Purchasing Seats in School Choice and Inequality*

Tong Wang† Congyi Zhou‡

July 13, 2024

Abstract

We study a mechanism that gives students the option of paying higher tuition to attend their preferred schools. This seat-purchasing mechanism is neither strategy-proof nor stable. Our paper combines administrative and survey data to estimate students' preferences and conducts welfare analysis. We find that changing from a deferred acceptance mechanism to the cadet-optimal stable mechanism reduces students' welfare but that adopting the observed seat-purchasing mechanism alleviates this welfare loss. Moreover, students from affluent communities prefer to pay higher tuition to stay at preferred schools, while those from less affluent communities are more likely be priced out to lower-quality schools.

JEL Code: C78, D82, I21, I28

*We would like to thank Atila Abdulkadiroglu, Hideo Akabayashi, Daniel Barron, Dan Black, Eric Budish, Yinghua He, Kohei Kawamura, Jacob Leshno, Haruko Noguchi, Elena Prager, Shigehiro Serizawa for helpful comments and workshop and conference participants at Keio University, Northwestern Kellogg Strategy Lunch, Osaka University, University of Chicago Industrial Organization Lunch, University of Tokyo, Waseda University, AEA annual conference, EEA annual conference, IIOC conference and APPAM annual conferences.

†Tong Wang, College of International Relations, Ritsumeikan University, tongw@fc.ritsumei.ac.jp

‡Corresponding author: Congyi Zhou, Wilf Department of Politics, New York University, cz536@nyu.edu

1 Introduction

The analysis of centralized school choice mechanisms has become a key focus of research in market design (Abdulkadiroglu and Sönmez 2003). Kelso Jr and Crawford (1982) and Hatfield and Milgrom (2005) have built the connection between auction and matching by introducing the matching with contracts model. Since then, analyzing how individuals respond to a “price menu” for an individual good in matching markets has attracted growing interest. Theoretical analysis has been used to address this question in practice (Sönmez and Switzer 2013; Biro et al. 2022). However, no clear empirical analysis has disentangled individual behaviors under the matching model with monetary transfer.

In extant literature on the school choice problem, the effect of monetary transfers between students and schools is seldom considered because public schools either offer free education or have a fixed (and usually low) tuition fee. Yet unlike public school choice systems in other countries, many Chinese cities have—starting in the 1990s—offered students the option of paying higher tuition and thereby gaining admission to public schools.[1] This procedure is referred to as the *Ze Xiao* (ZX) policy.[2] The ZX policy is a practical application of the *matching with contracts* model (Kelso Jr and Crawford 1982; Hatfield and Milgrom 2005; Hatfield and Kojima 2008, 2010; Hatfield, Kominers, and Westkamp 2017). However, this policy provoked controversy because it was perceived as an unfair policy to families that cannot afford higher tuition (Shen and Wu 2006). The controversy lasted for more than a decade and was somewhat defused in 2012, when the Ministry of Education announced the restrictions on the ZX policy and requested that public high schools stop using it within three years.[3]

We exploit a new dataset covering high school admissions for the period 2012–2014 in a

[1] Zhu Kaixuan, then chairman of the state education commission, publicly addressed the seat-purchasing problem in public schools. In 1995 he argued, in the *People's Daily*, against paying higher tuition to purchase admission to compulsory education.

[2] “Ze Xiao” is Chinese for “school selection”.

[3] Many cities, including Shanghai (which ceased using the ZX policy in 2012), Beijing (2014), and Shenzhen and Tianjin (2015), ceased the policy for high school admissions.

large Chinese city.[4] By combining these admission records with data from a 2014 survey, we fill two aspects of the research gap in education policy and market design. First, we provide empirical evidence to understand students' strategic behaviors in matching with contracts theory. Second, considering that the ZX policy directly involves monetary transfers between students and schools, evaluating this policy helps us understand whether offering a "price menu" in a centralized school choice procedure would increase the inequality in education among students from different backgrounds.

The high schools in our focal city adopted the typical ZX policy for their admissions procedure until that policy was cancelled in 2014 (see Appendix I for details of the ZX policy in various Chinese cities). The ZX policy specified the basic and higher tuition levels (i.e., the "price menu") and the number of seats for sale in each school (i.e., the ZX *quota*), and was fully controlled by the city government. From 2008 to 2013, *Chinese parallel purchasing seats* (CPPS) mechanism was used to assign students to schools; it was an indirect extension of the *Chinese parallel* (CP) mechanism (Chen and Kesten 2017).

The CPPS mechanism is not a direct mechanism. When ranking various schools, a potential student's rank-ordered list (ROL) needs to indicate whether she is willing to pay higher tuition to each school that would otherwise deny her admission. This mechanism has some undesirable features. It is not strategy-proof, which allows students to "game" the system by misreporting their true preferences with respect to schools.[5] Moreover, the equilibrium outcomes of this mechanism can be inefficient and unstable. One way to overcome these imperfections—while retaining the option to purchase admission—is to adopt the *cadet-optimal stable mechanism* (COSM) and its variation, the COSM-BRADSO mechanism proposed by Sönmez and Switzer (2013) and Greenberg, Pathak, and Sonmez (2021). These mechanisms are the extension of the *student-proposing deferred acceptance* (DA) mechanism (Gale and Shapley 1962), ensuring stability and strategy-proofness, wherein submitting true

[4] Confidentiality restrictions prevent this city from being identified by name.

[5] The true preference with respect to schools is referred as students' preferences on schools without considering the tuition.

preferences is a weakly dominant strategy. Since the COSM-BRADSO mechanism is better related to the Chinese mechanism in practice (it is formally defined in Section 3), we focus on this mechanism and use the COSM to denote COSM-BRADSO hereafter.

The theoretical properties of these mechanisms motivate us to investigate real-world student behavior and welfare consequences. One difficulty with any empirical analysis of the school choice problem is estimating students' preferences when only the submitted applications can be observed and the adopted mechanism is *not* strategy-proof. Our survey, which covered nearly half of those who graduated from middle school in 2014, aimed to uncover students' true preferences and thereby to some extent, solve the problems associated with assessing those preferences in the presence of strategic behavior.

The first contribution of our analysis is that students have heterogeneous preferences on the "price menu" of schools. We estimate students' preferences in two steps. In the first step, we use survey results to estimate students' preferences over schools without considering the strategic behavior in ROLs. Given that the ZX policy ceased and all students paid the same basic tuition after 2013, the survey data cannot be used to identify any ZX–related parameters (e.g., tuition). In the second step, we use the ROLs submitted in 2012 and 2013 to estimate other parameters. In this step, we assume that students have homogeneous beliefs about the likelihood of being admitted to each school and that they try to maximize their expected utilities in a rational manner. Our estimated results indicate that a one-unit increase in a school's positive reputation (see Section 4.2 for the definition) is associated with the willingness of high-scoring students from communities with high housing prices to pay an additional 296 yuan—or about US$48.5—to attend that school. In contrast, high-scoring students from communities with low housing prices are willing to pay only 184 yuan for the privilege. The competition for college admission in China is fierce and intense. Our results indicate that students from rich communities have higher desires to "consume" the high-quality schools that might help them attain admissions to colleges compared with others.

Our second contribution is the evaluation of the welfare consequences of different mechanisms. We use the simulated matching outcomes under the DA mechanism as a benchmark, then we measure the welfare change when the DA mechanism is replaced by a seat-purchasing mechanism. This replacement could affect a student's welfare in various ways. First, a student may take advantage of the ZX policy to attend a preferred school by paying higher tuition, which may increase her welfare. Second, if this student's score is high enough, then she may stay at the same school and pay the normal tuition. Otherwise, she will suffer a welfare loss by either paying higher tuition to save her seat in this school or being priced out and attending a less preferred school.

Our counterfactual analysis indicates that when the DA mechanism is replaced with the COSM, student welfare is reduced (on average) by 30 yuan when 10% of the seats are reserved for sale (referred to as the ZX quota). The welfare loss increases to 56 yuan when the ZX quota is increased to 30%. These results reflect that the direct influence of the seat-purchasing option decreases students' total welfare given that both of the two mechanisms in question are strategy-proof. If the DA mechanism is instead replaced by the CPPS mechanism, the average welfare increases slightly by 6 yuan when the ZX quota is 10%, and the welfare loss due to purchasing seats is 118 yuan when the ZX quota is 30%. The reason is that more students can attend their preferred schools by gaming the system. These results reveal an interesting phenomenon: If the ZX quota is limited, leaving room for students to game the system may reduce the average welfare loss. However, when the ZX quota is larger, the welfare loss is much larger.

Meanwhile, students from different communities react differently to seat-purchasing mechanisms. When suffering a welfare loss, students from high-housing-price communities prefer to pay higher tuition to keep seats at the same schools. However, those from low-housing-price communities would rather be priced out to less preferred schools. Among the few students who attend more preferred schools by taking advantage of the ZX policy, most are from rich communities. Interestingly, when the ZX quota is large, although more students from poor communities are priced out, high-scoring students (approximately 10%)

from these communities exhibit a stronger incentive to pay higher tuition to stay in better schools than medium- and low-scoring students (approximately 1%). In summary, our results imply that the ZX policy increases inequality among students in terms of their future educational opportunities, and is not determined solely by their welfare measure in monetary terms. Competitive students (high-scoring students), specifically those from poor communities, show a strong incentive to attend better schools compared to other student groups.

Finally, we investigate the impact of the ZX policy on schools. In China, an intense competition exists among high schools regarding admissions. Schools, which suffer a welfare loss under reforms, have a strong incentive to block such reforms. Therefore, our analysis of the policy impact on schools may provide references for policymakers about potential difficulties from reforms. We measure the impact on high schools in terms of: (a) the quality of admitted students and (b) the profit derived from collecting student tuition. The seat-purchasing option helps upper-tier schools collect significantly more tuition with only a limited decline (relative to the DA mechanism) in the quality of their admitted students. Yet for other schools, the seat-purchasing option leads to more uncertainty about both collected tuition and the resulting quality of their admitted students.

This paper is closely related to the theoretical work of Sönmez (2013), Sönmez and Switzer (2013) and Greenberg, Pathak, and Sonmez (2021), who investigate cadet–branch matching in the US military. We extend the theoretical results and complement these outcomes by offering an empirical analysis. Our work is also directly related to the extensive theoretical literature on the centralized school choice problem.[6] More specifically, there is a growing literature that discusses the role of multi-level financial aid in the school choice problem (Hassidim, Romm, and Shorrer 2016). Hassidim, Romm, and Shorrer (2017) discover that, in a matching procedure for Israeli Master's programs in psychology, many applicants make the mistake of highly ranking programs that offer less financial aid.

[6]See Pathak (2011) for a survey on the school choice problem from the perspective of market design.

The research undertaken here contributes to a growing body of empirical work on the school choice mechanism. Agarwal and Budish (2021) review the development of structural estimates of market design models. One strand of that literature uses the preferences reported under non–strategy-proof mechanisms to estimate students' preferences (Hwang 2015; He 2016; Agarwal and Somaini 2018; Calsamiglia, Fu, and Güell 2020). Other papers focus on strategy-proof mechanisms. Abdulkadiroğlu, Agarwal, and Pathak (2017) treat preferences reported under the DA mechanism as students' true preferences and then use those preferences to analyze the demand for particular schools in New York City. Fack, Grenet, and He (2019) propose an approach for estimating preferences that does not require truth-telling to be the unique equilibrium under the DA mechanism. Several empirical papers (e.g., Burgess et al. 2014; Akyol and Krishna 2017; Wang and Zhou 2020; Ajayi 2022) bear similarities to our strict priority setting. Others begin to investigate the effect of transfers in the market design (Agarwal 2015; Agarwal 2017; Bobba et al. 2021).

There is an increasing use of survey data in scholarly research exploring strategic behavior under matching mechanisms. Budish and Cantillon (2012) conduct a survey on students' preferences for offered courses to study the course allocation mechanism at Harvard Business School, and Rees-Jones (2018) provide survey-based evidence of preference misrepresentation. Burgess et al. (2014) use survey data to directly assess the preferences of students regarding schools. Surveys are also used by De Haan et al. (2023) to analyze the Boston mechanism's deficiencies and by Kapor, Neilson, and Zimmerman (2020) to study heterogeneous beliefs in the school choice problem.

Our estimation of students' preferences also underscores the importance of considering the cardinal preference. Abdulkadiroğlu, Che, and Yasuda (2011, 2015) suggest that from an ex-ante perspective, when schools have coarse preferences for students coupled with a symmetric tie-breaking rule, students could fare better under the Boston mechanism than under the DA mechanism, as assessed by their cardinal preferences. Our analysis finds that from an ex-post perspective, when schools have strict priorities for students, a manipulable mechanism such as the CPPS can still yield higher average student welfare for some types of students than the DA mechanism, especially when the number of seats for sale is limited.

The remainder of this paper proceeds as follows. Section 2 provides details on the local ZX policy's background. In Section 3, we present school choice mechanisms that incorporate seat-purchasing options. Section 4 describes our data and the summary statistics. We present the empirical model and our estimates of students' preferences in Section 5, and in Section 6, we conduct counterfactual experiments across mechanisms. Section 7 concludes with a summary of our findings.

2 Background on High School Admissions

The high schools in our focal city can be categorized into various types based on their educational goals for students after completing middle school. These types include general high schools, which can be public or private, aimed at preparing students for colleges and universities within China. Additionally, there are foreign language schools that focus on preparing students for studies at foreign institutions. Fine arts schools cater to students aspiring to attend fine arts colleges in China. Lastly, vocational schools prepare students with skills necessary for the labor market.

The City Education Bureau (hereafter referred to as "the Bureau") mandates that all schools, regardless of type or ownership, participate in the centralized admission system for middle school graduates. Moreover, each student going through this process must register at the school assigned by the system. Thus, no other options are available for students wishing to continue their education in this city.

Annually, at the end of March, the Bureau announces an admissions plan detailing the admission quota to each school. The quota for each public high school j consists of three parts: the quota for early admission (q_j^e),[7] the quota for normal admission (q_j^a), and the

[7] Students eligible for early admission are determined through a separate procedure, not directly impacting the normal admission process, so they are excluded from this analysis.

quota for the ZX policy (q_j^z). The Bureau, not the schools, controls these quotas. Students admitted under any category receive identical education within each high school. In mid-May, students must submit their rank-ordered lists (ROLs) of preferred schools. All students then take the centralized high school entrance exam in early June. From 2012 to 2014, the maximum score for this exam was 665.[8] After grading, a centralized matching mechanism assigns students to schools. Notably, all schools enforce the same strict priority based on exam scores during student admission.

Local public high schools are pivotal in preparing students for college in China. For many, entering a public high school is their sole opportunity for higher education. However, high school education in China extends beyond compulsory levels, and local public high schools can accommodate fewer than half of all middle school graduates. Before the matching procedure, the Bureau establishes and announces a public high school admission threshold (hereafter "the threshold") based on score distribution and seat availability. Only students scoring above this threshold are eligible for admission to public high schools. This threshold ensures the number of qualified students does not exceed available seats in public high schools.

Students can list up to three schools on their ROL and indicate their choice of the ZX option for each. They also need to state if they'll accept a random assignment if rejected by their chosen schools. Since 2008, the CPPS mechanism—with permanency-execution periods $(2, 1)$—has been employed to assign students (details on matching algorithms are described in Section 3). This mechanism concludes after considering each student's three choices[9]. Students admitted with only basic tuition fees are referred to as *normal* students, while those admitted with higher tuition fees are *ZX students*. Unmatched students open to random assignment are randomly placed in public high schools with vacancies. The rest explore alternatives to continue education or join the workforce.

