

In an ideal random assignment of treatment with no missing outcomes, winners and losers of a particular lottery are balanced, on average, in terms of both their observable and unobservable characteristics. Consequently, a simple difference between the observed outcomes of the winners and losers of a particular lottery provides a consistent estimate of the intent-to-treat (ITT) effect: the impact of being offered the option to attend that magnet school for students who applied to it in that year. Let Y_i denote an outcome measure of student i.

Then the ITT effect of winning magnet school j's lottery in year t can be expressed as:

$$\pi_{jt} = E[Y_i|D_i = 1, A_i = jt] - E[Y_i|D_i = 0, A_i = jt] \tag{1.4}$$

where A_i is a categorical variable denoting which lottery applicant i participated in. The parameter π_{jt} is of direct interest if the focus is the impact of having magnet school j in the choice set of students of cohort t who expressed an interest in it. The regression-adjusted ITT effect of winning a lottery on student outcome Y_i can be estimated using the following regression model:

$$Y_i = \pi D_i + X_i'\beta_2 + \sum_{j=1}^{J}\sum_{t=1}^{T}\kappa_{jt}\mathbf{1}(A_i = jt) + v_i \tag{1.5}$$

In this study, I construct five measures of students' ex post academic outcomes based on their performance on the HSEE and their secondary school enrollment status: total test scores on the HSEE, a dummy for scoring above the threshold for elite high school admission, a dummy for being admitted by an elite high school, a dummy for scoring above the threshold for regular high school admission, and a dummy for being admitted by any high school. Every year, the city education council announces the minimum score requirements for attending a regular high school and an elite high school, respectively. The two thresholds were usually around the 40th percentile and the 70th percentile of the HSEE test scores of that year.[13]

[13]In addition to these two thresholds, the city education council also sets minimum score requirement for different subcategories of vocational secondary schools. Each senior secondary school then announces a guaranteed admission score that cannot be lower than the threshold the city education council sets for the category it belongs to. After knowing their own HSEE test scores and the guaranteed admission scores of their interested schools, students then submit their ranked preferences for secondary schools – high schools and vocational secondary schools. As the eventual admission score of a school might be lower than its announced guaranteed admission score, students might be strategic in ranking their preferences. The admission process

If winning a lottery could improve students' (elite) high school enrollment status through channels beyond raising their HSEE test scores, we would expect a larger ITT effect on enrollment status than on scoring above the corresponding threshold.

Panels A through E of Table 1-4 present the losers' means and the ITT effects of winning a lottery for each of these five outcome measures. Each column reports an estimated ITT effect using a separate sample. Despite the significant and substantial positive effects of winning a lottery on magnet school attendance and peer achievement, I find little evidence that winning a lottery improved students' academic outcomes three years later. None of the 20 ITT coefficient estimates are significant at the 10 percent level, with 13 coefficients being estimated to be negative (though statistically insignificant). The ITT coefficient estimates of the full sample can reject a positive effect of 0.05 SD on the HSEE test scores and a gain of three percentage points for any of the other four binary outcome measures at the five percent level. The ITT effects on being admitted by an elite high school do not seem to be larger than the effects on scoring above the threshold for elite high school admission, suggesting no evidence that winning a lottery improved students' enrollment status at elite high schools through any nonacademic channel. The same conclusion holds when examining (elite and regular) high school enrollment status in general. The ITT coefficients using the District 2 sample are less precisely estimated, but are qualitatively similar to the results using the full sample. Estimates using the District 2 sample indicate that much of the variation in students' HSEE test scores can be explained by their 6th grade test score difference: a one-standard-deviation rise in the 6th grade test score would lead to a 0.61 SD increase in the

---

is computerized and under direct control of the admission office of the city education council. Even though "back door" admission is not eliminated, it plays a much smaller role in high school admission than it does in middle school admission.

HSEE test score.

1.5.2 Instrumental Variable Estimates of Magnet School Effects

Section 1.5.1 focuses on the ITT effects of winning a lottery on student academic outcomes. The lottery results, however, were not completely binding. As discussed in Section 1.4, a substantial fraction of lottery losers (about one-half) attended the magnet schools they chose through the "back door", while a small fraction (about one-tenth) of lottery winners did not exercise their choice option. Using applicants' lottery status as an instrument, this section presents the two-stage least squares (2SLS) estimates of the effects of attending a magnet school on students' academic outcomes. Under the exclusion assumption that winning a lottery only affected students' outcomes through its effect on their enrollment status in the magnet school they chose, the 2SLS regression estimates the following model:

$$Y_i = \gamma S_i + X_i'\beta_3 + \sum_{j=1}^{J}\sum_{t=1}^{T}\psi_{jt}\mathbf{1}(A_i = jt) + \mu_i \tag{1.6}$$

The associated first-stage relationship instruments the enrollment status at the magnet school they chose (S_i) using their lottery status (D_i):

$$S_i = \theta_{jt} D_i + X_i'\beta_4 + \sum_{j=1}^{J}\sum_{t=1}^{T}\tau_{jt}\mathbf{1}(A_i = jt) + \nu_i \tag{1.7}$$

In addition to including a vector of individual characteristics X_i' as controls, Equation (1.7) differs from equation (1.2) in that equation (1.7) allows the first-stage effect to vary across lotteries, while equation (1.2) estimates a single parameter for the weighted average first-stage effect.

Table 1-5 presents both the OLS and 2SLS estimates of the effects of attending the magnet school students applied to on the five outcome measures discussed in Section 1.5.1. The general picture of the OLS results is that applicants who attended the magnet school they chose outperformed those who did not. All the ten OLS coefficient estimates are positive and statistically significant at the 10 percent level, with eight being significant at the one percent level. Because of the positive selection on student ability via "back door" admission as discussed earlier, the OLS estimates of magnet school enrollment effects are biased upward. Following the previous literature on treatment effects under imperfect compliance (e.g., Imbens and Angrist, 1994), the 2SLS estimates of γ should be interpreted as the local average treatment effects (LATE) of attending a magnet school on the academic outcomes for a subgroup of lottery winners who would not have attended magnet schools had they lost their lottery. Similar to the results of the ITT coefficient estimates, none of the 20 2SLS coefficients (estimated for the five outcome variables using four samples) is significant, suggesting no evidence of any significant effect of attending a magnet school on students' academic outcomes after controlling for selection via "back door" entry. Though statistically insignificant, the 2SLS estimates of the treatment effects on the HSEE test scores are negative for all of the four samples. The point estimates and standard errors of these four samples all allow me to reject a modest gain of 0.1 SD in HSEE test scores at the five percent level.

Table 1-6 shows the quantile regressions and the quantile treatment effects of attending a magnet school on students' HSEE test scores. The quantile regression indicates that applicants who attended a magnet school had higher HSEE test scores than those who did not for all deciles, with the gap by enrollment status around the median being larger than that at the lower or upper tail. The quantile treatment effects are estimated using lottery status as an instrument based on the methodology presented in Abadie et al. (2002). There is some suggestive evidence that low-achieving students might have been worse off by attending these high-achieving magnet schools. But there is no evidence that students at the upper tail of the test score distribution benefited by attending these magnet schools.

1.6 Accounting for Differential Attrition

As discussed in Section 1.3, the outcome data are subject to a small degree of differential sample attrition between the winners and losers in the study. A small fraction of applicants – who would have opted for schooling outside the North Bank had they lost their lottery – were induced to enroll in the North Bank after winning their lottery. If the characteristics of these applicants (whose outcomes would be observed only if they had won their lottery) differed from those whose outcomes would be observed, irrespective of their lottery status, the main empirical results presented above would be biased.

In the following, I outline a framework using the easily analyzed difference in the means of the observed winner and loser samples to examine the magnitude and the likely sign of the potential biases in the estimated ITT effects due to the differential sample attrition. The framework is similar in spirit to earlier work on treatment effect analysis with missing outcomes (e.g., Manski, 1990; Lee, 2002).[14] Let T_{1i} and T_{0i} be the latent dummy variables denoting whether the outcome of an applicant is tracked when $D_i = 1$ and $D_i = 0$, re-

[14]The framework presented in the paper only considers the first-order issue of differential sample attrition between the lottery winners and losers and ignores the interaction between matching errors and differential sample attrition.

spectively. For example, $T_{1i} = 1$ and $T_{0i} = 0$ imply that the outcome would be tracked if $D_i = 1$, but would not be tracked if $D_i = 0$. Let Y_{1i} and Y_{0i} be the potential outcomes of interest when $D_i = 1$ and $D_i = 0$, respectively. In practice, we observe (Y_i, T_i, D_i), where $T_i = T_{0i} + (T_{1i} - T_{0i})D_i$; $Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$ if $T_i = 1$; and Y_i is missing if $T_i = 0$. Following the existing literature, I assume that winning a lottery has a monotone impact on sample selection.

Axiom 1 *Winning a lottery, if it affects sample selection at all, only induce some individuals who otherwise would drop out of the sample to stay in the sample, but not the opposite.*

$$T_{1i} \geq T_{0i}$$

The monotonicity assumption excludes the possibility that an applicant would be induced to drop out of the sample because of winning a lottery, but still allows $\Pr(T_{1i} = 1, T_{0i} = 0)$ to be positive, i.e., some individuals might be induced to stay in the sample after winning a lottery. The difference between the means of the observed treatment and control groups can be expressed as:

$$\begin{aligned}
&E[Y_{1i}|T_{1i} = 1] - E[Y_{0i}|T_{0i} = 1] = E[Y_{1i}|T_{0i} = 1] \Pr[T_{0i} = 1|T_{1i} = 1] + \\
&E[Y_{1i}|T_{1i} = 1, T_{0i} = 0] \Pr[T_{0i} = 0|T_{1i} = 1] - E[Y_{0i}|T_{0i} = 1] \\
= \;& \underbrace{E[Y_{1i} - Y_{0i}|T_{0i} = 1]}_{\text{parameter of interest}} + \underbrace{\{E[Y_{1i}|T_{1i} = 1, T_{0i} = 0] - E[Y_{1i}|T_{0i} = 1]\} \Pr[T_{0i} = 0|T_{1i} = 1]}_{\xi_1}
\end{aligned} \tag{1.8}$$

where the first component $E[Y_{1i} - Y_{0i}|T_{0i} = 1]$ is the parameter of interest – the ITT effect of a subpopulation whose outcomes would be observed, irrespective of their treatment assignment status – and the second component (ξ_1) is the bias of using the observed difference between the tracked winners and losers ($E[Y_{1i}|T_{1i} = 1] - E[Y_{0i}|T_{0i} = 1]$) as an estimate of $E[Y_{1i} - Y_{0i}|T_{0i} = 1]$.

Given the small degree of differential sample attrition ($\Pr[T_{0i} = 0|T_{1i} = 1] \approx 3\%$), the bias of using the difference in the means of observed winners and losers as an estimate of the ITT effect of the $T_{0i} = 1$ subpopulation is likely to be very small. In addition, I examine whether and to what extent the contamination group ($T_{0i} = 0, T_{1i} = 1$) differed from the group of interest ($T_{0i} = 1$) in terms of students' pre-lottery test scores for applicants from District 2. This is a highly relevant test, as students' ex post academic outcomes were highly correlated with their pre-lottery test scores (see Section 1.5.1). In order to conduct the comparison, I impose a second assumption:

Axiom 2 *The unidentified sampling error in the pre-lottery test scores (x_i) for the $T_{0i} = 1$ subpopulation is the same as the sampling error for the entire population (which is identified):*

$$E[x_i|D_i = 1, T_{0i} = 1] - E[x_i|D_i = 0, T_{0i} = 1] = E[x_i|D_i = 1] - E[x_i|D_i = 0]$$

Table 1-7 shows that, among the observed winners, the contamination group ($D_i = 1, T_{0i} = 0, T_{1i} = 1$) had an average 6th grade test score that was 0.53 SD higher than the average of the subpopulation of interest ($D_i = 1, T_{0i} = 1$), and the difference is statistically significant. This indicates that the estimated ITT effects and the corresponding 2SLS estimates of the treatment effects are likely to be biased upward. The magnitude of the potential upward bias, however, is very small. Given the point estimate of the impact of the pre-lottery test score on the HSEE test score (0.60) and the degree of differential sample attrition (0.03), the magnitude of the bias in the estimated ITT effect on the HSEE test scores is on the order of 0.01 SD (.53 × .60 × 03), a size that is almost negligible.

1.7 Parental School Choice

Given the high tuition cost and the lack of systematic academic benefits, why is competition for entry into these magnet schools so intense? One possibility is that parents and students may prefer these magnet schools for reasons other than academic performance, such as physical environments, classroom discipline, and peer behavior. Anecdotal evidence does indicate that magnet schools are superior in such nonacademic school attributes compared to neighborhood middle schools. Considering the rigorous exam-based high school admission in China, it is hence unlikely that Chinese parents choose magnet schools primarily out of concern for these nonacademic attributes of magnet schools.

Another possibility is that parents place high value on average student achievement when choosing a school, either because they value peer quality beyond its effect on value added, or because they confuse average achievement with value added. Even though the empirical evidence cannot distinguish these two underlying reasons, either of them being true would lead schools to be sought for their observed student achievement. I construct two popularity measures for these magnet schools: lottery winning rates and winner take-up rates. Table 1-8 presents school-level mean statistics on the HSEE test scores (column 3), estimated value added effects (column 4), lottery winning rates (column 5), winner take-up rates (column 6), loser "back door" entry rates (column 7), and first-stage effects on enrollment (column

8) for all the eight magnet schools in the North Bank. Figures 1-4-A and 1-4-B show the two popularity measures plotted against schools' average HSEE test scores, respectively. Results demonstrate that the highly popular magnet schools are those with high student achievement on the HSEE, suggesting that parents might indeed use achievement measures to guide their school choice. Columns 1 and 4 in Table 1-9 report the OLS regressions of the lottery winning rates and the winner take-up rates on the average student achievement, respectively. The estimates confirm the results shown in Figures 1-4-A and 1-4-B. Columns 3 and 6 further show that the impact of average student achievement on school popularity remains marginally significant even after controlling for the value added measure I estimate for each magnet school.

One explanation is that parents may value peer quality beyond its effect on test scores. It could be that parents place high intrinsic weight on value added, but modest intrinsic weight on peer quality (beyond its effect on value added), in choosing a school. However, because value added effects are very imprecisely measured, while peer quality can be observed accurately, the high intrinsic weight on value added is swamped by the noisy measure, resulting in schools being chosen mainly for their observed peer quality.

Another explanation is that, given the lack of any reliable value added measure, parents may instead use the easily obtainable achievement measure as a proxy for value added (Figlio and Lucas, 2004). Using achievement to proxy for value added is not a big problem if the former is indeed a good proxy for the latter, i.e., the two measures are highly positively correlated. Figure 1-5 plots the estimated value added effects against schools' average HSEE test scores for these magnet schools and shows that the two measures are largely uncorrelated (with a correlation coefficient of 0.08). This echoes the previous finding of a weak correlation between school grades and value added in the U.S. school accountability literature (see Kane and Staiger, 2002, for a survey). Figures 1-6-A and 1-6-B show the two popularity measures plotted against the estimated value added effects, respectively. If parents indeed preferred and were able to identify high value added schools, we would expect to see higher value added magnet schools to have lower lottery winning rates but higher take-up rates of lottery winners. Figure 1-6-A shows a positive correlation between lottery winning rates and schools' value added effects, opposite to what one would otherwise expect. Figure 1-6-B also shows little evidence that the take-up rates of lottery winners are higher for schools with higher value added – the correlation between the two measures is 0.07. These results are consistent with an explanation that parents using the easily obtainable achievement measure to proxy for value added fail to identify schools with higher value added when the quality of the proxy is poor.

1.8 Conclusion

This chapter evaluates the impact of attending a magnet school in China on student achievement by exploiting random school admissions lotteries in Wuhan. Magnet school graduates had much higher HSEE test scores compared to those graduated from neighborhood middle schools. However, the cross-sectional superiority is likely to be spurious, largely attributable to student selection in lottery participation and "back door" entry, as well as the advance admission policy that favored the gifted and talented students. Despite that lottery winners had access to a better peer group, better qualified teachers, and a school of their choice (and hence might better suit their learning needs), I find little evidence that winning a lottery is associated with any academic benefit to students. The point estimates of the two-stages least squares (2SLS) regressions of students' HSEE test scores on their magnet school enrollment status using lottery results as an instrument are all insignificant, and allow me to reject a modest gain of 0.1 standard deviation (SD) at the five percent level.

I find that magnet school popularity, measured by either the competitiveness of the admission lottery or the take-up rate of lottery winners, is highly positively correlated with the average student achievement, but largely unrelated to the treatment effect on student achievement that I estimate for each school. This evidence suggests that parents value peer quality beyond its effect on achievement gains, or confuse average student achievement with value added. The finding that magnet schools are sought mainly for their observed superiority in average student achievement rather than for their academic value added casts doubt on the potential of school choice to improve student achievement, at least in this context.

Bibliography

[1] Alberto Abadie, Joshua Angrist, and Guido Imbens. Instrumental variable estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117, 2002.

[2] Joshua D. Angrist, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. Vouchers for private schooling in colombia: Evidence from a randomized natural experiment. *American Economic Review*, 92(5):1535–1558, 2002.

[3] Joshua D. Angrist, Eric Bettinger, and Michael Kremer. Long-term educational consequences of secondary school vouchers: Evidence from administrative records in colombia. *American Economic Review*, 96(3):847–862, 2006.

[4] Martin Carnoy, Frank Adamson, Amita Chudgar, Thomas F. Luschei, and John F. Witte. *Vouchers and Public School Performance.* Economic Policy Institute, Washington,DC, 2007.

[5] John Chubb and Terry Moe. *Politics, Markets, and America's Schools.* Brookings Institution Press, 1990.

[6] Julie Berry Cullen, Brian A Jacob, and Steven Levitt. The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica*, 74(5):1191–1230, 2006.

[7] D. Figlio and C Rouse. Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics*, 92:239–55, 2006.

[8] David N. Figlio and Maurice E. Lucas. What's in a grade? school report cards and the housing market. *American Economic Review*, 94(3):591–604, 2004.

[9] Milton Friedman. *Capitalism and Freedom.* University of Chicago Press, Chicago, 1962.

[10] Jay P. Greene, Paul E. Peterson, and Jiangtao Du. Effectiveness of School Choice: The Milwaukee Experiment. *Education and Urban Society*, 31:190–213, 1999.

[11] Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. Parental preferences and school competition: Evidence from a public school choice program. Working Paper 11805, National Bureau of Economic Research, 2005. NBER Working Paper 11805.

[12] Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger. Preferences and heterogeneous treatment effects in a public school choice lottery. Working Paper 12145, National Bureau of Economic Research, 2006.

[13] W. Howell, P. Wolf, D. Campbell, and P. Peterson. School vouchers and academic performance: Results from three randomized field trials. *Journal of Policy Analysis and Management*, 21(2):191–217, 2002.

[14] Caroline M. Hoxby. "does competition among public schools benefit students and taxpayers. *American Economic Review*, 90(5):1209–38, 2000.

[15] Caroline M. Hoxby. School choice and school competition: Evidence from the united states. *Swedish Economic Policy Review*, 10:13–67, 2003.

[16] Caroline M. Hoxby and Jonah E Rockoff. The impact of charter schools on student achievement. Unpublished manuscript, Columbia University, 2004.

[17] Chang-Tai Hsieh and Miguel Urquiola. The effects of generalized school choice on achievement and stratification: Evidence from chile's voucher program. *Journal of Public Economics*, 90(8-9):1477–1503, 2006.

[18] Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62:467–76, 1994.

[19] Brian A. Jacob and Lars Lefgren. What do parents value in education? an empirical investigation of parents' revealed preferences for teachers. *Quarterly Journal of Economics*, 122(4):1603–1637, 2007.

[20] Brian A. Jacob and Steven D. Levitt. The impact of school choice on student outcomes: An analysis of the chicago public schools. *Journal of Public Economics*, 89(5-6):729–60, 2005.

[21] Thomas J. Kane and Douglas O. Staiger. The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, 16(4):91–114, 2002.

[22] Alan B. Krueger and Pei Zhu. Another look at the new york city school voucher experiment. *American Behavioral Scientist*, pages 658–98, 2004.

[23] Fang Lai, Elisabeth Sadoulet, and Alain de Janvry. Do school characteristics and teacher quality affect student performance? Evidence from a natural experiment in beijing middle schools. Unpublished manuscript, University of California, Berkeley, 2008.

