Information Access and College Admission Outcomes*

Le Kang†
Li Chen ‡
Yang Song§

February 2025

Abstract

This paper provides quasi-experimental evidence on the effects of information access in centralized college admissions. The variation in access to information stems from a batched dynamic matching mechanism adopted in the Chinese province of Inner Mongolia. Students are grouped into batches based on test scores and face different deadlines to finalize their choice submission: students just above the cutoffs (JACs) finalize their choices along with their higher-scoring peers and thus only observe the tentative choices, whereas students just below the cutoffs (JBCs) can observe the finalized choices of higher-scoring peers. Using administrative data linked with university post-graduation outcomes, we find that JBCs are more likely to apply for colleges considered reach options and are admitted to significantly more selective colleges. These colleges are more likely to be out-of-province and with graduates having higher starting salaries and higher graduate school attendance rates. The effects are more pronounced near batch cutoffs where access to such information is more important.

JEL Codes: C82, C83, I23, I28.

Keywords: College admissions; information; dynamic mechanism

*We would like to thank Inácio Bó, Juan Pereyra and Yan Song for suggestions, and seminar and conference participants at Colgate University, Nanjing University, Nankai University, Peking University, Renmin University of China, Shandong University, 2024 CCER-NCER Conference on Chinese Economy, 2024 China Economics of Education Annual Conference, 2024 China Educational Finance Annual Conference, 2024 China Forum on Empirical Education Research, 2024 Chinese Economists Society Annual Conference, and 2024 Greater Bay Area Market Design Workshop for helpful feedback.

†School of Education, Nanjing University. lekang@nju.edu.cn.

‡Corresponding author. Shanghai International College of Intellectual Property, Tongji University and Department of Economics, University of Gothenburg. chen_li@tongji.edu.cn.

§Department of Economics, Colgate University. 13 Oak Dr, Hamilton, NY 13346. ysong@colgate.edu.

1 Introduction

Information is critical in shaping decisions and outcomes. It improves market performance and welfare (Jensen, 2007; Galdon-Sanchez et al., 2023), influences people's retirement planning and financial decision-making (Duflo and Saez, 2003; Lusardi and Mitchell, 2014), and helps students and parents make more informed school choice (Hastings and Weinstein, 2008). In the context of college admissions, deciding where to apply is a complex and high-stake decision with profound and lasting impacts on students' labor market outcomes globally (Hoekstra, 2009; Kirkeboen et al., 2016; Chetty et al., 2020; Jia and Li, 2021). Yet, students often navigate this process under significant uncertainty.

We focus on centralized matching markets in this paper, where students are systematically assigned to colleges through matching mechanisms based on their submitted preferences and admission priority. To support their decision-making, students are usually provided with relevant information, including college quotas and historical admission statistics, such as past admission cutoff scores. In recent years, dynamic mechanisms have emerged in college admissions across Brazil, China, France, Germany, and Tunisia, as well as in school choice programs in Chile and the US. These mechanisms generate additional information about the intermediate outcomes during the application process, allowing students to adjust their choices. Theoretically, under the assumption of *full information*, dynamic mechanisms can yield more desirable matchings by encouraging students to truthfully report their preferences (Echenique et al., 2016; Klijn et al., 2019; Bó and Hakimov, 2020) and improving coordination efficiency (Dur et al., 2021). Despite the efforts to provide information, students may still face residual uncertainty in practice, particularly regarding *preferences and choices of the others*.

The college admissions system in Inner Mongolia provides a unique setting to examine the impact of access to information on college decisions and admission outcomes. Introduced in 2007, the mechanism provides students with real-time updates on their application status—including whether they are tentatively accepted, their ranking, and the admission cutoff score at the college they applied to—and allows them to revise their choices during a specified period. Students obtain these updates by applying to a single college at a time. Final admissions are based on the last submitted choices, with priorities determined by test scores.

While real-time updates help students gauge their admission chances and steer away from overly competitive or non-feasible choices, uncertainty about others' choices remains. Because the system design allows students to revise their choices at any time, applications and updates occur *not* necessarily sequentially, limiting students' ability to fully observe the choices of higher-scoring peers. The effect on students is asymmetric, with greater risks faced by lower-scoring students—particularly in an environment like Inner Mongolia where students are ranked uniformly across colleges based on test scores. Our theoretical analysis shows that when such uncertainty is present, lower-priority students have stronger incentives to apply to colleges where they have higher probabilities to be admitted, even if those colleges are less preferred. By doing so, they reduce the risk of being displaced by last-minute changes from higher-scoring applicants.

A key feature of the mechanism is the *staggered deadlines*. Students are grouped into batches, with highest-scoring students in the first batch finalizing their choices first, followed by students in the next batch, and so on. Later batches can revise choices after observing the finalized choices of earlier batches. This staggered design creates a quasi-experimental setting that allows us to evaluate the impact of information using a Regression Discontinuity (RD) approach, comparing students just above and below the cutoffs. Students just above the cutoff (JACs) must finalize choices sooner, when still facing greater uncertainty about higher-priority applicants' choices. In contrast, students just below the cutoff (JBCs) have access to information on the choices of higher-priority applicants in the previous batches.

Our results are two-fold. First, we investigate application strategies using data on students' submitted choices from 2020 to 2022. *Safe* options are colleges with larger safety margins, where as *reach* options are colleges with smaller or no safety margins. We define safety margin as the difference between number a college's quota and a student's ranking at that college. A safety margin of zero indicates the student is ranked last. We find that JBCs applied to colleges with smaller safety margins compared to JACs, aligning with the theoretical predictions. On average, JBCs left 3 to 4 fewer positions in their standardized safety margins and were more inclined to apply to reach options.

Second, these differential choices result in significant differences in admission outcomes. We show that despite having slightly lower test scores, JBCs have significantly better admissions outcomes, including average peer quality, desirability of the geographic location, and average under-

graduates' labor market outcomes and graduate school attendance rates.[1] Analyzing data on test scores and admission outcomes between 2019 and 2023, we find that JBCs were admitted to significantly more selective colleges, equivalent to around 9 to 17 points in test scores, or 0.17 to 0.32 standard deviations. These colleges not only admit higher-achieving peers but also are around 10 percent more likely to be in an economically advanced city, such as the capital city of Beijing or a major coastal city like Shanghai or Shenzhen. Importantly, JBCs are admitted to colleges that yield 3.9 to 6.36 percent higher graduate school attendance rates and 5 to 8 percent higher average starting salary for their graduates. If these effects are driven by information disparity around the cutoffs, we should expect to see stronger effects around batching cutoffs where the information matters more. Indeed, we find larger gains from information access for students applying to middle-range colleges, where uncertainty is greater, as indicated by higher variance in admission cutoff scores across years.

Our paper connects a few strands of literature. First, we contribute to the literature on the effects of information on education matching market by providing novel evidence on how students navigate uncertainty about others' preferences. Starting from the seminal paper by Gale and Shapley (1962) on the college admission problems, the Gale-Shapley deferred-acceptance (DA) mechanism has been the workhorse in centralized college admissions (Balinski and Sönmez, 1999) and in school choice (Abdulkadiroğlu and Sönmez, 2003). In this framework, students submit their preferences over colleges or schools once, in a static manner, and DA ensures that truthful preference submission is a dominant strategy under complete information. However, extending this result of truth-telling guarantee to settings with incomplete information presents challenges (Roth, 1989; Ehlers and Massó, 2007). Studies have documented that in such complex environments, students indeed engage in strategic behavior, deviating from the theoretical predictions of DA mechanism in simpler environments (Hassidim et al., 2017; Chen and Pereyra, 2019; Shorrer and Sóvágó, 2023; Artemov et al., 2023).

A growing body of research has explored the dynamic or sequential implementation of matching mechanisms, both in theory and in lab, showing that dynamic mechanisms generally improve truthful preference reporting (Echenique et al., 2016; Klijn et al., 2019; Bó and Hakimov, 2020, 2022).

[1] Li and Qiu (2023) document a left-digit bias in which students just below a 10-point interval (e.g., 629) tend to make more conservative choices compared to those just above the interval (e.g., 630). If this also holds in Inner Mongolia, our findings would represent a lower bound on the effect of information access.

Hakimov et al. (2023a) offers an alternative justification, showing that sequential mechanisms reduce search cost by guiding students towards feasible choices. Our paper provides empirical evidence that, across batches, the sequential structure of the mechanism helps students to gain better access to information, in particular on the feasible choices. Similar to our context, Gong and Liang (2024) demonstrate that under the Inner Mongolian mechanism, stable allocations can emerge if students have sufficient opportunities for revision—a prediction that is confirmed by their experimental results. In addition, Chen et al. (2022) show that when students face time constraints, dynamic mechanisms like the one in Inner Mongolia is not stable nor efficient. However, these studies largely maintain the assumption of complete information, which implies that students fully observe others' choices before submitting their own. They also typically focus on straightforward strategies, where students apply to the most preferred feasible choices and avoid the infeasible ones. In contrast, this paper consider an environment where students face uncertainty about others' preferences. Instead of focusing on straightforward strategies, we analyze students' equilibrium strategies when facing deadlines.

Our study is also relevant to an emerging empirical literature evaluating sequential mechanisms. Related to our paper, Luflade (2017) analyzes the college admissions in Tunisia, where a three-phase sequential DA mechanism is used. Using a structural approach, the paper finds that providing students with updated information on vacant capacities mitigates the adverse effect due to limited choices. Comparing to the Tunisian system, the Inner Mongolian system provides more granular real-time updates, including students' ranking within the applied college, and allows more frequent revisions. These features enable us to analyze students' application strategies, revealing the impact of information access and highlighting concerns about its unequal effects. Dur et al. (2018) also discuss the unequal benefits of sequential mechanisms by distinguishing between sophisticated students who respond to updated information versus naive students who do not, showing that sophisticated students gain at the cost of naive students. Likewise, Kang et al. (2023) evaluate empirically that the Inner Mongolian dynamic mechanism produces higher instability than alternative mechanisms such as the Parallel mechanism. Using choice data and admission outcomes on students around the batching cutoffs, our results offer a more detailed understanding of the specific types of instability that arise in the Inner Mongolian mechanism, shedding light on how students respond to information updates. Several studies explore the behavioral aspects of how students

process information and make decisions under such mechanisms. For instance, Grenet et al. (2022) study preference discovery in college admissions where students have incomplete information about their own preferences. Using data from the Germany university admissions, they show that a dynamic multi-offer mechanism reduces learning cost for students and yields an improved outcome compared to the single-offer DA mechanism, partly by mitigating confirmation bias in the application process. Similarly, Hakimov et al. (2023b) use data from the French college admissions and show that student confidence influences application strategies. By providing feedback throughout the process, sequential mechanism can help students make more informed choices and reduce social disparity in the competition for elite colleges.

Additionally, our paper contributes to the understanding of application strategies in education matching markets. In a different setting – the static Parallel mechanism, which also involves uncertainty about others' choices – Song et al. (2020) find that lower-scoring students tend to apply conservatively, reflecting their pessimistic beliefs. Similarly, we find that students around the batching cutoffs adopt application strategies consistent with the equilibrium strategies predicted in the decentralized college admissions models. Chade et al. (2014) analyze students' optimal application portfolios when admission probabilities are independent and show the optimal strategy is aggressive. Ali and Shorrer (2025) extend the analysis to the situation with correlated admission probabilities, as in settings with common scores like ours, showing the optimal strategy is a mix of both "reach" and "hedge" strategies. The mechanism we study is a special case of correlated admissions: depending on the position of students within a batch, higher-priority students observe choices of others, employing a reach strategy, whereas lower-priority students face uncertainty about others' choices employ a hedging strategy.

Methodologically, instead of using discrete choice models to conduct counterfactual welfare analysis, we use a RD approach to identify the impact of information on various outcomes. The RD approach allow us to focus on more comparable students near batching cutoffs, whose preferences are more likely to be identical. In addition, our analysis extend beyond the score-related outcomes, which has usually been the focus of the literature.

Second, beyond the education markets, the use of dynamic mechanism with real-time information is often seen in the auction design. Arbitrage in the forms of sniping—delaying bids until the last moment—are well-documented in these markets (Roth and Ockenfels, 2002; Budish et al.,

2015). While, in theory, delaying preference submission to the last minute in hope of getting a better assignment is possible, the fact that students are ranked the same by all colleges according to their test scores, rules out strategic advantage of late submission. Moreover, students have incentives to submit their choices and acquire information about their acceptance chances.

Lastly, our study contributes to the broader literature on the importance of information in markets (Stigler, 1961; Akerlof, 1970; Jensen, 2007; Galdon-Sanchez et al., 2023). For instance, Jensen (2007) demonstrates how the introduction of mobile phone services in India significantly reduced price dispersion in the fishing industry. In the context of our study, students effectively "bid" for limited seats at colleges using their test scores. However, despite having nearly identical test scores, JACs and JBCs, students are admitted into colleges of varying quality. The observed variation suggests that access to information, rather than ability alone, influence outcomes. This finding offers compelling evidence on the role of information in matching markets and suggests that improving information design could lead to better student placement and market efficiency.

2 The IM Dynamic Mechanism and Application Strategies

2.1 The IM Dynamic Mechanism

We first outline the key features of the IM dynamic mechanism. During the period covered by our data (2019-2023), the mechanism operated as follows.

- The application process *begins* at the same time for all students. They must apply to one college and can revise their choices as many times as they want, at any time. When students apply a choice, they observe information including whether they are accepted by the applied college, the minimum acceptance cutoff at the college, and their rankings within that college (if accepted).

- The application process *terminates* following a staggered closing rule. Students in the first batch finalize their choices first. Students in the second batch are given an additional hour to adjust their choices before they finalize them. Similarly, each subsequent batch receives an additional hour to adjust their choices before finalizing them.

- Students' final assignments are based on their last submitted preferences.

