Journal of Public Economics 142 (2016) 56–82

Contents lists available at ScienceDirect

Journal of Public Economics

journal homepage: www.elsevier.com/locate/jpube

Identification of treatment effects under imperfect matching with an application to Chinese elite schools ☆

Hongliang Zhang

Department of Economics, Hong Kong Baptist University, Hong Kong

A R T I C L E I N F O

Article history:
Received 7 May 2014
Received in revised form 17 February 2016
Accepted 15 March 2016
Available online 25 March 2016

JEL classification:
C26
C81
I21
I28

Keywords:
Local average treatment effect
Distributional treatment effects
Imperfect matching
Elite schools
Student achievement

A B S T R A C T

This paper extends the treatment effect framework for causal inference to contexts in which the instrument appears in a data set that can only be linked imperfectly to the treatment and outcome variables contained in another data set. To overcome this problem, I form all pairwise links between information on the instrument and information on the treatment and outcome matched by the commonly recorded personal characteristics in both data sets. I show how these imperfect conditional matches can be used to identify both the average and distributional treatment effects for compliers of the common units of the two data sets. This multiple data source approach is then applied to analyze the effect of attending an elite middle school in a Chinese city where schools' admissions lottery records can only be linked imperfectly to the administrative student records.

1. Introduction

Instrumental variable (IV) methods are widely used by economists and other social scientists in program evaluations for informing public policy. In a series of seminal papers, Imbens and Angrist (1994, hereafter IA) and Angrist et al. (1996, hereafter AIR) develop a causal framework with unrestricted heterogenous potential outcomes and lay out the assumptions under which the IV estimand has a precise and straightforward causal interpretation: the local average treatment effect (LATE) for compliers whose treatment status is affected by the instrument. Abadie (2002) shows that under the same set of assumptions, this framework can be used to identify the distributional treatment effects for compliers. This causal framework, which I refer to as the benchmark IV treatment effect (hereafter IVTE) framework, assumes concurrent observation of the instrument, treatment, and outcome for all individuals of interest. However, observational problems sometimes prevent researchers from obtaining all of these variables from a single data source. When they are contained in multiple data sources, researchers are confronted with the problem of whether this framework can still be invoked for causal inference by combining different data sources (see Ridder and Moffitt (2007) for a survey on data combination). A prominent example is the two-sample IV estimation proposed independently by Angrist and Krueger (1992) and Arellano and Meghir (1992), in which instrument Z is common to both data sets, but treatment D and outcome Y are included in only one or the other. Despite the lack of data on the joint distribution of Z, D, Y, under certain conditions moments derived separately from the two data sets can still be combined to form an IV estimator that carries the same LATE interpretation as that formed under concurrent observation of Z, D, Y.

☆ An earlier version of this paper was circulated under the title "The mirage of elite schools: Evidence from lottery-based school admissions in China." I thank Joshua Angrist, Abhijit Banerjee, James Berry, Weili Ding, Esther Duflo, Rongzhu Ke, Cynthia Kinnan, Frank Levy, Steven Lehrer, Weifeng Li, Haoming Liu, Alex Mas, Roger Moon, Benjamin Olken, Florian Ploeckl, Karen R. Polenske, Aaron Sojourner, Zhentao Shi, Christopher Taber, William Wheaton, and seminar participants at CUHK, Georgetown University, Hong Kong Baptist University, HKUST, MIT, Nanyang Technological University, Peking University, Shanghai Jiao Tong University, University of Kent, University of Toronto, the 2010 AEA Annual Meeting, and the 2012 Singapore Conference on Evidence-based Public Policy Using Administrative Data for valuable discussions and comments. I am also grateful to the co-editor and three anonymous referees for their helpful comments and feedback, and Amy Ru Chien Tseng for her excellent research assistance. I acknowledge financial support from the Hong Kong Research Grants Council General Research Fund (No. 458610). All remaining errors are my own.
E-mail address: zhang.hongliang@gmail.com (H. Zhang).

http://dx.doi.org/10.1016/j.jpubeco.2016.03.004
0047-2727/

In this paper, I consider another multiple data source approach to implement the IVTE framework in settings with a different observational problem to that in the two-sample IV estimation.[1] In the settings considered here, instrument Z is observed in one data set, while treatment D and outcome Y are observed in the other. The two data sets share a substantial number of common units and a set of commonly recorded personal characteristics C (e.g., name, gender, age). However, C does not constitute a unique identifier for *exact matching* of the records between the two data sets, resulting in *imperfect matching* when they are combined. To address this observational obstacle, I propose a data combination procedure that forms all pairwise links between the records in the two data sets that are matched by C, including both correctly and incorrectly linked record pairs that are observationally indistinguishable. Under an additional assumption of the independence of the treatment and outcome to the incorrectly linked instrument in case of false matching, I demonstrate that both the average and distributional treatment effects for compliers can be identified using the imperfect conditional matches formed in the above data combination procedure. For expository purposes, I follow AIR to outline the extension of the IVTE framework for causal inference under imperfect matching in a setup with a simple random assignment of the instrument, in which the aforementioned additional independence assumption is immediately satisfied by random assignment. It is worth noting that the extended IVTE framework for causal inference under imperfect matching developed here is also applicable to the regression discontinuity (RD) design, in which the variation in the running variable in a neighborhood around the discontinuity threshold (i.e., the instrument) is "as good as randomized" when individuals have only imprecise control over the running variable (Lee, 2008; Lee and Lemieux, 2010).

This multiple data source approach for implementing the IVTE framework under imperfect matching is then applied to analyze the effect of elite school attendance on students' academic achievements and school placements in a Chinese city where admissions lotteries are used to determine the allocation of places at the oversubscribed elite middle schools (hereafter elite schools). The data used in this analysis come from two sources: school admissions lottery records and administrative records from the Middle School Exit Exam (MSEE), which is compulsory for all middle school (grades 7–9) students and also serves as the entrance exam for secondary school (grades 10–12) admissions. The former contain information on the school choices and lottery assignments for all elite school applicants, and the latter contain information on the MSEE scores, middle schools attended, and secondary school placements of all MSEE takers. However, linking records between the two data sources is challenged because of the lack of a common unique identifier. I thus resort to the commonly recorded personal characteristics in the two data sources, i.e., name and gender, to form all pairwise links between the lottery and MSEE records, which sometimes leads to false links between records of different individuals who happen to possess the same name. Nonetheless, despite the presence of falsely linked record pairs, applying the extended IVTE framework to all linked lottery-MSEE pairs identifies the average and distributional treatment effects for lottery compliers who are retained in the MSEE data set if such retentions are not affected by lottery assignments.

Despite the considerable superiority of elite schools in terms of students' academic achievements and school placements, exploiting the exogenous variation in access to elite schools generated by admissions lotteries reveals little evidence that elite schools confer any general achievement or placement benefits to their attendees who are lottery compliers. While there is some evidence that winning a lottery may induce some applicants who would otherwise opt out of the MSEE sample to remain in the sample, whom I refer to as "*marginal compliers*," it is unlikely that the failure to establish evidence of any positive effects of elite school attendance is driven by biases arising from the small degree of lottery assignment-induced differential sample selection. Evidence from a subsample of applicants whose baseline scores at middle school entry are available suggests that the sign of the biases from lottery assignment-induced differential sample selection is likely to be positive, as the baseline scores of marginal compliers compare favorably with those of others. An analysis of heterogeneous effects by gender and baseline scores yields some suggestive evidence that attending an elite school might increase the chances of top-echelon high school admissions for girls and stronger applicants, although these estimates are not significant at conventional levels. Moreover, such placement benefits, if they indeed exist, seem to be attributable, by and large, to elite school attendees playing more successful strategies in their secondary school choices, which could increase their chances of gaining admission to a top-echelon high school conditional on MSEE scores under the Boston mechanism that the city uses to assign secondary school places.

A growing body of literature uses compelling research designs to examine the effect of attending a preferred school on students' achievements and academic outcomes. One strand of this literature exploits the exogenous variation in school access generated by admissions lotteries to examine the effects of attending a preferred, academically nonselective schools, albeit with mixed findings. For example, Cullen et al. (2006) find no test score gains from winning a lottery to a choice school in Chicago, whereas Deming et al. (2014) report substantial gains in postsecondary attainment and high school GPA for students who win the lottery to attend their first-choice school in Charlotte-Mecklenburg, concentrated for girls and students from low-quality neighborhood schools.[2] The other strand of this literature uses RD designs to examine the effects of attending a preferred, academically selective school, but again with mixed results. On the one hand, Jackson (2010) and Pop-Eleches and Urquiola (2013) discover large test score gains for students attending selective secondary schools in Trinidad and Tobago and Romania. On the other hand, Clark (2010) finds that grammar schools in the UK yield little achievement benefit to their attendees, and Abdulkadiroglu et al. (2014) and Dobbie and Fryer (2014) show that attending an exam school has little causal effect on test scores or college quality in the US.

Contrasting evidence has also been found in the Chinese context by studies exploiting discontinuity around the entrance exam cutoffs to examine the effects of attending selective high schools in China. Ding and Lehrer (2007) show positive effects of selective high school attendance on students' test scores on the College Entrance Exam in a county in Jiangsu province, whereas Dee and Lan (2015) find no such positive effects in a city in Inner Mongolia Autonomous Region.[3] If a preferred, better schooling environment indeed benefits students academically, there seem to be reasons to expect the achievement gains to be more salient in China than in the West. First, secondary

[1] In the two-sample IV estimation, the two data sets must be independent of each other but do not need to contain any overlapping units. The opposite is true in the current paper: the two data sets must share a substantial number of common units but do not necessarily need to be independent of each other.

[2] Deming et al. (2014) estimate a school's college "value added" as the school average residual from a student-level regression of an indicator for four-year college enrollment on a set of student characteristics, including demographics and prior math and reading scores. Based on these estimated college "value added," they label the four lowest-ranked schools as "low quality" and all other schools as "high quality."

[3] Ding and Lehrer (2007) attribute the substantial achievement gains in the Jiangsu context as the result, at least in part, of strong positive peer effects in selective high schools; Dee and Lan (2015) interpret the lack of achievement benefits in the Inner Mongolia context as due to the limited access of marginal admissions to better teachers and classmates as the result of subject tracking in selective high schools.

school and college admissions in China are almost solely determined by students' test scores in entrance exams. Thus, compared with many Western contexts, where such admissions are based on multi-dimensional criteria that go beyond standardized test scores (e.g., teacher recommendations, extracurricular activities, leadership potential, etc.), school competition in China should focus more on academic criteria. Second, student outcomes are measured by test scores on high-stakes entrance exams and school placements. As previous research in the US context (e.g., Jacob, 2007; Corcoran et al., 2011) suggests, achievement gains, if any, on high-stakes exams are likely to outpace those on low-stakes exams. Third, to prepare for the uniform entrance exams, Chinese schools adopt the same curriculum and textbooks,[4] offering a fair ground for comparing student achievement across schools. Given these distinct features of the Chinese education system, the null results in this paper and in Dee and Lan (2015) pose more serious challenges to the presence of significant academic value-added of the conventionally preferred elite schools, compared with similar results found in other contexts.

The remainder of the paper is organized as follows. Section 2 introduces the benchmark IVTE framework, illustrates the imperfect matching problem, develops a data combination procedure and a multiple data source approach to implement the IVTE framework in settings with such an observational obstacle, and formally establishes the conditions under which point or partial identification of average and distributional treatment effects can be achieved. Section 3 applies the proposed multiple data source approach to analyze the effect of elite school attendance in a Chinese city where schools' admissions lottery records can only be matched imperfectly to administrative records of students' elite school enrollment status and exit exam outcomes, and then presents the empirical results. Section 4 concludes.

2. Econometric framework

2.1. Benchmark IVTE framework

In two seminal papers, IA and AIR outline a causal inference framework with heterogeneous effects and formulate the IV methods in a potential outcome notation. Let $D_i(z)$ denote the potential treatment status of individual i were the individual to have instrument value $Z_i = z$ and $Y_i(d, z)$ denote the potential outcome of individual i were the individual to have treatment status $D_i = d$ and instrument value $Z_i = z$. Assuming concurrent observation of $(Z_i, D_i(Z_i), Y_i(D_i(Z_i), Z_i))$ for every individual i, IA and AIR show that under some plausible conditions, the IV estimand identifies the LATE for compliers whose treatment status is changed because of the instrument. Moreover, under the same set of assumptions, Abadie (2002) further shows that this framework can be used to identify the cumulative distributions of potential outcomes for compliers, thus also allowing the identification of distributional treatment effects for compliers. Proposition 1 summarizes the results of IA, AIR, and Abadie (2002) for cases where both the instrument and treatment are binary.[5]

Proposition 1. *Suppose the following assumptions hold:*

(A1). Exclusion I: $Y_i(d, z) = Y_i(d)$ for $d \in \{0, 1\}, z \in \{0, 1\}$;
(A2). Independence I: $(D_i(0), D_i(1), Y_i(0), Y_i(1)) \perp Z_i$;
(A3). First stage: $E[D_i(1) - D_i(0)] > 0$ and $0 < P(Z_i = 1) < 1$;
(A4). Monotonicity I: $D_i(1) \geq D_i(0) \forall i$.

Then, the LATE, cumulative distributions of potential outcomes, and distributional treatment effects for compliers can be identified as follows:

$$\begin{aligned}
\gamma_c &= E[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)] = \tfrac{E[Y_i|Z_i=1] - E[Y_i|Z_i=0]}{E[D_i|Z_i=1] - E[D_i|Z_i=0]};\\
F^C_{(1)}(y) &= E[1(Y_i(1) \leq y) | D_i(1) > D_i(0)] = \tfrac{E[1(Y_i \leq y)D_i|Z_i=1] - E[1(Y_i \leq y)D_i|Z_i=0]}{E[D_i|Z_i=1] - E[D_i|Z_i=0]};\\
F^C_{(0)}(y) &= E[1(Y_i(0) \leq y) | D_i(1) > D_i(0)] = \tfrac{E[1(Y_i \leq y)(1-D_i)|Z_i=0] - E[1(Y_i \leq y)(1-D_i)|Z_i=1]}{E[D_i|Z_i=1] - E[D_i|Z_i=0]};\\
F^C_{(1)}(y) &- F^C_{(0)}(y) = \tfrac{E[1(Y_i \leq y)|Z_i=1] - E[1(Y_i \leq y)|Z_i=0]}{E[D_i|Z_i=1] - E[D_i|Z_i=0]}.
\end{aligned}$$

2.2. Identification of treatment effects under imperfect matching

While the setup of the benchmark IVTE framework assumes concurrent observation of (*Z,D,Y*), sometimes there is no single data set that contains all variables of interest. Observational challenges arise when the relevant variables are in separate data sets. A particular challenge addressed in this paper is that treatment *D* and outcome *Y* belong to one data set, referred to as the primary data set, and instrument *Z* belongs to another, referred to as the instrument data set. Such separation in itself may not constitute a problem if exact matching based on a common unique identifier, such as a person's social security number, is possible between the two data sets; that is, if the primary data set consists of (*D,Y,i*) and the instrument data set (*Z,i*), where *i* is a unique identifier observed in both data sets. However, the two data sets may not contain such a common unique identifier. Instead, they may share only a set of personal characteristics, *C*, that pertain to the individuals of interest,

[4] In the city under study, classroom teaching in elite schools and neighborhood schools covers the same set of materials designated by the city education council. Nonetheless, the focus of instruction may be different, with elite schools placing more emphasis on advanced materials.
[5] For the identification of LATE, Angrist and Imbens (1995) and Angrist et al. (2000) also generalize the results to cases with multi-valued instruments and treatments.

such as name, gender, and age. In other words, the primary data set might contain (D,Y,C) while the instrument data set might contain (Z,C). The problem of imperfect matching arises when C does not constitute a unique identifier in one or both data sets. In this subsection, I demonstrate that if the two data sets share a substantial number of common units, then under a plausible additional assumption both the average and distributional treatment effects can be identified for a subset of the common units whose treatment status is affected by the instrument.

Let us assume that there exists a superpopulation of latent identifiers from which the latent identifiers of individuals in the instrument data set and the primary data set are drawn. The instrument data set contains (Z_i, C_i) for $i \in I$ and the primary data set contains (D_j, Y_j, C_j) for $j \in J$. In this notation, $i = j$ if the instrument record and the primary record pertain to the same individual and $i \neq j$ if the two records pertain to different individuals. Now consider a data combination procedure that forms *all pairwise links* between the records in the two data sets that are matched by the common characteristic variables C. That is, for each instrument record i and each primary record j such that $C_i = C_j$, construct a linked record pair $(Z_{i(j)}, D_{i(j)}, Y_{i(j)})$ in which $Z_{i(j)} = Z_i$, $D_{i(j)} = D_j$, and $Y_{i(j)} = Y_j$. Then, the combined data set can be written as

$$\Psi = \left\{ \left(Z_{i(j)},\ D_{i(j)},\ Y_{ij}\right)\ \forall\ C_i = C_j,\ i \in I, j \in J \right\}.$$

To start with, I assume that all individuals in the instrument data set are also observed in the primary data set (i.e., $I \subseteq J$) and restrict attention to those for whom the values of Z, D, and Y are all observed (although separately).[6] As shown in Proposition 1, with a binary instrument and treatment, the IV estimand can be expressed as the ratio of the intent-to-treat (ITT) effect of Z on Y and of Z on D. Under imperfect matching but without missing observations on treatments and outcomes for individuals with instrument values, these two ITT estimands can be constructed using the linked record pairs in Ψ as follows :

$$E\left[Y_{i(j)}|Z_{i(j)}=1\right] - E\left[Y_{i(j)}|Z_{i(j)}=0\right] = (1-p)\left(E\left[Y_i|Z_i=1\right] - E\left[Y_i|Z_i=0\right]\right) + p\left(E\left[Y_j|Z_i=1,\ C_i=C_j,\ i\neq j\right] - E\left[Y_j|Z_i=0,\ C_i=C_j,\ i\neq j\right]\right) \tag{1a}$$

and

$$E\left[D_{i(j)}|Z_{i(j)}=1\right] - E\left[D_{i(j)}|Z_{i(j)}=0\right] = (1-p)\left(E\left[D_i|Z_i=1\right] - E\left[D_i|Z_i=0\right]\right) + p\left(E\left[D_j|Z_i=1,\ C_i=C_j,\ i\neq j\right] - E\left[D_j|Z_i=0,\ C_i=C_j,\ i\neq j\right]\right), \tag{1b}$$

where $p = P[i \neq j | C_i = C_j]$, the proportion of false matches among all matches in Ψ. Eqs. (1a) and (1b) show that for both Y and D, the ITT estimand of Z's effect is a weighted average of the mean difference by Z of correct matches and that of false matches, with the weights equal to their corresponding proportions in Ψ.

Concurrent observation of (Z, D, Y) considered in Proposition 1 can be viewed as a special case with no false matches (i.e., $p = 0$), under which the second term is 0 in both equations and the LATE is identified as a Wald estimator by taking the ratio of the two ITT estimands. However, even when $p > 0$, the second term in both equations can still be eliminated if the outcome and treatment are *mean* independent of the falsely linked instrument among the erroneously formed observations, i.e., $E[Y_j | Z_i, C_i = C_j, i \neq j] = E[Y_j | C_i = C_j, i \neq j]$ and $E[D_j | Z_i, C_i = C_j, i \neq j] = E[D_j | C_i = C_j, i \neq j]$. In such cases, although the two ITT estimands are both attenuated by the proportion of false matches in Ψ, the biases are canceled out when their ratio is taken in calculating the Wald estimator. Although the mean independence of D_j and Y_j to Z_i among false matches is sufficient to identify the LATE under imperfect matching, a stronger independence condition – the independence of (D_j, Y_j) to Z_i under false matching – is needed for the identification of cumulative distributions of potential outcomes and distributional treatment effects for compliers. Assumption (A5) states formally this stronger independence condition:

(A5). Independence II: $(D_j, Y_j) \perp Z_i | (C_i = C_j, i \neq j)$.

Implicit in the notation adopted in this paper is the assumption of no interference between individuals (Cox, 1958), which implies that the observation of one individual is unaffected by the instrument value taken by any other individual. With no interference between individuals, Assumption (A5) is immediately satisfied if the instrument values are independently and randomly assigned. Given the nature of random assignment and the no interference assumption, the instrument value of an individual is independent of the observed treatments and outcomes of all other individuals, including those who happen to share the same values in C, thus ensuring Assumption (A5) is met. Another assignment mechanism worth mentioning is the RD design in which instrument Z is a binary indicator for the running variable to exceed the discontinuity threshold. Lee (2008) formally demonstrates that when individuals have only imprecise control over the running variable, this instrument is as good as randomly assigned around the discontinuity threshold. Thus, Assumption (A5) also holds within a neighborhood of the discontinuity threshold. However, this assumption is violated when the instrument assignment mechanism depends on the values of common variables used in data combination, in which case Z_i and Z_j will be correlated due to the influence of their shared values in C, i.e., $cov(Z_i, Z_j | C_i = C_j, i \neq j) \neq 0$. For example, a training program may randomly select applicants conditional on age according to some age-varying assignment probability schedule (i.e., applicants of the same age face the same assignment probability, whereas those of different ages may face different assignment probabilities); in the meantime, age is also used as one of the common variables to form imperfect links between the instrument records and the primary records. Then, because of the intervening role of age, the falsely linked instrument value (Z_i) and the true instrument value (Z_j) are positively correlated. Therefore, given that (D_j, Y_j) is affected by Z_j, (D_j, Y_j) is also correlated with the falsely linked instrument Z_i, violating the independence condition in Assumption (A5).