[24] David S. Lee. Trimming for bounds on treatment effects with missing outcomes. NBER Technical Working Papers 0277, National Bureau of Economic Research, 2002.

[25] Charles F Manski. Nonparametric bounds on treatment effects. *American Economic Review Papers and Proceedings*, 80(2):319–23, 1990.

[26] Alejandra Mizala and Miguel Urquiola. School markets: The impact of information approximating schools' effectiveness. NBER Working Papers No. 13676, 2007.

[27] Jesse M. Rothstein. Good principals or good peers? parental valuation of school characteristics, tiebout equilibrium, and the incentive effects of competition among jurisdictions. *American Economic Review*, 96(4):1333–1350, 2006.

[28] Cecilia Elena Rouse and Lisa Barrow. School vouchers and student achievement: Recent evidence, remaining questions. Federal Reserve Bank of Chicago Working Paper 2008-08, 2008.

[29] J. Douglas Willms and Frank Echols. Alert and inert clients: The scottish experience of parental choice of schools. *Economics of Education Review*, 11(4):339–350, 1992.

Table 1-1 Predetermined Individual Characteristics and Matching Outcomes, by Lottery Status[a]

Dependent Variable	All Districts		District 2	
	Losers' mean	Win/loss difference	Losers' mean	Win/loss difference
	(1)	(2)	(3)	(4)
Panel A Predetermined Individual Characteristics				
Female	0.475	-0.011	0.500	-0.022
		(0.010)		(0.021)
Qualified applicant[b]	0.190	-0.020	0.209	0.004
		(0.012)		(0.018)
6th grade test score available	-	-	0.864	-0.002
				(0.015)
6th grade combined Chinese and mathematics score (in s.d.)	-	-	0.287	0.034
				(0.037)
Panel B Matching Outcomes				
Overall match rate	0.893	0.022 ***	0.905	0.017
		(0.006)		(0.012)
Single match	0.676	-	0.652	-
2 to 5 matches	0.170	-	0.189	-
Number of observations[c]	9,630	13,769	2,730	3,484

Notes: [a] The table reports the losers' means for each dependent variable indicated by the row heading and the coefficients of separate regressions of each dependent variable on an indicator variable for winning a lottery and a full set of lottery fixed effects. Robust standard errors are in parentheses. A triple asterisk (***) denotes significant at the 1 percent level.

[b] Observations used in this row only include participants of the lotteries that had involved qualified applicants.

[c] Number of observations shows the maximum sample size in each column.

Table 1-2 First-Stage Effect on Enrollment in the Choice Magnet School[a]

	Full Sample				District 2 Sample			
	Single-matched applicants		All matched applicants		Single-matched applicants		All matched applicants	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fraction of lottery losers enrolled in their choice magnet school[b]	0.534				0.396			
Won lottery	0.334 ***	0.281 ***	0.342 ***	0.293 ***	0.494 ***	0.496 ***	0.506 ***	0.479 ***
	(0.008)	(0.012)	(0.008)	(0.012)	(0.018)	(0.044)	(0.017)	(0.040)
Female		-0.012		-0.019 *		0.008		-0.002
		(0.011)		(0.011)		(0.023)		(0.020)
Won lottery * female		-0.005		-0.012		-0.042		-0.016
		(0.016)		(0.016)		(0.036)		(0.034)
Qualified applicant		0.070 ***		0.085 ***		0.040		0.059 **
		(0.023)		(0.021)		(0.033)		(0.030)
Won lottery * qualified applicants		-0.022		-0.036		-0.008		0.010
		(0.037)		(0.037)		(0.047)		(0.053)
Pre-lottery test score (in s.d.)						0.054 ***		0.049 ***
						(0.014)		(0.013)
Won lottery * pre-lottery test score						-0.063 ***		-0.061 ***
						(0.024)		(0.020)
Number of observations	9,424		11,734		1,990		2,661	

Notes: [a] The table reports the coefficients of OLS regressions of a binary dependent variable denoting whether a student was enrolled in her choice magnet school on an indicator variable of winning a lottery, a full set of lottery fixed effects as well as covariates and their interactions with the lottery status as indicated by the row headings. An interaction between a dummy variable denoting whether the lottery involved qualified applicants and lottery status is also included in columns (2), (3), (4), and (5) so that the coefficients on the interaction between qualified applicant dummy and lottery status does not include across-lottery variation in the first-stage effects. Robust standard errors are in parentheses. A triple asterisk (***) denotes significant at the 1 percent level; a double asterisk (**) denotes significant at the 5 percent level; a single asterisk (*) denotes significant at the 10 percent level.

[b] Only the losers' means of the dependent variable of the single-matched samples are reported. The losers' means of the multi-matched applicants are uninformative because of the matching errors due to common names.

Table 1-3 The Impact of Winning a Lottery on Peer Achievement, Single-matched Applicants[a]

	Full Sample (1)	District 2 Sample (2)
Dependent variable: School-grade-level peer mean HSEE test score		
Won lottery	0.110 ***	0.188 ***
	(0.029)	(0.078)
Female	-0.002	-0.012
	(0.006)	(0.016)
Qualified applicant	0.060 **	0.060
	(0.026)	(0.041)
Pre-lottery test scores (in s.d.)		0.035 **
		(0.016)
Losers' mean in the dependent variable	0.190	0.103
Number of observations[b]	9,394	1,982

Notes: [a] The table reports the coefficients of the OLS regressions of the leave-one-out peer mean HSEE test score at the middle school-grade one attended on an indicator variable of winning a lottery, gender, qualification status, and pre-lottery test score, if available, and a full set of lottery fixed effects, as well as lottery losers' means in the dependent variable. Robust standard deviations are in parentheses. Standard errors are corrected for the within-school-grade-of-enrollment clustering. A triple asterisk (***) denotes significant at the 1 percent level; a double asterisk (**) denotes significant at the 5 percent level.

[b] The sample sizes are slightly smaller as the peer mean achievement is missing for a very small fraction (less than one percent) of the single-matched application.

Table1-4 Reduced-form Effect of Winning a Lottery[a]

	Full Sample		District 2 Sample	
	Single-matched applicants	All matched applicants	Single-matched applicants	All matched applicants
	(1)	(2)	(3)	(4)
Panel A HSEE test score				
Losers' mean in the dep var[b]	0.247		0.137	
	(0.999)		(1.009)	
Won lottery effect	-0.009	-0.004	-0.064	-0.036
	(0.024)	(0.026)	(0.061)	(0.090)
	[-.059, .040]	[-.058, .050]	[-.220, .092]	[-.266, .195]
Pre-lottery test scores (in s.d.)			0.607 ***	0.637 ***
			(0.013)	(0.030)
			[.574, .640]	[.560, .714]
Panel B Scoring above the threshold for elite high school admission				
Losers' mean in the dep var	0.415		0.354	
	(0.493)		(0.478)	
Won lottery effect	-0.001	0.001	-0.013	0.003
	(0.009)	(0.014)	(0.015)	(0.024)
	[-.019, .018]	[-.029, .032]	[-.053, .026]	[-.060, .066]
Panel C Being admitted by an elite high school				
Losers' mean in the dep var	0.310		0.262	
	(0.463)		(0.440)	
Won lottery effect	0.008	0.006	-0.005	0.001
	(0.010)	(0.018)	(0.018)	(0.025)
	[-.012, .028]	[-.031, .044]	[-.052, .043]	[-.064, .066]
Panel D Scoring above the threshold for regular high school admission				
Losers' mean in the dep var	0.665		0.624	
	(0.472)		(0.485)	
Won lottery effect	-0.005	-0.011	-0.016	0.033
	(0.013)	(0.021)	(0.039)	(0.052)
	[-.033, .023]	[-.056, .034]	[-.116, .084]	[-.101, .166]
Panel E Being admitted by any high school				
Losers' mean in the dep var	0.680		0.611	
	(0.466)		(0.037)	
Won lottery effect	-0.008	-0.015	-0.030	0.004
	(0.013)	(0.021)	(0.037)	(0.056)
	[-.035, .018]	[-.058, .028]	[-.126, .066]	[-.141, .149]
Number of observations	9,424	12,102	1,990	2,743

Notes: [a] Each panel of the table reports losers' means in the dependent variable as indicated by the panel headings and the the OLS regression estimates of the effect of winning a lottery on the dependent variable from models that include controls for gender, qualification status, and pre-lottery test score, if available, and a full set of lottery fixed effects. The estimated coefficients of the pre-lottery test scores are only reported in Panel A. Each column presents the estimates of a separate sample. Numbers in paratheses are standard deviations in rows of means and robust standard errors clustered by lotteries in rows of estimated won lottery effects. Numbers in brackets are the 95 percent confidence intervals. A triple asterisk (***) denotes significant at the 1 percent level.

[b] Only the losers' means of the dependent variable of the single-matched samples are reported. The losers' means of the multi-matched applicants are uninformative because of the matching errors due to common names.

Table 1-5 OLS and 2SLS Estimates of the Effect of Attending the Choice Magnet School[a]

Dependent variable	Full Sample				District 2 Sample		
	Losers' means	OLS	2SLSa	2SLSb	OLS	2SLSa	2SLSb
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
HSEE test score (in s.d.)	0.247	0.127 ***	-0.023	-0.037	0.092 **	-0.099	-0.103
	(0.999)	(0.022)	(0.058)	(0.068)	(0.038)	(0.085)	(0.104)
Scoring above the threshold for elite high school admission	0.415	0.078 ***	0.012	0.008	0.078 ***	-0.007	-0.012
	(0.493)	(0.011)	(0.028)	(0.033)	(0.020)	(0.042)	(0.051)
Being admitted by an elite high school	0.310	0.053 ***	0.026	0.005	0.048 ***	0.007	-0.018
	(0.463)	(0.010)	(0.026)	(0.030)	(0.019)	(0.039)	(0.044)
Scoring above the threshold for regular high school admission	0.665	0.049 ***	-0.013	-0.027	0.033 *	-0.017	0.024
	(0.472)	(0.011)	(0.029)	(0.047)	(0.020)	(0.043)	(0.082)
Being admitted by any high school	0.680	0.063 ***	-0.033	-0.052	0.057 ***	-0.047	-0.026
	(0.466)	(0.011)	(0.029)	(0.046)	(0.020)	(0.044)	(0.080)
Number of observations	6,510	9,424	9,424	11,734	1,990	1,990	2,661

Notes: [a] The table reports the OLS and IV estimates of the effect of attending the selected magnet school on each dependent variable indicated in the row heading. Each column presents the estimates of a separate sample. For columns 1 to 4, the exogenous control variables include gender, qualification status, and a full set of lottery fixed effect. For columns 5 to 8, the exogenous control variables include the pre-lottery test score, in addition to those used in columns 1 to 4. For all the IV estimates, the specifications allow the first-stage effect of winning a lottery on attendance at the selected magnet school to vary across lotteries. Robust standard errors are in parentheses. A tripe asterisk (***) denotes significant at the 1 percent level; a double asterisk (**) denotes significant at the 5 percent level; and a single asterisk (*) denotes significant at the 10 percent level.

Table 1-6 Quantile Regressions and IV Estimates of the Quantile Treatment Effects, Single-matched Applicatns[a]
(Dependent Variable: HSEE Test Scores in SD's)

		Quantile								
	OLS/2SLS	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
OLS/Quantile Regression										
Enrollment	0.116	0.142	0.099	0.114	0.196	0.170	0.169	0.156	0.118	0.092
	(0.051)	(0.043)	(0.042)	(0.032)	(0.033)	(0.034)	(0.026)	(0.023)	(0.022)	(0.019)
Y0	0.12	-1.22	-0.70	-0.30	-0.04	0.27	0.54	0.80	1.06	1.32
2SLS/Quantile Treatment Effect										
Enrollment	-0.122	-0.115	-0.215	-0.203	-0.170	-0.142	-0.099	-0.066	-0.049	-0.028
	(0.117)	(0.142)	(0.123)	(0.112)	(0.118)	(0.109)	(0.108)	(0.090)	(0.084)	(0.077)
Y0	0.12	-1.19	-0.68	-0.30	-0.02	0.26	0.54	0.78	1.05	1.33

Note: [a] Lottery status is used as an instrument for enrollment in estimating the quantile treatment effect. Other control variables include gender, qualification status and lottery fixed effects. Numbers in parentheses are robust standard errors. The quantile coefficients and standard errors are estimated using the program code used in Abadie, Angrist, and Imbens (2002).

Table 1-7 The Sign of Potential Biases

	x_i	s.d.	s.e.
$E[x_i \mid D_i = 0]$	0.287	(0.860)	
$E[x_i \mid D_i = 1] - E[x_i \mid D_i = 0]$	0.034		(0.037)
$E[x_i \mid D_i = 0, T_{0i} = 1]$	0.294	(0.864)	
$E[x_i \mid D_i = 0, T_{1i} = 1] - E[x_i \mid D_i = 0, T_{0i} = 1]$	0.046		(0.037)
$\widehat{E}[x_i \mid D_i = 1, T_{0i} = 1]$	0.328		
$\widehat{E}[x_i \mid D_i = 1, T_{0i} = 0, T_{1i} = 1]$	0.854		
$\widehat{E}[x_i \mid D_i = 1, T_{0i} = 0, T_{1i} = 1] - \widehat{E}[x_i \mid D_i = 1, T_{0i} = 1]$	0.527		(0.238)

Table 1-8 School Level Statistics [a]

School index (1)	District (2)	Mean HSEE score (3)	Estimated value added (LATE) (4)	Lottery winning rate (5)	Fraction enrolled, winners (6)	Fraction enrolled, losers (7)	First-stage effect (8)
1a	1	0.447	-0.072	0.288	0.928	0.770	0.159
1b	1	0.416	0.064	0.315	0.949	0.643	0.306
2c	2	0.755	0.027	0.168	0.890	0.441	0.449
2d	2	-0.061	0.214	0.378	0.855	0.380	0.475
2e	2	-0.080	-0.299	0.196	0.924	0.297	0.627
3f	3	0.430	-0.017	0.346	0.826	0.324	0.502
3g	3	0.333	-0.216	0.298	0.871	0.502	0.368
3h	3	0.035	-0.477	0.455	0.773	0.489	0.285

Notes: Column 3 reports the mean HSEE score of students who graduated from each magnet school during 2005-2007, including those who were admitted during the advance admission. Column 4 reports the LATE estimate of the value added effect of attending the magnet school on students' HSEE test scores. Columns 5 to 8 report the mean statistics of applicants who participated the main admission lotteries.

Table 1-9 The Effect of Achievement and Value Added on Popularity

	Dependent Variable					
	Lottery winning rate			Winner take-up rate		
	(1)	(2)	(3)	(4)	(5)	(6)
School mean HSEE test score (in s.d.)	-0.174 *		-0.193	0.083 *		0.083
	(0.109)		(0.110)	(0.045)		(0.048)
School value added measure (in s.d.)		0.102	0.137		0.015	0.000
		(0.135)	(0.126)		(0.059)	(0.055)
Number of observations	21	21	21	21	21	21

Notes: The table reports the OLS coefficient estimates of regression the dependent variable indicated by the column heading on the independent variable(s) indicated by the row headings and a full set of district * year fixed effect. Standard errors are reported in parentheses. A single asterisk denotes significant at the 10 percent level.