Before the introduction of the IM dynamic mechanism, Inner Mongolia used a static Immediate Acceptance (IA) mechanism, also known as the Boston mechanism. Under IA, students submit a rank-ordered list of preferences, the mechanism then allocates seats first to first choices according to priorities determined by test scores, before considering second choices. A key concern with IA is that truthful preference reporting can be penalized: if a student is rejected by their first choice, the chance of being accepted by their second choice is greatly diminished, leaving them at the risk of being placed eventually at a much less desired option. This creates incentives for students to misreport their preferences by top-ranking a safe option where they have a higher chance of being accepted, rather than their true top choice. As a result, IA has been criticized for its poor incentive properties compared to DA (Abdulkadiroğlu and Sönmez, 2003).[2]

The welfare implications of IA are nuanced. In fact, if students have the necessary information to play the "game" under IA, then IA can yield the same outcome as the celebrated DA (Ergin and Sönmez, 2006), which finds stable matchings.[3] Consider the safe choice that they top rank is the one which they are assigned to under DA, then indeed IA and DA produce the same outcome. But, if such information is lacking, or students do not have the ability to process such information, then those form less advantageous backgrounds might be less sophisticated at gaming. Theoretical results corroborate this concern, showing that sophisticated students can gain at the cost of naive students (Pathak and Sönmez, 2008). The empirical evidence is however mixed. While Dur et al. (2018) find support for the conjecture that sophisticated families manipulate their reports and benefit disproportionately, Calsamiglia and Güell (2018) show that sophisticated families tend to rank according to true preferences, due to their outside options. In addition, He (2015) finds that students top rank safe choices too often. News report from Inner Mongolia echoed the latter point, noting that students under IA applying to less selective colleges as a safer option, only to regret their choices later upon realizing that their scores could have secured admission to better colleges.[4]

The IM dynamic mechanism was introduced with the aim to provide students with the information that helps them to attend more desirable colleges. On the one hand, by allowing students to

[2]The standard models in the literature assumes complete information. Another strand of literature departs from the complete information setting, in a setting without priorities—students have incomplete information—IA can outperform DA, as manipulation reveals information about their cardinal preference intensity and allow students to sort themselves according to how strongly they prefer certain colleges (Abdulkadiroğlu et al., 2011).

[3]A matching is stable when there are no student-college pairs would prefer each other over their current matching.

[4]See for example, report by China Youth Daily http://www.moe.gov.cn/jyb_xwfb/s5147/201508/t20150812_199644.html (in Chinese), accessed December 1, 2024

explore their options freely, they can obtain information about their application status and choose a preferred college where they are likely to be accepted. However, this flexibility brings in a coordination externality, especially for lower-scoring students, as they cannot be sure of the choices by their higher-scoring peers and are at a risk of being displaced. The staggered closing rule partially alleviates this issue by allowing students in a next batch to observe the choices made by the previous batch, recovering coordination across batches. The number of batches has gradually increased since its inception. During the period of our analysis, the number is around 9 to 10. This sequential aspect of the mechanism resonates with findings by Dur et al. (2021), who demonstrate that when IA is implemented sequentially, with complete information, students' strategic behaviors—skipping non-feasible choices—can lead to more efficient assignment outcomes than the ones achieved under DA. Within a batch, though, students do not apply sequentially and therefore lack full information about the choices of their higher-scoring peers.

Example 1. Consider a college admission market with 6 students $\{1, 2, 3, 4, 5, 6\}$ ranked in a descending order according to their test scores, with student 1 having the highest score and 6 the lowest. There are 2 colleges, where college A has a quota of 2 seats and B has a quota of 4 seats. The students are divided in two groups. The first batch consists of students 1 to 3, who must finalize their choices first. The second batch consists of students 4 to 6, who finalize their choices afterwards.

Preferences are private information, but are drawn independently from a known distribution. To capture the uncertainty within a batch, we assume when ending time arrives, everyone has time to make exactly one more change, without knowing perfectly what other students are applying to.

Table 1: Application status

(a) Application status observed by student 3

College	Total Seats	Applicants	Ranking	Safety Margin
A	2	1,3	2	0
B	4	2,3	2	2

(b) Application status observed by student 4

College	Total Seats	Applicants	Ranking	Safety Margin
A	2	1,4	2	0
B	4	2,3,4	3	1

Student 1 has the highest priority and can apply to the most preferred choice without worry

about the others' choices. Let us consider student 3, who ranks last in the first batch. Suppose the last application status update she observes is as indicated in Table 1a.

At college A, student 3 observes that one higher-priority student, student 1, has applied there, and she ranks second out of the two available seats. Will she be finally accepted there? It depends on the choice of student 2, who is ranked higher than her. If student 2 prefers college B and finalizes the choice there, student 3 will be safely assigned to college A. However, if student 2 prefers college A and decides to apply to it in the end, student 3 will be displaced.

At college B, student 3 observes that one higher-priority student, student 2, has applied there, and she ranks second out of the four available seats. Regardless of whether the other higher-priority student, student 1, ultimately prefers college A or B, student 3 will secure a seat at college B. The remaining quota of two seats provide a sufficient safety margin to accommodate both higher-priority students.

Since students can apply at any time and applications are not coordinated within a batch, uncertainty about others' choices persists. When student 3 see that student 2 applying to college B, it may simply be an attempt to gather information rather than a final decision. Thus, student 3 must weigh the expected payoff of applying to college A, where the risk of being displaced is significant, against the safer option of applying to college B, which has sufficient capacity to accommodate higher-priority students. This decision hinges on the trade-off between student 3's utilities and admission probabilities. The admission probabilities, in turn, depend on her belief about the choices of higher-priority students.

For illustration, suppose student 3 believes that student 1 will remain at college A, and student 2 will choose college A with a probability of 0.6 (our theoretical analysis extends this assumption by allowing students to update their beliefs based on observed rankings in the applied college). Then, the expected utility for student 3 of applying to college 1 and getting accepted there is $0.4u_A$, where u_A represents the utility of attending college A. Similarly, the expected utility of applying to college B is u_B, where u_B represents the utility of attending college 2. If $u_A > u_B > 0.4u_A$, student 3 prefers applying to college B in the end, even if student 3 prefers college A over college B. The greater safety margin at college B serves as a buffer, reducing the risk against being displaced by higher-priority students.

Now, consider student 4, who is ranked lower than student 3, but has the highest priority in the second batch. By the time to finalize choices, student 4 observes the full information of all higher-priority students as indicated in Table 1b. If student 4 prefers college A over college B, she

can safely apply to college A despite having no safety margin, as there is no risk of displacement from higher-priority applicants. □

The example above sheds light on how information availability affects students application strategies in the IM dynamic mechanism. First, higher-priority students, like student 1, face no or little risk of displacement. In contrast, lower-priority students, like student 3, must consider the possibility that higher-priority students might have similar preferences and compete for the same colleges. This increased uncertainty motives them to apply for safe options with higher admission probability.

Second, the staggered closing rule in the IM mechanism, give students at the top of the next group (JBCs), like student 4, an information advantage, who can observe the choices of the previous batches. This allows them to apply to more competitive colleges where they narrowly get in, "stealing" seats that might have been unattainable without this information advantage.

2.2 Theoretical Analysis

We formalize the above observations and allow students to learn about their admission probabilities. Consider a set of students $I = \{1, 2, \ldots, n\}$, who are to be assigned to a set of colleges $C = \{1, 2, \ldots, m\} \cup \{\emptyset\}$, where $\emptyset$ represents the option of being unassigned. Each college $c \in C$ has a quota of q_c.

Students have strict preferences over the set of colleges, including being unassigned, based on the utility they derive from each option. For a given student i, attending college j, with $j \in C$, gives a utility of u_{ij}, while being unassigned gives a utility of zero. Preferences are private information, meaning that each student knows their own preferences but not those of others. However, the distribution of preferences in the population is common knowledge.

Student priorities at each college are determined by their test scores, which are strict and common across all colleges. Each student knows their own test scores.

A matching μ is a mapping from I to C, $\mu : I \to C$, such that no more students would be accepted than the available quota, that is, $|\mu^{-1}(c)| \leq q_c$, $\forall c \in C$, where $|\mu^{-1}(c)$ is the set of students accepted to college c. A matching is *blocked* by a pair (i, c) if i prefers c to her assignment $\mu(i)$, and either c has some empty seats under μ, or there is a lower priority student i' who is

assigned to c under μ. That is, $c \succ_i \mu(i)$, and $|\mu^{-1}(c)| < q_c$ or $i \succ_c i'$ for some $i' \in \mu^{-1}(c)$. A matching is *stable* if it is not blocked by any such a pair (i, c).

Analyzing the strategic interactions under a dynamic mechanism like the IM dynamic mechanism is challenging. Previous research has pointed out that in such environment, dominant strategy might not exist and multiple equilibrium strategies can arise (Bó and Hakimov, 2022; Chen et al., 2022; Gong and Liang, 2024). Much of the research consider *straightforward strategies*, where students apply to the most preferred college first, and if rejected apply to the best college that they can be accepted. When using such strategies and when there is enough time to adjust choices, stable matching arise under the dynamic mechanism.

To simplify our analysis, we consider only one batch, and contrast the information asymmetry between high-priority students and low-priority students. In addition, we focus on the final-decision making of students before the deadlines—the choices that ultimately determines their assignment under the IM dynamic mechanism. Students know their overall priority ranking among all students, R_i. Student i can apply to a college c at any time $t \in [0, 1) \cup \{1\}$, and observe updates on their ranking within applied colleges, $r_{ic}(t)$ (we defer the formal definitions to Appendix A). The updates of $r_{ic}(t)$ allow students to assess the likelihood that higher-priority students have competing choices, thereby inferring their admission probability at college c. When the deadline approaches, $t = 1$, they must make a final choice that maximizes their expected utility.

As a benchmark, consider the ranking within an applied college c that student i would observe if applications are submitted sequentially in order of their priority ranking. This corresponds to the ranking in the stable matching, and we define it as:

$$\tilde{r}_{ic} = \sum_{j \in I \setminus \{i\}} \mathbf{1}\{R_j < R_i | j \in \tilde{\mu}(c)\} + 1,$$

where $\tilde{\mu}(c)$ represents the set of students matched to college c in the stable matching produced by the sequential application. The indicator function equals 1 if student j, with higher priority than i's, applied to the same college.

In the sequential application process, students with higher priority always apply first, ensuring that students who apply later can accurately observe their position among competing applicants. This allows them to know precisely where they are admitted to. However, when students do not

apply sequentially, the observed ranking, $r_{ic}(t)$, can change over time as higher-priority students change their choices, resulting in uncertainty about the final rankings.

Proposition 1. *In the IM dynamic mechanism, suppose for each student i, their observed ranking within the applied college, $r_{ic}(t)$, converges to $\tilde{r}_{ic}$, then when deadline approaches, applying to the most preferred college where their ranking falls within the quota, $\lim_{t\to 1} r_{ic}(t) \leq q_c$ is a Bayesian Nash equilibrium.*

The intuition of Proposition 1 is straightforward. When one's ranking updates within the applied college does not change, the uncertainty about others' choices vanishes, then the chance that a higher-priority student would apply to the same choice and displace her also diminishes. In this case, every student's final choice is effectively the assignment under the stable matching. Conversely, if the ranking updates continue to change, then uncertainty persists and applying to a less preferred option with higher admission probability might be optimal.

Proposition 2. *Consider two colleges c and k, where student i prefers college c over college k ($u_{ic} > u_{ik}$). If the observed ranking $r_{ic}(t)$ changes over time, applying to college k is a symmetric Bayesian Nash equilibrium, if the admission probability at k is higher than that at c to compensate the relative preference for college c.*

In the IM dynamic mechanism, if ranking continues to change until the final application period, which is more likely for students with lower-priority in a common priority environment, then they are more likely to apply to less preferred colleges with higher admission probabilities.

Denote the safety margin of student i at college c at the ending time $t = 1$ the number of quota at the college minus the student's rank, $s_{ic}(1) = \max(0, q_c - r_{ic}(1))$. Thus, the safety margin captures the buffer between the student's current position and the last available seat at the applied college. A student with a positive safety margin has enough capacity to avoid being displaced, even if higher-priority students apply to the same college. We have the following corollary, which is implied by Proposition 2.

Corollary 1. *In the IM dynamic mechanism, suppose the observed ranking continues to change until the ending time. Consider two students, i and j, where i has lower priority than j ($R_i > R_j$), and both have the same ordinal preferences over colleges. Then, the safety margins at their chosen colleges satisfies $s_{ic(i)}(1) \geq s_{jc(j)}(1)$, where $c(i)$ and $c(j)$ are the final choices of student i and j, respectively.*

Corollary 1 offers testable implications for our empirical analysis of student application strategies. While our theoretical analysis focus on a single batch, our empirical analysis will extend this framework to compare lower-priority students in an earlier batch to higher-priority students in the subsequent batch, who have identical ordinal preferences. This comparison allows us to assess the impact of information asymmetry on students' application strategies and their outcomes.

3 Background and Data

3.1 Centralized College Admissions in Inner Mongolia

China uses centralized procedures for college admissions. Each province is allocated quotas from colleges all over the country and runs a within-province matching based on students' submitted choice lists and college quotas. Student priority is almost solely based on their scores from the National College Entrance Exam (also known as "Gaokao"), which happens once a year in June. Colleges are differentiated by tiers, with Tier I colleges considered to have the best reputation, teaching and research quality, and infrastructure. The matching of students to colleges is done sequentially from Tier I to the next two tiers. The IM dynamic mechanism is applied to all tiers. Within the Tier I matching, there are two rounds. If a student fails to be admitted in the main round, there is a second round using the same procedure to allocate the remaining seats. If the student fails again to be admitted by a Tier I college, they will then participate in the matching of the next tiers. We focus on the matching of Tier I colleges in our main results and provide additional analyses on Tier II colleges in the appendix.

Provinces have the autonomy to choose their own matching mechanisms. Inner Mongolia, a northern Autonomous Region in China, introduced a dynamic matching mechanism in 2007.

We describe a typical procedure under Inner Mongolia's dynamic matching mechanism, using the year of 2019 as an example. The first round of Tier I matching took place on July 13th. The system opens up at 8:00, and students are not allowed to enter a new choice after 10:30 or delete their choice after 10:00. The former rule ensures students to enter some information in the system before the first batch has to finalize. The latter rule prevents students from emptying their choices once they submitted it once—they can revise their choices up till their batching deadline but not delete their choices. The batching deadlines are outlined in Table 2 below: students in the first batch can no longer revise their choices by 11:00, and the second batch by 12:00, and so

on. Unassigned students can enter a second round following the same procedure for the remaining quota.

Table 2: Batching Deadlines in 2019

Deadline	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00
Score Batch	≥670	640-669	610-639	580-609	550-579	520-549	490-519	460-489	≤460

Every hour, a tentative assignment of all students is published on the official website, with information on remaining quota for each college, including the test scores of those who are tentatively admitted and rejected.[5] Therefore, students from a later batch can observe the choices made by students from a previous batch, who can no longer revise their choices. However, students do not observe such information about their higher-scoring peers within the same batch.