The ZX policy was designed exclusively for public high schools and not for other school

[8]Before 2012, the highest score was 650, and it increased to 780 after 2014.

[9]This mechanism is a *constrained* mechanism as described by Haeringer and Klijn (2009)

types.[10] Tuition for public high schools is based on the student's exam score. Since scores are the only admission criteria, schools set a cutoff for normally admitted students. Normal students pay an annual tuition of 1,600 yuan (roughly $260 in 2013), ensuring affordability. ZX students' tuition depends on their score: scores within 10 points of the cutoff result in a 3,333.3 yuan annual fee; scores 11-20 points above the cutoff pay 5,000 yuan; and scores 21-30 points above the cutoff pay 6,000 yuan.[11] No school can admit a ZX student with a score more than 30 points below its cutoff. On their ROLs, students can only answer "yes" or "no" to the ZX option, meaning they cannot select a specific ZX tuition but must accept the entire package if admitted as ZX students. Following the Ministry of Education's directive, the local education bureau ceased the ZX option post the 2013 admissions process.

3 Chinese Parallel Purchasing Seats (CPPS) Mechanism

In this section, we provide a formal definition of a school choice problem that incorporates the option of purchasing seats. We consider a finite set of students, denoted by I, and a finite set of schools, denoted by $J \cup \emptyset$, where $\emptyset$ represents the situation where a student does not attend any school. Each school offers two tuition levels: c_0 and c_1, where c_0 is the basic tuition paid by normal students, and c_1 is the higher tuition for ZX students.[12] Each school j has two quotas, q_j^a and q_j^z, for normal and ZX students, respectively. It holds that $\sum_{j \in J}(q_j^a + q_j^z) \geq n$, where n represents the total number of students. Each student i has a strict preference, denoted by π_i, over schools and tuition. The notation $(j, c_0)\pi_i(j, c_1)$

[10] The college admission rate for the best private high school is below 1%, making it lower than even low-quality public high schools. Private schools charge a regulated flat tuition fee, with admission cutoffs equal to the public high school threshold every year. Essentially, private high schools mainly cater to students scoring below the public high school threshold but still wishing to continue their high school education. Our analysis doesn't delve into these schools.

[11] ZX students pay a lump-sum for all three high school years, unlike normal students who pay annually.

[12] The model can be readily expanded to incorporate multiple levels of tuition, as discussed in Sönmez and Switzer (2013). However, in our focal city, ZX students are presented with a singular ZX tuition package and can only decide whether to accept or reject it, rather than selecting a specific tuition level. For the sake of simplicity, we focus on two tuition levels in our model.

indicates that student i strictly prefers to pay the basic tuition for a seat in school j rather than paying the higher tuition for the same school. All schools employ a uniform strict priority ranking, denoted by $\succ$, to order students based on their exam scores. When student i is allocated a seat in school j by paying tuition c, the pair (j, c) is termed as student i's *assignment*. A matching X is defined as a collection of student-to-assignment pairings that satisfies two conditions: (a) each student has only one assignment, and (b) no school admits more students than its total quota. We denote the set of all matching outcomes as $\mathcal{X}$.

A *mechanism* is defined as a strategy space Π_i for each student i, accompanied by an outcome function $\psi : (\Pi_{i_1} \times \Pi_{i_2} \times \cdots \times \Pi_{i_n}) \to \mathcal{X}$, which selects a matching outcome for each strategy vector $\mathbf{a} \in (\Pi_{i_1} \times \Pi_{i_2} \times \cdots \times \Pi_{i_n})$. A *direct* mechanism is a function ψ that selects a matching outcome for each preference profile.

The CPPS mechanism is an extension of the Chinese parallel (CP) mechanism (Chen and Kesten 2017). However, unlike the CP mechanism, the CPPS mechanism is not a direct mechanism. Specifically, under the CPPS mechanism, each student is required to provide (i) her ranked preferences for schools and (ii) indicate, for each ranked school, whether she would opt for the ZX option (i.e., paying higher tuition) to attend that school if she is not assigned a seat as a normal student.

Under the CPPS mechanism, schools allocate the normal seats based on the normal priority ($\succ$). For the allocation of ZX seats, each school j employs the ZX priority $\succ^+$, which is constructed as follows: All applicants for ZX seats at school j are divided into two groups: the ZX-qualified group A_j, comprising students who opt for the ZX option for school j and meet the predetermined qualification rule (related to the school's normal priority) and the remaining applicants in group B_j. When school j compares two students i and i', the following rules apply: If $i \in A_j$ and $i' \in B_j$, then $i \succ^+ i'$, indicating that student i is given higher priority over i'. If $i, i' \in A_j$ or $i, i' \in B_j$, then $i \succ^+ i'$ if and only if $i \succ i'$. In other words, school j's preference is solely determined by students' exam scores in this case.

The ZX priority indicates that a qualified ZX applicant has a higher priority for receiving a ZX seat compared to an applicant who either doesn't choose the ZX option or selects the ZX option but does not meet the qualification criteria. In all other cases, school j utilizes the normal priority to allocate ZX seats.[13] As mentioned in Section 2 concerning our focal city, a ZX applicant is considered qualified for a ZX seat only if her exam score falls within a range of 30 points below the school's normal admission cutoff (Group A). Students whose scores are more than 30 points below the cutoff are not qualified for admission as ZX students. In essence, opting for the ZX option can give a student the privilege to take a ZX seat under the condition that her score is not too low.

The CPPS mechanism with a *permanency-execution period* vector, $\mathbf{e}$=$(e_1, e_2, \ldots)$, selects the matching outcome as described below.

Round 1:

• Every student applies to her first choice. Each school j applies the normal priority to tentatively reserve the top q_j^a applicants in the normal pool. Among the remaining applicants, the school tentatively reserves the top q_j^z applicants in its ZX pool, following the ZX priority. All other applicants are rejected.

In general:

• Any rejected student i who has not yet applied to her (e_1)th-choice school applies to her next-preferred school. A student who has been rejected by all her first e_1 choices does not apply to any other school until the next round. Each school j evaluates the new applicants, along with those already held in the normal pool, and tentatively reserves the top q_j^a applicants in its normal pool based on the normal priority. Subsequently, school j considers all remaining applicants, along with those already held in its ZX pool, and tentatively reserves the top q_j^z applicants based on the ZX priority. All other applicants are rejected.

• The round terminates when each student is either held in a school's pool or has been

[13] All other cases encompass the following scenarios: (1) when both applicants choose the ZX option and are qualified, (2) when both applicants choose the ZX option but neither are qualified, (3) when one applicant chooses the ZX option but is not qualified, and the other applicant doesn't choose the ZX option, and (4) when neither of them chooses the ZX option.

rejected by all her first e_1 choices. At this stage, all tentative assignments become final. The remaining quotas for each school are denoted as $q^a_{j,2}$ and $q^z_{j,2}$ for normal and ZX students, respectively.

Round $k > 1$

• Each remaining student applies to her $\left(\sum_{j=1}^{k-1} e_j + 1\right)$th-choice school, while each school begins to utilize its remaining quota $q^a_{j,k}$ and $q^z_{j,k}$. The algorithm follows the same procedure as in Round 1, considering a student's choices up to her $\sum_{j=1}^{k} e_j$th preference in round k.

• The algorithm terminates when each student is admitted to a school, and all tentative assignments become final. A student who is allocated a normal seat pays tuition c_0. If a student chooses the ZX option for certain schools and is qualified as a ZX student, and subsequently receives a ZX seat, she will pay the higher tuition c_1. On the other hand, if a student either does not choose the ZX option or chooses it but is not qualified, and still receives a ZX seat, she will only pay the basic tuition c_0.

The "Boston mechanism with purchasing seats option" (BMPS), is a special case of the CPPS mechanism when $e_j = 1$ for all j. In that case, the assignments made after each choice are final.

It's evident that revealing one's true preferences regarding schools under the CPPS mechanism may not be a weakly dominant strategy (refer to Appendix A for an example).[14] Owing to this flaw, Sönmez and Switzer (2013) and Greenberg, Pathak, and Sonmez (2021) have examined a similar system that matches cadets to military bases in the United States. They propose the "cadet-optimal stable mechanism" and its variation, the "cadet-optimal mechanism with BRADSO program" (referred to as COSM), which are strategy-proof, meaning that reporting the true preference becomes a weakly dominant strategy for each student. These mechanisms also maintain the option for players to consider the "purchasing" option.

In our specific context under the COSM mechanism, each student's strategy space is

[14] Since the CPPS mechanism is not a direct mechanism, we can only discuss a student's true preferences regarding schools. In a direct mechanism, a student's true preference involves comparing different pairs (j, c), reflecting the student's preference for both school and tuition.

denoted as Π, making it a direct mechanism. Under this mechanism, the ZX priority of school j, represented by $\widetilde{\succ}$, is adjusted according to the following rules: (i) When i's application is (j, c_1) and i''s application is (j, c_0), then $i \widetilde{\succ} i'$ if and only if i is in the ZX-qualified group A_j. (ii) When both applicants choose either (j, c_0) or (j, c_1), or if student i is in the group B_j, then $i \widetilde{\succ} i'$ if and only if $i \succ i'$, meaning that student i has a higher exam score than student i'.

Based on the submitted preference lists, the COSM selects the outcome as follows.

Round 1. Each student applies to her first choice. Each school j applies the normal priority to tentatively reserve the top q_j^a applicants whose first choices are (j, c_0) in the normal pool. Among the remaining applicants, the school tentatively reserves the top q_j^z applicants whose first choices are either (j, c_1) or (j, c_0) in its ZX pool, following the ZX priority ($\widetilde{\succ}$). All other applicants are rejected.

Round $k > 1$. Any rejected student then applies to her next choice. Each school j considers the new applicants whose choices are (j, c_0) along with those who are held in the normal pool from the previous round; then each j tentatively reserves the top q_j^a applicants in the normal pool based on the normal priority. Among the remaining applicants, j considers the new applicants whose choice is (j, c_1) or (j, c_0) along with those who are held in its ZX pool from the previous round; it then reserves the top q_j^z applicants based on the ZX priority ($\widetilde{\succ}$). The other applicants are rejected.

This algorithm terminates when each student is tentatively held by a school, at which point the tentative assignments become final. A student i who is assigned a seat in j pays tuition c_0 if her assignment is (j, c_0) or pays c_1 if the assignment is (j, c_1).

The COSM is a stable mechanism wherein the matching outcome meets the following conditions (i) there are no unselected assignments (i, j, c) where student i prefers assignment (j, c) over her current assignment and has a high enough priority to be selected by school j after paying cost c; (ii) no student prefers a pair (j, c) with an unfilled quota to her current assignment; and (iii) no school would prefer to reject one of the assignments that includes it. Moreover, the matching outcome under the COSM is weakly preferred by all students compared to any stable matching.[15] Conversely, the CPPS and BMPS mechanisms have some deficiencies. Firstly, the matching outcome under the CPPS mechanism may not be stable and can be Pareto-dominated by the COSM mechanism. Secondly, although the BMPS mechanism can reach an equilibrium outcome that is stable, it can still be Pareto inferior to the outcomes achieved under the COSM mechanism (The formal statement of these theoretical results and proofs is in Appendix A).

4 Data Description

4.1 Data Source and Sample Selection

As our analysis centers on the ZX policy, specifically designed for public high schools, we direct our attention to students qualified for admission to these schools.

We utilize a dataset comprising two components: administrative data and survey data (City Bureau of Education 2014). The 2012–2014 admission records include a total of 41,939 students. First, we exclude the 13.3% of students admitted by schools under special quotas, as these did not influence the standard admission procedure.[16] Next, we exclude the 48.6% of students whose exam scores fell below the threshold, rendering them unqualified for admission to public high schools. Finally, we exclude 11.35% of students either deviating from the assignment's official rules or lacking a home address.[17] Following these exclusions, our final sample size from the administrative data is 11,217.

We surveyed middle school graduates in early May 2014. The survey asked each student to list five high schools that they might attend and to rank them based on their preferences

[15]Additional theoretical properties of the COSM are proven by Sönmez and Switzer (2013) and Greenberg, Pathak, and Sonmez (2021).

[16]This category includes students admitted early or through fine arts schools, as well as those on sports or art scholarships.

[17]For example, a few students were assigned to schools at which the cutoff was higher than their actual exam scores.

(the survey questions can be found in Appendix H). Students were explicitly instructed to report their genuine preferences, with no compelling reasons to do otherwise. Given that the survey was conducted just two weeks prior to students submitting their ROLs, a shift in their preferences within that brief window seems unlikely (the detailed discussion of our survey is in Section 4.3). We surveyed 49.17% (6,980) of the middle school graduates in 2014. After matching these students with the final administrative data sample and removing invalid observations[18], we were left with 1,447 survey observations for subsequent analysis. Therefore, our survey covered 43.74% of the selected sample in 2014.

The subsequent section provides a brief summary of school characteristics. Following that, we discuss the students' strategies in the admission procedure. In Section 4.4, we explore the correlation between living areas and admission results.

4.2 School Characteristics

In the administrative data, nonpublic high schools were assigned a single number for coding purposes. Consequently, we treated all these schools as a collective entity without making any distinction among them. Table 1 presents an overview of the characteristics of public high schools throughout the study period. There were 13 public high schools and three special classes in 2012. Special classes are established to accommodate gifted students and operate independently, with their own admission quotas in the matching process. The last row of the table shows changes in the total number of public high schools, accounting for the inclusion of special classes in certain years.

There is a significant variation in the normal admission quotas.[19] The decline in the average normal admission quota over the years can be attributed to the introduction of new special classes. The average quota for ZX students ranges between 95 and 100 across years, with a standard deviation of about 35. Four public high schools and all special classes do not

[18]Invalid surveys include responses such as students who ranked no school or only one school in the survey

[19]The largest school can admit 600 students, while at the other end of the spectrum, a small, "special class" school admits only 40 or 80 students each year. School quotas are listed in Table B.3 of Appendix.

admit ZX students.[20] The table's fourth row indicates that the number of schools providing dormitories increased from nine in 2012 to thirteen in 2014.[21]

To assess the reputation of public high schools, we use college admission rates as a proxy. Chinese students and parents commonly rely on these rates as popular indicators to assess a school's reputation or quality. The average reputation score for schools is 0.8, with a standard deviation of 0.12.[22] Although other reputation-related factors like teachers and facilities are not directly observable in our dataset, we account for them by incorporating schools' fixed effects into our estimation. When assessing school reputation, we do not employ a separate approach to measure schools' added value. This is because students rarely consider this added value when evaluating schools, instead relying on straightforward indicators such as school rankings or college admission records. For instance, empirical evidence from Abdulkadiroğlu et al. (2020) demonstrates that parents' preferences show no correlation with the school effectiveness and academic match quality after controlling for the quality of peers. In our estimation, we aim to replicate students' decision-making strategies; thus, little can be gained by considering the "true quality" of schools.[23]

A school's admission cutoff is indicative of its popularity among students. We classify a school as "popular" if its first-round cutoff is higher than the threshold, indicating that

[20]In fact, special classes and one public high school are not *allowed* to admit ZX students. The other three public high schools are termed as the "leftover" schools in the admission procedure. These schools admit students with scores above the threshold and then, if any unassigned seats remain, they admit ZX students with scores below the threshold.

[21]Whether to live in a dorm is optional for admitted students.