Figure 1-1 Map of Wuhan

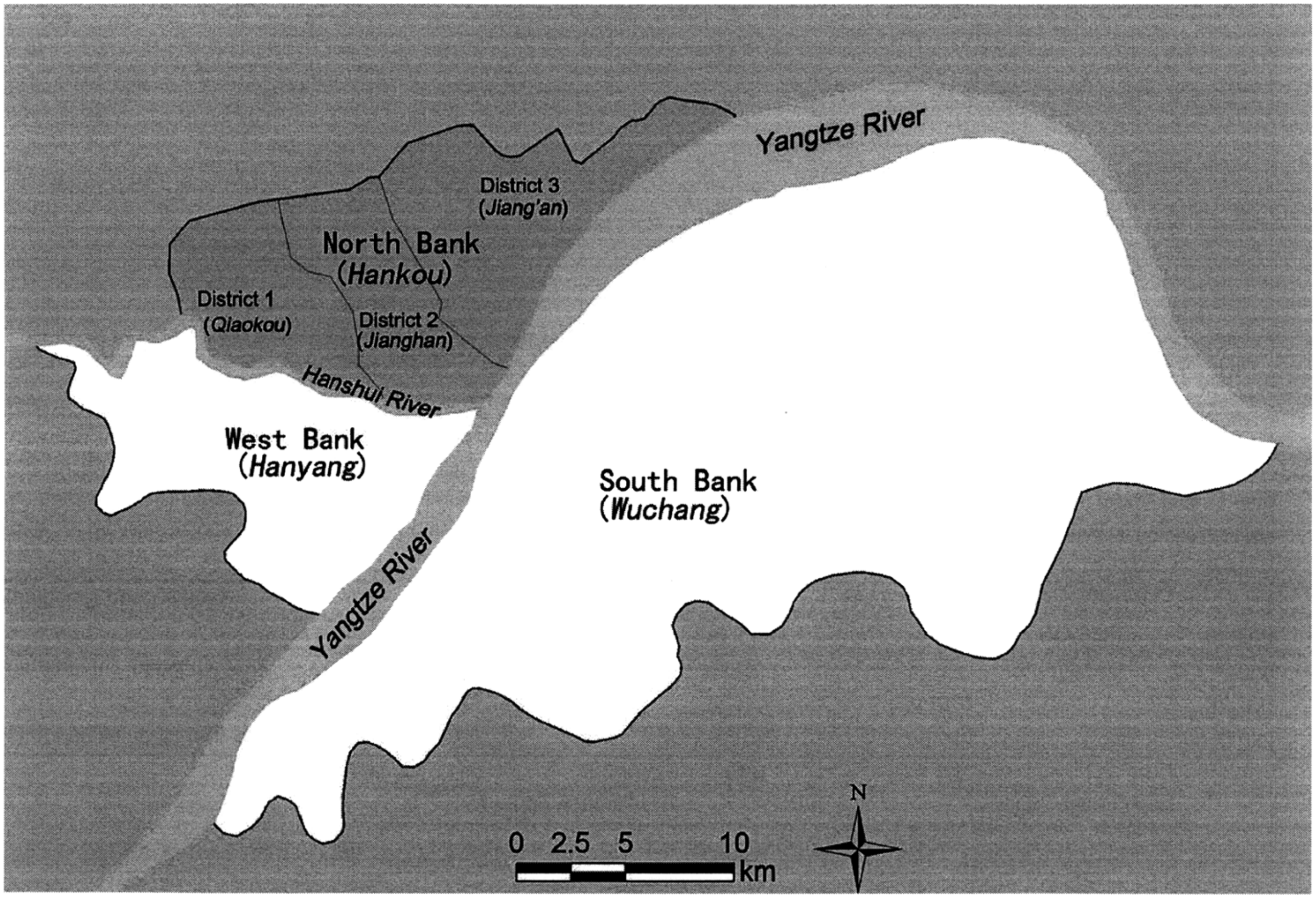
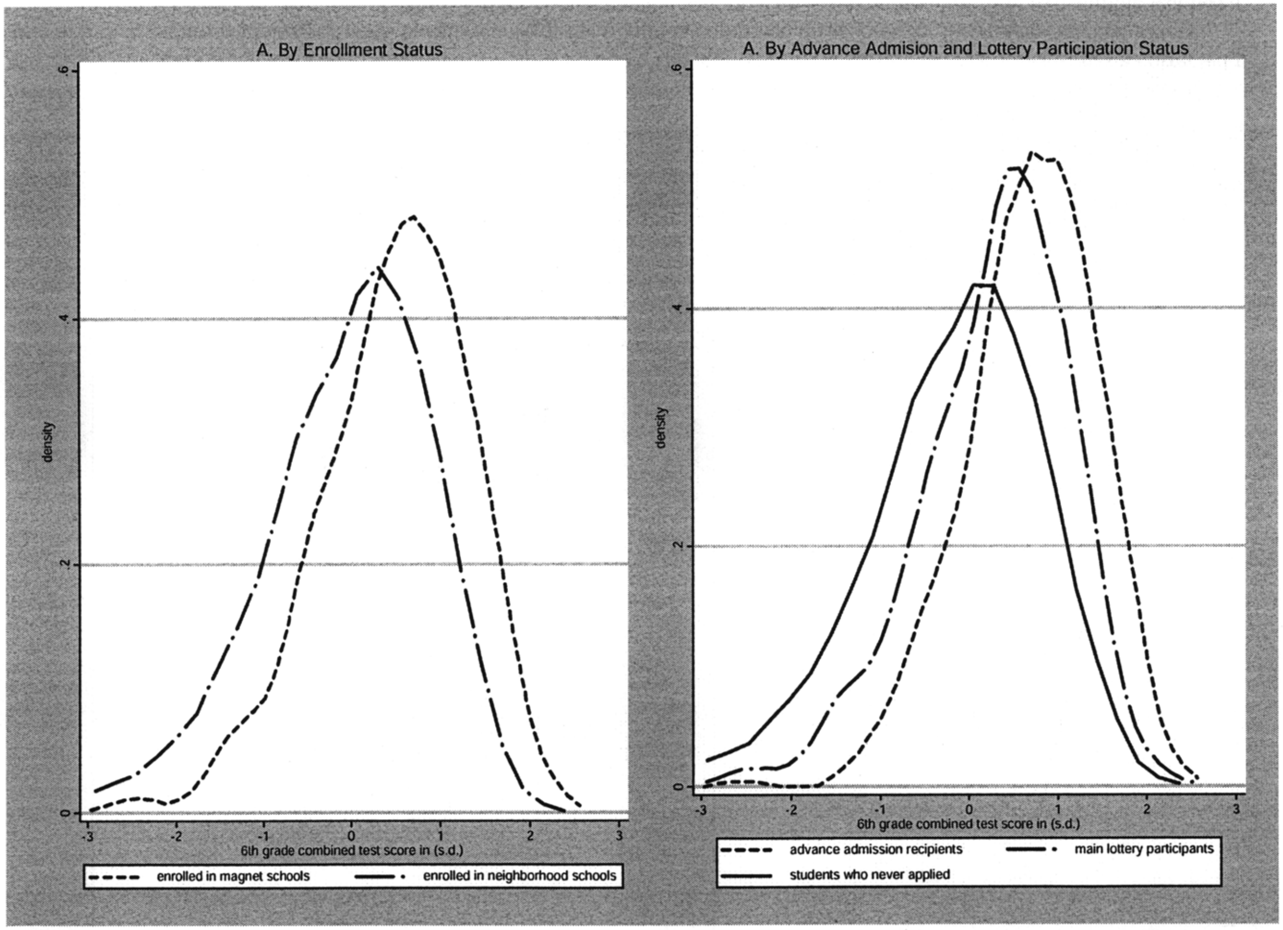
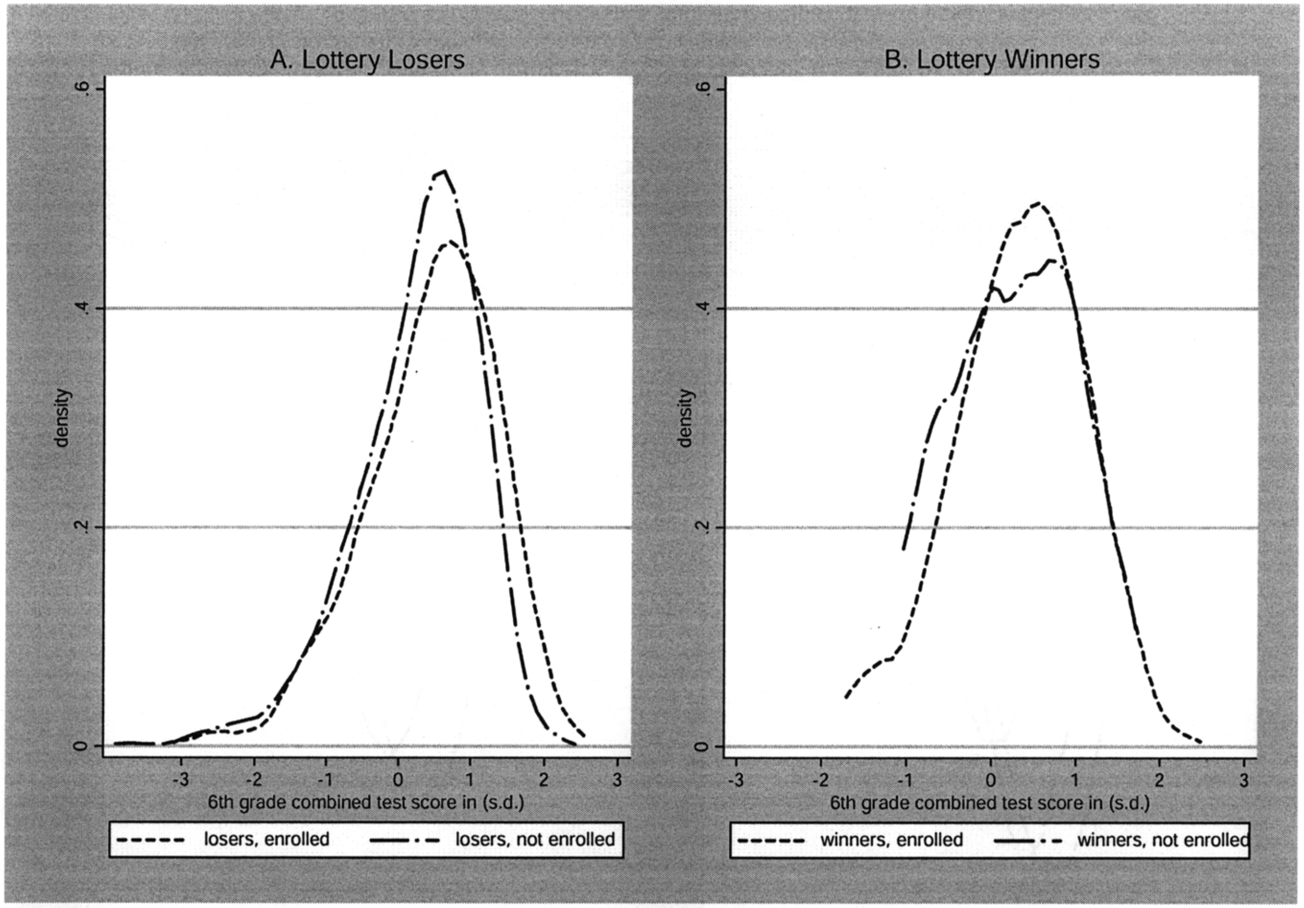


Figure 1-2 Pre-lottery Test Score Distributions, District 2 Sample



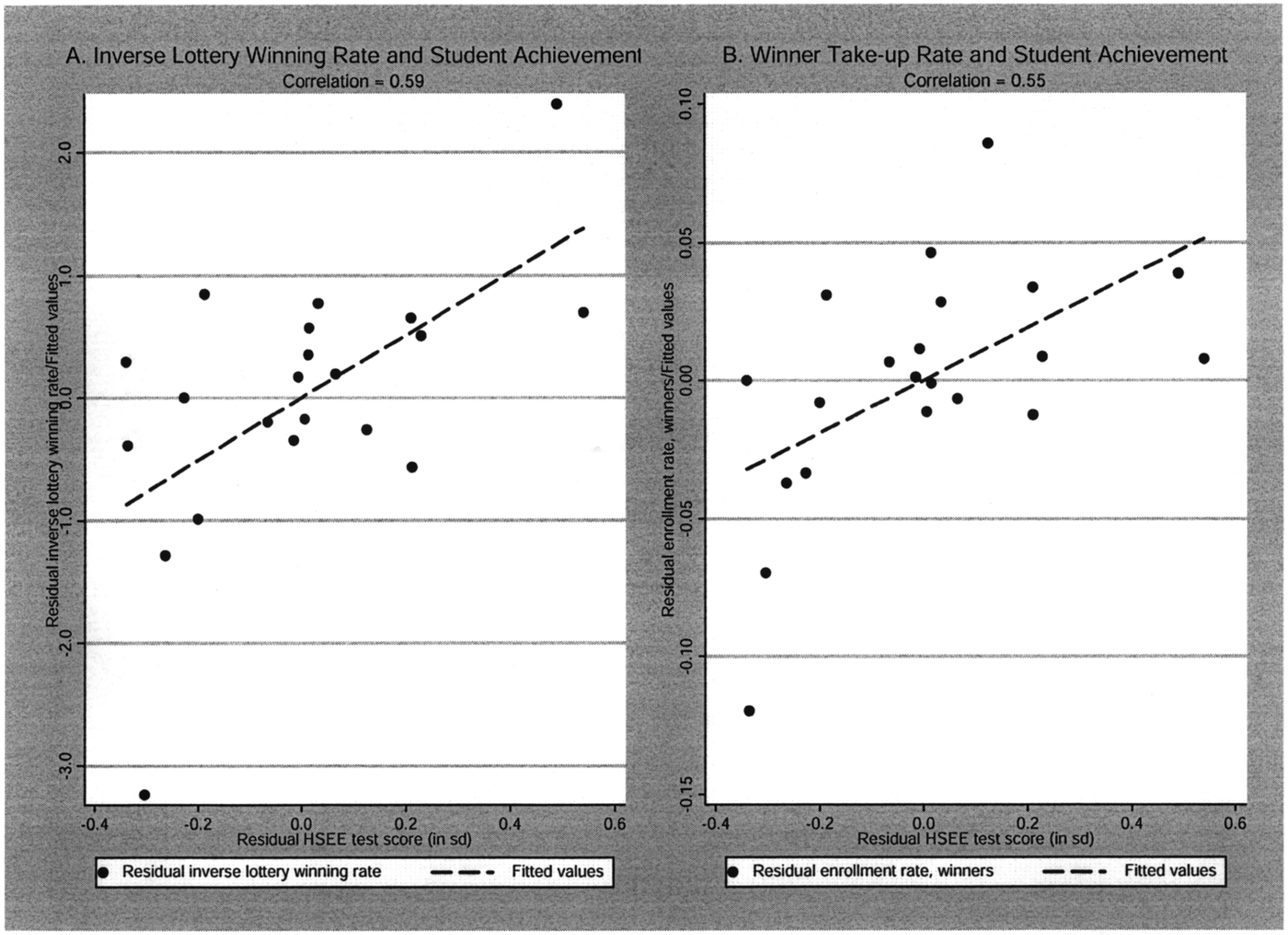
Notes: Figure A plots the Kernel density curve of students' 6th grade combined test scores by their magnet school enrollment status for students from District2. The Kolmogorov-Smirnov two-sample test has a p-value of 0.000. Figure 2 plots the Kernel density curve of the 6th grade combined test scores for advance admission recipients, main lottery participants, and nonparticipants in District 2, respectively. The Kolmogorov-Smirnov two-sample tests show the three distributions are all different to each other with a p-value of 0.000.

Figure 1-3 Pre-lottery Test Score Distributions by Lottery Status and Enrollment Status, District 2 Sample



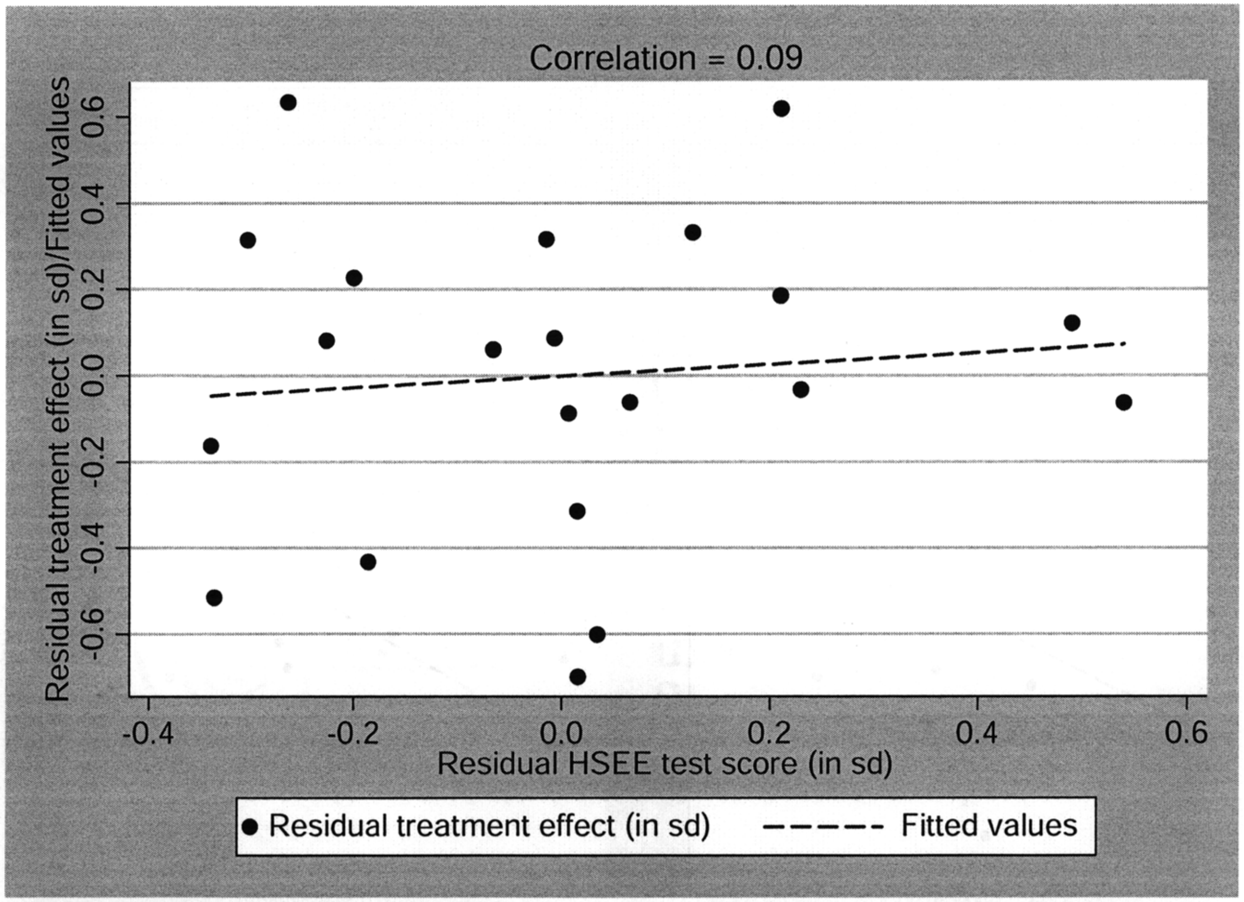
Notes: Figure A plots the Kernel density curve of the 6th grade combined test scores for the single-matched lottery losers in District 2 by their enrollment status in their choice magnet school. The Kolmogorov-Smirnov two-sample test has a corrected p-value of 0.000. Figure B plots the Kernel density curve of the 6th grade combined test scores for the single-matched lottery winners in District 2 by their enrollment status in their choice magnet school. The Kolmogorov-Smirnov two-sample test has a corrected p-value of 0.948.

Figure 1-4 School Popularity and Student Achievement



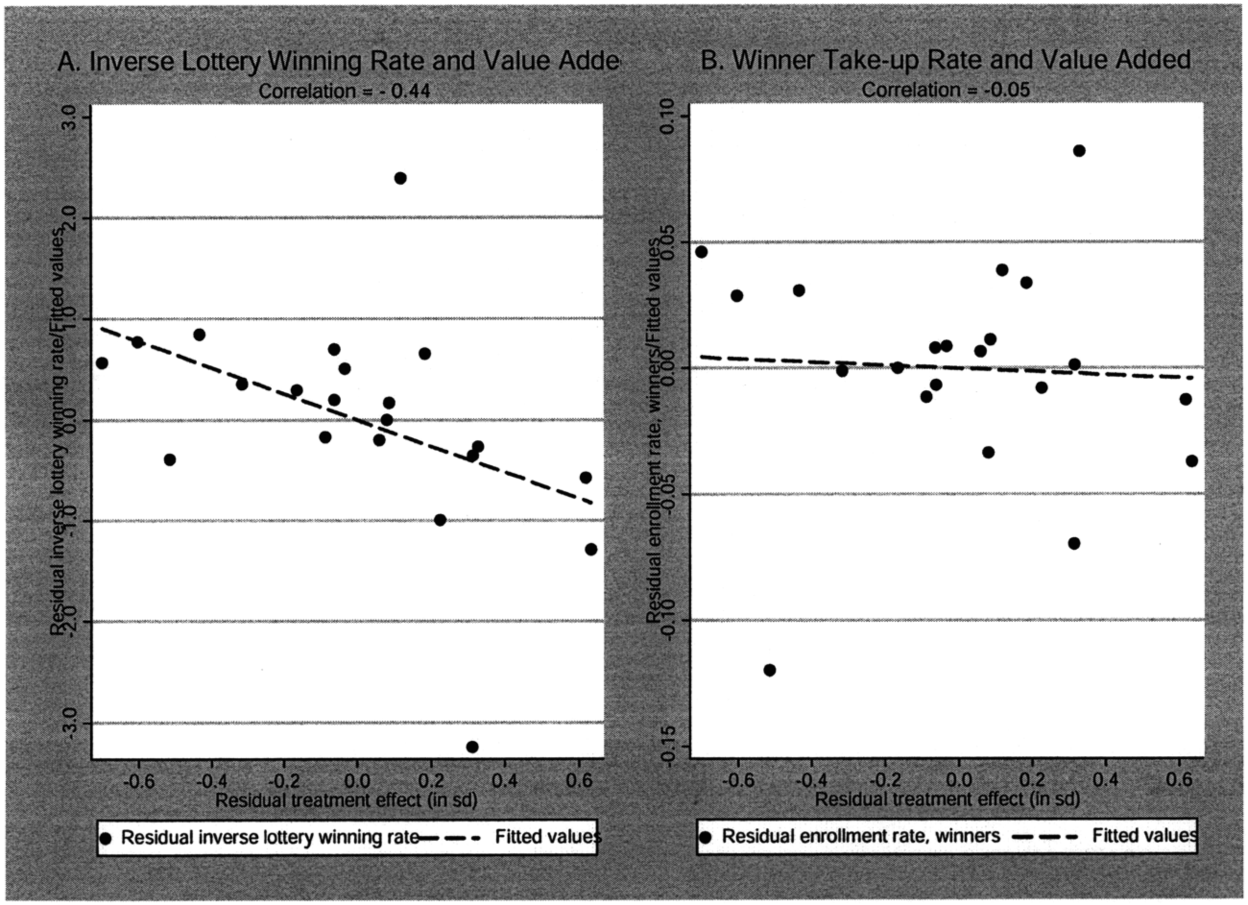
Notes: Figure 4A plots the residual inverse lottery winning rate and the residual school average HSEE test score (excluding the cohort of lottery participants) after controlling for full interactions of district and year fixed effects. Figure 6B plots the residual winner take-up rates and the residual school average HSEE test scores (excluding the cohort of lottery participants) after controlling for full interactions of district and year fixed effects.

Figure 1-5 Treatment Effects and Student Achievement



Notes: The graph plots the residual treatment effect and the residual school average HSEE test score (excluding the cohort of lottery participants) after controlling for full interactions of district and year fixed effects.

Figure 1-6 School Popularity and Treatment Effect



Notes: Figure 6A plots the residual inverse lottery winning rate and the residual treatment effect after controlling for full interactions of district and year fixed effect. Figure 6B plots the residual winner take-up rate and the residual treatment effect after controlling for full interactions of district and year fixed effect.

Chapter 2

Peer Effects on Student Achievement: Evidence from Middle School in China

2.1 Introduction

The effects of peer groups on students' academic performance play a prominent role in various education policy debates. Many current education interventions – for example, school choice, ability tracking, and affirmative action – have the potential to influence student outcomes through their impacts on peer composition. Understanding the structure and the magnitude of peer effects is therefore a critical ingredient in evaluating these policies. However, despite the importance of peer influences for education policies, empirical research has not yet reached a consensus on the existence and the nature of peer effects. While some studies show large positive effects of peer quality on academic achievement (e.g., McEwan, 2003; Kang, 2007), others find small or insignificant effects (e.g., Angrist and Lang, 2004; Lefgren, 2004).