With the additional independence condition, Assumption (A5), the LATE, cumulative distributions of potential outcomes, and distributional treatment effects for compliers can be identified using the imperfectly linked record pairs in Ψ in spite of imperfect matching, which is summarized in Proposition 2.

[6] Under this assumption, some individuals may be observed only in the primary data set but not the instrument data set. These individuals, however, are excluded from the focus of the treatment effect analysis because of missing information on the instrument.

Proposition 2. *In the presence of imperfect matching, if Assumptions (A1)–(A5) hold and $I \subseteq J$, then the LATE, cumulative distributions of potential outcomes, and distributional treatment effects for compliers can be identified as follows:*

(2a) $\gamma_c = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)] = \frac{E[Y_{i(j)}|Z_{i(j)}=1]-E[Y_{i(j)}|Z_{i(j)}=0]}{E[D_{i(j)}|Z_{i(j)}=1]-E[D_{i(j)}|Z_{i(j)}=0]}$;
(2b) $F^C_{(1)}(y) = \frac{E[1(Y_{i(j)} \le y)D_{i(j)}|Z_{i(j)}=1]-E[1(Y_{i(j)} \le y)D_{i(j)}|Z_{i(j)}=0]}{E[D_{i(j)}|Z_{i(j)}=1]-E[D_{i(j)}|Z_{i(j)}=0]}$;
(2c) $F^C_{(0)}(y) = \frac{E[1(Y_{i(j)} \le y)(1-D_{i(j)})|Z_{i(j)}=0]-E[1(Y_{i(j)} \le y)(1-D_{i(j)})|Z_{i(j)}=1]}{E[D_{i(j)}|Z_{i(j)}=1]-E[D_{i(j)}|Z_{i(j)}=0]}$;
(2d) $F^C_{(1)}(y) - F^C_{(0)}(y) = \frac{E[1(Y_{i(j)} \le y)|Z_{i(j)}=1]-E[1(Y_{i(j)} \le y)|Z_{i(j)}=0]}{E[D_{i(j)}|Z_{i(j)}=1]-E[D_{i(j)}|Z_{i(j)}=0]}$.

Proof. First, applying Assumption (A5) to eliminate the contaminating term due to false matches in Eqs. (1a) and (1b) yields

$$E[Y_{i(j)}|Z_{i(j)}=1]-E[Y_{i(j)}|Z_{i(j)}=0]=(1-p)(E[Y_i|Z_i=1]-E[Y_i|Z_i=0])$$

and

$$E[D_{i(j)}|Z_{i(j)}=1]-E[D_{i(j)}|Z_{i(j)}=0]=(1-p)(E[D_i|Z_i=1]-E[D_i|Z_i=0]).$$

Taking the ratio of the above two equations cancels out the attenuation biases and identifies γ_c in Eq. (2a). Second, similar to Eq. (1a), the numerator of Eq. (2b) can be expressed as follows:

$$\begin{aligned}E[1(Y_{i(j)} \le y)D_{i(j)}|Z_{i(j)} = 1] - E[1(Y_{i(j)} \le y)D_{i(j)}|Z_{i(j)} = 0] &= (1-p)(E[1(Y_i \le y)D_i|Z_i = 1] - E[1(Y_i \le y)D_i|Z_i = 0]) \\ &\quad + p(E[1(Y_j \le y)D_j|Z_i = 1,\ C_i = C_j, i \neq j]) - E[1(Y_i \le y)D_j|Z_i = 0,\ C_i = C_j,\ i \neq j]) \\ &= (1-p)(E[1(Y_i \le y)D_i|Z_i = 1] - E[1(Y_i \le y)D_i|Z_i = 0]),\end{aligned}$$

in which the elimination of the second term from the first equality follows the independence condition in Assumption (A5). Compared with the formula for $F^C_{(1)}(y)$ in Proposition 1 under perfect observation, the numerator and denominator of Eq. (2b) are both subject to attenuation due to false matches in Ψ. Nonetheless, taking their ratio cancels out the attenuation biases and identifies $F^C_{(1)}(y)$. Analogously, compared with the formula for $F^C_{(0)}(y)$ in Proposition 1, the numerator and denominator of Eq. (2c) are both subject to the same extent of attenuation due to false matches, which is also canceled out when their ratio is taken. Finally, given that the denominators of Eqs. (2b) and (2c) are the same, it is straightforward to see that subtracting the numerator of Eq. (2c) from that of Eq. (2b) eliminates the interaction terms between $1(Y_{i(j)} \le y)$ and $D_{i(j)}$, leading to a simple Wald estimator for the distributional treatment effects for compliers in Eq. (2d). ■

Remark 1. Define Ψ_1 and Ψ_0 as two mutually exclusive and collectively exhaustive subsets of Ψ matched to individuals with $Z_i = 1$ and $Z_i = 0$, respectively, and denote the cumulative distribution of outcomes in these two subsets as $F_1(y)$ and $F_0(y)$, such that $F_1(y) = E[1(Y_{i(j)} \le y)|Z_{i(j)} = 1]$ and $F_0(y) = E[1(Y_{i(j)} \le y)|Z_{i(j)} = 0]$. Eq. (2d) of Proposition 2 shows that $\frac{F^C_{(1)}(y)-F^C_{(0)}(y)}{F_1(y)-F_0(y)}$ equals to a constant parameter $\frac{1}{E[D_{i(j)}|Z_{i(j)}=1]-E[D_{i(j)}|Z_{i(j)}=0]}$. Therefore, as illustrated in Proposition 2.1 of Abadie (2002), testing the distributional hypotheses (e.g., equality, first-order stochastic dominance) between $F^C_{(1)}(y)$ and $F^C_{(0)}(y)$ is equivalent to testing the same distributional hypotheses between $F_1(y)$ and $F_0(y)$. In a setting with perfect observation of (Z, D, Y), Abadie (2002) suggests Kolmogorov–Smirnov type nonparametric tests for distributional hypotheses of potential outcomes for compliers based on the discrepancy of the two empirical distributions $F_1(y)$ and $F_0(y)$, and proposes a bootstrap strategy to approximate the distributions of test statistics under the null hypotheses. The same idea also carries over to the setting with imperfect matching considered in the current paper, which I discuss in further detail in the empirical application in Section 3.3.5.

Remark 2. While Assumption 5 ensures the exogeneity of the instrument value among false matches, the relevance of the instrument is still weakened due to the presence of false matches in Ψ. Thus, even if there is a strong underlying first-stage relationship (i.e., $E[D_i|Z_i = 1] - E[D_i|Z_i = 0]$) between the instrument and treatment among correct matches, the observed first-stage relationship among the pairwise linked records in Ψ (i.e., $E[D_{i(j)}|Z_i = 1]-E[D_{i(j)}|Z_i = 0]$) can be seriously weakened and lead to imprecise, uninformative, and even biased IV estimates if the proportion of false matches in Ψ (i.e., p) is large. When noise from false matches outweighs useful information from correct matches in Ψ, a possible way to overcome the resulting weak instrument and imprecise estimation problems is to restrict the analysis to a subset of individuals from the instrument data set with relatively distinct values in the common variables used in matching. For example, when names are used in matching, the proportion of false matches in Ψ can be substantially reduced by excluding individuals with popular names. Although restricting the analysis to individuals with relatively low-frequency names limits the interpretation of the results to those individuals only, it does not affect the internal validity of the estimates among this subsample. In the empirical application in Section 3.3.3, I examine the sensitivity of the estimates to the extent of false matching by subdividing individuals into subsamples based on name character length and the existence of duplicate names in I, both of which are exogenous characteristics highly correlated with the extent of false matching in data combination. Note that it is inappropriate to define subsamples based on the existence or frequency of duplicate names in J because the observability of an individual (and thus the existence or frequency of his/her name) in J is potentially affected by his/her instrument value, which I discuss in detail below.

2.3. Partial identification under both imperfect matching and differential sample selection

By assuming that I is a subset of J, Proposition 2 considers complete, albeit imperfect, observation of (Z_i, D_i, Y_i) in Ψ for all $i \in I$. However, in practice, sometimes not all individuals in the instrument data set are contained in the primary data set, leading to missing observations of (Z_i, D_i, Y_i) in Ψ for some $i \in I$. Moreover, the process for determining the selection of an individual into the primary data set may be related to the individual's treatment and instrument. Let $S_i(d, z)$ denote the potential observability of individual i in the primary data set were the individual to have treatment status $D_i = d$ and instrument value $Z_i = z$. Note that a key assumption in the IVTE framework is the exclusion of the instrument from the potential outcome notation, i.e., $Y_i(d, z) = Y_i(d)$. In the same spirit, I also focus on sample selection processes in which the potential observability indicator can be indexed solely against the treatment:

(A6). Exclusion II: $S_i(d, z) = S_i(d)$ for $d \in \{0, 1\}, z \in \{0, 1\}$.

With this exclusion restriction on the sample selection process, the realized observability of an individual with instrument value Z_i in the primary data set (and thus in Ψ) can be expressed as follows:

$$S_i\left(D_i\left(Z_i\right)\right) = S_i\left(D_i(0)\right) + \left(S_i(1) - S_i(0)\right)\left(D_i(1) - D_i(0)\right) Z_i,$$

where $S_i(.)$ is indexed against the treatment and $D_i(.)$ is indexed against the instrument. For noncompliers (i.e., $D_i(0) = D_i(1)$), sample selection is independent of the instrument as the second term is always 0. However, for compliers (i.e., $D_i(1) > D_i(0)$), sample selection could be affected by the instrument if $S_i(1) \neq S_i(0)$ for some individuals. In a setting with perfect compliance, similar to the monotonicity condition that Assumption (A4) imposes on how instrument value may affect treatment status (i.e., $D_i(1) \geq D_i(0)$), Lee (2009) invokes a monotonicity assumption for how treatment intake may affect sample selection, i.e., $S_i(1) \geq S_i(0)$. While in the imperfect compliance setting of concern in the current paper, how D_i might affect S_i is irrelevant for noncompliers as their treatment status never changes. Therefore, I impose this monotone sample selection assumption for compliers only:

(A7). Monotonicity II: $S_i(1) \geq S_i(0) \forall i$ s.t. $D_i(1) > D_i(0)$.

Under the above exclusion restriction and monotone sample selection assumption, the matched record pairs $(Z_{i(j)}, D_{i(j)}, Y_{i(j)})$ in Ψ can be partitioned into three principle strata as follows:

1. $AR = \{i = j, S_i(D_i(0)) = 1\}$, correctly matched pairs pertaining to individuals who are *always retained* in Ψ regardless of their instrument values.
2. $FM = \{i \neq j\}$, *falsely matched* pairs.
3. $MC = \{i = j, S_i(D_i(1)) > S_i(D_i(0))\}$, correctly matched pairs pertaining to individuals who are retained in Ψ if and only if $Z_i = 0$. I refer to these individuals as *marginal compliers* as they will be treated and retained in Ψ when $Z_i = 1$ (i.e., $D_i(1) = 1$ and $S_i(1) = 1$) and will be untreated and not retained in Ψ when $Z_i = 0$ (i.e., $D_i(0) = 0$ and $S_i(0) = 0$).

Note that the AR and FM strata are contained in both Ψ_0 and Ψ_1, whereas the MC stratum is contained in Ψ_1 only. Moreover, following AIR, individuals in the AR stratum can be further partitioned by their potential treatment status into *always takers* (i.e., $D_i(0) = D_i(1) = 1$), *compliers* (i.e., $D_i(0) = 0, D_i(1) = 1$), and *never takers* (i.e., $D_i(0) = D_i(1) = 0$), which I denote respectively as the AR_{at}, AR_c, and AR_{nt} substratum. Since the counterfactual outcome $Y_i(0)$ is never observed for marginal compliers, I define both the average and distributional treatment effects of interest for the subpopulation of compliers pertaining to the AR_c substratum (i.e., $D_i(1) > D_i(0), S_i(D_i(0)) = 1$). That is,

$$\gamma_{AR_c} = E\left[Y_i(1) - Y_i(0) | AR_c\right]$$

and

$$F_{(1)}^{AR_c} - F_{(0)}^{AR_c} = E[1Y_i(1) \leq y | AR_c] - E[1Y_i(0) \leq y | AR_c].$$

Note that with differential sample selection into Ψ_1 and Ψ_0 , point identifications of γ_{AR_c} and $F_{(1)}^{AR_c}(y) - F_{(0)}^{AR_c}(y)$ are generally not feasible. Specifically, Appendix A derives the bias in the IV estimand using all matched record pairs in Ψ for γ_{AR_c} under differential sample selection. As formulated in Eq. (A3), the bias in the IV estimand is determined by three factors: (i) the ratio of the proportional share of the MC stratum in Ψ_1 and the first-stage effect of Z on D in Ψ; (ii) the extent of ability selection of the MC stratum relative to that of the AR and FM strata combined; and (iii) the difference in the average treatment effect (ATE) for the MC stratum and the weighted average of the ATEs for the AR and FM strata.

Define Ψ_{11} and Ψ_{10} as two mutually exclusive and collectively exhaustive subsets of Ψ_1 with $D_{i(j)} = 1$ and $D_{i(j)} = 0$, respectively. Given Assumptions (A6) and (A7), all marginal compliers in Ψ_1 are treated and thus pertain to Ψ_{11} only. Let p_1^m and p_{11}^m denote the proportional share of the MC stratum in Ψ_1 and Ψ_{11}, respectively, which can be expressed as follows:

$$p_1^m = \Pr\left[MC | \Psi_1\right] = \frac{E\left[n_i | Z_i = 1\right] - E\left[n_i | Z_i = 0\right]}{E[n_i | Z_i = 1]} \tag{3a}$$

and

$$p_{11}^{m} = \Pr[MC|\Psi_{11}] = \frac{\Pr[MC|\Psi_1]}{\Pr[\Psi_{11}|\Psi_1]} = \frac{p_1^m}{E\left[D_{i(j)}|Z_{i(j)} = 1\right]}, \tag{3b}$$

where n_i is the number of matched record pairs in Ψ for instrument record i (i.e., $\sum_{j\in J}\mathbf{1}(C_j = C_i)$) and is affected by the instrument value Z_i for marginal compliers only.[7] Given that the sample analogs of $E[n_i|Z_i = 1]$, $E[n_i|Z_i = 0]$, and $E[D_{i(j)}|Z_{i(j)} = 1]$ are all observable in the data, both p_1^m and p_{11}^m are point identified. Let Ψ_{11}^* denote a subset of Ψ_{11} containing only the matched record pairs pertaining to the AR and FM strata (i.e., excluding the MC stratum):

$$\Psi_{11}^* = \{(Z_i, D_i, Y_i)\ \forall\ Z_i = 1, D_i = 1, S_i(D_i(0)) = 1\} \cup \{(Z_i, D_j, Y_j)\ \forall\ Z_i = 1, D_j = 1, C_i = C_j, i \neq j\}.$$

Proposition 3. *In the presence of imperfect matching, suppose Assumptions (A1)–(A7) hold and Ψ_{11}^* is identified, then the LATE, cumulative distributions of potential outcomes, and distributional treatment effects for compliers who are always retained in Ψ regardless of their instrument values (i.e., the AR_c substratum) can be identified as follows:*

$$\gamma_{AR_c} = E[Y_i(1) - Y_i(0)|AR_c] = \frac{E[Y_{i(j)}|\Psi_{11}^* \cup \Psi_{10}] - E[Y_{i(j)}|\Psi_0]}{E[D_{i(j)}|\Psi_{11}^* \cup \Psi_{10}] - E[D_{i(j)}|\Psi_0]};$$

$$F_{(1)}^{AR_c}(y) = \frac{E[1(Y_{i(j)} \le y)D_{i(j)}|\Psi_{11}^* \cup \Psi_{10}] - E[1(Y_{i(j)} \le y)D_{i(j)}|\Psi_0]}{E[D_{i(j)}|\Psi_{11}^* \cup \Psi_{10}] - E[D_{i(j)}|\Psi_0]};$$

$$F_{(0)}^{AR_c}(y) = \frac{E[1(Y_{i(j)} \le y)(1 - D_{i(j)})|\Psi_0] - E[1(Y_{i(j)} \le y)(1 - D_{i(j)})|\Psi_{11}^* \cup \Psi_{10}]}{E[D_{i(j)}|\Psi_{11}^* \cup \Psi_{10}] - E[D_{i(j)}|\Psi_0]};$$

$$F_{(1)}^{AR_c}(y) - F_{(0)}^{AR_c}(y) = \frac{E[1(Y_{i(j)} \le y)|\Psi_{11}^* \cup \Psi_{10}] - E[1(Y_{i(j)} \le y)|\Psi_0]}{E[D_{i(j)}|\Psi_{11}^* \cup \Psi_{10}] - E[D_{i(j)}|\Psi_0]}.$$

The proof of Proposition 3 is provided in Appendix B. A special case worth noting is that when $S_i(1) = S_i(0) = 1$ for all compliers, Ψ_{11}^* is equal to Ψ_{11} and thus is empirically identified. However, when Assumption (A7) holds with inequality for some compliers, Ψ_{11}^* is a latent subset of Ψ_{11} that cannot be exactly identified as it is not possible to distinguish between the AR, FM, and MC strata in Ψ_{11}. Nonetheless, the distribution of outcomes in the latent subset Ψ_{11}^* can still be bounded using the trimming method proposed by Lee (2009). Specifically, using the proportional share of the MC stratum in Ψ_{11} (p_{11}^m) identified in Eq. (3b), the distribution of outcomes in Ψ_{11}^* can be bounded by trimming the lower and upper p_{11}^m tail of the outcome distribution in Ψ_{11}, as can the two causal parameters of interest, γ_{AR_c} and $F_{(1)}^{AR_c}(y) - F_{(0)}^{AR_c}(y)$. I summarize this formally in Proposition 4.

Proposition 4. *Suppose Assumptions (A1)–(A7) hold under imperfect matching. Define two subsets of Ψ_{11} as follows:*

$$\underline{\Psi}_{11} = \left\{\left(Z_{i(j)}, D_{i(j)}, Y_{i(j)}\right)|Z_{i(j)} = 1, D_{i(j)} = 1, Y_{i(j)} \le y_{\left(1-p_{11}^m\right)}\right\};$$

$$\overline{\Psi}_{11} = \left\{\left(Z_{i(j)}, D_{i(j)}, Y_{i(j)}\right)|Z_{i(j)} = 1, D_{i(j)} = 1, Y_{i(j)} \ge y_{p_{11}^m}\right\};$$

where $y_{p_{11}^m}$ and $y_{\left(1-p_{11}^m\right)}$ denote the $100p_{11}^m$ th and $100\left(1 - p_{11}^m\right)$ th percentile, respectively, of the distribution of Y in Ψ_{11} with p_{11}^m defined in Eq. (3b). Then, the LATE and distributional treatment effects for compliers who are always retained in Ψ regardless of their instrument values (i.e., the AR_c substratum) can be partially identified as follows:

$$\frac{E\left[Y_{i(j)}|\underline{\Psi}_{11}\cup\Psi_{10}\right] - E\left[Y_{i(j)}|\Psi_0\right]}{E\left[D_{i(j)}|\underline{\Psi}_{11}\cup\Psi_{10}\right] - E\left[D_{i(j)}|\Psi_0\right]} \le \gamma_{AR_c} \le \frac{E\left[Y_{i(j)}|\overline{\Psi}_{11}\cup\Psi_{10}\right] - E\left[Y_{i(j)}|\Psi_0\right]}{E\left[D_{i(j)}|\overline{\Psi}_{11}\cup\Psi_{10}\right] - E\left[D_{i(j)}|\Psi_0\right]};$$

$$\frac{E\left[1(Y_{i(j)} \le y)D_{i(j)}|\underline{\Psi}_{11} \cup \Psi_{10}\right] - E\left[1(Y_{i(j)} \le y)D_{i(j)}|\Psi_0\right]}{E\left[D_{i(j)}|\underline{\Psi}_{11} \cup \Psi_{10}\right] - E\left[D_{i(j)}|\Psi_0\right]} \le F_{(1)}^{AR_c}(y) \le \frac{E\left[1(Y_{i(j)} \le y)D_{i(j)}|\overline{\Psi}_{11} \cup \Psi_{10}\right] - E\left[1(Y_{i(j)} \le y)D_{i(j)}|\Psi_0\right]}{E\left[D_{i(j)}|\overline{\Psi}_{11} \cup \Psi_{10}\right] - E\left[D_{i(j)}|\Psi_0\right]};$$

$$F_{(0)}^{AR_c}(y) = \frac{E\left[1\left(Y_{i(j)} \le y\right)\left(1 - D_{i(j)}\right)|\Psi_0\right] - E\left[1\left(Y_{i(j)} \le y\right)\left(1 - D_{i(j)}\right)|\left(\underline{\Psi}_{11} \text{ or } \overline{\Psi}_{11}\right)\cup\Psi_{10}\right]}{E\left[D_{i(j)}|\left(\underline{\Psi}_{11} \text{ or } \overline{\Psi}_{11}\right) \cup \Psi_{10}\right] - E\left[D_{i(j)}|\Psi_0\right]};$$

$$\frac{E\left[1\left(Y_{i(j)} \le y\right)|\underline{\Psi}_{11} \cup \Psi_{10}\right] - E\left[1\left(Y_{i(j)} \le y\right)|\Psi_0\right]}{E\left[D_{i(j)}|\underline{\Psi}_{11} \cup \Psi_{10}\right] - E\left[D_{i(j)}|\Psi_0\right]} \le F_{(1)}^{AR_c}(y) - F_{(0)}^{AR_c}(y) \le \frac{E\left[1\left(Y_{i(j)} \le y\right)|\overline{\Psi}_{11} \cup \Psi_{10}\right] - E\left[1\left(Y_{i(j)} \le y\right)|\Psi_0\right]}{E\left[D_{i(j)}|\overline{\Psi}_{11} \cup \Psi_{10}\right] - E\left[D_{i(j)}|\Psi_0\right]}.$$

[7] Specifically, for marginal compliers $n_i(1) = n_i(0) + 1$, where $n_i(1)$ and $n_i(0)$ denote the potential number of linked record pairs in Ψ to instrument record i when $Z_i = 1$ and $Z_i = 0$, respectively.