[22]The college admission rate includes admissions to both four-year colleges (benke) and three-year specialized postsecondary colleges (dazhuan). To standardize the measurement, we multiply the percentage grade by 100. For example, if a school's reputation (college admission rate) is 80%, we record it as 80 rather than 0.8.

[23]The assumption that students seldom consider schools' added-value when evaluating schools follows from the prior literature in school choice studies (e.g., He 2016; Calsamiglia, Fu, and Güell 2020), who utilize schools' exam scores to measure school quality. Furthermore, Lai, Sadoulet, and De Janvry (2011) and Lai, Sadoulet, and Janvry (2009) provide evidence on how students' parents evaluated schools based on academic achievement using data from Beijing. One reason students and parents may struggle to assess schools' added value is the lack of information available about schools. In the city under study, comprehensive information about aspects such as teacher quality is not accessible even to researchers like us. Instead, schools' college admission records are reported by local media every year after the college entrance exam, and some high schools employ these records as a marketing tool to promote their performance to potential applicants.`https://www.sohu.com/a/157569093_278291`

the demand for admission to these schools exceeds the available seats.[24] Among the popular schools, two institutions consistently exhibit the highest cutoffs and maintain notable gaps relative to other schools (Figure B.1 in Appendix B). We refer to these two schools as “upper-tier” schools, while the remaining popular schools are considered “middle-tier” schools. Additionally, schools whose cutoffs are equal to the threshold are referred to as “leftover” or “lower-tier” schools.

Table 1: School Characteristics

	2012		2013		2014	
	mean	s.d	mean	s.d	mean	s.d.
Reputation	0.8	0.12	0.81	0.12	0.83	0.11
Normal Quota	215.9	182.6	197	183.8	186.4	164.5
ZX Quota	94.8	37.9	101.3	33.1		
♯ of schools with dorms	9		11		13	
♯ of schools†	16		18		19	

Notes: The schools and special classes that did not admit zx students are excluded when the zx quota is calculated. † includes both high schools and special classes.

4.3 Student Characteristics and Behaviors

The exam score distributions are summarized in the first panel of Table 2. The first column presents the percentile benchmarks, and the next three columns display the corresponding absolute scores for each year. Although exam scores in 2013 were slightly lower than those in 2012 and 2014, the difference in absolute scores for the same percentile level never surpassed 1.7% of the total possible score. This finding confirms the stability of score distributions across the years.

Our analysis focuses on students who were eligible for placement in public high schools. In 2012, approximately 94.3% of these students secured seats in public high schools, compared to 95.1% in 2013 and 90.3% in 2014. These values indicate that the majority of qualified students chose to attend public high schools rather than opting for other types of schools.

[24] Since no school’s second-round cutoff was higher than the threshold when its first-round cutoff was equal to the threshold, there is no confusion in considering only the first-round cutoffs to determine popularity.

The second panel of Table 2 presents the distribution of the number of schools listed on students' ROLs. More than 93% of students provided a complete list of three schools, whereas approximately 5% listed two schools. Fewer than 1% of all students included just one school in their list.[25]

The third panel of the table provides insights into the assignment results, showing comparable patterns between 2012 and 2013. Approximately 30% (resp. 37%) of students were assigned to their first (resp. second) choice, while around 11% to 13% of students were rejected by all three of their listed schools. Roughly 14% (resp. 6%) of students were assigned to their first (resp. second) choice as ZX students. Notably, no ZX student was assigned to their third choice. Following the cancellation of the ZX policy in 2014, fewer students (26%) were assigned to their first choices and more students (17%) faced rejection by all three choices.

The Chinese parallel mechanism is not strategy-proof, and the ROL in our focal city is subject to constraints. Experimental findings from Chen and Kesten (2017) and Calsamiglia, Haeringer, and Klijn (2010) suggest students might opt for reachable schools on their ROLs rather than indicating their genuine preferences. Furthermore, considerable score disparities between different choices on the ROLs might influence students' chances of admission. However, assessing the extent of students misrepresenting their true preferences solely based on submitted ROLs is challenging.[26] Our survey data allow for direct comparisons between each student's actual ordinal preferences and their strategic behavior. More than 60% of the surveyed middle school graduates ranked five schools, while 17% ranked four schools, and approximately 21% ranked fewer than four schools (see Table B.1 in Appendix B).

[25]Schools listed multiple times in the same ROL are counted as one school.

[26]Students can misreport their preferences in various ways. For example, if a student's true preference is $j_1 \pi_i j_2 \pi_i j_3$ and she needs to submit the ROL under the Chinese parallel mechanism with permanency-execution periods $(2, 1)$, she may choose to submit a ROL as (j_2, j_3, j_1) if she believes that attending j_1 is unlikely and placing j_2 and j_3 before j_1 may yield a better expected outcome. Alternatively, she could submit a ROL as (j_1, j_3, j_2) if she thinks she has a chance of being admitted to j_1 and selecting j_3 as the second choice is a safe option since the second-round assignment could be final under the current mechanism. It is also possible for her to report her true preferences if she is confident in being admitted to any of these schools.

Table 2: Student Characteristics

	2012 (1)		2013 (2)		2014 (3)	
Score Distributions						
Percentile	Abs. Scores		Abs. Scores		Abs. Scores	
90th	597		590.5		598	
80th	579.5		572		578	
70th	562		553.5		557.5	
60th	542.5		531.5		532.5	
Threshold	535		530		535	
	Freq.	Percent	Freq.	Percent	Freq.	Percent
Rank Ordered Lists						
3 Schools	3696	94.33%	3793	95.04%	3100	93.71%
2 Schools	191	4.87%	167	4.18%	189	5.71%
1 Schools	31	0.79%	31	0.78%	19	0.57%
Assignment Results						
1st Choice	1153	29.43%	1227	30.74%	875	26.45%
ZX students	542	13.83%	599	15.01%		
2nd Choice	1441	36.78%	1545	38.71%	1290	39%
ZX students	217	5.54%	262	6.56%		
3rd Choice	803	20.50%	751	18.82%	565	17.08%
ZX students	0	0	0	0		
Rejected by all 3	521	13.30%	468	11.73%	578	17.47%
Total observations	3918	100%	3991	100%	3308	100%

Notes: The first panel shows the score distribution. The first data column gives the percentile benchmarks, and the next three columns report the corresponding absolute scores. The second panel reports the number of schools on students' submitted ROLs. The third panel shows the assignment results.

Unlike typical surveys that aim to uncover students' preferences regarding their favorite choices (Budish and Cantillon 2012; Kapor, Neilson, and Zimmerman 2020), we did not ask them to simply rank their favorite schools. Instead, we asked them to rank schools they think they might attend based on their true preferences. It is important to note that the exam score is the sole admission criterion, and the highest admission cutoff could be more than 80 points higher than the lowest cutoff. Our survey design was intended to prevent situations where a low-scoring student ranks schools at which they have no chance of admission, even though they could include three schools with low cutoffs in their ROL. This approach aimed to minimize the underreporting of lower-tier schools. However, if such underreporting does occur, it may result in less accurate estimates of preferences for lower-tier schools, particularly regarding school fixed effects. Focusing solely on a few favorite schools would make it challenging to differentiate students' preferences among lower-tier schools.[27] When conducting the counterfactual analysis to assess the welfare of low-scoring students, understanding their preferences regarding attainable schools is more valuable than their preferences for favorite but unattainable schools. In our survey, all schools received substantial representation in students' responses. It's noteworthy that each of the remaining three schools was chosen by more than 100 low-scoring students. In contrast, these institutions were seldom mentioned by their high-scoring counterparts (Table B.2 in Appendix B).

Figure 1 shows the average admission cutoffs of schools chosen by students both in the survey and in their ROLs.[28] Students are categorized into four groups based on their score percentiles. In the survey, high-scoring students (with exam scores above the 90th percentile) have average school cutoffs of 606.1 and 599.4 for their first and second choices, respectively. The average cutoff for third choices is 593.2, which is 6 points lower. Additionally, the survey reveals that the cutoff gaps between the third and fourth choices, and between the fourth and fifth choices, are 5 and 9 points, respectively. Students in the other three score percentile

[27]For example, suppose there are four schools, namely A, B, C and D. Most students' most favorite school is A, and B is the second favorite, then we can only infer that students prefer A to B to C and D, but it is difficult to tell students' preference between C and D.

[28]The corresponding table can be found in Appendix B.

groups exhibit similar patterns. Within each group, the gap in average cutoffs between consecutive choices is around 6 points and never surpasses 10 points. When comparing between groups, the average cutoff for the first choices of students in the 80th to 90th percentiles is 6 points lower than that of the highest decile of students. Furthermore, this average cutoff decreases by an additional 9 points (to 591) for students in the 70th to 80th percentile range. Students below the 70th percentile of exam scores have an average first-choice cutoff of 585. With each additional choice, the average cutoffs decrease similarly (at a rate of 4 to 10 points) based on exam scores.

The observed decrease in the average cutoff for students' first choices as their scores decrease suggests that the surveyed students provided truthful responses by listing and ranking schools to which they had realistic admission chances. The gaps between consecutive choices within groups in the survey indicate that students' preferences for schools decrease with the popularity of those schools. For example, in 2014, the gaps in consecutive cutoffs for two sought-after schools ranged from 3 to 9 points. Additionally, the small cutoff gaps (4 to 10 points) between consecutive choices within each group implies that the preferences reported in the survey are reliable enough to be viewed as the students' true preferences.

In the rank-ordered lists, the average cutoffs for the first choices of students whose exam scores were above the 70th percentile nearly coincide with the corresponding parts in the survey. However, the average cutoffs for the first choices of low-scoring students (i.e., with exam scores below the 70th percentile) are 6 points lower than that in the survey. Notably, the gap between the first and second choices increases significantly as exam scores decline. While the gap in average cutoffs between the first and second choices for the top 10% of students remains almost the same as in the survey, it increases to 19 points for students in the 80th to 90th percentile range, and approximately 25 points for the two groups of low-scoring students. Furthermore, the average cutoffs for third choices remain consistently around the 535 threshold across all groups in the ROLs.

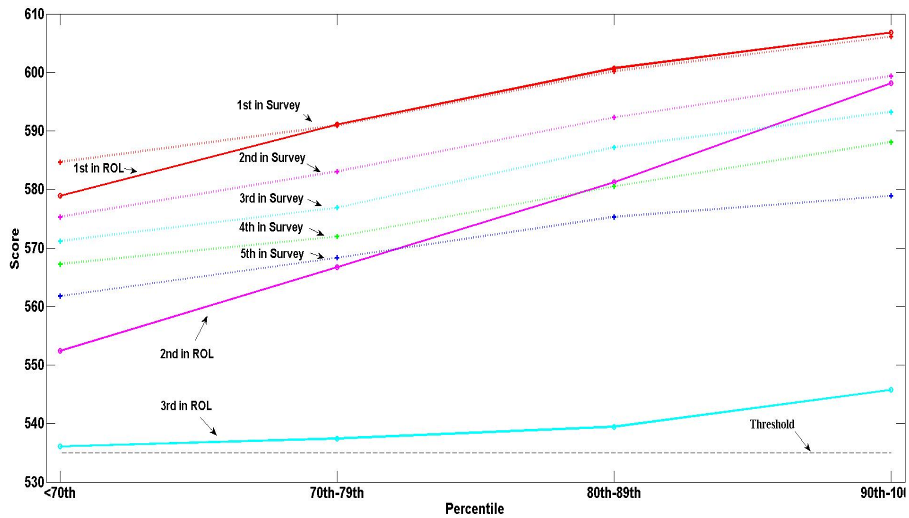
When compared to the survey data, the significant gaps between consecutive choices in the ROLs indicate students' strategic behavior in their submitted preferences: maintaining a substantial gap between choices with the intention of increasing their chances of admission to some school.[29] The correspondence between the first choices in the survey and the ROLs implies that students prefer to apply to their favorite attainable schools. This coincidence, along with the small cutoff gaps among choices reported in the survey, provides further evidence that the surveyed students accurately reported their five favorite attainable schools. However, it is evident that students, especially those outside the top-scoring group, strategically manipulate their reported preferences in the ROLs to increase their overall likelihood of admission—that is, particularly when faced with rejection from their top choices. Thus the second choices in the ROLs for students in 80th–90th percentile (resp., 70th–80th percentile) closely resemble their fourth (resp., fifth) choices in the survey. Furthermore, a majority of students (across all four groups) selected a leftover school as their third choice because the ROL is restricted to only three choices.

One drawback of a non–strategy–proof mechanism is that students who strategically modify their ROLs may exploit naïve students who reveal their true preferences (Pathak and Sönmez 2008).[30] We directly compare the schools listed in the survey and in the ROLs. Only 20 students (1.38% of all observations) submitted ROLs that matched their survey lists. However, it is possible that the number of naïve students is even smaller, as reporting true preferences could be a weakly dominant strategy for some students, especially those in the top-scoring groups. Additionally, some students may exhibit risk aversion by reporting their true preferences. In both cases, these students are not naïve players in the game. Therefore, the 1.38% figure can be considered an upper bound for the number of naïve students. These findings are consistent with previous research, suggesting that only a small number of students submit a rank-ordered list without strategic considerations, particularly

[29] This finding is consistent with the literature that suggests students behave strategically under non–strategy-proof mechanisms (see e.g. Abdulkadiroğlu, Pathak, and Roth 2005; Chen and Sönmez 2006; Abdulkadiroğlu et al. 2020).

[30] Calsamiglia, Fu, and Güell (2020) indicate that, in Barcelona's local school choice setting, the proportion of such naïve students is less than 4%.

Figure 1: Average Admission Cutoffs of Schools: Survey versus ROLs



Notes: The y-axis represents absolute scores, and the x-axis represents four student groups categorized by exam scores in percentile. The dotted curves represent the average cutoffs of the chosen schools in the survey. The solid curves represent the average cutoffs of schools in the ROLs. The threshold for public high school admission is 535 (60.95 percentile) in 2014.

when a strict criterion is used for student assignment.[31]

4.4 Housing Prices and Admission Distributions

Our dataset does not contain any individual- or household-related information, such as household income. Additionally, the city does not provide sub-city level aggregate information regarding local residents' income.[32] To analyze whether the ZX policy has heterogeneous effects on households with varying economic statuses, we match students' home addresses with the local housing market information as proxy. Urban economists have previously examined the positive correlation between local housing prices and residents' income (Goodman 1988; Hwang and Quigley 2006). The assumption that higher-income families reside in

[31] In our context, unlike situations where students are assigned based on coarser criteria (e.g., walking zones or siblings), high school admission offers no safe choice for students until their exam scores are known. Consequently, estimating their exam scores becomes a student's initial strategic move. Therefore, one can expect an extremely low percentage of na"ıve students when subjected to admission procedures like those described here.

[32] The city level annual per capita disposable income of an urban household in 2013 was 35227 yuan ($\approx$ 5775 US dollar)

areas with higher housing prices has also been employed in economic analysis (Abraham and Hendershott 1996; Capozza et al. 2002). Similarly, studies in China have detected a positive relationship between housing price and residents' income (Zhang, Jia, and Yang 2016). In our study, the focal city is divided into 85 communities, which are neither government administrative units nor school districts. This division is based on the local housing market and traditional living areas identified by a real estate website (Wang and Zhou 2024).[33] Local public schools receive full funding from the city-level government, and students are not restricted to specific zones when choosing high schools. Therefore, there is no direct connection between housing prices and school quality.

