Gaming the Boston School Choice Mechanism in Beijing

YINGHUA HE

April 2012

(First Version: November, 2009)

(The up-to-date version is available at: http://sites.google.com/site/yinghuahe/)

1 Toulouse School of Economics. Email address: yinghua.he@gmail.com. I am deeply indebted to Fang Lai for her generous help that makes this project possible. For their advice, constant encouragement and support, I am grateful to Kate Ho, W. Bentley MacLeod, Bernard Salanié, Miguel Urquiola, and Eric Verhoogen. For their helpful comments and suggestions, I thank Atïla Abdulkadiroğlu, Jushan Bai, Yeon-Koo Che, Pierre-André Chiappori, Navin Kartik, Wojciech Kopczuk, Dennis Kristensen, Greg Lewis, Sanxi Li, Hong Luo, Thierry Magnac, Serena Ng, Brendan O’Flaherty, Parag Pathak, Michael Riordan, Jonah Rockoff, Johannes Schmieder, Herdis Steingrimsdottir, Priscilla K. Yen, and seminar participants at Autonoma de Barcelona, Columbia, Copenhagen Business School, Nanyang, Paris School of Economics, Toulouse, Western Ontario, Wisconsin-Madison, WUSTL, WZB, the 2009 Econometric Society NASM in Boston, 2010 NBER Market Design Meeting, 2010 China Economics Summer Institute, and 2010 Econometric Society World Congress in Shanghai. Financial supports from Columbia’s Wueller and Vickrey research prizes and the Program for Economic Research are gratefully acknowledged.
Abstract

The Boston mechanism is criticized for its poor incentive and welfare performances compared to the Gale-Shapley deferred acceptance mechanism (DA). Using school choice data from Beijing, I investigate parents’ behavior under the Boston mechanism, taking into account parents’ possible mistakes when they strategize. Evidence shows that parents are overcautious as they play "safe" strategies too often. Wealthier/more educated parents are less overcautious and perform slightly better because they have better outside options while being equally or even less sophisticated. Counterfactual analyses show that switching to the DA at best helps 44% parents who do not strategize but also hurts 35% of them.

Keywords: the Boston Mechanism, the Gale-Shapley Deferred-Acceptance Mechanism, School Choice, Bayesian Nash Equilibrium, Strategy-Proofness, Simulated Maximum Likelihood
To allocate students to public schools, a centralized assignment mechanism is often necessary as monetary transfers are usually precluded. Despite the increasing popularity of school choice programs, the question about which assignment mechanism should be used has recently been debated among researchers and policy makers.

At the center of the debate is the Boston mechanism which is one of the most popular mechanisms in practice. It was used by the Boston Public Schools (BPS) during 1989-2005 and then was abolished in 2005.\footnote{There are many schools districts that still use the mechanism, for example Cambridge, MA, Charlotte-Mecklenburg, NC, St. Petersburg, FL, Minneapolis, MN, Providence, RI. The mechanism is also popular in other countries and in other contexts, for example, China’s college admissions.} The main criticism on the mechanism is that it encourages parents to “game the system.” Namely, parents may have incentives to misreport their preferences when submitting rank-order lists of schools (Abdulkadiroglu and Sonmez (2003)). On the other side, schools also form a strict priority ordering of students usually with lotteries as tie-breakers. Each school first considers students who rank it first, and assigns seats in order of their priority at that school. Then, each school that still has available seats considers unmatched students who rank it second. This process continues until the market is cleared. If a student ranks a popular school first and gets rejected, her chance of getting her second choice is greatly diminished because she can only be accepted after everyone who lists that school as their first choice.

Since the mechanism is not strategy-proof, the ability to strategize, or the level of sophistication, might affect parents/students welfare. Experimental and empirical evidence in previous literature suggests that parents strategize at different levels (e.g., Abdulkadiroglu, Pathak, Roth, and Sonmez (2006), Chen and Sonmez (2006), Lai, Sadoulet, and de Janvry (2009), and Pais and Pinter (2008)). Theoretically, Pathak and Sonmez (2008) considers two types of parents: sincere (or naive) parents who always reveal their preferences truthfully and sophisticated parents who always play a best response against others. They show that the mechanism may give an advantage to sophisticated parents.

\footnote{In real life, this is well known to some parents. For instance, the West Zone Parents Group in Boston, recommends two types of strategies to its members in 2003: “One school choice strategy is to find a school you like that is undersubscribed and put it as a top choice, OR, find a school that you like that is popular and put it as a first choice and find a school that is less popular for a ‘safe’ second choice.”}
These results were instrumental to the decision of BPS. The Boston School Committee voted in 2005 to replace the Boston mechanism with the student-proposing Deferred-Acceptance mechanism (henceforth, DA; Gale and Shapley (1962)), which is strategy-proof – reporting true preferences is always a weakly dominant strategy (Dubins and Freedman (1981); Roth (1982)). A description of the reform can be found in Abdulkadiroglu, Pathak, Roth, and Sonmez (2005).

One of the key arguments for the reform is that the Boston mechanism might hurt less sophisticated parents, while the DA provides protection. For example, the BPS Strategic Planning Team claimed in 2005 that "a strategy-proof algorithm ‘levels the playing field’ by diminishing the harm done to parents who do not strategize or do not strategize well." More importantly, policy makers are worried that poor and/or less educated parents are less sophisticated. Therefore, under the Boston mechanism, "the need to strategize provides an advantage to families who have the time, resources and knowledge to conduct the necessary research," as stated by Thomas Payzant then BPS Superintendent (Payzant (2005)).

However, the literature has not reached a consensus about these arguments. There is no evidence relating parents’ sophistication level to family background, and there are mixed, mainly theoretical or experimental, results on naive parents’ welfare. A recent strand of literature provides results in favor of the Boston mechanism (e.g. Featherstone and Niederle (2008), Miralles (2008)). In particular, Abdulkadiroglu, Che, and Yasuda (2011) show that even some naive parents can be better off in the Boston mechanism.

Using field data from Beijing where 914 students apply to four middle schools under a version of the Boston mechanism, this paper fills the gap by answering two questions: (i) whether poor/less educated parents are more likely to be naive and (ii) whether the Boston harms naive parents relative to the DA.

In Beijing, the middle schools use a random lottery to rank students without any predetermined priorities. To evaluate parents’ welfare, this paper adopts the concepts of Bayesian Nash equilibrium and \textit{ex ante} efficiency\footnote{In terms of \textit{ex ante} efficiency, Zhou (1990) shows that it is impossible to have a strategy-proof and efficient mechanism that treats the same type of parents equally. Therefore, the DA mechanism is not \textit{ex ante} efficient.}. At the time of application, parents’
preferences are private information and the random lottery is unknown. They maximize their expected utility by choosing different rankings of schools under uncertainties from two sources: other parents’ behavior and the random lottery.

The data contain parents’ submitted lists of schools and family background but not the true preferences. The major challenge is to estimate true preferences when parents are not necessarily truth-telling. I assume a random utility model for parents’ preferences over schools with the stochastic part being normally distributed as in a multinomial probit model. Besides, parents’ potentially heterogeneous sophistication is explicitly considered.

Under the assumption that everyone understands the uncertainty from the lottery, a parent’s sophistication depends on her assessment of other parents’ behaviors which are determined by the joint distribution of their preferences and sophistication. A parent is sophisticated if she does so correctly and therefore her subjective beliefs – the perceived probabilities of being accepted by each school when submitting different choice lists – match what are implied by the true distribution. Less sophisticated parents have inaccurate beliefs, while naive ones disregard the uncertainty and are always truth-telling.

Although beliefs may be wrong, they satisfy some properties imposed by the mechanism, for example, moving a school upward in a list always (weakly) improves the probability of being accepted by that school. These properties lead to a set of dominated strategies, for instance, ranking an unacceptable school as first choice is dominated. Assuming that these dominated strategies are not played in equilibrium, I group certain lists together and, loosely speaking, make the necessary equilibrium conditions both necessary and sufficient with respect to the new choice set.

In previous literature, some papers assume complete information, for example Ergin and Sonmez (2006), Kojima (2008), and Pathak and Sonmez (2008). They focus on Nash equilibrium and ex post efficiency. Recently, the ex ante view becomes more common, for example, Abdulkadiroglu, Che and Yasuda (2008, Forthcoming), Featherstone and Niederle (2008), and Miralles (2008).

Hastings, Kane, and Staiger (2008) estimate the demand for schools under the assumption that students are truth-telling under the Boston mechanism. Their data is from Charlotte-Mecklenburg Public School District in 2002 when the Boston mechanism had just been implemented. The truth-telling assumption may be more likely to be valid in their setting than others. In the current study, I also estimate the model under the truth-telling assumption, and it is rejected when tested against the model with strategic behavior.
In practice, indeterminacy regarding parents’ behavior arises because (i) there are schools worse than the outside option and (ii) some probabilities in parents’ beliefs might be zero. I provide evidence that the indeterminacy exists in the data and propose solutions while maintaining the model’s point identification. A method of simulated maximum likelihood, similar to that in a multinomial probit model, is used for estimation.

Results reject the hypothesis that everyone is naive and also reject that everyone is sophisticated. Parents understand the rules well, but they are overcautious, as they avoid top ranking the school with best quality and smallest quota too often than what they would do if best responding. Their income and education offset the overcautiousness partially, because wealthier/more educated parents’ true preference order is more likely to be a best response. There is no evidence of them being more sophisticated. These findings are driven by the fact that wealthier/more educated parents have a better outside option.

The data has information on how much attention parents pay to uncertainty in the game. Poorer parents pay more attention, and it implies they try to find a best response. However, paying more attention does not mitigate and sometimes even worsen their overcautiousness.

To evaluate the effect of replacing the Boston with the DA, I simulate the outcomes under both mechanisms assuming the preferences do not change across mechanisms. Given parents’ behavior, especially their overcautiousness, in the data, both naive and sophisticated parents suffer a utility loss in the DA which on average roughly amounts to tripling the distance to a school. For naive parents, only 8% are better off in the DA, while 71.5% are worse off. The negative effects are larger for sophisticated parents, and they decrease with parents’ income and education because of the outside option.

If every parent is either sophisticated or naive and no one is overcautious, switching from the Boston to the DA has mixed effects. For sophisticated parents, their utility loss is on average 16.48 percentage point, equivalent to increasing the distance to a school by 91.29%. Among them, only 11% are better off, while about 70% are worse off.

---

6This overcautiousness is related to the "small school bias" in experimental studies (Chen and Sonmez (2006), Calsamiglia, Haeringer, and Klijn (2010)). Namely, schools with fewer slots are ranked at lower positions than those in the true preference. However, overcautiousness compares observed behaviors with best responses.
parents on average have a utility gain in the DA, although the gain is rather small and only amounts to decreasing the distance to a school by 6.74%. The DA helps about 44% of naive parents but hurts about 35% of them as well.

**Other Related Literature**

There is a growing literature using real life data to study school choice or assignment problems. For example, Budish and Cantillon (Forthcoming) compare a strategy-proof mechanism with a non-strategy-proof one using a data set on MBA students’ course-allocation problem. Braun, Dwenger, and Kubler (2010) study the strategic behavior in the centralized university admissions in Germany, and Carvalho and Magnac (2009) investigate the college admission with exams in Brazil.

This study relates to the literature on testing whether an equilibrium is played in real life games. For example, Chiappori, Levitt, and Groseclose (2002) and Kovash and Levitt (2009) study professional sports, and Hortacsu and Puller (2008) examine the strategic bidding in an electricity spot market auction. Hortacsu and Puller characterize a Bayesian-Nash equilibrium model and compare actual bidding behavior to theoretical benchmarks. The difficulty in estimating a Bayesian-Nash equilibrium lies in specifying the beliefs. Under some technical assumptions, they show the best response is also ex post optimal, i.e., seeing other players’ behavior would not change one’s behavior. Thus, they can just look at the ex post optimality without evaluating the beliefs. In contrast, the current study allows players to mistake and derives identification independent of beliefs.

Another related strand of literature is the estimation of simultaneous games of incomplete information. Most studies need the condition of consistent beliefs to derive moment conditions or choice probabilities, e.g., Seim (2006), Bajari, Hong, Krainer, and Nekipelov (2010), Aradillas-Lopez (2007a), and Aradillas-Lopez (2007b). Given the small number of players, identification requires multiple game plays and equilibrium beliefs which are correct and stable across game plays. I relax these assumptions and allow inaccurate beliefs.

In the following, Section 2 describes the two school choice mechanisms and the data from Beijing. Section 3 formalizes the school choice problem under the Boston mech-
anism as a Bayesian game. Restrictions on parents’ behavior are derived under various assumptions, and I also characterize choice probabilities and propose a method of simulated log-likelihood for estimation. Section 4 presents reduced-from results, while Section 5 shows the model estimation. In particular, I present the correlation between sophistication and family background in both sections. Section 6 shows the counterfactual analysis for replacing the Boston with the DA mechanism. The paper concludes in Section 7.

1 The Two Mechanisms, Background, and Data

1.1 Deferred-Acceptance Mechanism

The DA mechanism works as follows:

(i) Each school forms a strict priority ordering of students which is determined by state or local laws. For example, as in Boston, it depends on sibling’s enrollment, distance to schools, and a lottery.

(ii) Schools announce enrollment quota and students submit a rank-order list of schools.

(iii) With priority orderings and submitted lists, the matching process has several rounds:

Round 1. Every student applies to her first choice. Each school rejects the lowest-priority students in excess of its capacity and temporarily holds the other students.

Generally, in

Round k. Every student who is rejected in Round \((k - 1)\) applies to the next choice on her list. Each school pools new applicants and those who are held from Round \((k - 1)\) together and rejects the lowest-priority students in excess of its capacity. Those who are not rejected are temporarily held by the schools.

The process terminates after any Round \(k\) when no rejections are issued. Each school is then matched with students it is currently holding.

If schools’ priority is determined by the same factor, for example, test scores or one single lottery, then schools rank students in the same way. In this case, the DA is equivalent to the serial dictatorship mechanism (Abdulkadiroglu and Sonmez (1998)). Following their
priority order, students sequentially choose their most preferred among schools which still have available seats.

1.2 Boston Mechanism

Similarly, the Boston mechanism asks students to submit rank-order lists, uses pre-defined rules to determine schools’ ranking over students, and has multiple rounds:

Round 1. Each school considers all the students who rank it first and assigns seats in order of their priority at that school until either there is no seat left or no such student left.