While I delegate the formal proof of Proposition 4 to Appendix C, it is useful to note a few remarks here. First, by construction, $\underline{\Psi}_{11}$ and $\overline{\Psi}_{11}$ are identical in size to Ψ^*_{11} and they all contain matched record pairs with $D_{i(j)}=1$. Therefore, replacing the latent subset Ψ^*_{11} in Proposition 3 by the constructed subset $\underline{\Psi}_{11}$ or $\overline{\Psi}_{11}$ in Proposition 4 leaves the denominators of all formulas unchanged. Second, also by construction, in terms of the distribution of Y, $\overline{\Psi}_{11}$ first-order stochastically dominates Ψ^*_{11} and Ψ^*_{11} first-order stochastically dominates $\underline{\Psi}_{11}$. Thus, replacing Ψ^*_{11} by $\underline{\Psi}_{11}\left(\overline{\Psi}_{11}\right)$ weakly decreases (increases) the numerator of the formulas for γ_{AR_c} and $F^{AR_c}_{(1)}(y)$. Third, $F^{AR_c}_{(0)}(y)$ is still point identified in Proposition 4 because both the distribution of outcomes in Ψ_{10} and the size of Ψ_{10} proportional to Ψ^*_{11} remain unchanged when Ψ^*_{11} is replaced by $\underline{\Psi}_{11}$ or $\overline{\Psi}_{11}$.

3. Analysis of the effect of elite school attendance in a Chinese city

This section applies the ideas for the identification of the average and distributional treatment effects under imperfect matching developed in the previous section to analyze the effect of elite school attendance in a Chinese city. Section 3.1 provides the background on the middle school system, the elite school admissions procedures, and the differences between elite schools and other schools in the city. Section 3.2 introduces the two data sources – the school admissions lottery records and the administrative MSEE records – and summarizes the data combination outcomes. Section 3.3 presents the main empirical results obtained from estimations that ignore the small degree of differential sample selection between lottery winners and losers. Section 3.4 performs a partial identification analysis to account for differential sample selection and constructs bounds on the treatment effects. Section 3.5 discusses the implications of the results for the debate over school choice.

3.1. Background

The empirical assessment in this paper focuses on middle schools in an affluent provincial capital city in China.[8] In this city, students are first assigned to a primary school based on the attendance area of their residence and then, upon graduation from primary school, to a neighborhood middle school (hereafter neighborhood school) through a probabilistic assignment mechanism that works at the neighborhood level. Elite schools and private schools exist outside the residence-based middle school assignment system and offer alternatives for students to opt out of their assigned neighborhood school. All elite schools in the city were historically exam schools that admitted students based on entrance exam scores, and thus are widely considered to be more advantageous than neighborhood schools. In contrast, most private schools are newly founded boarding schools that due to their lack of an established reputation are generally regarded as inferior to elite schools. Nonetheless, private schools remain a feasible and important alternative particularly for students who are interested in attending an elite school but fail to gain a place.

Elite schools' history as exam schools ended in the late 1990s when the city education council banned the use of entrance exams for middle school admissions,[9] although they remain in use for secondary school and university admissions. The shift away from entrance exams prompted elite schools to switch to a two-tier admissions scheme: advance admissions (open to gifted and talented students only)[10] and general admissions (open to all interested students willing to pay the elite school tuition fee). For the 2002–2004 entering cohorts investigated in this paper, all elite schools in the city set their tuition fees at the maximum allowed by the city education council, that is, RMB3,000 (∼ USD360) per year, or about one-tenth of the average annual disposable income of a three-person family, whereas neighborhood schools were tuition-free.[11] All of these elite schools were nonetheless oversubscribed, even though each student was allowed to apply to only one elite school.[12] Consequently, these schools all resorted to using lotteries to determine the allocation of their quotas for general admissions. Specifically, for every school, a computer program designated by the city education council was used to assign a random lottery number to each applicant and admitted students with the lowest numbers first until the school's quota was filled. To prevent tampering, all admissions lotteries were certified by public notaries. Within a few weeks after these lotteries, the winners were required to pay the entire three-year tuition fee, which was nonrefundable even if they later switched schools,[13] and those who did not pay on time were regarded as having declined their offers. Many applicants who lost out in the lottery, however, still gained admission to their selected elite school through the "back door channels."[14] As a result, the final enrollment in an elite school was much larger than its official quota. For the period investigated in this paper, a typical elite school in the city admitted approximately one-third of its students through the advance admissions, one-third through the general admissions lottery, and the remaining third through the "back door."

Panel A of Table 1 summarizes the enrollment and student outcomes of different types of middle schools using the MSEE takers in 2005,[15] the year in which the first cohort of students examined in this paper finished middle school. The city had 181 middle schools, including 160 neighborhood schools, 16 elite schools, and 5 private schools. The average school-grade size was 591 for elite schools, 223 for neighborhood schools, and 114 for private schools. Accordingly, the 16 elite schools and 5 private schools accounted for 20.7% and 1.3%,[16] respectively, of the city's total middle school enrollment. Elite schools, with mean MSEE scores of 0.52σ, were

[8] The city's per capita GDP is about twice the national average, which ranks it in the top quartile among the nation's 27 provincial capital cities (excluding 4 provincial-level municipalities).

[9] This decision was a reaction to the central government's call to reduce the excesses of exam-based assessment in its reform promoting quality education (Dello-Iacovo, 2009).

[10] The recipients of the advance admissions offers were selected based on award records in city- or district-level academic, artistic and athletic contests, and were offered with full or half tuition waivers depending on award rank.

[11] Neighborhood schools were entirely publicly funded. In contrast, elite schools relied on public funding for basic personnel expenses only and used tuition fees for operating and benefit expenses, and provided a higher level of overall compensation to their teachers through school-funded benefits.

[12] Those caught making multiple applications were disqualified from enrolling in any elite school.

[13] The nonrefundable nature of the tuition payment implies that once enrolled, students rarely switched out of an elite school.

[14] Anecdotal evidence from conversations with elite school principals, teachers, and parents indicates that the key determinant of a lottery loser's chances of admission to an elite school through the "back door" was whether he/she was referred by someone (e.g., a government official) who could exert influence over the school principal. The student's academic performance in primary school was another factor, albeit of secondary importance. The tuition paid by a student admitted through the "back door" could also be affected by the importance of the referee and the student's own academic performance in primary school, and was subject a formal ceiling imposed by the city education council to be no more than twice as much as the regular tuition paid by students admitted through the lotteries.

[15] All statistics are calculated using a random sample of MSEE takers in 2005, as the census data are not available.

[16] Because some private boarding schools located outside the city's boundaries are not included in the sample, the actual enrollment share of private schools is larger than the reported 1.3%; nonetheless, the increase would be marginal even if the these schools were included.

Table 1
Summary statistics by school type.

	All schools	Elite schools	Neighborhood schools	Private schools
	(1)	(2)	(3)	(4)
Panel A. All districts[a]				
Number of schools	181	16	160	5
Average number of 9th-grade students	252	591	223	114
Enrollment share (%)	100.0	20.7	78.0	1.3
Mean MSEE scores (in s.d.)	0	0.523	−0.137	−0.093
% of students admitted to top-echelon high schools	21.7	40.0	16.9	10.5
% of students admitted to regular high schools	30.5	40.9	27.5	26.3
Panel B. District 3[b]				
Number of teachers	933	279	654	–
% teachers with master's degree	3.8	4.7	3.4	–
% teachers with bachelor's degree	88.2	88.9	87.9	–
% superior-class teachers	39.7	41.9	38.7	–
% first-class teachers	40.5	40.9	40.4	–
% teachers aged 35 or below	45.4	47.0	44.8	–
% teachers aged 36 to 45	37.4	41.9	35.5	–
Student–teacher ratio	17.9	18.7	17.6	–

Notes:
[a] Statistics reported in Panel A are calculated from a random sample of the city's MSEE takers in 2005.
[b] Statistics reported in Panel B are summary statistics from the census of all middle school teachers from District 3, which happens to have no private middle schools.

far more advantageous in terms of student achievement than neighborhood schools (-0.14σ) and private schools (-0.09σ), whereas the latter two were largely comparable. Based on their MSEE scores, middle school graduates were tracked into three types of secondary schools: top-echelon high schools, regular high schools, and vocational secondary schools, ranked in descending order of entrance scores requirements. The superior outcomes of students from elite schools are also reflected in the larger proportions of graduates admitted to top-echelon and regular high schools (40% and 41%, respectively) compared with neighborhood schools (17% and 28%, respectively) and private schools (11% and 26%, respectively). Panel B of Table 1 provides further comparisons of elite schools and neighborhood schools using information on teachers obtained from one school district, District 3. In this district, elite school teachers were superior to their neighborhood school counterparts in both educational attainment and rank. For example, elite school teachers were 1.3 percentage points (or 38.5%) more likely to have a master's degree and 3.2 percentage points (or 8.4%) more likely to have a superior-class rank. It is worth noting that elite schools gained an advantage in teacher ranking despite hiring disproportionately more young (aged 35 or below) and middle-aged (aged 36 to 45) teachers who were usually ranked lower than older teachers. However, the student–teacher ratio was higher in elite schools (18.7) than neighborhood schools (17.6), suggesting a slightly less favorable class-size environment in elite schools as a result of their excess demand.

There are several reasons that elite schools may confer achievement benefits to their attendees. First, elite school students are exposed to higher-achieving peers, whose mean MSEE scores are about two-thirds of a standard deviation higher than those of students from neighborhood schools. Second, elite school teachers are superior to their neighborhood school counterparts in observed characteristics (i.e., educational attainment, rank), and possibly also in unobserved characteristics. Third, although not reflected in the summary statistics in Table 1, elite schools are usually better equipped with modern technologies and facilities (e.g., computers, language centers, science labs) that can facilitate learning. However, despite these advantages, it is not unambiguous that elite school attendance ensures higher achievement because it is also associated with potential adverse effects, especially for weaker students admitted from general admissions lotteries. First, exposure to higher-achieving peers pushes a student's relative position to a lower rank, which may be demoralizing and thus diminish achievement (Hoxby and Weingarth, 2006). Second, because of the over-representation of high-achieving students and the exam school tradition, classroom instruction in elite schools tends to emphasize more advanced materials, which may have adverse effects on weaker students (Duflo et al., 2011; Bui et al., 2014). Third, the student–teacher ratio is higher in elite schools. Last but not least, parents who succeeded in enrolling their children in elite schools may lower their own effort (see Pop-Eleches and Urquiola (2013) for evidence in the Romanian context), whereas parents who failed to gain an elite school place for their children may increase their own effort and/or make compensatory investments in private tutoring. Both of these behavioral responses from parents can offset the positive achievement effects (if any) of elite school attendance. The empirical analysis in this paper provides a reduced-form assessment of the effect of elite school attendance that reflects the combined influences of the various pathways, and their separate influences are not distinguishable.

3.2. Data

The city investigated in this paper has seven municipal districts whose boundaries coincide with school district boundaries. The Yangtze River runs through the city and divides it into two parts: the North Bank (Districts 1–3) and the South Bank (Districts 4–7), which are largely independent enrollment areas as students rarely commute across the river for schooling. With the cooperation of the notary public office, I obtained the lottery records of three cohorts of applicants to all eight elite schools in the North Bank between 2002 and 2004. After excluding three lotteries in 2003 for which the notary public records contain only winners' information and the approximately 4% of applicants enrolled in primary schools outside the North Bank, the final lottery sample comprises 13,768 applicants for 21 admissions lotteries.[17] The lottery records contain each applicant's name, gender, primary school attended, and lottery assignment, but no information on family background or baseline scores. However, for students from District 3, I obtained separate test score records for a district-wide uniform exam taken in grade 6 (the

[17] Students in this sample accounted for 23% of all students in the North Bank who transitioned from primary school to middle school in the 2002–2004 period.

final grade at primary school) and matched them to 85% of the applicants in the lottery sample by name, gender, primary school, and year.[18]

Columns 1–3 in Table 2 report the descriptive statistics for lottery assignments and applicants' predetermined characteristics at the time of application. These admissions lotteries were quite competitive: on average only 3 out of 10 applicants won the lottery and the odds of winning (22%) were even lower in District 3. If these lotteries were indeed random, the winners and losers of a given lottery would be expected to have similar predetermined characteristics. Accordingly, I check the validity of the randomization by examining the association between applicants' win/loss status and predetermined characteristics. In column 4 of Table 2, I regress the lottery winner dummy on gender, a set of dummy indicators for primary school attended, and lottery fixed effects for the full sample. Neither the coefficient on the female indicator nor any of the coefficients on the primary school dummies (omitted in the table) are statistically significant. The F-test of the joint significance of these coefficients (excluding lottery fixed effects) is very small and insignificant ($F = 0.78$, $p\text{-}value = 0.986$), suggesting little evidence that gender or primary school attended was associated with the odds of winning. For applicants from District 3, I further include the availability of baseline scores and, if available, the combined 6th-grade math and Chinese scores (standardized to have zero mean and unit variance for each cohort) as additional regressors. The results, reported in columns 5 and 6 of Table 2, show that neither the availability nor the level of scores was associated with the lottery assignment. In particular, the fact that these lotteries did not favor applicants with higher baseline scores provides compelling evidence for the validity of their randomness.

The other data source is the administrative MSEE records, which contain information on each exam taker's MSEE scores, middle school attended, and secondary school placement outcome. The lottery and MSEE records are connected through two common variables: name and gender. As skipping or repeating a grade rarely occurs in middle school in the city under study and few students commute across the Yangtze River for schooling, for each lottery the target MSEE records are restricted to those pertaining to exam takers from the North Bank three years after that lottery.[19] Implementing the data combination algorithm described in Section 2.2 under this restriction yields all pairwise links between the lottery and MSEE records that share the same name conditional on gender and cohort.

With the assignment probability ($\Pr(Z_i = 1)$) varying across lotteries, Panel A of Table 3 separately compares the lottery-adjusted[20] matching statistics between winners and losers for the full sample (in columns 1–3) and the District 3 subsample with baseline scores (in columns 4–6). Let $\bar{n}_1$ and $\bar{n}_0$ denote the average number of matches to each lottery winner and loser, respectively. For the full sample, Panel A of Table 3 yields sample estimates of 1.592 and 1.567 for $\bar{n}_1$ and $\bar{n}_0$, respectively. Given Eq. (3a), an estimator for the proportional share of the MC stratum in Ψ_1 can be formulated as $\hat{p}_1^m = \frac{\bar{n}_1 - \bar{n}_0}{\bar{n}_1} = \frac{1.592 - 1.567}{1.592} = 0.016$. Let nf_i denote the number of false matches to applicant i, which is linked to the number of total matches n_i and the sample selection indicator S_i such that $n_i = nf_i + S_i$. Because nf_i is independent of lottery assignments, the win/loss difference in the average number of matches corresponds to the proportion of *marginal compliers* in *I*, i.e., $\theta_{MC} = \Pr(S_i(D_i(1)) > S_i(D_i(0))) = \bar{n}_1 - \bar{n}_0 = 0.025$. However, the statistical power of the test of the win/loss difference in n_i is quite weak: with a standard error of 0.038, this test is unable to detect as significant any small degree of differential sample selection as indicated by the observed difference. To address this imprecision problem, I further compare the matching rate between winners and losers. Note that sample attrition ($S_i = 0$) leads to non-matching ($n_i = 0$) if and only if the applicant has no false matches ($nf_i = 0$). Thus, the win/loss difference in matching rate corresponds only to the proportion of *marginal compliers with no false matches* in *I*, which is hereafter denoted as the MC_0 substratum with the subscript 0 indicating $nf_i = 0$. This exercise yields a win/loss difference (i.e., an estimate of θ_{MC_0}) of 0.02 only. While smaller in magnitude, this difference is much more precisely estimated and is significant at the 1% level,[21] suggesting that differential sample selection does exist, even though it is not detected at a level of statistical significance in the previous test.

The small degree of differential sample selection documented above indicates the contamination of *marginal compliers* in Ψ_1. As they are indistinguishable from other matched record pairs in Ψ_1 pertaining to the AR and FM strata, it is not possible to evaluate the bias in the IV estimator directly as formulated in Eq. (A3) in Appendix A. Nonetheless, it is possible to gauge the sign and extent of *ex ante* ability selection for the MC stratum relative to the AR stratum and the FM stratum based on baseline scores prior to the admissions lotteries under some proxy assumptions. Table 4 illustrates the working of this exercise using the subsample of District 3 applicants with baseline scores. First, column 1 in Panel A of Table 4 replicates the sample estimates reported in Table 3 for the proportional share and mean baseline scores for each of the four observed subgroups: unmatched winners ($p_{01}, \bar{a}_{01}$), matched winners ($p_{11}, \bar{a}_{11}$), unmatched losers ($p_{00}, \bar{a}_{00}$), and matched losers ($p_{10}, \bar{a}_{10}$). Moreover, for each observed subgroup, column 2 in Panel A of Table 4 further links its proportional share to those of applicants belonging to each stratum or substratum. For example, since the unmatched winners contain only *never retained individuals with no false matches* (denoted as the NRI_0 substratum), $p_{01} = \theta_{NRI_0}$; in contrast, as the unmatched losers contain both the NRI_0 substratum and the MC_0 substratum, $p_{00} = \theta_{NRI_0} + \theta_{MC_0}$. Given the sample statistics and compositional relationships listed in Panel A, three proxy assumptions are imposed in Panel B to gauge the ability selection in the MC stratum, AR stratum, and FM stratum, respectively. First, the mean baseline scores of the MC_0 substratum are estimated as $\bar{a}_{MC_0} = \frac{p_{00}\bar{a}_{00} - p_{01}\bar{a}_{01}}{p_{00} - p_{01}} = 0.685\sigma$ (row 2). Under the not very unrealistic assumption that ability selection is independent of the presence of false matches among *marginal compliers* (i.e., $\bar{a}_{MC_0} = \bar{a}_{MC_1}$), this estimator is also used to proxy for the extent of the ability selection in the entire MC stratum ($\bar{a}_{MC}$). Second, assuming that the NRI stratum and the MC stratum have the same proportion of individuals with/without false matches (i.e., $\frac{\theta_{NRI_1}}{\theta_{NRI_0}} = \frac{\theta_{MC_1}}{\theta_{MC_0}}$), the proportional share of the NRI_1 substratum (θ_{NRI_1}) is estimated to be 0.021 (row 4). Subtracting both θ_{NRI_1} and θ_{MC} from p_{11} yields an estimate of θ_{AR} of 0.91 (row 5), indicating that the vast majority (95.2%) of the matched winners belong to the AR stratum. Therefore, the mean baseline scores of the matched winners ($\bar{a}_{11} = 0.332\sigma$) are a close proxy for those of the AR stratum ($\bar{a}_{AR}$). Third, there seems to be little reason to expect any ability selection in the false matches in the FM stratum (i.e., $\bar{a}_{FM} = 0$). Given that the estimated mean baseline scores for the contaminating MC stratum ($\cong 0.685\sigma$) are larger than those of both the AR stratum ($\cong 0.332\sigma$) and the FM stratum ($\cong 0$), the ability selection component η_1 in Eq. (A3) is positive, leading to an upward bias in the IV estimate.

[18] Duplicate names are very unusual within the same grade in the same primary school. Non-matching is largely due to name misspellings or gender misidentifications in either the lottery or baseline score records.

[19] Expanding the target MSEE records to include students from the South Bank increases the overall match rate by only two percentage points but significantly increases the chances of multiple matches.

[20] Specifically, for a variable of interest (x), the lottery-specific mean for lottery j ($\bar{x}_j$) is adjusted to be the same as the mean for the pooled sample across all lotteries ($\bar{x}$).

[21] The sharp increase in precision is the result of a large reduction in the standard deviation of the variable compared, from 2.0 for the number of matches to only 0.3 for the matching rate.

Table 2
Predetermined individual characteristics and lottery assignment.

	Means			Coefficients		
	Full sample	District 3 subsample	District 3 subsample w/ baseline scores	Full sample	District 3 subsample	District 3 subsample w/ baseline scores
	(1)	(2)	(3)	(4)	(5)	(6)
Lottery winner	0.301	0.216	0.215	*Dependent variable: lottery winner*		
	(0.459)	(0.412)	(0.411)			
Female	0.469	0.493	0.489	−0.008	−0.013	−0.021
	(0.499)	(0.500)	(0.500)	(0.008)	(0.014)	(0.015)
6th-grade baseline scores available	–	0.854	1.000	–	−0.002	-
		(0.354)	(0.000)		(0.020)	
6th-grade baseline scores (in s.d.)	–	–	0.292	–	-	0.011
			(0.759)			(0.010)
F-statistics				F(176,13591) = 0.78	F(35,3447) = 0.58	F(34,2939) = 0.68
Prob > F				0.985	0.979	0.919
Number of applicants	13,768	3483	2973	13,768	3483	2973

Notes: Columns (1)–(3) report the mean of each variable indicated by the row heading for each sample. Columns (4)–(6) report the coefficients of a linear regression of the lottery winner dummy on the independent variables indicated by the row headings and the full set of primary school and lottery fixed effects. The F-statistics and Prob > F report, respectively, the F-test statistic and*p-value* for the test of the hypothesis that the coefficients on all of the predetermined individual characteristics including primary school fixed effects (but excluding lottery fixed effects) are zero. The numbers reported in parentheses are standard deviations in columns (1)–(3) and standard errors in columns (4)–(6).