For instance, at 10:55, students in the first batch need to carefully monitor their application status, as higher-scoring student might displace them from their current choice. This is especially the case for students near the batching cutoff of 670. The staggered deadlines create an advantage, though, for students in the second batch, who adjust their choices based on the finalized choices of the first batch. This leads to information asymmetry around the batching cutoff: JBCs posses more precise information about others' choices compared to JACs. News articles and online forums tell stories about "crunch time in Internet bars right before cutoff time" and that "having a score ending with 9 at the top of a batch is worth more than getting a better score." These anecdotes highlight that families and students are aware of, to some extent, the importance of information on feasible choices and opportunities to "steal" leftover seats at selective colleges by the previous batch of higher-scoring students. This awareness, and the question of whether students can effectively utilize this information advantage, motivate our empirical analysis.

3.2 Data

We use the administrative data on Inner Mongolia college admissions (2019 - 2023) together with the choice data (2020 - 2022) to first analyze application strategies and then examine subsequent admission outcomes. Our main sample is students whose scores are above Tier I university cutoff.[6]

[5] See https://www.nm.zsks.cn/22gkwb/index_12.html for an example.

[6] We repeat the same analyses on students in Tier II score ranges. Results on Tier II are qualitatively similar, and we discussed them in Appendix B.

For the choice data, we focus on the first round final choice data to analyze students school choice behavior.[7] Safety margin is measured as the number of students ranked behind a given student at a college, in other words, the difference between the college's quota and the student's rank at that college. When this number is zero, it means that the student herself is ranked last; if any student with a higher priority decides to change their choice to her choice school, she would be bumped out. Note that because quotas are much larger for local schools, which are usually less selective, this safety margin is much larger for local schools than non-local schools. To account for the variations in quotas, we divide the measure by the quota of school admitted and multiply it with the median of quotas in the sample. Table 3 shows the summary statistics for the choice data. Panel A shows statistics for the entire sample, and Panel B shows those for the RD sample, i.e. students whose scores are within five points of batching cutoffs.

For the admissions data, we observe the test scores used for ranking students by the algorithm.[8] We also observe the university and major every student was assigned to, as well as which round the student was assigned.

Our measures for the admission outcomes include average peer quality, university cutoff, university location indicators, average starting wage of graduates, and the percentage of graduate school attendance for the assigned university.[9] Average peer quality is measured by the average test score of all students admitted by the university, and university admission cutoff is the lowest test score among that group. University location indicators include whether the university is out-of-province and whether the university is in a coastal city or in the capital city of China. We choose these two measures because attending a university in these areas is a strong predictor of migration to more economically advanced regions in the long term.

Finally, the information on the average starting wage and the percentage of graduate school attendance comes from hand-collected university reports on post-graduation outcomes. Since not all universities provide these reports on their websites, we have data for only 48% of graduate school attendance and 34% of starting wages. We perform a balance check, presented in Table B1, and find that students just below a batching cutoff are equally likely to have missing wage data

[7] Note that roughly 98% of students are admitted in the first round.

[8] This score is the sum of raw total test score and any bonus score a student may receive, with the most common bonus score being the ten-point bonus for ethnic minority students.

[9] We use "college" and "university" interchangeably in the paper.

for their admitted university. However, they are 3% less likely to have missing graduate school attendance data. Students with a higher test score are also less likely to have missing graduate school attendance. Based on these findings, we place greater emphasis on the results using the starting wage data.

Table 4 presents summary statistics for the admissions outcomes data. Among the 115,761 students, over 80% are in the STEM track, with scores ranging from 418 to 717 out of a possible 750. Post-graduation outcomes vary widely across universities, with graduate school attendance rate ranging from as low as 2% to as high as 80%. Similarly, starting wages show significant variations, with the lowest average reported at 3,940 yuan (around 540 USD) per month and the highest at 12,183 yuan (around 1,690 USD) per month.

Figure 1 shows the raw distributions of minimum and average scores for universities admitted in Inner Mongolia, while Figure 2 shows the same for Guangxi, a province using the parallel mechanism as their matching mechanism without batching cutoffs. As we can see, the distributions for Guangxi are smooth, while Inner Mongolia exhibits steep jumps just below the batching cutoffs, indicated by the red vertical lines. Additional raw distributions for other variables of interest, including student choice behavior, university locations, and average post-graduation outcomes, are presented in Figure B2, all of which show noticeable bumps around the batching cutoffs.

3.3 Empirical Strategy

As discussed in the theoretical and institutional background, scoring just below the batching cutoffs enables students at the top of the next batch to gain valuable information about the finalized choices of higher-scoring students and the remaining slots at selective universities. To investigate how access to this information influences student choice strategy and admission outcomes, we use proximity to the batching cutoffs as an instrument for access to valuable information in college admissions and estimate the following RD specification:

$$Y_{icgt} = \beta I(\text{Score}_{icgt} < \text{Cutoff}_{cgt}) + \gamma_1 f(\text{Score}_{icgt} - \text{Cutoff}_{cgt})$$
$$+\gamma_2 I(\text{Score}_{icgt} < \text{Cutoff}_{cgt}) * f(\text{Score}_{icgt} - \text{Cutoff}_{cgt}) + \theta X_i + \mu_{cgt} + \epsilon_{igt}, \quad (1)$$

where Y_{icgt} is a set of choice and admission outcome variables for student i in year t and track g (science or social science), whose score is closest to cutoff c. The variable Score_{icgt} is student i's test

score, and Cutoff_{cgt} is the closest deadline cutoff score for student i. The indicator $I(\text{Score}_{icgt} < \text{Cutoff}_{cgt})$ equals 1 if student i's test score is below the closest deadline cutoff. We include a function of the running variable, $(\text{Score}_{icgt} - \text{Cutoff}_{cgt})$, the distance between the test score and the closest deadline cutoff, and its interaction with the indicator of below the cutoff. We use linear and quadratic functions, as well as the local polynomial non-parametric estimation and inference procedure (Calonico et al., 2014). For the main result, we use a 5-point bandwidth and uniform kernel weights. Our results are robust to alternative bandwidths and kernel weights. Lastly, we control for cutoff-by-track-by-year fixed effects μ_{cgt}, and the standard errors are clustered at the county level.

A key assumption in RD design is that students cannot manipulate their scores around the cutoff. Since students only take the exam once a year and the score is the sum of multiple subjects, it is virtually impossible to target a score precisely within one or two points of the cutoff. Nevertheless, we perform a standard validity check by plotting student characteristics, such as gender and ethnicity, around the cutoff. If Han ethnic students were more advantaged and better able to position themselves just below the cutoffs, then we would expect to see a spike in this characteristic right below the cutoff. Reassuringly, Figure 3 shows that the distributions of these characteristics are smooth around the cutoffs.

4 Effects of Information on Choice Strategy and Admission Outcomes

4.1 Information and Safety Margin

To understand how information influences student choices, we gather student choice data between 2020 and 2022 and use the final first-round choice data to analyze students' choice strategy. Given the information uncertainty, we hypothesize that students just below the cutoffs are less cautious in leaving safety margins. Indeed, as shown in Figure 4, students just below the batching cutoffs (JBCs) are less cautious, as evidenced by the significantly smaller safety margin, i.e. the number of students behind them at the choice school. We see much larger average safety margin for non-standardized measures, driven by significantly larger quotas by local universities, which are usually less selective. On the other hand, the standardized measure allows us to better compare across cutoffs without being heavily influenced by quota variations. On average, a JBC leaves three to

four fewer students in terms of standardized safety margin. This means that even if two students with higher priority change their choice to this school, she would still be admitted.

If this pattern is driven by access to the finalized choice information of higher-priority students, we would not see such patterns before JBCs gain access to such information. Since we obtained hourly record of student submitted choices, we can test whether the RD results for differences in choice behavior hold across hours. Denote the last hour before JBCs can change their choice as $t = 0$, with $t = -1$, $t = -2$, and $t = -3$ representing the hours preceding this moment. We expect that the differences in choices between students below and above the cutoffs will emerge only during the hour when JBCs gain access to the finalized information, i.e. at $t = 0$. Indeed, this is what we observe. Figure 5 presents the coefficients for "below", showing the difference in safety margins between JBCs and JACs at four time points. The coefficients for $t = 0$ are statistically different from those for the other three time points at the 1% significance level.

So who is driving the results: the JACs or the JBCs? Does being the last in your batch significantly disadvantage you, or does being the first in your batch give you a notable advantage? Or is it both factors in play? Since the RD results, comparing across batches, don't allow us to distinguish between these potential drivers, we turn to the raw data for comparisons both within and across batches. First, by examining the raw distributions of admission outcomes in Figure 1, we observe a relatively smooth and almost linear relationship for both average and minimum scores. This roughly linear relationship supports our theoretical discussions that students have relatively homogeneous preferences and are assortatively matched to universities.

If JACs are responsible for the differences in admission outcomes between JBCs and JACs, we would expect to see a significant drop in admission outcomes just above the batch cutoffs when comparing JACs to higher-scoring students within the same batch. However, when comparing JACs with other students within the same batch, we find no significant disadvantage after adjusting for the positive linear relationship between scores and admission outcomes, as shown in Figure 1. In contrast, JBCs are significantly better off when compared to others within their own batch, even after adjusting for the same linear relationship between scores and outcomes. It supports our earlier argument about the coordination problem: students lack full information about the preferences of higher-priority students and their admission chances. When JBCs gained access to the final choices of higher-priority students, they can strategically secure spots in the last available positions at

selective universities, without the risk of being overtaken by a higher-priority student.

4.2 Information and Admission Outcomes

Students scoring just below the batching cutoffs receive more information on the finalized choices of higher-scoring peers and are able to apply to reach schools without fearing the scenario of getting bumped out last minute. How much gain does this information advantage translate into? We use several measures of admission outcomes and a RD design to answer this question.

Figure 6 plots the admission outcomes, including peer quality and post-graduation outcomes of the assigned university, against the distance to the batching cutoffs. The top left panel of the figure plots the average admission score, and the top right panel of the figure plots the minimum admission score as the outcome variable. Students with a lower score than the batching cutoffs were admitted to universities with significantly higher average and minimum admission scores. The second and third rows plot the average post-graduation outcomes and the geographic location of the assigned university. We see a significantly higher likelihood of being out-of-province and better average post-graduation outcomes for the universities assigned to students just below the cutoffs (JBCs). Since their true feasible sets are almost the same, this significantly better admission outcome is a result of access to complete information on the feasible set for JACs, and choices based on this complete information on feasible set improves the admission outcomes substantially.

Table 6 shows the RD results on the benefits of crucial information on admission outcomes. Even though JBCs have lower test scores than JACs, they were admitted to significantly more selective colleges, equivalent to around 9 to 17 points in test scores, or 0.17 to 0.32 standard deviations. These colleges not only admit higher-achieving peers but also are 13% to 19% more likely to be out-of-province and 8% to 12% more likely to be in the capital city of Beijing or on a coastal city, such as Shanghai or Shenzhen. They are also admitted to universities that yield higher average starting wages for their graduates and higher graduate school attendance rate. The magnitudes are economically meaningful: the schools they are assigned to have 3.9 to 6.36% higher graduate school attendance rates and 5 to 8% higher average starting salary for their graduates.

We test the robustness of our results to alternative bandwidths and kernel weights. Figure 8 shows the coefficients for the "below batching cutoff" dummy across different specifications. The results remain statistically significant and similar in magnitude. We also examine robustness with

respect to sample choice and school selectivity measures. First, we expand the analysis to include Tier II students, with sample statistics and results discussed in Appendix B. While the coefficients are slightly smaller, the findings still emphasize the importance of information for students just below the batching cutoffs. Next, we adjust the sample to ensure an equal number of score points above and below the cutoffs (e.g., 595-604). A robustness check including scores from -5 to +5 (e.g., 595-605) yields similar results, as shown in Table B2. We also restrict the analysis to students admitted in the first round of matching, excluding those admitted later. Results remain consistent, as shown in Panel A of Table 7. Finally, we use previous year's school average and minimum admission scores as selectivity measures, with results presented in Panel B of Table 7. The sample size is smaller due to missing data from 2019, but results remain statistically significant at the 1% level, with slightly larger magnitudes.

4.3 Heterogeneity by Gender, Income, and Ethnicity

We test whether the effects of information access differ by demographics. For all the heterogeneity analyses, we include an interaction term between a dummy variable for the demographic of interest (male, ethnic majority, and high-income) and the "below cutoff" dummy in our baseline regression as well as the interaction term between the dummy and the linear or quadratic controls. Figure 7 shows the coefficients for the main variable of interest, i.e. the interaction term between below and the corresponding demographic variable, with average and minimum peer quality measured by the college entrance exam score as the outcome variables.

The top panel shows the results by gender and ethnicity. One may suspect that male students just above the batch cutoffs might be less risk averse and leave lower safety margins, thus the benefit of concrete choice information about higher priority students could be smaller, leading to a negative interaction term. We find no statistically significant gender difference in the improvement of college peer quality for those just below the cutoffs, and the coefficients are positive.[10] Results for Han ethnicity are similar.

Another aspect of potential heterogeneity is sophistication. Disadvantaged students, such as those from low-income background, may not be as sophisticated to use the additional information

[10] Table B3 presents detailed results for the heterogeneity analyses. Male and Han ethnicity students leave more safety margin, on average, possibly due to stronger emphasis on major choice. Scoring below the batching cutoffs and gaining access to information allow these students to leave less safety margin and gaining admissions to better universities, although these coefficients are significantly insignificant.

when they score just below the batching cutoffs, compared with the more advantaged students. Our results provide some support to this hypothesis. The bottom panel of Figure 7 shows that students from counties with higher income per capita see a larger benefit from access to information.[11] This additional benefit from access to information is only significant for Top 20% county income and Top 40% county income, but not for Top 60% and Top 80% county income.

4.4 Larger RD Effects Where Information is More Valuable

If the results are driven by information, we would expect stronger effects around batching cutoffs, where information is more valuable. Which score range creates the most uncertainty for students in their admissions chances? Figure 10 shows the relationship between the standard deviation and mean of university cutoff score percentiles across sample years. The left figure presents data for all universities in Tier I admissions, while the right figure focuses on universities consistently admitting students in Tier I between 2005 and 2011. Both the highest and lowest ranked universities in Tier I exhibit low variation in cutoff scores over time. However, universities in the middle of the ranking have more variation in their cutoff scores, leading to greater uncertainty in admissions and higher informational value for students targeting these high-variance schools. Based on this, we hypothesize that the RD effects will be larger around batching cutoffs in the middle score ranges.