Generally, in

Round $k$. The $k$th choice of the students who have not yet been assigned is considered. Each school that still has available seats assigns the remaining seats to students who rank it as $k$th choice in order of their priority at that school until either there is no seat left or no such student left.

The process terminates after any round $k$ when every student is assigned a seat at a school, or if the only students who remain unassigned listed no more than $k$ choices. Unassigned students are then matched with available seats randomly.

1.3 Boston Mechanism in Beijing

I study school choice in the largest neighborhood of Beijing’s Eastern City District in 1999. The assignment mechanism allowed students to apply to four middle schools with a total quota of 960 which was determined by the Education Bureau. To be included in this neighborhood, a student must be enrolled as a 6th grader in one of the seven given elementary schools in 1999. A more detailed description of the data and background is available in Lai, Sadoulet, and de Janvry (2009).

The neighborhood adopted a version of the Boston mechanism for middle school admission in which schools’ priority ranking over students was solely determined by a random lottery (single tie-breaker). Students could submit a choice list ranking up to four schools.
After the submission, a computer-generated 10-digit number was randomly assigned to each student, and then the admission proceeded exactly as described above.

Students’ outside option was mainly the 28 public schools in the district, including the four to which they could apply through the mechanism. They might attend a public school without going through the mechanism in three ways. (i) Schools admit students directly if their parents are employed by the school, if they have received at least a city-level prize in academic or special skill achievements, or if a considerable payment is made to the school. (ii) Besides the quota announced, some top schools admit more students by offering an admission exam. (iii) Schools admit some transfer students who are not satisfied with their assignment and make a payment to the accepting school.

Other possible outside options were not very relevant at that time. Specifically, private schools were not well developed in 1999. Besides, there was no strong incentive for students to transfer out of the district, because the Eastern City District had both an advantageous location and a very good reputation in educational quality among all districts in Beijing. Such transfers were only possible when there was a formal relocation of parents or the out-of-district accepting school received a payment which was usually higher than what was required by the within-district schools.

1.4 Data

There are two data sources: one being the administrative data which provides submitted lists, elementary school enrollment, grade 6 test scores, and home addresses in 1999; and the other being a district-level survey in early 2002.

---

7 The same mechanism was used in all Beijing’s neighborhoods in 1999, including those in other districts.
8 The amount of this kind of payments, or "ze xiao fei" (literally, "fees for choosing a school"), may depend on the student’s ability and parents’ connection. Unfortunately, information on these payments is not publicly available. Since 2008, the education authority of Beijing regulates that such fees cannot be more than 30,000 yuan (Source: http://www.bjedu.gov.cn/publish/portal0/tab67/info11554.htm). It is slightly higher than the average disposable income among the urban residents in Beijing in 2008, or 24,725 yuan. The out-of-pocket cost for parents may easily exceed this limit. For example, a blog post claims that some people paid 250,000 yuan to get into a very good elementary school in 2011 (Source: http://blog.sina.com.cn/s/blog_6ce3959f0102dr2x.html).
The middle schools in China provide three years of education – grades 7-9, so the survey covered all students enrolled in the last year of middle schools in the district, as well as their parents. Dropping out or repeating grades was negligible in these schools; inter-district transfers were extremely rare as discussed above. Hence, the survey population is close to the population of students who entered middle schools in 1999. A questionnaire directed to parents collected information on their educational attainment and income and retrospective information on their preparedness for making school choice decisions in 1999.

1.4.1 Heterogeneity among the Middle Schools

The four "inside" middle schools are highly differentiated on two dimensions: enrollment quota and quality. Table 1 shows that School 1 has the smallest quota, 63 seats, although its size is not small because it also enrolls students from other neighborhoods. School 4 has the largest quota, 360 seats.

<table>
<thead>
<tr>
<th>Schools</th>
<th>Quota</th>
<th>( School _Score_s ): Average Test Score</th>
<th>Ranking in the district</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>63</td>
<td>559.27</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>227</td>
<td>522.91</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>310</td>
<td>508.47</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>360</td>
<td>470.13</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td>960</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Average test score of the graduating class in high school entrance exam in 1999, out of 600.
b. Ranking based on average test score among all 28 public schools in the district.

School 1 also has the best quality measured by the performance of the school’s graduating class in the high school entrance exam in 1999. The exam is city-wide and high-stake, and thus it is an important factor that parents use to determine school quality. As column 3 shows, these schools spread out in the distribution of quality among all 28 schools in the district, while better schools have smaller quotas.
1.4.2 Students’ Characteristics and Behavior: A First Look

Using their elementary school enrollment, I identify 914 students as in this neighborhood in 1999. The 46 "missing" students, i.e., the difference between the total quota (960) and the number of observed students (914), may have come from three sources: (i) enrollment quota is usually larger than the number of students; (ii) some may have skipped the mechanism and gone to schools outside the district in 1999; and (iii) some may have transferred to schools outside the district after 1999.

The last two sources, especially (iii), may create sample selection issue because it is possible that they felt unsatisfied with the expected or realized school assignment and made that decision. However, as discussed above, this issue is plausibly negligible, although unfortunately there is no way to verify. In the following, I focus on the observed 914 students.

The distribution of submitted lists in 1999 and middle school enrolment in 2002 are summarized in Table 2. About 20% did not submit any list and took their outside option directly, while among them 60.77% were still enrolled in one of the four schools in 2002. The majority submitted a full list with three or four schools, and only 7.44% submitted a partial list. Overall, only 10.07% of the students were enrolled in a school other than the four schools. Given that School 1 was the best in the district, its total enrolment among the 914 students is also very high – 147 students which is more than two times of the quota. For any of the other three schools, the total enrollment is lower than its announced quota.

For the variables which are used in the estimation, Table 3 presents their definition. I mainly focus on students’ family background (Parent_Inci, Parent_Edui), their ability (Own_Scorei, Awardsi) and gender (Girli), and distance to each school (Distanceisi).

Table 4 further summarizes these variables. In the estimation, most of the variables are taken logarithm and de-meaned. With the raw data, I present summary statistics for the

---

9 As a robustness check, I impute 46 additional students by drawing from the observed 914 students to "complete" the data. Based on the experiments that I’ve done, the results are not sensitive to the imputation as long as students are not imputed in an extreme way.

10 This is another piece of indirect evidence that transferring to another district is rare, as it is already rare for students to choose a within-district school other than the four "inside" ones.
Table 2: Distribution of Submitted Lists in 1999 and Middle School Enrollment in 2002

<table>
<thead>
<tr>
<th>Submitted Lists in 1999</th>
<th>Middle School Enrollment in 2002</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq.</td>
</tr>
<tr>
<td>Not Participating</td>
<td>181</td>
</tr>
<tr>
<td>Full Lists</td>
<td>665</td>
</tr>
<tr>
<td>4 Schools</td>
<td>558</td>
</tr>
<tr>
<td>3 Schools</td>
<td>107</td>
</tr>
<tr>
<td>Partial Lists</td>
<td>68</td>
</tr>
<tr>
<td>2 Schools</td>
<td>58</td>
</tr>
<tr>
<td>1 School</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>914</td>
</tr>
</tbody>
</table>

a. "Other" means one of the other 24 public middle schools in the district.

Table 3: Definition of Main Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Definition</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent_Inc_i</td>
<td>Parents’ income yuan/month in 2002</td>
<td>Survey in 2002</td>
</tr>
<tr>
<td>Parent_Edu_i</td>
<td>Parents’ average years of education</td>
<td>Survey in 2002</td>
</tr>
<tr>
<td>Girl_i</td>
<td>=1 if student i is a girl</td>
<td>Survey in 2002</td>
</tr>
<tr>
<td>Own_Score_i</td>
<td>Elementary Chinese + math, out of 200.</td>
<td>Administrative data</td>
</tr>
<tr>
<td>Awards_i</td>
<td>District level awards in elementary school</td>
<td>Survey in 2002</td>
</tr>
<tr>
<td>Distance_{i,s}</td>
<td>Walking distance to school s in 1999, km</td>
<td>Administrative data</td>
</tr>
</tbody>
</table>

full sample and 3 subsamples – non-participants, participants submitting partial lists, and those submitting full lists. Non-participants have richer and more educated parents than average, and they have higher test scores and have earned more awards. This is consistent with the above discussion that parents’ income and students’ ability increase the quality of their outside option. The same pattern for parents is observed for participants submitting partial lists, while they have lower test scores and have earned fewer awards than average.

2 Model: Boston School Choice as A Bayesian Game

In this section, the school choice problem under the Boston mechanism, as students in Beijing face, is formalized as a Bayesian game. There are:

(i) a set of students/parents, \{i\}_{i=1}^{I};
(ii) a set of schools, \{s\}_{s=0}^{S}, S \geq 3, where School 0 is the outside option.\[^{11}\]

\[^{11}\] When the number of schools is less than 3, truth-telling is a dominant strategy.
### Table 4: Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>Transformation</th>
<th>Full Sample</th>
<th>Non-Participant</th>
<th>Partial List</th>
<th>Full List</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Used in Estimation</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Raw Data</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td>Parent Inc&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Log, de-mean&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.00</td>
<td>3664.01</td>
<td>4249.07</td>
<td>4191.12</td>
</tr>
<tr>
<td></td>
<td>(0.82)</td>
<td>(3468.85)</td>
<td>(2457.23)</td>
<td>(2069.04)</td>
<td>(3782.76)</td>
</tr>
<tr>
<td>Parent Edu&lt;sub&gt;i&lt;/sub&gt;</td>
<td>De-mean</td>
<td>0.00</td>
<td>13.44</td>
<td>14.28</td>
<td>14.19</td>
</tr>
<tr>
<td></td>
<td>(2.24)</td>
<td>(2.24)</td>
<td>(2.57)</td>
<td>(2.13)</td>
<td>(2.07)</td>
</tr>
<tr>
<td>Girl&lt;sub&gt;i&lt;/sub&gt;</td>
<td>None</td>
<td>0.52</td>
<td>0.52</td>
<td>0.51</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(0.50)</td>
<td>(0.50)</td>
<td>(0.50)</td>
<td>(0.50)</td>
</tr>
<tr>
<td>Own Score&lt;sub&gt;i&lt;/sub&gt;</td>
<td>Log, de-mean</td>
<td>0.00</td>
<td>183.56</td>
<td>187.09</td>
<td>178.41</td>
</tr>
<tr>
<td></td>
<td>(0.08)</td>
<td>(11.66)</td>
<td>(7.84)</td>
<td>(15.55)</td>
<td>(11.83)</td>
</tr>
<tr>
<td>Awards&lt;sub&gt;i&lt;/sub&gt;</td>
<td>De-mean</td>
<td>0.00</td>
<td>0.73</td>
<td>1.12</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>(1.00)</td>
<td>(1.00)</td>
<td>(1.29)</td>
<td>(0.82)</td>
<td>(0.90)</td>
</tr>
<tr>
<td>Distance&lt;sub&gt;i,1&lt;/sub&gt;</td>
<td>Log, de-mean&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.02</td>
<td>2.31</td>
<td>2.64</td>
<td>1.94</td>
</tr>
<tr>
<td></td>
<td>(0.76)</td>
<td>(2.27)</td>
<td>(2.31)</td>
<td>(1.25)</td>
<td>(2.32)</td>
</tr>
<tr>
<td>Distance&lt;sub&gt;i,2&lt;/sub&gt;</td>
<td>Log, de-mean&lt;sup&gt;b&lt;/sup&gt;</td>
<td>-0.12</td>
<td>2.22</td>
<td>2.55</td>
<td>1.89</td>
</tr>
<tr>
<td></td>
<td>(0.86)</td>
<td>(2.29)</td>
<td>(2.31)</td>
<td>(1.41)</td>
<td>(2.35)</td>
</tr>
<tr>
<td>Distance&lt;sub&gt;i,3&lt;/sub&gt;</td>
<td>Log, de-mean&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.13</td>
<td>2.51</td>
<td>2.95</td>
<td>2.21</td>
</tr>
<tr>
<td></td>
<td>(0.66)</td>
<td>(2.21)</td>
<td>(2.17)</td>
<td>(1.34)</td>
<td>(2.28)</td>
</tr>
<tr>
<td>Distance&lt;sub&gt;i,4&lt;/sub&gt;</td>
<td>Log, de-mean&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.02</td>
<td>2.41</td>
<td>2.84</td>
<td>1.84</td>
</tr>
<tr>
<td></td>
<td>(0.82)</td>
<td>(2.20)</td>
<td>(2.14)</td>
<td>(1.43)</td>
<td>(2.28)</td>
</tr>
<tr>
<td># of Obs.</td>
<td>914</td>
<td>914</td>
<td>181</td>
<td>665</td>
<td>68</td>
</tr>
</tbody>
</table>

Standard deviations in parentheses.  

<sup>a</sup> More precisely, it is "log(Parent_Inc<sub>i</sub> + 1), de-mean";  

<sup>b</sup> The mean here is the one of all 4 distances.

(iii) a capacity vector, \( \{q_s\}_{s=1}^S \);  
\( \sum_{s=1}^S q_s \geq I, \sum_{s=1}^S q_s - q_{s'} < I, \) and \( q_{s'} < I, \forall s' \neq 0. \)

(iv) students’ choice lists, \( \{C_i\}_{i=1}^I \), where \( C_i = (c_i^1, ..., c_i^S), c_i^k \in \{s\}_{s=0}^S, \forall k = 1, ..., S; \)

(v) schools’ priorities over students which are determined solely by a random lottery.

At the start of the game, each school announces its capacity, \( q_s \). There are enough seats to accommodate all the students, i.e. \( \sum_{s=1}^S q_s \geq I, \) and no school has enough seats to enroll all students, \( q_{s'} < I, \forall s' \), while every school is significantly big in the sense that not all students can be accommodated by the other \( S - 1 \) schools, \( \sum_{s=1}^S q_s - q_{s'} < I. \)

Parents then submit their choice lists, \( C_i = (c_i^1, ..., c_i^S) \) where \( c_i^k \) is the \( k \)th choice. \( C_i \) is a full list if it includes all \( S \) schools; otherwise, it is a partial list. Parents are allowed to submit partial lists or not submit anything at all. In the latter case, the student is not considered in the mechanism.

After collecting \( \{C_i\}_{i=1}^I \), the mechanism assigns each student a random number which determine her priority at all schools. Therefore, all students have the same \textit{ex ante} priority,
although pre-determined priorities can be considered as well. With the lists and the random lottery, the admission process goes as described in the previous section. After students get their assignments, they can choose the outside option if they are not satisfied.