3.3. Main empirical results

3.3.1. First-stage estimates

I begin by examining the effect of winning a lottery on elite school enrollment for the full sample in Panel A of Table 5. Column 1 presents the results of the first-stage regression that includes no controls except for lottery fixed effects using all linked lottery-MSEE pairs in Ψ. The coefficient on the lottery winner dummy shows that the MSEE records matched to winners (i.e., Ψ_1) are 19.7 percentage points more likely to pertain to students from the applicant's selected elite school than the records matched to losers (i.e., Ψ_0). As Eq. (1b) illustrates, this first-stage relationship is attenuated because of the presence of false matches in Ψ. Although the consistency property of the IV estimator is unaffected by this attenuation, one may still be interested in the first-stage relationship for the applicants only, which is not attenuated because of imperfect matching. Column 3 of Table 5 presents the results of such an exercise, in which the unit of analysis is changed to each successfully matched lottery applicant and the dependent variable to the number of matched MSEE records from the applicant's selected elite school (i.e., $\sum_{j:C_j=C_i} D_j$). Because the number of false matches from this school (i.e., $\sum_{j:C_j=C_i, j\neq i} D_j$) is independent of the applicant's lottery assignment, the win/loss difference in $\sum_{j:C_j=C_i} D_j$ is entirely attributable to the win/loss difference in the applicant's own enrollment status in the selected elite school. The coefficient on the lottery winner dummy in this regression indicates winners to be 34.0 percentage points more likely to enroll in their selected elite schools than losers. Note that in the absence of differential sample selection, the ratio of the attenuated first-stage estimate in column 1 and the unattenuated first-stage estimate in column 3 $\left(\frac{0.197}{0.340} = 0.579\right)$ corresponds to the proportion of correct matches in Ψ.

Table 3
Matching outcomes and sample selection by lottery assignment.

	Full sample			District 3 subsample w/ baseline scores		
	Winners' mean	Losers' mean	Win/loss difference	Winners' mean	Losers' mean	Win/loss difference
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Matching outcomes						
Number of matches (n_i)	1.592	1.567	0.025	1.905	1.880	0.025
	(1.960)	(2.046)	(0.038)	(2.254)	(2.405)	(0.106)
Matching rate ($n_i \geq 1$)	0.911	0.891	0.020***	0.956	0.939	0.017*
	(0.287)	(0.310)	(0.006)	(0.208)	(0.238)	(0.010)
Number of applicants	4138	9630	13,768	638	2335	2973
Panel B. Ability selection in baseline scores						
Unmatched applicants ($n_i = 0$)	–	–	–	0.046	0.224	−0.178#
				(0.732)	(0.643)	(0.134)
Number of applicants				29	141	170
Matched applicants only ($n_i \geq 1$)	–	–	–	0.322	0.291	0.032
				(0.693)	(0.715)	(0.033)
Number of applicants				609	2194	2803

Notes: Each row of the table reports the lottery-adjusted winners' and losers' means and their difference for the dependent variable denoted in the row heading. For each dependent variable, the mean of each lottery is adjusted to be the same as that of the pooled sample of all lotteries. The numbers reported in parentheses are standard deviations for means and standard errors for differences.

Significant at 20%.
* Significant at 10%.
*** Significant at 1%.

Table 4
Decomposition and ability selection proxy, District 3 subsample w/ baseline scores.

Panel A. Sample statistics and decomposition

	Sample statistics	Decomposition
	(1)	(2)
(1) Winners	Average number of matches: $\bar{n}_1 = 1.905$	$\bar{n}_1 = \theta_{AR} + \theta_{MC} + \theta_{FM}$
(1a) Unmatched	Proportional share: $p_{01} = 0.044$ Mean baseline scores: $\bar{a}_{01} = 0.046\sigma$	$p_{01} = \theta_{NRI_0}$
(1b) Matched	Proportional share: $p_{01} = 0.956$ Mean baseline scores: $\bar{a}_{11} = 0.322\sigma$	$p_{01} = \theta_{NRI_1} + \theta_{MC} + \theta_{AR}$
(2) Losers	Average number of matches: $\bar{n}_0 = 1.88$	$\bar{n}_0 = \theta_{AR} + \theta_{FM}$
(2a) Unmatched	Proportional share: $p_{00} = 0.061$ Mean baseline scores: $\bar{a}_{00} = 0.224\sigma$	$p_{00} = \theta_{NRI_0} + \theta_{MG_0}$
(2b) Matched	Proportional share: $p_{10} = 0.939$ Mean baseline scores: : $\bar{a}_{10} = 291\sigma$	$p_{10} = \theta_{NRI_1} + \theta_{MC_1} + \theta_{AR}$

Panel B. Estimates of (sub)stratum statistics

	Proportional share	Mean baseline scores size
	(1)	(2)
(1) NRI_0	$\theta_{NRI_0} = p_{01} = 0.044$	$\bar{a}_{NRI_0} = \bar{a}_{01} = 0.046\sigma$
(2) MC_0	$\theta_{MC} = p_{00} - p_{01} = 0.017$	$\bar{a}_{MC_0} = \frac{p_{00}\bar{a}_{00} - p_{01}\bar{a}_{01}}{p_{00}p_{01}} = 0.685\sigma.$
(3) MC	$\theta_{MC} = \bar{n}_1 - \bar{n}_0 - \theta_{AR} = 0.025$	$\bar{a}_{MC} \cong \bar{a}_{MC_0} = 0.695\sigma.$ (Proxy assumption 1: $\bar{a}_{MC_1} \cong \bar{a}_{MC_0}$)
(4) NRI_1	$\theta_{NRI_1} = \theta_{NRI_0}\frac{\theta_{NRI_1}}{\theta_{NRI_0}} \cong \theta_{NRI_0}\frac{\theta_{MC_1}}{\theta_{MC_0}}$ $= \theta_{NRI_0}\frac{\theta_{MC} - \theta_{MC_0}}{\theta_{MC_0}} = 0.021$ (Proxy assumption 2: $\frac{\theta_{NRI_1}}{\theta_{NRI_0}} \cong \theta_{NRI_0}\frac{\theta_{MC_1}}{\theta_{MC_0}}$)	
(5) AR	$\theta_{AR} = p_{11} - \theta_{NRI_1} - \theta_{MC} = 0.91$	$\bar{a}_{AR} \cong \bar{a}_{11} = 0.322\sigma$ $\left(\frac{\theta_{AR}}{p_{11}} = 0.952\right.$: vast majority of matched winners belong to the AR stratum$\left.\right)$
(6) FM	$\theta_{AR} = \bar{n}_0 - \theta_{AR} = 0.97$	$\bar{a}_{FM} \cong 0$ (Proxy assumption 3: no ability selection of the false matches.)

The considerable deviation of this coefficient from unity suggests a high degree of noncompliance as the result of either losers enrolling in their selected elite schools through the "back door" or winners declining their admission offers. Quantifying these two types of noncompliance requires information on each applicant's enrollment status at his/her selected elite school, which cannot be identified exactly as correct and false matches are indistinguishable in Ψ. Nonetheless, I conduct a proxy exercise in which I infer this information by whether an applicant's name appears among the MSEE takers from his/her selected elite school (conditional on gender and cohort), i.e., $\max\limits_{j:C_j=C_i}\{D_j\}$. Although false matching by name is quite common among a cohort of over 20,000 MSEE takers from the North Bank, such a possibility becomes very rare (albeit not entirely zero) when the target universe is restricted to a few hundred MSEE takers from a particular school. Therefore, the chances (if any) of mistakenly inferring an unenrolled applicant as enrolled are quite small. In column 5, the unit of analysis remains each successfully matched applicant, but the dependent variable changes to the inferred enrollment status at the selected elite school: 52.8% of the losers are inferred to have enrolled in their selected elite schools, whereas only 10.5% of the winners are inferred to have declined their admission offers. The regression-adjusted win/loss difference in the inferred enrollment status is 33.2 percentage points, very close to the 34.0-percentage-point difference found in column 3. Controlling for the applicant's gender in the even columns in Panel A of Table 5 has almost no effect on the estimated first-stage relationships throughout all three specifications, although the coefficients on the female dummy are always negative and significant, suggesting somewhat smaller chances of girls enrolling in their selected elite schools through the "back door."

Panel B of Table 5 reports separate results for the District 3 subsample with baseline scores. For all regressions, this subsample yields substantially larger estimates of the coefficient on the lottery winner dummy than the full sample, showing a higher degree of compliance. Moreover, when applicants' baseline scores are included as a control variable in the even columns, the coefficient is always positive and significant, indicating that (non)compliance is selective in baseline scores. To further investigate this selection, Figs. 1A and 1B plot the cumulative distribution of baseline scores by inferred enrollment status at the selected elite school for lottery winners and losers, respectively. For winners, the two cumulative distribution curves resemble each other closely: the two-sample Kolmogorov–Smirnov test statistics cannot reject the equality of the two distributions (with a *p-value* of 0.799), showing no evidence for selective (non)compliance. However, for losers, the curve for noncompliers (i.e., enrolled in the selected elite school) always lies below that for compliers (i.e., not enrolled in the selected elite school), suggesting more favorable treatment of higher ability losers in the "back door" admissions: the two-sample Kolmogorov–Smirnov test statistics reject the equality of the two distributions (with a *p-value* of 0.000) and suggest the first-order stochastic dominance of noncomplying losers over complying losers in baseline scores (with a *p-value* of 0.999).

3.3.2. IV estimates

The positive ability selection in the "back door" admissions of lottery losers indicates a positive correlation between elite school enrollment status and *ex ante* ability among lottery participants. To the extent that elite school attendees' advantage in *ex ante* ability remains at middle school exit, it yields a spurious positive relationship between applicants' elite school enrollments and *ex post* academic outcomes. This positive relationship carries over to the matched MSEE takers in Ψ, although it is somewhat attenuated by the presence of false matches. To circumvent this spurious positive

Table 5
First-stage estimates of lottery effect on elite school enrollment.

	Unit of analysis: each lottery-MSEE pair		*Unit of analysis: each successfully matched lottery applicant*			
	Dependent var: $D_{i(j)}$		*Dependent var:*sum $(D_{i(j)})$		*Dependent var:* max $\{D_{i(j)}\}$	
	w/o controls	w/ controls	w/o controls	w/ controls	w/o controls	w/ controls
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Full sample						
Lottery winner	0.197***	0.196***	0.340***	0.340***	0.332***	0.332***
	(0.007)	(0.007)	(0.009)	(0.009)	(0.009)	(0.009)
Female	–	−0.020***	–	−0.019**	–	−0.014*
		(0.006)		(0.008)		(0.008)
Loser's mean in the inferred enrollment status	–		–		0.528	
Winner's mean in the inferred enrollment status	–		–		0.895	
Number of lottery-MSEE pairs	21,676		–		–	
Number of applicants	12,347		12,347		12,347	
Panel B. District 3 subsample w/ baseline scores						
Lottery winner	0.258***	0.257***	0.519***	0.517***	0.498***	0.496***
	(0.014)	(0.014)	(0.021)	(0.021)	(0.020)	(0.020)
Female		0.002		0.001		−0.003
		(0.011)		(0.017)		(0.016)
Baseline scores	–	0.034***	–	0.055***	–	0.056***
		(0.008)		(0.012)		(0.011)
Loser's mean in the inferred enrollment status	–		–		0.396	
Winner's mean in the inferred enrollment status	–		–		0.936	
Number of lottery-MSEE pairs	5606		–		-	
Number of applicants	2803		2803		2803	

Notes: In columns (1) and (2), the unit of analysis is each linked lottery-MSEE pair and the dependent variable is whether the matched MSEE taker was enrolled in the applicant's selected elite school. In columns (3) and (4), the unit of analysis is each successfully matched lottery applicant, and the dependent variable is the number of matched MSEE records from the applicant's selected elite school. In columns (5) and (6), the unit of analysis is each successfully matched lottery applicant, and the dependent variable is whether there exists any matched MSEE record from the applicant's selected elite school. All regressions include lottery fixed effects. Standard errors are reported in parentheses.

* Significant at 10%.

** Significant at 5%.

*** Significant at 1%.

correlation, I use the applicant's lottery assignment as an instrument for the matched MSEE takers' elite school enrollment status in examining the effects of attending an elite school on students' *ex post* academic outcomes at middle school exit. Specifically, based on the MSEE scores and secondary school placements, I construct five measures of students' *ex post* academic outcomes: total scores on the MSEE, a dummy for admission to a top-echelon high school, a dummy for scoring above the minimum requirement for top-echelon high school admissions, a dummy for admission to any high school (including both top-echelon and regular high schools), and a dummy for scoring above the minimum requirement for regular high school admissions.[22] Note that within each type of secondary school (i.e., top-echelon high schools, regular high schools, and vocational secondary schools), the Boston mechanism is used for the assignment of school places to students. Under this mechanism, a school first admits students who list it as their first choice in descending order of MSEE scores, and then moves to those who rank it second only if there are places remaining after the first round, and so forth. It is well known that the Boston mechanism is not strategy proof.[23] Therefore, it is possible that elite school attendance may improve students' secondary school placements by playing more successful strategies in their school choices. The inclusion of the two dummy indicators for scoring above the two thresholds can help to examine whether elite school attendance improves their attendees' secondary school placements through channels beyond entrance exam scores.

Table 6 reports the results of the IV estimations using all linked lottery-MSEE pairs in Ψ. Each row uses a separate academic outcome as the dependent variable, whereas each column uses either a different specification or a different sample. All regressions include lottery fixed effects, and those in the even columns also include some covariates.[24] Columns 1–2 show the IV coefficients for the full sample, estimated without and with covariates. Eight of the ten coefficients are negative and the other two, although positive, are small and insignificant, showing little evidence of positive academic benefits of elite school attendance in general. For the achievement effects on the MSEE scores in particular, both coefficients (0.041 and −0.016) are negative and can reject a modest achievement gain of 0.15σ at the 10% level. However, in both specifications it is also worth noting the opposite sign of the coefficients on top-echelon high school admissions and scoring above the corresponding admissions threshold. Although the quantitative differences (0.027 vs. −0.005 in column 1 and 0.031 vs. −0.005 in column 2) are not significant, taking the signs at face value would indicate that attending an elite school helps to place students in top-echelon high schools, although it does not help to raise their MSEE scores above the threshold for top-echelon high school admissions. I delegate further discussion on this discrepancy to Section 3.3.4 on treatment heterogeneity, where such discrepancy is more salient for some subgroups. Columns 3–4 of Table 6 report the same IV estimates for the District

22 Students with MSEE scores below the minimum requirement for regular high school admissions either drop out of school after completing nine years of compulsory education or attend a vocational secondary school to obtain job-oriented training.

23 See Abdulkadiroglu and Sonmez (2003) for a more detailed discussion of the Boston mechanism and He (2014) for an example of its application to school choice in China.

24 Specifically, column 2 controls for gender and column 4 controls for both gender and baseline scores.

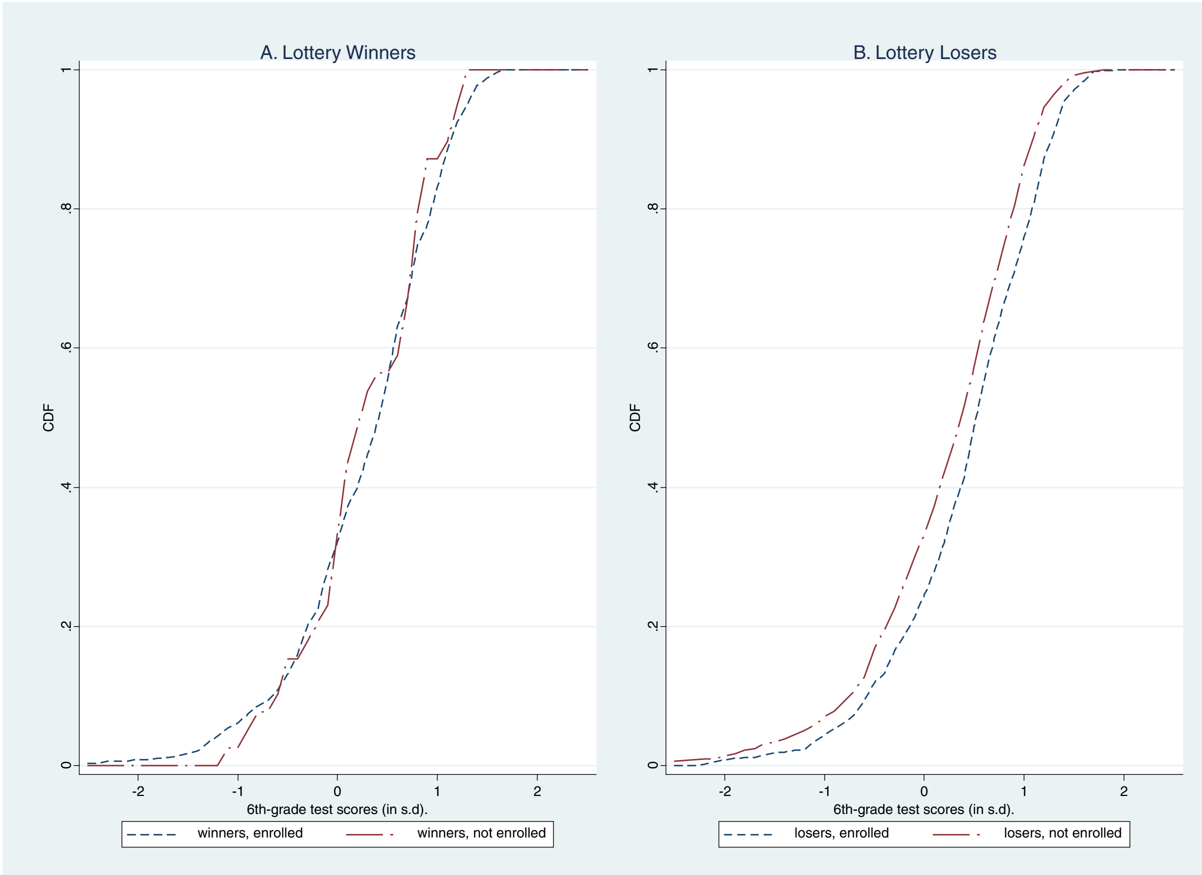

Fig. 1. Empirical distribution of baseline scores by lottery assignment and enrollment.
Notes: Panel A plots the cumulative distribution curves of baseline scores of lottery winners in District 3 by their inferred elite school enrollment status. Panel B plots the cumulative distribution curves of baseline scores of lottery losers in District 3 by their inferred elite school enrollment status.

3 subsample. The IV coefficients for this subsample are always positive and larger than the corresponding estimates for the full sample. However, these coefficients are too small to reject the null hypothesis that there is no academic benefit from attending an elite school in general, yielding qualitatively the same conclusion as the results of the full sample.

3.3.3. *Sensitivity to the extent of false matching*

To examine the sensitivity of the IV estimates to the extent of false matching, I subdivide lottery applicants into different subsamples that are subject to varying degrees of false matching in the data combination. Let q_i denote the frequency of the values of the common variables of individual i that appear in the lottery data set, i.e. $q_i = \sum_{i' \in I} 1(C_{i'} = C_i)$, and let l_i denote the number of Chinese characters used in individual i's name, which takes the value of 2, 3, or 4 in the sample. Note that both q_i and l_i are unaffected by individual i's instrument value Z_i but are highly correlated ($+$ for q_i and $-$ for l_i) with the number of false matches (nf_i) in the data combination. Based on the values of q_i and l_i, I partition lottery applicants into three subsamples that are subject to increasing degrees of false matching in the data combination as follows: (i) individuals whose names are unique (conditional on gender and cohort) in the lottery data set and contain three or four Chinese characters (i.e., $q_i = 1$, $l_i \geq 3$); (ii) individuals whose names are unique (conditional on gender and cohort) in the lottery data set and contain only two Chinese characters (i.e., $q_i = 1$, $l_i = 2$); and (iii) individuals whose names are non-unique (conditional on gender and cohort) in the lottery data set (i.e., $q_i \geq 2$).[25] As noted in Panel A of Table 7, these three subsamples account for 48.5%, 40.1%, and 11.4% of all matched applicants, respectively. The average number of matched MSEE records increases monotonically from only 1.043 for subsample 1, to 1.594 for subsample 2, and to 5.362 for subsample 3. (In comparison, for the full sample the average number of matches is 1.756.) As the pairwise matches for subsample 1 contain few false matches, the estimated first-stage relationship in the subset of Ψ for this subsample is 0.307, only slightly smaller than the unattenuated first-stage effect (0.340) estimated in column 3 of Table 5. However, because the pairwise matches for subsample 3 are predominantly false matches, the estimated first-stage effect in the subset of Ψ for this subsample (0.060) is seriously attenuated and is only approximately one-sixth of the true first-stage effect.

Panel B of Table 7 reports separate IV estimates for each subsample. Similar to Table 6, each row in Panel B of Table 7 uses a separate academic outcome as the dependent variable. For comparison, column 1 in Panel B of Table 7 replicates the IV coefficients in column 1 of Table 6 estimated from the specification that includes no controls except for lottery fixed effects, using the entire set of linked

[25] Note that there are only 121 applicants with $q_i \geq 2$ and $l_i \geq 3$. Because of the small number of applicants falling in this category, I consolidate them with those with $q_i \geq 2$ and $l_i = 2$ into a single subsample $q_i \geq 2$ here. The results remain qualitatively the same if this small number of applicants with $q_i \geq 2$ and $l_i \geq 3$ are instead grouped with those with $q_i = 1$ and $l_i \geq 3$.