The lack of consensus on peer influences reflects various challenges confronted by empirical research on peer effects (Manski, 1993; Moffitt, 2001; Brock and Durlauf, 2001). The first challenge is to isolate peer effects from "correlated effects" due to the correlation between peer composition and the omitted individual or institutional characteristics that can affect student outcomes. The second challenge, known as the "reflection problem," arises from the reciprocal nature of peer interactions: a student influences her peers and is also influenced by her peers, which causes a classical simultaneity problem of econometrics. These two challenges have engaged much of the attention of the peer effects literature and have been treated intensively. The past decade has seen the development of a variety of empirical strategies to identify exogenous sources of variation in peer characteristics to deal with the endogeneity problem. These recent studies have exploited within-school (grade) variation (e.g., Hoxby, 2000; Hanusheck et al., 2003; McEwan, 2003; Vigdor and Nechyba, 2004, 2006; Lavy and Schlosser, 2007; Ammermueller and Pischke, 2009; Gould, Lavy, and Paserman, forthcoming), within-student variation (e.g., Betts and Zau, 2004; Lavy, Silva, and Weinhardt, 2009), subgroup reassignment (e.g., Angrist and Lang, 2004; Hoxby and Weingarth, 2005), instrumental variables (IV) (e.g., Kang, 2007; Zabel, 2008), and random assignment (e.g., Sacerdote, 2001; Zimmerman, 2003; Duflo, Dupas, and Kremer, 2007, 2008). Identifying the structural parameters under the simultaneity problem, however, has proved difficult or impossible without imposing severe restrictions on the econometric model.[1] The empirical peer effects literature has often resorted to estimating the exogenous relationship between

[1]Necessary conditions for identification of the structural parameters can be found in Brock and Durlauf (2001). Two principle methods for identification are either to introduce some type of nonlinearity into the model (e.g., suppose individual behavior varies with other moments of group behavior) or to assume there exists one individual variable whose group-level average has no direct influence on individual outcomes or vice versa.

individual outcomes and predetermined measures of peer composition to circumvent the reflection problem (Nechyba, 2006). While some studies have focused on the relationship between exogenous peer characteristics (such as race, gender, and family background) and individual outcomes (e.g., Hoxby, 2000; Angrist and Lavy, 2004; Ammermueller and Pischke, 2009; Lavy and Schlosser, 2007; Gould, Lavy, and Paserman, forthcoming), other studies have benefited from panel data to include lagged outcome measures rather than contemporary values (e.g., Hanusheck et al., 2003; Ding and Lehrer, 2003; Vigdor and Nechyba, 2004, 2006; Lavy, Silva, and Weinhardt, 2009).[2]

In this chapter, we examine peer effects on students' math scores using a matched panel data set from China. The data set consists of 7,435 students from three successive cohorts of all the 15 middle schools in a school district and tracks their academic histories from finishing primary school (grades 1-6) to completing middle school (grades 7-9). By taking advantage of the panel data set, we address the reflection problem by focusing our interest on the exogenous relationship between predetermined peer characteristics – gender and lagged achievement in particular – and individual outcomes. We base the identification on within-school, between-cohort variation in peer composition, thereby controlling for omitted variables due to unobserved school characteristics and student sorting across schools. In terms of the identification strategy, the papers closest to ours are the ones by Hoxby (2000) and Gould, Lavy, and Paserman (forthcoming), both of which use comparisons in adjacent cohorts' peer composition within schools. Our identification strategy is also similar in spirit to studies that assume random classroom assignment within schools and use comparisons

[2] Hanushek et al. (2003) provide a thorough discussion of how this does and does not fully address the reflection problem.

across classrooms for the same cohort (grade) in the same school (e.g., McEwan, 2003; Vigdor and Nechayba, 2004, 2006; Kang, 2007; Ammermueller and Pischke, 2009).

A surprising result for our data is that the within-school estimate of the coefficient of the average lagged peer test score is negative and significant. We argue that the unexpected negative estimate of peer coefficient is explained by correlation in measurement errors between the individual- and the peer-level regressors. In our within-school estimation, we simultaneously control for lagged individual test score and the average lagged peer test score. These two variables, however, are subject to transitory common shocks due to the continuing presence of former peers in a student's current peer group. In this chapter, we refer to transitory common shocks as group-specific contextual, or environmental, influences that have only transitory effects on students' observed outcomes, i.e., these influences affect the observed test scores of all students in a group, but not their abilities. For example, if a teacher happens to cover in the classroom some materials that for random reasons are tested in the exam, the test scores of all students in the class will be inflated for this particular exam. The presence of such transitory common shocks will lead to a positive correlation in measurement errors between the individual- and the peer-level lagged test score variables. We derive formally that such a positive correlation in measurement errors will lead to a negative bias in the estimate of peer coefficient. In our context, this negative bias due to transitory common shocks on lagged test scores dominates the within-school estimator, making the point estimate negative and significant. The source of this bias is the presence of a student's former peers in her current peer group. The longitudinal structure of our data allows us to track the primary school origins of a student's peers and distinguish between new peers and old peers. In our sample, on average three-quarters of a student's middle school peers are new peers. Our way to address this transitory-common-shock problem is to use the lagged test score measures of new peers to instrument for the corresponding lagged test score measures of all peers. Under the assumption that transitory shocks in lagged test scores are uncorrelated for students from different primary schools, the measurement error in the lagged test scores of new peers is expected to be unrelated to the measurement error in lagged individual test score. Hence, the transitory-common-shock problem can be circumvented by using the lagged test score measures of new peers as instruments. The existence of transitory common shocks has been well documented in the school accountability literature, in which it leads a regression-to-the-mean problem in school or teacher assessment (e.g., Kane and Staiger, 2002; Betts and Dannenberg, 2002). The potential effect of the transitory common shocks on lagged individual and peer test scores, however, has largely been ignored in the peer effects literature. This chapter clarifies the econometric problem of transitory common shocks on lagged test scores and makes an important methodological contribution to the existing peer effects literature by proposing an IV strategy to correct this problem.

In our linear-in-means peer effects models, we examine the effect of peer gender mix and average lagged peer test score on a student's 9th-grade math score in a school fixed-effect framework. As discussed earlier, we instrument the average lagged peer test score with the average lagged test score of new peers to circumvent the transitory-common-shock problem. Unlike some previous studies that find positive spillover effects of girls on math scores (Hoxby, 2000; Whitemore, 2003; Lavy and Schlosser, 2009), we find no evidence that peer gender composition has an impact on students' 9th-grade math scores. Our within-school IV estimate of the linear-in-means model also shows little evidence that having peers of higher average lagged test score significantly improves a student's test score in math, although the IV coefficient is quite imprecisely estimated. This finding contrasts with some existing studies on educational peer effects in China, which have found significant and positive effects of average peer achievement (Ding and Lehrer, 2007; Lai, 2007; Carman and Zhang, 2008). We believe, however, that the within-school, between-cohort variation in peer composition we rely on for identification is more credibly exogenous than those in these previous studies.[3] Some recent well-identified studies also find no evidence of a significant positive effect of average peer achievement in the linear-in-means specifications and suggest other alternative peer effects models (e.g., Hoxby and Weingarth, 2005; Duflo, Dupas and Kremer, 2008).

We then explore some simple nonlinear peer effects models, allowing peer influences to operate through the dispersion of the distribution of lagged peer test scores or through the interaction between the distribution of lagged peer test scores and a student's initial achievement. Our results on the effect of peer heterogeneity, measured by the inter-quartile range (IQR) of lagged peer test scores, show that students benefit from having more homogeneous peers: the point estimate indicates that a 0.2σ reduction in the IQR of 6th-grade peer test scores, a magnitude of change over two-thirds of the middle schools in our sample had experienced among three adjacent cohorts, can increase a student's test score by 0.1σ. Estimates of heterogeneous peer effects models also show some interesting findings. First, a rightward shift in the distribution of lagged peer test scores benefits high-achieving students relative to low-achieving students, making the overall effect of the average lagged peer test score

[3]Ding and Lehrer (2007) rely on variation in peer quality across schools for identification. Even though they include observed school and teacher characteristics as control variables, the observed peer quality might still be endogenous to the unobserved school and teacher characteristics, resulting in omitted-variable biases in their estimates. Lai (2007) and Carman and Zhang (2008) both exploit variation in peer quality across classrooms within the same school-grade (cohort). They both argue that students are randomly assigned into classrooms within schools. However, even if the assignment rule is indeed random, imperfect compliance with initial classroom assignment would still lead to upward biases in their estimates. Zhang (2009) shows that a substantial proportion of students opt out of their assigned middle school in China.

insignificant. Second, a mean-preserving contraction in the distribution of lagged peer test scores benefits all students, but to a greater extent for those in the middle of a school's lagged test score distribution. Both of these findings are in favor of ability tracking for math learning.

The remainder of this chapter proceeds as follows: Section 2.2 provides background and describes the data; Section 2.3 discusses the econometric problem and presents the empirical strategy; Section 2.4 presents the empirical results on peer effects on student test scores; and Section 2.5 provides some concluding remarks.

2.2 Background and Data

2.2.1 Institutional Background and Data Construction

The cornerstone of this research is the analysis of peer effects on student achievement in middle school. The data come from administrative records of a school district in Wuhan. Based on the district's administrative records, we construct a matched panel data set that tracks three successive cohorts of middle school students in the district who had completed middle school between 2005 and 2007. The middle school system of the district includes 13 public neighborhood schools and two semi-private magnet schools. Upon graduation from primary school, each student is assigned to one of the 13 neighborhood middle schools based on their residency. The zoning scheme for middle school, however, is not fixed over time. The school district creates a new zoning scheme every year and announces it in June after 6th-grade students complete their primary school. As proximity has been taken into consideration in creating the zoning scheme, students know in advance the set of nearby neighborhood middle schools they might be assigned to, but do not know the exact school until the announcement by the school district. Since primary school enrollment is also based on residency, a student usually has some of her former schoolmates from primary school assigned to the same neighborhood middle school with her. Students also have the option to apply to one of the two semi-private magnet schools in the district and will be selected based on an admission lottery (Zhang, 2009).[4] On average, about 30 percent of the students in the district opt out from their assigned neighborhood middle school to a semi-private magnet school.

Our panel data set is constructed by matching administrative student records from two sources. Student information at the end of middle school comes from the city's High School Entrance Exam (HSEE) database, which includes each student's middle school of graduation and test scores in four subjects examined in the citywide uniform HSEE: math, science, Chinese, and English. Student information before the start of middle school comes from the district's records of students' primary school of graduation and their math scores in a district-wide uniform exam taken at the end of 6th grade.[5]

Some limitations remain in the structure of the matched panel data set. First, the two databases do not share perfect individual identification information to guarantee unique tracking of the academic histories of all students. Specifically, we can only use the combination of name and gender to match student records in the two databases. Some students

[4]Neighborhood middle schools are tuition-free under the compulsory education law, but magnet schools charge additional tuition. Zhang (2009) provides details about the magnet school admission process.

[5]Students also take an exam in Chinese at the end of 6th grade. The Chinese exam includes a writing section. Students' Chinese test scores are largely non-comparable across schools as grading standards differ substantially across schools.

cannot be uniquely tracked due to multiple matches to common names. In addition, some students in the HSEE database have no matched primary school records, either because they attended a primary school outside the district or because their names were misspelled in one or both databases.[6] Second, we can only identify peer composition at the cohort (grade) level but not the classroom level. The classroom-level measures of peer composition may be more desirable if peer externalities take place mainly through classroom interactions. However, classroom-level measures are likely to be endogenous as school administrators and parents can have some discretion in placing students in different classes within a grade. Because of the potential sorting of students across classrooms within a grade, we would still use the cohort-level measures even if classroom-level measures were available.

Our sample consists of 7,435 students from three successive cohorts in the district whose academic histories are uniquely tracked. Students in our sample account for about 86 percent of the universe of 8,620 students who had completed middle school in the district during 2005-2007. Our data are ideal for analyzing peer effects in education for two reasons. First, we measure individual and peer abilities by lagged test scores, which are much more precise proxies than other individual and peer characteristics such as mother's education (McEwan, 2003) and number of books at home (Ammermueller and Pischke, 2009). Second, there is a large amount of reshuffling of peers during the transition from primary school to middle school and the longitudinal structure of the data allows us to distinguish between new peers and old peers. As we will discuss in further detail in Section 2.3.3, separating new peers from old peers is very important for the identification of peer effects when lagged test scores

[6] Misspelling is more likely to occur in the primary school information records as we have obtained these records in handwritten paper documents and coded them into an electronic database. The HSEE records are obtained in electronic format.

are used. For concerns about the sensitivity of our results to the inclusion of two magnet schools, we replicate all our analyses to a subsample of 5,191 students from neighborhood middle schools only. Results of this subsample remain qualitatively the same as the full sample. We therefore report only the results of the full sample in this chapter.

2.2.2 Descriptive Statistics

Table 2-1 presents descriptive statistics for the matched sample. Panel A shows statistics for three individual-level variables: gender, 6th-grade math score, and 9th-grade math score. Panel B reports four exogenous measures of peer composition in middle school: proportion of female peers, proportion of new peers, average 6th-grade peer math score, and inter-quartile range (IQR) of 6th-grade peer math scores. For the latter two measures of lagged test scores, Panel B also reports separate statistics for new peers only. Column 1 shows the means for these variables. The sample is balanced in gender, consisting of roughly 50 percent girls. On average, three-quarters of a student's peers in middle school are new peers from other primary schools. For ease of interpretation, we normalize student test scores by cohorts to have zero means and standard deviations of one. As some students are not included in our matched sample for reasons discussed above, any observed deviation of our sample mean test score from zero reflects selection into the matched sample. For instance, the average 6th-grade test score (0.035σ) in our sample is slightly higher than the district average (which is normalized to be zero). Our explanation of this difference is that a disproportionate share of students from high mobility families (e.g., rural migrants), who on average have lower academic achievement, opt out of the district's middle school system and are therefore not tracked in our matched sample. The IQR of lagged peer test scores in our sample is 1.04σ. For reference, the IQR of a standard normal distribution is 1.35σ, which is what we would expect to see had students been randomly assigned to middle schools. The observation of a smaller IQR of lagged peer test scores than the case of random assignment indicates student sorting in peer group formation.

Column 2 reports the standard deviations of these individual- and peer-level variables. As any between-school variation is removed in the school fixed-effect framework, our source for identification is variation across cohorts within the same school. Hence, column 3 reports a measure of the within-school dispersion of the individual- and the peer-level variables: the standard deviation of the residual of each variable after removing school and cohort fixed effects. Figures 2-1 to 2-3 plot the within-school variation in peer composition measured by gender mix, average 6th-grade test scores, and IQR of 6th-grade math scores, respectively. These figures show that there is a fair amount of cohort-to-cohort variation in peer composition within schools. For example, over the three consecutive cohorts observed in the sample, 11 out of a total of 15 schools have experienced a more than seven-percentage-point change in the proportion of female peers, 10 schools have experienced a more than 0.30σ change in the average 6th-grade math scores, and 11 schools have experienced a 0.20σ change in the IQR of 6th-grade math scores.

2.3 Empirical Strategy

2.3.1 The Model Framework

We start from a simple linear-in-means education production function of the following form:

$$Y_{ics} = \beta A_{ics} + \lambda \overline{A}_{(-i)cs} + \phi_s + \kappa_{cs} + \upsilon_{ics} \tag{2.1}$$

where Y_{ics} is a student outcome, such as a test score, for student i of cohort c in school s; A_{ics} is the exogenous predetermined ability of student i; $\overline{A}_{(-i)cs}$ is the average ability of student i's peers; ϕ_s represents the school-specific common-shock effects, arising from school-level unobserved common contextual, or environmental, influences that affect the outcomes of all students in that school; κ_{cs} represents variation in the common-shock effects across cohorts within schools and has a zero-mean within each school; and υ_{ics} is an individual-level stochastic error term that has a zero-mean within each cohort in each school. Note that the model is set up by assuming no cohort-to-cohort evolution in student ability A_{ics} or outcome Y_{ics}. In practice, such cohort-to-cohort evolution can be easily controlled by including a cohort fixed effect.

The identification of peer effects λ in equation (2.1) faces two major challenges. First, A_{ics} and $\overline{A}_{(-i)cs}$ are latent variables and cannot be observed directly. Second, the two common-shock effects ϕ_s and κ_{cs} reflect correlated effects and will give rise to a bias in the estimated peer coefficient $\widehat{\lambda}$ if they are correlated with $\overline{A}_{(-i)cs}$. Let us pretend for a moment that we have perfect measures of A_{ics} and $\overline{A}_{(-i)cs}$ and focus on the second challenge of isolating peer effects from correlated effects. Random assignment of students to groups, where a group refers to a cohort in a school in our context, can solve this problem because randomization breaks the potential link between peer composition ($\overline{A}_{(-i)cs}$) and the common shock effects (ϕ_s and κ_{cs}). However, true random assignment rarely exists outside experimental settings (Sacerdote, 2001; Zimmerman, 2003; Duflo et al., 2007, 2008). In practice, parents choose a school based on its quality and the composition of its peers, and schools also have some discretion in choosing students for admission. Hence, peer quality $\overline{A}_{(-i)cs}$ will be systematically correlated with common shock effects ϕ_s at the school level, causing the OLS estimator of λ to be biased. A possible way to account for such school-level correlated effects is to use within-school estimation that exploits variation in peer composition across adjacent cohorts within the same school.

As shown in Appendix A, the within-school specification of the education function can be written as

$$y_{ics} = \beta a_{ics} + \lambda \overline{a}_{(-i)cs} + \kappa_{cs} + v_{ics} \tag{2.2}$$

where y_{ics}, a_{ics}, and $\overline{a}_{(-i)cs}$ are derivations from their school means. The basic idea behind the within-school estimation is to examine whether, for students from adjacent cohorts in the same school, those who have more favorable peers (in terms of average peer ability) in their cohort score higher conditional on their own abilities. The identification assumption of the within-school estimation is that the within-school, between-cohort variation in peer composition $\overline{a}_{(-i)cs}$ is uncorrelated with the within-school, between-cohort variation in common-shock effects κ_{cs}. Under this identification assumption, the cohort-to-cohort variation in common-shock effects κ_{cs} can be subsumed into a general error term ϵ_{ics} such that

$\epsilon_{ics} = \kappa_{cs} + v_{ics}$. Consequently, the within-school model estimates the following equation:

$$y_{ics} = \beta a_{ics} + \lambda \overline{a}_{(-i)cs} + \epsilon_{ics} \tag{2.3}$$

where $cov(a_i, \epsilon_{ics}) = cov(\overline{a}_{(-i)cs}, \epsilon_{ics}) = 0$.

Although equation (2.3) is not confounded by correlated effects given the above assumption, it still cannot be estimated directly because the de-meaned ability measures a_i and $\overline{a}_{(-i)cs}$ are not directly observed. Lagged test scores are often used as proxies for latent abilities. Let x_{ics} denote the deviation of the observed lagged individual test score from its school mean, and w_{ics} denote the deviation of the observed average lagged peer test score from its school mean. Appendix A shows that x_{ics} and w_{ics} are related to a_i and $\overline{a}_{(-i)g}$ as follows:

$$x_{ics} = a_{ics} + v_{ics} \tag{2.4a}$$

$$w_{ics} = \overline{a}_{(-i)cs} + u_{ics} \tag{2.4b}$$

where v_{ics} is a stochastic error term that has a zero mean within each school and is uncorrelated with ϵ_{ics}, and $u_{ics} = \overline{v}_{(-i)cs}$. Substituting equations (2.4a) and (2.4b) into equation (2.3) yields

$$y_{ics} = \beta x_{ics} + \lambda w_{ics} + \psi_{ics} \tag{2.5}$$

where $\psi_{ics} = \epsilon_{ics} - \beta v_{ics} - \lambda u_{ics}$. Note that x_{ics} (w_{ics}) and ψ_{ics} are correlated because they both contain v_{ics} (u_{ics}).

Ammermueller and Pischke (2009) consider a similar within-school estimation of peer effects in the presence of measurement errors. While we rely on variation across cohorts within the same schools for identification, they use comparisons across classes within the same grade in the same school. As they do not have students' lagged test scores, they use parents' reports of number of books at home as a measure of peer composition. They argue that classes are formed roughly randomly in European primary schools and the measurement errors in books at homes are uncorrelated within classes. In our context, these assumptions would imply that the within-school, between-cohort variation in peer quality $\overline{a}_{(-i)cs}$ is idiosyncratic and not related to a_{ics} or ϵ_{ics}, and that the error terms v_{ics} and u_{ics} are uncorrelated. Under these assumptions, the within-school estimators $\widehat{\beta}_W$ and $\widehat{\lambda}_W$ converge to:

$$p \lim \widehat{\beta}_W = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_v^2}\beta \tag{2.6a}$$

$$p \lim \widehat{\lambda}_W = \frac{\sigma_{\overline{a}}^2}{\sigma_{\overline{a}}^2 + \sigma_u^2}\lambda \tag{2.6b}$$

where σ_a^2, σ_v^2, $\sigma_{\overline{a}}^2$, and σ_u^2 denote the variances of a_{ics}, v_{ics}, $\overline{a}_{(-i)cs}$, and u_{ics}. Although x_{ics} and w_{ics} are both correlated with the error term ψ_{ig} in equation (2.5), they are uncorrelated with each other under the above assumptions. Hence, $\widehat{\beta}_W$ and $\widehat{\lambda}_W$ are both subject to attenuation biases in the classical errors-in-variables (EIV) problem. Specifically, the plims of $\widehat{\beta}_W$ and $\widehat{\lambda}_W$ are regression coefficients in the following model:

$$y_{ics} = \widetilde{\beta}x_{ics} + \widetilde{\lambda}w_{ics} + \mu_{ics} \tag{2.7}$$

where $\widetilde{\beta} = \frac{\sigma_a^2}{\sigma_a^2+\sigma_v^2}\beta_0$, $\widetilde{\lambda} = \frac{\sigma_{\overline{a}}^2}{\sigma_{\overline{a}}^2+\sigma_u^2}\lambda$, and $\mu_{ics} = (\beta - \widetilde{\beta})x_{ics} + (\lambda - \widetilde{\lambda})w_{ics} + (\epsilon_{ics} - \beta v_{ics} - \lambda u_{ics})$.