In the following, "student" and "parent" are used interchangeably. I first present the set-up and the benchmark case where everyone is (equally) sophisticated with a common prior. I then extend the model to the case where parents have heterogeneous sophistications. The definition of sophistication is formalized along the discussion.

2.1 Set-Up

The utility of student $i$ attending $s$ ($s = 1, \ldots, S$) is

$$u_{i,s} = \alpha_s + X_i \beta_X + Z_{i,s} \beta_Z + \varepsilon_{i,s}$$

where $\alpha_s$ is the school $s$’s fixed effect; $X_i \in \mathbb{R}^{K_1}$ are $i$’s characteristics such as $i$’s test score and parents’ income and education, etc.; $Z_{i,s} \in \mathbb{R}^{K_2}$ are student-school specific attributes, e.g., the distance from $i$’s home to $s$ or $i$’s characteristics interacted with $s$’s attributes, and $Z_i \equiv \{Z_{i,s}\}_{s=1}^{S}; \varepsilon_{i,s} \in \mathbb{R}$ includes all other factors, and $\varepsilon_i \equiv \{\varepsilon_{i,s}\}_{s=1}^{S}$.

The utility when choosing the outside option is normalized to be zero. Equivalently, the utility of attending any school $s \neq 0$ should be interpreted as the utility difference between attending $s$ and choosing the outside option. If a school is worse than her outside option, $u_{i,s} < 0$, it is unacceptable. Otherwise, it is acceptable.

The following assumptions are maintained throughout the paper:

**AM.1.** Parents are expected utility maximizers who know their own true preferences, $\{u_{i,s}\}_{s=0}^{S}$, and the function of $u_{i,s}$ and its parameters.

**AM.2.** $(X_i, Z_i)$ are i.i.d. over $i$ with C.D.F. $G(X, Z)$ which is common knowledge, while $(X_i, Z_i)$ being private information of $i$.

**AM.3.** $\varepsilon_i \perp (X_i, Z_i)$ and $\varepsilon_i \sim N(0, \Sigma)$ and i.i.d. over $i$, with C.D.F. $F_\Sigma(\varepsilon_i)$ and $Var(\varepsilon_{i,1}) = 1$. $\varepsilon_i$ is private information of $i$, and its distribution is common knowledge.
AM.4. If no school is acceptable, the parent does not participate and submits $(0, \ldots, 0)$.

The assumption of $(X_i, Z_i)$ being private information in AM.2. is for the ease of exposition. When $I$ is large, similar results hold true if $(X_i, Z_i)$ is common knowledge, or if $i$ knows a fixed number of other parents’ $(X_j, Z_j)$. Appendix A.3 provides a detailed discussion.

AM.3 allows an arbitrary correlation between any $\varepsilon_{i,s}$ and $\varepsilon_{i,s'}$. For example, some schools are better at sciences, while others are better at arts. Students who like sciences more than arts have positive shocks for some of the schools and negative ones for the others.

2.2 Benchmark: homogeneous Sophistication

In the following, everyone is (equally) sophisticated and is endowed with a common prior. Namely, they have the same information and correctly use the information in the same way.

2.2.1 Strategy, Payoff, and Decision Making

A strategy $\sigma_i (X_i, Z_i, \varepsilon_i)$, possibly in mixed strategies, is a mapping from $i$’s "type" space to the set of all probability distributions over possible lists: $\mathbb{R}^{K_1 + SK_2 + S} \to \Delta (C)$. The total number of pure strategies or possible lists in $C$ is finite, $L \equiv S! \left( \frac{1}{S!} + \frac{1}{(S-1)!} + \cdots + \frac{1}{1!} \right)$.

Each element in $C$ is a rank-order list of $k$ different schools, where $k = 0, \ldots, S$.

The payoff to $i$ can be characterized in two steps: (i) other parents actions, $C_{-i}$, are given; and (ii) instead of $C_{-i}$, other parents strategies, $\sigma_{-i}$, are given.

Given $C_{-i}$ and if $\sigma_i = C$ is a pure strategy, the expected payoff to $i$ is:

$$V_i (C; C_{-i}) \equiv \sum_{s=1}^{S} \left[ a_s (C, C_{-i}) \max (u_{i,s}, 0) \right],$$

where only $\max (u_{i,s}, 0)$ matters because parents choose the outside option whenever the assigned one is unacceptable; and $a_s (C, C_{-i})$ is the probability of student $i$ being accepted.

\[12\] Notice that those lists in which one school appears multiple times are excluded, as they are obviously not optimal. Other obviously dominated lists are also not considered.
by \( s \) given \((C, C_{-i})\). It is completely determined by the random lottery, and the following lemma summarizes its properties.

**Lemma 1** Given any \( C \) and \( C_{-i} \), \( a_s(C, C_{-i}) \) has the following properties:

(i) A seat is guaranteed if participating: \( \forall C \neq (0, \ldots, 0), \sum_{s=1}^{S} a_s(C, C_{-i}) = 1 \);

(ii) In any two lists, if a school is listed after a same ordering of schools, the probability of being accepted by that school is the same when submitting either of the two lists:
\[
a_s(C, C_{-i}) = a_s(C', C_{-i}), \forall C, C' \in C, \text{ s.t. } c_K = c'_K = s \text{ and } c_k = c'_k, \forall k \leq K \leq S.
\]

(iii) Moving up a school (or including an otherwise omitted one) in the list (weakly) increases the probability of being accepted by that school: \( a_s(C', C_{-i}) \geq a_s(C, C_{-i}), \forall C, C' \in C, \text{ s.t. } c_K = c'_K = s, K' < K \leq S, \text{ and } c_k = c'_k, \forall k < K' \).

(iv) If school \( s \) is top ranked, the probability of being accepted by that school is at least \( q_s/I \):
\[
a_s(C, C_{-i}) \geq q_s/I, \forall C \in C, \text{ s.t. } c^1 = s.
\]

Proofs are collected in Appendix [A.1], while these properties can be easily verified given the mechanism. Similarly, since \( \sigma_i \) is a probability distribution over pure strategies, \( a_s(\sigma_i, C_{-i}) \) has the same properties and \( V_i(\sigma_i, C_{-i}) \) can be defined in the same way.

Now suppose that \((\sigma_i, \sigma_{-i})\) is given. \( i \)'s expected payoff is defined as:

\[
V_i(\sigma_i, \sigma_{-i}) \equiv \sum_{n=1}^{L(I-1)} \left\{ \Pr(C_{-i}^n \text{ played under } \sigma_{-i}) V_i(\sigma_i, C_{-i}^n) \right\}
\]

\[
= \sum_{s=1}^{S} \sum_{n=1}^{L(I-1)} \left[ \Pr(C_{-i}^n \text{ played under } \sigma_{-i}) a_s(\sigma_i, C_{-i}^n) \right] \max(u_{i,s}, 0),
\]

where the probability that other parents choose \( C_{-i}^n \), \( \Pr(C_{-i}^n \text{ played under } \sigma_{-i}) \), equals:

\[
\int \Pr(C_{-i}^n \text{ played under } \sigma_{-i}(X_{-i}, Z_{-i}, \varepsilon_{-i})) dG(X_{-i}, Z_{-i}) dF_\Sigma(\varepsilon_{-i})
\]

Given others play \( \sigma_{-i} \), \( i \)'s probability of being accepted by \( s \) when playing \( \sigma_i \) can be written.
as:

\[ A_s(\sigma_i, \sigma_{-i}) \equiv \sum_{n=1}^{L(i-1)} \Pr(C^n_{-i} \text{ played under } \sigma_{-i}) a_s(\sigma_i, C^n_{-i}), \]

which may be individual-specific because \( \Pr(C^n_{-i} \text{ played under } \sigma_{-i}) \) might differ across \( i \).

The expected payoff is simplified as

\[ V_i(\sigma_i, \sigma_{-i}) = \sum_{s=1}^{S} A_s(\sigma_i, \sigma_{-i}) \max(u_{i,s}, 0). \]

Furthermore, denote

\[ B(\sigma_i, \sigma_{-i}) \equiv (A_1(\sigma_i, \sigma_{-i}), \ldots, A_s(\sigma_i, \sigma_{-i})): \Delta(C) \rightarrow [0, 1]^S \]

as \( i \)'s beliefs. By definition, \( A_s(\sigma_i, \sigma_{-i}) \) is a probability weighted sum of \( a_s(C, C_{-i}), \forall C \) and \( C_{-i} \). Therefore, it is straightforward to verify that the properties of \( a_s(C, C_{-i}) \) in Lemma 1 still hold for \( A_s(\sigma_i, \sigma_{-i}) \). With the beliefs, one may define sophistication as follows.

**Definition 1** Given homogeneous sophistication, \( i \) is sophisticated if her beliefs are \( B(\sigma_i, \sigma_{-i}) \).

With her beliefs, parent \( i \) choose an strategy to maximize her expected utility:

\[ \sigma_i(X_i, Z_i, \epsilon_i) \in \arg \max_{\sigma_i \in \Delta(C)} \sum_{s=1}^{S} A_s(\sigma_i, \sigma_{-i}) \max(u_{i,s}, 0). \]

\( i \)'s optimal strategy may not be unique because (i) the operator \( \max(u_{i,s}, 0) \) creates multiple payoff-equivalent lists if some schools are unacceptable; and (ii) more payoff-equivalent lists are generated when \( A_s(\sigma_i, \sigma_{-i}) \) is zero for some \( s \). This indeterminacy brings a big challenge to empirical analysis, since it complicates characterization of choice probabilities.

### 2.2.2 Symmetric Bayesian Nash Equilibrium

To mitigate the indeterminacy problem, I consider a symmetric equilibrium in which all parents employ the same strategy, i.e., \( \sigma_i(X_i, Z_i, \epsilon_i) = \sigma_j(X_j, Z_j, \epsilon_j) \forall i \neq j \), if \( u_{i,s} = u_{j,s} \forall s \). Given that everyone is an expected-utility maximizer, the symmetry only requires that when there are multiple solutions to their problem (1), they use the same rule to choose one strategy, pure or mixed.
**Definition 2**  A mixed-strategy symmetric Bayesian Nash equilibrium in the Boston school choice game with homogeneous sophistication is a common strategy \( \sigma^* \in \Delta(C) \), s.t.,

\[
\sigma^*(X_i, Z_i, \varepsilon_i) \in \arg \max_{\sigma \in \Delta(C)} \sum_{s=1}^{S} A_s(\sigma, \sigma^*_{-i}) \max(u_{i,s}, 0), \text{ given } (X_i, Z_i, \varepsilon_i), \forall i;
\]

and there are common equilibrium beliefs, \( B^*(C, \sigma^*) \equiv B(C, \sigma^*_{-i}), \forall i \) and \( C \).

The existence and a characterization of such an equilibrium is in Proposition 1.

**Proposition 1**  There always exists a symmetric Bayesian Nash equilibrium in the Boston school choice game. In any symmetric equilibrium,

(i) equilibrium beliefs are such that \( A_s(C, \sigma^*_{-i}) \in (0, 1) \forall s, \forall C \neq (0, ..., 0) \);

(ii) if at most one school is unacceptable, \( i \) plays a pure strategy with probability one;

(iii) if \( i \) plays mixed strategies, with probability one she has at least two unacceptable schools and she only mixes lists in which the unacceptables are excluded or included after the acceptables.

This paper estimates a large game being played once, and thus multiplicity of equilibrium is not a concern for estimation, as there is only one equilibrium being played in the data. However, it matters when I do counterfactual analyses where I have to pick an equilibrium. In that case, I focus on the equilibrium that has been played in the data.

Proposition 1 leaves some indeterminacy: Parents may rank unacceptable schools unsystematically in equilibrium, as long as all unacceptables are ranked after the acceptables. This claim is formalized later in Proposition 2, and I make the following assumption.

**Assumption UNACCEPTABLES**  In a symmetric Bayesian Nash equilibrium, if some or all of the unacceptable schools are included in the list, they are ranked according to their true preference order among themselves after the acceptables. Moreover, the excluded unacceptables are always less desirable than included unacceptables.
It is not implausible that parents follow this strategy in real life to play safe. The symmetric equilibrium also requires that parents play the same mixed strategy. The following assumption further clarifies the possible pure strategies in equilibrium mixed strategies.

**Assumption MIXING** When playing mixed strategies, everyone follows the same mixing rule in equilibrium. Namely, if, without loss of generality, \( u_{i,1} > u_{i,2} > ... > u_{i,S} \), where \( K < (S - 2) \), \( i \) only submits the following lists with positive probabilities:

\[
(c_1^1, ..., c_K^1, 0, ..., 0) \text{ where all included schools are acceptable;}
(c_1^1, ..., c_K^1, (K + 1), 0, ..., 0);
(c_1^1, ..., c_K^1, (K + 1), (K + 2), 0, ..., 0); \text{ ... ; and (c}_1^1, ..., c_K^1, (K + 1), (K + 2), ..., S);
\]

while every of them is a best response. Let \( m_{K,l} \), \( l \geq K \), denote the probability that a \( l \)-school list is submitted while only \( K \) schools are acceptable. \( m_{K,l} \) is common to everyone and is independent of the identities of the acceptables.

Ideally, one should show the existence of a symmetric equilibrium under the two assumptions. Unfortunately, I have not found a proof or a disproof and thus leave it to future research.

### 2.2.3 Estimation

With Proposition 1 and Assumptions UNACCEPTABLES and MIXING, I characterize the probabilities that each list is played in equilibrium. I assume equilibrium beliefs, \( B^* \), are known and set. \( S \) equal to 4 in the following.

