

Endogenous information acquisition in matching markets: China's college admission mechanisms

Yao Luo

May 3, 2024

1 Motivation

In the matching markets it has long been assumed that agents have full information. Some recent work show that people do not know their full preferences and it is costly to acquire information (Corcoran et al. 2018, Dynarski et al. 2020, Grenet et al. 2019). Therefore, market designers should pay attention to the acquisition and flow of information besides stability.

To match the stable outcomes under full information, Immorlica et al (2020) defines regret-free stable outcomes in the context of college admission and school choice. In short, it requires students to conduct optimal search while generating stable matching outcomes. Follow the work by Azevedo and Loshno (2016), Immorlica et al (2020) also shows that finding regret-free stable outcomes is equivalent to finding market-clearing cutoffs. However, students need the information of market-clearing cutoffs to form their budget set and conduct information acquisition within the set, which will determine their preferences and hence affect the

market-clearing cutoffs. Because of this information deadlocks, it is impossible to achieve regret-free matching outcomes in real life mechanism implementations.

According to Immorlica et al (2020), we can achieve approximately regret-free stable outcomes by providing external historical data with respect to perturbed capacities. However, China's two college admission mechanisms only provide historical cutoffs without the promise to perturb capacities. So we know certainly some students have regret because of the costly information acquisition. Every year over 10 million high school graduates join in the National College Entrance Examination (gaokao in mandarin) competing for about 7 million seats. Gaokao is currently the largest centralized matching market in the world. Achieving stable and efficient matching outcomes is extremely important in this market since it affects the welfare of so many students.

Most provinces in China use the Parallel admission mechanism (PA). It is a variant of direct serial dictatorship (DirSD) with the restriction on the length of the submitted rank-ordered list (ROL). The important feature is that students submit their ROL simultaneously.

Inner Mongolia uses a dynamic admission mechanism (IM), ? has a detailed description. It can be seen as sequential serial dictatorship (SeqSD) with time constraints and the restriction that groups of students move sequentially instead of individuals.

Both PA and IM provides historical cutoff scores, but IM offers additional information regarding matching outcomes. Normally we think more information helps, so my first guess is that IM achieves higher student welfare than PA. However, in 2025 Inner Mongolia will give up IM and use PA. This seems counterintuitive to me, especially the government has invested a lot of money on IM mechanism and

has used it since 2007.

This puzzling fact leads to my research question: Does PA indeed achieves higher student welfare than IM? If so why does the additional information provided by IM harm student welfare instead? Who are the winners and who are the losers? Maybe the time constraints in IM (one hour per group) make the price discovery process too costly? Maybe the information communication in IM is not effective? Maybe the additional information is too noisy and detriment students' welfare instead? Or maybe IM is indeed better than PA and the policy change is purely of political intention. Moreover, are there better ways to communicate information to students to improve matching outcomes of IM?

2 Literature Review

? question the common assumption in the matching literature that agents have full information on their own preferences and conclude that students' costly discovery of preferences exist. Similarly, ? don't believe that agents precisely know their preferences and introduce the assumption of endogenous information acquisition into otherwise standard house allocation problems.

Theoretically, ? examines the cost of information acquisition in centralized matching markets and how to reduce wasteful information acquisition. Similarly, ? look at information acquisition in matching markets but with a focus on the role of price discovery.

? and ? compares both theoretically and experimentally the matching outcomes of PA and DA under the assumption of complete information and find that DA is better than PA in terms of stability.

Gong and Liang (2023) shows experimentally that IM mechanism achieves similar stability as DA mechanism and similar efficiency as Boston mechanism when the correlation of preferences is low. They assume incomplete information.

? focus on the impact of time-constraint imposed by the authority, and find the matching outcomes of time-constrained dynamic mechanism (TCDM) can be worse than the student-proposing deferred acceptance (DA) mechanism.

? look at the university admissions in Brazil and show that dynamic mechanism can outperform the standard ones since matchings are implemented sequentially and students can revise their choices after receiving information or feedback about their current allocations. (Hakimov et al 2023) shows both theoretically and experimentally that a sequential serial dictatorship mechanism leads to higher student welfare than a direct serial dictatorship mechanism.

3 Contribution

This paper will provide empirical and experimental comparison of real-life implementation of DirSD (PA) and SeqSD (IM).

This paper sheds light on the importance of information flows in market design and validate costly endogenous information acquisition.

4 Data

Students' exam scores, rank, ethnicity, gender and address. University's past cutoff scores and address.

5 Empirical Strategy

$$y_i = \alpha_0 + \alpha_1 X_i + \beta Y_{2025} + \varepsilon_i$$

y_i : college prestige index (determined by cutoff scores of year 2022 and 2023).

X_i : students' gender, ethnicity, rank by exam scores being normalized to be within (0,1).

Implicitly assumes higher-ranked students prefer more prestigious colleges. Only care about big names without considering majors.

I assume that colleges with higher cutoffs are more prestigious, and hence the mechanism that maps higher-scored students to more prestigious colleges achieves better matching outcomes. To control for the impact of geography on students' preferences (some students may prefer to be closer to home while some may prefer colleges that are out of their home province), I include a binary variable which equals one if the student is assigned to a college in his/her home province and zero otherwise. For robustness checks, I try different measures of college prestige, for example the QS ranking.

6 Experiment design

General setup:

Exam scores are randomly and independently drawn from the IM empirical distribution between 0 and 100. 30 students competing for 15 seats in 10 colleges. Admission rate is 50%. Students are told that any ranking of the university is equally possible. Preferences are randomly drawn from the space of rankings.

Students only know their own exam scores and ranks. University's quotas and historical cutoffs are common knowledge. Historical cutoffs are gotten by running DirSD. Students need to pay search costs to know their own preferences. Preferences are private knowledge. Students receive more rewards for being assigned to more preferred university.

Environments:

Dimension 1: The degree of correlation of preferences among students. One tier, two tiers and three tiers. Dimension 2: The cost of information acquisition. Low and high costs.

Predictions:

Hypothesis 1: Lower-ranked students gain more from IM compared to PA. Hypothesis 2: IM has a higher probability of being unmatched. Hypothesis 3: Students are more likely to oversearch in IM. Hypothesis 4: Welfare: on average $IM < PA$, but not hold for all students. Hypothesis 5: Smaller group size produces better matching outcomes. Hypothesis 6: Increasing the ROL in PA improves the matching outcomes.

7 Caveats

In this research proposal I ignored the fact that students are applying for a specific program in the college, so implicitly I assumed that the preferences are only over college choices and students do not care about their major. It's a sensible assumption to make in the past when Chinese care a lot about the reputation about the college and do not care much about majors. However, over the years people became more realistic and care more about majors. The dispersion of ad-

mitted students' exam scores is increasing. Therefore, the objective of matching mechanism might be altered to include the preference over majors.

Moreover, the universities determine the cutoff scores by their total quota instead of quota of each major. In many cases some major might have too many applicants while some have too few. The university would reassign the lower-ranked students to the unfilled seats of another major even when the students do not want it. This sometimes result in terrible assignment result and the student has to go back to high school and retake the Gaokao. Matching students to their preferred majors might be a better objective to achieve in the Chinese Gaokao context. But the current literature has not looked in this direction because of the complication.

One method I propose to accommodate students' major preferences is to allow for full rank of all the majors the university offers instead of only rank 5 majors in each college choice. Pros and cons???