The median housing price in these communities is 13,636 yuan/m^2,[34] with the highest price being 30,405 yuan/m^2, and the lowest price being 3,968 yuan/m^2 (Figure 2).[35] To simplify the analysis, we classify communities with housing prices above the third quartile as high housing price (HHP) communities,[36] communities with housing prices below the first quartile as low housing price (LHP) communities,[37] and the remaining communities as moderate housing price (MHP) communities.

On average, each community has 54.48 students, with a standard deviation of 26. Approximately 35.6% of the students live in HHP communities, 49.8% live in MHP communities, and the remaining 14.7% come from LHP communities. After the cancellation of the ZX policy, the percentage of admissions from HHP communities to upper-tier schools in 2014 was 38.2% (Table 3). This number was lower than the percentages in 2012 (41.9%) and in 2013 (45.8%). Instead, upper-tier schools admitted more students from MHP and LHP communities. These changes cannot be simply attributed to the fluctuation of exam performance (Table 4), as 43.7% of high-scoring students (scoring above the 90th percentile) were

[33] The authors have collected and processed the data on local housing prices from `https://www.58.com/`. This dataset can be downloaded from the replication package: `https://doi.org/10.5281/zenodo.12735964`

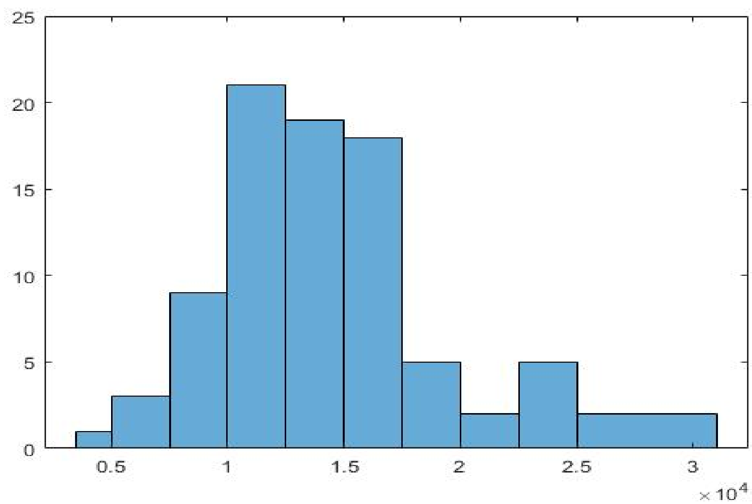
[34] 1 m^2 is equal to 10.76 sq.ft.

[35] The average housing price of this city was 12,187 yuan/m^2 in 2014.

[36] Communities with housing price greater or equal to 16,161 yuan/m^2

[37] Communities with housing price less than 10,609 yuan/m^2

Figure 2: Distribution of the Housing Price



Notes: This is the histogram of communities' housing prices in 2014. The unit of the x-axis is 10,000 yuan/m^2. The unit of the y-axis is the number of communities. Each bin represents 2,500 yuan/m^2 except that the first bin includes the housing price lower than 5,000 yuan/m^2 and the last one includes the housing price above 25,000 yuan/m^2.

from HHP areas in 2014, compared to 46.7% in 2013 and 42% in 2012.[38]

Prior to 2014, upper-tier schools admitted a comparable percentage of students from HHP and MHP communities. However, the HHP area accounted for a larger proportion of ZX students compared to the combined contributions from MHP and LHP communities. Conversely, more than half of the high-scoring students came from the MHP and LHP communities.

These summary statistics at the aggregate level indicate that the cancellation of the ZX policy has provided students from moderate and low housing price areas with increased opportunities to enroll in well-regarded schools. However, the impact of this policy change on middle- and lower-tier schools is not as evident as it is on upper-tier schools. To further analyze all these questions, we estimate students' preferences in the next section.

[38]The upper-tier schools' admission cutoff was set at the 93rd percentile.

Table 3: Community Admission Distribution (%)

	Upper-tier Schools			Middle-tier schools			Lower-tier schools		
	2012	2013	2014	2012	2013	2014	2012	2013	2014
HHP communities.	41.9 (13.3)	45.8 (14.5)	38.2	31.4 (7.8)	38.9 (10.7)	36.9	26.8 (0)	31.0 (0)	30.9
MHP communities	46.5 (9.3)	42.0 (8.3)	48.2	54.7 (11.3)	47.8 (10.9)	49.3	53.7 (0)	48.0 (0)	51.3
LHP communities	11.4 (1.0)	12.1 (1.1)	13.6	14.0 (1.7)	13.2 (2)	13.9	19.4 (0)	20.8 (0)	17.8

Notes: This table indicates the distribution of admitted students from different communities. In each column, the first number represents the percentage of students who live in high, moderate or low housing price communities; the number in the parenthesis represents the percentage of students who are the ZX students in the corresponding communities.

Table 4: Community Score Distribution (%)

	High-scoring students			Median-scoring students			Low-scoring students		
	2012	2013	2014	2012	2013	2014	2012	2013	2014
HHP communities	42.1	46.7	43.7	29.8	37.2	34.1	29.2	33.2	31.7
MHP communities	48.9	44.5	46.4	55.0	47.6	50.5	52.2	46.6	49.9
LHP communities	9.0	8.8	9.9	15.2	15.2	15.4	18.7	20.2	18.5

Notes: This table decomposes each scoring group into different housing price communities.

5 Empirical Model and Preference Estimate

To estimate students' preferences, we adjust the tuition fee structure, based on the local admission rule, as described in the school choice problem from Section 3. Recall that there is a set of tuition fees $C = \{c_0, c_1, c_2, c_3\}$, where c_0 is the basic tuition for normal students while c_1, c_2 and c_3 are the higher tuition amounts paid by ZX students; here, $c_0 < c_1 < c_2 < c_3$.

Student i's (indirect) utility from being assigned to public high school j with tuition $c_{ij} \in C$ is

$$u_{i,j} = \sum_l \beta^l y_j^l + \sum_w \beta^w x_i^w y_j^w + \beta^D d_{ij} + \sum_k \alpha^k x_i^k (c_{ij} - c_0) + \varepsilon_{ij} \tag{1}$$

and that the utility from being assigned to nonpublic high school o is

$$u_{i,o} = F_o + \varepsilon_{io}. \tag{2}$$

Here $\{y_j\}$ represents a vector of observed characteristics for school j; $\{x_i\}$ is a vector of student i's observed characteristics; d_{ij} is the home–school distance;[39] F_o is the fixed effect of nonpublic high schools; and ε_{ij} and ε_{io} are i's idiosyncratic taste for (respectively) public high school j and nonpublic high schools. In the estimate, we assume that the home–school distance is additively separable and independent of unobserved students' preferences; in addition, we normalize the coefficient of d_{ij} for the home–school distance to be -1.[40]

The utility function of students in Equation (2) is similar to that in Abdulkadiroğlu, Agarwal, and Pathak (2017) and Agarwal and Somaini (2018), with the exception that we do not present the random coefficient model for estimating students' heterogeneous preferences for observed school characteristics due to limited variation in our data. In China, the primary goal of general high schools is to prepare students for the college entrance exams. Apart from reputation, the observed characteristics of schools, such as facilities, are fairly homogeneous. The teaching programs are fully controlled by the local education bureau. Additionally, students who are qualified for local public high schools exhibit similar preferences for schools (see Appendix H for details of students' survey responses).[41]

Consistent with Abdulkadiroğlu, Agarwal, and Pathak (2017), we do not explicitly model an outside option. It is because no outside option can be observed in the current admission record, as mentioned in Section 2. In addition, we make the following assumption.

Assumption 1. *The terms ε_{ij} and ε_{io} are independent of the explanatory variables, x_i, y_j, d_{ij}, C, and F_o. Both ε_{ij} and ε_{io} are independent and identically distributed (i.i.d.) and exhibit a type I extreme value distribution with cumulative distribution function (CDF) $F(\varepsilon)$.*

Our estimation of students' preferences employs both administrative and survey data. A key benefit of the survey data is its ability to yield estimates without factoring in the strategic

[39] The road distance d_{ij} is calculated via Google Maps by inputting the focal school's address and the student's home address.

[40] Unlike admission to elementary and middle schools, the high school admission procedure does not consider the locations of school districts or homes. Hence we assume that, in this city, the school choice mechanism does not directly influence residential decisions or local housing prices.

[41] To avoid choosing the wrong empirical model, we consider an alternative random coefficient model and compare the resulting estimates. However, the random coefficient model performs worse than the non-random coefficient model in both within-sample and out-of-sample tests.

behaviors students might exhibit when submitting their ROLs. However, the survey data cannot provide insights into students' preferences regarding ZX options because the ZX policy was discontinued after 2013. Consequently, in 2014, all students paid the same basic tuition for all public high schools. As a result, the survey data alone cannot estimate the parameters α^k in Equation (1). Therefore, we divide our estimation procedure into two steps. First, we use the survey data from 2014 to estimate the vector of parameters *unrelated* to the ZX option—that is, $\{\beta\}$. Second, we estimate the vector $\{\alpha\}$ of parameters related to the ZX policy using the student ROLs submitted prior to 2014.

5.1 Step One: Estimating the Non-ZX-Related Parameters $\boldsymbol{\beta}$

In this step, we focus on the survey data without considering students' strategic behavior when submitting their ROLs. Each surveyed student ranked five schools she believed she had a realistic chance of attending. This selection process implies that the student first identifies the schools for which admission is feasible and then ranks them accordingly. This procedure complicates the construction of the model that explains how these middle school graduates initially select schools. For instance, if a school with a high admission cutoff is not included in a surveyed student's list, it becomes challenging to distinguish between two possibilities: (a) the student prefers the listed schools over the high-cutoff school, or (b) the student believes that admission to the high-cutoff school is not possible. Based on the evidence presented in Section 4.3, we conclude that the survey responses reflect students' true preferences, conditioned on their beliefs about the possibility of admission. To simplify the estimation process, we focus solely on the rankings of the listed schools in the survey and do not consider unlisted schools. In other words, we do not attempt to infer the relative rankings of listed and unlisted schools. While this approach may result in a less efficient estimate, it remains consistent when surveyed students report their true rankings. For example, consider a student i with a score s who lists school j before school j' in her survey; The admission cutoffs for these two schools are denoted as $\bar{S}_j$ and $\bar{S}_{j'}$, respectively. Then the probability that i prefers j to j' conditional on both schools being attainable for her is $Pr(u_{i,j} > u_{i,j'} | s > \bar{S}_j \cap s > \bar{S}_{j'})$. This conditional probability equals the unconditional probability, i.e.

$$\begin{aligned} Pr(u_{i,j} > u_{i,j'} | s > \bar{S}_j \cap s > \bar{S}_{j'}) &= \frac{Pr(u_{i,j} > u_{i,j'} \cap s > \bar{S}_j \cap s > \bar{S}_{j'})}{Pr(s > \bar{S}_j \cap s > \bar{S}_{j'})} \\ &= \frac{Pr(u_{i,j} > u_{i,j'}) Pr(s > \bar{S}_j \cap s > \bar{S}_{j'})}{Pr(s > \bar{S}_j \cap s > \bar{S}_{j'})} \\ &= Pr(u_{i,j} > u_{i,j'}) \end{aligned}$$

The second equality arises from the fact that students' beliefs about admission probabilities only affect whether researchers can observe students' preferences in the survey, but do not alter the relative positions of these preferences. For instance, suppose student i has a true preference for all schools as $j^1 \pi_i j^2 \pi_i j^3 \pi_i \cdots$, i.e., $j^k \pi_i j^{k'}$ when $k < k'$. If the selected and ranked schools in the survey are $j^2 \pi_i j^7 \pi_i j^9$, the relative rank of any two of these schools still preserves the relationship: $j^k \pi_i j^{k'}$ with $k < k'$, regardless of how these schools are selected into the survey. The rankings are independent of the set of schools selected when the top five schools in a feasible set are chosen. We assume that the process of selecting the feasible set is independent of preferences. The equation indicates that one's relative preference over any two schools is independent of the set of schools picked in the survey. Note that we do not assume the selection of feasible schools in the survey is independent of students' scores or that students with the same scores rank the same set of schools.

Given Assumption 1, while referring to the survey data, we use the rank-ordered logit model (Beggs, Cardell, and Hausman 1981) to estimate $\boldsymbol{\beta}$.[42] Given a surveyed student i's ranked school list $(j^1, \ldots, j^l)$ of length $l \leq 5$, we conclude that j^1 is her favorite school among all the l schools on her survey list, that j^2 is her second-favorite school, and so on. The joint probability of these choices is

[42] Because $c_{ij} = c_0$ in this step, $\boldsymbol{\alpha}$ does not appear in the utility function.

$$\Pr(u_{i,j^1} > u_{i,j^2} > \cdots > u_{i,j^l}) = \prod_{k=1}^{l-1} \frac{e^{\mu_{i,j^k}}}{e^{\mu_{i,j^k}} + e^{\mu_{i,j^{k+1}}} + \cdots + e^{\mu_{i,j^l}}}, \tag{3}$$

where $\mu_{i,j}$ is the deterministic component of $u_{i,j}$ or $u_{i,o}$.[43] Then the log-likelihood function can be written as

$$\log L_1(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k=1}^{l_i-1} \mu_{i,j^k} - \sum_{i=1}^{n} \sum_{k=1}^{l_i-1} \log\left(\sum_{s=k}^{l_i} e^{\mu_{i,j^s}}\right). \tag{4}$$

Now we can estimate $\boldsymbol{\beta}$ using maximum likelihood estimation.[44]

5.2 Step Two: Estimating the ZX-Related Parameters $\boldsymbol{\alpha}$

In the second step, we estimate $\boldsymbol{\alpha}$ while considering students' strategic behavior in the admission procedure. After plugging the estimated $\hat{\boldsymbol{\beta}}$ into Equations (1) and (2), we can rewrite student i's utility function as

$$u_{i,j} = \hat{u}_{i,j} + \sum_{k} \alpha^k x_i^k (c_{ij} - c_0) + \varepsilon_{ij}, \tag{5}$$

$$u_{i,o} = \hat{F}_o + \varepsilon_{io}, \tag{6}$$

where $\hat{u}_{i,j} = \sum_l \hat{\beta}^l y_j^l + \sum_w \hat{\beta}^w x_i^w y_j^w + \hat{\beta}^D d_{ij}$.

Calsamiglia, Fu, and Güell (2020) find that 96% of students in Barcelona are strategic players, and Abdulkadiroğlu, Agarwal, and Pathak (2017) use students' reported preferences as their true preferences under the DA mechanism to estimate the parameters based on the assumption that students maximize their expected utilities given their beliefs about admission probability (see Section VI.A in Abdulkadiroğlu, Agarwal, and Pathak (2017)). Therefore, in light of the evidence (from Section 4.3) that few students report their true

[43] More precisely, $\mu_{i,j} = \sum_l \beta_l y_j^l + \sum_w \beta_w x_i^w y_j^w + \beta_D d_{ij}$ when j is a public high school and $\mu_{i,j} = F_o$ when j is not a public high school.

[44] We assume that the utility function has an additively separable form; It is thus easy to demonstrate that $\log L_1$ is globally concave in $\boldsymbol{\beta}$—from which it follows that there exists a unique maximum of the likelihood function.

preferences when submitting ROLs, we model their strategic behavior by assuming that they submit ROLs that are optimal given their beliefs about their likelihood of admission. The assumption of rational expectations will be further supported when we calculate the likelihood function. Our analysis reveals that only a small number of students (1-2%) adopt weakly dominated strategies in their ROLs. The literature suggests that students may hold heterogeneous beliefs about these probabilities (Kapor, Neilson, and Zimmerman 2020) or make mistakes in their ROLs (Hwang 2015; Artemov, Che, and He 2017). However, defining "mistakes" in our administrative data is challenging since students submit their ROLs before taking the exams, and a student may accurately estimate admission cutoffs but face uncertainty about their exam performance. Therefore, we make the following assumption.