**Characterization of Choice Probabilities** Given \((X_i, Z_i, B^*; \theta)\) where \( \theta \) are the unknown parameters, the conditional probability of \( i \) choosing \( C_i \), \( \Pr(C_i|X_i, Z_i, B^*; \theta) \), is:

(i) if \( C_i = (0, 0, 0, 0) \), \( \Pr(u_{i,s} < 0, \text{ for all } s|X_i, Z_i, B^*; \theta) \);

(ii) if \( C_i = (c_1^1, 0, 0, 0) \), \( m_{1,1} \ast \Pr(u_{i,c_1} > 0 > u_{i,s}, \text{ for } s \neq c_1^1|X_i, Z_i, B^*; \theta) \);

(iii) if \( C_i = (c_1^1, c_2^1, 0, 0) \), \( m_{2,2} \ast \Pr(C_i \text{ is a best response}; u_{i,c_1}, u_{i,c_2} > 0 > u_{i,s}, \text{ for } s \neq c_1^1, c_2^1|X_i, Z_i, B^*; \theta) \)

\[ + m_{1,2} \ast \Pr(u_{i,c_1} > 0 > u_{i,c_2} > u_{i,s}, \text{ for } s \neq c_1^1 \neq c_2^1|X_i, Z_i, B^*; \theta) \);

(iv) if \( C_i = (c_1^1, c_2^1, c_3^1, c_4^1) \),
\[ \Pr (C_i \text{ is a best response}; u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > 0|X_i, Z_i, B^*; \theta) \]
\[ + m_{1,4} \times \Pr (u_{i,c^1} > 0 > u_{i,c^2} > u_{i,c^3} > u_{i,c^4}|X_i, Z_i, B^*; \theta) \]
\[ + m_{2,4} \times \Pr (C_i \text{ is a best response}; u_{i,c^1}, u_{i,c^2} > 0 > u_{i,c^3} > u_{i,c^4}|X_i, Z_i, B^*; \theta). \]

Part (i) says that the probability of not participating equals the probability that all schools are unacceptable. The probability of submitting a one-school list, in part (ii), is the mixing probability \( m_{1,1} \) times the probability that only one school is acceptable, because parents may submit a two-school or a full list \( (m_{1,2}, m_{1,4} \geq 0) \).

Part (iii) shows the likelihood of submitting a two-school list coming from two sources: (a) there are two schools acceptable, and (b) there is only one acceptable. In (a), students may submit either a two-school or a full list\(^{13}\) and thus the probability of optimally ranking two acceptable schools is weighted by the mixing probability, \( m_{2,2} \).

In (b), the first choice must be acceptable and the second is unacceptable. The omitted schools are unacceptable and worse than the second choice.

Similarly, in part (iv), parent submits a full list in three cases: (a) there are at least three schools acceptable, (b) there are two acceptable, and (c) there is only one acceptable. Again, the last two cases contribute to the likelihood because of the mixing assumption, while in case (a) there is no possibility of mixing.

The mixing probabilities are independent of \((X_i, Z_i, B^*)\) due to Assumption MIXING. Moreover, \(m_{1,1} + m_{1,2} + m_{1,4} = 1\) and \(m_{2,2} + m_{2,4} = 1\).

**Simulated Maximum Likelihood Estimation** Since the equilibrium beliefs, \( B^* \), are unknown, given the large number of players, I use the empirical beliefs, \( \hat{B} \), as an approxi-

\(^{13}\)It is also possible to submit a three-school list, but it is equivalent to submitting a four-school/full list.
The model is estimated by the following (simulated) maximum likelihood:

$$\max_{\theta} \sum_{i=1}^{I} \ln \left[ \Pr \left( C_i \mid X_i, Z_i, \hat{B}; \theta \right) \right]$$  \hspace{1cm} (2)

The choice probabilities are simulated by the smoothed logit-smoothed accept-reject simulator (Chapter 5, Train (2009)) which is described in Appendix A.5.

### 2.3 General Case: Heterogeneous Sophistication

Below I relax the sophistication assumption and consider that parents may make mistakes when forming their beliefs. This may be due to information differences and/or different abilities to process the information. Under the maintained assumptions, a particular structure on the beliefs and some dominated strategies are identified.

#### 2.3.1 Equilibrium Concept, Definition of Sophistication, and Dominated Strategies

To highlight heterogeneity, denote i’s belief as $B_i (C, \sigma_{-i}) \equiv \{ A_{i,s} (C, \sigma_{-i}) \}_{s=1}^{S} \in [0, 1]^S$, $\forall C$, where $A_{i,s} (C, \sigma_{-i}) \equiv \sum_{n=1}^{L(i-1)} \Pr_i (C_{n_i} \text{ played under } \sigma_{-i}) a_s (C, C_{n_i})$ and the probability measure $\Pr_i ()$ is i’s subjective assessment of an event’s likelihood.

By the above notation, the extent to which parents can make mistakes is limited. They may mistake when assessing others’ behavior and thus $\Pr_i ()$ is individual specific, while they know the rules of the game and thus know $a_s (C, C_{n_i})$ precisely.

---

Footnotes:

14 The distribution of equilibrium strategies is approximated by the 914 observations plus 9 choice lists which are not played by anyone in the data. 5,000 samples of random draws from the distribution are created. Each sample consists of 914 random draws from the 923 data points, with replacement. Fixing other parents’ submitted lists in each sample, I then calculate $\hat{B}^n$ for parent 1. Namely, parent 1 experiments the 24 (= 4!) full lists, and probabilities of being accepted by each school given any list are calculated by drawing 1,000 independent sets of lotteries and running the mechanism 1,000 times. It is sufficient to consider the full lists only, because either the beliefs associated with partial lists can be derived from those associated with the full lists, or the partial lists are dominated. After repeating this for the 5,000 samples, I calculate $\hat{B} = \sum_{n=1}^{5000} \hat{B}^n / 5000$. Note that $\hat{B}$ may have many elements equal to 0 or 1 given the observed data. This contradicts the results in Proposition 11 and as a remedy I perturb the system a little to make all elements in (0, 1) – the maximum absolute difference between the original and the perturbed is $7.282 \times 10^{-8}$.

15 In Appendix A.3, I also discuss an alternative approach which uses equilibrium constraints and solves $B^*$ as a function of $\hat{\theta}$, a fixed point.
Since $A_{i,s}(\sigma_i, \sigma_{-i})$ is a probability weighted average of $a_s(C, C_{-i})$, the properties of $a_s(C, C_{-i})$ still hold for $A_{i,s}(\sigma_i, \sigma_{-i})$, and the proof for the following lemma is omitted.

**Lemma 2** Given $\sigma_{-i}$, $A_{i,s}(\sigma_i, \sigma_{-i})$ has the same properties as $a_s(C, C_{-i})$ in Lemma 1.

I rewrite $i$'s strategy as an explicit correspondence of beliefs, $\sigma_i([X_i, Z_i, \varepsilon_i; B_i(\cdot, \sigma_{-i})])$ and again consider a symmetric equilibrium.

**Definition 3** A mixed-strategy symmetric Bayesian Nash equilibrium in the Boston school choice game with heterogeneous sophistication is a common strategy $\sigma^* \in \Delta(C)$ s.t.,

$$
\sigma^*[X_i, Z_i, \varepsilon_i; B_i(\cdot, \sigma^*_{-i})] = \arg \max_{\sigma \in \Delta(C)} \sum_{s=1}^{S} A_{i,s}(\sigma, \sigma_{-i}^*) \max(u_{i,s}, 0), \text{ given } (X_i, Z_i, \varepsilon_i), \forall i.
$$

In this definition, the only restriction is that everyone is a subjective expected utility maximizer, while there is no requirement on subjective beliefs. The existence of such an equilibrium is thus guaranteed. This definition also provides a way to measure sophistication in terms of how correct one’s prediction of the game play is.\(^{16}\)

**Definition 4** With heterogeneous sophistication, in equilibrium $\sigma^*$, $i$ is sophisticated if

$$
A_{i,s}(C, \sigma^*_{-i}) = \sum_{n=1}^{L(i-1)} \Pr(C^a_{-i} \text{ played under } \sigma^*_{-i} \text{ given } B_{-i}) a_s(C, C^a_{-i}), \text{ for all } s \text{ and } C,
$$

where $\Pr(C^a_{-i} \text{ played under } \sigma^*_{-i} \text{ given } B_{-i})$ is the objective (correct) probability of $C^a_{-i}$ being played under $\sigma^*_{-i}$ given $B_{-i}$.

The above definition implies that if $i$ is sophisticated, she plays a best response against others with the knowledge of their beliefs and the distribution of their types.

Given the properties of beliefs in Lemma 2, a set of dominated strategies can be identified. More importantly, these dominated strategies are independent of the beliefs as long as they satisfy the properties in Lemma 2.

\(^{16}\)One common approach to model levels of sophistication is the level-k model introduced by Crawford and Iriberri (2007). In that model no player plays best response and therefore no one is fully sophisticated.
Proposition 2  Suppose $i$ has at least one acceptable school, given beliefs $B_i \left( \cdot, \sigma^*_i \right)$,

(i) listing an unacceptable or the worst school as the first choice is strictly dominated;

(ii) listing an unacceptable or the worst before an acceptable school is weakly dominated;

(iii) excluding an acceptable school from the list is weakly dominated;

(iv) if $A_{i,s} \left( C, \sigma^*_i \right) \in (0, 1)$, $\forall s$ and $C \neq (0, ..., 0)$, moving $s$ upward strictly increases the probability of being accepted and the dominances in (ii) and (iii) become strict.

Intuitively, ranking a school first always gives a strictly positive probability of being assigned to that school, and a parent should thus aim at better schools first. Besides, the worst outcome of participating is being accepted by the worst school. By ranking better schools before the worst school, a parent increases her child’s chance of being assigned to a better school. If a school is unacceptable, putting it at the bottom or omitting it also increases the likelihood of getting into better schools.

Moreover, the results above are also satisfied in the case of homogeneous sophistication, and a truth-telling strategy is not dominated according to Proposition 2.

Assuming parents do not play dominated strategies, I characterize the choice probabilities. The term, "choice probabilities", is defined in a broader sense. With heterogeneous beliefs, the model cannot predict the probability of each parent choosing a particular list. Instead, it can only predict the probability that a choice falls into a group of lists.

Since weak dominance creates more indeterminacy, given the results in part (iv) of Proposition 2, I consider two cases: (a) for any parent, all elements in her beliefs are in $(0, 1)$; and (b) some elements in some parents’ beliefs may be zero.

2.3.2 All Elements in Beliefs Are Positive

Given $S = 4$, I assign the lists into 15 groups, $g_n, n = 1, ..., 15$. The criteria of grouping are the number and identities of schools included in the list while the order among the listed schools does not necessarily matter. They belong to three types: (a) 5 groups in which the lists include no more than one school; (b) 6 groups which include only two-school lists; and
(c) 4 groups of full lists. Appendix A.4 shows the detailed grouping and characterization of choice probabilities which is now defined with respect to the groups.

For type-(a) groups, the choice probabilities are the same as those in the Bayesian Nash equilibrium, since they are independent of beliefs. For the 6 groups of type (b), the grouping is only based on which two schools are included but not their ranking. For example, \((s, s', 0, 0)\) and \((s', s, 0, 0)\) are in the same group, while \((s, s'', 0, 0)\) and \((s'', s, 0, 0)\) are in a different one, given \(s \neq s' \neq s''\). The choice probabilities for these groups have two sources: The two included schools are the only acceptables, and only one of the two is acceptable while the other is unacceptable but better than the two excluded. Similar to those in the Bayesian Nash equilibrium, the contributions of both sources to the choice probabilities are weighted by mixing probabilities.

The remaining 4 groups of type (c) are differentiated by their last school. Namely, \((s, s', s'', s''')\) and \((s', s, s'', s''')\) are in the same group, while \((s, s', s''', s''')\) is in a different one, given that \(s, s', s'', \) and \(s'''\) are different. The choice probabilities for these groups come from three sources: The top three schools are all acceptable, two of the top three are acceptable, and only one of the top three is acceptable. Again, the contributions of last two sources are weighted by the mixing probabilities.

Three points should be highlighted here: (i) none of the choice probabilities involves beliefs; (ii) after grouping, the model is complete as it implies a unique distribution of groups given a distribution of preferences\(^{17}\) and (iii) the model implies a unique distribution of preferences given a distribution of groups, under the maintained assumptions AM.1-AM.4 and Assumptions UNACCEPTABLES and MIXING. The model is thus estimated by the (simulated) maximum likelihood:

\[
\max_\theta \sum_{i=1}^I \ln [\Pr (C_i \in g_n | X_i, Z_i; \theta)], n = 1, \ldots, 15.
\]

Similar to a multinomial probit, given the assumptions, there is a unique solution to the

\(^{17}\)The choice probabilities of the 15 groups add up to 1.
above maximization problem, and thus the model is identified. This is better illustrated in full-list groups. Given any of these groups, one knows which school is the worst, while in a multinomial probit, the best school is known. One complication is from the mixing probabilities. Fortunately, they are also identified since the choice probability of the group, \((0, 0, 0, 0)\), is independent of the mixing probabilities.

### 2.3.3 Some Elements in Beliefs May Be Zero

When parents are allowed to make mistakes, some elements in their beliefs are likely to be zero. However, the more zeros are allowed in the beliefs, the less tractable the characterization of choice probabilities becomes. Facing this trade-off, I consider the following:

**Assumption ZERO-PROB** Some parents may expect that the probabilities of being assigned to School 1 are zero if it is ranked third, fourth, or omitted, while other elements in their beliefs are always in \((0, 1)\) for every parent.

School 1 has the smallest quota, 63 seats and is also the "best" school with the highest average test score of graduating students. In the data, there are 228 parents ranking School 1 top, and it is impossible to get into School 1 unless it is ranked top.

I assume that all parents assign positive values to the probabilities of being assigned to School 1 when it is ranked 2nd because: (i) 157 parents ranking School 1 second implies that many have assigned a positive value to the probability; and (ii) otherwise, it would require possibly too restrictive assumptions to characterize the choice probabilities. For similar reasons, I do not consider zero probabilities for School 2, the second "best" school.\(^{18}\)

Under Assumption ZERO-PROB, if a parent does have these zero probabilities, it does not matter if School 1 is ranked third or fourth as long as it is optimal not to rank it top 2.

**Lemma 3** Under Assumption ZERO-PROB, for parent \(i\) with at least three acceptable schools, there exists a cutoff, \(\bar{u}_i \geq 0\), which is a function of \(i\)'s beliefs and preferences but

\(^{18}\)School 2 has the second smallest quota, 227 seats. In the data, there are 431 parents ranking School 2 top. However, there are 206 parents ranking School 2 second, while 75 ranking it 3rd or 4th. Further discussion is in Subsection 3.2.
not of \( u_{i,1} \), such that School 1 is not ranked as top 2 choices if and only if \((u_{i,1} - u_{i,s'}) < \bar{u}_i\), where \( s' \) is the second best school among schools \( s \neq 1 \).

If \( \bar{u}_i = 0 \), Lemma 3 is consistent with truth-telling behavior. When \( \bar{u}_i > 0 \), \( i \) takes cautions by ranking School 1 low if it is not significantly better than the others.

Since the zero probabilities create many payoff-equivalent lists, to simplify the analysis, I make the following assumption regarding the mixed strategies in equilibrium.