In Figure 9, we plot the RD coefficients for "below" using subsamples around specific batching cutoff scores, pooling samples across years. The dots represent results with linear controls, and the diamonds represent results with quadratic controls. We observe an overall inverted-U shape in the benefits of scoring just below the cutoff (and thus gaining complete information about the feasible set). The fact that middle-ranked students benefited more from access to information supports our hypothesis.

An alternative explanation for this inverted-U pattern could be the density distribution of students across test scores. If the number of students above the Tier I cutoff follows a normal distribution, we might expect a crowding effect that drives a similar pattern. In Figure B3, we plot the density distribution of student test scores in 2023, with a vertical line marking the Tier I cutoff. As shown, the number of students at the lower end of Tier I scores exceeds that in the middle and top ranges, as the normal distribution of scores spans the entire range, not just above the Tier I

[11] Table B4 presents detailed results for the heterogeneity analyses by county income. Students from higher income counties are admitted to schools with significantly better peer quality.

cutoff.

5 Conclusions

This paper presents novel empirical evidence on the effect of information provision in centralized college admissions and the potential unequal effects when information provided differs across students. Using administrative data from Inner Mongolia, where the matching mechanism provides real-time application information to students and processes matching batch by batch, we observe that students scoring just below the batch cutoffs were admitted to significantly more selective colleges compared to those scoring just above the batch cutoffs. These universities are more likely to be in an economically advantaged location and yield higher average starting salaries and higher graduate school attendance rates for their graduates. The effects are more pronounced in batches where information is more critical.

While information provision is generally seen as welfare-enhancing, our findings underscore that what information is provided and how it is structured fundamentally shape student choices and outcomes. The design of the matching mechanism plays a crucial role in how students process and respond to available information. In theory, if the dynamic mechanism allowed for multiple choices with a deferred acceptance mechanism, the discontinuity in outcomes around the batch cutoffs would likely diminish.

Our findings highlight a broader implication of information design in education matching markets. While theoretically certain mechanisms have great properties, realistically students may face cognitive constraints and information-processing hurdles (Arteaga et al., 2022). Designs that provides personalized information (Mulhern, 2021; Arteaga et al., 2022) or minimize the amount of information requested from and provided to the participants (Ashlagi et al., 2020), such as the iterative DA (Bó and Hakimov, 2022), could be useful and welfare-enhancing.

References

Atila Abdulkadiroğlu and Tayfun Sönmez. School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, 2003.

Atila Abdulkadiroğlu, Yeon-Koo Che, and Yosuke Yasuda. Resolving conflicting preferences in school choice: The "Boston mechanism" reconsidered. *American Economic Review*, 101(1):399–410, 2011.

George A Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500, 1970.

S Nageeb Ali and Ran I Shorrer. Hedging when applying: Simultaneous search with correlation, 2025.

Felipe Arteaga, Adam J Kapor, Christopher A Neilson, and Seth D Zimmerman. Smart matching platforms and heterogeneous beliefs in centralized school choice. *The Quarterly Journal of Economics*, 137(3):1791–1848, 2022.

Georgy Artemov, Yeon-Koo Che, and YingHua He. Stable matching with mistaken agents. *Journal of Political Economy Microeconomics*, 1(2):270–320, 2023.

Itai Ashlagi, Mark Braverman, Yash Kanoria, and Peng Shi. Clearing matching markets efficiently: informative signals and match recommendations. *Management Science*, 66(5):2163–2193, 2020.

Michel Balinski and Tayfun Sönmez. A tale of two mechanisms: Student placement. *Journal of Economic theory*, 84(1):73–94, 1999.

Inácio Bó and Rustamdjan Hakimov. Iterative versus standard deferred acceptance: Experimental evidence. *The Economic Journal*, 130(626):356–392, 2020.

Inácio Bó and Rustamdjan Hakimov. The iterative deferred acceptance mechanism. *Games and Economic Behavior*, 135:411–433, 2022.

Eric Budish, Peter Cramton, and John Shim. The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4):1547–1621, 2015.

Sebastian Calonico, Matias D Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

Caterina Calsamiglia and Maia Güell. Priorities in school choice: The case of the boston mechanism in barcelona. *Journal of Public Economics*, 163:20–36, 2018.

Hector Chade, Gregory Lewis, and Lones Smith. Student portfolios and the college admissions problem. *Review of Economic Studies*, 81(3):971–1002, 2014.

Li Chen and Juan Sebastián Pereyra. Self-selection in school choice. *Games and Economic Behavior*, 117:59–81, 2019.

Li Chen, Juan S Pereyra, and Min Zhu. Time-constrained dynamic mechanisms for college admissions. *arXiv preprint arXiv:2207.12179*, 2022.

Raj Chetty, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. Income segregation and intergenerational mobility across colleges in the united states. *The Quarterly Journal of Economics*, 135(3):1567–1633, 2020.

Esther Duflo and Emmanuel Saez. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. *The Quarterly journal of economics*, 118(3):815–842, 2003.

Umut Dur, Robert G Hammond, and Thayer Morrill. Identifying the harm of manipulable school-choice mechanisms. *American Economic Journal: Economic Policy*, 10(1):187–213, 2018.

Umut Dur, Robert G Hammond, and Onur Kesten. Sequential school choice: Theory and evidence from the field and lab. *Journal of Economic Theory*, 198:105344, 2021.

Federico Echenique, Alistair J Wilson, and Leeat Yariv. Clearinghouses for two-sided matching: An experimental study. *Quantitative Economics*, 7(2):449–482, 2016.

Lars Ehlers and Jordi Massó. Incomplete information and singleton cores in matching markets. *Journal of Economic Theory*, 136(1):587–600, 2007.

Haluk Ergin and Tayfun Sönmez. Games of school choice under the boston mechanism. *Journal of public Economics*, 90(1-2):215–237, 2006.

Jose E Galdon-Sanchez, Ricard Gil, and Guillermo Uriz Uharte. The value of information in competitive markets: Evidence from small and medium enterprises. *Available at SSRN 4022675*, 2023.

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69(1):9–15, 1962.

Binglin Gong and Yingzhi Liang. A dynamic matching mechanism for college admissions: Theory and experiment. *Management Science*, 2024.

Julien Grenet, YingHua He, and Dorothea Kübler. Preference discovery in university admissions: The case for dynamic multioffer mechanisms. *Journal of Political Economy*, 130(6):1427–1476, 2022.

Rustamdjan Hakimov, Dorothea Kübler, and Siqi Pan. Costly information acquisition in centralized matching markets. *Quantitative Economics*, 14(4):1447–1490, 2023a.

Rustamdjan Hakimov, Renke Schmacker, and Camille Terrier. Confidence and college applications: Evidence from a randomized intervention. Technical report, Working Paper, 2023b.

Avinatan Hassidim, Déborah Marciano, Assaf Romm, and Ran I Shorrer. The mechanism is truthful, why aren't you? *American Economic Review*, 107(5):220–224, 2017.

Justine S Hastings and Jeffrey M Weinstein. Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly journal of economics*, 123(4):1373–1414, 2008.

Yinghua He. Gaming the boston school choice mechanism in beijing. *TSE Working Paper*, 2015.

Mark Hoekstra. The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics*, 91(4):717–724, 2009.

Robert Jensen. The digital provide: Information (technology), market performance, and welfare in the south indian fisheries sector. *The Quarterly Journal of Economics*, 122(3):879–924, 2007.

Ruixue Jia and Hongbin Li. Just above the exam cutoff score: Elite college admission and wages in china. *Journal of Public Economics*, 196:104371, 2021.

Le Kang, Wei Ha, Marty Haoyuan Chen, and Yuhao Deng. Dynamic matching mechanism and matching stability in college admissions: Evidence from inner mongolia. *Available at SSRN 4555471*, 2023.

Lars J Kirkeboen, Edwin Leuven, and Magne Mogstad. Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111, 2016.

Flip Klijn, Joana Pais, and Marc Vorsatz. Static versus dynamic deferred acceptance in school choice: Theory and experiment. *Games and Economic Behavior*, 113:147–163, 2019.

Hongbin Li and Xinyao Qiu. Heuristics in self-evaluation: Evidence from the centralized college admission system in china. *Review of Economics and Statistics*, pages 1–36, 2023.

Margaux Luflade. The value of information in centralized school choice systems. *Working paper, Duke University*, 2017.

Annamaria Lusardi and Olivia S Mitchell. The economic importance of financial literacy: Theory and evidence. *American Economic Journal: Journal of Economic Literature*, 52(1):5–44, 2014.

Christine Mulhern. Changing college choices with personalized admissions information at scale: Evidence on naviance. *Journal of Labor Economics*, 39(1):219–262, 2021.

Parag A Pathak and Tayfun Sönmez. Leveling the playing field: Sincere and sophisticated players in the boston mechanism. *American Economic Review*, 98(4):1636–52, 2008.

Alvin E Roth. Two-sided matching with incomplete information about others' preferences. *Games and Economic Behavior*, 1(2):191–209, 1989.

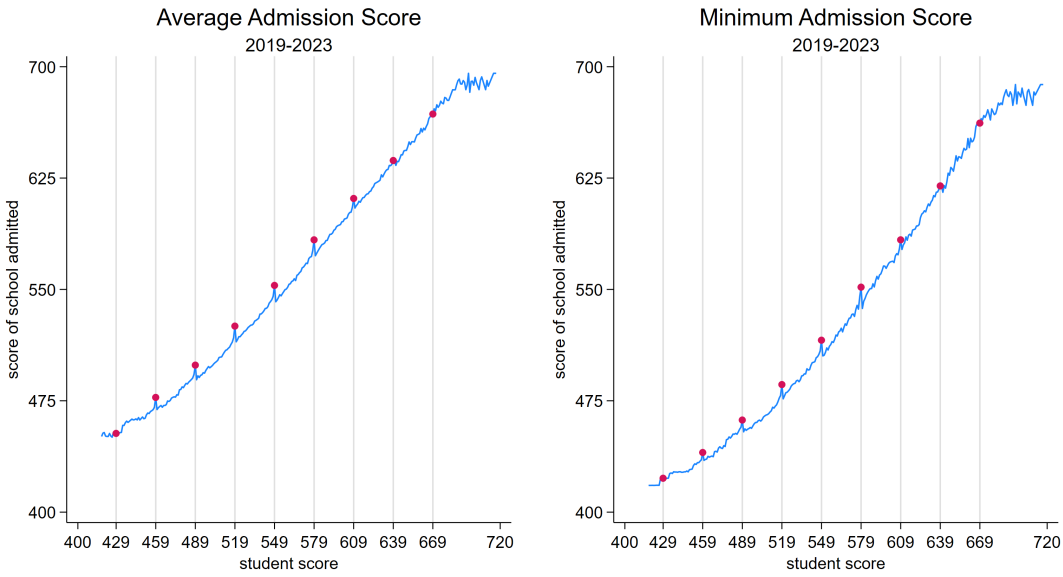
Alvin E Roth and Axel Ockenfels. Last-minute bidding and the rules for ending second-price auctions: Evidence from ebay and amazon auctions on the internet. *American economic review*, 92(4):1093–1103, 2002.

Ran I Shorrer and Sándor Sóvágó. Dominated choices in a strategically simple college admissions environment. *Journal of Political Economy Microeconomics*, 1(4):781–807, 2023.

Yan Song, Kentaro Tomoeda, and Xiaoyu Xia. Sophistication and cautiousness in college applications. *Available at SSRN 3711118*, 2020.

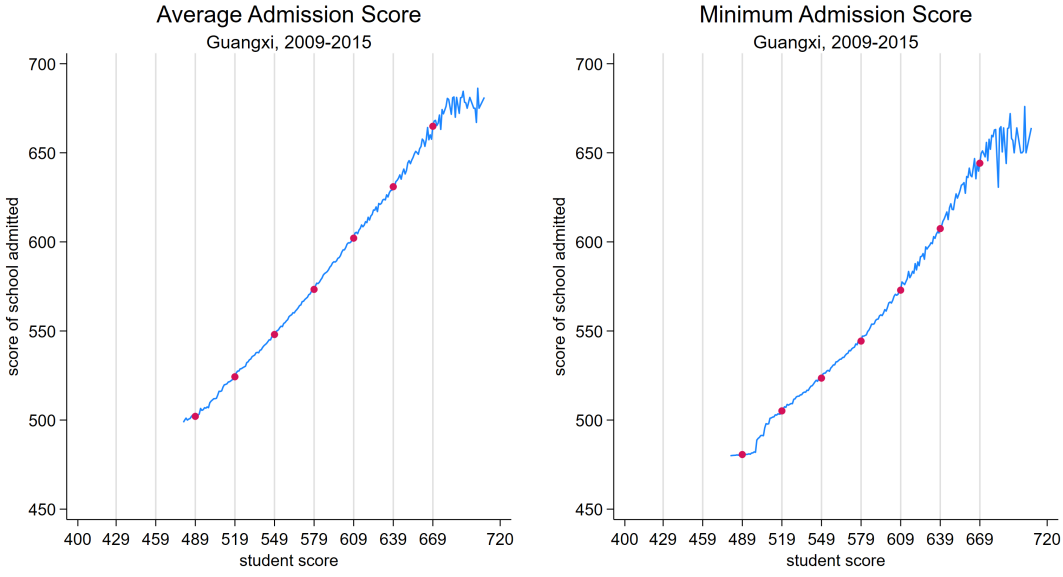
George J Stigler. The economics of information. *Journal of political economy*, 69(3):213–225, 1961.