Because of the unobservable nature of individual and peer abilities, the structural coefficients β and λ in the ability model equation (2.1) cannot be estimated directly. The regression coefficients $\widetilde{\beta}$ and $\widetilde{\lambda}$ in equation (2.7), however, are estimable and can be interpreted as important policy parameters of interest: the marginal effect of an individual's lagged test score and the marginal effect of the average lagged peer test score.

Ammermueller and Pischke (2009) argue that, if there exists another set of independent measures of the same individual and peer variables x'_{ics} and w'_{ics} (e.g., students' reports of number of books at home in addition to parents' reports), using x'_{ics} and w'_{ics} as instruments for x_{ics} and w_{ics} can correct the measurement error problem and provide consistent estimates of β and λ. However, we do not always have two measurements of the same variables, and, even if we do, the errors in the two measurements may well be correlated. Hence, their IV approach to correct the measurement error problem may not always be feasible. Despite the classical attenuation biases, the within-school estimators $\widehat{\beta}_W$ and $\widehat{\lambda}_W$ are still informative for at least two reasons. First, they provide consistent estimators of policy parameters $\widetilde{\beta}$ and $\widetilde{\lambda}$ as defined in the previous paragraph. Second, the attenuation biases decrease with the precision of the proxy ability measures. We expect students' lagged test scores used in our study are more precise measures of abilities than other indirect measures, such as mother's schooling and number of books at home, used in some previous studies.

2.3.2 The Problem of Transitory Common Shocks on Lagged Test Scores

Table 2-2 reports the OLS and the within-school estimations that regress students' 9th-grade math scores on their 6th-grade math scores and the average 6th-grade math scores of their peers. Each column corresponds to a separate regression and the standard errors reported in parentheses are adjusted for clustering within each cohort in each school. Column 1 reports the least square estimation that does not control for school fixed effects. The OLS estimator $\widehat{\lambda}_{OLS}$ (0.591 with a standard error of 0.104) shows a very strong positive relationship between one's 9th-grade test score and the average 6th-grade test score of one's peers in a cross-sectional setting. The large positive OLS estimator of peer coefficient, however, almost certainly confounds peer effects with "correlated effects" because of student sorting across schools based on the unobserved school characteristics. The fact that the estimated peer coefficient $\widehat{\lambda}_{OLS}$ (0.591) is even larger than the estimated coefficient of own lagged test scores $\widehat{\beta}_{OLS}$ (0.442) also implies that $\widehat{\lambda}_{OLS}$ is biased upward due to correlated effects and that the magnitude of the bias may be quite large. As we have discussed earlier, introducing school fixed effects to the model can mitigate the bias due to school-level correlated effects. Column 2 reports the results of the within-school estimation that includes both school and cohort fixed effects. The F-test of the joint significance of the school fixed effects has a p-value below 0.001, showing evidence of the existence of school-level correlated effects. Not surprisingly, $\widehat{\lambda}_W$ is reduced considerably in the within-school estimation. What is perhaps surprising is that $\widehat{\lambda}_W$ is now negative and significant (with a point estimate of -0.250 and a standard error of 0.112). Although the empirical literature has not reached a consensus on the existence and the magnitude of peer effects, the true peer coefficient λ is unlikely to be negative. Hence, we take the negative and significant point estimate of λ as evidence that our within-school estimator $\widehat{\lambda}_W$ is subject to a negative bias that cannot be simply explained by the attenuation bias.

Next, we revisit the assumptions used to derive the within-school estimators equation (2.6b) to examine the potential sources of such a negative bias. First, we assume that within-school, between-cohort variation in peer quality $\overline{a}_{(-i)cs}$ is unrelated to the de-meaned individual ability a_{ics}. This assumption implies that student sorting only occurs across schools but not across cohorts within the same school. This is plausible in our context as parents are unlikely to be well informed and sophisticated enough to condition their school choice decision on the cohort-to-cohort variation in peer quality within a school. Moreover, with classical measurement errors, within-school student sorting will introduce an upward bias in $\widehat{\lambda}_W$, opposite to what we have seen in the data. Specifically, if there exists within-school student sorting by ability, i.e., $\pi = cov(a_{ics}, \overline{a}_{(-i)cs}) > 0$, the within-school estimator $\widehat{\lambda}_W$ will converge to $\lambda - \frac{\sigma_u^2}{(\sigma_{\overline{a}}^2+\sigma_u^2)-\frac{\pi^2}{(\sigma_a^2+\sigma_v^2)}}\lambda + \frac{\sigma_v^2}{(\sigma_a^2+\sigma_v^2)}\frac{\pi}{(\sigma_{\overline{a}}^2+\sigma_\mu^2)-\frac{\pi^2}{(\sigma_a^2+\sigma_v^2)}}\beta$. The third component can be interpreted as a correlation bias, which arises from correlation between a_{ics} and $\overline{a}_{(-i)cs}$, and has the same sign as own ability effect β. Our second assumption is that, within the same school, the cohort-to-cohort variation in peer quality $\overline{a}_{(-i)cs}$ is uncorrelated with the cohort-to-cohort variation in common-shock effects κ_{cs}. A downward bias in $\widehat{\lambda}_W$ would arise if $\overline{a}_{(-i)cs}$ is instead negatively correlated with κ_{cs}. This would be the case if, when a cohort quality is relatively poor in a school, a principal who cares about within-school equity assigns high-quality teachers to that cohort to partly compensate for the poor student quality. However, the extent of such endogenous teacher assignment, if it exists at all, is likely to be quite limited as teachers usually rotate their grade assignment on a three-year basis (grades 7 to 9). Moreover, we would not expect a principal to manipulate teacher assignment to the extreme extent to more than fully compensate the difference in peer quality such that the net effect $(\lambda\overline{a}_{(-i)cs} + \kappa_{cs})$ is negatively correlated with peer quality $\overline{a}_{(-i)cs}$.

The third assumption to derive equation (2.6b) is that measurement errors in the individual- and the peer-level lagged test scores are uncorrelated, i.e., $\rho = cov(v_{ics}, u_{ics}) = 0$, a condition that would hold if the error terms of lagged individual test scores are i.i.d. within each cohort in each school. However, students usually take some former peers with them when moving to the next schooling phase. In our sample, about a quarter of a student's peers in middle school are her former peers from the same primary school. To the extent that the lagged test scores of students from the same primary school are subject to transitory common shocks, the presence of a student's former peers in her current peer group leads to a positive correlation between v_{ics} and u_{ics}, i.e., $\rho > 0$. In this chapter, we refer to transitory common shocks as group-specific contextual, or environmental, influences that have only transitory effects on students' observed outcomes, i.e., these influences affect the observed test scores of all students in a group, but not their permanent abilities.[7] As we have mentioned earlier, random overlapping between testing contents and teachers' instructions is one source of such transitory common shocks. Another source of such transitory common shocks is the across-school difference in grading standards, causing students' grades to be inflated in some schools but deflated in others. Note that random assignment of teachers to grading at the

[7] The literature usually uses the terminology "common shocks" to refer to common contextual factors, such as school resources and teacher quality, that affect students' test scores through their effects on students' abilities. However, the transitory common shocks we consider here have no (permanent) influences on students' abilities, but only transitory effects on test scores through their effects on the measurement errors.

school or class level cannot alleviate the second type of transitory common shocks, although random assignment of grading at the individual level will work. Moreover, there is no direct way to correct the first type of transitory common shocks.

When the imperfect individual and peer ability measures x_{ics} and w_{ics} are subject to transitory common shocks, the correlation between the error terms v_{ics} and u_{ics} will carry over to x_{ics} and w_{ics}. Hence, the standard attenuation formulation no longer applies. Appendix A shows that, in the presence of transitory common shocks in lagged test scores, the within-school estimator of peer coefficient $\widehat{\lambda}_W$ converges as follows:

$$p\lim(\widehat{\lambda}_W - \lambda) = -\frac{\sigma_u^2 - \frac{\rho^2}{\sigma_{\bar{a}}^2+\sigma_v^2}}{\sigma_{\bar{a}}^2 + \sigma_u^2 - \frac{\rho^2}{\sigma_{\bar{a}}^2+\sigma_v^2}}\lambda - \frac{\sigma_a^2}{\sigma_a^2 + \sigma_v^2}\frac{\rho}{\sigma_{\bar{a}}^2 + \sigma_u^2 - \frac{\rho^2}{\sigma_{\bar{a}}^2+\sigma_v^2}}\beta \tag{2.8}$$

Assuming that β and λ are both positive, the within-school estimator $\widehat{\lambda}_W$ underestimates λ as both bias components in equation (2.8) are negative (see proof in Appendix A). The first bias component, which we refer to as the "attenuation bias," is similar to the classical attenuation bias formula except that it has an additional adjustment component $(-\frac{\rho^2}{\sigma_{\bar{a}}^2+\sigma_v^2})$ in both the numerator and denominator to correct for the correlation between the v_{ics} and u_{ics}. Note that when $\rho > 0$, this attenuation bias is smaller in magnitude than the classical attenuation bias $(-\frac{\sigma_u^2}{\sigma_{\bar{a}}^2+\sigma_u^2}\lambda)$. The second component, which we refer to as the "transitory-common-shock bias," arises because the peer-level regressor w_{ics} is negatively correlated with ψ_{ics} $(= \epsilon_{ics} - \beta v_{ics} - \lambda u_{ics})$ through its correlations with both v_{ics} and u_{ics} in the presence of transitory common shocks. Specifically, $cov\left(w_{ics}, \psi_{ics}\right) = -\lambda\sigma_u^2$ when measurement errors v_{ics} and u_{ics} are independent, while $cov\left(w_{ics}, \psi_{ics}\right) = -\beta\rho - \lambda\sigma_u^2$ when v_{ics} and u_{ics} are correlated. The second bias component in equation (2.8) can dominate the true coefficient λ and reverse the sign of $\widehat{\lambda}_W$ when β is sufficiently large compared to λ. Hence, our explanation of the negative and significant within-school estimator $\widehat{\lambda}_W$ is that the lagged individual and peer test scores used in our estimation are subject to transitory common shocks because of the presence of former peers in one's current peer group.

2.3.3 The IV Approach

Our foremost concern about the within-school estimator $\widehat{\lambda}_W$ is that it is subject to a transitory-common-shock bias. The empirical evidence indicates that the transitory-common-shock bias dominates the within-school estimator $\widehat{\lambda}_W$ and reverses its sign. The source of this bias is the correlation between v_{ics} and u_{ics} arising from the presence in a student's current peer group of her former peers, whose lagged test scores are subject to transitory common shocks just like her own lagged test score. An idea for overcoming the transitory-common-shock bias is to use the average lagged test scores of a student's new peers ($w_{ics,new} = \overline{x}_{(-i)cs,new}$) as an instrument for the average lagged test score of all peers (w_{ics}), so as to consistently estimate an intermediate model similar to equation (2.7). The error term $u_{ics,new}$ in the peer-level regressor $w_{ics,new}$ is expected to be uncorrelated with the error term v_{ics} in the individual-level regressor x_{ics}, under the assumption that transitory shocks in lagged test scores are uncorrelated for students from different primary schools. However, there remains the question of what is the formulation of the peer coefficient λ^* we actually estimate in this IV approach and to what extent λ^* is informative regarding the true structural coefficient λ. This is shown in Proposition 1.

Proposition 1 *Let y_{ics} denote an outcome of interest for student i of cohort c in school s;*

let x_{ics} denote the lagged test score of student i, which is an imperfect measure of student i's latent ability a_{ics} such that $x_{ics} = a_{ics} + v_{ics}$, where v_{ics} is a stochastic individual error term; let w_{ics} denote the average lagged test score of student i's peers such that $w_{ics} = \overline{x}_{(-i)cs} = \overline{a}_{(-i)cs} + u_{ics}$, where $u_{ics} = \overline{v}_{(-i)cs}$; and let $w_{ics,new}$ denote the average lagged test score of student i's new peers such that $w_{ics,new} = \overline{x}_{(-i)cs,new} = \overline{a}_{(-i)cs,new} + u_{ics,new}$. Suppose the latent education production function takes the following linear-in-means form:

$$y_{ics} = \beta a_{ics} + \lambda \overline{a}_{(-i)cs} + \epsilon_{ics}$$

Assume all the covariances between a_{ics}, v_{ics}, $\overline{a}_{(-i)cs}$, u_{ics}, $\overline{a}_{(-i)cs,new}$, $u_{ics,new}$, and ϵ_{ics} are zero except for $cov(v_{ics}, u_{ics})$, which is denoted as ρ and is assumed to be positive. Then, using $w_{ics,new}$ as an instrument for w_{ics} can provide consistent IV estimators of the following intermediate model:

$$y_{ics} = \beta^* x_{ics} + \lambda^* w_{ics} + \varphi_{ics} \tag{2.9}$$

where $\beta^* = \frac{\sigma_a^2}{\sigma_a^2+\sigma_v^2}\beta - \frac{\rho}{\sigma_a^2+\sigma_v^2}\frac{\sigma^2_{\overline{a}_{new}}}{\sigma^2_{\overline{a}_{new}}+\sigma^2_{u_{new}}}\lambda$, $\lambda^* = \frac{\sigma^2_{\overline{a}_{new}}}{\sigma^2_{\overline{a}_{new}}+\sigma^2_{u_{new}}}\lambda$, and $\varphi_{ics} = (\beta - \beta^*)x_{ics} + (\lambda - \lambda^*)w_{ics} + (\epsilon_{ics} - \beta v_{ics} - \lambda u_{ics})$. In the preceding formulas, σ_a^2, σ_v^2, $\sigma^2_{\overline{a}_{new}}$, and $\sigma^2_{u_{new}}$ denote, respectively, the variances of a_{ics}, v_{ics}, $\overline{a}_{(-i)cs,new}$, and $u_{ics,new}$.

The proof of Proposition 1 is provided in Appendix B. As lagged test scores are imperfect measures of abilities, attenuation biases remain in both $\widehat{\beta}_{W,IV}$ and $\widehat{\lambda}_{W,IV}$. However, using $w_{ics,new}$ as an instrument for w_{ics} removes the transitory-common-shock bias in $\widehat{\lambda}_{W,IV}$ because the error term $u_{ics,new}$ in the instrument $w_{ics,new}$ is uncorrelated with the error term in lagged individual test score v_{ics}. The transitory-common-shock bias component remains in $\widehat{\beta}_{W,IV}$ as we do not have an instrument for individual lagged test scores. Despite the attenuation bias, $\widehat{\lambda}_{W,IV}$ is still informative because it is a consistent estimator of an interesting intermediate parameter $(\frac{\sigma^2_{\bar{a}_{new}}}{\sigma^2_{\bar{a}_{new}}+\sigma^2_{\bar{u}_{new}}}\lambda)$, which converges to the policy parameter of interest $\widetilde{\lambda}(=\frac{\sigma^2_{\bar{a}}}{\sigma^2_{\bar{a}}+\sigma^2_{\bar{u}}}\lambda)$ if the information-to-noise ratio in the lagged test scores of new peers $(\frac{\sigma^2_{\bar{a}_{new}}}{\sigma^2_{\bar{u}_{new}}})$ is the same as that in the lagged test scores of all peers $(\frac{\sigma^2_{\bar{a}}}{\sigma^2_{\bar{u}}})$.

2.4 Empirical Results

2.4.1 First-stage Results

Results in Table 2-2 show a possibly serious negative transitory-common-shock bias in the within-school estimator of peer coefficient $\widehat{\lambda}_W$. Our approach to correct this transitory-common-shock problem is to use the lagged test score measures of new peers to instrument for the corresponding measures of all peers. Specifically, we are interested in the causal effects of two lagged peer test score measures: the mean and the IQR. In order to overcome the transitory-common-shock problem, we instrument the average lagged peer test score using the average lagged test score of new peers, and instrument the IQR of lagged peer test scores with the IQR of lagged test scores of new peers. Table 2-3 shows the first-stage relationships. All specifications include as covariates the student's own lagged test score, the proportion of female peers, a female dummy, middle school dummies, and cohort dummies. Column 1 shows that the two average lagged test score measures are highly positively correlated: a one standard deviation increase in the average lagged test score of new peers is associated with 0.56 standard deviation increase in the average lagged peer test score. The first-stage coefficient estimate is less than the average proportion of new peers in our sample (75 percent), suggesting a negative association between the average lagged test score of old peers and that of new peers.[8] This observation indicates that the school district may have some equity concern in mind when determining the middle school zoning scheme every year. In such case, studies examining solely the reduced-form effect of new peers (e.g., Gibbons and Telhaj, 2006; Lavy, Silva, and Weinhardt, 2009) may underestimate the magnitude of peer effects. Column 3 estimates the first-stage relationship between two IQR measures: a one standard deviation increase in the IQR of lagged test scores of new peers is associated with a 0.30 standard deviation increase in the IQR of lagged test scores of all peers. Columns 2 and 4 include both instruments and estimate the first-stage models, respectively, for the average lagged peer test score and the IQR of lagged peer test scores. The first-stage coefficient of the relevant instrument (i.e., the average lagged test score of new peers in column 2 and the IQR of lagged test scores of new peers in column 4) remains virtually unchanged after the inclusion of the other instrument.

2.4.2 Basic Results on Peer Effects

Table 2-4 summarizes our basic results on peer effects from homogenous models. Columns 1-3 present results from reduced-form regressions and columns 4-6 report the corresponding IV results. Let us first consider the coefficients of two individual-level regressors: the female dummy and the lagged individual test score. The coefficients of both these individual-level regressors are consistently estimated across all the reduced-form and IV specifications. The

[8] Another way to examine such a negative association is to regress the lagged average test score of old peers on that of new peers and the same set of covariantes. The estimated coefficient of the lagged average test score of new peers in that regression is -0.166 (with a standard error of 0.022).

coefficient of the female dummy is statistically insignificant in all specifications, indicating no significant gender gap in 9th-grade math scores. The coefficient of lagged individual test score is highly significant and is estimated to be virtually the same (around 0.43) across all the six specifications. As shown in Proposition 1, $\widehat{\beta}_{W,IV}$ puts a lower bound of the structural coefficient of own ability effect β when the peer ability effect λ is nonnegative.

We now turn to the results on peer effects. We are interested in three measures of peer composition: peer gender mix, average lagged peer test score (a proxy measure for the average peer ability), and IQR of the lagged peer test scores (a proxy measure for the spread of the distribution of peer ability). Unlike some previous studies that find positive spillover effects of girls on math scores (Hoxby, 2000; Whitemore, 2003; Lavy and Schlosser, 2009), we find no evidence that peer gender composition has an impact on students' 9th-grade math scores. The coefficient of the proportion of female peers variable is insignificant in all the reduced-form and the IV specifications. Columns 1 reports the reduced-form effect of the average lagged test score of new peers. Once we replace average lagged achievement of a student's peers with the same measure of her new peers, the negative peer coefficient for the within-school estimator (-0.250 with a standard error of 0.112) disappears. Column 4 presents the corresponding IV estimator of the average lagged peer test score. Unfortunately, both the reduced-form (0.121 with a standard error of 0.087) and the IV coefficients (0.218 with a standard error of 0.175) are very imprecisely estimated. Although we cannot reject the null hypothesis of no linear-in-means peer effects, we also cannot reject very large peer effects. The imprecise reduced-form and IV estimators are likely to be because of the relatively small number of clusters (cohorts $\times$ schools) in our sample. Despite the imprecise results, the pattern of change from the within estimator to the reduced-form and IV estimators still shows strong evidence for the existence of a severe negative bias of using the average lagged peer test score measure when this measure is subject to transitory common shocks just like the student's own lagged test score.