**Assumption ZERO-PROB-MIXING** Given Assumption ZERO-PROB, in addition to the mixed strategies specified in Assumption MIXING, if \( i \)'s preferences are such that the list \((c_1, c_2, 1, c_4)\) is a best response, and that \( u_{i,c_4} = \min \{ u_{i,1}, u_{i,2}, u_{i,3}, u_{i,4} \} \) and \( u_{i,s} > 0 \) for \( s \neq c_4 \), then \( i \) mixes among:

(i) \((c_1, c_2, 1, c_4), (c_1, c_2, c_4, 1), \) and \((c_1, c_2, 0, 0)\), if \( u_{i,c_4} > 0 \);

(ii) \((c_1, c_2, 1, c_4)\) and \((c_1, c_2, 0, 0)\), if \( u_{i,c_4} < 0 \).

If \( i \) has zeros in her beliefs as specified in Assumption ZERO-PROB, \((c_1, c_2, c_4, 1)\) and \((c_1, c_2, 0, 0)\) are also best responses whenever \((c_1, c_2, 1, c_4)\) is a best response. Therefore, the mixing patterns should be interpreted as a combination of parents without zeros in their beliefs and those with zeros and playing these mixed strategies. The common strategy assumption then implicitly requires that the group of parents with zeros is exogenously determined. This may not be too restrictive if all parents expect a very small probability of being accepted by School 1 when it is ranked third or fourth.

Moreover, if no one has zeros in beliefs, the mixing probabilities should be such that \((c_1, c_2, 1, c_4)\) is always being played with probability one given \( i \)'s preferences. This provides a test for Assumption ZERO-PROB.

**Choice Probabilities and Estimation** Putting together Proposition 2, Assumptions UNACCEPTABLES, MIXING, ZERO-PROB, and ZERO-PROB-MIXING, I re-assign the lists into 18 groups. The grouping now depends on how School 1 is ranked.

The new groups can be summarized into six types: (a) 5 groups in which the lists include no more than one schools; (b) 3 groups in which the lists only include School 1
and another school; (e) 3 groups in which the lists only include two schools and exclude School 1; (d) 3 groups where the lists rank all four schools and School 1 is ranked top 2; (e) 3 groups where the lists rank all four schools and School 1 is ranked third; (f) 1 group where the lists rank all four schools and School 1 is ranked fourth.

The detailed characterization is again shown in Appendix A.4. The choice probabilities for groups of types (a) and (b) can be formulated the same as in the previous case. The main difference is that for types (c)-(f), one has to consider how School 1 is ranked. Namely, based on Lemma 3 and Assumption ZERO-PROB-MIXING, the choice probability takes into account if School 1 is optimally ranked top 2 or otherwise.

For example, for groups of type (c), now there is a possibility that School 1 is acceptable but is optimally not ranked top 2. Thus, it may be excluded from the list because some parents expect the probability of being accepted by School 1 is zero if it is ranked third. The characterization takes this into account, so do those of types (d)-(f) groups.

I further assume the cutoff in Lemma 3, $\bar{u}_i$, has the following form:

$$\bar{u}_i = \exp (\eta_0 + X_i \eta_X + (Z_{i,2}, Z_{i,3}, Z_{i,4}) \eta_Z + Y_i \eta_Y),$$

where $Y_i$ is correlated with beliefs but not with preferences; and the exponential function ensures $\bar{u}_i$ being non-negative. The characteristics of School 1 are excluded due to Lemma 3 and together with $Y_i$, preferences and the cutoff are separately identified.

Let $\tilde{\theta}$ be the set of parameters including the additional mixing probabilities and the coefficients in $\bar{u}_i$. Similar to previous cases, the model is estimated by the (simulated) maximum likelihood: $\max_{\tilde{\theta}} \sum_{i=1}^{T} \ln \left[ \Pr \left( C_i \in \tilde{g}_n | X_i, Z_i, Y_i; \tilde{\theta} \right) \right], n = 1, ..., 18.$

2.4 Relationship among the Different Cases

The relationship among the above three cases is such that: $\text{BNE} \subset \text{Positive Belief} \subset \text{Zero Belief}$. The Bayesian Nash equilibrium (BNE) under homogeneous sophistication is nested in the other two cases where sophistication may be heterogeneous. Besides, the
case where all elements in everyone’s beliefs are positive (Positive_Belief) is nested in
the other case where I allow some elements to be possibly zero (Zero_Belief).

Case TT where everyone is truth-telling is also considered in estimation. It is nested in
the two cases with heterogeneous sophistication: TT ⊂ Positive_Belief ⊂ Zero_Belief.
However, there is no clear nesting structure between TT and BNE. I will present some
model selection tests along with the estimation results.

2.5 Sophistication and Incentives

To see who is more sophisticated/strategic in the game, it is necessary to measure parents’
sophistication. Measures on individual sophistication are ideal. However, since our esti-
mates of preferences can only tell us the distribution of preferences conditional on \((X_i, Z_i)\),
the sophistication can only be measured conditional on \((X_i, Z_i)\) as well.

With either heterogeneous or homogeneous beliefs, a parent is defined as sophisticated
if her beliefs are the same as the (correct) equilibrium beliefs. With the large number of
players, the empirical beliefs, \(\hat{B}\), which is discussed in details previously in Section 2.2.3,
provide a good approximation of the equilibrium beliefs. Especially, with heterogeneity in
sophistication, it is impossible to solve the equilibrium given that the joint distribution of
preference and sophistication is unknown and not estimated. All the following measures
can be calculated by replacing equilibrium beliefs \(B^*\) with \(\hat{B}\), and \(\theta\) with estimates \(\hat{\theta}\).

2.5.1 Probability of Observing A Given Action

Under the assumption that everyone with \((X_i, Z_i)\) plays a best response, the model can pre-
dict the choice probability for each list, \(P_{i,k}^{BR} \equiv \Pr (C_i = C^k | X_i, Z_i, B^*; \theta), k = 1, \ldots, 41,\)
for \(S = 4\). Given \(i\) chooses \(C_i\), define \(d_{i,k}\), such that \(d_{i,k} = 1\) if \(C_i = C^k\), and 0 otherwise.
If \(i\) always plays a best response,

\[
E[d_{i,k} | X_i, Z_i, B^*; \theta] - P_{i,k}^{BR} = 0, \forall k.
\]
One may test the hypothesis that \( i \) plays a best response by running the 41 regressions:

\[
d_{i,k} - P_{i,k}^{BR} = \delta_0 + X_{i} \delta_X + Z_{i} \delta_Z + W_{i,k} \delta_W + \nu_{i,k}, \forall k,
\]

where \( W_{i,k} \) is a vector of variables other than \( X_{i} \) and \( Z_{i} \). Under the null, all coefficients, \((\delta_0, \delta_X, \delta_Z, \delta_W)\), should be zero.

If \( C^k \) is a one-school list or is \((0, 0, 0, 0)\), all the coefficients should always be zero, because the model assumes parents do not make mistakes when playing these strategies. I therefore use these five regressions as placebo tests.

Under the assumption that \( i \) is always truth-telling, \( C^k \)'s choice probability is:

\[
P^{TT}_{i,k} \equiv \Pr \left( C^k \text{ is chosen under truth-telling} \mid X_i, Z_i, B^*; \theta \right).
\]

Similarly, I regress \((d_{i,k} - P^{TT}_{i,k})\) on \((X_i, Z_i, W_{i,k})\) and test the truth-telling hypothesis.

2.5.2 Incentives to Be Strategic

In real life, it is not implausible that it is costly to find a best response. The incentive to be strategic, or to play a best response, would thus affect parents’ behavior.

The first incentive measure is the probability that truth-telling is a best response:

\[
P^{TT=BR}_{i} \equiv \Pr \left( \text{truth-telling is a best response} \mid X_i, Z_i, B^*; \theta \right).
\]

I assume that the cost of finding a best response is lower if truth-telling itself is a best response. Therefore, a high \( P^{TT=BR}_{i} \) means a greater incentive for \( i \) to play a best response.

The second measure is the expected utility gain if \( i \) changes from truth-telling to best responding:

\[
Gain_{i} \equiv \left( V_{i}^{BR} - V_{i}^{TT} \right) / V_{i}^{BR},
\]

where \( V_{i}^{BR} \) is the expected utility if \( i \) always plays a best response, and \( V_{i}^{TT} \) is the one
when she is always truth-telling. If Gain$_i$ is higher, $i$ has a greater incentive to find her best response.

I later include $P_{i}^{TT=BR}$ and Gain$_i$ the 41 regressions in (3) to test if parents’ behavior is affected by these incentives.

3 Reduced-Form Results

Before reporting the model estimates, I present evidence directly from the data which are consistent with the assumptions and model predictions.

3.1 Understanding the Rules of the Game

One of the important assumptions is that parents understand the rules of the game, and therefore their beliefs follow the structure specified in Lemma 2. I examine parents’ responses to two questions in the 2002 survey: "On a scale of 0-10, what is the probability that your child is admitted into your 1st (2nd) choice?" Table 5 shows the summary statistics. The empirical beliefs are calculated from the submitted lists. The empirical beliefs and self-reported beliefs share the same pattern, although they do not exactly match.

Also consistent with Lemma 2, parents on average expect that moving up a school in the list increases the probability of being accepted by that school.

More precisely, $V_{i}^{BR}$ is defined as follows: Given any realization of $\varepsilon_i$, $i$ plays a best response. I calculate the expected utility given $\varepsilon_i$ and then integrate it over all possible $\varepsilon_i$. Namely,

$$V_{i}^{BR} = \int_{\varepsilon_i} \max_{\sigma \in \Delta(C)} \sum_{s=1}^{S} A_s \left( \sigma, \sigma_{-i} \right) \max \left( u_{i,s}, 0 \right) dF_{\varepsilon_i}(\varepsilon_i).$$

$V_{i}^{TT}$ is similarly defined.

Since these questions are asked after the assignment is realized, the results may just show the ex post probability, namely, whether or not the student has been accepted by that school. For this reason, I use their self-reported top 2 choices which are not necessarily the ones submitted by them.
Table 5: Empirical Beliefs and Parents’ Self-Reported Beliefs

<table>
<thead>
<tr>
<th>School</th>
<th>Empirical Beliefs</th>
<th>Survey Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Beliefs</td>
<td>Mean</td>
</tr>
<tr>
<td>1</td>
<td>26.7%</td>
<td>4.35</td>
</tr>
<tr>
<td>2</td>
<td>50.7%</td>
<td>6.72</td>
</tr>
<tr>
<td>3</td>
<td>100%</td>
<td>8.11</td>
</tr>
<tr>
<td>4</td>
<td>100%</td>
<td>8.32</td>
</tr>
</tbody>
</table>

a. Calculated from the actual submitted lists. Each entry shows the probability being accepted the school when that school is ranked 1st or 2nd, given all other students’ submitted lists.
b. Responses to the survey question: "On a scale of 0-10, what is the probability that your child is admitted into your 1st (2nd) choice?"
c. The 1st and 2nd choices are self-reported and thus are not necessarily the submitted ones.

3.2 Undominated Strategies, Truth-Telling, and Zero Probabilities

Lemma 2 leads to the dominated strategies in Proposition 2, and parents should not play these strategies in equilibrium. Table 6 shows the distribution of parents’ first choice: 24.9% ranking School 1 top, while 47.0% ranking School 2 top.

Table 6: Parents’ First Choices and Claimed Best Schools

<table>
<thead>
<tr>
<th>School</th>
<th>Quota</th>
<th># Parents</th>
<th>Rank It #1</th>
<th># Parents Claim It as the Best</th>
<th>Rank the Claimed Best</th>
<th># Omitted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>63</td>
<td>228</td>
<td>(24.8%)</td>
<td>579 (82.8%)</td>
<td>186</td>
<td>107</td>
</tr>
<tr>
<td>2</td>
<td>227</td>
<td>431</td>
<td>(47.0%)</td>
<td>58 (8.3%)</td>
<td>49</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>310</td>
<td>66 (7.2%)</td>
<td>26 (4.3%)</td>
<td>3 (0.4%)</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>360</td>
<td>8 (0.9%)</td>
<td>3 (0.4%)</td>
<td>0 (0.4%)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Non-Particip.</td>
<td>181 (19.8%)</td>
<td>33 (4.7%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>960</td>
<td>914 (100%)</td>
<td>699 (100%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Responses to a survey question: "Among those to which you could apply, which school was the best?"
b. Among all the parents who claim a given school as the best school, these five columns show how they rank it in the application, conditional on participating.
c. "Other" means schools other than the four schools. This may be due to misreporting/misunderstanding.

A survey question asks, "Among the schools to which you could apply, which school was the best?" Among 699 valid responses, 82.8% claim School 1 as the best. Comparing the first-choice school in submitted lists with the most recognized school, the difference is significant. This is evidence against the truth-telling hypothesis.

21 This question is not necessarily asking the parent’s favorite or her most preferred school.
If everyone understands the rules, the first-choice school should never be the worst school (Proposition 2). This is consistent with the data in Table 6; only 8 parents top rank School 4, while even fewer people claim it as the best school.

Proposition 2 also predicts that the last-choice school (or the omitted school conditional on participating) should either be an unacceptable/the worst or a school which is impossible to get in if it is ranked low. The last five columns in Table 6 show how parents rank the claimed best school. For Schools 2, 3 and 4, only two parents rank the best school as fourth or omit it. Following the discussion in Section 2.3.3, this implies that zero probability is less of a concern for these schools.

However, there are 36 (6.2%) parents ranking School 1 fourth, while another 25 (4.3%) participants excluding School 1 from their list. Since School 1 has the smallest quota, the only chance to get into School 1 is to top rank it, and the probability is merely 26.7% in the data. It is very likely that a parent expects that there is no chance to get into School 1 when ranking it third or fourth. This confirms the discussion in Section 2.3.3.

3.3 Attention on Uncertainty

In survey, there are questions regarding parents’ perception of the importance of 12 factors. Parents rate them on a scale of 1-5, where 5 being very important. Three factors are related to uncertainty of the game: (i) admission quota and possibility of being accepted; (ii) reducing the probability of being assigned to bad schools; and (iii) considering other parents’ applications. Since considering others’ applications may also be correlated with school quality because other parents’ applications reveal their preferences over school, I create Attn\_U\_i (attention on uncertainty) as the average response to the first two factors and Attn\_Others\_i (attention on others’ application) as the response to the third.

Another nine factors are about school quality, e.g., teachers’ quality, peer quality, school facilities, etc. I take the average of the responses to these questions and create Attn\_Q\_i (attention on quality).