Table 6
The IV estimates of the elite school treatment effects.

	Full sample		District 3 subsample w/ baseline scores	
	w/o controls	w/ controls	w/o controls	w/ controls
Dependent variable	(1)	(2)	(3)	(4)
MSEE scores (in s.d.)	−0.041	−0.016	0.052	0.038
	(0.102)	(0.100)	(0.184)	(0.182)
Admission to a top-echelon high school	0.027	0.031	0.048	0.029
	(0.035)	(0.035)	(0.056)	(0.053)
Scoring above the threshold for top-echelon high school admissions	−0.005	−0.005	0.021	0.001
	(0.033)	(0.033)	(0.044)	(0.043)
Admission to any high school	−0.022	−0.015	0.017	0.008
	(0.045)	(0.045)	(0.087)	(0.086)
Scoring above the threshold for high school admissions	−0.013	−0.008	0.084	0.067
	(0.042)	(0.042)	(0.093)	(0.091)
Number of lottery-MSEE pairs	21,676		5606	
Number of observations	12,347		2803	

*Notes:*Each cell in the table corresponds to a separate IV estimate of the elite school treatment effect on the dependent variable indicated in the row heading using the applicant's lottery assignment as an instrument for the elite school enrollment status of the matched MSEE record. All regressions include lottery fixed effects. The control variables are gender for column (2) and both gender and baseline scores for column (4). Robust standard errors clustered by middle school attended interacted with graduation year are reported in parentheses.

lottery-MSEE pairs in Ψ. Columns 2–4 in Panel B of Table 7 report the IV coefficients estimated from the same specification for each subsample using the respective subset of linked lottery-MSEE pairs in Ψ. As noted previously, in the absence of differential sample selection one minus the ratio of the attenuated first-stage effect in Ψ (or its subset) and the unattenuated first-stage effect (estimated to be 0.340 in column 3 of Table 5) corresponds to the proportion of false matches among all matches, and this proportion is estimated to be 42.1% for the full sample, 9.6% for subsample 1, 25.8% for subsample 2, and 82.3% for subsample 3. When only the subset of Ψ for subsample 1 is used, all coefficients in column 2 are similar in both size and precision to the corresponding coefficients in column 1, suggesting that the IV estimates are not sensitive to the increase in the proportion of false matches from below 10% in this subsample to over 40% in the full sample. While estimations using the subset of Ψ for subsample 2 consistently yield smaller coefficients and larger standard errors in column 3 than column 1, testing the equality of the corresponding coefficients always yields very large *p-values* and the larger standard errors in column 3 are attributable to the reduction in sample size, the effect of which is only partially offset by the slight increase in the first-stage relationship. However, when the subset of Ψ for subsample 3 is used in column 4, all coefficients become very imprecisely estimated with standard errors more than four times those in column 1, suggesting that the proposed multiple data source approach lacks statistical precision to produce informative estimates for this subsample, for which the pairwise matches in Ψ consist of predominantly (more than 80%) false matches. Besides the proportion of correct matches in Ψ, the working of the proposed multiple data source approach also depends on the size of Ψ and the underlying first-stage relationship among correct matches. *Ceteris paribus*,

Table 7
Sensitivity analysis for subsamples subject to varying extents of false matching.

	Full sample	Subsample 1	Subsample 2	Subsample 3
		($q = 1, l \geq 3$)	($q = 1, l = 2$)	($q \geq 2$)
	(1)	(2)	(3)	(4)
Panel A. Matching statistics and first-stage estimates				
Proportional share of all matched applicants	100.0%	48.5%	40.1%	11.4%
Average number of matches	1.756	1.043	1.594	5.362
Estimated first-stage effect	0.197	0.307	0.252	0.060
	(0.023)	(0.044)	(0.026)	(0.014)
Estimated proportional share of false matches	42.1%	9.6%	25.8%	82.3%
Panel B. IV estimates				
MSEE scores (in s.d.)	−0.041	0.008	−0.137	0.395
	(0.102)	(0.098)	(0.120)	(0.443)
Admission to a top-echelon high school	0.027	0.018	0.006	0.156
	(0.035)	(0.043)	(0.050)	(0.189)
Scoring above the threshold for top-echelon high school admissions	−0.005	−0.033	−0.012	0.172
	(0.033)	(0.036)	(0.042)	(0.173)
Admission to any high school	−0.022	0.002	−0.079	0.196
	(0.045)	(0.042)	(0.058)	(0.213)
Scoring above the threshold for high school admissions	−0.013	0.003	−0.039	0.129
	(0.042)	(0.043)	(0.054)	(0.210)
Number of lottery-MSEE pairs	21,676	6243	7899	7534
Number of applicants	12,347	5985	4957	1405

Notes: Each column of this table uses a different sample of matched lottery-MSEE pairs defined by the column heading. Panel A reports the matching statistics and first-stage estimates of each sample. Panel B conducts IV estimations of the elite school treatment effects using the applicant's lottery assignment as an instrument for the elite school enrollment status of the matched MSEE record. Each cell in Panel B corresponds to a separate IV estimate for the dependent variable indicated in the row heading. All regressions control for lottery fixed effects. Robust standard errors clustered by middle school attended interacted with graduation year are reported in parentheses.

the statistical precision of the IV estimation using pairwise matches in Ψ increases with the size of Ψ, the proportion of correct matches in it, and the degree of instrument compliance among correct matches.

3.3.4. Treatment heterogeneity

Table 8 examines the heterogeneity in the effects of elite school attendance on all five academic outcomes by applicants' gender (for the full sample) and relative position in their selected elite school in terms of baseline scores (for the District 3 subsample). For each outcome indicated in the row heading, columns 1 and 2 present separate IV estimates of the effect of elite school attendance for boys and girls, and column 3 reports the *p-value* of the test for the equality of the two coefficients. Although the coefficients are never significantly different from each other at conventional levels (with the smallest *p-value* being 0.131), they are always positive for girls but negative for boys. For the dummy indicator for top-echelon high school admissions in particular, both the coefficient for girls itself and the difference between it and the coefficient for boys are significant at the 15% level, providing suggestive evidence that attending an elite school may increase the chances of top-echelon high school admissions, but only for girls. This finding is consistent with previous work showing that girls benefit academically more than boys from a better schooling environment (Lavy et al., 2012; Deming et al., 2014).

As discussed in Section 3.1, the learning environment at elite schools may be more beneficial to stronger students than weaker students. To test treatment heterogeneity by applicants' relative initial achievement at middle school entry, I subdivide applicants in the District 3 subsample by whether their baseline scores are above or below the median scores of attendees in their selected elite school in the same cohort, and conduct and compare separate estimates in columns 4–6 of Table 8. Although rather imprecisely estimated due to the small sample size, the coefficients for applicants whose baseline scores fall in the top half in their selected elite school are always positive and larger than the corresponding coefficients for those in the bottom half, which are often negative. Again, for the dummy indicator for top-echelon high school admissions in particular, the coefficient is positive with a *p-value* of less than 20% for the top half but negative for the bottom half. The equality of the two coefficients can be rejected at the 10% level, suggesting that attending an elite school may benefit stronger applicants in terms of top-echelon high school admissions, but not weaker ones. It is also worth noting that for both girls and stronger applicants, when the dummy indicator for scoring above the threshold for top-echelon high school admissions is used as the outcome variable, the IV coefficients become much smaller (although still positive) and statistically indistinguishable from zero even at very generous levels of significance. Besides achievement improvement, a plausible alternative channel for elite schools to improve their attendees' top-echelon high school admissions is to help students to make more informed school-choice decisions and/or play safer strategies, both of which could increase the chances of gaining admission to a top-echelon high school conditional on test scores under the Boston mechanism used in the city. For both girls and stronger applicants, the quantitative differences in the coefficients on top-echelon high school admissions and scoring above the corresponding admissions threshold seem to suggest that the placement benefits of elite school attendance on top-echelon high school admissions (if indeed any) are driven more by school-choice strategy improvements than by achievement gains.

3.3.5. Distributional treatment effects

For students' test scores on the MSEE, another way to examine treatment heterogeneity is to assess the distributional treatment effects by comparing the distributions of potential MSEE scores for compliers with and without elite school attendance. Given that the lottery data set I contains different admissions lotteries with varying winning probabilities and Z_i is randomly assigned only within each lottery, Assumptions (A2), (A3), and (A5) only hold in their conditional versions after controlling for lottery membership. Let k denote a lottery index, $I(k)$ denote the subset of applicants in I from lottery k, and $\Psi(k)$ denote the subset of linked lottery-MSEE pairs in Ψ for applicants from lottery k, which can be further partitioned into $\Psi_1(k)$ and $\Psi_0(k)$ for applicants from lottery k with $Z_i = 1$ and $Z_i = 0$, respectively. Although under the aforementioned conditional assumptions, the propositions presented in Section 2 can only be applied to each subset of lottery applicants $I(k)$ to identify the lottery-specific parameter of interest for the subpopulation of compliers from lottery k (i.e., $D_i(1) > D_i(0), i \in I(k)$), the corresponding parameter of interest for all compliers (i.e., $D_i(1) > D_i(0), i \in I$) can still be identified by aggregating these lottery-specific parameters. Specifically, assuming there is no differential sample selection

Table 8
Treatment heterogeneity analysis by gender and relative position in baseline scores.

	Full sample			District 3 subsample w/ baseline scores		
	Elite school treatment effect		Test (girls = boys)	Elite school treatment effect		Test (bottom = top)
	Boys	Girls	$p-value$	Bottom half	Top half	$p-value$
Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)
MSEE scores (in s.d.)	−0.072	0.061	0.412	−0.091	0.054	0.642
	(0.116)	(0.138)		(0.248)	(0.263)	
Admission to a top-echelon high school	−0.015	0.083†	0.136†	−0.051	0.141#	0.099*
	(0.040)	(0.054)		(0.052)	(0.109)	
Scoring above the threshold for top-echelon high school admissions	−0.019	0.013	0.611	−0.026	0.033	0.557
	(0.039)	(0.050)		(0.055)	(0.081)	
Admission to any high school	−0.029	0.009	0.620	−0.059	0.052	0.490
	(0.054)	(0.062)		(0.100)	(0.144)	
Scoring above the threshold for high school admissions	−0.055	0.048	0.131†	0.015	0.103	0.557
	(0.050)	(0.057)		(0.109)	(0.139)	
Number of lottery-MSEE pairs	11,207	10,469	21,676	3052	2554	5606
Number of observations	6509	5838	12,347	1469	1334	2803

*Notes:*Each cell of this table reports the IV estimate of the elite school treatment effect (using lottery assignments as an instrument) on each dependent variable indicated by the row heading for a subsample defined by the column heading. All regressions control for lottery fixed effects, whereas those in columns (3) and (4) further control for a female dummy and baseline scores. Robust standard errors clustered by middle school attended interacted with graduation year are reported in parentheses.

Significant at 20%.

† Significant at 15%.

* Significant at 10%.

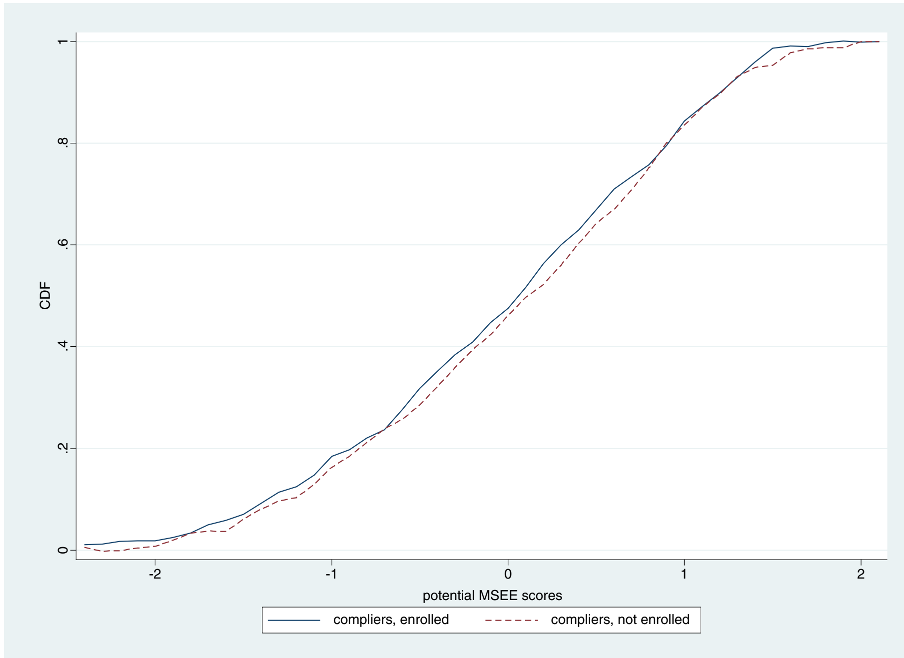

Fig. 2. Estimated distributions of potential MSEE scores for compliers by treatment status.

(i.e., no marginal compliers), Proposition 2 can be applied to each subset $I(k)$ to identify the lottery-specific distributions $E[1(Y_i(1) \leq y)|D_i(1) > D_i(0), i \in I(k)]$ and $E[1(Y_i(0) \leq y)|D_i(1) > D_i(0), i \in I(k)]$, which I denote respectively as $F^C_{(1),k}(y)$ and $F^C_{(0),k}(y)$. Then, the distributions of potential MSEE scores for all compliers $F^C_{(1)}(y)$ and $F^C_{(0)}(y)$ can be expressed as the weighted average of the lottery-specific distributions $F^C_{(1),k}(y)$ and $F^C_{(0),k}(y)$, with the weights equal to each lottery's proportional share of compliers, i.e.,

$$\begin{aligned} F^C_{(1)}(y) &= \sum_k \Pr(i \in I(k)|D_i(1) > D_i(0), i \in I) E[1(Y_i(1) \leq y)|D_i(1) \\ &\quad > D_i(0), i \in I(k)] \\ &= \sum_k \frac{N_k \delta_k}{\sum_k N_k \delta_k} F^C_{(1),k}(y) \end{aligned} \tag{4a}$$

and

$$\begin{aligned} F^C_{(0)}(y) &= \sum_k \Pr(i \in I(k)|D_i(1) > D_i(0), i \in I) E[1(Y_i(0) \leq y)|D_i(1) \\ &\quad > D_i(0), i \in I(k)] \\ &= \sum_k \frac{N_k \delta_k}{\sum_k N_k \delta_k} F^C_{(0),k}(y), \end{aligned} \tag{4b}$$

where N_k denotes the total number of linked record pairs in $\Psi(k)$ and δ_k denotes the first-stage effect of Z on D in $\Psi(k)$, i.e., $\delta_k = E[D_{i(j)}|Z_{i(j)} = 1, i \in I(k)] - E[D_{i(j)}|Z_{i(j)} = 0, i \in I(k)]$.

Fig. 2 plots the results of using Eqs. (4a) and (4b) to aggregate the lottery-specific estimates $\hat{F}^C_{(1),k}(y)$ and $\hat{F}^C_{(0),k}(y)$ obtained from applying Proposition 2 to each subset $\Psi(k)$.[26] At the lower quantiles of MSEE scores $\hat{F}^C_{(1)}(y)$ always lies above $\hat{F}^C_{(0),k}(y)$, whereas at the upper quantiles of MSEE scores the two curves converge and become indistinguishable. The differences between the two curves in Fig. 2, if taken at face value, would suggest that elite school attendance yields negative achievement effects to low-scoring compliers but has no achievement effects on high-scoring compliers.

Next, I perform distributional hypotheses testing for $F^C_{(1)}(y)$ and $F^C_{(0)}(y)$ under stratified randomization with imperfect matching. Note that Eq. (2d) of Proposition 2 shows that under a simple randomization with imperfect matching, $F^C_{(1)}(y) - F^C_{(0)}(y)$ can be expressed as the difference in the distributions of outcomes between Ψ_1 and Ψ_0 (i.e., $F_1(y) - F_0(y)$) multiplied by a constant (i.e., the inverse of the first-stage relationship in Ψ), and thus testing distributional hypotheses for $F^C_{(1)}(y)$ and $F^C_{(0)}(y)$ is equivalent to testing the same distributional hypotheses for $F_1(y)$ and $F_0(y)$, both of which are observed directly. However, this property does not carry over exactly under stratified randomization, because under stratified random assignment of the instrument the proportional share of $\Psi_1(k)$ in Ψ_1 generally differs from the proportional share of $\Psi_0(k)$ in Ψ_0. Thus, to the extent that lottery membership is correlated with the distribution of outcomes, differences between $F_1(y)$ and $F_0(y)$ may arise from the differences between Ψ_1 and Ψ_0 in their lottery compositions (even in the absence of treatment effects). Nonetheless, with stratified random assignment of the instrument, $F^C_{(1)}(y) - F^C_{(0)}(y)$ can still be expressed as the difference between two *reweighted* distributions of outcomes in Ψ_1 and Ψ_0 multiplied by a constant as follows:

$$F^C_{(1)}(y) - F^C_{(0)}(y) = C \cdot \left(F_1\left(y; \frac{1}{\pi_k}\right) - F_0\left(y; \frac{1}{1-\pi_k}\right) \right) \tag{5}$$

where $C = \frac{\sum_k N_k}{\sum_k N_k \delta_k}$ is the inverse of the (pooled) first-stage relationship in Ψ; $\pi_k = Pr(\Psi_1(k)|\Psi(k))$ is the lottery-specific proportional share of $\Psi_1(k)$ in $\Psi(k)$; $F_1\left(y; \frac{1}{\pi_k}\right)$ is the reweighted cumulative distribution of Y in Ψ_1 with a lottery-specific weight $\frac{1}{\pi_k}$; and $F_0(y; \frac{1}{1-\pi_k})$ is the reweighted cumulative distribution of Y in Ψ_0 with a lottery-specific weight $\frac{1}{1-\pi_k}$. The derivation of Eq. (5) is formulated in Appendix D. Note that Eq. (2d) in Proposition 2 can be considered

[26] Note that in finite samples, the IV estimates of potential cumulative distributions for compliers may not be non-decreasing, as discussed in Imbens and Rubin (1997).

as a special case of Eq. (5) when $k = 1$. Given Eq. (5), testing distributional hypotheses between $F^C_{(1)}(y)$ and $F^C_{(0)}(y)$ is thus equivalent to testing the same distributional hypotheses between the two reweighted distributions $F_1\left(y; \frac{1}{\pi_k}\right)$ and $F_0\left(y; \frac{1}{1-\pi_k}\right)$. In a setting with simple randomization and perfect observation of (Z, D, Y), Abadie (2002) suggests using Kolmogorov–Smirnov type nonparametric tests of the distributional hypotheses for $F^C_{(1)}(y)$ and $F^C_{(0)}(y)$ and proposes a bootstrap strategy to approximate the distributions of test statistics under the null hypotheses. Following the same spirit as Abadie (2002), I define three nonparametric distance test statistics for three distributional hypotheses for $F^C_{(1)}(y)$ and $F^C_{(0)}(y)$ as follows:

H1. Equality of distributions $\left(F^C_{(1)}(y) = F^C_{(0)}(y)\,\forall\, y \in R\right)$:

$$T^{eq} = \sup_{y \in R}\left|F_1\left(y; \frac{1}{\pi_k}\right) - F_0\left(y; \frac{1}{1-\pi_k}\right)\right|;$$

H2. $F^C_{(1)}(y)$ weakly first-order stochastically dominates $F^C_{(0)}(y)$ $\left(F^C_{(1)}(y) \le F^C_{(0)}(y)\,\forall\, y \in R\right)$:

$$T^{1d0} = \sup_{y \in R}\left(F_1(y; \frac{1}{\pi_k}) - F_0(y; \frac{1}{1-\pi_k})\right);$$

H3. $F^C_{(0)}(y)$ weakly first-order stochastically dominates $F^C_{(1)}(y)$ $\left(F^C_{(0)}(y) \le F^C_{(1)}(y)\,\forall\, y \in R\right)$:

$$T^{0d1} = \sup_{y \in R}\left(F_0\left(y; \frac{1}{1-\pi_k}\right) - F_1\left(y; \frac{1}{\pi_k}\right)\right).$$

To approximate the distributions of these test statistics under the null hypotheses, I adapt the bootstrap strategy proposed by Abadie (2002) to the setting encountered with both imperfect matching and stratified randomization considered in the current paper as follows.