We next examine the effect of peer group heterogeneity, measured by the IQR of lagged peer test scores, on student achievement. Columns 2 and 5, report, respectively, the reduced-form and the IV estimates of the effect of the IQR of lagged peer test scores. Both estimators are negative and significant at the five percent level, suggesting that students benefit from having homogeneous peers. The point estimate of the IV coefficient (-0.566 with a standard error of 0.327) indicates that a 0.2σ reduction in the IQR of 6th-grade peer test scores, a magnitude of change 11 out 15 schools in our sample had experienced, can increase a student's test score by 0.1σ. Column 3 presents the results of reduced-form estimation that includes both the mean and the IQR of lagged test scores of new peers. Column 6 shows the corresponding IV results that control for both the mean and the IQR of lagged peer test scores. The IV estimate of the coefficient of the IQR of lagged peer test scores (-0.571 and a standard error of 0.321) is insensitive to the inclusion of average peer test score. The point estimate of the IV coefficient of the average lagged peer test score (-0.007 with a standard error of 0.199), however, has been reduced substantially once we control for the IQR of lagged peer test scores. The reduction in the point estimate is due to a negative correlation between the residual average peer test score and the residual IQR of lagged test scores (after controlling for school fixed effects).

2.4.3 Allowing Heterogeneity for Peer Effects

Peer influences, however, may be heterogeneous and operate through the interaction between the distribution of peer ability and a student's own ability. For instance, some existing computational models of peer sorting in schools assumes that peer effects exhibit single crossing, i.e., an increase in average peer ability affects high-achieving students more than low-achieving students (Nechyba, 2006). In addition, several recent empirical studies find that students seem to benefit from having peers with similar characteristics as themselves, evidence in support of tracking (e.g., Hoxby and Weingarth, 2005; Duflo et al., 2008).

We explore these alternative models of heterogeneous peer effects in this subsection by interacting measures of lagged peer test scores – the mean and the IQR in particular – with student's own lagged test scores. To implement the estimation of these heterogeneous peer effects models in an IV framework, we instrument each interaction term between measures of lagged peer test scores and a student's lagged test score with the corresponding interaction term between measures of lagged test score of new peers and the student's lagged test score. Tables 2-5 and 2-6 present the first-stage and reduced-form results of these heterogeneous models. We focus our discussion in the text on the IV results reported in Table 2-7. Column 1 examines whether peer effects exhibit the single-crossing property. The IV coefficient of the interaction term between the average lagged peer test score and a student's own lagged test score (0.113 with a standard error of 0.052) is positive and significant at the five percent level, evidence in support of the single-crossing property. Column 2 examines whether a change in the dispersion of lagged peer test scores in a school affects students at the middle of the school's lagged test score distribution differently than those at the two tails. To do so, we interact the IQR of the distribution of lagged peer test scores with the absolute deviation of a student's own lagged test score from the school-cohort median, and instrument this interaction term with the corresponding interaction term using the IQR of her new peers. The IV coefficient of this interaction is positive and significant at the one percent level (0.093 with a standard error of 0.024), indicating that students in the middle of the lagged test score distribution benefit most from a contraction in the spread of the distribution of lagged peer test scores. Column 3 provides estimates of the full specification that includes both the mean and the IQR of lagged peer test scores as well as their interactions with students' own lagged test scores. Results of column 3 can summarize our findings. First, a rightward shift in the distribution of lagged peer scores benefits high-achieving students relative to low-achieving students, making the overall effect of average lagged test score insignificant. Second, a mean-preserving contraction in the distribution of lagged peer scores benefits all students, but to a greater extent for those in the middle of a school's lagged test score distribution. Both of these findings are in favor of ability tracking.

2.5 Conclusion

We provide empirical evidence on the existence and the structure of peer effects in middle school using a unique longitudinal data set from China. The peer effects literatue seems to be dominated by discussions on the relection problem and the selection issues, whereas little attention is being paid to the potential correlation in measurement errors between the individual- and the peer-level regressors, which we find important in our data. Such a correlation in measurement errors would arise if we simultaneously control for the lagged individual and peer test scores as the two measures are subject to transitory common shocks due to the continuing presence of former peers in a student's current peer group. An important contribution of this chapter is to clarify the impact of lagged transitory common shocks on estimates of peer effects. We derive formally that a positive correlation in measurement errors between the individual- and the peer-level regressors will lead to a negative bias in the estimate of peer coefficient, and provide empirical evidence that the transitory-common-shock problem is more than theoretical. We propose an empirical strategy to circumvent the transitory-common-shock problem by using the lagged test score measures of new peers, whose measurement error is uncorrelated with the measurement error in lagged individual test score, to instrument for the corresponding lagged test score measures of all peers.

Our main identification strategy uses within-school variation in peer composition across adjacent cohorts to control for student sorting across schools and the unobserved school characteristics that affect student outcomes. Our within-school IV estimate of the linear-in-means model shows little evidence that having peers of higher average lagged test score significantly improves a student's test score in math. The coefficients of the average lagged peer test scores, however, are not very precisely estimated. While we cannot reject the null hypothesis of no peer effects, we also cannot reject relatively large peer effects that have been found in the previous literature. Estimates of heterogeneous peer effects models show some evidence in favor of ability tracking for math learning. We find that a rightward shift in the distribution of lagged peer test scores benefits high-achieving students relative to low-achieving students, while a mean-preserving contraction in the distribution of lagged peer test scores benefits all students, but to a greater extent for those in the middle of a school's lagged test score distribution.

2.6 Appendices

2.6.1 Appendix A The Within-School Estimation

We are interested in the within-school estimation of equation (2.1) in the text

$$Y_{ics} = \beta A_{ics} + \lambda \overline{A}_{(-i)cs} + \phi_s + \kappa_{cs} + \upsilon_{ics} \tag{A1}$$

Taking average of equation (A1) for all students in school s yields

$$\overline{Y}_s = \beta \overline{A}_s + \lambda \overline{A}^*_s + \phi_s \tag{A2}$$

where $\overline{Y}_s = \frac{1}{n_s}\sum\limits_i Y_{ics}$, $\overline{A}_s = \frac{1}{n_s}\sum\limits_i A_{ics}$, $\overline{A}^*_s = \frac{1}{n_s}\sum\limits_i \frac{n_{cs}-1}{n_{cs}} A_{ics}$, and n_s and n_{cs} represent, respectively, the total number of students in school s and the total number of students in cohort c in school s. Here $\overline{A}^*_s$ differs from $\overline{A}_s$ because we use leave-out average peer ability in equation (A1). Note that κ_{cs} and υ_{ics} are not included in equation (A2) as they both have zero means at the school level. We can transform Y_{isc}, A_{isc}, and $\overline{A}_{(-i)cs}$ into derivations from their school means such that $y_{ics} = Y_{isc} - \overline{Y}_s$, $a_{ics} = A_{ics} - \overline{A}_s$, and $\overline{a}_{(-i)cs} = \overline{A}_{(-i)cs} - \overline{A}^*_s$. Substracting equation (A2) from equation (A1), the within-school specification of the education production function is

$$y_{ics} = \beta a_{ics} + \lambda \overline{a}_{(-i)cs} + \kappa_{cs} + \upsilon_{ics} \tag{A3}$$

Consider the following model generating the lagged test score (X_{ics}):

$$X_{ics} = A_{ics} + V_{ics}$$

where V_{ics} is a stochastic individual error term that is uncorrelated with A_{ics} and ϵ_{ics}. Let W_{ics} denote the average lagged test score of student $i's$ peers such that:

$$W_{ics} = \overline{X}_{(-i)cs} = \overline{A}_{(-i)cs} + U_{ics}$$

where $U_{ics} = \overline{V}_{(-i)cs}$. The within-school transformation of X_{ics} and W_{ics} can be written as follows

$$x_{ics} = X_{ics} - \overline{X}_s = (A_{ics} - \overline{A}_s) + (V_{ics} - \overline{V}_s) = a_{ics} + v_{ics} \quad \text{(A4a)}$$

$$w_{ics} = W_{ics} - \overline{W}_s = (\overline{A}_{(-i)cs} - \overline{A}^*_s) + (U_{ics} - \overline{U}_s) = a_{(-i)cs} + u_{ics} \quad \text{(A4b)}$$

Note that the above within-school transformation allows the possibility that $\overline{V}_s$ and $\overline{U}_s$ are nonzero. For example, if a middle school always draw students from a primary school that manipulate the test scores of its students by lowering the grading standards, $\overline{V}_s$ and $\overline{U}_s$ would both be positive. Such across-school variation in measurement errors, however, is accounted for in the within-school estimation. Substituting equations (A4a) and (A4b) into equation (A3) yields

$$y_{ics} = \beta x_{ics} + \lambda w_{ics} + \psi_{ics} \quad \text{(A5)}$$

where $\psi_{ics} = \epsilon_{ics} - \beta v_{ics} - \lambda u_{ics}$. We assume all covariances between a_{ics}, v_{ics}, $\overline{a}_{(-i)cs}$, u_{ics}, and

ϵ_{ics} are zero except for $cov(v_{ics}, u_{ics})$, which is denoted as ρ and is assumed to be nonnegative.

Let n denotes the total number of students in the sample, the plims of the variance and covariance terms are

$$\begin{aligned}
p\lim \sum \frac{(x-\overline{x})^2}{n} &= \sigma_a^2 + \sigma_v^2 \\
p\lim \sum \frac{(w-\overline{w})^2}{n} &= \sigma_{\overline{a}}^2 + \sigma_u^2 \\
p\lim \sum \frac{(w-\overline{w})(x-\overline{x})}{n} &= \rho \\
p\lim \sum \frac{(x-\overline{x})(y-\overline{y})}{n} &= \beta\sigma_a^2 \\
p\lim \sum \frac{(w-\overline{w})(y-\overline{y})}{n} &= \lambda\sigma_{\overline{a}}^2
\end{aligned}$$

The within-school estimator $\widehat{\beta}_W$ is

$$\widehat{\beta}_W = \frac{\sum(w-\overline{w})^2 \sum(x-\overline{x})(y-\overline{y}) - \sum(w-\overline{w})(x-\overline{x}) \sum(w-\overline{w})(y-\overline{y})}{\sum(x-\overline{x})^2 \sum(w-\overline{w})^2 - (\sum(w-\overline{w})(x-\overline{x}))^2} \tag{A6}$$

Taking the plim of (A6) and substituting the above plims of the variance and covariance terms yield,

$$\begin{aligned}
p\lim \widehat{\beta}_W &= \frac{(\sigma_{\overline{a}}^2 + \sigma_u^2)\beta\sigma_a^2 - \rho\lambda\sigma_{\overline{a}}^2}{(\sigma_a^2+\sigma_v^2)(\sigma_{\overline{a}}^2+\sigma_u^2) - \rho^2} \\
&= \beta - \frac{\sigma_v^2 - \frac{\rho^2}{(\sigma_{\overline{a}}^2+\sigma_u^2)}}{(\sigma_a^2+\sigma_v^2) - \frac{\rho^2}{(\sigma_{\overline{a}}^2+\sigma_u^2)}}\beta - \frac{\sigma_{\overline{a}}^2}{(\sigma_{\overline{a}}^2+\sigma_u^2)} \frac{\rho}{(\sigma_a^2+\sigma_v^2) - \frac{\rho^2}{(\sigma_{\overline{a}}^2+\sigma_u^2)}}\lambda
\end{aligned} \tag{A7}$$

By the same argument,

$$\begin{aligned}
p\lim \widehat{\lambda}_W &= p\lim \frac{\sum(x-\overline{x})^2 \sum(w-\overline{w})(y-\overline{y}) - \sum(w-\overline{w})(x-\overline{x}) \sum(x-\overline{x})(y-\overline{y})}{\sum(x-\overline{x})^2 \sum(w-\overline{w})^2 - (\sum(w-\overline{w})(x-\overline{x}))^2} \\
&= \frac{(\sigma_a^2+\sigma_v^2)\lambda\sigma_{\overline{a}}^2 - \rho\beta\sigma_a^2}{(\sigma_a^2+\sigma_v^2)(\sigma_{\overline{a}}^2+\sigma_u^2) - \rho^2} \\
&= \lambda - \frac{\sigma_u^2 - \frac{\rho^2}{(\sigma_a^2+\sigma_v^2)}}{(\sigma_{\overline{a}}^2+\sigma_u^2) - \frac{\rho^2}{(\sigma_a^2+\sigma_v^2)}}\lambda - \frac{\sigma_{\overline{a}}^2}{(\sigma_{\overline{a}}^2+\sigma_\mu^2)} \frac{\rho}{(\sigma_a^2+\sigma_v^2) - \frac{\rho^2}{(\sigma_{\overline{a}}^2+\sigma_\mu^2)}}\beta \qquad \text{(A8)}
\end{aligned}$$

As $\rho^2 = [cov(v,u)]^2 < var(v)var(u) = \sigma_v^2\sigma_\mu^2$, the last two terms in (A7) and (A8) are negative when $\rho > 0$. For the special case in which $\rho = 0$,

$$\begin{aligned}
p\lim \widehat{\beta}_W &= \beta - \frac{\sigma_v^2}{(\sigma_a^2+\sigma_v^2)}\beta \\
p\lim \widehat{\lambda}_W &= \lambda - \frac{\sigma_u^2}{(\sigma_{\overline{a}}^2+\sigma_u^2)}\lambda
\end{aligned}$$

2.6.2 Appendix B Proof of Proposition 1

We are interested in using $w_{ics,new}$ as an instrument for w_{ics} to estimate the following intermediate model of interest:

$$y_{ics} = \beta^* x_{ics} + \lambda^* w_{ics} + \phi_{ics} \qquad \text{(A9)}$$

where $\phi_{ics} = (\beta-\beta^*)x_{ics} + (\lambda-\lambda^*)w_{ics} + (\epsilon_{ics} - \beta v_{ics} - \lambda u_{ics})$. For $w_{ics,new}$ to be a valid instrument for w_{ics}, the formulations of β^* and λ^* need to satisfy the following two conditions:

$$cov(x_{ics}, \phi_{ics}) = 0 \tag{1A}$$

$$cov(w_{ics,new}, \phi_{ics}) = 0 \tag{1B}$$

Condition (1A) implies

$$\begin{aligned}
& cov(x_{ics}, \phi_{ics}) \\
= & (\beta - \beta^*)var(x_{ics}) + (\lambda - \lambda^*)cov(x_{ics}, w_{ics}) - \beta cov(x_{ics}, v_{ics}) - \lambda cov(x_{ics}, u_{ics}) \\
= & (\beta_0 - \beta^*)(\sigma_a^2 + \sigma_v^2) + (\lambda - \lambda^*)\rho - \beta\sigma_v^2 - \lambda_0\rho \\
= & \beta\sigma_a^2 - \beta^*(\sigma_a^2 + \sigma_v^2) - \lambda^*\rho \\
= & 0 \qquad \text{(1A')}
\end{aligned}$$

Condition (1B) implies

$$
\begin{aligned}
& cov(w_{ics,new}, \phi_{ics}) \\
= \ & (\beta - \beta^*)cov(w_{ics,new}, x_{ics}) + (\lambda - \lambda^*)cov(w_{ics,new}, w_{ics}) - \beta cov(w_{ics,new}, v_{ics}) - \lambda cov(w_{ig,new}, u_{ics}) \\
= \ & (\beta - \beta^*)0 + (\lambda - \lambda^*)(1-p)(\sigma^2_{\bar{a}_{new}} + \sigma^2_{u_{new}}) - \beta 0 - \lambda(1-p)\sigma^2_{u_{new}} \\
= \ & (1-p)[\lambda\sigma^2_{\bar{a}_{new}} - \lambda^*(\sigma^2_{\bar{a}_{new}} + \sigma^2_{u_{new}})] \\
= \ & 0 \qquad \text{(1B')}
\end{aligned}
$$

The formulations of β^* and λ^* that satisfies both (1A') and (1B') are:

$$
\begin{aligned}
\beta^* &= \frac{\sigma^2_a}{\sigma^2_a + \sigma^2_v}\beta - \frac{\rho}{\sigma^2_a + \sigma^2_v}\frac{\sigma^2_{\bar{a}_{new}}}{\sigma^2_{\bar{a}_{new}} + \sigma^2_{u_{new}}}\lambda \\
\lambda^* &= \frac{\sigma^2_{\bar{a}_{new}}}{\sigma^2_{\bar{a}_{new}} + \sigma^2_{u_{new}}}
\end{aligned}
$$

Therfore, using $w_{ics,new}$ as an instrument for w_{ics} provides consistent IV estimates of β^* and λ^* as defined above.

Bibliography

[1] Andreas Ammermueller and Jorn-Steffen Pischke. Peer effects in European primary schools: Evidence from the Progress in International Reading Literacy Study. *Journal of Labor Economics*, 27(3):315–348, 2009.

[2] Joshua D. Angrist and Kevin Lang. Does school integration generate peer effects? Evidence from Boston's Metco program. *American Economic Review*, 94(5):1613–1634, 2004.

[3] Julian R. Betts and Anne Danenberg. School accountability in california: An early evaluation. *Brookings Papers on Education Policy*, 2002:123–197, 2002.

[4] Julian R. Betts and Andrew Zau. Peer groups and academic achievement: Panel evidence from administrative data. Unpublished manuscript, University of California, San Diego, 2004.

[5] William A. Brock and Steven N. Durlauf. Interactions-based models. In James J. Heckman and Edward Leamer, editors, *Handbook of Econometrics*, volume 5, chapter 54, pages 3297–3367. Elsevier Science, Amsterdam, 2001.

[6] Katherine Carman and Lei Zhang. Classroom peer effects and academic achievement: Evidence from a Chinese middle school. Unpublished manuscript, Clemson University, 2008.

[7] Weili Ding and Steven F Lehrer. Do peers affect student achievement in China's secondary schools? *Review of Economics and Statistics*, 89(2):300–312, 2007.

[8] Esther Duflo, Pascaline Dupas, and Michael Kremer. Peer effects, pupil-teacher ratios, and teacher incentives: Evidence from a randomized evaluation in Kenya. unpublished manuscript, 2007.

[9] Esther Duflo, Pascaline Dupas, and Michael Kremer. Peer effects and the impact of tracking: Evidence from a randomized evaluation in Kenya. unpublished manuscript, 2008.

[10] Stephen Gibbons and Shqiponje Telhaj. Peer effects and pupil attainment: Evidence from secondary school transition. CEE Discussion Papers 0063, Centre for the Economics of Education, LSE, May 2006. LSE Center for the Economics of Education Discussion Papers No. 0063.

[11] Eric D. Gould, Victor Lavy, and M. Daniele Paserman. Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence. *The Economic Journal*, Forthcoming.

[12] Eric A. Hanushek, John F. Kain, Jacob M. Markman, and Steven G. Rivkin. Does peer ability affect student achievement? *Journal of Applied Econometrics*, 18:527–544, 2002.

[13] Caroline M. Hoxby. Peer effects in the classroom: Learning from gender and race variation. NBER Working Papers No. 7867, 2000.

[14] Caroline M. Hoxby and Gretchen Weingarth. Taking race out of the equation: School reassignment and the structure of peer effects. Unpublished manuscript, Harvard University, 2005.

[15] Thomas J. Kane and Douglas O. Staiger. The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, 16(4):91–114, 2002.

[16] Changhui Kang. Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics*, 61:458–495, 2007.

[17] Fang Lai. How do classroom peers affect student outcomes? Evidence from a natural experiment in Beijing's middle schools. Unpublished manuscript, New York University, 2007.

[18] Victor Lavy and Analia Schlosser. Mechanisms and impacts of gender peer effects at school. NBER Working Papers No. 13292, 2007.

[19] Victor Lavy, Olmo Silva, and Felix Weinhardt. The good, the bad and the average: Evidence on the scale and nature of ability peer effects in schools. Unpublished manuscript, Hebrew University of Jerusalem, 2009.

[20] Lars Lefgren. Educational peer effects and the Chicago public schools. *Journal of Urban Economics*, 56(2):169–191, 2004.

[21] Charles F. Manski. Identification of endogenous social effects: The reflection problem. *Journal of Economic Studies*, 14(3):115–136, 1993.

[22] Patrick J. McEwan. Peer effects on student achievement: Evidence from Chile. *Economics of Education Review*, 22:131–141, 2003.