For a sophisticated parent, she understands the uncertainty of other parents’ behavior
This implies a positive correlation between sophistication and \( \text{Attn}_i \). While I investigate of the correlation between \( \text{Attn}_U \) and parents’ performance in Section 4.2, I first explore how family background is correlated with \( \text{Attn}_U \).

Table 7 presents the regression results of \( \text{Attn}_U \) on family background and students’ characteristics while controlling \( \text{Attn}_Q \) and \( \text{Attn}_{\text{Others}} \). The first column contains the full sample, and family background has no significant correlation with \( \text{Attn}_U \). I exclude non-participants (column 2) and then those who submitted partial lists (columns 3 and 4). The negative coefficient on parent’s income becomes significant at 10% level and larger in magnitude, especially in the subsample of parents who submitted full lists. As a comparison, this coefficient is significantly positive (at 5% level) for the sample of non-participants (column 5).

<table>
<thead>
<tr>
<th>Dependent Variable: Attention on Uncertainty</th>
<th>Full Sample(^a)</th>
<th>Participant(^b)</th>
<th>( \geq 2 ) Schools(^a)</th>
<th>Full List(^c)</th>
<th>Non-Participant(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean(Dep V)</td>
<td>4.339</td>
<td>4.357</td>
<td>4.361</td>
<td>4.350</td>
<td>4.190</td>
</tr>
<tr>
<td>Std Dev(Dep V)</td>
<td>0.743</td>
<td>0.721</td>
<td>0.708</td>
<td>0.698</td>
<td>0.900</td>
</tr>
<tr>
<td>Parent(<em>Edu</em>) (_i)</td>
<td>0.005</td>
<td>0.007</td>
<td>0.006</td>
<td>0.012</td>
<td>-0.015</td>
</tr>
<tr>
<td>( (0.013) )</td>
<td>( (0.014) )</td>
<td>( (0.014) )</td>
<td>( (0.015) )</td>
<td>( (0.054) )</td>
<td></td>
</tr>
<tr>
<td>Parent(<em>Inc</em>) (_i)</td>
<td>-0.037</td>
<td>-0.053</td>
<td>-0.056*</td>
<td>-0.069*</td>
<td>0.550**</td>
</tr>
<tr>
<td>( (0.035) )</td>
<td>( (0.034) )</td>
<td>( (0.034) )</td>
<td>( (0.035) )</td>
<td>( (0.228) )</td>
<td></td>
</tr>
<tr>
<td>Own(<em>Score</em>) (_i)</td>
<td>0.231</td>
<td>0.288</td>
<td>0.312</td>
<td>0.944*</td>
<td>1.739</td>
</tr>
<tr>
<td>( (0.414) )</td>
<td>( (0.403) )</td>
<td>( (0.402) )</td>
<td>( (0.512) )</td>
<td>( (3.567) )</td>
<td></td>
</tr>
<tr>
<td>Awards(_i)</td>
<td>0.020</td>
<td>0.038</td>
<td>0.038</td>
<td>0.061*</td>
<td>0.048</td>
</tr>
<tr>
<td>( (0.026) )</td>
<td>( (0.029) )</td>
<td>( (0.029) )</td>
<td>( (0.032) )</td>
<td>( (0.080) )</td>
<td></td>
</tr>
<tr>
<td>Girl(_i)</td>
<td>-0.047</td>
<td>-0.036</td>
<td>-0.026</td>
<td>-0.023</td>
<td>-0.248</td>
</tr>
<tr>
<td>( (0.050) )</td>
<td>( (0.051) )</td>
<td>( (0.051) )</td>
<td>( (0.057) )</td>
<td>( (0.205) )</td>
<td></td>
</tr>
<tr>
<td>Attn(<em>Others</em>) (_i)</td>
<td>-0.003</td>
<td>-0.011</td>
<td>-0.013</td>
<td>-0.016</td>
<td>0.073</td>
</tr>
<tr>
<td>( (0.021) )</td>
<td>( (0.021) )</td>
<td>( (0.021) )</td>
<td>( (0.024) )</td>
<td>( (0.093) )</td>
<td></td>
</tr>
<tr>
<td>Attn(<em>Q</em>)</td>
<td>0.823***</td>
<td>0.822***</td>
<td>0.798***</td>
<td>0.750***</td>
<td>0.755***</td>
</tr>
<tr>
<td>( (0.056) )</td>
<td>( (0.056) )</td>
<td>( (0.056) )</td>
<td>( (0.063) )</td>
<td>( (0.237) )</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.011</td>
<td>-0.169</td>
<td>-0.143</td>
<td>-3.230</td>
<td>-12.600</td>
</tr>
<tr>
<td>( (2.138) )</td>
<td>( (2.085) )</td>
<td>( (2.078) )</td>
<td>( (2.649) )</td>
<td>( (18.430) )</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>676</td>
<td>605</td>
<td>597</td>
<td>457</td>
<td>71</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.270</td>
<td>0.294</td>
<td>0.281</td>
<td>0.279</td>
<td>0.364</td>
</tr>
</tbody>
</table>

\( a. \) The full sample includes every parent whose relevant variables are not missing. Participants are those who submits a list which is not \((0,0,0,0)\). The subsample \((\geq 2 \) schools\) includes participants whose submitted lists have at least 2 schools. And the subsample with full list are those who submit a full list. Elementary school fixed effects included. Standard errors in parentheses. *** \( p<0.01\), ** \( p<0.05\), * \( p<0.1\).
The pattern of the coefficients of parents’ income across subsamples is consistent with the model prediction that parents submitting partial lists are insensitive to the uncertainty in the game because of their good outside option.

The above negative correlation between $Attn_U_i$ and parents income is robust if $Attn_U_i$ is break into attention on admission quota and attention on probability of getting into bad schools, as shown in Table A-1 in appendix. The coefficient on income is negative in both regressions, although one of them is not significant. The same regression is run for $Attn_Others_i$ and the coefficient on parents’ education is significantly negative, although the one on parents’ income is insignificant. I also regress $Attn_Q_i$ on the same set of variables. Results in Table A-1 show that, contrary to those from $Attn_U_i$ regressions, parents’ income is significantly positively correlated with $Attn_Q_i$.

In short, there is a negative correlation between $Attn_U_i$ and parents’ income, and how $Attn_U_i$ affects their performance in the game is investigated in Section 4.2.

4 Model Estimation and Test Results

This section presents the estimates from four cases: (i) BNE: symmetric Bayesian Nash equilibrium with homogeneous sophistication; (ii) Positive_Belief: heterogeneous sophistication with all belief elements being positive; (iii) Zero_Belief: heterogeneous sophistication with some belief elements being possibly zero; and (iv) TT: truth-telling. The utility function is specified as:

$$u_{i,s} = \alpha_s + \beta_{X,1}Own\_Score_i + \beta_{X,2}Parent\_Inc_i + \beta_{X,3}Parent\_Edu_i + \beta_{X,4}Girl_i + \beta_{X,5}Awards_i + \beta_{Z,1}Distance_i + \beta_{Z,2}Own\_Score_i \times School\_Score_s + \beta_{Z,3}Parent\_Inc_i \times School\_Score_s + \beta_{Z,4}Parent\_Edu_i \times School\_Score_s + \beta_{Z,5}Awards_i \times School\_Score_s + \beta_{Z,6}Girl_i \times School\_Score_s + \varepsilon_{is},$$
where $\alpha_s$ is the middle school fixed effect; $School\_Score_s$ is the (log) average test score of school $s$; and other variables are defined in Table 3. The part $\beta_{X,1} Own\_Score_i + \beta_{X,2} Parent\_Inc_i + \beta_{X,3} Parent\_Edu_i + \beta_{X,4} Awards_i + \beta_{X,5} Girl_i$, which is constant for any inside school, captures the quality of outside option. $(\varepsilon_{i1}, \ldots, \varepsilon_{i4}) \sim N(0, \Sigma)$ are i.i.d. across students and the variance of $\varepsilon_{i,1}$ is 1.

In Case Zero Belief, the cutoff $\tilde{u}_i \geq 0$ as in Lemma 3 is specified as:

$$\tilde{u}_i = \exp \left( \eta_{i,e} + \eta_{X,1} Own\_Score_i + \eta_{X,2} Parent\_Inc_i + \eta_{X,3} Parent\_Edu_i + \eta_{X,4} Awards_i + \eta_{X,5} Girl_i + \sum_{s=2}^{4} \eta_{Z,s} Distance_{i,s} \right), \quad (4)$$

where, $\eta_{i,e}$ is the elementary school $e$’s fixed effect. Elementary schools may matter as teachers usually help them with filling out applications. I also estimate the model with $Attn\_U_i$ and $Attn\_Q_i$ in $\tilde{u}_i$ which are possibly correlated with beliefs.

In all four cases, the choice probabilities are simulated as in a multinomial probit model by the logit-smoothed accept-reject simulator with 300 draws (details in Appendix A.5).

### 4.1 Estimation Results and Model Selection

Table 8 presents the coefficients of main variables in utility function for all four cases. Standard errors are calculated by the robust asymptotic approximation as in McFadden and Train (2000). In Zero Belief, results from excluding and including $Attn\_U_i$ and $Attn\_Q_i$ in $\tilde{u}_i$ are reported in columns (1) and (2) respectively.

As discussed previously, to distinguish between Zero Belief and Positive Belief, the mixing probabilities which are unique to Zero Belief provide a test. Precisely, based on Assumption ZERO-PROB-MIXING, $(c_1, c_2, c_4, 1)$ and $(c_1, c_2, 0, 0)$ are also best responses whenever $(c_1, c_2, 1, c_4)$ is a best response, while this is not true in Positive Belief. As column (1) shows, when the worst, $u_{c_1}$, is acceptable, parents submit $(c_1, c_2, 0, 0)$ with probability 40.3% and $(c_1, c_2, c_4, 1)$ with probability 59.7%.

Interestingly, when the worst school is unacceptable, $u_{c_1} < 0$, parents do not mix as they play $(c_1, c_2, 1, c_1)$ with a probability close to one, or play $(c_1, c_2, 0, 0)$ with almost zero probability, as columns (1) and (2) show. (Note: The exact percentages and probabilities are derived from the estimated model coefficients and the standard errors are provided in the appendix. The 95% confidence intervals are also calculated based on the asymptotic normality of the estimators. For instance, the 95% confidence interval for the probability of choosing $(c_1, c_2, 0, 0)$ is approximately (0.4, 0.5) and for $(c_1, c_2, c_4, 1)$, it is (0.59, 0.6). These intervals are calculated using the standard errors of the estimated parameters and the normal distribution.)
Table 8: Preferences over Schools: Model Estimation Results from Different Cases

<table>
<thead>
<tr>
<th></th>
<th>Zero_Belief</th>
<th>Posit_Belief</th>
<th>BNE</th>
<th>TT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Distance(_{i,s})</td>
<td>-0.254***</td>
<td>-0.246***</td>
<td>-0.242***</td>
<td>-0.026</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.007)</td>
<td>(0.018)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Own_Score(<em>{i} \times )School_Score(</em>{s})</td>
<td>11.516***</td>
<td>19.600***</td>
<td>65.545***</td>
<td>21.008***</td>
</tr>
<tr>
<td></td>
<td>(0.182)</td>
<td>(0.021)</td>
<td>(0.152)</td>
<td>(0.023)</td>
</tr>
<tr>
<td>Parent_Inc(<em>{i} \times )School_Score(</em>{s})</td>
<td>-0.047***</td>
<td>0.329***</td>
<td>0.663***</td>
<td>0.195***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.023)</td>
<td>(0.015)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Parent_Edu(<em>{i} \times )School_Score(</em>{s})</td>
<td>-0.256***</td>
<td>-0.263***</td>
<td>-0.450***</td>
<td>0.159***</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Girl(<em>{i} \times )School_Score(</em>{s})</td>
<td>0.266</td>
<td>0.271***</td>
<td>-0.359***</td>
<td>-0.492***</td>
</tr>
<tr>
<td></td>
<td>(0.164)</td>
<td>(0.012)</td>
<td>(0.063)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Awards(<em>{i} \times )School_Score(</em>{s})</td>
<td>0.631***</td>
<td>0.545***</td>
<td>0.688***</td>
<td>0.855***</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.002)</td>
<td>(0.017)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Own_Score(_{i})</td>
<td>-74.678***</td>
<td>-126.712***</td>
<td>-409.774***</td>
<td>-131.443***</td>
</tr>
<tr>
<td></td>
<td>(0.682)</td>
<td>(0.176)</td>
<td>(0.811)</td>
<td>(0.159)</td>
</tr>
<tr>
<td>Parent_Inc(_{i})</td>
<td>-0.019</td>
<td>-2.402***</td>
<td>-4.537***</td>
<td>-1.288***</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
<td>(0.142)</td>
<td>(0.107)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>Parent_Edu(_{i})</td>
<td>1.538***</td>
<td>1.626***</td>
<td>2.728***</td>
<td>-1.005***</td>
</tr>
<tr>
<td></td>
<td>(0.038)</td>
<td>(0.024)</td>
<td>(0.028)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Awards(_{i})</td>
<td>-4.198***</td>
<td>-3.550***</td>
<td>-4.512***</td>
<td>-5.406***</td>
</tr>
<tr>
<td></td>
<td>(0.104)</td>
<td>(0.020)</td>
<td>(0.091)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>Girl(_{i})</td>
<td>-1.647</td>
<td>-1.572***</td>
<td>2.260***</td>
<td>3.112***</td>
</tr>
<tr>
<td></td>
<td>(1.025)</td>
<td>(0.074)</td>
<td>(0.362)</td>
<td>(0.040)</td>
</tr>
<tr>
<td>Mixing Probabilities:</td>
<td><img src="image_url" alt="Image" /></td>
<td><img src="image_url" alt="Image" /></td>
<td><img src="image_url" alt="Image" /></td>
<td><img src="image_url" alt="Image" /></td>
</tr>
<tr>
<td>m(_{1,4})</td>
<td>0.766</td>
<td>0.794</td>
<td>0.752</td>
<td>0.766</td>
</tr>
<tr>
<td></td>
<td>[0.689, 0.833]</td>
<td>[0.725, 0.852]</td>
<td>[0.696, 0.803]</td>
<td>[0.677, 0.835]</td>
</tr>
<tr>
<td>m(_{1,2})</td>
<td>0.179</td>
<td>0.156</td>
<td>0.212</td>
<td>0.190</td>
</tr>
<tr>
<td></td>
<td>[0.119, 0.254]</td>
<td>[0.102, 0.223]</td>
<td>[0.165, 0.267]</td>
<td>[0.122, 0.272]</td>
</tr>
<tr>
<td>m(_{2,4})</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td>[0.000, 1.000]</td>
<td>[0.111, 1.000]</td>
<td>[1.000, 1.000]</td>
<td>[0.912, 0.987]</td>
</tr>
</tbody>
</table>

Given \(u_{c4} \leq u_{s} \forall s\), Prob. submitting:\(a\)

|                  |当日（c1, c2, 0, 0）if \(u_{c4} > 0\) | \(c1, c2, c4, 1\) if \(u_{c4} > 0\) | \(c1, c2, 0, 0\) if \(u_{c4} < 0\)|                 |
|------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| LR Test: \(\chi^2_{(11)}\) | 213.560 | 217.999 | 127.698 | 681.616 | 244.072 |
| p-value           | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |

In Case Zero_Belief, column (1) are estimates without attention measures, while column (2) are those when including attention measures in the cutoff function, \(\tilde{u}_i\). Middle school fixed effects are included in all cases. \(m_{k,l}, l \geq k\), denotes the probability that a \(l\)-school is submitted when only \(k\) schools are acceptable.

a. These are the probabilities of submitting each list when \((c1, c2, 1, c4)\) is a best response in Zero_Belief.