Figure 1: Inner Mongolia: Distribution of Admission Outcomes Across Test Scores



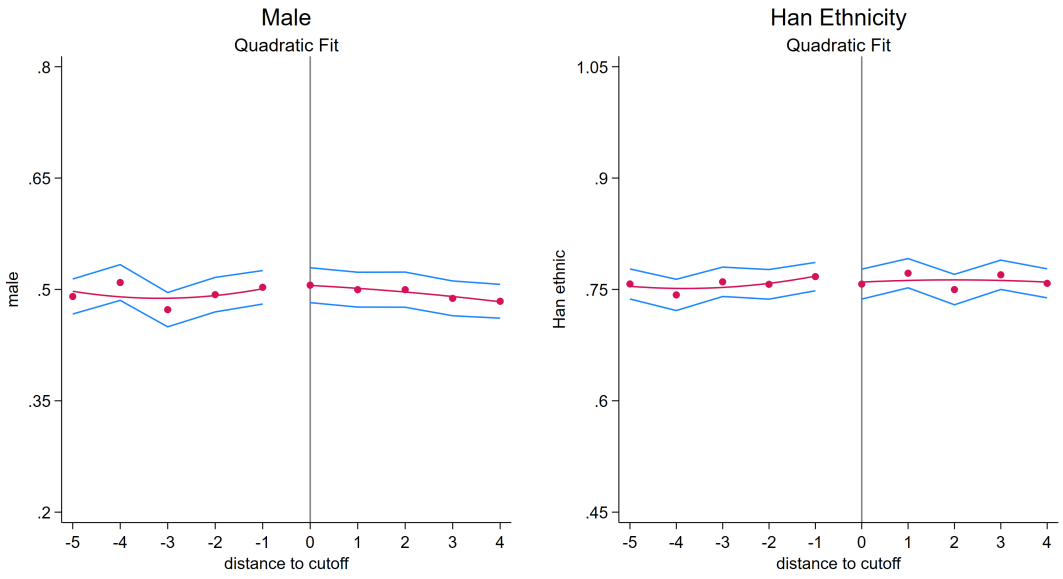
Notes: This figure plots the distribution of average scores and lowest scores for the university admitted for students with different test scores for Inner Mongolia between 2019 and 2023.

Figure 2: Guangxi as a Placebo: Distribution of Admission Outcomes Across Test Scores



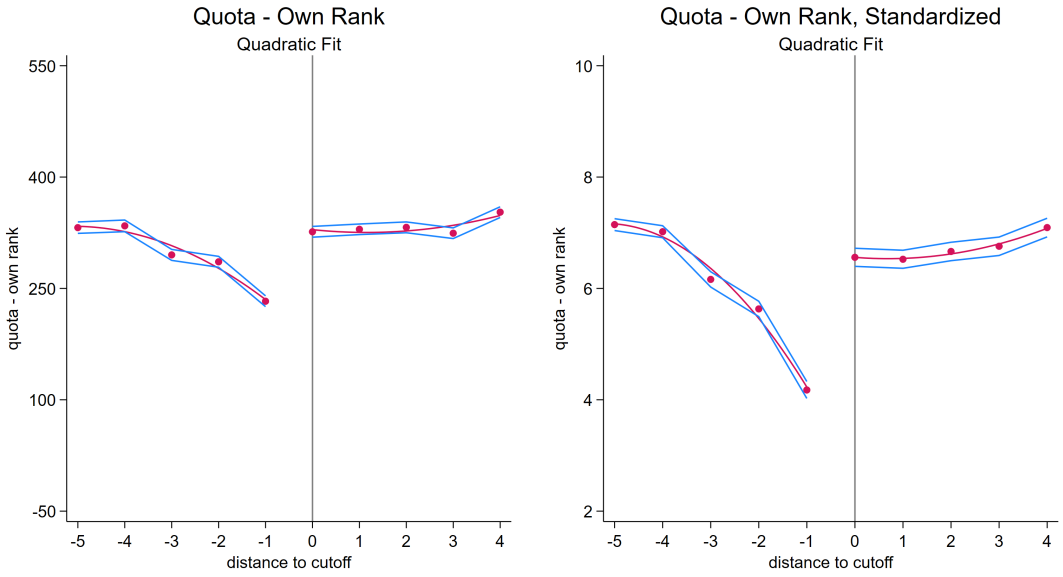
Notes: This figure plots the distribution of average scores and lowest scores for the university admitted for students with different test scores for Guangxi between 2009 and 2015. During this time, Guangxi adopted the parallel mechanism.

Figure 3: Validity Check for Regression Discontinuity Design



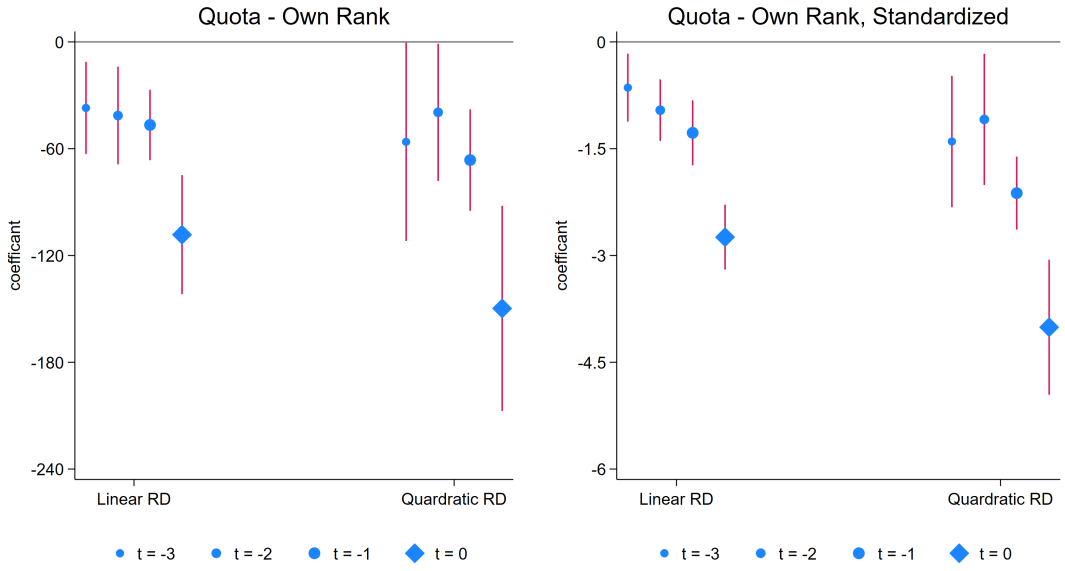
Notes: This figure plots gender and ethnicity against the distance to the batching deadline cutoff scores. The sample consists of observations within the 5-point bandwidth around the cutoffs. Each circle corresponds to one point in the test score. The red lines represent the fitted linear functions to the left and to the right of the cutoff. The blue lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Figure 4: Regression Discontinuity: Student Choice Data



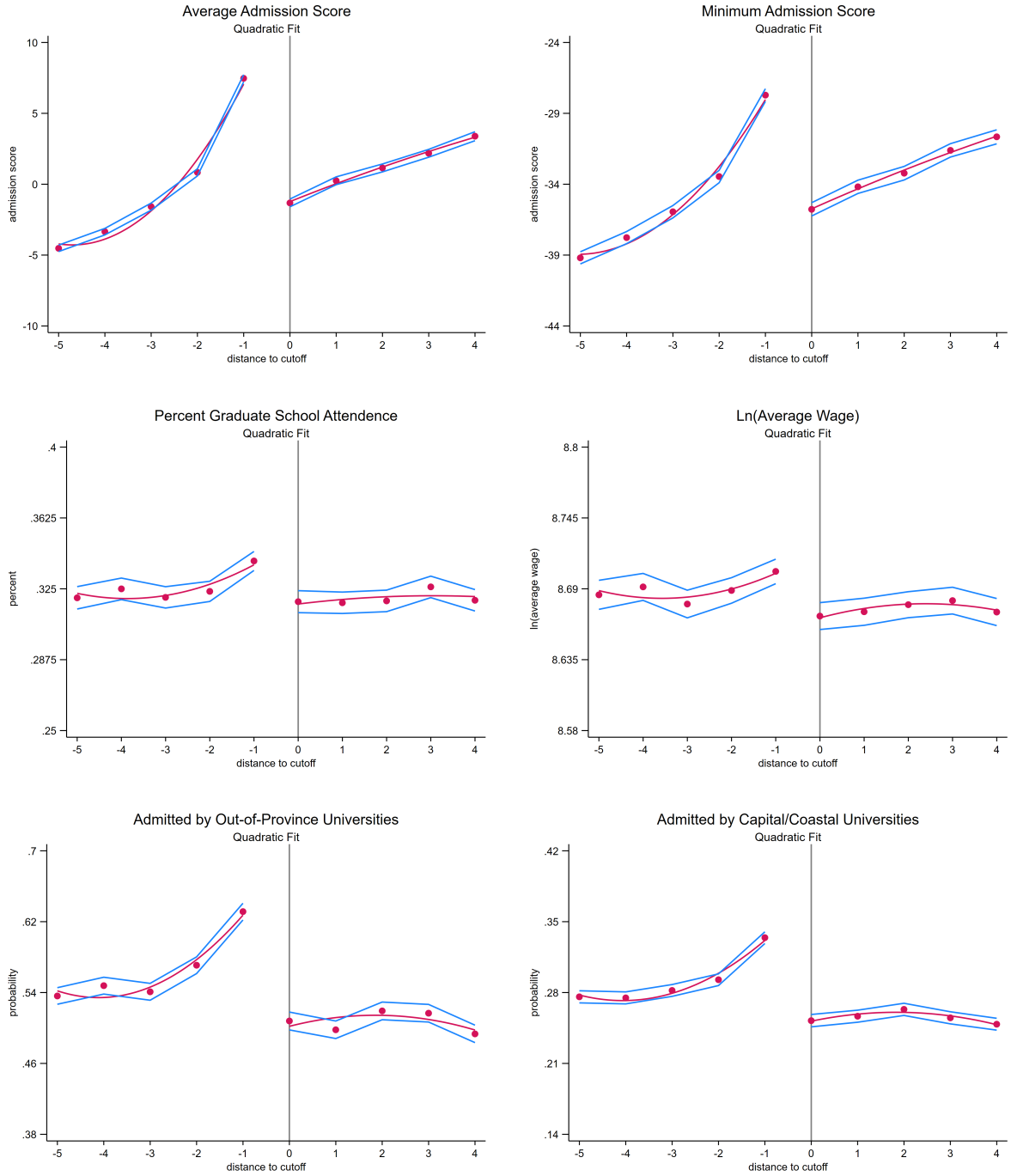
Notes: This figure plots the admission outcomes against the distance to the batching cutoffs. The left panel plots the quota of the admitted university minus own rank at the admitted university, our measure for safety margin. The right panel plots the standardized measure, by dividing the quota of the admitted school and multiplying the median of quota in the sample (17). The sample consists of observations within the 5-point bandwidth around the batching cutoffs among Tier I students between 2020 and 2022. Each circle corresponds to one point in the test score. The red straight lines represent the fitted linear functions to the left and to the right of the cutoff. The blue lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Figure 5: Regression Discontinuity: Student Choice Data



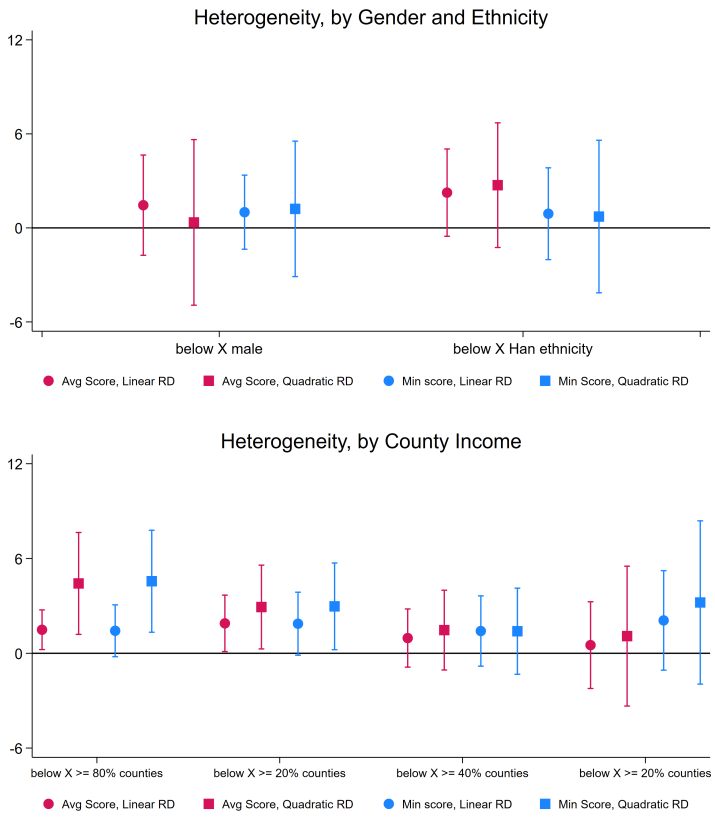
Notes: This figure plots the admission outcomes against the distance to the batching cutoffs. The left panel plots the quota of the admitted university minus own rank at the admitted university, our measure for safety margin. The right panel plots the standardized measure, by dividing the quota of the admitted school and multiplying the median of quota in the sample (17). The sample consists of observations within the 5-point bandwidth around the batching cutoffs among Tier I students between 2020 and 2022. Each circle corresponds to one point in the test score. The red straight lines represent the fitted linear functions to the left and to the right of the cutoff. The blue lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Figure 6: Regression Discontinuity: Student Outcomes