Step 1: From the original combined data set Ψ, for each lottery k count the total number of observations in $\Psi(k)$, $\Psi_1(k)$, and $\Psi_0(k)$, and denote them as N_k, N_{1k}, and N_{0k}, respectively. Note that by definition $N_{1k} = N_k\pi_k$ and $N_{0k} = N_k(1-\pi_k)$.
Step 2: From the original combined data set Ψ, compute the two reweighted distributions $F_1\left(y; \frac{N_k}{N_{1k}}\right)$ and $F_0\left(y; \frac{N_k}{N_{0k}}\right)$ using Ψ_1 and Ψ_0, respectively. Compute the empirical test statistic T based on $F_1\left(y; \frac{N_k}{N_{1k}}\right)$ and $F_0\left(y; \frac{N_k}{N_{0k}}\right)$, where T is a generic notation for T^{eq}, T^{1d0}, or T^{0d1}.
Step 3: Perform a stratified resampling of the original combined data set Ψ as follows: (i) for each subset $\Psi(k)$, resample N_k observations $\left(\hat{Y}_{1,k}, \hat{Y}_{2,k}, \ldots, \hat{Y}_{N_k,k}\right)$ from $\left(Y_{1,k}, Y_{2,k}, \ldots, Y_{N_k,k}\right)$ with replacement; (ii) for each resampled subset $\hat{\Psi}(k)$, assign the first N_{1k} elements of $\left(\hat{Y}_{1,k}, \hat{Y}_{2,k}, \ldots, \hat{Y}_{N_k,k}\right)$ to $\hat{\Psi}_1(k)$ and the remaining N_{0k} elements of $\left(\hat{Y}_{1,k}, \hat{Y}_{2,k}, \ldots, \hat{Y}_{N_k,k}\right)$ to $\hat{\Psi}_0(k)$; and (iii) pool $\hat{\Psi}_1(k)$ for all k to construct $\hat{\Psi}_1$, and pool $\hat{\Psi}_0(k)$ for all k to construct $\hat{\Psi}_0$. Use the two generated samples $\hat{\Psi}_1$ and $\hat{\Psi}_0$ to compute the two reweighted distributions $\hat{F}_1\left(y; \frac{N_k}{N_{1k}}\right)$ and $\hat{F}_0\left(y; \frac{N_k}{N_{0k}}\right)$. Compute the test statistic $\hat{T}_b$ based on $\hat{F}_1\left(y; \frac{N_k}{N_{1k}}\right)$ and $\hat{F}_0\left(y; \frac{N_k}{N_{0k}}\right)$, where $\hat{T}_b$ is generic notation for $\hat{T}^{eq}_b, \hat{T}^{1d0}_b$, or $\hat{T}^{0d1}_b$.
Step 4: Repeat step 3, B times. Note that N_k, N_{1k} , and N_{0k} are all constant across bootstrap repetitions.
Step 5: Calculate the *p-values* of the tests with *p-value* $= \sum_{b=1}^{B}(1(\hat{T}_b \ge T))/B$. Reject the null hypothesis if the *p-value* is smaller than some significance level α.

Following the procedure described above, I perform stratified bootstrap resampling of the original combined data set Ψ 500 times and compute the distributions of the test statistics $\left\{(\hat{T}^{eq}_b, \hat{T}^{1d0}_b, \hat{T}^{0d1}_b)\right\}_{b=1}^{500}$ under the null hypotheses of (H1)–(H3). Column 1 of Table 9 reports the *p-values* for the empirical test statistics for (H1)–(H3). The test of (H1) yields a *p-value* of 0.788, suggesting little evidence to reject the null hypothesis that the two distributions are equal. Consistent with the pattern in Fig. 2 that $\hat{F}^C_{(0)}(y)$ almost always lies below $\hat{F}^C_{(1)}(y)$, the test of (H3) generates the largest *p-value* (0.964) and is most favorable for the null hypothesis of the weak first-order stochastic dominance of $F^C_{(0)}(y)$ over $F^C_{(1)}(y)$. While the test of (H2) produces the smallest *p-value* (0.408) and is least favorable for the null hypothesis, it still cannot reject the weak first-order stochastic dominance of $F^C_{(1)}(y)$ over $F^C_{(0)}(y)$ at conventional significance levels. Taken together, this exercises yields little evidence for the existence of any distributional consequences of elite school attendance on MSEE scores.

3.4. Bounds analysis

Note that the empirical analysis in Section 3.3 ignores the small degree of differential sample selection between winners and losers documented in Section 3.2 and the resulting contamination of the *MC* stratum in Ψ_1. Moreover, given Assumptions (A6) and (A7), all marginal compliers are treated and thus pertain to the subset Ψ_{11}

Table 9
Tests on distributional hypotheses of elite school treatment effects, *p-values*.

	Untrimmed data	Upper-tail trimmed data	Lower-tail trimmed data
	(1)	(2)	(3)
(H1) *Equality of distributions* $(F^C_{(1)}(y) = F^C_{(0)}(y))$	0.788	0.070	0.480
(H2) *Weak first-order stochastic dominance of treatment* $\left(F^C_{(1)}(y) \le F^C_{(0)}(y)\right)$	0.408	0.030	0.598
(H3) *Weak first-order stochastic dominance of non-treatment* $\left(F^C_{(1)}(y) \ge F^C_{(0)}(y)\right)$	0.964	1.000	0.260

*Notes:*Each cell of the table reports the *p-value* for the empirical test statistic compared to the distribution of test statistics under the null hypothesis computed from 500 repetitions of bootstrapped resampling with replacement described in Section 3.3.5. In column (1), both the empirical test statistic and the bootstrapped distribution of test statistics under the null hypothesis are constructed based on the original data. In column (2), both the empirical test statistic and the bootstrapped distribution of test statistics under the null hypothesis are constructed based on the upper-tail trimmed data. In column (3), both the empirical test statistic and the bootstrapped distribution of test statistics under the null hypothesis are constructed based on the lower-tail trimmed data.

Table 10
Bounds on the average treatment effect of elite school attendance.

	Lower bound	Upper bound
Dependent variable	(1)	(2)
MSEE scores (in s.d.)	−0.190	0.121
	(0.237)	(0.172)
Admission to a top-echelon high school	−0.032	0.049
	(0.093)	(0.044)
Scoring above the threshold for top-echelon high school admissions	−0.062	0.014
	(0.076)	(0.040)
Admission to any high school	−0.059	0.024
	(0.081)	(0.066)
Scoring above the threshold for high school admissions	−0.048	0.035
	(0.073)	(0.072)

*Notes:*The estimated lower and upper bounds are obtained by adapting Lee's (2009) trimming procedure to settings with imperfect matching and stratified randomization as described in Section 3.4. Standard errors reported in parentheses are calculated from 500 bootstrap replications of these coefficients.

consisting of treated record pairs matched to lottery winners. In this subsection, I conduct a partial identification exercise by constructing bounds on the treatment effects for compliers who are always retained in Ψ regardless of their lottery assignments (i.e., the AR_c substratum). Specifically, this exercise is implemented as follows. First, for each lottery k, the number of marginal compliers in Ψ_1 is estimated as $p_1^m N_{1k}$, where p_1^m is calculated from Eq. (3a) and N_{1k} is the total number of matched record pairs in Ψ_1 for lottery k. Second, for each lottery k, the lower/upper bound on the distribution of outcomes for the latent subset Ψ_{11}^* that excludes marginal compliers is constructed by trimming the upper/lower tail of the outcome distribution for the observed subset Ψ_{11} corresponding to this number. Third, Proposition 4 is applied to the trimmed data to yield bounds on the lottery-specific parameters of interest, which are then aggregated to generate the corresponding parameters of interest for all compliers in the AR stratum.

Columns 2 and 3 of Table 9 report the *p-values* for the empirical test statistics for (H1)–(H3), constructed using the upper- and lower-tail trimmed data, respectively. In column 2, the test statistics are constructed by comparing the point estimate of $F_0^C(y)$ to the least favorable estimate of $F_1^C(y)$ constructed after trimming the upper tail of the distribution of MSEE scores in Ψ_{11}. The tests of both (H1) and (H2) yield *p-values* of less than 10%, suggesting that $F_0^C(y)$ strictly first-order stochastically dominates the lower bound of $F_1^C(y)$. In column 3, none of the three null hypotheses can be rejected at the conventional significance levels, indicating a lack of evidence for any positive distributional treatment effects of elite school attendance on MSEE scores even with the use of the most favorable estimate of $F_1^C(y)$. Table 10 reports estimates of the lower and upper bounds on the ATE of elite school attendance for each of the five academic outcomes. For every outcome considered, the estimated lower- and upper-bound interval always includes zero, yielding an inconclusive bounding inference on the sign of the ATE, which is not surprising given that the corresponding simple IV coefficient estimated without trimming in column 1 of Table 6 is always small and insignificant. The bootstrapped standard errors of the estimated bounds also indicate large confidence intervals for these identified sets.[27]

[27] The literature considers alternative approaches to constructing confidence intervals (CIs) for partially identified parameters (see, e.g., Imbens and Manski, 2004; Chernozhukov et al., 2007). One approach constructs CIs that cover the entire identified set with a fixed probability. The other approach constructs CIs that cover the true value of the parameter with a fixed probability, which are strictly tighter than those constructed in the former approach. However, regardless of which approach is used here, the CIs are always too wide to be informative.

3.5. School choice and school competition

The results of this paper are also relevant to the broader debate over school choice,[28] as the magnet-type elite schools investigated herein provide schooling alternatives for students to opt out of their assigned neighborhood schools. Proponents of school choice argue that increasing parental choice can improve educational outcomes by raising the demand for schools that are more effective in a value-added sense. This argument hinges on the assumption that *parents are able to identify and indeed attempt to enroll their children in more effective schools.* However, this assumption has been challenged on several grounds in the literature. First, parents may choose schools for reasons other than learning effectiveness, which will dilute the incentives for efficiency competition that the choice mechanism might otherwise create. The existing literature provides both theoretical and empirical evidence consistent with parents choosing schools for reasons other than academic concerns. For example, MacLeod and Urquiola (2015) present a model of school competition and show that even without peer externalities, a preference for good peers can emerge endogenously from reputational concerns arising from signaling under imperfect information, and Jacob and Lefgren (2007), Hastings et al. (2009), and Cellini et al. (2010) present evidence that parents choose schools based on racial composition, school facilities, or student satisfaction. Second, parents may lack information or face high decision-making costs to act on their academic preference, as suggested by studies showing that receiving simplified school performance information has significant effects on school/housing markets. For example, Hastings and Weinstein (2008) find that providing simplified information on school performance leads more parents to choose higher-achieving schools in Charlotte-Mecklenburg, and Figlio and Lucas (2004) show that both residential choices and house prices in Florida respond to the assignment of school grades, even though the informational content used in the assignment is made public before the introduction of the grading system. Third, while solid research yields consistent evidence that parents value schools' achievement levels in making school choice and residential location decisions, there is less evidence that such decisions are influenced by schools' value-added effects on achievement. For example, Downes and Zabel (2002) show that the value of neighborhood schools in Chicago, as capitalized in house prices, is based on schools' mean contemporaneous test scores but not on the value-added they produce. in the Chilean context, Mizala and Urquiola (2013) also fail to find consistent evidence that information on schools' effectiveness affects market outcomes such as enrollment, tuition levels, and socioeconomic composition.

To shed light on this debate, among the sample of elite schools investigated in this paper, I examine the empirical relationships between a school's popularity, value-added, and achievement level. I construct two alternative measures to represent a school's popularity: the oversubscription rate[29] (ratio of total number of applicants to general admissions quota) and the winner take-up rate

[28] Recent years have seen a surge in the empirical literature investigating the effects of school choice on student outcomes, and the debate has permeated various cultural contexts, including Chile (Hsieh and Urquiola, 2006), China (Lai et al., 2011; He, 2014), Colombia (Angrst et al., 2002; Angrist et al., 2006; Bettinger et al., 2010), and Israel (Lavy, 2010).

[29] As each student can only apply to one elite school, a school's oversubscription rate reflects both its popularity in parents' true preferences and the strategic behavior of parents in selecting which school to apply to. With some parents not choosing their most preferred school, differences in schools' oversubscription rates are a compressed transformation of differences in schools' popularities in parents' true preferences. Nonetheless, such a transformation, as long as it is rank-preserving, should still reflect the relative popularities of these elite schools.

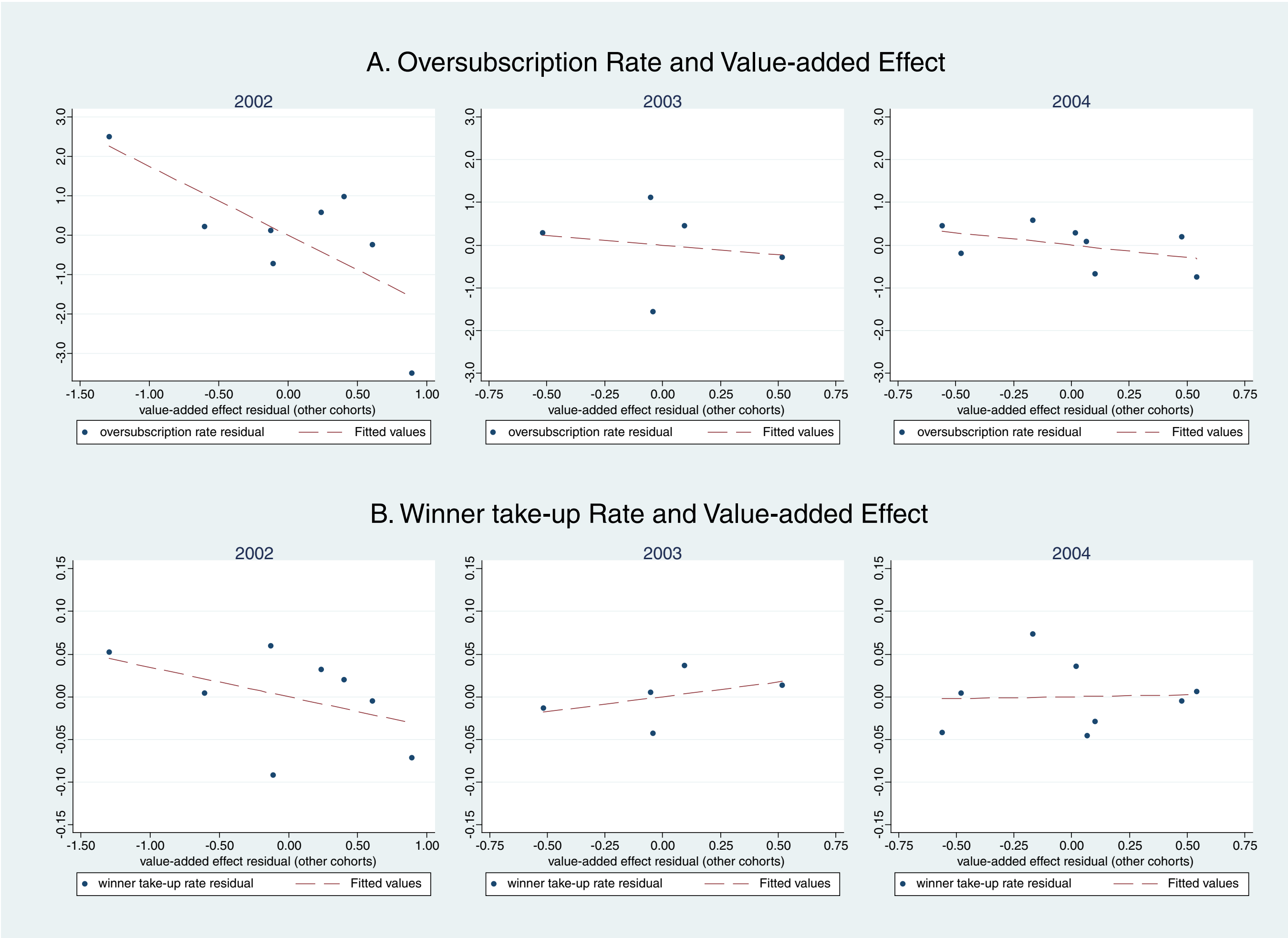

Fig. 3. School popularity and value-added effect.
Notes: Panel A plots the school's residual oversubscription rate in the contemporary cohort against the school's residual value-added effect on lottery compliers from other cohorts after controlling for district fixed effects. Panel B plots the school's residual winner take-up rate in the contemporary cohort against the school's residual value-added effect on lottery compliers from other cohorts after controlling for district fixed effects. For both panels, the 2003 entering cohort graph contains only five schools as three schools with incomplete lottery records in 2003 are excluded from the analysis throughout the paper.

(proportion of lottery winners enrolled). For a school's value-added, I use the estimated ATE on the MSEE scores for the school's lottery compliers. While the generalization of this estimated ATE to all of the school's applicants must be undertaken with considerable caution, it is nonetheless the only internally valid measure of a school's value-added given the assumptions maintained in this paper. To examine the relationship between a school's popularity and value-added, Fig. 3 plots the two sets of school popularity measures in the contemporary cohort against the school's estimated value-added effect on lottery compliers from other cohorts.[30] If parents indeed prefer and are able to identify high value-added schools, one would expect to see higher value-added schools possessing heavier oversubscription rates and higher take-up rates among lottery winners. Contrary to this expectation, four out of six panels in Fig. 3 illustrate a negative relationship between a school's popularity and estimated value-added, with the remaining two showing only weak positive correlations. Analogously, Fig. 4 plots the two sets of school popularity measures in the contemporary cohort against the average MSEE scores of attendees from other cohorts. For all three cohorts, the two popularity measures are both found to be positively correlated with the average MSEE scores, although the strength of the positive correlation decreases over time. The decline in the relative popularity of higher-achieving elite schools is consistent with two possibilities. The first is that over time, parents may have learned that higher-achieving elite schools are no more advantageous in yielding larger achievement gains, although they are still unable to identify more effective schools in the value-added sense, as the results in Fig. 3 indicate. The second is that under the restriction that each student can apply to only one elite school, parents may have learned to play safer school-choice strategies to shift away from higher-achieving elite schools that used to be more highly sought-after and have lower winning rates.

[30] As elite schools compete with each other mainly within district, all plotted relationships in this subsection (i.e., Figs. 3-5) control for district fixed effects. Also, the 2003 entering cohort sample consists of only five schools because three schools with incomplete lottery records in 2003 are excluded from the sample, as noted previously.

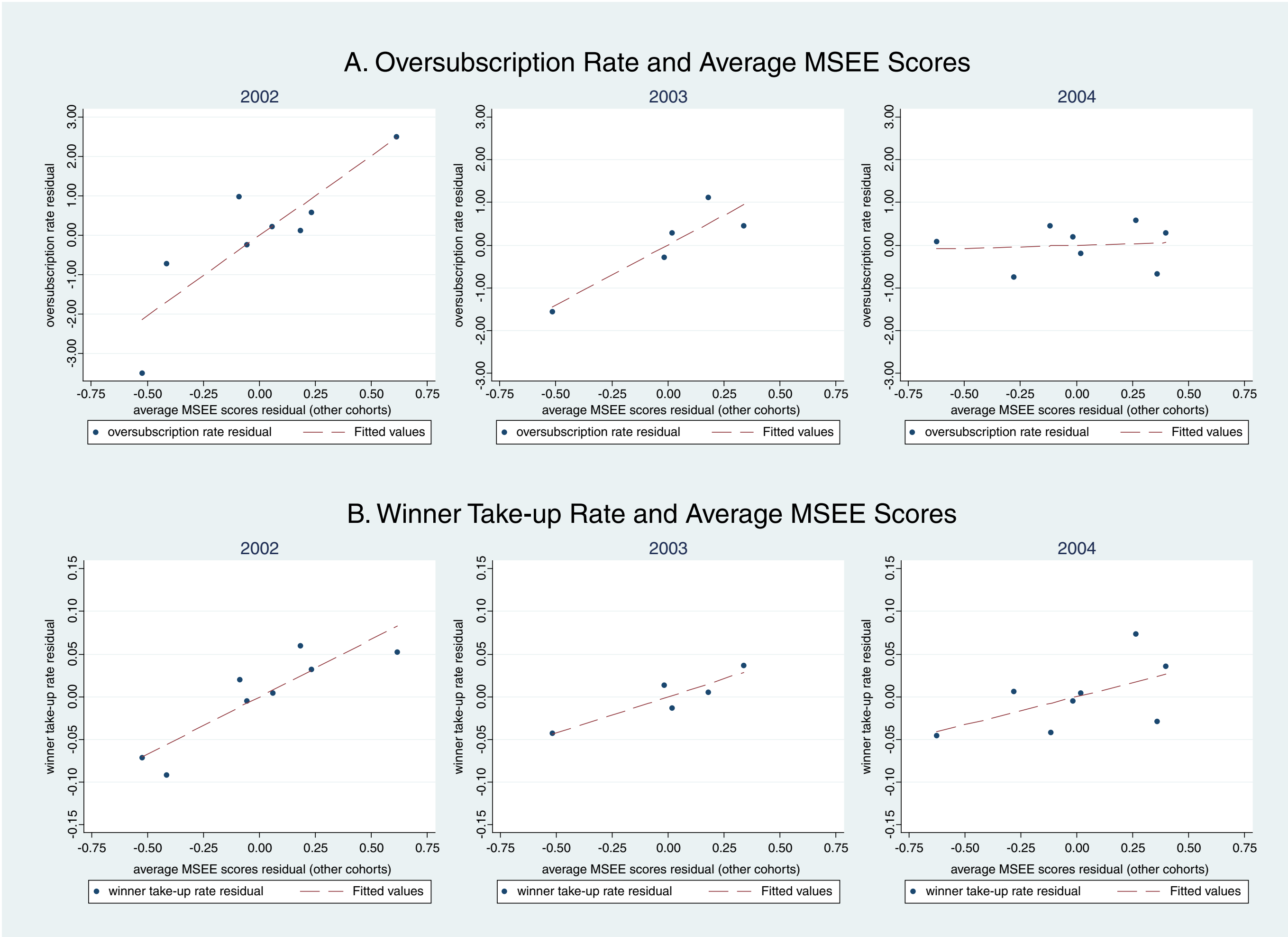

Fig. 4. School popularity and average MSEE scores.
Notes: Panel A plots the school's residual oversubscription rate in the contemporary cohort against the school's residual average MSEE scores of attendees from other cohorts after controlling for district fixed effects. Panel B plots the school's residual winner take-up rate in the contemporary cohort against the school's residual average MSEE scores of attendees from other cohorts after controlling for district fixed effects. For both panels, the 2003 entering cohort graph contains only five schools as three schools with incomplete lottery records in 2003 are excluded from the analysis throughout the paper.