The likelihood test is for the hypothesis that coefficients of the 11 individual characteristics equal zero. 95% confidence intervals in brackets for mixing probabilities, as logistic functions are used in the estimation to ensure all values in \([0,1]\). Standard errors in parentheses. *** \(p<0.01\), ** \(p<0.05\), * \(p<0.1\).
Table 9: Marginal Effects of Individual Characteristics: Estimates from Case Zero_Belief

<table>
<thead>
<tr>
<th>School</th>
<th>Parent_Inc</th>
<th>Own_Score</th>
<th>Parent_Edu</th>
<th>Awards</th>
<th>Girl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.318</td>
<td>-1.822</td>
<td>-2.713</td>
<td>-0.079</td>
<td>-0.203</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.535)</td>
<td>(0.217)</td>
<td>(0.007)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>2</td>
<td>-0.315</td>
<td>-2.597</td>
<td>-4.030</td>
<td>-0.062</td>
<td>-0.263</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.523)</td>
<td>(0.216)</td>
<td>(0.007)</td>
<td>(0.020)</td>
</tr>
<tr>
<td>3</td>
<td>-0.313</td>
<td>-2.919</td>
<td>-4.579</td>
<td>-0.055</td>
<td>-0.263</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.519)</td>
<td>(0.215)</td>
<td>(0.007)</td>
<td>(0.019)</td>
</tr>
<tr>
<td>4</td>
<td>-0.310</td>
<td>-3.822</td>
<td>-6.116</td>
<td>-0.035</td>
<td>-0.312</td>
</tr>
<tr>
<td></td>
<td>(0.042)</td>
<td>(0.506)</td>
<td>(0.214)</td>
<td>(0.007)</td>
<td>(0.019)</td>
</tr>
</tbody>
</table>

For each variable, columns (1) are calculated using estimates from the Zero_Belief case without attention measures, while columns (2) are calculated using those from the Zero_Belief case when including attention measures as independent variables. For Parent_Inc and Own_Score, the table reports the change in each school’s utility, in percentage points, if there is a 1% increase in them. For other 3 variables, it reports the change in utility when the variable increase by 1 unit. Standard errors in parentheses.

For variables in the interaction terms, Table 9 present their marginal effects which are calculated with the two sets of estimates from Zero_Belief. For the five variables except Girl, they have a negative effect on the utility of any school, since wealthier and more educated parents and students with better achievements have better outside options.

In the following, I use estimates from Zero_Belief without attention measures because (i) the estimates are not very different as shown in Tables 8 and 9 and (ii) the original attention measures are missing for 26% observations and results may be sensitive to imputation.

4.2 Sophistication and Incentives to Be Strategic

This section investigates who strategizes better and how parents response to incentives. Measures of sophistication and incentives are constructed using estimated preferences and

and (2) show. Therefore, Zero_Belief is not an issue when uc < 0. With estimates in Column (1) of Table 8 on average, the probability that School 1 is optimally ranked below top 2 and uc < 0 < u1 is 21.27%. The probability that School 1 is optimally ranked below top 2 and 0 < uc < u1 is 6.92%. This means ignoring Zero_Belief and using Positive_Belief would misinterpret 63 parents’ revealed preferences.
the empirical equilibrium beliefs.

### 4.2.1 Deviation from Best Responding and Truth-Telling: Overcautiousness

As Section 2.5.1 shows, deviations from best responding is on average zero if everyone plays best responses, so are deviations from truth-telling if everyone is truth-telling. For the 24 full lists, Table 10 presents the summary statistics on how observed behaviors deviate from best responding and truth-telling. I run a t-test for the null hypothesis that each mean independently equals zero, and in only 4 cases I do not reject the null. This is true for both deviations from best responding and from truth-telling.

Table 10 highlights the importance to distinguish between best responding and truth telling because of their different predictions. For example, \((1, 2, 3, 4)\) is the most common true preference order, and it is played by 13.89% of the parents. The truth-telling hypothesis predicts that 25.19% should choose that list, but the best-responding hypothesis only predicts 10.65%. A similar discrepancy is found for the most under-used list, \((1, 3, 2, 4)\).

I also do the same exercise for the partial lists including \((0, \ldots, 0)\). Among the 12 two-school lists, the null is not rejected in 5 cases. More importantly, as the model predicts, deviations for the 4 one-school lists and \((0, \ldots, 0)\) have means which are not significantly different from zero, since by assumption parents do not make mistakes when playing these strategies. Details are collected in the appendix Table A-2.

These results provide another piece of evidence that parents are neither all-best-responding nor all-truth-telling in the game. Instead, parents are overcautious because they low rank School 1 and top rank School 2 too often. In the data, 24.51% of the parents top rank School 1, while the model predicts that 54.88% should optimally do that given other parents’ behavior. For School 2, 47.16% rank it top, but the model only predicts 6.69%. To top rank School 2 while avoiding School 1 is *ex ante* rational, because School 1 is the best and only has 63 slots, while School 2 is still a very good school and has 227 slots. However, as many parents choose a "safe" strategy, the overcautiousness leads to a coordination failure.
Table 10: Deviation from Best-Responding and Truth-Telling Predictions: Full Lists

<table>
<thead>
<tr>
<th>Rank-Order Lists</th>
<th>Observed Percent</th>
<th>Deviation from the Prediction of</th>
<th>Best Responding</th>
<th>Truth-Telling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (10^-2)</td>
<td>Std. Dev.</td>
<td>Mean (10^-2)</td>
</tr>
<tr>
<td>(1, 2, 3, 4)</td>
<td>13.89%</td>
<td>3.24***</td>
<td>0.346</td>
<td>-11.30***</td>
</tr>
<tr>
<td>(1, 2, 4, 3)</td>
<td>2.41%</td>
<td>-2.31**</td>
<td>0.153</td>
<td>-2.72***</td>
</tr>
<tr>
<td>(1, 3, 4, 2)</td>
<td>0.77%</td>
<td>-0.15</td>
<td>0.087</td>
<td>-0.15</td>
</tr>
<tr>
<td>(1, 3, 2, 4)</td>
<td>5.14%</td>
<td>-30.20***</td>
<td>0.235</td>
<td>-17.70***</td>
</tr>
<tr>
<td>(1, 4, 3, 2)</td>
<td>0.00%</td>
<td>-0.21***</td>
<td>0.003</td>
<td>-0.21***</td>
</tr>
<tr>
<td>(1, 4, 2, 3)</td>
<td>0.22%</td>
<td>-0.30**</td>
<td>0.047</td>
<td>-0.30*</td>
</tr>
<tr>
<td>(2, 1, 3, 4)</td>
<td>12.58%</td>
<td>10.76***</td>
<td>0.331</td>
<td>10.91***</td>
</tr>
<tr>
<td>(2, 1, 4, 3)</td>
<td>2.74%</td>
<td>2.24***</td>
<td>0.163</td>
<td>2.42***</td>
</tr>
<tr>
<td>(2, 3, 1, 4)</td>
<td>21.33%</td>
<td>18.50***</td>
<td>0.406</td>
<td>19.92***</td>
</tr>
<tr>
<td>(2, 3, 4, 1)</td>
<td>4.70%</td>
<td>4.01***</td>
<td>0.211</td>
<td>4.29***</td>
</tr>
<tr>
<td>(2, 4, 3, 1)</td>
<td>1.09%</td>
<td>0.93***</td>
<td>0.103</td>
<td>1.01***</td>
</tr>
<tr>
<td>(2, 4, 1, 3)</td>
<td>0.77%</td>
<td>0.60**</td>
<td>0.087</td>
<td>0.70**</td>
</tr>
<tr>
<td>(3, 1, 4, 2)</td>
<td>0.22%</td>
<td>-0.03</td>
<td>0.047</td>
<td>-0.18</td>
</tr>
<tr>
<td>(3, 1, 2, 4)</td>
<td>0.88%</td>
<td>-2.55***</td>
<td>0.094</td>
<td>-6.18***</td>
</tr>
<tr>
<td>(3, 2, 4, 1)</td>
<td>0.88%</td>
<td>-2.12***</td>
<td>0.093</td>
<td>-1.56***</td>
</tr>
<tr>
<td>(3, 2, 1, 4)</td>
<td>4.16%</td>
<td>-3.96***</td>
<td>0.199</td>
<td>-1.26*</td>
</tr>
<tr>
<td>(3, 4, 1, 2)</td>
<td>0.22%</td>
<td>0.09</td>
<td>0.047</td>
<td>0.12</td>
</tr>
<tr>
<td>(3, 4, 2, 1)</td>
<td>0.22%</td>
<td>-0.23</td>
<td>0.047</td>
<td>-0.16</td>
</tr>
<tr>
<td>(4, 1, 3, 2)</td>
<td>0.00%</td>
<td>0.00***</td>
<td>0.000</td>
<td>-0.01***</td>
</tr>
<tr>
<td>(4, 1, 2, 3)</td>
<td>0.00%</td>
<td>-0.01***</td>
<td>0.000</td>
<td>-0.02***</td>
</tr>
<tr>
<td>(4, 2, 3, 1)</td>
<td>0.00%</td>
<td>-0.03***</td>
<td>0.001</td>
<td>-0.03***</td>
</tr>
<tr>
<td>(4, 2, 1, 3)</td>
<td>0.00%</td>
<td>-0.02***</td>
<td>0.001</td>
<td>-0.02***</td>
</tr>
<tr>
<td>(4, 3, 1, 2)</td>
<td>0.00%</td>
<td>-0.01***</td>
<td>0.000</td>
<td>-0.01***</td>
</tr>
<tr>
<td>(4, 3, 2, 1)</td>
<td>0.55%</td>
<td>0.51**</td>
<td>0.074</td>
<td>0.52**</td>
</tr>
</tbody>
</table>

A t-test is run for the null hypothesis that each mean equals zero. *** p<0.01, ** p<0.05, * p<0.1.

a. It is the percentage out of the total of 914 students.
b. The means should be interpreted as percentages points, i.e., 3.24 means 3.24 percentage points.

4.2.2 Incentives to Be Strategic

Table 11 shows summary statistics of the incentive measures and how they are correlated with individual characteristics. The main result is that wealthier and more educated parents and students with better achievements have a lower cost for finding best responses (higher \( P_{TT=BR} \)) and a lower incentive to move away from always truth-telling (lower \( Gain_i \)).

The probability that truth-telling is a best response, \( P_{TT=BR} \), is relatively high – 77.4%. However, the variation is not low – min = 42.9%, max = 99.4%, and standard deviation
9.4%. In both regressions, \( Parent_{Inc_i} \), \( Parent_{Edu_i} \), \( Own_{Score_i} \), and \( Awards_i \) are positively correlated with \( P^{TT=BR}_i \), with or without controlling for \( Gain_i \).

\( Gain_i \) is the utility gain if changing from always truth-telling to always best responding. The mean gain is 0.029, equivalent to reducing the distance to a school by 10.8%, and its variation is also high – min = 0.000, max = 0.167 (reducing the distance by 48.2%), and standard deviation 0.019. When not controlling \( P^{TT=BR}_i \), \( Gain_i \) is negatively correlated with \( Parent_{Inc_i} \), \( Parent_{Edu_i} \), \( Own_{Score_i} \), and \( Awards_i \). However, when conditional on \( P^{TT=BR}_i \), all the correlations are moved to the positive side and those of \( Parent_{Edu_i} \), \( Own_{Score_i} \), and \( Awards \) become significantly positive.

Below, regressions are used to investigate how the incentives affect parents’ behavior. It is tempting to include both \( P^{TT=BR}_i \) and \( Gain_i \) in the same regression. However, this might cause multicollinearity, as the correlation between \( P^{TT=BR}_i \) and \( Gain_i \) is -0.945.

<table>
<thead>
<tr>
<th></th>
<th>( P^{TT=BR}_i )</th>
<th>( Gain_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean(Dep V)</td>
<td>0.774</td>
<td>0.029</td>
</tr>
<tr>
<td>Std Dev(Dep V)</td>
<td>0.094</td>
<td>0.019</td>
</tr>
<tr>
<td>( Gain_i )</td>
<td>-3.469***</td>
<td>-0.230***</td>
</tr>
<tr>
<td></td>
<td>(0.138)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>( P^{TT=BR}_i )</td>
<td>0.046*** (0.004)</td>
<td>-0.009*** (0.002)</td>
</tr>
<tr>
<td>( Parent_{Inc_i} )</td>
<td>0.013*** (0.001)</td>
<td>-0.002*** (0.000)</td>
</tr>
<tr>
<td>( Parent_{Edu_i} )</td>
<td>0.334*** (0.024)</td>
<td>-0.041*** (0.008)</td>
</tr>
<tr>
<td>( Own_{Score_i} )</td>
<td>0.039*** (0.002)</td>
<td>-0.006*** (0.000)</td>
</tr>
<tr>
<td>( Awards_i )</td>
<td>0.872*** (0.003)</td>
<td>0.023*** (0.001)</td>
</tr>
<tr>
<td>( Girl_i )</td>
<td>-0.000</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>( Constant )</td>
<td>0.792*** (0.004)</td>
<td>0.206*** (0.004)</td>
</tr>
<tr>
<td>Obs.</td>
<td>914</td>
<td>914</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.797</td>
<td>0.654</td>
</tr>
</tbody>
</table>

\( P^{TT=BR}_i \): probability that truth-telling is a best response.
\( Gain_i \): utility gain if changing from always truth-telling to best responding.
Elementary school fixed effects are included in all regressions.
Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
4.2.3 Who Strategizes Better?