Assumption 2. *Students are fully informed about their own preferences, and they maximize their expected utility in a rational manner.*

Students' Decision Problem

Initially, student i submits the ROL $a_i = \{(j^1, v^1), (j^2, v^2), (j^3, v^3), r\}$; here $v^k \in \{0, 1\}$ indicates whether i selects the ZX option for her kth choice j^k, and $r \in \{0, 1\}$ indicates whether i accepts a random assignment if she is rejected by all three of her chosen schools.[45] Subsequently, each student takes the entrance exam and receives a score s_i. The decision problem for student i is to maximize her expected payoff by selecting the best choice a_i from the set of all possible choices A_i.

$$\max_{a_i \in A_i} EU(a_i, s_i) \tag{7}$$

Since the choice set for each individual is larger than 100,000, we follow the backward induction method developed by Calsamiglia, Fu, and Güell (2020) to address the curse of dimensionality for the empirical school choice problem. This approach follows the logic that although the entire ROL is submitted all at once, the ranked schools on the list are considered

[45] We drop subscript i for schools and ZX options for simplicity.

sequentially in the procedure. The kth listed school and its ZX option are relevant only if the student is rejected by all previously listed schools. Therefore, the kth choice should be the student's best choice conditional on reaching this stage, and consequently, the problem can be solved via backward induction. The student's decision problem can be rewritten as:

$$V^k(s_i) = \max_{(j^k, v^k)} \left\{ \bar{\mathbf{P}}_i^k \cdot \mathbf{U}_i^k + [1 - \bar{\mathbf{P}}_i^k \cdot \hat{\mathbf{I}}] \cdot V^{k+1}(s_i) \right\} \tag{8}$$

with $k \in \{1, 2, 3\}$, and

$$V^4(s_i) = \max_{t \in \{0,1\}} \{ I_{r=1} \cdot EU_i^l + (1 - I_{r=1}) \cdot \widetilde{u}_i \} \tag{9}$$

Equation (8) indicates that student i needs to choose school j^k and its ZX option v^k to maximize her value function V^k. $\bar{\mathbf{P}}_i^k$ and $\mathbf{U}_i^k$ are vectors of probabilities and utilities, respectively, associated with i attending school j^k at different tuition levels.[46] Here $\hat{\mathbf{I}} = (1, 1, 1, 1)'$. Equation (9) indicates that if student i is rejected by all three choices, she needs to decide whether to accept a random assignment to a leftover school. EU_i^l represents the expected payoff when i is randomly assigned to a leftover school, $\widetilde{u}_i$ represents the payoff when i has lost all chances to attend a school in the matching procedure, and $I_{r=1}$ is an indicator function of whether student chooses to accept the random assignment.

Admission Probabilities and Beliefs

Given student i's ROL a_i, the conditional probability of her being admitted to her kth choice as a normal student is

$$P_{i,c_0}^k = \Pr\left(\bar{S}_{j^k} \leq s_i | i \text{ is rejected by } j^{k-1}\right). \tag{10}$$

[46] More precisely, $\bar{\mathbf{P}}_i^k = (P_{i,c_0}^k, I_k \cdot P_{i,c_1}^k, I_k \cdot P_{i,c_2}^k, I_k \cdot P_{i,c_3}^k)$. P_{i,c_t}^k represents the probability of student i being admitted to her kth choice with tuition c_t, given that she has been rejected by her previous choices and/or tuition levels. I_k is an indicator function that determines whether i chooses the ZX option for j^k. $\mathbf{U}_i^k = (u_{i,j^k,c_0}, u_{i,j^k,c_1}, u_{i,j^k,c_2}, u_{i,j^k,c_3})'$. It's worth noting that we slightly abuse the notation here: if student i chooses a nonpublic school at any position, we can replace u_{i,j,c_0} in Equation (8) with $u_{i,o}$ as defined in Equation (6) and set $v^k = 0$.

Here $\bar{S}_{j^k}$ represents the cutoff of j^k.[47] Equation (10) indicates that i attends the kth choice school as a normal student if and only if her score is no less than the school's cutoff conditional on being rejected by $k-1$th choice, i.e., $s_i < \bar{S}_{j^{k-1}}$ if i does not choose the ZX option of j^{k-1} or $s_i < \bar{S}_{j^{k-1}} - 30$ otherwise.[48]

The conditional probability of student i being admitted to j^k as a ZX student, with tuition c_t where $t \in \{1, 2, 3\}$, is

$$P^k_{i,c_t} = \Pr\left(\bar{S}_{j^k} - 10t \le s_i | s_i < \bar{S}_{j^k} - 10(t-1)\right). \tag{11}$$

From the perspective of student i, we assume that she anticipates her exam score to be $m_i + \eta_i$; here m_i represents either i's mock exam score or her true ability (by which she estimates her exam score) and η_i is the uncertainty. We assume that η_i is i.i.d. and distributed normally as $N(0, \delta)$. Note that m_i cannot be directly observed from the data. Instead, we use the student's actual exam score s_i as the proxy of m_i. We simplify our estimation process by setting $\delta = 20$, which is 3% of the full mark.[49] After we replace s_i with $s_i + \eta_i$ in Equation (10) to (11), the admission probabilities can be expressed as the CDF of the standard normal distribution (see Appendix D for the functional forms).

Students assess their chances of being admitted to each school before submitting their ROLs. In line with much of the school choice literature, we assume that students consider admission probabilities as exogenous, meaning they can precisely forecast the equilibrium admission cutoffs.[50] While this assumption is not without its complexities and limitations, it enables us to estimate students' preferences without the need to explicitly solve for the

[47] Note that because the CPPS mechanism is an extension of the CP mechanism with an executive period (2,1), the schools' cutoffs used for the first two choices are generated after considering all students' second choices. For the third choice, the schools' cutoff is generated after considering all students' third choices. The similar calculation approach for the general Chinese parallel mechanism can be found in Calsamiglia, Fu, and Güell (2020) and their online Appendix.

[48] When $k = 1$, Equation (10) becomes the unconditional probability: $\Pr\left(\bar{S}_{j^1} \le s_i\right)$

[49] The estimated results when $\delta = 13.3$ (2% of the full mark), when $\delta = 26.6$ (4% of the full mark), and when $\delta = 33.35$ (5% of the full mark) are reported in Section 5.4.

[50] Abdulkadiroğlu et al. 2011; Azevedo and Hatfield 2018; Kojima 2017; Agarwal and Somaini 2018; and Calsamiglia, Fu, and Güell 2020

equilibrium, which can be non-unique in many non-strategy-proof matching mechanisms. Solving for the equilibrium and selecting the appropriate one can be computationally intensive. The price-taking assumption offers the advantage of simplifying the analysis and is commonly employed in the literature for this reason.

In our study, the admission cutoffs of schools are made public after the annual admission season. Analyzing the previous year's data, we observed that the majority of popular schools' cutoff scores (with one exception) in the period between 2011 and 2013 showed no increase greater than 4 units or decrease greater than 2 units in terms of percentile scores (refer to Figure B.1 in Appendix B). Specifically focusing on 2012 and 2013, the absolute change in cutoff scores for each school never exceeded 6 points, which is less than 1% of the total exam scores. Additionally, the set of popular schools and leftover schools remained unchanged across these years. Based on these findings, we assume that students have accurate beliefs about admission cutoffs in the current year, given the stability of cutoffs and the distribution of exam scores.

Alternatively, we can assume students shape their beliefs based on the previous year's admission cutoffs, embracing the concept known as "adaptive expectations." This assumption is reasonable because previous year's cutoffs are announced after the admission season, and therefore can serve as informative reference points for students. To ensure the robustness of our findings, we provide estimated results based on this assumption in Section 5.4.

Likelihood Function and Identification

Using the backward induction approach, we rule out some weakly dominated strategies, thereby limiting the choice set to $A_i' \subset A_i$, as described next.

First, if a student lists a leftover school as her first or second choice, then the remaining choices should be left blank. This ensures that no student will be admitted to a school that is listed after a leftover school.[51] Second, if a student lists a popular school as her first or

[51] We observed that no student was admitted by a school listed after a leftover school.

second choice, then her subsequent choice should not be blank.[52] Third, no student's ROL can include the same school more than once.[53] Fourth, no student selects the ZX option for her third choice. According to the admission records, students are admitted by their third choices only when those choices are leftover schools, which admit all students as normal students.[54] Fifth, a student accepts the randomly assigned school if she is rejected by all of her listed schools. So if a student in those circumstances does not accept the randomly assigned school, then her only option is to attend a nonpublic high school. In the admission records, all nonpublic high schools have admission cutoffs that are below the threshold; in other words, their admission probability is equivalent to that of leftover high schools. The implication is that, if a student would rather attend a nonpublic high school than be randomly assigned to a leftover school, then she should list that nonpublic school as one of her three choices.[55] After excluding these weakly dominated strategies, our expression for a student i's submitted ROL simplifies as follows: $a_i = \{(j^1, v^1), (j^2, v^2), j^3\}$. Hence the choice set A'_i incorporates alternatives and so is significantly smaller than its parent set A_i.

Although we can observe the students' three choices in ROLs, their choices regarding the ZX option cannot be directly observed from the admission records. We can only determine whether a student is assigned to a school as a normal student or a ZX student. However, we can infer the students' ZX options partially or sometimes fully from their assignment results. For example, if a student is assigned to her first choice as a ZX student, we can infer that she must have selected the ZX option for that choice. If she is assigned to the second choice but is qualified for admission by the first choice as a ZX student, we can infer that she did not select the ZX option for the first choice. Therefore, we can categorize the observations into three groups. The first group (G_1) consists of students whose ZX options in ROLs can be

[52]Only 2.4% of students violated this assumption, and merely 1% were not admitted to their first or second choices. It implies that among those who violate this assumption, most were confident they would get into their first or second choices.

[53]In the data set, no student selected a particular school more than once.

[54]Choosing the ZX option as one's third preference doesn't influence the admission outcome.

[55]In the admission records, 1.21% of the students did not accept the random assignment after being rejected by their preferred schools.

unambiguously inferred from the admission records data. The second group (G_2) comprises students whose decisions regarding ZX options can be observed or inferred for either their first choice or second choice but not for both.[56] We use $\widetilde{a}_i$ to denote the partially inferred or observed choice of student i (i.e., when v^1 or v^2 is unknown).

For an observation in G_1, the probability of observing an ROL a_i is written as $\Pr(a_i \in A_i^*)$; here $A_i^* \subset A_i'$ is the optimal solution set of the student's problem in Equation (7). For student i in G_2, we can infer whether i selected the ZX option for her first choice but not for her second choice; however, we do know that she either: (a) selected the ZX option for her second choice, $a_i^+ = \{(j^1, v^1), (j^2, 1), (j^3)\}$; or (b) did not select that option, $a_i^- = \{(j^1, v^1), (j^2, 0), (j^3)\}$. Hence the probability of observing $\widetilde{a}_i$ is $\Pr(a_i^+ \in A_i^*) + \Pr(a_i^- \in A_i^*)$. Similarly, if the ZX choice for the first choice is unknown then the probability of observing $\widetilde{a}_i$ is $\Pr(a_i^+ \in A_i^*)+\Pr(a_i^- \in A_i^*)$, where $a_i^+ = \{(j^1, 1), (j^2, v^2), j^3\}$ and $a_i^- = \{(j^1, 0), (j^2, v^2), j^3\}$.

The total log-likelihood function for the entire sample can be expressed as follows:

$$\log L_2(\boldsymbol{\alpha}) = \sum_{i \in G_1} \log(\Pr(a_i \in A_i^*)) + \sum_{i \in G_2} \log[\Pr(a_i^+ \in A_i^*) + \Pr(a_i^- \in A_i^*)]. \tag{12}$$

Should each student be able to select only one school in the ROL without the ZX option, the identification of the model's parameters could resemble a multinomial discrete choice model. This is established under general conditions (Matzkin 1993). Our model deviates in that every student evaluates the admission probabilities of schools and then selects the option yielding the highest expected payoff (Calsamiglia, Fu, and Güell 2020). In the realm of school choice models, Agarwal and Somaini (2018) offer a comprehensive method, including the necessary identification conditions to discern the distribution of student preferences when all of them are strategic players.[57] Furthermore, Calsamiglia, Fu, and Güell (2020) utilize

[56] Formally, let $\tilde{a}_i = (j_i^1, v_i^1), (j_i^2, v_i^2), j_i^3)$ denote the part of the action i takes. We extend the definition of v_i^k a little bit. Let $v_i^k \in 0, 1, \emptyset$, where $v_i^k = \emptyset$ represents that the researcher doesn't observe or infer the value of v_i^k. Then $G_1 = \{i | v_i^k \neq \emptyset \text{ for } k = 1, 2\}$ and $G_2 = \{i | (v_i^1 \in \{0, 1\} \text{ and } v_i^2 = \emptyset) \text{ or } (v_i^1 = \emptyset \text{ and } v_i^2 \in \{0, 1\})\}$.

[57] Agarwal and Somaini (2018) demonstrate that the distribution of students' preferences can be identified using two types of variation. First, they consider variation in choice environments, where a student's optimal choice is determined based on their belief about the probability of school admission. When these

the discontinuous change of admission probabilities along the boundary of school zones to identify both strategic and non-strategic players. Drawing from the existing literature, our model's identification hinges on the observable variation in decisions made by students who were denied regular admission by either their first or second choices $\cdots$. When students submit their ROLs, they must make a decision regarding whether or not to pay higher tuition in order to increase the probability of being admitted by their first or second choices. To illustrate this, consider a simplified example with only two schools (A and B). Suppose student i lists School A before School B. If she chooses the ZX option for School A, it implies that she prefers attending School A as a ZX student (and paying higher tuition) over attending School B as a normal student (assuming School A rejects her as a normal student); otherwise, she should not choose the ZX option for School A (see Appendix C for a proof of the identification for a general case). There is no closed-form solution to Equation (12).[58] Therefore, we estimate the parameters using the maximal simulated likelihood estimate with the logit-smoothed accept-reject simulator (Train 2009, ch. 5). For detailed information, refer to Appendix D.

5.3 Estimation Results

Table 5 presents the estimated coefficients of the utility function. Columns 1 and 2 display results without considering student-school interaction terms. Column 4 provides the results for the full model without the school fixed effect. Our primary focus is on Column 3, which represents the full model incorporating school fixed effects. Rows 2–4 of Column 3 present students' preferences regarding school reputation. Students are categorized based on their exam scores as follows: high-scoring students (scores above the 90th percentile), medium-

environments change, such as with adjustments to school capacities, students are faced with different sets of lottery choices, leading to a shift in their optimal choices. Second, the authors discuss the use of a regressor that is additively separable in the utility function, such as the distance between home and school. This regressor can be employed to alter preferences, facilitating the identification of the distribution of students' preferences.

[58] No such solution exists because (a) the distribution of the summation of a type I extreme distribution does not itself follow a type I extreme distribution and (b) the ROLs are only partially observed.

scoring students (scores between the 70th and 90th percentile), and low-scoring students (scores below the 70th percentile but above the threshold).