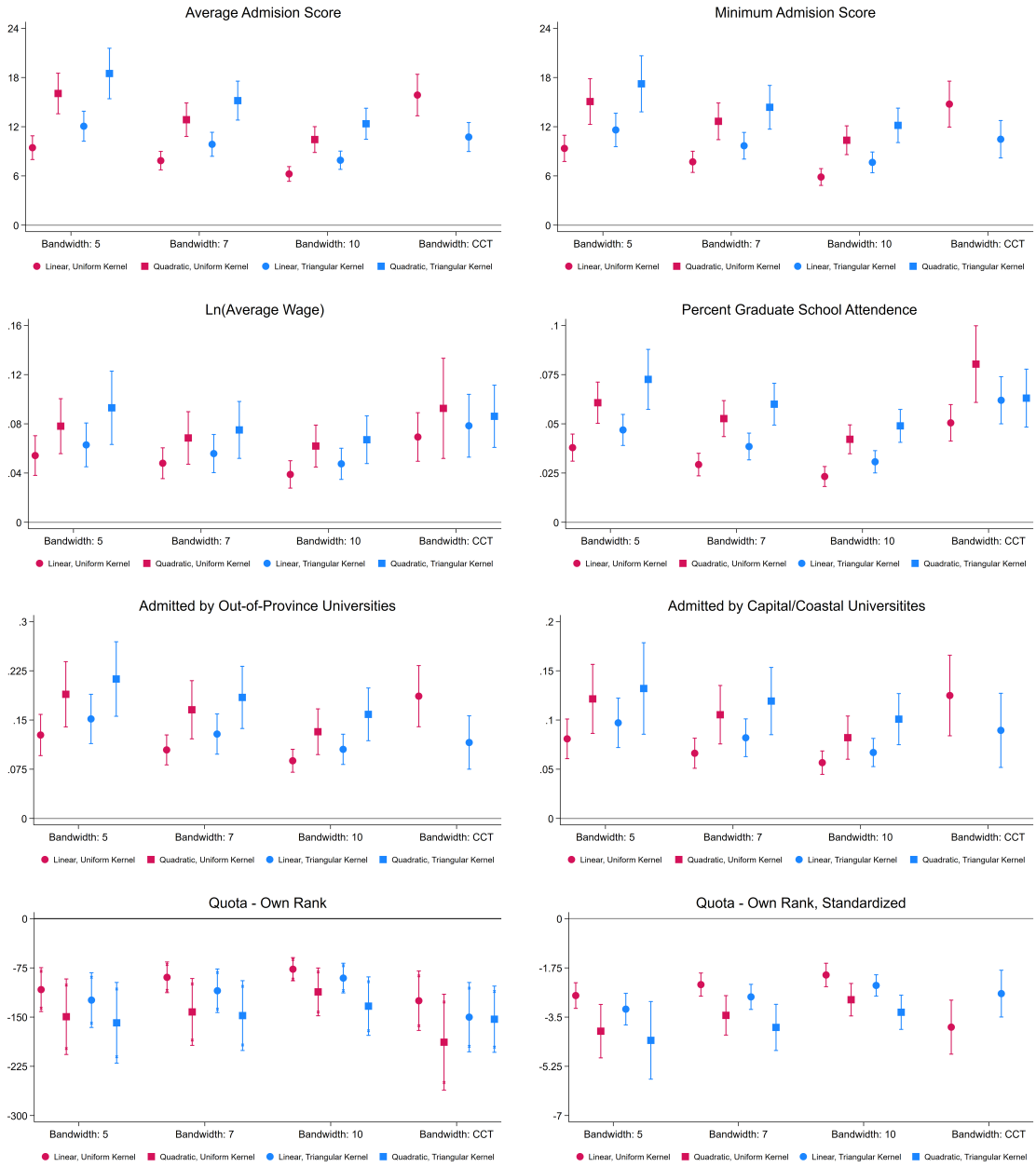
Notes: This figure plots the admission outcomes against the distance to the batching cutoffs. The top left panel of the figure plots the average admission score, and the top right panel of the figure plots the minimum admission score as the outcome variable. The bottom two panels plot the average postgraduation outcomes of the assigned university. The sample consists of observations within the 5-point bandwidth around the batching cutoffs among Tier I students between 2019 and 2023. Each circle corresponds to one point in the test score. The red straight lines represent the fitted linear functions to the left and to the right of the cutoff. The blue lines represent the lower and upper bounds of the 95% confidence interval for the sample mean of the outcome variable within the corresponding bin.

Figure 7: Heterogeneity of RD Results



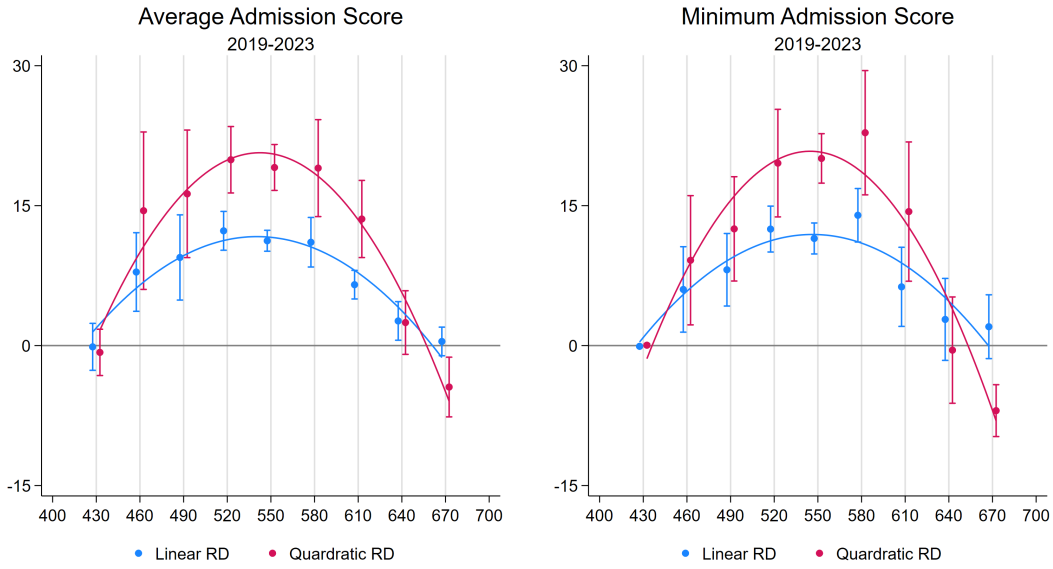
Notes: This figure plots the coefficients of interaction terms between "below" and the corresponding demographic variable of interest. A positive coefficient means that the focal group sees a larger benefit from access to information because of scoring just below the batching cutoffs. The sample consists of observations within 5-point bandwidth around the cutoffs. 95% confidence intervals are included around the coefficients.

Figure 8: RD Robustness Checks



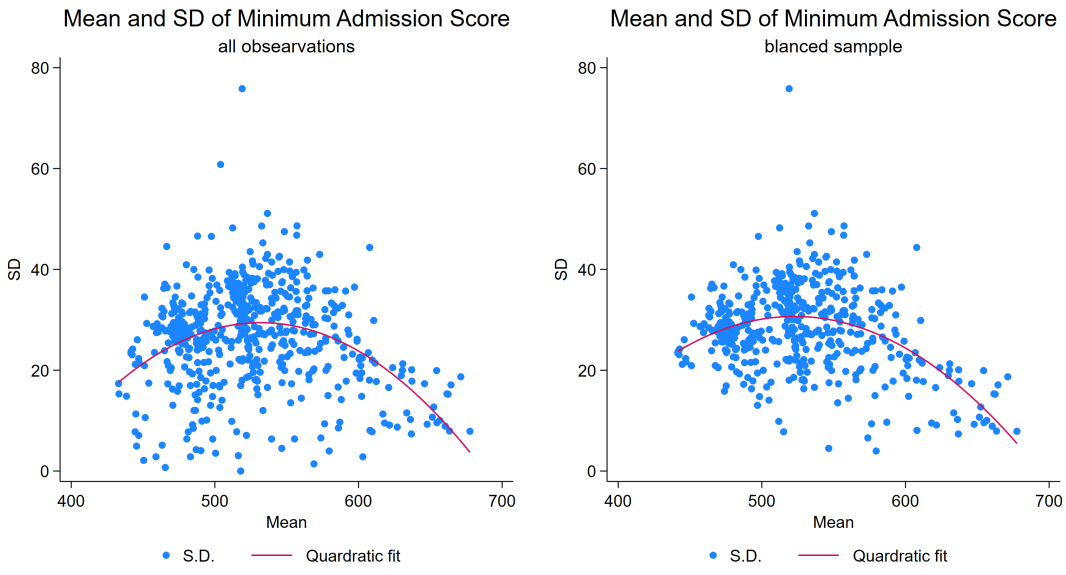
Notes: This figure visualizes admission outcomes relative to distance from the batching cutoffs. The top-left panel shows the average admission score, while the top-right panel presents the minimum admission score. The bottom-left panel displays the average graduate school attendance rate of the assigned university, and the bottom-right panel shows the average logged starting salary. Each point represents the estimated coefficient for "below", with 95% confidence intervals indicated by the extended lines. Red dots correspond to the uniform kernel, and blue dots to the triangular kernel. For the CCT bandwidth, only linear estimates are provided, as the selected bandwidth of two prevents the application of a quadratic function.

Figure 9: Value of Information: Effects of Access to Information by Batching Cutoffs



Notes: These figures plot the RD coefficients for "below" using subsamples around specific batching cutoff score, pooling samples across years. Each point represents the estimated coefficient for "below", with 95% confidence intervals indicated by the extended lines. The blue lines/dots are using linear controls, and the red are using quadratic controls.

Figure 10: Value of Information: Variability in Cutoffs and Selectivity



Notes: These figures present the relationship between the standard deviation and mean of university cutoff score percentile across the sample years. The left figure shows data for all universities that admit students in Tier I admissions; the right figure shows data for universities that consistently admit students in Tier I admissions between 2005 and 2011. They show that universities that are ranked the highest and lowest among the Tier I universities have low variations of cutoff score changes across years, indicating less uncertainty for students whose scores fall into these ranges.

Table 3: Summary Statistics: Data on Student Choice Behavior

Panel A. Student Choice Sample					
VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
stem	90,593	0.800	0.400	0	1
score	90,593	512.9	54.15	418	717
score - cutoff	90,593	-0.355	8.539	-15	47
quota	90,593	620.8	784.7	1	2,813
rank	90,593	293.1	442.3	1	2,482
quota - own rank	90,593	327.7	516.5	-245	2,812
quota - own rank, standardized	90,593	7.185	8.697	-175.7	16.99

Panel B. Student Choice RD Sample					
VARIABLES	(6) N	(7) mean	(8) sd	(9) min	(10) max
stem	31,545	0.799	0.401	0	1
score	31,545	509.5	54.77	425	674
score - cutoff	31,545	-0.417	2.843	-5	4
quota	31,545	631.3	796.4	1	2,813
rank	31,545	317.4	475.6	1	2,178
quota - own rank	31,545	314.0	505.8	-212	2,811
quota - own rank, standardized	31,545	6.354	10.05	-175.7	16.99

Notes: This table presents summary statistics for student choice data between 2020 and 2022. We standardize (quota-own rank) by dividing the quota of admitted school and multiplying the median of quota in the sample (17).

Table 4: Summary Statistics: Data on Student Admissions Outcomes

Panel A. Full Sample					
VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
STEM track	126,527	0.809	0.393	0	1
score	126,527	523.9	52.76	418	717
score - cutoff	126,527	-0.452	8.636	-15	47
average admission score	126,527	523.9	49.63	423	695.6
minimum admission score	126,527	489.7	55.19	418	688
non-IM universities	126,527	0.527	0.499	0	1
capital/coastal universites	126,527	0.269	0.443	0	1
miss first round	126,527	0.0317	0.175	0	1
grad-school attendance	63,903	0.321	0.161	0.0235	0.804
average wage	41,331	6,057	1,571	3,940	12,183
ln(average wage)	41,331	8.678	0.245	8.279	9.408
male	53,317	0.497	0.500	0	1
Han ethnicity	53,317	0.762	0.426	0	1
Panel B. Full RD Sample					
VARIABLES	(6) N	(7) mean	(8) sd	(9) min	(10) max
STEM track	42,663	0.810	0.392	0	1
score	42,663	523.1	52.79	425	674
score - cutoff	42,663	-0.498	2.868	-5	4
average admission score	42,663	524.1	49.85	423	684.5
minimum admission score	42,663	489.7	54.98	418	674
non-IM universities	42,663	0.537	0.499	0	1
capital/coastal universites	42,663	0.274	0.446	0	1
miss first round	42,663	0.0331	0.179	0	1
grad-school attendance	21,546	0.323	0.159	0.0235	0.804
average wage	13,797	6,080	1,565	3,940	12,183
ln(average wage)	13,797	8.682	0.243	8.279	9.408
male	17,669	0.494	0.500	0	1
Han ethnicity	17,669	0.759	0.428	0	1

Notes: Panel A presents summary statistics for students in Tier I from 2019 to 2023. Panel B presents summary statistics for the main sample for regression discontinuity analysis, i.e. those whose scores are within five points from the batching cutoffs.

Table 5: Information Uncertainty and Choice Safety Margin: Regression Discontinuity Result

VARIABLES	(1)	(2)	(3)	(4)
	quota - own rank		quota - own rank (std)	
below	-108.10***	-149.42***	-2.73***	-4.00***
	(10.26)	(18.11)	(0.23)	(0.42)
score - cutoff	4.29	-4.87	0.18***	-0.02
	(2.61)	(9.14)	(0.05)	(0.17)
below X (score - cutoff)	-25.82***	-48.54***	-0.59***	-1.41***
	(3.68)	(15.91)	(0.07)	(0.33)
(score - cutoff)^2		2.29		0.05
		(2.21)		(0.04)
below X (score - cutoff)^2		-7.63**		-0.22***
		(3.10)		(0.06)
Observations	31,544	31,544	31,544	31,544
R-squared	0.18	0.18	0.19	0.19
Bandwidth	5	5	5	5
Fitting	Linear	Quadratic	Linear	Quadratic
Year-track FE	Yes	Yes	Yes	Yes

Notes: This table presents regression discontinuity results for student choice behavior between 2020 and 2022. We standardize (quota-own rank) by dividing the quota of admitted school and multiplying the median of quota in the sample (17).