[23] Robert A. Moffitt. Policy interventions, low-level equilibria, and social interactions. In Steven N. Durlauf and H. Peyton Young, editors, *Social Dynamics*, chapter 3, pages 45–82. Brookings Institution Press, Washington, D.C., 2001.

[24] Thomas J. Nechyba. Income and peer quality sorting in public and private schools. In Eric A. Hanushek and Finis Welch, editors, *Handbook of the Economics of Education*, volume 2, chapter 22, pages 1327–1368. Elsevier Science, Amsterdam, 2006.

[25] Bruce Sacerdote. Peer effects with random assignment: Results from dartmouth roomates. *Quarterly Journal of Economics*, 116(2):681–704, 2001.

[26] Jacob L. Vigdor and Thomas Nechyba. Peer effects in elementary school: Learning from "apparent" random assignment. Unpublished manuscript, Duke University, 2004.

[27] Jacob L. Vigdor and Thomas Nechyba. Peer effects in North Carolina public schools. In Ludger Woessmann and Paul E. Peterson, editors, *Schools and the Equal Opportunity Problem*. MIT Press, Cambridge, MA, 2006.

[28] Diane Whitemore. Resource and peer impact on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review Papersand Proceedings*, 95(2):199–203, 2005.

[29] Jeffrey E. Zabel. The impact of peer effects on student outcomes in New York City public schools. *Education Finance Policy*, 3:197–249, 2008.

[30] Hongliang Zhang. Magnet schools and student achievement: Evidence from a randomized natural experiment in China. Unpublished manuscript, Massachusetts Institute of Technology, 2009.

[31] David Zimmerman. Peer effects in academic outcomes: Evidence from a natural experiment. *Journal of Urban Economics*, 85(1):9–23, 2003.

Table 2-1 Descriptive Statistics

	Mean	s.d.	within-school s.d.
Panel A: Individual characteristics:			
Female	0.497	0.500	0.500
6th-grade math score	0.035	0.927	0.865
9th-grade math score	-0.005	1.002	0.914
Panel B: Peer-group characteristics			
All peers:			
Proportion of female peers	0.497	0.036	0.029
Proportion of new peers	0.749	0.281	0.151
Average 6th-grade peer math score	0.035	0.363	0.140
Inter-quartile range of 6th-grade peer math scores	1.043	0.233	0.118
New peers:			
Average 6th-grade math score	0.010	0.383	0.183
Inter-quartile range of math scores	1.065	0.292	0.210
Number of observations		7,435	

Table 2-2 OLS and Within-School Estimation

	Dependent variable: 9th-grade math score	
	OLS	Within
	(1)	(2)
Own lagged test score	0.442***	0.436***
	(0.019)	(0.019)
Average lagged peer test score	0.591***	-0.250***
	(0.104)	(0.112)
Middle school fixed effects	no	yes
Number of observations	7,435	

Notes: All specifications control for a female dummy, the proportion of female peers, and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported in parentheses. A triple asterisk (***) denotes significant at the 1 percent level.

Table 2-3 The First-Stage Effects

	Dependent Variables			
	Average lagged peer test score		IQR of lagged peer test scores	
	(1)	(2)	(3)	(4)
Average lagged test score	0.555***	0.546***		-0.058
of new peers	(0.071)	(0.068)		(0.078)
IQR of lagged test scores		-0.015	0.297***	0.271***
of new peers		(0.042)	(0.060)	(0.058)

Notes: All specifications control for own lagged test score, a female dummy, the proportion of female peers, middle school fixed effects and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported in parentheses. A triple asterisk (***) denotes significant at the 1 percent level.

Table 2-4 Basic Results of Within-school IV Estimates of Peer Effects

	Reduced-form			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.020 (0.023)	-0.020 (0.022)	-0.020 (0.023)	-0.019 (0.022)	-0.025 (0.024)	-0.025 (0.023)
Proportion of female peers	-0.182 (0.614)	0.330 (0.592)	0.313 (0.587)	0.281 (0.696)	-0.747 (0.839)	-0.759 (0.869)
Own lagged test score	0.433*** (0.019)	0.432*** (0.019)	0.432*** (0.019)	0.427*** (0.020)	0.426*** (0.020)	0.426*** (0.020)
Average lagged peer test score				0.218 (0.175)		
Average lagged test score of new peers	0.121 (0.087)		0.030 (0.087)			-0..007 (0199)
IQR of lagged peer test scores					-0..566* (0.327)	-0..571* (0.321)
IQR of lagged test scores of new peers		-0.168** (0.080)	-0.155* (0.077)			

Notes: All specifications control for middle school fixed effects and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported ·arentheses. A triple asterisk (***) denotes significant at the 1 percent level. A single asterisk (*) denotes significant at the 10 percent level.

Table 2-5 The First-stage Results of Heterogeneous Peer Effects Models

	Model 1	Model 2	Model 3
Average lagged test score of new peers	0.550***		0.541***
	(0.072)		(0.067)
Average lagged test score of new peers * lagged individual test score	0.821***		0.821***
	(0.083)		(0.083)
IQR of lagged test scores of new peers		0.291***	0.265***
		(0.060)	(0.056)
IQR of lagged test scores of new peers * Deviation of own lagged test score from the school-cohort median		0.955***	0.953***
		(0.032)	(0.031)

Notes: Each cell of the table reports the coefficient of regressing the corresponding instrumented variable on the instrument. All specifications control for the set of other instruments used in the model, own lagged test score, a female dummy, the proportion of female peers, middle school fixed effects, and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported in parentheses. A triple asterisk (***) denotes significant at the 1 percent level.

Table 2-6 The Reduced-form Results of Heterogeneous Peer Effects Models

	Dependent variable: 9th-grade math scores		
	(1)	(2)	(3)
Average lagged test score of new peers	0.140		0.052
	(0.090)		(0.090)
Average lagged test score of new peers * Own lagged test score	0.087**		0.091*
	(0.041)		(0.042)
IQR of lagged test scores of new peers		-0.221*	-0.206***
		(0.080)	(0.056)
IQR of lagged test scores of new peers * Deviation of own lagged test score from the school-cohort median		0.084***	0.086***
		(0.023)	(0.022)

Notes: All specifications control for own lagged test score, a female dummy, the proportion of female peers, middle school fixed effects, and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported in parentheses. A triple asterisk (***) denotes significant at the 1 percent level. A double asterisk (**) denotes significant at the 5 percent level. A single asterisk (*) denotes significant at the 10 percent level.

Table 2-7 Results of Within-school IV Estimates of Heterogeneous Peer Effects

	Dependent variable: 9th-grade math scores		
	(1)	(2)	(3)
Average lagged peer test score	0.246		0.027
	(0.180)		(0.208)
Average lagged peer test score * Own lagged test score	0.113**		0.097*
	(0.052)		(0.057)
IQR of lagged peer test scores		-0.613*	-0.585*
		(0.327)	(0.330)
IQR of lagged peer test scores * Deviation of own lagged test score from the school-cohort median		0.093***	0.094***
		(0.024)	(0.024)

Notes: All specifications control for own lagged test score, a female dummy, the proportion of female peers, middle school fixed effects, and cohort fixed effects. Robust standard errors, adjusted for within-school-cohort clustering, are reported in parentheses. A triple asterisk (***) denotes significant at the 1 percent level. A double asterisk (**) denotes significant at the 5 percent level. A single asterisk (*) denotes significant at the 10 percent level.

Figure 2-1 Within-school Variation in Peer Gender Mix

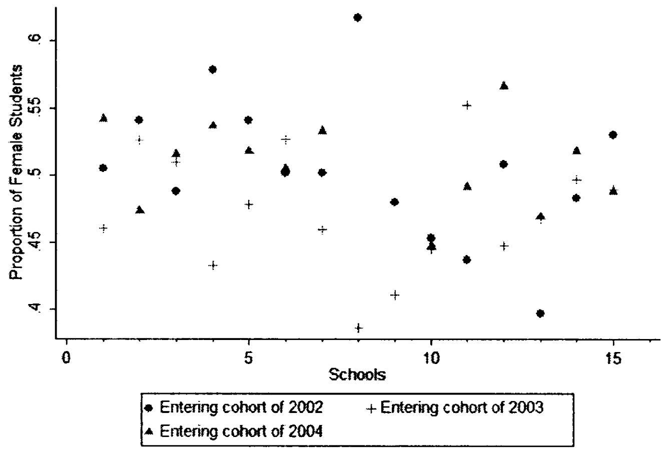


Figure 2-2 Within-school Variation in the Average 6th-Grade Peer Test Score

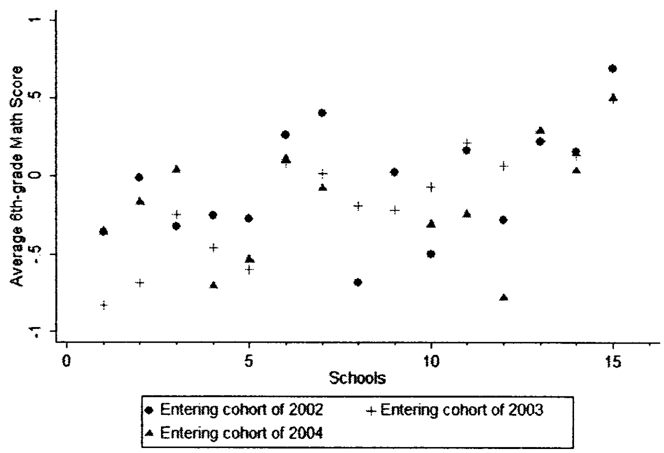
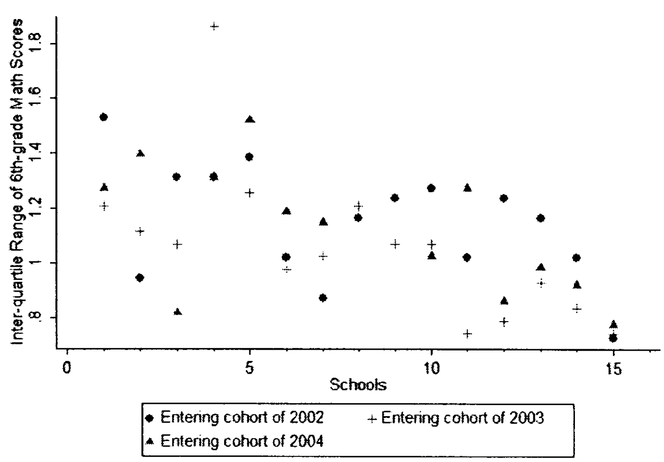


Figure 2-3 Within-school Variation in the Inter-Quartile Range of 6th-Grade Peer Test Scores



Chapter 3

Nonparametric Bounds on Treatment Effects with Missing and Mismatched Outcome Data

3.1 Introduction

This chapter concerns the nonparametric analysis of randomized experiments with observational problems that prevent researchers from collecting complete and error-free outcome data. Evaluation of a randomized experiment requires tracking the post-treatment outcomes of the program participants. When outcome data come from a separate source (e.g., an administrative file), it usually requires linking the program records to records in the outcome database. The problem of missing outcomes arises if there is incomplete overlapping of common units between the program database and the outcome database (see Ridder and Moffitt (2007) for a survey). In addition, often common variables besides the identifier are used to match records, e.g., we match records by name, gender, and age. Even if all program participants are observed in the outcome database, matching based on common variables that do not constitute a unique identifier in the outcome database may lead to imperfect matching, i.e., some program participants are matched to more than one record in the outcome database, leading to data contamination in the matched sample.

The existing literature on this subject is concerned mainly with the former problem of missing outcomes. The population parameters of interest are in general not identified in the presence of missing outcomes unless one makes untestable assumptions about the distribution of missing data (e.g., missing at random). Horowitz and Manski (2000) provide a general framework for constructing bounds for population treatment effect parameters when outcome and covariate data are missing nonrandomly. The idea is to construct "worst-case scenario" bounds of the treatment effect by imputing the missing data to either the largest or the smallest possible values of the outcome. However, their approach is applicable only if the outcome variable has a bounded support. Lee (2002, 2009) proposes a trimming procedure for bounding average treatment effects for a specific subpopulation whose outcomes would be observed irrespective of treatment assignment status, under the assumption of monotone treatment selection (MTS), i.e. treatment assignment only affects sample selection in "one direction". By restricting interest to this subpopulation whose outcomes would always be observed, Lee's trimming procedure no longer requires a bounded support of the outcome.

In this chapter, I extend Lee's trimming strategy to construct nonparametric bounds on average treatment effects when sample attrition and imperfect matching both exist and interact with each other. Section 3.2 shows that average treatment effects can still be identified if imperfect matching is the only problem. This is because data errors due to multiple matching are statistically independent of treatment assignment status conditional on the common variables, under random assignment and complete observation of program participants. Throughout this chapter, we will refer to "treatment effect" as the effect of being assigned treatment status, or "intention-to-treat effect", to bypass the problem of imperfect compliance with treatment assignment (Imbens and Angrist, 1994; Angrist, Imbens, and Rubin, 1996). Section 3.3 replicates Lee's trimming procedure, and shows that sharp bounds on average treatment effects for a subpopulation can be connstructed if matching is based on a unique identifier (i.e., no imperfect matching) and sample selection is monotone to treatment assignment status.

While Sections 3.2 and 3.3 separately consider each of these two observational problems, Section 3.4 considers the presence of both problems. The interaction between sample attrition and imperfect matching leads to another challenge for analyzing treatment effects: mismatching, which arises when a program participant of interest drops out of the outcome database but is matched to one or more erroneous records in the outcome database based on the common variables. The difference between imperfect matching (or multiple matching) and mismatching is that the true record of interest is included in the matched set in the former case but not in the latter case. Since the true outcome of the program participant is not included in the matched set, mismatching leads to corrupted data in the matched sample. When the observability of a program participant in the outcome database depends on treatment assignment, data errors due to mismatching are not independent of treatment assignment status. We show that average treatment effects cannot be bounded in general in the presence of mismatching. However, if prior knowledge or a consistent estimator of the mismatching probability is available, we can exploit this piece of information to construct bounds for average treatment effects under the MTS assumption.

3.2 Identification under Imperfect Matching

Let y denote the outcome of interest, d denote the randomly assigned binary treatment status, and z denote the common variable(s) used to match program participants in the outcome database. Define the random variables $(y^1, y^0, \widetilde{y}_k, d, z)$ such that y^1 denotes the potential outcome of a participant when $d = 1$, y^0 denotes the potential outcome of a participant when $d = 0$, and $\widetilde{y}_k$ denotes the mean outcome of the k records in the outcome database that are **incorrectly** matched to the participant based on z. *Throughout this chapter, we assume that both the number of incorrect matches k and the mean outcome of these incorrect matches $\widetilde{y}_k$ are independent of d being conditional on z.* This assumption is expected to hold under random assignment of treatment status.

For each program participant, the realized outcome is $y^d = y^1 d + y^0(1-d)$. In this section, we assume the outcome database is a population register in which all program participants are observed, i.e., no sample attrition. Let $\overline{y}^d_{k+1}$ denote the mean outcome of the $k + 1$ records matched to a program participant based on z. $\overline{y}^d_{k+1}$ can be linked to y^d and $\widetilde{y}_k$ by the following equation:

$$\overline{y}^d_{k+1} = \frac{1}{k+1} y^d + \frac{k}{k+1} \widetilde{y}_k \ \text{ for } k \geq 0, d = \{0, 1\}$$

Hence, the average population treatment effect can be identified as the difference in the weighted average outcomes by treatment assignment status as follows:

$$E[y^1 - y^0] = (k+1)(E[\overline{y}^1_{k+1}] - E[\overline{y}^0_{k+1}]) \tag{3.1}$$

3.3 Trimming with Missing Outcomes

This section replicates the results from Lee's trimming procedure which allows missing outcomes but assumes perfect matching in the outcome database based on an identifier. Let a denote an indicator of whether the outcome is observed or not. Let a^1 and a^0 be two latent variables denoting whether the outcome would be observed when $d = 1$ and $d = 0$, respectively. Note that y^d is observed if $a^d = 1$ and y^d is missing if $a^d = 0$. Furthermore, it is assumed that assignment to d, if it affects a at all, can affect a in only "one direction." Without loss of generality, we assume that $\Pr[a^0 = 1, a^1 = 0] = 0$. This monotonicity assumption excludes the possibility that some individuals are induced to drop out of the sample because of the treatment, but still allows some individuals who otherwise would drop out of the sample to be induced to stay in the sample because of the treatment.

Lee (2002) shows that the difference between the means of the observed outcomes of the treatment and the control groups is:

$$\begin{aligned} & E[y^1|a^1 = 1] - E[y^0|a^0 = 1] \\ = \; & \frac{\Pr(a^0 = 1)}{\Pr(a^1 = 1)} E[y^1|a^0 = 1] + \frac{\Pr(a^0 = 0, a^1 = 1)}{\Pr(a^1 = 1)} E[y^1|a^0 = 0, a^1 = 1] - E[y^0|a^0 = 1] \end{aligned} \tag{3.2}$$

In general, equation (3.2) will be a biased estimator for a particular parameter of interest: $E[y^1 - y^0|a^0 = 1]$, the average treatment effect for the subpopulation whose outcomes will be observed irrespective of treatment status. However, the observed data can yield lower and upper bounds $\underline{E}$ and $\overline{E}$ such that $\underline{E} \leq E[y^1|a^0 = 1] \leq \overline{E}$. As the matching rates for the treatment group ($\Pr(a^1 = 1)$) and the control group ($\Pr(a^0 = 1)$) can be identified in the observed data, the proportion of the selected treatment group that is induced to have non-missing outcome because of assignment to treatment can be identified as:

$$p = \frac{\Pr(a^0 = 0, a^1 = 1)}{\Pr(a^1 = 1)} = \frac{\Pr(a^1 = 1) - \Pr(a^0 = 1)}{\Pr(a^1 = 1)} \tag{3.3}$$

Therefore, the lower and upper bounds for the average outcome of the $a^0 = 1$ subpopulation can be identified as:

$$\underline{E} = \frac{1}{1-p}\int_{-\infty}^{F^{-1}(1-p)} y^1 f(y^1)dy^1 \leq E[y^1|a^0 = 1] \leq \frac{1}{1-p}\int_{F^{-1}(p)}^{\infty} y^1 f(y^1)dy^1 = \overline{E} \tag{3.4}$$

where $\underline{E}$ is constructed by excluding the top p percentile from the observed treatment group and $\overline{E}$ is constructed by excluding the bottom p percentile.

Given equation (3.4), the lower and upper bounds for the average treatment effect of the $a^0 = 1$ subpopulation is:

$$\underline{E} - E[y^0|a^0 = 1] \leq E[y^1 - y^0|a^0 = 1] \leq \overline{E} - E[y^0|a^0 = 1] \tag{3.5}$$

3.4 Bounding Average Treatment Effects with Missing and Mismatched Outcomes

3.4.1 Sample Attrition is Independent of Treatment Status

In this subsection, we consider the case that sample attrition exists but is independent of the treatment status:

Assumption 1A: Independence Assumption

$$\Pr(a^0 = a^1) = 1$$

Under Assumption 1A, we can suppress the superscript of a. Let m denote the number of records incorrectly matched based on z and let n be the number of total matches to z. Then, by construction, we must have $n = m + a$. The following statements summarize the matching outcomes and how they are related to sample attrition and imperfect matching:

(i) Missing outcome: $(a = 0, m = 0)$

(ii) Mismatching: $(a = 0, m \geq 1)$;

(iii) Unique and correct matching: $(a = 1, m = 0)$;

(iv) Imperfect matching: $(a = 1, m \geq 1)$.