Students' sensitivity to tuition decreases with their living locations. Specifically, students from high housing price (HHP) communities exhibit less sensitive to changes in tuition compared to students from moderate- and low-housing price (MHP and LHP) communities (Rows 25-27 in Column 3). Moreover, as shown in Rows 2-4 of Column 3, high-scoring students demonstrate a stronger sensitivity to school reputation than their medium- and low-scoring counterparts. If school reputation increases by 1 unit, high-scoring students from HHP communities are willing to pay an additional 352 yuan, whereas medium- and low-scoring students from the same type of communities would be willing to pay an extra 334 yuan and 278 yuan, respectively. For students from LHP communities, high-scoring students indicate a willingness to pay 220 yuan for a 1 unit increase in school reputation, while low-scoring students are willing to pay 188 yuan.

The valuation of a school's overall capacity, standardized to 100 seats, varies among students. All students have a preference for smaller schools when other variables are held constant. Notably, medium-scoring students show the most pronounced aversion to larger schools. When school capacity decreases by 100 seats, medium-scoring students are willing to pay an additional 1308 yuan. On the other hand, high-scoring and low-scoring students are willing to pay 954 yuan and 713 yuan, respectively, for the same reduction in school capacity.

Table 5 also presents our estimates for other parameters. In Column 3, Rows 10–12 show that high-scoring students have a somewhat unfavorable attitude toward special classes, while the other two student groups view them positively. In Column 3, Rows 17–18 indicate that a school's provision of dormitory accommodations can alleviate students' concerns about travel distance, particularly for girls.

In summary, student preferences regarding school characteristics are diverse. Interestingly, the estimates reveal that high-scoring students value school quality-related factors, such as reputation and special classes, more than other student groups. This may be attributed to the intense competition for college admissions in China. Highly competitive students have strong incentives to enroll in high-quality schools that can enhance their chances of gaining admission to prestigious colleges in the future. On the other hand, students from low-housing price communities are more sensitive to tuition costs compared to their counterparts. This suggests that they are less willing to pay higher tuition fees to attend a more preferred school.

5.4 Model Fit and Robustness Check

Next, we examine the extent to which our preference estimates align with the data by conducting both within-sample and out-of-sample tests to evaluate aggregate-level matching patterns.[59] Table 6 compares the actual to the predicted admission cutoffs for each high school.[60]

For the within-sample test, Column 3 of the table presents the predicted cutoffs for the schools in 2012. In this year, the gaps between the predicted and actual cutoffs are less than 7.5 points, representing 1.13% of the full mark (665) for 15 of the 16 schools. For the last school, the discrepancy approximates 1.8% of the full mark. Column 6 displays the predicted cutoffs for the schools in 2013. With the exception of one school, the gaps between the actual and predicted cutoffs are less than 1% of the full mark. Additionally, the predicted results accurately identify all the leftover schools, for which the cutoffs are 530 in 2012 and 535 in 2013.

For the out-of-sample test, we employ the parameter estimation procedure from Section 5, but this time we exclude the 2012 data. Using the newly estimated parameters, we simulate the behavior of students based on their preference profiles from 2012. The results, shown in

[59] The simulation of students' behaviors in the sample test follows the same procedure as the counterfactual analysis in Section 6. A detailed description of the procedure can be found in Appendix E.

[60] The reported results pertain to the first-round admission cutoffs for all schools. The actual second-round cutoffs for all popular schools are infinite, while for all leftover schools they are equal to the threshold. Since our predicted results accurately identify all popular and leftover schools, we only report the first-round cutoffs.

Table 5: Preference Parameters

	No student interactions		With student interactions	
	(1)	(2)	(3)	(4)
Reputation	0.835	0.296		
	(0.044)	(0.021)		
Reputation × HS			0.592	0.699
			(0.149)	(0.158)
Reputation × MS			0.209	0.351
			(0.037)	(0.037)
Reputation × LS			0.175	0.327
			(0.029)	(0.044)
Capacity	-0.011	-1.969		
	(0.193)	(0.125)		
Capacity × HS			-0.999	0.158
			(0.850)	(0.847)
Capacity × MS			-1.489	-0.731
			(0.286)	(0.278)
Capacity × LS			-1.064	-0.607
			(0.231)	(0.191)
Special class	-1.006	-2.121		
	(1.201)	(0.910)		
Special class × HS			-6.925	-2.747
			(1.925)	(1.409)
Special class × MS			0.597	1.198
			(1.529)	(0.830)
Special class × LS			6.365	5.294
			(5.575)	(6.764)
Distance	-1	-1	-1	-1
Distance × Male			0.804	0.931
			(0.037)	(0.036)
Same district			-1.905	-2.418
			(0.238)	(0.284)
Same district × Male			1.759	2.606
			(0.301)	(0.320)
Dorm	-3.924	4.253	4.389	-0.741
	(0.329)	(0.967)	(1.021)	(0.332)
Dorm × Male			0.633	0.612
			(0.273)	(0.326)
Cost	-3.072	-2.468		
	(0.002)	(0.001)		
Cost × HS			-1.047	-0.935
			(0.011)	(0.025)
Cost × MS			-1.138	-1.169
			(0.010)	(0.024)
Cost × LS			-1.492	-2.033
			(0.011)	(0.024)
Cost × HHP			-0.637	-0.496
			(0.011)	(0.025)
Cost × MHP			-1.027	-1.060
			(0.011)	(0.024)
Cost × LHP			-1.649	-1.798
			(0.020)	(0.025)
Non-public high school	43.909	2.005	1.052	10.739
	(3.596)	(0.457)	(0.571)	(3.428)
School Fixed Effect		Y	Y	

Notes: Standard errors are reported in parentheses. Distance is measured by kilometer. The coefficient of female's attitude to home-school distance is normalized to -1. Capacity is measured by 100 seats. The unit of cost(Tuition) is 1000 Yuan. "HS", "MS", and "LS" represent high-, medium- and low-scoring students respectively. "HHP", "MHP", and "LHP" represent students from high-, moderate- and low-housing price communities respectively.

Column 9, are similar to those obtained in the within-sample test. Furthermore, we utilize the preferences estimated solely from the survey data to simulate students' choices in their ROLs for the year 2014. All the predicted cutoffs in Column 12 are less than 1% of the full mark. These findings suggest our survey aptly captures the genuine preferences of students for the schools listed.

We further evaluate the aggregate-level matching patterns corresponding to the initial two choices on students' ROLs (see Table F.1 in Appendix F).[61] In the within-sample test, the data show that 29.4% (resp. 30.7%) of students were admitted by their first-choice schools in 2012 (resp. 2013). Our predictions are 32.8% and 35% respectively, which are close to the actual values. More specifically, for 2012, we predict that 12% of students would be admitted by their first choices as ZX students and 3.8% would be admitted by their second choices as ZX students. These predictions are closely aligned with the observed values of 13.8% and 5.7%, respectively. Most discrepancies between predicted patterns and actual values are reasonably close. Similar results are observed in the out-of-sample test.

For the robustness check, we use an alternative proxy for measuring school reputation, which is based on the quality of students admitted by schools in previous years. Specifically, we quantify a school's reputation using the average high school entrance exam scores (percentile grade) of students admitted over the past three years, but we limit this to the scores within the 10th to 90th percentile range.[62] The results are presented in the first panel of Table 7. Continuing the trend, students from HHP communities show the least sensitivity to tuition costs, suggesting their readiness to pay premium prices to join reputable schools. As housing prices in these communities decrease, the sensitivity towards tuition increases. The estimated coefficients for other variables are consistent with the results reported in Section 5.3.

The results presented in Section 5.3 operate under the assumption of rational expec-

[61]We focus on the first two school choices because no student is admitted as a ZX student for the third choice.

[62]Our school quality measure is highly correlated (0.96) with the schools' college admission rate.

Table 6: Admission Cutoffs

	Within Sample						Out of Sample					
	2012			2013			2012			2014		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
School ID	True	Predicted	Diff	True	Predicted	Diff.	True	Predicted	Diff.	True	Predicted	Diff.
141	607.0	603.4	3.6	604.0	598.1	5.9	607.0	603.2	3.8	605.0	600.2	4.8
142	535.0	535.0	0.0	530.0	530.0	0.0	535.0	535.0	0.0	535.0	535.0	0.0
147	555.5	555.0	0.5	552.5	558.0	-5.5	555.5	551.6	3.9	558.0	555.9	2.1
167	592.5	592.3	0.2	590.0	587.2	2.8	592.5	591.8	0.7	593.5	589.6	3.9
173	535.0	535.0	0.0	530.0	530.0	0.0	535.0	535.0	0.0	552.0	547.3	4.7
177	597.0	591.7	5.3	590.5	585.8	4.7	597.0	591.2	5.8	600.5	597.3	3.2
179	571.5	570.7	0.8	565.0	567.0	-2.0	571.5	570.4	1.1	573.5	568.2	5.3
181	535.0	535.0	0.0	530.0	530.0	0.0	535.0	535.0	0.0	535.0	535.0	0.0
183	617.0	613.0	4.0	611.0	606.5	4.5	617.0	613.0	4.0	611.0	609.6	1.4
184	535.0	535.0	0.0	530.0	530.0	0.0	535.0	535.0	0.0	535.0	535.0	0.0
185	583.0	579.5	3.5	580.0	574.0	6.0	583.0	579.7	3.3	583.0	576.5	6.5
186	583.0	576.7	6.3	578.0	571.8	6.2	583.0	575.9	7.1	576.0	573.3	2.7
187	599.5	598.5	1.0	594.5	594.9	-0.4	599.5	598.3	1.2	596.5	593.3	3.2
188	571.5	583.7	-12.2	575.0	571.7	3.3	571.5	582.2	-10.7	580.0	577.2	2.8
28†	594.5	587.0	7.5	589.0	579.3	9.7	594.5	586.4	8.1			
165†	608.5	613.5	-5.0	605.5	608.6	-3.1	608.5	613.5	-5	609.0	607.8	1.2
166†										595.0	594.5	0.5
169†				599.0	596.6	2.4				604.0	603.7	0.3
180†				576.5	577.9	-1.4				584.5	588.6	-4.1
200†										607.0	607.8	-0.8

Notes: This table indicates the within- and out-of-sample tests for the schools' cutoffs, using the estimated coefficients in Columan 3 of Table 5. The full mark is 665. The threshold is 535 in 2012 and 2014, and 530 in 2013. † indicates the special class. The number of special classes varies with years

tations, wherein students are believed to precisely predict schools' admission probabilities. Given that students finalize their ROLs prior to taking the entrance exam, it stands to reason that their main information source leans on data from previous years (adaptive expectations). In the second panel of Table 7, In Column 1, we present the parameters estimated in relation to the ZX option, assuming students base their estimated chances of admission on the prior year's cutoffs.[63] These results align closely with those stemming from rational expectations. This alignment indicates consistent stability in admission cutoffs over the years and affirms that students' perceptions, formed using the preceding year's cutoffs, yield fairly accurate estimates.

Uncertainty tied to their exam scores can shape the strategic behavior of students. In the second panel of Table 7, Columns 2–4 in the second panel of Table 7 showcase parameters estimated in relation to the ZX option. Here, the standard deviation δ of the exam scores is

[63]Note that our estimation of the non-ZX-related parameters does not rely on any assumptions about student beliefs.

designated as 13.3 (2% of the full mark), 26.6 (4% of the full mark), and 33.35 (5% of the full mark), respectively. These results exhibit a similar pattern to the findings presented in Table 5.

6 Counterfactual Analysis

To assess the impact of the ZX policy on student welfare, we conduct simulations to compare different mechanisms. Specifically, we use two benchmark mechanisms for comparison: the DA mechanism and the cadet-optimal stable mechanism (COSM). Economists widely recommend the DA mechanism, known for its strategy-proof nature. Under the DA mechanism, students have a weakly dominant strategy of truthfully reporting their preferences. Empirical studies suggest that students' strategies in practice align closely with theoretical predictions (Abdulkadiroğlu, Pathak, and Roth 2009; Abdulkadiroğlu, Agarwal, and Pathak 2017; Pathak and Shi 2021). Therefore, the DA mechanism is commonly used as a benchmark in counterfactual analyses of different mechanisms (Agarwal and Somaini 2018). Pathak and Shi (2021) analyze the effectiveness of such counterfactual analyses of the school choice problem. Following a similar approach to Calsamiglia, Fu, and Güell (2020), we simulate the students' true-telling strategy under the DA mechanism to compare the welfare from other mechanisms. The COSM, another benchmark mechanism, is an extension of the DA mechanism. Under the COSM, players still have a weakly dominant strategy to truthfully report their preferences. Sönmez and Switzer (2013) provide evidence that the strategies of US military cadets largely conform to the theoretical predictions. Therefore, we assume that students will truthfully report their preferences in their ROLs under the COSM mechanism.

Based on the estimated preferences, we simulate students' application lists. For the simulation, we utilize the student and school profiles from the 2014 administrative data. Given the absence of ZX students that year, we treat the normal admission quota as the school's total capacity. To analyze the welfare impact of different ZX quotas, we conduct experiments under two setups: one with the ZX quota representing 10% of the total quota

Table 7

Reputation × HS	0.225		Capacity × HS	0.365
	(0.054)			(0.545)
Reputation × MS	0.094		Capacity × MS	-0.808
	(0.022)			(0.248)
Reputation × LS	0.052		Capacity × LS	-0.569
	(0.023)			(0.176)
Special class × HS	-6.625		Distance	-1
	(2.312)			
Special class × MS	1.060		Distance × Male	0.794
	(2.072)			(0.036)
Special class × LS	6.377		Dorm	4.893
	(8.139)			(1.209)
Score range	-0.043		Dorm× Male	0.791
	(0.445)			(0.318)
Score range × Male	0.534			
	(0.554)			
Same district	-1.888			
	(0.234)			
Same district × Male	1.794			
	(0.298)			
Non-public high school	-9.113		School Fixed Effect	Y
	(1.192)			
	(1)	(2)	(3)	(4)
Cost × HS	-1.242	-1.011	-1.162	-1.163
	(0.008)	(0.017)	(0.009)	(0.010)
Cost × MS	-1.430	-1.171	-1.391	-1.391
	(0.006)	(0.015)	(0.013)	(0.013)
Cost × LS	-1.580	-1.423	-1.674	-1.673
	(0.007)	(0.015)	(0.012)	(0.012)
Cost × HHP	-0.671	-0.635	-0.694	-0.694
	(0.006)	(0.012)	(0.013)	(0.013)
Cost × MHP	-1.051	-1.044	-1.121	-1.121
	(0.007)	(0.012)	(0.010)	(0.012)
Cost × LHP	-1.992	-1.801	-1.957	-1.955
	(0.008)	(0.018)	(0.015)	(0.013)
School Fixed Effect	Y	Y	Y	Y

Notes: The first panel is the estimated results based on the admitted student qualities from previous years as the school reputation measure. The second panel is the estimated results for different assumptions about students' behaviors in ROLs. Col 1 represents the adaptive expectation assumption. Col 2-4 represents the s.d. of the uncertainty of exam score are 13.3, 26.6, and 33.35 respectively. Standard errors are reported in parentheses. Distance is measured by kilometer. The coefficient of female's attitude to home-school distance is normalized to -1. Cost(tuition) is measured by 1000 yuan. "HS", "MS", and "LS" represent high-, medium- and low-scoring students respectively. "HHP", "MHP", and "LHP" represent students from high-, moderate- and low-housing price communities respectively.

and another with the ZX quota representing 30% of the total quota when the focal mechanism allows the option to purchase seats.[64]

For both the DA and COSM mechanisms, we assume that students' ROLs reflect their genuine preferences. Under the CPPS mechanism, we create ROLs that reflect each student's best response in equilibrium. Specifically, we begin with the "telling the truth" strategy as the initial point and iteratively calculate each student's best response while keeping all other students' strategies fixed. If a student has an incentive to adopt a new strategy, we replace their old strategy with this new one and recalculate the matching outcome, setting it as the new starting point. We repeat this iteration until no student deviates to a new strategy (see Appendix E for details). We then carry out 5,000 simulations, with each student subjected to a unique vector of random utility shocks.