Table 6: Regression Discontinuity Result: Effect of Access to Information

Panel A. Effect on Peer Quality				
	(1)	(2)	(3)	(4)
VARIABLES	average admission score		minimum admission score	
below	9.44***	16.05***	9.36***	15.07***
	(0.40)	(0.81)	(0.41)	(1.01)
Observations	42,661	42,661	42,661	42,661
R-squared	0.90	0.90	0.87	0.87
Panel B. Effect on University Location				
	(1)	(2)	(3)	(4)
VARIABLES	non-IM universites		capital/coastal universities	
below	0.127***	0.189***	0.081***	0.121***
	(0.007)	(0.013)	(0.007)	(0.013)
Observations	42,661	42,661	42,661	42,661
R-squared	0.425	0.425	0.198	0.198
Panel C. Effect on University Average Postgraduation Outcomes				
	(1)	(2)	(3)	(4)
VARIABLES	% grad school attendance		ln(average wage)	
below	0.038***	0.061***	0.054***	0.078***
	(0.002)	(0.004)	(0.005)	(0.009)
Observations	21,544	21,544	13,797	13,797
R-squared	0.694	0.695	0.657	0.657
Bandwidth	5	5	5	5
Fitting	Linear	Quadratic	Linear	Quadratic
Year-track-cutoff FE	Yes	Yes	Yes	Yes

Notes: This table presents the OLS results from equation (1). Below is a binary variable indicating whether the student is right below a batching score cutoff. Sample is restricted to students whose scores are within five points from the batching cutoffs. Uniform Kernel is used for all regressions. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Robustness checks: Sample and Selectivity Measure

VARIABLES	(1) average admission score	(2) average admission score	(3) minimum admission score	(4) minimum admission score
Panel A. Sample restricted to first round admitted students				
below	9.83***	16.57***	9.70***	15.45***
	(0.40)	(0.77)	(0.41)	(1.00)
Observations	41,248	41,248	41,248	41,248
R-squared	0.90	0.90	0.87	0.87
Panel B. Previous year's school selectivity measures				
below	10.27***	16.87***	10.71***	17.23***
	(0.49)	(0.87)	(0.50)	(1.09)
Observations	34,566	34,566	34,566	34,566
R-squared	0.88	0.88	0.85	0.85

Notes: In Panel A, we restrict our analysis to those who are admitted during the first day of matching, excluding students who missed the first round and got admitted during the second or later rounds. In Panel B, we use the previous year's school average and minimum admission scores as the selectivity measures. All specifications follow Table 6.

Appendix A: Proofs

A.1 The extensive-form game

The IM dynamic mechanism extensive-form game contains the following elements:

- There are n students, $I = \{1, \ldots, n\}$.
- For student i, a ranking history at college c contains a sequence of observed rankings

$$H_{ic}(t) = \{r_{ic}(\tau) | \tau \in [0, t]\},$$

  and let $H_i(t) = \bigcup_{c \in C} H_{ic}(t)$ be all the histories observed by i.
- Students can apply to a college $c \in C$ at any time $t = [0, 1) \cup \{1\}$.
- A student observes their rankings at t and can apply to another college at t' , and the change is not instantaneous, that is, $t < t' < 1$.
- At $t = 1$, everyone observes their ranking histories up to t, and has time to make one more choice, without knowing what other students are choosing simultaneously.

Let $P_i\left(c|H_i(1), R_i\right)$ denote the admission probability to college c for student i given observed histories up to $t = 1$, $H_i(1)$, with priority R_i. Let P_i denote the vector of admission probabilities over all colleges. A strategy at the ending time $t = 1$ is a function $\sigma_i : P_i \times u_i \to C$, where u_i is a vector of student i's utilities over colleges.

A.2 Proof of Proposition 1

Consider the final decision stage at $t = 1$, where each student i chooses a college based on their beliefs about admission probabilities, P_i, given the observed ranking history $H_i(1)$ and the priority ranking R_i. If the observed rankings converge to their true values as if they apply sequentially, that is, $\lim_{t \to 1} r_{ic}(t) \to \tilde{r}_{ic}$, $\forall i, c$, then their admission probability for college c becomes:

$$P_i(c|H_i(1), R_i) = \begin{cases} 1, & \text{if } r_{ic} \leq q_c \\ 0, & \text{otherwise.} \end{cases}$$

This means that if i selects a college c where their final ranking falls within the quota, they are guaranteed admission, and they are certainly rejected otherwise. Then, every student strictly prefers

to apply to the most preferred college where their ranking is within the quota. Any deviation leads to a worse college.

A.3 Proof of Proposition 2

If the observed ranking $r_{ic}(t)$ continues to change, it indicates that the admission probability at college c remains uncertain when ending time approaches. This implies that the probability of displacement at c is nonzero because higher-priority students may still apply.

Given that other students follow their equilibrium strategies, the expected utility for student i from applying to college c is:

$$E[u_i|c] = P_i\left(c|H_i(1), R_i\right) u_{ic}.$$

Similarly, the expected utility for student i from applying to college k is:

$$E[u_i|k] = P_i\left(k|H_i(1), R_i\right) u_{ik}.$$

At equilibrium, student i chooses the college that maximizes her expected utility given others' strategies fixed. Then, applying to college k is optimal if and only if:

$$E[u_i|k] \geq E[u_i|c],$$

which simplifies to:

$$P_i(k|H_i(1), R_i)u_{ik} \geq P_i(c|H_i(1), R_i)u_{ic}.$$

Rearranging this inequality and combining with the assumption that $u_{ik} < u_{ic}$, we obtain the threshold condition:

$$1 \geq \frac{u_{ik}}{u_{ic}} \geq \frac{P_i(c|H_i(1), R_i)}{P_i(k|H_i(1), R_i)}.$$

Thus, student i applies to college k when the relative admission probability is high enough to offset their relative preference for college c. In a symmetric equilibrium, no student would deviate from this strategy. This establishes that applying to the less preferred but more accessible college k is a Bayesian Nash equilibrium strategy.

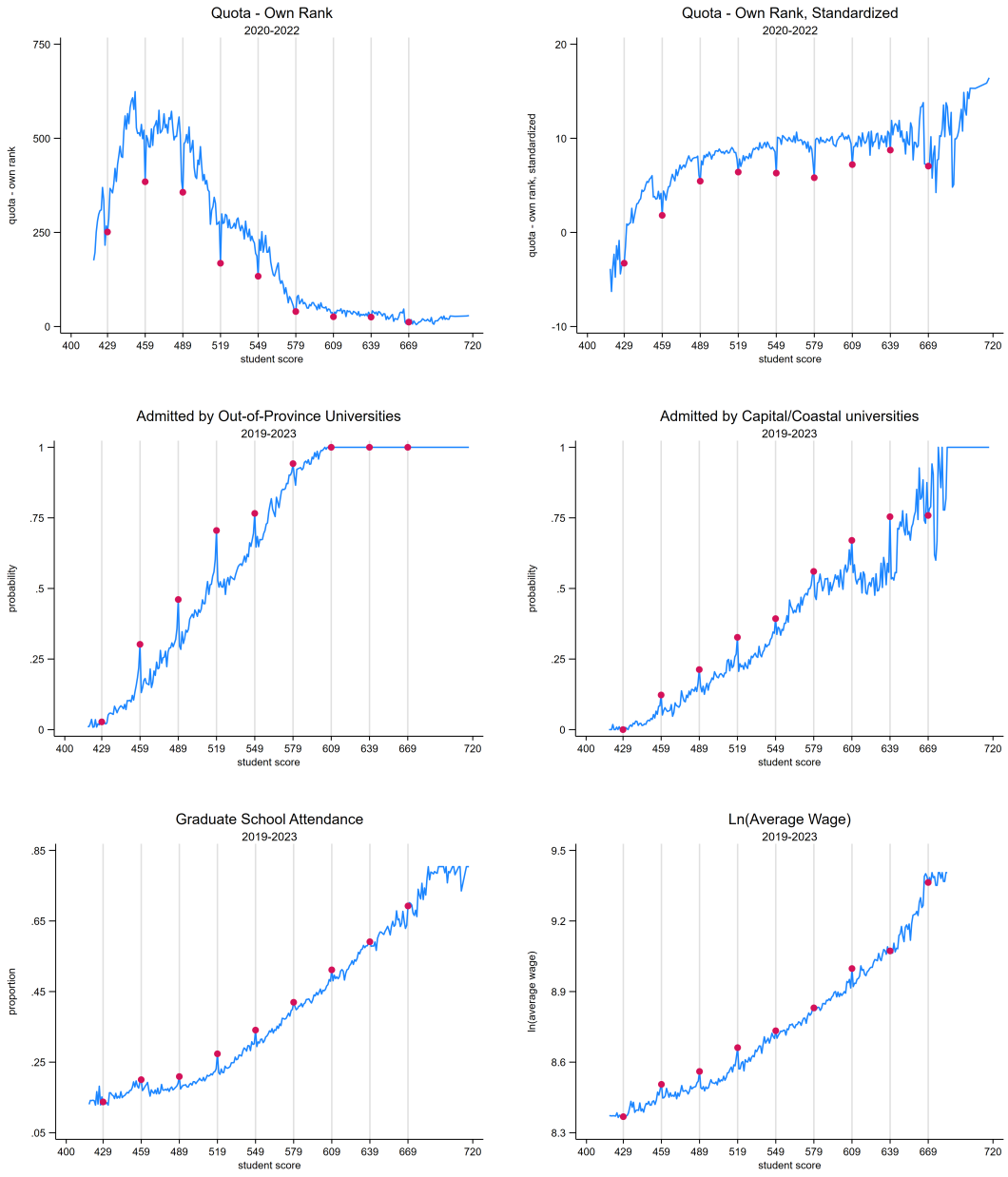
Appendix B: Supplementary Figures and Tables

Table B1: Balance Check: Missing Post-graduation Outcome

VARIABLES	(1) Missing graduate school attendance	(2)	(3) Missing wage	(4)
below cutoff	-0.082***	-0.131***	-0.039***	-0.072***
	(0.008)	(0.014)	(0.007)	(0.015)
score - cutoff	-0.003	-0.002	-0.000	0.004
	(0.002)	(0.007)	(0.002)	(0.008)
below cutoff X (score - cutoff)	-0.020***	-0.063***	-0.011***	-0.046***
	(0.003)	(0.015)	(0.003)	(0.014)
(score - cutoff)^2		-0.000		-0.001
		(0.002)		(0.002)
below cutoff X (score - cutoff)^2		-0.007***		-0.004
		(0.002)		(0.003)
Observations	42,661	42,661	42,661	42,661
R-squared	0.279	0.280	0.103	0.103
Bandwidth	5	5	5	5
Fitting	Linear	Quadratic	Linear	Quadratic
Year-track-cutoff FE	Yes	Yes	Yes	Yes

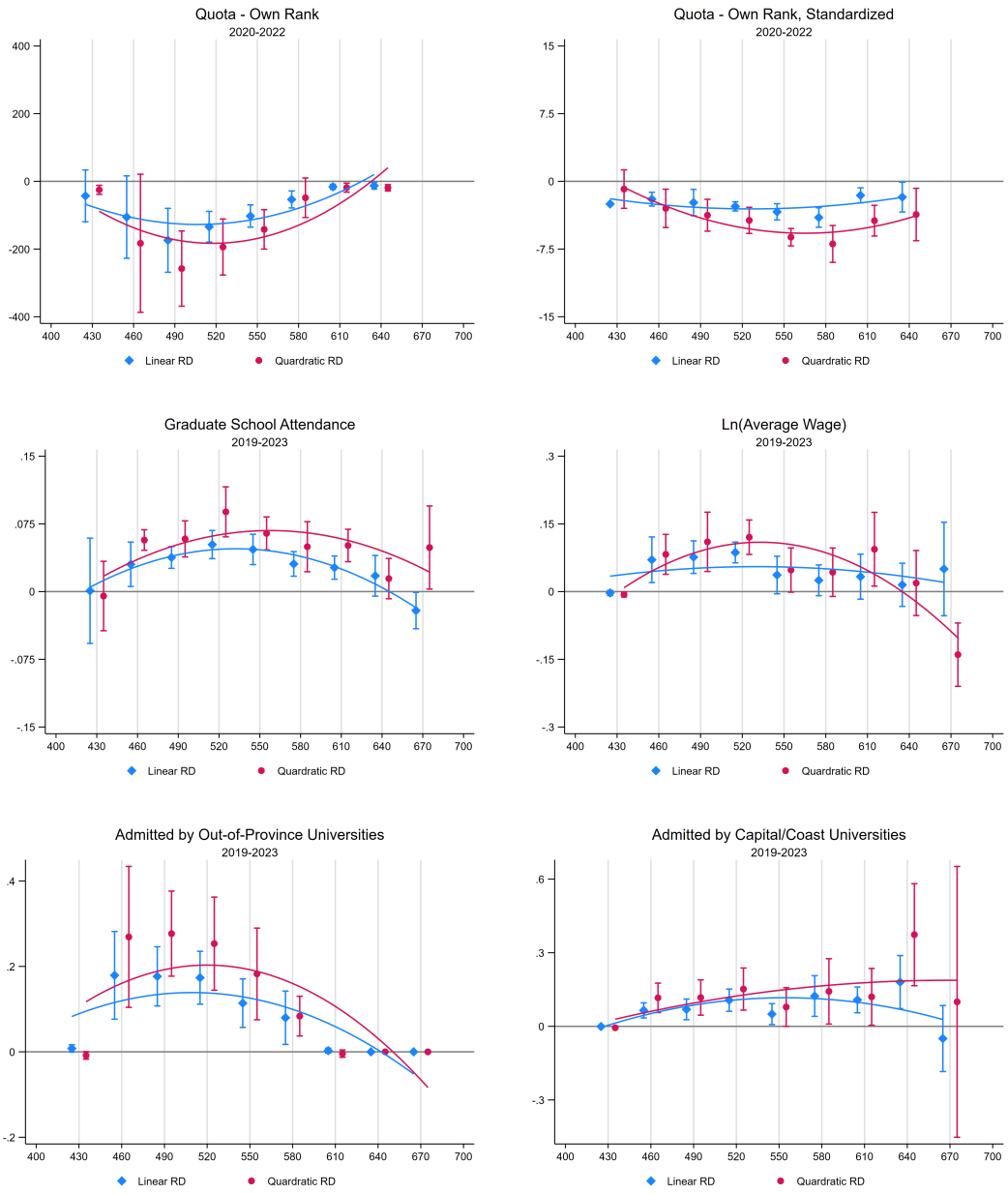
Notes: This table presents the OLS results from equation (1), intending to test whether missing values for postgraduation outcomes are balanced for above or below the batching cutoffs. Below is a binary variable indicating whether the student is right below a batching score cutoff. Sample is restricted to students whose scores are within five points from the batching cutoffs. Uniform Kernel is used for all regressions. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Figure B1: Inner Mongolia: Distributions of University Average Outcomes and Choice Behavior



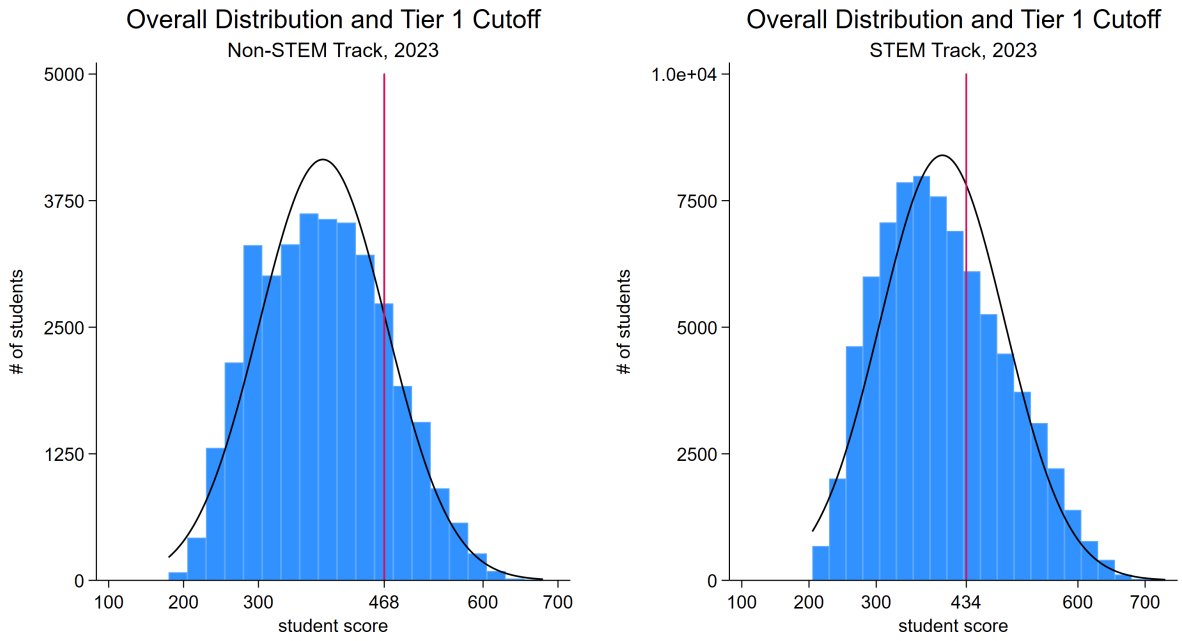
Notes: The figures show the raw distributions of student choice behavior and university characteristics of admitted colleges, averaging across our sample from 2019 to 2023.

