The impact of information acquisition in college admissions

Yao Luo

April 29, 2024

1 Motivation

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 adopted the parallel mechanism (PA), where stduents submit a rank-ordered list (ROL) of certain length to the clearinghouse and colleges admit students solely depending on their test scores until their seats are filled. Different provinces might have different requirement for the ROL length. For example, Hebei province allows for 5 colleges 5 majors under each college. Students have the option to agree or disagree with being allocated to another major not in their list within a college in their list.

The only exception from the PA mechanism is Inner Mongolia province, which

adopted a dynamic mechanism in 2006. I will refer to it as IM in this paper. A nice paper which reviews the history of college admission mechenisms in Inner Mongolia is **?**. Students report their first choice on the online system, and they know their ranking in that major as well as the capacity right after submission. Students have unlimited times of reporting, but with time constraints. For example in 2023, the application process begins at8am, and students with grades above 670 need to finish their application by 11am. Students whose grade fall in the range of 640-669 need to finish their application by 12pm etc. The process ends at 20pm.

The most salient distinction between the PA and IM mechanism is the amount of information students have regarding their chances of admission. Under the PA mechanism, students have access to historical cutoffs of each college. The current year's cutoffs are only achieved after the matching outcomes are realized. Students have to apply based on their expected chances of admission, which requires to form beliefs on cutoffs ex-ante. In contrast, under the IM mechanism students acquire information about cutoffs during the process of application by revising their choices. In this setting, the real-time information of cutoffs for each college plays the role of prices in a normal market, and in equilibrium the stable-matching outcomes shoule be achieved and the final cutoffs work as the market clearing prices.

? shows theoretically mechanisms that facilitates efficient price discovery or offer relevant information on cutoffs yield approximately regret-free stable outcomes. In this paper I want to design a model to examine the impact of acquiring information of cutoffs on college admission outcomes empirically, using data from the Inner Mongolia dynamic mechanism. The additional information of college admission chances in IM intuitively should help students to improve their admission

outcomes. However, since the information acquisition is costly and students have unlimited times of submission, the extra burden may harm students' welfare and students who fail to search efficiently may have unsatisfactory matching results. This tradeoff leads to ambiguity in welfare analysis and identifying winners and losers when switch from PA to IM mechanism.

2   Literature Review

There are theoretical and empirical papers on the parallel mechanism (PA) in China, for example **?** and **?**.

There have been a few empirical papers on the Inner Mongolia College admission mechanism. **?** focus on the impact of time-constraint imposed by the authority, and compare the matching outcomes of time-constrained dynamic mechanism (TCDM) with the stduent-proposing deferred acceptance (DA) mechanism. **?** compares the dynamic matching mechanism with the parallel mechanism using justified envy as a criteria, but their conclusions are counterintuitive.

? 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.

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.

? 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.

3   Data

Ideally, the dataset should include all student-major pairs and the whole time-path of college choice revisions.

However, the time series woule be enormous to store and the Chinese education authority does not record this data.

I have data on the full empirical distribution of the Gaokao exam grades, as well as the number of stduents each major in a specific college admits, and their maximum and minimum grades.

The Inner Mongolia education authority publish hourly the current cutoffs, the number of stduents applied to each college and their grades for all college majors. The hourly updated information contains rich information about stduents' application behaviours and preferences.

4   Empirical Strategy

Since the Inner Mongolia Ministry of Education does not record the historical choice submissions and only records the final choice, there is no data on students' rank-ordered list(ROL) of colleges. Moreover, there is no guarantee that students search their more preferred colleges first. Therefore, the order in the list wouldn't

reveal students' preferences. I plan to use reduced form analysis in this paper to answer my researcg question.

My key explanatory variable is students' exam scores. 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.

Another dimension for quantifying the quality of the matchings is the number of students who reject their offers and go back to high school for the next year's College Entrance Exam (Gaokao).

In 2025 Inner Mongolia will switch from IM to PA mechanism, which serves as a natrual experiment. Since students couldn't control which year they would participate in Gaokao, which is decided by their birth year. I can do a difference in difference analysis using the exogenous variation by the mechanism change in year 2025.

5 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 admitted students' exam scores is increasing. Therefore, the objective of matching mechamism 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 stduents 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 prefered 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???