We consider two comparisons in this section. First, we use the matching outcome under the DA mechanism as our benchmark. Then we replace the DA mechanism with the COSM, which serves as an alternative strategy-proof and stable mechanism. Without considering students' strategies, this comparison evaluates the seat-selling policy itself, and using different ZX quotas further enables the study of welfare under various policies. To evaluate the mechanisms actually adopted, we also analyze the welfare changes when the CPPS mechanism replaces the DA mechanism. This replacement may reflect how students' strategic behaviors may change under different mechanisms and the welfare consequences. These two comparisons provide a comprehensive analysis of the entire ZX policy. The transition from the DA mechanism to the COSM evaluates the effects of implementing a price menu in the seat-purchasing policy, while the transition from the DA mechanism to the CPPS mechanism assesses the influence of students' strategic behaviors under the CPPS.

[64] The local government required that no school could admit ZX students totaling more than 20% of its capacity.

Students' Welfare

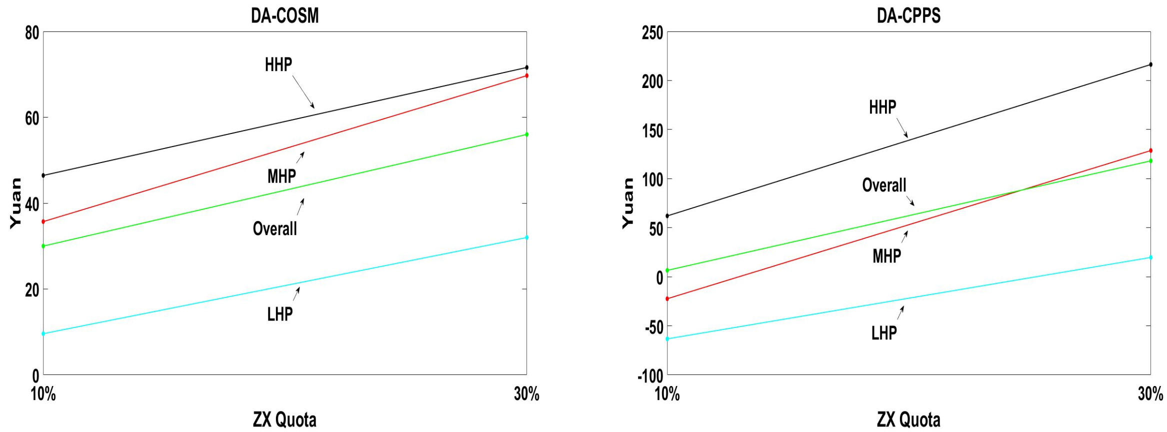
For each tested mechanism, we employ the welfare-equalizing tuition adjustment Δyuan, as proposed by Calsamiglia, Fu, and Güell (2020), to quantify the welfare change. This adjustment represents the tuition amount that a student would need to pay (or be credited) under the DA mechanism to attain the same utility level as under the replacement mechanism being evaluated.[65]

First, we examine the comparison between the DA mechanism and the COSM, which directly evaluates the ZX policy without considering students' strategies. As shown in Figure 3a, when the DA mechanism is replaced by the COSM, the average welfare of students decreases as the ZX quota increases. On average, students under the DA mechanism need to pay an additional 30 yuan to reach the same utility level as under the COSM when the ZX quota is 10% of the total quota. This loss increases to 57 yuan when the ZX quota rises to 30%. Students from different communities exhibit a similar trend of welfare loss as the ZX quota increases, with students from the LHP communities experiencing relatively less loss compared to students from other communities.

To further analyze the impact of the ZX policy on different students, we examine the school assignments when the DA mechanism is replaced by the COSM. The first panel of Table 8 shows that with a 10% ZX quota, approximately 1.8% to 2.2% of students from the HHP and MHP communities choose to pay higher tuition to secure their seats in the same schools, while only 0.7% of students from the LHP communities make the same choice. Under the same ZX quota, around 4% of students in each community are unable to attend their more preferred schools due to the increased competition for seats. On the other hand, 4.8% of students from the HHP communities are able to secure spots in their more preferred schools, a higher percentage compared to students from other communities (4% in the MHP

[65] All other parameters (except for tuition) remain fixed. Formally, let $u_{ij} = U(c_{ij})$ be i's utility derived from admittance to school j when paying tuition c_{ij} under the DA mechanism. If that mechanism is replaced by the focal new mechanism—in which case student i is assigned to school j' and achieves utility $u_{ij'}$—then the welfare-equalizing tuition adjustment (Δyuan) is the solution to $U(c_{ij} + \Delta\text{yuan}) = u_{ij'}$.

Figure 3: Welfare Change



(a) DA-COSM (b) DA-CPPS

Notes: These figures represents the welfare change when DA is replaced by another mechanism measured by the welfare-equalizing tuition adjustment. The y-axis represents the change of Yuan, and the x-axis represents the ZX quota. To keep the welfare level under the DA mechanism, a positive position indicates a student needs to pay additional tuition (loss), a negative position indicates a student receives a tuition deduction (gain). "HHP", "MHP" and "LHP" represent students from high-, moderate- and low-housing price communities.

communities and 3.5% in the LHP communities). When the ZX quota is increased to 30%, an additional 2% to 3% of students from each community choose to pay higher tuition to stay in the same schools, and a similar increase is observed for students who are unable to attend their more preferred schools and are instead assigned to less preferred ones. However, 9% of students from the HHP communities may get into their more preferred schools by paying higher tuition, and only 6% of students from the LHP and MHP communities can take advantage of the ZX policy in the same way.

In each type of community, the impact of the ZX policy varies for students with different scores. When the ZX quota increases from 10% to 30%, high-scoring students from the LHP area are the most affected. Around 14% of these students choose to pay higher tuition to secure their seats at the same schools, while 21% of them are unable to attend their more preferred schools and are priced out. Another interesting finding is that the ZX policy has a significant influence on medium-scoring students across all communities. Compared to high- and low-scoring students, a larger proportion of medium-scoring students choose to pay higher tuition to attend their preferred schools under the ZX policy. However, at the same time, many medium-scoring students are also affected negatively and are priced out, resulting in them being assigned to their less preferred schools.

Table 8: Changes of Matching Assignments under the Purchasing Seats Option(%)

	ZX quota 10%					ZX quota 30%				
	Same	Better		Worse		Same	Better		Worse	
	ZX	Normal	ZX	Normal	ZX	ZX	Normal	ZX	Normal	ZX
DA-COSM										
HHP	2.2	0.0	4.8	3.9	0.0	4.3	0.0	9.0	5.9	0.0
HS	2.2	0.0	2.6	3.2	0.0	6.6	0.0	7.6	7.2	0.0
MS	3.7	0.0	7.7	6.7	0.0	4.8	0.0	12.1	8.1	0.0
LS	0.4	0.0	4.2	0.9	0.0	0.3	0.1	7.0	0.8	0.0
MHP	1.8	0.0	4.0	4.1	0.0	3.3	0.0	6.6	7.1	0.0
HS	2.1	0.0	3.0	2.4	0.0	6.0	0.0	6.5	9.6	0.0
MS	3.4	0.0	6.2	8.9	0.0	4.6	0.0	10.0	11.6	0.0
LS	0.1	0.0	2.8	1.1	0.0	0.1	0.1	3.6	0.9	0.0
LHP	0.7	0.0	3.5	4.6	0.0	2.4	0.0	5.9	7.4	0.0
HS	2.0	0.0	0.1	6.7	0.0	14.3	0.0	0.4	21.6	0.0
MS	1.3	0.0	9.4	10.3	0.0	2.2	0.0	15.7	13.3	0.0
LS	0.0	0.0	0.4	0.4	0.0	0.0	0.0	0.6	0.4	0.0
DA-CPPS										
HHP	0.6	3.4	5.7	10.5	0.0	6.2	2.0	10.3	10.8	0.0
HS	1.0	0.0	3.4	9.7	0.0	8.5	0.0	9.5	10.4	0.1
MS	0.5	6.5	9.7	16.8	0.0	8.0	3.6	14.1	18.6	0.0
LS	0.0	4.2	3.8	2.8	0.0	0.1	3.0	6.2	0.6	0.0
MHP	0.5	4.3	4.1	8.0	0.0	3.8	2.9	7.0	11.1	0.0
HS	1.2	0.0	3.4	7.7	0.0	7.4	0.0	7.4	13.7	0.1
MS	0.3	8.1	7.8	15.0	0.0	4.8	5.2	11.5	20.2	0.0
LS	0.0	4.4	1.4	1.9	0.0	0.0	2.9	2.6	0.6	0.0
LHP	0.0	4.0	3.0	7.0	0.0	2.2	2.8	4.6	10.3	0.0
HS	0.1	0.0	0.0	9.4	0.0	16.8	0.0	1.1	22.3	0.1
MS	0.0	8.5	7.9	15.7	0.0	1.0	5.9	11.7	21.5	0.0
LS	0.0	1.9	0.3	0.8	0.0	0.0	1.4	0.5	0.3	0.0

Notes: This table indicates the percentage change in the number of students whose assignments are different under the purchasing seats option, when the DA mechanism is replaced by the COSM and CPPS mechanisms. When DA is replaced by another mechanism, "Same" means the student is assigned to the same school, "Better" represents the student is assigned to a more preferred school, and "Worse" represents the student is assigned to a less preferred school. "ZX" and "Normal"represents the student pays the basic and higher tuition respectively. "HHP", "MHP", and "LHP" represent students from high-, moderate- and low-housing price communities respectively. "HS", "MS", and "LS" represent high-, medium- and low-scoring students respectively.

When the DA is replaced by the COSM, high-scoring students experience a significant welfare loss under the COSM, primarily because most of them were already assigned to their most preferred schools under the DA mechanism. Under the COSM, these students must either pay higher tuition to secure their seats in the same schools or face being priced out and assigned to less preferred schools. In contrast, medium-scoring students are influenced in different ways. They have a higher probability of attending their preferred schools by paying higher tuition, but they are also susceptible to being priced out and assigned to their less preferred schools.

The estimated coefficients suggest that the willingness to pay higher tuition increases with both students' scores and housing prices. As a result, when the ZX quota is increased, the high-scoring students from the LHP have a stronger incentive to secure their seats compared to students from the same communities, hence a large proportion of them choose to pay higher tuition. However, they are also more likely to be priced out since their incentive to pay higher tuition is lower than that of high-scoring students from affluent communities. Consequently, a larger proportion of high-scoring students from LHP communities are priced out. The vacant seats left by these high-scoring students are mostly occupied by medium-scoring students from various communities. At the same time, a comparable number of medium-scoring students are also priced out due to the increased ZX quota. Therefore, the influence of the ZX policy on medium-scoring students is substantial in both directions.

Table 9 identifies the percentage of "winners" (students whose welfare increases) and "losers" (students whose welfare decreases) when the DA mechanism is replaced. Regardless of the ZX quota, the proportion of winners under the COSM never exceeds 8.7% for any student group. However, the proportion of losers exceeds that of winners in all cases.

When the ZX quota is increased from 10% to 30%, the proportions of losers in the HHP and MHP communities experience a substantial rise, while the change is relatively small for the LHP communities. The analysis of the welfare change in each type of community (Table F.2) further confirms the explanation of the effect of the ZX policy. More than 35%

of high-scoring students from poor communities become losers and experience an average welfare loss of 1587 yuan when the ZX quota increases. This welfare loss primarily arises from a large proportion of students who must either pay higher tuition to secure their seats or be priced out. The average welfare gain for medium-scoring students is slightly higher than the losses, except for one community.

In summary, when the DA mechanism is replaced by the COSM, the average impact on students from different residential areas is similar. The number of students who experience a welfare loss due to the ZX policy is higher than the number of students who potentially benefit from it, and this loss is magnified as the ZX quota increases. However, students' reactions to the policy vary. Medium-scoring students are the most affected group by the ZX policy. They are more likely than low-scoring students to attend their preferred schools by paying higher tuition, but they are also more susceptible to being priced out and assigned to less preferred schools. Top-performing students from economically disadvantaged communities bear the greatest burden under the ZX policy. A significant proportion of them either have to pay higher tuition to secure seats in their preferred schools or are priced out altogether. Comparatively, students from affluent communities are more likely to stay at their desired schools by paying higher tuition, while students from other communities are more likely to be priced out and assigned to less preferred schools. These findings indicate that while the ZX policy intensifies educational inequality among students, its overall impact on their welfare, when assessed in monetary terms, isn't as pronounced.

Next, we investigate how the practical implementation of the mechanism may impact students' welfare and their strategic behaviors. When the DA mechanism is replaced by the CPPS mechanism, the changes in student welfare exhibit a similar pattern as observed in the COSM case, but with a notable difference (Figure 3b). Students from the MHP and LHP communities experience a welfare gain of 22 yuan and 63 yuan, respectively, when the ZX quota is 10%.[66]

[66] Abdulkadiroğlu, Che, and Yasuda (2011, 2015) suggest that from an ex-ante perspective, when schools have coarse preferences for students coupled with a symmetric tie-breaking rule, students could fare better

However, as the ZX quota increases to 30%, all student groups face welfare losses, particularly students from the HHP and MHP communities. Students from the HHP communities endure a welfare loss equivalent to a 216 yuan increase in tuition, while students from the MHP communities experience a loss of 128 yuan.

The second panel of Table 8 explains the reasons for the improved student welfare under the CPPS mechanism compared to the COSM at lower ZX quotas. Unlike under the COSM, a significantly lower proportion of students choose to save their seats in the same schools by paying higher tuition under the CPPS mechanism, regardless of student groups. Meanwhile, more students are priced out to their less preferred schools under the CPPS mechanism compared to the COSM. Additionally, a positive number of students from every type of community are able to secure spots in their more preferred schools without paying extra tuition, which is not the case under the COSM. However, when the ZX quota is increased to 30%, a larger number of students, particularly those from the HHP and MHP communities, choose to pay higher tuition to stay in the same schools. Simultaneously, more students across all communities are priced out compared to the COSM. Consequently, all students experience greater welfare losses when the ZX quota is high.

Table 9 further confirms that the number of winners and losers increases in all student groups. These findings suggest that when the CPPS mechanism is used to replace the DA mechanism, students may have more opportunities to strategically manipulate their preferences, leading to greater variations in student welfare across different communities.

To further investigate students' strategic behaviors under the CPPS mechanism, we also simulate the students' strategies under the Chinese parallel mechanism as an intermediate step. For high-scoring students from HHP communities, their first choice under the Chinese parallel mechanism is, on average, their 1.12 choice in their true preference (Table F.3). However, under the CPPS mechanism, their first choice moves slightly closer to their true

under the Boston mechanism than under the DA mechanism, as assessed by their cardinal preferences. In contrast, our results show that when schools have strict priorities for students, a manipulable mechanism such as the CPPS can still yield higher average student welfare for some types of students than the DA mechanism from an ex-post perspective.

first choice at 1.01. This indicates that more high-scoring students are inclined to choose their true first choice under the CPPS mechanism. Additionally, almost 50% of these students choose the ZX options for their first choice. The average second choice for this group is similar to their 3rd choice in their true preference under both the Chinese parallel and CPPS mechanisms. However, under the CPPS mechanism, 80% of students opt for the ZX options for their second choices.