The lack of evidence of any general achievement benefits conferred by elite schools on marginal admissions from lotteries, together with the positive correlation between a school's popularity and achievement level, indicates that elite schools may be sought after primarily for their observed superiority in student achievement rather than academic value-added. One possibility is that parents choose elite schools for reasons other than their effect on learning, such as for school facilities and peer quality (beyond peer externalities on learning, if any). Another possibility is that parents may confuse schools' achievement levels with achievement gains and therefore use the former to proxy for the latter. Fig. 5 plots a school's estimated value-added effect on achievement for the contemporary cohort against the average MSEE scores of attendees from other cohorts. These two measures are never strongly positively correlated with each other, echoing previous findings of a weak correlation between school grades and value-added in the US school accountability literature (see Kane and Staiger (2002) for a review). Thus, when schools' achievement levels constitute a poor proxy for achievement gains (as in the present case), parents are likely to misidentify higher-achieving schools as having more value-added. A third possibility is that parents may indeed care about achievement gains much more than peer quality (net of peer externalities on learning) in choosing a school, but are unable to effectively exercise such a preference because of the lack of reliable information on schools' value-added effects on achievement. Note that the value-added measures used here are constructed by myself using students' *ex post* MSEE scores and there is no similar measure available to parents when making school-choice decisions for their children. Any anecdotal information that parents may possess about a school's value-added is likely to be very noisy, whereas in contrast, peer quality is observed directly with accuracy. The sharp differences in the precision of anecdotal information on a school's value-added (if any) and peer quality may thus lead parents to rationally place heavier expressed weights on peer quality in choosing a school for their children, even if they care about the value-added the most. Although the available empirical evidence is unable to distinguish between these three underlying reasons, each is sufficient to lead to elite schools being sought after mainly for their

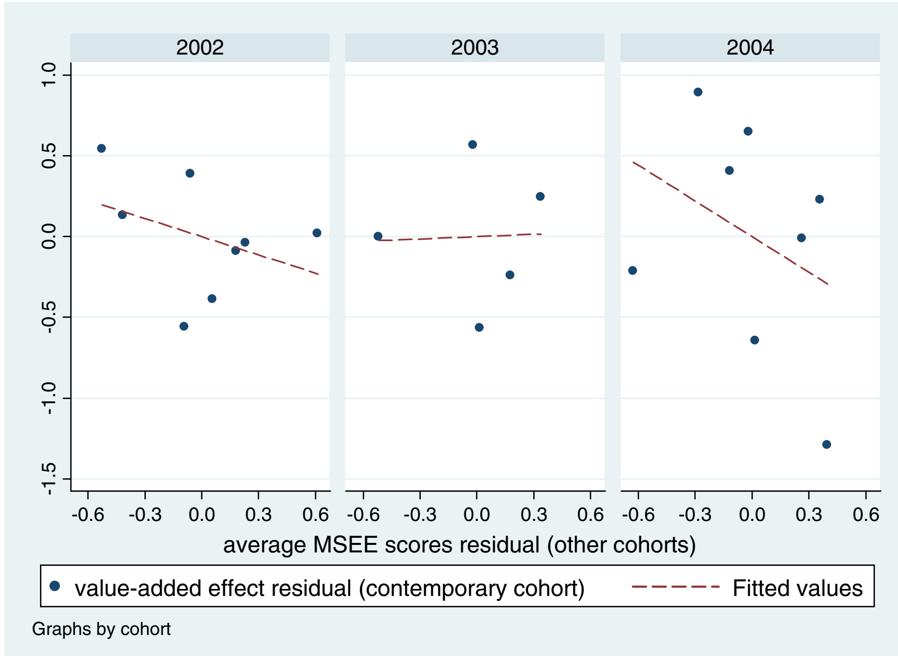

Fig. 5. Value-added effect and average MSEE scores.
Notes: For each cohort, the graph plots the school's residual value-added effect on lottery compliers in the contemporary cohort against the school's residual average MSEE scores of attendees from other cohorts after controlling for district fixed effects. The 2003 entering cohort graph contains only five schools as three schools with incomplete lottery records in 2003 are excluded from the analysis throughout the paper.

observed superiority in student achievement rather than for their effective value-added, thus casting doubt on the potential of school choice to increase demand-side pressure for schools to improve their effectiveness.

4. Conclusion

This paper contributes to the treatment effect analysis methods by extending the IVTE framework to be applicable to contexts where the instrument appears in a data set that can only be linked imperfectly to the treatment and outcome variables contained in another data set. I develop a data combination procedure that forms all pairwise links between the records in the two data sets matched by the set of commonly recorded personal characteristics, including both correct and incorrect links that are observationally indistinguishable. Under the relatively weak assumption of the independence of the treatment and outcome to the incorrectly linked instrument in case of false matching, I demonstrate that both the average and distributional treatment effects for compliers can be identified using the imperfect conditional matches. As this additional independence assumption is sustained under both the random assignment with imperfect compliance and the RD design, the extended IVTE framework for treatment effect analysis under imperfect matching developed herein can be applied to a wide range of empirical contexts encountered with similar observational obstacles.

This multiple data source approach for applying the IVTE framework for the identification of treatment effects is then applied to analyze the effect of attending an elite school in a Chinese city by exploiting exogenous variation in access to elite schools generated by school admissions lotteries. Although winning a lottery substantially increases students' chances of enrolling in their selected elite schools, which are far more advantageous in terms of peer achievement than neighborhood schools, I find little evidence that elite school attendance confers any general achievement or placement benefits to marginal admissions from lotteries. While there is some suggestive evidence that attending an elite school might increase the chances of admission to a top-echelon high school for girls and stronger applicants, such placement benefits (if indeed any) seem attributable primarily to strategy improvement in secondary school choices rather than achievement improvement. However, a number of caveats should be taken into account when interpreting these findings. First, I only estimate the effect of elite school attendance for lottery compliers; it is not possible establish counterfactuals to evaluate the effect of elite school attendance on students admitted through advance admissions or "back door" channels, who account for the majority of elite schools' attendees. Second, the empirical analysis considers only students' immediate academic outcomes and cannot rule out the possibility that these elite schools yield nonacademic or later-life benefits.[31] Third, the assessment of the effect of elite school attendance in this paper is reduced-form in nature, reflecting the combined influences of various forces whose separate effects are indistinguishable, some of which may be value-adding while others value-reducing.

[31] Examples of such benefits found in previous research in other countries include reducing crime (Cullen et al., 2006; Deming, 2011), decreasing teen cohabitation/marriage (Angrst et al., 2002), increasing college enrollment (Clark, 2010; Deming et al., 2014), reaching top management positions (Zimmerman, 2015).

Appendix A. Derivation of the IV bias under differential sample selection

This appendix section illustrates the bias in the IV estimand under differential sample selection, in which the compositions of Ψ_0 and Ψ_1 are partitioned as follows:

	AR	*FM*	*MC*
Ψ_0	$1-p_0$	p_0	0
Ψ_1	$(1-p_1^m)(1-p_0)$	$(1-p_1^m)p_0$	p_1^m

where p_1^m denotes the proportional share of the *MC* stratum in Ψ and p_0 denotes the proportional share of the *FM* stratum in Ψ_0 . Moreover, the *AR* stratum can be further partitioned into three substrata by potential treatment status, and the *FM* stratum (i.e. $i \neq j$) into two substrata by treatment status as follows:

AR	*FM*
(1a) $AR_{at} = \{D_i(0) = D_i(1) = 1\}$: *always takers*	(2a) $FM_{tr} = \{D_j = 1\}$: *treated*
(1b) $AR_c = \{D_i(1) > D_i(0)\}$: *compliers*	(2b) $FM_{ntr} = \{D_j = 0\}$: *non-treated*
(1c) $AR_{nt} = \{D_i(1) = D_i(0) = 0\}$: *never takers*	

With the foregoing partitions, the first-stage effect of Z on D in Ψ can be expressed as

$$\begin{aligned}
&E[D|\Psi_1] - E[D|\Psi_0] \\
&= (1-p_1^m)(1-p_0)E[D|AR, Z=1] + (1-p_1^m)p_0E[D|FM, Z=1] + p_1^mE[D|MC, Z=1] - (1-p_0)E[D|AR, Z=0] - p_0E[D|FM, Z=0] \\
&= (1-p_1^m)(1-p_0)E[D_1|AR] + (1-p_1^m)p_0E[D|FM] + p_1^mE[D_1|MC] - (1-p_0)E[D_0|AR] - p_0E[D|FM] \\
&= (1-p_0)(E[D_1 - D_0|AR] + p_1^m\{E[D_1|MC] - (1-p_0)E[D_1|AR] \ -p_0E[D|FM]\} \\
&= (1-p_0)\pi_{AR_c} + p_1^m[1-(1-p_0)(\pi_{AR_{at}} + \pi_{AR_c}) - p_0\pi_{FM_{tr}}],
\end{aligned} \tag{A1}$$

where π_{AR_c} and $\pi_{AR_{at}}$ denote the proportional share of the AR_c and AR_{at} substratum in the *AR* stratum, respectively, and $\pi_{FM_{tr}}$ denotes the proportional share of the FM_{tr} substratum in the *FM* stratum. The first term in Eq. (A1), $(1-p_0)\pi_{AR_c}$, corresponds to the first-stage effect of Z on D in Ψ without differential sample selection (i.e., $p_1^m = 0$), whereas the second term is the bias owing to the contamination of the *MC* stratum in Ψ_1. This bias term is always positive as the average treatment status of the *MC* stratum (1) exceeds that of both the *AR* stratum ($\pi_{AR_{at}} + \pi_{AR_c}$) and the *FM* stratum ($\pi_{FM_{tr}}$).

Analogously, the ITT effect of Z on Y in Ψ can be written as

$$\begin{aligned}
&E[Y|\Psi_1] - E[Y|\Psi_0] \\
&= (1-p_1^m)(1-p_0)E[Y|AR, Z=1] + (1-p_1^m)p_0E[Y|FM, Z=1] + p_1^mE[Y|MC, Z=1] - (1-p_0)E[Y|AR, Z=0] - p_0E[Y|FM, Z=0] \\
&= (1-p_0)(E[Y|AR, Z=1] - E[Y|AR, Z=0]) + p_1^m\{E[Y|MC, Z=1] - (1-p_0)E[Y|AR, Z=1] - p_0E[Y|FM]\} \\
&= (1-p_0)\Pr(AR_c|AR)E[Y_1 - Y_0|AR_c] + p_1^m\{(E[Y_0|MC] + E[Y_1 - Y_0|MC]) - (1-p_0)(E[Y_0|AR] + \Pr(AR_{at}|AR)E[Y_1 - Y_0|AR_{at}] + \Pr(AR_c|AR)E[Y_1 \\
&\quad -Y_0|AR_c]) - p_0(E[Y_0|FM] + \Pr(FM_{tr}|FM)E[Y_1 - Y_0|FM_{tr}])\} \\
&= (1-p_0)\pi_{AR_c}\gamma_{AR_c} + p_1^m\{(y_{MC} + \gamma_{MC}) - (1-p_0)(y_{AR} + \pi_{AR_{at}}\gamma_{AR_{at}} + \pi_{AR_c}\gamma_{AR_c}) - p_0(y_{FM} + \pi_{FM_{tr}}\gamma_{FM_{tr}})\},
\end{aligned} \tag{A2}$$

where y_{MC}, y_{AR}, and y_{FM} denote the mean potential outcome *without treatment* for the *MC*, *AR*, and *FM* stratum, respectively;[32] and γ_{MC}, $\gamma_{AR_{at}}$, γ_{AR_c}, and $\gamma_{FM_{tr}}$ denote the ATE for the *MC* stratum, the AR_{at} substratum, the AR_c substratum, and the FM_{tr} substratum, respectively.[33] The first term in Eq. (A2), $(1-p_0)\pi_{AR_c}\gamma_{AR_c}$, corresponds to the ITT effect of Z on Y in Ψ without differential sample selection (i.e., $p_1^m = 0$), which is entirely attributable to the treatment effect for the AR_c substratum. The second term is the bias owing to the contamination of the *MC* stratum in Ψ_1, which is nonzero if in Ψ_1 the mean outcome of the *MC* stratum ($y_{MC} + \gamma_{MC}$) differs from the weighted average of the mean outcomes of the *AR* stratum ($y_{AR} + \pi_{AR_{at}}\gamma_{AR_{at}} + \pi_{AR_c}\gamma_{AR_c}$) and the *FM* stratum ($y_{FM} + \pi_{FM_{tr}}\gamma_{FM_{tr}}$).

Taking the ratio of Eqs. (A2) and (A1) yields the IV estimand as follows:

$$\begin{aligned}
\gamma_{IV} = \tfrac{E[Y|\Psi_1]-E[Y|\Psi_0]}{E[D|\Psi_1]-E[D|\Psi_0]} = \gamma_{AR_c} + \tfrac{p_1^m}{E[D|\Psi_1]-E[D|\Psi_0]}\Big\{&\underbrace{[y_{MC} - (1-p_0)y_{AR} - p_0y_{FM}]}_{\eta_1} + \\
&\underbrace{[\gamma_{MC} - (1-p_0)\pi_{AR_{at}}\gamma_{AR_{at}} - p_0\pi_{FM_{tr}}\gamma_{FM_{tr}} - (1-(1-p_0)\pi_{AR_{at}} - p_0\pi_{FM_{tr}})\gamma_{AR_c}]}_{\eta_2}\Big\}.
\end{aligned} \tag{A3}$$

Besides γ_{AR_c}, the remaining term in Eq. (A3) is the bias in the IV estimand due to the contamination of the *MC* stratum in Ψ_1 caused by differential sample selection (if any). A few remarks are worth mentioning regarding the bias term.

[32] That is, $y_{MC} = E[Y_0|MC]$, $y_{AR} = E[Y_0|AR]$, and $y_{FM} = E[Y_0|FM]$.

[33] That is, $\gamma_{MC} = E[Y_1 - Y_0|MC]$, $\gamma_{AR_{at}} = E[Y_1 - Y_0|AR_{at}]$, $\gamma_{AR_c} = E[Y_1 - Y_0|AR_c]$, and $\gamma_{FM_{tr}} = E[Y_1 - Y_0|FM_{tr}]$.

Remark 1. The magnitude of the bias is proportional to the ratio of the *MC* stratum's share in Ψ_1 (p_1^m) and the first-stage effect of *Z* on *D* in Ψ. The smaller the value of this ratio, the smaller the bias. The point identification of γ_{AR_c} in Proposition 3 can be considered as a special case of Eq. (A3) in which the bias term is eliminated in the absence of differential sample selection (i.e., $p_1^m = 0$).

Remark 2. The ability selection component, η_1, reflects the extent to which ability selection in the contaminating *MC* stratum differs from that in the *AR* and *FM* strata combined. Specifically, the size of this ability selection term is determined by the extent to which the mean potential outcome without treatment of the *MC* stratum (y_{MC}) differs from a weighted average of the mean potential outcome without treatment of the *AR* stratum (y_{AR}) and the *FM* stratum (y_{FM}), with the weights equal to their respective proportional shares in Ψ_0.

Remark 3. The treatment heterogeneity component, η_2, reflects the extent to which the ATE of the *MC* stratum (γ_{MC}) differs from the weighted average of the ATEs of the AR_{at} substratum ($\gamma_{AR_{at}}$), the FM_{tr} substratum ($\gamma_{FM_{tr}}$), and the AR_c substratum (γ_{AR_c}). In particular, if treatment effects are homogeneous across all individuals, η_2 becomes 0 as the ATEs are the same for all subgroups.

Appendix B. Proof of Proposition 3

Proof of Proposition 3. In this appendix section, I show the proof for the identifications of γ_{AR_c} and $F_{(1)}^{AR_c}(y)$ in Proposition 3. The proofs for the identifications of $F_{(0)}^{AR_c}(y)$ and $F_{(1)}^{AR_c}(y) - F_{(0)}^{AR_c}(y)$ are skipped here as they follow immediately from the proof for the identification of $F_{(1)}^{AR_c}(y)$ and Proposition 2. First, the numerator and denominator of the formula for γ_{AR_c} can be expressed as follows:

$$
\begin{aligned}
&E[Y_{i(j)}|\Psi_{11}^* \cup \Psi_{10}] - E[Y_{i(j)}|\Psi_0] \\
&\quad = (1-p_0)E[Y|AR, Z=1] + p_0 E[Y|FM, Z=1] - (1-p_0)E[Y|AR, Z=0] - p_0 E[Y|FM, Z=0] \\
&\quad = (1-p_0)(E[Y|AR, Z=1] - E[Y|AR, Z=0])
\end{aligned}
\tag{A4}
$$

and

$$
\begin{aligned}
&E[D_{i(j)}|\Psi_{11}^* \cup \Psi_{10}] - E[D_{i(j)}|\Psi_0] \\
&\quad = (1-p_0)E[D|AR, Z=1] + p_0 E[D|FM, Z=1] - (1-p_0)E[D|AR, Z=0] - p_0 E[D|FM, Z=0] \\
&\quad = (1-p_0)(E[D|AR, Z=1] - E[D|AR, Z=0]),
\end{aligned}
\tag{A5}
$$

where p_0 denotes the proportional share of the *FM* stratum in Ψ_0 as previously defined. Taking the ratio of Eqs. (A4) and (A5) eliminates the term $(1-p_0)$ and yields a Wald estimator for the ATE for compliers in the *AR* stratum.

Second, similar to Eq. (A4), the numerator of the formula for $F_{(1)}^{AR_c}(y)$ can be expressed as

$$
\begin{aligned}
&E\left[1(Y_{i(j)} \le y)D_{i(j)}|\Psi_{11}^* \cup \Psi_{10}\right] - E\left[1(Y_{i(j)} \le y)D_{i(j)}|\Psi_0\right] \\
&\quad = (1-p_0)E[1(Y \le y)D|AR, Z=1] + p_0 E[1(Y \le y)D|FM, Z=1] - (1-p_0)E[1(Y \le y)D|AR, Z=0] - p_0 E[1(Y \le y)D|FM, Z=1] \\
&\quad = (1-p_0)(E[1(Y \le y)D|AR, Z=1] - [1(Y \le y)D|AR, Z=0]),
\end{aligned}
\tag{A6}
$$

whereas the denominator of the formula for $F_{(1)}^{AR_c}(y)$ is Eq. (A5). Substituting Eqs. (A6) and (A5) into the formula for $F_{(1)}^{AR_c}(y)$ eliminates the term $(1-p_0)$ and yields

$$
\frac{E\left[1\left(Y_{i(j)} \le y\right)D_{i(j)}|\Psi_{11}^* \cup \Psi_{10}\right] - E\left[1\left(Y_{i(j)} \le y\right)D_{i(j)}|\Psi_0\right]}{E\left[D_{i(j)}|\Psi_{11}^* \cup \Psi_{10}\right] - E\left[D_{i(j)}|\Psi_0\right]} = \frac{E[1(Y \le y)D|AR, Z=1] - [1(Y \le y)D|AR, Z=0]}{E[D|AR, Z=1] - E[D|AR, Z=1]}.
\tag{A7}
$$

Note that the right-hand side of Eq. (A7) is a conditional version of the formula for $F_{(1)}^C(y)$ in Proposition 2, as all expectations are expressed conditional on pertaining to the *AR* stratum. Thus, Eq. (A7) identifies the cumulative distribution of potential outcomes for the subpopulation of compliers in the *AR* stratum. ■

Appendix C. Proof of Proposition 4

Proof of Proposition 4. First of all, by construction $\underline{\Psi}_{11}$ and $\overline{\Psi}_{11}$ are identical in size proportional to Ψ_{11}^*, indicating the following relationship:

$$
\Pr(\underline{\Psi}_{11}|\underline{\Psi}_{11} \cup \Psi_{10}) = \Pr\left(\overline{\Psi}_{11}|\overline{\Psi}_{11} \cup \Psi_{10}\right) = \Pr\left(\Psi_{11}^*|\Psi_{11}^* \cup \Psi_{10}\right).
\tag{A8}
$$

Moreover, the matched record pairs in $\underline{\Psi}_{11}$, $\overline{\Psi}_{11}$, and Ψ_{11}^* all have $D_{i(j)} = 1$, whereas those in Ψ_{10} all have $D_{i(j)} = 0$. Therefore,

$$
E\left[D_{i(j)}|\tilde{\Psi} \cup \Psi_{10}\right] = \Pr\left(\tilde{\Psi}|\tilde{\Psi} \cup \Psi_{10}\right) \text{ for } \tilde{\Psi} \in \left\{\underline{\Psi}_{11}, \overline{\Psi}_{11}, \Psi_{11}^*\right\}.
\tag{A9}
$$

Combining Eqs. (A8) and (A9) indicates that the denominators of all formulas in Propositions 3 and 4 are identical, i.e.,

$$
E\left[D_{i(j)}|\underline{\Psi}_{11} \cup \Psi_{10}\right] - E\left[D_{i(j)}|\Psi_0\right] = E\left[D_{i(j)}|\overline{\Psi}_{11} \cup \Psi_{10}\right] - E\left[D_{i(j)}|\Psi_0\right] = E\left[D_{i(j)}|\Psi_{11}^* \cup \Psi_{10}\right] - E\left[D_{i(j)}|\Psi_0\right].
\tag{A10}
$$

Also by construction, in terms of the distribution of Y, $\overline{\Psi}_{11}$ first-order stochastically dominates Ψ^*_{11} and Ψ^*_{11} first-order stochastically dominates $\underline{\Psi}_{11}$, i.e.,

$$F\left(y;\underline{\Psi}_{11}\right)\le F\left(y;\Psi^*_{11}\right)\le F\left(y;\overline{\Psi}_{11}\right)\forall y. \tag{A11}$$