Figure B2: Larger RD Effects Where Information Matters More



Notes: The figures show the RD results by batching cutoff scores for additional outcomes not presented in the main texts, using subsamples around each batching cutoff and pooling across years. Each point on the plots represents the coefficient of interest for "below", with 95% confidence interval indicated by the lines extending around it. The blue lines/dots are using linear controls, and the red are using quadratic controls.

Figure B3: Inner Mongolia: Distributions of Student Test Score in Relation to Tier I Cutoff in 2023



Notes: This figure plots the density distribution of student test score in 2023, with a vertical line indicating the Tier I cutoff. It shows a normal distribution of scores over the entire score range, and it is not a normal distribution above Tier I cutoff.

Table B2: Robustness Check: Including Students with Scores of Cutoff + 5

Panel A. Effect on Peer Quality				
	(1)	(2)	(3)	(4)
VARIABLES	average admission score		minimum admission score	
below cutoff	9.45***	15.90***	9.37***	14.90***
	(0.40)	(0.78)	(0.43)	(0.97)
Observations	46,897	46,897	46,897	46,897
R-squared	0.90	0.90	0.87	0.87
Panel B. Effect on University Location				
	(1)	(2)	(3)	(4)
VARIABLES	Non-IM universites		Capital/Coastal universities	
below cutoff	0.127***	0.188***	0.082***	0.120***
	(0.007)	(0.013)	(0.007)	(0.013)
Observations	46,897	46,897	46,897	46,897
R-squared	0.426	0.426	0.197	0.197
Panel C. Effect on University Postgraduation Outcomes				
	(1)	(2)	(3)	(4)
VARIABLES	% grad school attendance		ln(average wage)	
below cutoff	0.038***	0.061***	0.054***	0.077***
	(0.002)	(0.004)	(0.005)	(0.009)
Observations	23,624	23,624	15,121	15,121
R-squared	0.695	0.696	0.658	0.658
Panel D. Choice Safety Margin				
	(1)	(2)	(3)	(4)
VARIABLES	quota-own rank		quota-own rank (std)	
below cutoff	-111.43***	-145.40***	-2.74***	-3.96***
	(10.01)	(17.98)	(0.22)	(0.42)
Observations	34,610	34,610	34,610	34,610
R-squared	0.17	0.18	0.19	0.19
Fitting	Linear	Quadratic	Linear	Quadratic

Notes: This table presents the OLS results from equation (1), using a slightly different sample from our main result. In the main result, we include individuals with scores five points below and four points above. For example, we keep students with scores 595-604, with 595-599 being the below students and 600-604 being the above students. For this table, we include students with scores from minus five to plus five, and that would be 595-605 in the example above. Results here are remarkably similar to main results. Uniform Kernel is used for all regressions, and year-by-track-by-cutoff fixed effects are included in all specifications. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table B3: Heterogeneous Effects of Scoring Below Batching Cutoffs

Panel A. Heterogeneity by Gender								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	quota - own rank		quota - own rank (std)		average admission score		minimum admission score	
belowXmale	-37.53	-51.55	-0.62	0.42	1.45	0.50	1.00	1.34
	(25.50)	(38.22)	(0.39)	(0.78)	(1.63)	(2.69)	(1.21)	(2.15)
below	-91.99***	-129.41***	-2.47***	-4.16***	9.50***	16.90***	9.06***	14.53***
	(16.14)	(27.28)	(0.33)	(0.47)	(1.27)	(1.72)	(1.49)	(2.32)
male	78.62***	71.79**	0.67***	0.58**	-1.84*	-1.36	0.50	0.87
	(27.97)	(28.86)	(0.23)	(0.27)	(1.03)	(1.05)	(0.63)	(0.64)
Observations	29,414	29,414	29,414	29,414	17,668	17,668	17,668	17,668
R-squared	0.18	0.18	0.19	0.19	0.86	0.86	0.83	0.83
Panel B. Heterogeneity by Ethnicity								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	quota - own rank		quota - own rank (std)		average admission score		minimum admission score	
belowXhan	-15.61	-32.95	0.39	0.13	2.25	2.40	0.91	0.07
	(25.13)	(32.44)	(0.47)	(0.75)	(1.42)	(2.04)	(1.49)	(2.46)
below	-99.54***	-128.42***	-3.08***	-4.04***	8.51***	15.32***	8.86***	15.17***
	(29.67)	(35.75)	(0.41)	(0.67)	(1.78)	(2.89)	(1.86)	(3.18)
han	53.45***	47.08***	0.80***	0.96***	-2.03***	-2.00***	-1.53	-1.08
	(17.20)	(14.43)	(0.23)	(0.33)	(0.71)	(0.68)	(1.00)	(0.79)
Observations	29,414	29,414	29,414	29,414	17,668	17,668	17,668	17,668
R-squared	0.18	0.18	0.19	0.19	0.86	0.86	0.83	0.83
Fitting	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic

Note: The odd-numbered columns control for linear distance from the cutoff score, interaction term between below and linear distance (allowing for differential relationship below and above the cutoff), and interactions between the demographic dummy and these two variables. The even-numbered columns control for the quadratic terms of distance from the cutoff score (separated for above and below the cutoff), as well as the interaction terms with the demographic dummy. Robust standard errors clustered at the year-by-stem-by-cutoff in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table B4: Heterogeneous Effects of Scoring Below Batching Cutoffs by County Income

High SES proxy	County Income Top 20%				County Income Top 40%			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	average admission score		minimum admission score		average admission score		minimum admission score	
belowXhighSES	1.49**	4.42***	1.43*	4.56***	1.90**	2.93**	1.87*	2.97**
	(0.64)	(1.65)	(0.84)	(1.65)	(0.91)	(1.35)	(1.02)	(1.40)
below	9.16***	15.18***	9.09***	14.17***	8.48***	14.53***	8.41***	13.52***
	(0.76)	(1.33)	(0.83)	(1.43)	(1.03)	(1.61)	(1.04)	(1.62)
high SES	1.55***	1.89***	1.98***	2.30***	1.39***	1.21***	1.94***	1.81***
	(0.54)	(0.66)	(0.68)	(0.81)	(0.43)	(0.45)	(0.57)	(0.62)
R-squared	0.90	0.90	0.87	0.87	0.90	0.90	0.87	0.87
High SES proxy	County Income Top 60%				County Income Top 80%			
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
VARIABLES	average admission score		minimum admission score		average admission score		minimum admission score	
belowXhighSES	0.96	1.47	1.41	1.40	0.52	1.09	2.08	3.22
	(0.94)	(1.29)	(1.13)	(1.39)	(1.40)	(2.26)	(1.61)	(2.64)
below	8.76***	15.02***	8.36***	14.07***	8.95***	15.03***	7.44***	12.10***
	(1.21)	(1.81)	(1.26)	(1.88)	(1.71)	(2.56)	(1.87)	(3.00)
high SES	0.97**	1.16**	1.62***	1.86**	2.49***	2.81***	2.38***	2.35**
	(0.42)	(0.52)	(0.57)	(0.72)	(0.66)	(0.82)	(0.85)	(1.06)
R-squared	0.90	0.90	0.87	0.87	0.90	0.90	0.87	0.87
Fitting	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic

Note: Obs = 42, 661 for all regressions. The odd-numbered columns control for linear distance from the cutoff score, interaction term between below and linear distance (allowing for differential relationship below and above the cutoff), and interactions between the demographic dummy and these two variables. The even-numbered columns control for the quadratic terms of distance from the cutoff score (separated for above and below the cutoff), as well as the interaction terms with the demographic dummy. We cannot analyze choice behavior for the county-income based heterogeneity analysis because the choice data do not include necessary information for us to merge in county income. Robust standard errors clustered at the year-by-stem-by-cutoff in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Appendix C: Admissions to Tier II Universities

Inner Mongolia's matching process follows a sequential order based on the official university Tier designation, which categorizes schools into three tiers. Our primary analysis focuses on students whose scores exceed the Tier I university admission cutoff, allowing them to participate in the initial rounds of admissions. Once the market clears for these students, the online platform then opens to those with scores within the Tier II admissions range, operating in a similar batch-based manner with distinct choice-finalizing deadlines. Below, we present the summary statistics and RD results for these students. Note that we only have data on admission outcomes, based on administrative data, but do not have data on student choice, which were web-scraped for Tier I admissions process only.

Table C1 show that student scores range from 301 and 521, below our main analytical sample. A similarly small percentage (2.7%) of students is not assigned to any school in the first round (i.e. one full day of dynamic submissions). It is harder to collect postgraduation outcomes from lower ranked schools, thus we have a much smaller percentage of students for whom we observe their university's average starting wage and graduate school attendance. We caution to focus on the results related to average selectivity and the geographic location of the schools (Panel A and B) in Table C2. The benefit of scoring just below the batching cutoffs is slightly smaller in terms of higher selectivity and higher likelihood of out-of-province university, but the magnitude is still non-negligible and statistically significant at the one percent level.

Table C1: Summary Statistics: Students in Tier II Universities Admission Range

Panel A. Full Sample					
VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
year	175,527	2,021	1.445	2,019	2,023
STEM track	175,527	0.710	0.454	0	1
score	175,527	401.4	45.05	301	521
score - cutoff	175,527	-0.807	6.276	-29	11
average admission score	175,527	401.4	42.65	303	521
minimum admission score	175,527	376.3	43.78	301	521
non-IM universities	175,527	0.703	0.457	0	1
capital/coastal universites	175,527	0.0123	0.110	0	1
miss first round	175,527	0.0274	0.163	0	1
grad-school attendance proportion	20,068	0.101	0.0653	0.0342	0.415
average wage	10,491	4,373	204.5	3,940	6,708
ln(average wage)	10,491	8.382	0.0428	8.279	8.811
male	72,649	0.431	0.495	0	1
Han ethnic	72,649	0.766	0.423	0	1
Panel B. Full RD Sample					
VARIABLES	(6) N	(7) mean	(8) sd	(9) min	(10) max
year	85,749	2,021	1.458	2,019	2,023
STEM track	85,749	0.707	0.455	0	1
score	85,749	403.7	43.66	325	514
score - cutoff	85,749	-0.516	2.889	-5	4
average admission score	85,749	403.6	41.40	306.6	517.8
minimum admission score	85,749	378.0	43.07	301	514
non-IM universities	85,749	0.695	0.460	0	1
capital/coastal universites	85,749	0.0144	0.119	0	1
miss first round	85,749	0.0254	0.157	0	1
grad-school attendance proportion	10,178	0.102	0.0669	0.0342	0.415
average wage	5,250	4,381	218.3	3,940	6,708
ln(average wage)	5,250	8.384	0.0456	8.279	8.811
male	36,374	0.428	0.495	0	1
Han ethnic	36,374	0.767	0.423	0	1

Notes: Panel A presents summary statistics for years 2019 to 2023, which is the sample we use for most of our analysis. Panel B presents summary statistics for the RD analysis.

Table C2: Regression Discontinuity Results for Tier II Universities

Panel A. Effect on Peer Quality				
	(1)	(2)	(3)	(4)
VARIABLES	average admission score		minimum admission score	
below	5.07***	7.52***	4.01***	6.03***
	(0.20)	(0.37)	(0.21)	(0.46)
Observations	85,749	85,749	85,749	85,749
R-squared	0.92	0.92	0.90	0.90
Panel B. Effect on University Location				
	(1)	(2)	(3)	(4)
VARIABLES	Non-IM universites		Capital/Coastal universities	
below	0.044***	0.034***	0.011***	0.016***
	(0.006)	(0.010)	(0.002)	(0.003)
Observations	85,749	85,749	85,749	85,749
R-squared	0.171	0.171	0.040	0.040
Panel C. Effect on University Postgraduation Outcomes				
	(1)	(2)	(3)	(4)
VARIABLES	% grad school attendance		ln(average wage)	
below	0.009***	0.015***	0.012***	0.020***
	(0.002)	(0.005)	(0.003)	(0.007)
Observations	10,177	10,177	5,249	5,249
R-squared	0.161	0.161	0.167	0.169
Bandwidth	5	5	5	5
Fitting	Linear	Quadratic	Linear	Quadratic
Year-track-cutoff FE	Yes	Yes	Yes	Yes

Notes: This table presents the OLS results from equation (1). The sample is first restricted to students within Tier II university score ranges, then further restricted to students whose scores are within five points from the batching cutoffs. Below is a binary variable indicating whether the student is right below a batching score cutoff. Uniform Kernel is used for all regressions. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.