Let p_k denote the probability that a program participant has k incorrectly matched records based on z, $\Pr(m = k)$, and let θ_k denote the subgroup-specific sample attrition rate for participants with k incorrect matches, $\Pr(a = 0|m = k)$. Define $q_k = \Pr(n = k)$, the proportion of program participants with k matched records in the outcome database such that $\sum_{k=0}^{\infty} q_k = 1$. The composition of q_k can be written as the following:

The unmatched sample: $q_0 = \Pr(n=0) = \Pr(a=0, m=0) = \theta_0 p_0$

The matched sample: $q_{k+1} = \Pr(n=k+1) = \Pr(a=1, m=k) + \Pr(a=0, m=k+1)$

$$= (1-\theta_k)p_k + \theta_{k+1}p_{k+1}, \quad \forall\, k \geq 0$$

Following Section 3.2, the observed mean outcome of program participants with $k+1$ matches can be decomposed as:

$$\begin{aligned}
& E[\overline{y}^d_{k+1} | n = k+1] = E[\overline{y}^d_{k+1} | (a=1, m=k) \text{ or } (a=0, m=k+1)] \\
= \ & \frac{(1-\theta_k)p_k}{q_{k+1}} \frac{1}{k+1} E[y^d | a=1, m=k] + \frac{(1-\theta_k)p_k}{q_{k+1}} \frac{k}{k+1} E[\widetilde{y}_k | m=k] \\
+ \ & \frac{\theta_{k+1}p_{k+1}}{q_{k+1}} E[\widetilde{y}_{k+1} | m = k+1], \quad \forall\, k \geq 0, d = \{0, 1\} \qquad (3.7)
\end{aligned}$$

In equation (3.7), the first component is the outcome of interest, the second component is data contamination due to imperfect matching, while the third component is data corruption due to mismatching (i.e., the presence of both sample attrition and imperfect matching). Since both $\widetilde{y}_k$ and $\widetilde{y}_{k+1}$ are independent of d, the last two terms in equation (3.7) can be canceled out when we compare the observed difference between the mean outcomes of the treatment and control groups with $k+1$ total matches:

$$E[\overline{y}^1_{k+1}|n=k+1]-E[\overline{y}^0_{k+1}|n=k+1]$$
$$=\frac{(1-\theta_k)p_k}{q_{k+1}}\frac{1}{k+1}E[y^1-y^0|a=1,m=k]$$

Therefore, a particular parameter of interest, the difference in the mean potential outcomes for a subpopulation whose outcomes are observed together with k mismatches ($E[y^1-y^0|a=1,m=k]$), can be written as:

$$E[y^1-y^0|a=1,m=k]=\frac{q_{k+1}}{(1-\theta_k)p_k}(k+1)(E[\overline{y}^1_{k+1}|n=k+1]-E[\overline{y}^0_{k+1}|n=k+1]) \quad (3.8)$$

Note that q_{k+1}, $k+1$, and $E[\overline{y}^1_{k+1}|n=k+1]-E[\overline{y}^0_{k+1}|n=k+1]$ are all observable. The parameter of interest, $E[y^1-y^0|a=1,m=k]$, can be bounded if the denominator $(1-\theta_k)p_k$ is bounded. Define $M_k=\sum_{l=0}^{k}p_l$ and $Q_k=\sum_{l=0}^{k}q_l$. Q_k and M_k can be linked by the following relationship:

$$Q_k=\sum_{l=0}^{k}q_l=\sum_{l=0}^{k-1}p_l+\theta_k p_k=\sum_{l=0}^{k}p_l-(1-\theta_k)p_k=M_k-(1-\theta_k)p_k$$

i.e.,

$$(1-\theta_k)p_k=M_k-Q_k$$

Note that Q_k is observable but M_k is not. Bounding $(1-\theta_k)p_k$ requires prior knowledge about the mismatching probability based on z in the outcome database. Often, the structure of the common variables z could allow us to construct such prior knowledge. For example, if we match students' test scores by name and gender, the match is usually conducted within each cohort. But if we match name and gender of one cohort to the outcomes of another cohort, we can get a consistent estimator of the mismatching probability q_k because all matches from a different cohort are erroneous matches.

Assume that prior knowledge, or a consistent estimator of the bound of M_k – $[\underline{M_k}, \overline{M^k}]$ – is available, $(1-\theta_k)p_k$ can be bounded by $[\underline{M_k} - Q_k, \overline{M^k} - Q_k]$. Therefore, the parameter of interest, $E[y^1 - y^0|a = 1, m = k]$, is bounded by:

$$q_{k+1}(k+1)(E[\overline{y}^1_{k+1}|n = k+1] - E[\overline{y}^0_{k+1}|n = k+1])[\frac{1}{\overline{M^k} - Q_k}, \frac{1}{\underline{M_k} - Q_k}]$$

I define the final parameter of interest as the difference in the mean potential outcomes of all program participants whose outcomes are observed (possibly with imperfect matching): $E[y^1 - y^0|a = 1]$.

$$
\begin{aligned}
& E[y^1 - y^0 | a = 1] \\
= & \frac{\sum\limits_{k=0}^{\infty} \Pr(a = 1, m = k) E[y^1 - y^0 | a = 1, m = k]}{\Pr(a = 1)} \\
= & \frac{\sum\limits_{k=0}^{\infty} \Pr(a = 1, m = k) E[y^1 - y^0 | a = 1, m = k]}{\sum\limits_{k=0}^{\infty} \Pr(a = 1, m = k)} \\
= & \frac{\sum\limits_{k=0}^{\infty} q_{k+1}(k+1)(E[\overline{y}_{k+1}^1 | n = k+1] - E[\overline{y}_{k+1}^0 | n = k+1])}{\sum\limits_{k=0}^{\infty} (1 - \theta_k) p_k} \qquad (3.9)
\end{aligned}
$$

Note that q_{k+1}, $k+1$, and $E[\overline{y}_{k+1}^1|n = k+1] - E[\overline{y}_{k+1}^0|n = k+1]$ can all be observed directly from the data. With $\sum\limits_{k=0}^{\infty}(1-\theta_k)p_k$ bounded by $[\sum\limits_{k=0}^{\infty}(\underline{M_k}-Q_k), \sum\limits_{k=0}^{\infty}(\overline{M^k}-Q_k)]$, $E[y^1-y^0|a = 1]$ can be bounded as well.

It is important to note here that, if both the actual treatment status (instead of the treatment assignment status) and the outcomes are either observed or unobserved simultaneously from the same data source, the Wald estimate of the local average treatment effect (LATE) would be unbiased because both the first-stage and the reduced-form effects are subject to the same degree of bias. Letting s denotes the actual treatment status, the Wald estimator of LATE is identified as the following:

$$E[y^1 - y^0|S^0 = 0, S^1 = 1, a = 1] = \frac{E[y^1 - y^0|a = 1]}{E[S^1 - S^0|a = 1]}$$
$$= \frac{\sum_{k=0}^{\infty} q_{k+1}(k+1)(E[\overline{y}^1_{k+1}|n = k+1] - E[\overline{y}^0_{k+1}|n = k+1])}{\sum_{k=0}^{\infty} q_{k+1}(k+1)(E[\overline{s}^1_{k+1}|n = k+1] - E[\overline{s}^0_{k+1}|n = k+1])} \tag{3.10}$$

3.4.2 Monotone Sample Attrition by Treatment Status

In this subsection, we consider the case that sample attrition exists and is monotone by treatment status. Without loss of generality, we assume that assignment to treatment only induces some individuals who otherwise would drop out of the sample to stay in the sample, but not the opposite.

Assumption 1B: Monotonicity Assumption

$$\Pr(a^0 = 1, a^1 = 0) = 0$$

Let $\theta_k = \Pr(a^0 = 0, a^1 = 0|m = k) = \Pr(a^1 = 0|m = k)$, the proportion of program participants with k erroneous matches in the outcome database who would drop out of the sample irrespective of treatment status, and $\psi_k = \Pr(a^0 = 0, a^1 = 1|m = k)$, the proportion of program participants with k erroneous matches in the outcome database who would be induced to stay in the sample because of the treatment. The composition of q_k, the proportion of program participants with k matched records in the outcome database, can be written as follows.

$$\text{Treatment group with } k \text{ matches} \quad : \quad q_k^1 = \Pr(n^1 = k) = \Pr(a^1 = 1, m = k-1) + \Pr(a^1 = 0, m = k)$$

$$= \quad (1-\theta_{k-1})p_{k-1} + \theta_k p_k, \quad \forall\ k \geq 1$$

$$\text{Control group with } k \text{ matches} \qquad q_k^0 = \Pr(n^0 = k) = \Pr(a^0 = 1, m = k-1) + \Pr(a^0 = 0, m = k)$$

$$= \quad (1-\theta_{k-1} - \psi_{k-1})p_{k-1} + (\theta_k + \psi_k)p_k, \quad \forall\ k \geq 1$$

The observed mean outcomes of the treatment and control groups with $k+1$ total matches are, respectively,

$$\begin{aligned}
& E[\overline{y}_{k+1}^1 | n = k+1] = E[\overline{y}_{k+1}^1 | (a^1 = 1, m = k) \text{ or } (a^1 = 0, m = k+1)] \\
= \; & \frac{(1-\theta_k)p_k}{q_{k+1}} E[\overline{y}_{k+1}^1 | a^1 = 1, m = k] + \frac{\theta_{k+1}p_{k+1}}{q_{k+1}} E[\widetilde{y}_{k+1} | a^1 = 0, m = k+1] \\
= \; & \frac{(1-\theta_k - \psi_k)p_k}{q_{k+1}} E[\overline{y}_{k+1}^1 | a^0 = 1, m = k] + \frac{\psi_k p_k}{q_{k+1}} E[\overline{y}_{k+1}^1 | a^0 = 0, a^1 = 1, m = k] + \\
& \frac{\theta_{k+1}p_{k+1}}{q_{k+1}} E[\widetilde{y}_{k+1} | a^1 = 0, m = k+1] \\
= \; & \frac{(1-\theta_k - \psi_k)p_k}{q_{k+1}} \Big(\frac{1}{k+1} E[y^1 | a^0 = 1, m = k] + \frac{k}{k+1} E[\widetilde{y}_k | a^0 = 1, m = k]\Big) + \\
& \frac{\psi_k p_k}{q_{k+1}} \Big(\frac{1}{k+1} E[y^1 | a^0 = 0, a^1 = 1, m = k] + \frac{k}{k+1} E[\widetilde{y}_k | a^0 = 0, a^1 = 1, m = k]\Big) + \\
& \frac{\theta_{k+1}p_{k+1}}{q_{k+1}} E[\widetilde{y}_{k+1} | a^1 = 0, m = k+1]
\end{aligned} \tag{3.11a}$$

$$
\begin{aligned}
& E[\overline{y}^0_{k+1}|n = k+1] = E[\overline{y}^0_{k+1}(a^0 = 1, m = k) \text{ or } (a^0 = 0, m = k+1)] \\
= \;& \frac{(1-\theta_k-\psi_k)p_k}{q_{k+1}} E[\overline{y}^0_{k+1}|a^0 = 1, m = k] + \frac{(\theta_{k+1}+\psi_{k+1})p_{k+1}}{q_{k+1}} E[\widetilde{y}_{k+1}|a^0 = 0, m = k+1] \\
= \;& \frac{(1-\theta_k-\psi_k)p_k}{q_{k+1}} (\frac{1}{k+1} E[y^0|a^0 = 1, m = k] + \frac{k}{k+1} E[\widetilde{y}_k|a^0 = 1, m = k]) + \\
& \frac{\theta_{k+1}p_{k+1}}{q_{k+1}} E[\widetilde{y}_{k+1}|a^1 = 0, m = k+1] + \frac{\psi_{k+1}p_{k+1}}{q_{k+1}} E[\widetilde{y}_{k+1}|a^0 = 0, a^1 = 0, m = k+1]
\end{aligned}
\tag{3.12a}
$$

Define an intermediate parameter of interest, $E[\overline{y}^d_{k+1}|(a^0 = 1, m = k) \text{ or } (a^1 = 0, m = k+1)]$, the common components between (11a) and (11b). Note that $\psi_k p_k$ can be identified from the observed data as the difference in the proportion of participants with k matches or fewer between the non-treated and treated groups:

$$
\begin{aligned}
\psi_k p_k &= \Pr(a^0 = 0, a^1 = 1, m = k) \\
&= \Pr(n^0 \leq k) - \Pr(n^1 \leq k) = Q^0_k - Q^1_k
\end{aligned}
\tag{3.13}
$$

Therefore, the upper (or lower) bound of $E[\overline{y}^1_{k+1}|(a^0 = 1, m = k) \text{ or } (a^1 = 0, m = k+1)]$ can be estimated by excluding the bottom (or top) $\psi_k p_k$ percentile of the outcomes of the treatment group with $k+1$ matches as follows:

$$\overline{\Omega^1_{k+1}} = \frac{1}{1-\frac{\psi_k p_k}{q^1_{k+1}}}\int_{F^{-1}(\frac{\psi_k p_k}{q^1_{k+1}})}^{+\infty} \overline{y}^1_{k+1} f(y)dy$$

$$\underline{\Omega^1_{k+1}} = \frac{1}{1-\frac{\psi_k p_k}{q^1_{k+1}}}\int_{-\infty}^{F^{-1}(1-\frac{\psi_k p_k}{q^1_{k+1}})} \overline{y}^1_{k+1} f(y)dy$$

Analogously, the upper (or lower) bound of $E[\overline{y}^0_{k+1}|(a^0 = 1, m = k) \text{ or } (a^1 = 0, m = k+1)]$ can be estimated by excluding the bottom (or top) $\psi_{k+1}p_{k+1}$ percentile of the outcomes of the control group with $k+1$ matches as follows:

$$\overline{\Omega^0_{k+1}} = \frac{1}{1-\frac{\psi_{k+1} p_{k+1}}{q^0_{k+1}}}\int_{F^{-1}(\frac{\psi_{k+1} p_{k+1}}{q^0_{k+1}})}^{+\infty} \overline{y}^0_{k+1} f(y)dy$$

$$\underline{\Omega^0_{k+1}} = \frac{1}{1-\frac{\psi_{k+1} p_{k+1}}{q^0_{k+1}}}\int_{-\infty}^{F^{-1}(1-\frac{\psi_{k+1} p_{k+1}}{q^0_{k+1}})} \overline{y}^0_{k+1} f(y)dy$$

Denote as Δ_k: the difference in the intermediate parameters between the treatment and control groups

$$\begin{aligned}
\Delta_k = \; & E[\overline{y}^1_{k+1}|(a^0 = 1, m = k) \text{ or } (a^1 = 0, m = k+1)] - \\
& E[\overline{y}^0_{k+1}|(a^0 = 1, m = k) \text{ or } (a^1 = 0, m = k+1)] \\
= \; & \frac{(1-\theta_k-\psi_k)p_k}{(1-\theta_k-\psi_k)p_k + \theta_{k+1}p_{k+1}} \frac{1}{k+1} E[y^1 - y^0|a^0 = 1, m = k]
\end{aligned}$$

Therefore, the particular parameter of interest – the average treatment effect for a subpopulation whose outcomes are always observed together with k erroneous matches, irrespective of their treatment assignment status, can be identified as:

$$
\begin{aligned}
E[y^1 - y^0 | a^0 &= 1, m = k] = \frac{(1-\theta_k-\psi_k)p_k + \theta_{k+1}p_{k+1}}{(1-\theta_k-\psi_k)p_k}(k+1)\Delta_k \\
&= \frac{q^0_{k+1} - Q^1_{k+1} + Q^0_k}{(1-\theta_k-\psi_k)p_k}(k+1)\Delta_k
\end{aligned}
\tag{3.14}
$$

where $Q^d_k = \sum_{k=0}^{\infty} q^d_k$. Note that q^0_{k+1}, Q^1_{k+1}, Q^0_k, and $k+1$ are all observable in the data. Δ_k is bounded by $[\underline{\Omega^1_{k+1}} - \underline{\Omega^0_{k+1}}, \overline{\Omega^1_{k+1}} - \underline{\Omega^1_{k+1}}]$. The denominator can be written as:

$$(1-\theta_k-\psi_k)p_k = M_k - Q^0_k$$

Analogously with the discussion in Section 3.4.1, when prior knowledge or a consistent estimator of the bound of M_k – $[\underline{M_k}, \overline{M^k}]$ – is available, $(1-\theta_k-\psi_k)p_k$ can be bounded by $[\underline{M_k} - Q^0_k, \overline{M^k} - Q^0_k]$. Therefore, $E[y^1 - y^0|(a^0 = 1, m = k)]$ is bounded by the following formula:

$$(q^0_{k+1} - Q^1_{k+1} + Q^0_{k+1})(k+1)[\min\{\frac{\underline{\Omega^1_{k+1}} - \overline{\Omega^0_{k+1}}}{\overline{M_k} - Q^0_k}, \frac{\underline{\Omega^1_{k+1}} - \overline{\Omega^0_{k+1}}}{\underline{M_k} - Q^0_k}\}, \max\{\frac{\overline{\Omega^1_{k+1}} - \underline{\Omega^0_{k+1}}}{\overline{M_k} - Q^0_k}, \frac{\overline{\Omega^1_{k+1}} - \underline{\Omega^0_{k+1}}}{\underline{M_k} - Q^0_k}\}]^1 \tag{3.15}$$

The final parameter of interest, the difference in the mean potential outcomes for all participants whose outcomes are always observed (possibly with imperfect matching), irrespective of their treatment status, can be written as:

[1]We take into account that $\underline{\Omega^1_{k+1}} - \overline{\Omega^0_{k+1}}$ or $\overline{\Omega^1_{k+1}} - \underline{\Omega^0_{k+1}}$ may be negative. Note that $\underline{M_k} - Q^0_k$ and $\overline{M_k} - Q^0_k$ must be positive.

$$
\begin{aligned}
& E[y_1 - y_0 | a^0 = 1] \\
= \ & \frac{\sum_{k=0}^{\infty} \Pr(a^0 = 1, m = k) E[y^1 - y^0 | a = 1, m = k]}{\sum\limits_{k=0}^{\infty} \Pr(a^0 = 1, m = k)} \\
= \ & \frac{\sum_{k=0}^{\infty} ((1 - \theta_k - \psi_k) p_k + \theta_{k+1} p_{k+1})(k+1)\Delta_k}{\sum\limits_{k=0}^{\infty} (1 - \theta_k - \psi_k) p_k} \\
= \ & \frac{\sum_{k=0}^{\infty} (q^0_{k+1} - Q^1_{k+1} + Q^0_{k+1})(k+1)\Delta_k}{\sum_{k=0}^{\infty} (1 - \theta_k - \psi_k) p_k}
\end{aligned} \tag{3.16}
$$

Denoting the numerator of equation (3.15) as A and the denominator of equation (15) as B.

$$
\underline{A} = \sum_{k=0}^{\infty} (q^0_{k+1} - Q^1_{k+1} + Q^0_{k+1})(k+1)(\overline{\Omega^1_{k+1}} - \underline{\Omega^0_{k+1}}) \leq A \leq \sum_{k=0}^{\infty} (q^0_{k+1} - Q^1_{k+1} + Q^0_{k+1})(k+1)(\overline{\Omega^1_{k+1}} - \underline{\Omega^0_{k+1}}) = \overline{A}
$$

$$
\underline{B} = \sum_{k=0}^{\infty} \underline{M_k} - \sum_{k=0}^{\infty} Q^0_k \leq B = \sum_{k=0}^{\infty} M_k - \sum_{k=0}^{\infty} Q^0_k \leq \sum_{k=0}^{\infty} \overline{M_k} - \sum_{k=0}^{\infty} Q^0_k = \overline{B}
$$

Then, the final parameter of interest $E[y_1 - y_0 | a^0 = 1]$ is bounded by

$$
[\min\{\underline{A}/\underline{B}, \underline{A}/\overline{B}\}, \max\{\overline{A}/\underline{B}, \overline{A}/\overline{B}\}]
$$

3.5 Summary

In this chapter, we show that average treatment effects in general cannot be identified or even bounded mathematically when sample attrition and imperfect matching both exist. However, if prior knowledge or a consistent estimator of the mismatching probability is available and the treatment status only affects sample selection in "one direction," we can construct bounds for average treatment effect for a specific subpopulation whose outcomes would be observed irrespective of treatment assignment status. The bounds can be very tight when the sample attrition rate is small and the probability of mismatches declines rapidly with the number of mismatches.

Bibliography

[1] Joshua Angrist, Guido Imbens, and Donald Rubin. Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91:444–455, 1996.

[2] Joel L Horowitz and Charles F Manski. Identification and robustness with contaminated and corrupted data. *Econometrica*, 63(2):281–302, 1995.

[3] Joel L Horowitz and Charles F Manski. Nonparametric analysis of randomized experiments with missing covariate and outcome data. *Journal of the American Statistical Association*, 95(449):77–84, 2000.

[4] Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62:467–76, 1994.

[5] David S. Lee. Trimming for bounds on treatment effects with missing outcomes. NBER Technical Working Papers 0277, National Bureau of Economic Research, 2002.

[6] David S. Lee. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies*, forthcoming.

[7] Geert Ridder and Robert Moffitt. The econometrics of data combination. In James J. Heckman and Edward Leamer, editors, *Handbook of Econometrics*, volume 6B, chapter 75, pages 5469–5547. Elsevier Science, Amsterdam, 2007.