Now consider the numerator of the lower bound formula for γ_{AR_c}, which can be expressed as

$$\begin{aligned}
&E\left[Y_{i(j)}|\underline{\Psi}_{11}\cup\Psi_{10}\right]-E\left[Y_{i(j)}|\Psi_0\right]\\
&\quad=\Pr\left(\underline{\Psi}_{11}|\underline{\Psi}_{11}\cup\Psi_{10}\right)E\left[Y_{i(j)}|\underline{\Psi}_{11}\right]+\left(1-\Pr(\underline{\Psi}_{11}|\underline{\Psi}_{11}\cup\Psi_{10})\right)E\left[Y_{i(j)}|\Psi_{10}\right]-E\left[Y_{i(j)}|\Psi_0\right]\\
&\quad\le\Pr(\Psi^*_{11}|\Psi^*_{11}\cup\Psi_{10})E\left[Y_{i(j)}|\Psi^*_{11}\right]+\left(1-\Pr(\Psi^*_{11}|\Psi^*_{11}\cup\Psi_{10})\right)E\left[Y_{i(j)}|\Psi_{10}\right]-E\left[Y_{i(j)}|\Psi_0\right]\\
&\quad=E\left[Y_{i(j)}|\Psi^*_{11}\cup\Psi_{10}\right]-E\left[Y_{i(j)}|\Psi_0\right].
\end{aligned} \tag{A12}$$

Note that the inequality follows from two facts: $\Pr(\underline{\Psi}_{11}|\underline{\Psi}_{11}\cup\Psi_{10}) = \Pr\left(\Psi^*_{11}|\Psi^*_{11}\cup\Psi_{10}\right)$ and $E\left[Y_{i(j)}|\underline{\Psi}_{11}\right] \le E\left[Y_{i(j)}|\Psi^*_{11}\right]$, the latter of which is implied by the first-order stochastic dominance relationship in Eq. (A11). Dividing the left- and right-hand sides of Eq. (A12) by $E\left[D_{i(j)}|\underline{\Psi}_{11}\cup\Psi_{10}\right]-E\left[D_{i(j)}|\Psi_0\right]$ and $E\left[D_{i(j)}|\Psi^*_{11}\cup\Psi_{10}\right]-E\left[D_{i(j)}|\Psi_0\right]$, respectively, which are shown to be identical in Eq. (A10), yields the following inequality relationship:

$$\frac{E\left[Y_{i(j)}|\underline{\Psi}_{11}\cup\Psi_{10}\right]-E\left[Y_{i(j)}|\Psi_0\right]}{E\left[D_{i(j)}|\underline{\Psi}_{11}\cup\Psi_{10}\right]-E\left[D_{i(j)}|\Psi_0\right]}\le\frac{E\left[Y_{i(j)}|\Psi^*_{11}\cup\Psi_{10}\right]-E\left[Y_{i(j)}|\Psi_0\right]}{E\left[D_{i(j)}|\Psi^*_{11}\cup\Psi_{10}\right]-E\left[D_{i(j)}|\Psi_0\right]}=\gamma_{AR_c}.$$

Replacing $\underline{\Psi}_{11}$ with $\overline{\Psi}_{11}$ in the above equation reverses the sign of the inequality relationship and yields an upper bound for γ_{AR_c}:

$$\frac{E\left[Y_{i(j)}|\overline{\Psi}_{11}\cup\Psi_{10}\right]-E\left[Y_{i(j)}|\Psi_0\right]}{E\left[D_{i(j)}|\overline{\Psi}_{11}\cup\Psi_{10}\right]-E\left[D_{i(j)}|\Psi_0\right]}\ge\frac{E\left[Y_{i(j)}|\Psi^*_{11}\cup\Psi_{10}\right]-E\left[Y_{i(j)}|\Psi_0\right]}{E\left[D_{i(j)}|\Psi^*_{11}\cup\Psi_{10}\right]-E\left[D_{i(j)}|\Psi_0\right]}=\gamma_{AR_c}.$$

Similar to Eq. (A12), the numerator of the lower bound formula for $F^{AR_c}_{(1)}(y)$ can be expressed as

$$\begin{aligned}
&E\left[1\left(Y_{i(j)}\le y\right)D_{i(j)}|\underline{\Psi}_{11}\cup\Psi_{10}\right]-E\left[1\left(Y_{i(j)}\le y\right)D_{i(j)}|\Psi_0\right]\\
&\quad=\Pr(\underline{\Psi}_{11}|\underline{\Psi}_{11}\cup\Psi_{10})E\left[1\left(Y_{i(j)}\le y\right)|\underline{\Psi}_{11}\right]-E\left[1\left(Y_{i(j)}\le y\right)D_{i(j)}|\Psi_0\right]\\
&\quad\le\Pr\left(\Psi^*_{11}|\Psi^*_{11}\cup\Psi_{10}\right)E\left[1\left(Y_{i(j)}\le y\right)|\Psi^*_{11}\right]-E\left[1\left(Y_{i(j)}\le y\right)D_{i(j)}|\Psi_0\right]\\
&\quad=E\left[1\left(Y_{i(j)}\le y\right)D_{i(j)}|\Psi^*_{11}\cup\Psi_{10}\right]-E\left[1\left(Y_{i(j)}\le y\right)D_{i(j)}|\Psi_0\right].
\end{aligned} \tag{A13}$$

Note that the inequality follows the fact that $\Pr(\underline{\Psi}_{11}|\underline{\Psi}_{11}\cup\Psi_{10})=\Pr(\Psi^*_{11}|\Psi^*_{11}\cup\Psi_{10})$ and $E\left[1\left(Y_{i(j)}\le y\right)|\underline{\Psi}_{11}\right]\le E\left[1\left(Y_{i(j)}\le y\right)|\Psi^*_{11}\right]$, the latter of which is also implied by the first-order stochastic dominance relationship in Eq. (A11). Combining Eqs. (A13) and (A10) yields the following inequality relationship:

$$\frac{E\left[1\left(Y_{i(j)}\le y\right)D_{i(j)}|\underline{\Psi}_{11}\cup\Psi_{10}\right]-E\left[1\left(Y_{i(j)}\le y\right)D_{i(j)}|\Psi_0\right]}{E\left[D_{i(j)}|\underline{\Psi}_{11}\cup\Psi_{10}\right]-E\left[D_{i(j)}|\Psi_0\right]}\le\frac{E\left[1\left(Y_{i(j)}\le y\right)D_{i(j)}|\Psi^*_{11}\cup\Psi_{10}\right]-E\left[1\left(Y_{i(j)}\le y\right)D_{i(j)}|\Psi_0\right]}{E\left[D_{i(j)}|\Psi^*_{11}\cup\Psi_{10}\right]-E\left[D_{i(j)}|\Psi_0\right]}=F^{AR_c}_{(1)}(y).$$

Replacing $\underline{\Psi}_{11}$ with $\overline{\Psi}_{11}$ in the above equation reverses the sign of the inequality relationship and yields an upper bound for $F^{AR_c}_{(1)}(y)$.

Finally, let us derive the equivalence of the two formulas for the point identification of $F^{AR_c}_{(0)}(y)$ in Propositions 3 and 4. As Eq. (A10) already indicates the equivalence of the denominators of the two formulas, it is sufficient to show the equivalence of the two numerators:

$$\begin{aligned}
&E\left[1\left(Y_{i(j)}\le y\right)\left(1-D_{i(j)}\right)|\Psi_0\right]-E\left[1\left(Y_{i(j)}\le y\right)\left(1-D_{i(j)}\right)\ |(\underline{\Psi}_{11}\text{ or }\overline{\Psi}_{11})\cup\Psi_{10}\right]\\
&\quad=E\left[1\left(Y_{i(j)}\le y\right)\left(1-D_{i(j)}\right)|\Psi_0\right]-\Pr(\Psi_{10}|(\underline{\Psi}_{11}\text{ or }\overline{\Psi}_{11})\cup\Psi_{10})E\left[1\left(Y_{i(j)}\le y\right)\left(1-D_{i(j)}\right)|\Psi_{10}\right]\\
&\quad=E\left[1\left(Y_{i(j)}\le y\right)\left(1-D_{i(j)}\right)|\Psi_0\right]-\Pr(\Psi_{10}|\Psi^*_{11}\cup\Psi_{10})E\left[1\left(Y_{i(j)}\le y\right)\left(1-D_{i(j)}\right)|\Psi_{10}\right]\\
&\quad=E\left[1\left(Y_{i(j)}\le y\right)\left(1-D_{i(j)}\right)|\Psi_0\right]-E\left[1\left(Y_{i(j)}\le y\right)\left(1-D_{i(j)}\right)|\Psi^*_{11}\cup\Psi_{10}\right].
\end{aligned}$$

Note that the second equality in the above equation follows from the fact that $\Pr(\Psi_{10}|\Psi^*_{11}\cup\Psi_{10})=\Pr(\Psi_{10}|(\underline{\Psi}_{11}\text{ or }\overline{\Psi}_{11})\cup\Psi_{10})$ as indicated by Eq. (A8). With the lower and upper bounds constructed for $F^{AR_c}_{(1)}(y)$ and the point identification of $F^{AR_c}_{(0)}(y)$, it is straightforward to show the lower and upper bounds for the distributional treatment effects $F^{AR_c}_{(1)}(y)-F^{AR_c}_{(0)}(y)$ in Proposition 4. ■

Appendix D. Derivation of Eq. (5)

Given Eqs. (4a) and (4b), $F^C_{(1)}(y) - F^C_{(0)}(y)$ can be expressed as follows:

$$F^C_{(1)}(y) - F^C_{(0)}(y) = \sum_k \frac{N_k \delta_k}{\sum_k N_k \delta_k}\left(F^C_{(1),k}(y) - F^C_{(0),k}(y)\right) = \sum_k \frac{N_k \delta_k}{\sum_k N_k \delta_k} \frac{F_{1,k}(y) - F_{0,k}(y)}{\delta_k} = \sum_k \frac{N_k}{\sum_k N_k \delta_k}\left(F_{1,k}(y) - F_{0,k}(y)\right), \tag{A14}$$

where $F_{1,k}(y)$ denotes the cumulative distribution of outcomes in $\Psi_1(k)$, $E[1(Y_{i(j)} \le y)|Z_{i(j)} = 1, i \in I(k)]$; and $F_{0,k}(y)$ denotes the cumulative distribution of outcomes in $\Psi_0(k)$, $E[1(Y_{i(j)} \le y)|Z_{i(j)} = 0, i \in I(k)]$. Note that the second equality in Eq. (A14) follows from Eq. (2d) in Proposition 2. Substituting $F_{1,k}(y) = \frac{\sum_{i:Z_i=1, i \in I(k)} \sum_{j:C_j=C_i} 1(Y_j \le y)}{N_{1k}}$ and $F_{0,k}(y) = \frac{\sum_{i:Z_i=0, i \in I(k)} \sum_{j:C_j=C_i} 1(Y_j \le y)}{N_{0k}}$ into Eq. (A14) yields:

$$\begin{aligned}
&F^C_{(1)}(y) - F^C_{(0)}(y) \\
&= \sum_k \frac{N_k}{\sum_k N_k \delta_k}\left(\frac{\sum_{i:Z_i=1, i \in I(k)} \sum_{j:C_j=C_i} 1(Y_j \le y)}{N_{1k}} - \frac{\sum_{i:Z_i=0, i \in I(k)} \sum_{j:C_j=C_i} 1(Y_j \le y)}{N_{0k}}\right) \\
&= \sum_k \frac{N_k}{\sum_k N_k \delta_k}\left(\frac{\sum_{i:Z_i=1, i \in I(k)} \sum_{j:C_j=C_i} 1(Y_j \le y)}{N_k \pi_k} - \frac{\sum_{i:Z_i=0, i \in I(k)} \sum_{j:C_j=C_i} 1(Y_j \le y)}{N_k(1-\pi_k)}\right) \\
&= \frac{1}{\sum_k N_k \delta_k}\left(\sum_k \sum_{i:Z_i=1, i \in I(k)} \sum_{j:C_j=C_i} \frac{1}{\pi_k} 1(Y_j \le y) - \sum_k \sum_{i:Z_i=0, i \in I(k)} \sum_{j:C_j=C_i} \frac{1}{1-\pi_k} 1(Y_j \le y)\right) \\
&= \frac{\sum_k N_k}{\sum_k N_k \delta_k}\left(\frac{\sum_k \sum_{i:Z_i=1, i \in I(k)} \sum_{j:C_j=C_i} \frac{1}{\pi_k} 1(Y_j \le y)}{\sum_k N_k} - \frac{\sum_k \sum_{i:Z_i=0, i \in I(k)} \sum_{j:C_j=C_i} \frac{1}{1-\pi_k} 1(Y_j \le y)}{\sum_k N_k}\right) \\
&= \frac{\sum_k N_k}{\sum_k N_k \delta_k}\left(F_1\left(y; \frac{1}{\pi_k}\right) - F_0\left(y; \frac{1}{1-\pi_k}\right)\right)
\end{aligned} \tag{A15}$$

Recall that π_k is defined as the lottery-specific proportional share of $\Psi_1(k)$ in $\Psi(k)$, i.e., $\Pr(\Psi_1(k)|\Psi(k))$. Given this definition, the second equality follows immediately from the fact that $N_{1k} = N_k \pi_k$ and $N_{0k} = N_k(1-\pi_k)$. Also given this definition, the sum of the weights in Ψ_1 (i.e., $\sum_k \sum_{i:Z_i=1, i \in I(k)} \sum_{j:C_j=C_i} \frac{1}{\pi_k}$) and the sum of the weights in Ψ_0 (i.e., $\sum_k \sum_{i:Z_i=0, i \in I(k)} \sum_{j:C_j=C_i} \frac{1}{1-\pi_k}$) are both equal to $\sum_k N_k$. Therefore, dividing the two summations in the parenthesis in the third equality by $\sum_k N_k$ yields the two reweighted cumulative distribution functions $F_1(y; \frac{1}{\pi_k})$ and $F_0(y; \frac{1}{1-\pi_k})$, respectively.

References

Abadie, A., 2002. Bootstrap tests for distributional treatment effects in instrumental variables models. J. Am. Stat. Assoc. 97 (457), 284–292.

Abdulkadiroglu, A., Angrist, J.D., Pathak, P.A., 2014. The elite illusion: achievement effects at Boston and New York exam schools. Econometrica 82 (1), 137–196.

Abdulkadiroglu, A., Sonmez, T., 2003. School choice: a mechanism design approach. Am. Econ. Rev. 93 (3), 729–747.

Angrist, J., Bettinger, E., Bloom, E., King, E., Kremer, M., 2002. Vouchers for private schooling in Colombia: evidence from a randomized natural experiment. Am. Econ. Rev. 92 (5), 1535–1558.

Angrist, J., Bettinger, E., Kremer, M., 2006. Long-term educational consequences of secondary school vouchers: evidence from administrative records in Colombia. Am. Econ. Rev. 96 (3), 847–862.

Angrist, J.D., Graddy, K., Imbens, G.W., 2000. The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. Rev. Econ. Stud. 67 (3), 499–527.

Angrist, J.D., Imbens, G.W., 1995. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. J. Am. Stat. Assoc. 90 (430), 431–442.

Angrist, J.D., Imbens, G.W., Rubin, D.B., 1996. Identification of causal effects using instrumental variables. J. Am. Stat. Assoc. 91 (434), 444–455.

Angrist, J.D., Krueger, A.B., 1992. The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. J. Am. Stat. Assoc. 87 (418), 328–336.

Arellano, M., Meghir, C., 1992. Female labour supply and on-the-job search: an empirical model estimated using complementary data sets. Rev. Econ. Stud. 59 (3), 537–559.

Bettinger, E., Kremer, M., Saavedra, J.E., 2010. Are educational vouchers only redistributive? Econ. J. 120 (546). F204–F228.

Bui, S.A., Imberman, S.A., Craig, S.G., 2014. Is gifted education a bright idea? Assessing the impact of gifted and talented programs on achievement. Am. Econ. J. Econ. Pol. 6 (3), 30–62.

Cellini, S.R., Ferreira, F., Rothstein, J., 2010. The value of school facility investments: evidence from a dynamic regression discontinuity design. Q. J. Econ. 125 (1), 215–261.

Chernozhukov, C., Hong, H., Tamer, E., 2007. Estimation and confidence regions for parameter sets in econometric models. Econometrica 75 (5), 1243–1284.

Clark, D., 2010. Selective schools and academic achievement. The B.E. J. Econ. Anal. Policy 10 (1), 1–38.

Corcoran, S.P., Jennings, J.L., Beveridge, A.A., 2011. Teacher effectiveness on high- and low-stakes tests. Working Paper.

Cox, D.R., 1958. Planning of Experiments. John Wiley and Sons, New York.

Cullen, J.B., Jacob, B.A., Levitt, S., 2006. The effect of school choice on participants: evidence from randomized lotteries. Econometrica 74 (5), 1191–1230.

Dee, T., Lan, X., 2015. The achievement and course-taking effects of magnet schools: regression-discontinuity evidence from urban China. Econ. Educ. Rev. 47, 128–142.

Dello-Iacovo, B., 2009. Curriculum reform and “quality education” in China: an overview. Int. J. Educ. Dev. 29 (3), 241–249.

Deming, D.J., 2011. Better schools, less crime. Q. J. Econ. 126 (4), 2063–2115.

Deming, D., Hastings, J., Kane, T., Staiger, D., 2014. School choice, school quality, and postsecondary attainment. Am. Econ. Rev. 104 (3), 991–1013.

Ding, W., Lehrer, S.F., 2007. Do peers affect student achievement in China's secondary schools? Rev. Econ. Stat. 89 (2). 300–312.

Dobbie, W., Fryer, R.G.J., 2014. The impact of attending a school with high-achieving peers: evidence from the New York City exam schools. Am. Econ. J. Appl. Econ. 6 (3), 58–75.

Downes, T.A., Zabel, J.E., 2002. The impact of school characteristics on house prices: Chicago 1987–1991. J. Urban Econ. 52 (1), 1–25.

Duflo, E., Dupas, P., Kremer, M., 2011. Peer effects, teacher incentives, and the impact of tracking: evidence from a randomized evaluation in Kenya. Am. Econ. Rev. 101 (5), 1739–1774.
Figlio, D.N.F., Lucas, M.E., 2004. What's in a grade? School report cards and the housing market. Am. Econ. Rev. 94 (3), 591–604.
Hastings, J.S., Kane, T.J., Staiger, D.O., 2009. Heterogeneous preferences and the efficacy of public school choice. Working Paper.
Hastings, J.S., Weinstein, J.M., 2008. Information, school choice, and academic achievement: evidence from two experiments. Q. J. Econ. 123 (4), 1373–1414.
He, Y., 2014. Gaming the Boston school choice mechanism in Beijing. Working Paper.
Hoxby, C.M., Weingarth, G., 2006. Taking race out of the equation: school reassignment and the structure of peer effects. Working Paper.
Hsieh, C.T., Urquiola, M., 2006. The effects of generalized school choice on achievement and stratification: evidence from Chile's voucher program. J. Public Econ. 90 (8-9), 1477–1503.
Imbens, G.W., Angrist, J.D., 1994. Identification and estimation of local average treatment effects. Econometrica 62 (2), 467–476.
Imbens, G.W., Manski, C.F., 2004. Confidence intervals for partially identified parameters. Econometrica 72 (6), 1845–1857.
Imbens, G.W., Rubin, D.B., 1997. Estimating outcome distributions for compliers in instrumental variables models. Rev. Econ. Stud. 64 (4), 555–574.
Jackson, C.K., 2010. Do students benefit from attending better schools? Evidence from rule-based student assignments in Trinidad and Tobago. Econ. J. 120 (549), 1399–1429.
Jacob, B.A., 2007.Test-based accountability and student achievement: an investigation of differential performance on NAEP and State assessments. National Bureau of Economic Research Working Paper No 12817.
Jacob, B.A., Lefgren, L., 2007. What do parents value in education? An empirical investigation of parents' revealed preferences for teachers. Q. J. Econ. 122 (4), 1603–1637.
Kane, T.J., Staiger, D.O., 2002. The promise and pitfalls of using imprecise school accountability measures. J. Econ. Perspect. 16 (4), 91–114.
Lai, F., Sadoulet, E., de Janvry, A., 2011. The contributions of school quality and teacher qualifications to student performance: evidence from a natural experiment in Beijing Middle Schools. J. Hum. Resour. 46 (1), 123–153.
Lavy, V., 2010. Effects of free choice among public schools. Rev. Econ. Stud. 77 (3), 1164–1191.
Lavy, V., Silva, O., Weinhardt, F., 2012. The good, the bad, and the average: evidence on ability peer effects in schools. J. Labor Econ. 30 (2), 367–414.
Lee, D.S., 2008. Randomized experiments from non-random selection in U.S. House elections. J. Econ. 142 (2), 675–697.
Lee, D.S., 2009. Training, wages, and sample selection: estimating sharp bounds on treatment effects. Rev. Econ. Stud. 76 (3), 1071–1102.
Lee, D.S., Lemieux, T., 2010. Regression discontinuity designs in economics. J. Econ. Lit. 48 (2), 281–355.
MacLeod, W.B., Urquiola, M., 2015. Reputation and school competition. Am. Econ. Rev. (forthcoming).
Mizala, A., Urquiola, M., 2013. School markets: the impact of information approximating schools' effectiveness. J. Dev. Econ. 103, 313–335.
Pop-Eleches, C., Urquiola, M., 2013. Going to a better school: effects and behavioral responses. Am. Econ. Rev. 103 (4), 1290–1324.
Ridder, G., Moffitt, R., 2007. The econometrics of data combination. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics. 6B. Elsevier Science, Amsterdam, pp. 5469–5547.
Zimmerman, S., 2015. Making top Managers: the role of elite universities and elite peers. Working Paper.