Students' first choice under the CPPS mechanism shows a slight upward shift for high- and medium-scoring students in all communities, but a slight downward shift for low-scoring students. High-scoring students exhibit a higher likelihood of choosing the ZX options, particularly for their second choices, and this pattern decreases with housing price. Meanwhile, low-scoring students are less inclined to choose the ZX options, especially for their second choices. However, when the ZX quota rises from 10% to 30%, the change in strategic behaviors isn't substantial. This phenomenon indicates that students' strategic behavior under the CPPS mechanism improves their welfare under a low ZX quota. However, the same strategies lead to a substantial welfare loss when the ZX quota is increased.

Table 9: Winners and Losers

	DA-COSM				DA-CPPS			
	10%		30%		10%		30%	
	W	L	W	L	W	L	W	L
HHP %	4.5	8.3	8.7	14.3	8.4	14.9	8.4	27.6
MHP %	4.2	7.2	7.3	12.9	8.7	10.7	8.3	20.8
LHP %	3.9	4.0	5.8	6.6	7.2	5.8	6.2	9.4
Total %	3.9	6.3	6.9	10.9	7.8	10	7.3	18.6
HHP ¥	895	-1040	1043	-1147	1297	-1148	1153	-1138
MHP ¥	1064	-1106	1170	-1203	1459	-980	1307	-1142
LHP ¥	734	-943	766	-1182	1477	-754	1253	-1026
Total ¥	924	-1052	1036	-1076	1399	-1021	1237	-1124

Notes: The first panel of this table indicates the percentage change in the number of students whose utilities increase (winners) or decrease (losers) when the DA mechanism is replaced by the COSM and CPPS mechanisms. The second panel indicates the welfare change measured by yuan. "W" represents winners, and "L" represents losers. For each mechanism change, utility changes are measured in three scenarios in which the ZX quotas are 10% and 30% of the total quotas. "HHP" represents students from high housing price communities, 'MHP" represents students from moderate housing price communities, and "LHP" students from low housing price communities.

6.1 Impact on Schools

In this final section, we examine the impact of the ZX policy on schools, considering two factors: the quality of admitted students and the tuition collected by schools. Schools face a trade-off in implementing this policy. On one hand, allowing students to buy seats may increase the schools' income. On the other hand, seat purchasing can lead to the dispersion of high-quality students across different schools. Under the ZX policy, some high-scoring students who might have attended upper-tier schools under the DA mechanism may be priced out and end up in middle-tier schools if they choose not to pay the higher tuition. Conversely, some low-scoring students who are willing to pay more tuition for their preferred schools may displace high-scoring students and secure seats in those schools. As a result, upper-tier schools may collect more tuition but experience a decline in the overall quality of admitted students. While middle-tier schools may admit more high-quality students.

For *upper-tier* school #183 (F.4 in Appendix F), the collected fees increase proportionally with the ZX quota when the DA mechanism is replaced by the COSM. When the CPPS is adopted, this school may gain even more in terms of tuition collection, with the gain exceeding 40% when the ZX quota is 30%. Importantly, when the DA mechanism is replaced by either the COSM or the CPPS, this school experiences only a negligible decline in student quality. Considering the findings for other upper-tier schools (see Table F.4 in Appendix F), it becomes evident that there is a significant demand for elite schools. This allows them to profit substantially from selling seats without compromising the quality of admitted students.

For *middle-tier* school #179, the seat-purchasing option has the potential to generate profits. The impact on student quality can vary depending on the mechanism adopted. When the ZX quota increases to 30% under the COSM, student quality slightly decreases compared to its level under the DA mechanism. On the other hand, if the DA mechanism is replaced by the CPPS mechanism, there is a small increase in student quality. Consequently, these schools may experience significant variations in the quality of their admitted students, with some experiencing positive changes and others negative changes.

7 Conclusion

Our paper examines a contentious but previously overlooked Chinese school choice policy, *Ze Xiao*. This policy allows students to "purchase" seats at their desired schools by paying higher tuition. We find that the corresponding matching mechanisms employed in this policy are not strategy-proof and may lead to unstable outcomes. We combine data from high school admission records with survey responses from students in China to estimate their preferences for schools and tuition. Our findings reveal that high-scoring students are more willing than other students to incur additional costs, such as higher tuition fees, to attend their preferred schools. Furthermore, students from communities with high housing prices are less motivated to bear the financial burden of higher tuition compared to students from communities with low housing prices.

Using estimated preferences, we conduct counterfactual experiments to evaluate the welfare consequences of the ZX policy. We find that, when the strategy-proof COSM replaces the DA mechanism, students' welfare decreases across all student groups. However, when the DA mechanism is substituted by a non-strategy-proof mechanism like CPPS, it may alleviate the welfare losses, particularly when the ZX quota is low. This is because more students can exploit the system to secure admission to their preferred schools.

When experiencing a welfare loss, students from high housing price communities tend to opt for paying higher tuition to retain their seats at the same schools. Students from communities with low housing prices are more inclined to be priced out and settle for less preferred schools. As the ZX quota increases, high-scoring students from economically disadvantaged communities demonstrate a greater motivation to pay higher tuition in order to remain in higher-ranked schools compared to medium- and low-scoring students.

From the school's point of view, the seat-purchasing option proves beneficial for upper-tier schools as it enables them to collect a substantial amount of additional tuition while experiencing only a minor decline in the quality of admitted students compared to the DA mechanism. However, for other schools, the seat-purchasing option introduces greater uncertainty regarding both the amount of collected tuition and the quality of students admitted.

References

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag A Pathak (2017). "The welfare effects of coordinated assignment: Evidence from the New York City high school match". *American Economic Review* 107.12, pp. 3635–89.

Abdulkadiroğlu, Atila, Yeon-Koo Che, and Yosuke Yasuda (2011). "Resolving conflicting preferences in school choice: The "Boston mechanism" reconsidered". *American Economic Review* 101.1, pp. 399–410.

— (2015). "Expanding "choice" in school choice". *American Economic Journal: Microeconomics* 7.1, pp. 1–42.

Abdulkadiroğlu, Atila, Parag A Pathak, and Alvin E Roth (2005). "The new york city high school match". *American Economic Review*, pp. 364–367.

— (2009). "Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match". *American Economic Review* 99.5, pp. 1954–1978.

Abdulkadiroglu, Atila and Tayfun Sönmez (2003). "School choice: A mechanism design approach". *American Economic Review* 93.3, pp. 729–747.

Abdulkadiroğlu, Atila et al. (2011). "Accountability and flexibility in public schools: Evidence from Boston'Leveling the Playing charters and pilots". *The Quarterly Journal of Economics* 126.2, pp. 699–748.

Abdulkadiroğlu, Atila et al. (2020). "Do parents value school effectiveness?" *American Economic Review* 110.5, pp. 1502–1539.

Abdulkadiroğlu, Atila et al. (2020). "Efficiency, justified envy, and incentives in priority-based matching". *American Economic Review: Insights* 2.4, pp. 425–442.

Abraham, Jesse M and Patric H Hendershott (1996). "Bubbles in Metropolitan Housing Markets". *Journal of Housing Research* 7.2, p. 191.

Agarwal, Nikhil (2015). "An empirical model of the medical match". *American Economic Review* 105.7, pp. 1939–1978.

— (2017). "Policy analysis in matching markets". *American Economic Review* 107.5, pp. 246–50.

Agarwal, Nikhil and Eric Budish (2021). "Market design". *Handbook of Industrial Organization*. Vol. 5. 1. Elsevier, pp. 1–79.

Agarwal, Nikhil and Paulo Somaini (2018). "Demand Analysis using Strategic Reports: An application to a school choice mechanism". *Econometrica* 86.2, pp. 391–444.

Ajayi, Kehinde (2022). "School choice and educational mobility: Lessons from secondary school applications in Ghana". *Journal of Human Resources*.

Akyol, Pelin and Kala Krishna (2017). "Preferences, selection, and value added: A structural approach". *European Economic Review* 91, pp. 89–117.

Artemov, Georgy, Yeon-Koo Che, and Yinghua He (2017). *Strategic 'Mistakes': Implications for Market Design Research*. Tech. rep. mimeo.

Azevedo, Eduardo M and John William Hatfield (2018). "Existence of equilibrium in large matching markets with complementarities". *Available at SSRN 3268884*.

Beggs, Steven, Scott Cardell, and Jerry Hausman (1981). "Assessing the potential demand for electric cars". *Journal of Econometrics* 17.1, pp. 1–19.

Biro, Péter et al. (2022). "The large core of college admission markets: Theory and evidence". *Proceedings of the 23rd ACM Conference on Economics and Computation*, pp. 958–959.

Bobba, Matteo et al. (2021). *Teacher compensation and structural inequality: Evidence from centralized teacher school choice in Perú*. Tech. rep. National Bureau of Economic Research.

Budish, Eric and Estelle Cantillon (2012). "The multi-unit assignment problem: Theory and evidence from course allocation at Harvard". *American Economic Review* 102.5, pp. 2237–71.

Burgess, Simon et al. (2014). "What parents want: school preferences and school choice". *The Economic Journal*.

Calsamiglia, Caterina, Chao Fu, and Maia Güell (2020). "Structural estimation of a model of school choices: The Boston mechanism vs. its alternatives". *Journal of Political Economy*, pp. 642–680.

Calsamiglia, Caterina, Guillaume Haeringer, and Flip Klijn (2010). "Constrained school choice: An experimental study". *American Economic Review* 100.4, pp. 1860–1874.

Capozza, Dennis R et al. (2002). *Determinants of real house price dynamics*. Tech. rep. National Bureau of Economic Research.

Chen, Yan and Onur Kesten (2017). "Chinese college admissions and school choice reforms: A theoretical analysis". *Journal of Political Economy* 125.1, pp. 000–000.

Chen, Yan and Tayfun Sönmez (2006). "School choice: an experimental study". *Journal of Economic Theory* 127.1, pp. 202–231.

City Bureau of Education (2014). *High school admission record and middle school graduates survey in 2014 [dataset]*. Unpublished data. Accessed Sep. 20, 2014.

De Haan, Monique et al. (2023). "The performance of school assignment mechanisms in practice". *Journal of Political Economy* 131.2, pp. 388–455.

Fack, Gabrielle, Julien Grenet, and Yinghua He (2019). "Beyond truth-telling: Preference estimation with centralized school choice and college admissions". *American Economic Review* 109.4, pp. 1486–1529.

Gale, David and Lloyd S Shapley (1962). "College admissions and the stability of marriage". *The American Mathematical Monthly* 69.1, pp. 9–15.

Goodman, Allen C (1988). "An econometric model of housing price, permanent income, tenure choice, and housing demand". *Journal of Urban Economics* 23.3, pp. 327–353.

Greenberg, Kyle, Parag A Pathak, and Tayfun Sonmez (2021). "Mechanism design meets priority design: Redesigning the US Army's Branching Process". *arXiv preprint arXiv:2106.06582*.

Haeringer, Guillaume and Flip Klijn (2009). "Constrained school choice". *Journal of Economic Theory* 144.5, pp. 1921–1947.

Hassidim, Avinatan, Assaf Romm, and Ran I Shorrer (2016). "Strategic behavior in a strategy-proof environment". *Proceedings of the 2016 ACM Conference on Economics and Computation*. ACM, pp. 763–764.

— (2017). "Redesigning the Israeli Psychology Master's Match". *American Economic Review: Papers & Proceedings* 107.5, pp. 205–209.

Hatfield, John William and Fuhito Kojima (2008). "Matching with Contracts: Comment". *American Economic Review* 98.3, pp. 1189–1194.

— (2010). "Substitutes and stability for matching with contracts". *Journal of Economic Theory* 145.5, pp. 1704–1723.

Hatfield, John William, Scott Duke Kominers, and Alexander Westkamp (2017). "Stability, strategy-proofness, and cumulative offer mechanisms". *Proceedings of the 2017 ACM Conference on Economics and Computation*. ACM, pp. 673–674.

Hatfield, John William and Paul R Milgrom (2005). “Matching with contracts”. *American Economic Review* 95.4, pp. 913–935.

He, Yinghua (2016). “Gaming the Boston school choice mechanism in Beijing”. *Manuscript, Toulouse School of Economics*.

Hwang, Min and John M Quigley (2006). “Economic fundamentals in local housing markets: evidence from US metropolitan regions”. *Journal of Regional Science* 46.3, pp. 425–453.

Hwang, Sam Il Myoung (2015). “A Robust Redesign of High School Match”. *University of Chicago, Working Paper*.

Kapor, Adam, Christopher Neilson, and Seth Zimmerman (2020). “Heterogeneous beliefs and school choice mechanisms”. *American Economic Review* 110.5, pp. 1274–1315.

Kelso Jr, Alexander S and Vincent P Crawford (1982). “Job Matching, Coalition Formation, and Gross Substitutes”. *Econometrica* 50.6, pp. 1483–1504.

Kojima, Fuhito (2017). “Recent developments in matching theory and their practical applications”. *Advances in Economics and Econometrics* 1, p. 138.

Lai, Fang, Elisabeth Sadoulet, and Alain De Janvry (2011). “The contributions of school quality and teacher qualifications to student performance: Evidence from a natural experiment in Beijing middle schools”. *Journal of Human Resources* 46.1, pp. 123–153.

Lai, Fang, Elisabeth Sadoulet, and Alain de Janvry (2009). “The adverse effects of parents’ school selection errors on academic achievement: Evidence from the Beijing open enrollment program”. *Economics of Education Review* 28.4, pp. 485–496.

Matzkin, Rosa L (1993). “Nonparametric identification and estimation of polychotomous choice models”. *Journal of Econometrics* 58.1-2, pp. 137–168.

Pathak, Parag (2011). “The Mechanism Design Approach to Student Assignment”. *Annual Review of Economics* 3.1, pp. 513–536.

Pathak, Parag A and Peng Shi (2021). “How well do structural demand models work? Counterfactual predictions in school choice”. *Journal of Econometrics* 222.1, pp. 161–195.

Pathak, Parag A and Tayfun Sönmez (2008). "Leveling the playing field: Sincere and sophisticated players in the Boston mechanism". *American Economic Review* 98.4, pp. 1636–1652.

Rees-Jones, Alex (2018). "Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match". *Games and Economic Behavior* 108, pp. 317–330.

Shen, Junqiang and Zun-min Wu (2006). "School Choice and Education Equity——Changes in School Choice Policies and New Direction in Public School Reform in China". *Tsinghua Journal of Education* 6.

Sönmez, Tayfun (2013). "Bidding for army career specialties: Improving the ROTC branching mechanism". *Journal of Political Economy* 121.1, pp. 186–219.

Sönmez, Tayfun and Tobias B Switzer (2013). "Matching With (Branch-of-Choice) Contracts at the United States Military Academy". *Econometrica* 81.2, pp. 451–488.

Train, Kenneth E (2009). *Discrete choice methods with simulation*. Cambridge university press.

Wang, Tong and Congyi Zhou (2020). "High school admission reform in China: a welfare analysis". *Review of Economic Design* 24.3-4, pp. 215–269.

— (2024). *Replication package for: "Purchasing Seats in School Choice and Inequality." Quantitative Economics, data deposited at Zenodo*. Accessed July, 2024. URL: `https://doi.org/10.5281/zenodo.12735964`.

Zhang, Chuanchuan, Shen Jia, and Rudai Yang (2016). "Housing affordability and housing vacancy in China: The role of income inequality". *Journal of Housing Economics* 33, pp. 4–14.