Table 12 reports how family background affects their deviation from best responding. I focus on two lists, the most under-used list (1, 3, 2, 4) and the most over-used (2, 3, 1, 4). Neither of them is bad \textit{ex ante} since they both rank a popular school top and a safe one second. However, parents are overcautious. They choose (2, 3, 1, 4) too often – 18.50 percentage points more often than what best responding parents would do, and the list, (1, 3, 2, 4), is under-used by 30.20 percentage points.

I regress the deviations on 5 sets of regressors with or without controlling incentive measures and/or attention measures (columns 1-5 in Table 12). For the most under-used list, without controlling for incentive and attention measures, wealthier and more educated parents and students with better achievements play (1, 3, 2, 4) more often. However, family background cannot eliminate the underutilization of the strategy – increasing Parent\_Inc\_i and Parent\_Edu\_i at the same time by two standard deviations would only reduce the deviation by 11.88 percentage point.

When including incentive measures, other variables become insignificant, although Own\_Score\_i remains significantly positive. \(P_{i}^{TT=BR}\) reduces the deviation, Gain\_i increases the deviation, while \(P_{i,k}^{TT}\) (the probability that (1, 3, 2, 4) is the true preference order) also increases the deviation marginally. When adding attention measures, the coefficients on these three variables are insignificant, and other coefficients do not change.

This implies that family background offsets some underutilization of the strategy because wealthier and more educated parents’ true preference order is more likely to be a best response (higher \(P_{i}^{TT=BR}\)). It is not because they are more sophisticated, as they respond to incentives in the same way as others.\(^{23}\) Indeed, when deviations from truth-telling is regressed on family background, wealthier and more educated parents’ report true preferences at a similar or marginally higher rate (Table 3 in appendix).

The last 5 columns in Table 12 show the results for the most over-used list (2, 3, 1, 4).

\(^{23}\)If Gain\_i is interacted with Parent\_Inc\_i and Parent\_Edu\_i, the two terms have insignificant coefficients – negative and positive respectively. Separately, if \(P_{i}^{TT=BR}\) is interacted with Parent\_Inc\_i and Parent\_Edu\_i, the two terms are insignificant as well – positive and negative respectively.
Table 12: Who Strategizes Better: Regression Analysis of Sophistication Measures

<table>
<thead>
<tr>
<th></th>
<th>Most Under-Used List: (1,3,2,4)</th>
<th></th>
<th>Most Over-Used List: (2,3,1,4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Deviation from Best-Responding Prediction</td>
<td></td>
<td>Deviation from Best-Responding Prediction</td>
</tr>
<tr>
<td></td>
<td>Depend. V – Mean: -0.302; Std Dev: 0.235</td>
<td></td>
<td>Depend. V – Mean: 0.185; Std Dev: 0.406</td>
</tr>
<tr>
<td>Parent_Inci</td>
<td>0.026*** 0.001 -0.001 -0.003 -0.004</td>
<td></td>
<td>-0.012 0.022 0.019 0.032 0.028</td>
</tr>
<tr>
<td>(0.008) (0.010) (0.010) (0.012) (0.011)</td>
<td></td>
<td>(0.021) (0.027) (0.021) (0.029) (0.021)</td>
<td></td>
</tr>
<tr>
<td>Parent_Edui</td>
<td>0.017*** 0.008 0.008 0.005 0.005</td>
<td></td>
<td>-0.004 0.005 0.003 0.010 0.007</td>
</tr>
<tr>
<td>(0.004) (0.006) (0.006) (0.007) (0.006)</td>
<td></td>
<td>(0.007) (0.008) (0.007) (0.008) (0.008)</td>
<td></td>
</tr>
<tr>
<td>Own_Scorei</td>
<td>0.389*** 0.242*** 0.310*** 0.278*** 0.364***</td>
<td></td>
<td>-0.352 -0.101 -0.243 0.028 -0.166</td>
</tr>
<tr>
<td>(0.088) (0.092) (0.081) (0.106) (0.096)</td>
<td></td>
<td>(0.223) (0.248) (0.222) (0.268) (0.237)</td>
<td></td>
</tr>
<tr>
<td>Awardsi</td>
<td>0.037*** 0.013 0.016 0.009 0.013</td>
<td></td>
<td>-0.028** 0.001 -0.011 0.007 -0.008</td>
</tr>
<tr>
<td>(0.009) (0.015) (0.013) (0.017) (0.014)</td>
<td></td>
<td>(0.012) (0.017) (0.013) (0.018) (0.014)</td>
<td></td>
</tr>
<tr>
<td>Girli</td>
<td>-0.001 -0.001 -0.000 -0.007 -0.006</td>
<td></td>
<td>0.023 0.023 0.022 0.041 0.041</td>
</tr>
<tr>
<td>(0.014) (0.014) (0.014) (0.016) (0.016)</td>
<td></td>
<td>(0.027) (0.027) (0.027) (0.028) (0.028)</td>
<td></td>
</tr>
<tr>
<td>$P_{TT=BR}^i$</td>
<td>0.482*** 0.563***</td>
<td></td>
<td>-0.880* -1.189***</td>
</tr>
<tr>
<td>(0.188) (0.206)</td>
<td></td>
<td>(0.482) (0.520)</td>
<td></td>
</tr>
<tr>
<td>$Gain_i$</td>
<td>-2.379*** -2.731***</td>
<td></td>
<td>4.631*** 6.176***</td>
</tr>
<tr>
<td>(0.667) (0.741)</td>
<td></td>
<td>(2.104) (2.272)</td>
<td></td>
</tr>
<tr>
<td>Attn_Ui</td>
<td>-0.006 -0.006</td>
<td></td>
<td>0.039** 0.039**</td>
</tr>
<tr>
<td>(0.012) (0.012)</td>
<td></td>
<td>(0.017) (0.017)</td>
<td></td>
</tr>
<tr>
<td>Attn_Qi</td>
<td>-0.002 -0.002</td>
<td></td>
<td>-0.026 -0.026</td>
</tr>
<tr>
<td>(0.017) (0.017)</td>
<td></td>
<td>(0.035) (0.035)</td>
<td></td>
</tr>
<tr>
<td>Attn_Othersi</td>
<td>-0.008 -0.008</td>
<td></td>
<td>0.018 0.018</td>
</tr>
<tr>
<td>(0.006) (0.006)</td>
<td></td>
<td>(0.012) (0.012)</td>
<td></td>
</tr>
<tr>
<td>$P_{TT}^i$</td>
<td>-0.510 -0.677* -0.603 -0.781*</td>
<td></td>
<td>-1.788 -3.317 -3.153 -5.060</td>
</tr>
<tr>
<td>(0.372) (0.385) (0.399) (0.413)</td>
<td></td>
<td>(2.856) (3.022) (3.009) (3.110)</td>
<td></td>
</tr>
<tr>
<td>Obs.</td>
<td>914 914 914 810 810</td>
<td></td>
<td>914 914 914 810 810</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.154 0.161 0.164 0.172 0.175</td>
<td></td>
<td>0.048 0.053 0.055 0.060 0.063</td>
</tr>
</tbody>
</table>

Definitions of $P_{TT=BR}^i$ and $Gain_i$ are in Table 11. $P_{TT}^i$ is the probability of the list (in the dependent variable) being true preference order. When only adding $P_{TT}^i$ in the first regressions for each list, the coefficients remain similar, and the coefficient of $P_{TT}^i$ is similar to the one in other regressions. Elementary school fixed effects are included. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Without controlling incentive and attention measures, only $Awards_i$ has a significant effect. The two incentive measures still have an significant effect. Similar to the previous results, $P_{i}^{TT=BR}$ reduces the deviation, $Gain_i$ extends the deviation, and $P_{i,k}^{TT}$ has negative but insignificant coefficients. Surprisingly, attention on uncertainty, $Attn\_U_i$, increases the deviation (significant at 5%).

The same regressions are run for all other lists as well. The coefficients on family background and incentive and attention measures are mostly insignificant. For the most likely true preference order, $(1, 2, 3, 4)$, results are presented in Table A-4 in appendix.

As a placebo test, the same regressions are run for the one-school lists and $(0, 0, 0, 0)$. By assumption, nobody makes mistakes when playing these strategies, and thus all the coefficients should be zero. Indeed, very few of them are significant.

To summarize, these results shows that on average, parents are overcautious, as they rank School 1 low more often than they should. Paying more attention on uncertainty does not help, and sometimes exacerbate the overcautiousness. Wealthier and more educated parents are not more sophisticated, but they do slightly better because they their true preferences happen to be best responses more often.

5 Counterfactual Analyses

To analyze the welfare effect of replacing the Boston by the DA mechanism, I consider two experiments: (i) I take the empirical beliefs as the equilibrium beliefs, or equivalently, other parents behave as in the data; and (ii) parents are either naive or sophisticated.

5.1 Simulating Outcomes under the Two Mechanisms

Every parent reports her true preferences in the DA. The probabilities of being assigned to each school when submitting any list are obtained by drawing 20,000 profiles of prefer-

---

24 Among the 25 regressions, Parent Inc$_i$ and Parent Edu$_i$ never have a significant coefficient; Attn$_U_i$, $P_{i}^{TT=BR}$, and Gain$_i$ are significant only in 2 regressions; and $P_{i,k}^{TT}$ has 6 significant coefficients out of 25.
ences rankings, simulating their outcomes, and weighing the outcomes by the probability of obtaining each profile.

Under the Boston mechanism, the equilibrium outcome needs to be solved in the second experiment. Eleven cases with 0, 10%, 20%, ..., or 100% naive parents are considered. Which group of parents are naive in each case is randomly chosen, and the others are sophisticated. A naive parent always report her true preferences, while a sophisticated one plays a best response as if she knows the joint distribution of others’ preferences and beliefs. In the first ten cases, the equilibrium beliefs are solved by finding a fixed point as described in Appendix A.5. When all the parents are naive, the probabilities of being assigned to each school are simulated similar to that in the DA.

After solving the equilibrium, I simulate parents’ behavior and use the equilibrium beliefs to calculate their welfare. For each parent, 1,000 profiles of preferences are constructed by drawing the errors. Each of them plays two types of games – the Boston mechanism and the DA, and in the Boston mechanism, parents play each case of the Boston mechanism being sophisticated and then being naive.

5.2 Parents Behave as in the Data

The first experiment considers the equilibrium in the data, so the empirical beliefs, \( \hat{B} \), are taken as the equilibrium beliefs. The following results measure the effect when changing the Boston to the DA in Beijing, while parents behave as they do in the data under the Boston mechanism. Table 13 reports the results.

Switching to the DA hurts both naive and sophisticated parents on average (columns (1) and (2)), and the utility loss is sizable – for naive parents, the average utility loss is equivalent to increasing the distance to a school by 183.9%; 218.2% for the sophisticated. Regression results show that this loss is smaller for wealthier and more educated parents.

The utility loss implicitly assumes inter-personal welfare comparison, so in columns (3)-(8) I consider the probabilities of being better off, indifferent, or worse off. On average, 20.5% of naive or sophisticated parents achieve the same welfare in either mechanism, as
Table 13: Welfare Effects of Replacing the Boston with the DA: Regression Analyses Given the Observed Equilibrium

<table>
<thead>
<tr>
<th></th>
<th>Mean Utility Diff&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Prob(Better off in DA)</th>
<th>Prob(Indiff.)&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Prob(Worse off in DA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean(Dep V)</td>
<td>-0.265</td>
<td>-0.294</td>
<td>0.080</td>
<td>0.062</td>
</tr>
<tr>
<td>S.D(Dep V)</td>
<td>(0.059)</td>
<td>(0.071)</td>
<td>(0.049)</td>
<td>(0.035)</td>
</tr>
<tr>
<td>Parent&lt;sub&gt;Inc&lt;/sub&gt;&lt;sub&gt;i&lt;/sub&gt;</td>
<td>0.029***</td>
<td>0.033***</td>
<td>-0.013***</td>
<td>-0.007***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Parent&lt;sub&gt;Edu&lt;/sub&gt;&lt;sub&gt;i&lt;/sub&gt;</td>
<td>0.008***</td>
<td>0.011***</td>
<td>-0.006***</td>
<td>-0.004***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Own&lt;sub&gt;Score&lt;/sub&gt;&lt;sub&gt;i&lt;/sub&gt;</td>
<td>0.236***</td>
<td>0.277***</td>
<td>-0.128***</td>
<td>-0.078***</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.018)</td>
<td>(0.020)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Awards&lt;sub&gt;i&lt;/sub&gt;</td>
<td>0.023***</td>
<td>0.029***</td>
<td>-0.016***</td>
<td>-0.012***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>Girl&lt;sub&gt;i&lt;/sub&gt;</td>
<td>-0.001</td>
<td>-0.002</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.275***</td>
<td>-0.298***</td>
<td>0.063***</td>
<td>0.049***</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>Obs</td>
<td>914</td>
<td>914</td>
<td>914</td>
<td>914</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.786</td>
<td>0.850</td>
<td>0.510</td>
<td>0.472</td>
</tr>
</tbody>
</table>

In the DA, everyone is truth-telling. In the Boston, naive parents are truth-telling, and sophisticated ones play best response given the empirical beliefs, $B$.

a. The mean utility difference is defined as the average expected utility obtained in the DA minus the one obtained in the Boston.

b. The probability of being indifferent in the Boston and in the DA is the same for naive and sophisticated parents, because these are parents who do not participate at all in either mechanism.

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

they do not participate in either mechanism. This probability is positively correlated with family background, because the outside option increases with it.

Among naive parents, only 8.0% are better off in the DA, while, surprisingly, 71.5% are worse off. Among sophisticated parents, 6.2% are better off and 73.3% are worse off. For any parent, family background reduces probabilities of being better off and of being worse off, but it reduces the latter more quickly. This is again due to the outside option.

5.3 Equilibrium with only Naive and Sophisticated Parents

The second experiment considers that there are only two types of parents – naive and sophisticated, while the percentage of naive parents ranging from 0%, 10%, to 100%. Given the equilibrium, every parent behaves naively and then sophisticatedly in the Boston.
In Figure 1, similar to the first experiment, I report the average expected utility change and probabilities of being better off, indifferent, and worse off. Overall, the results are not sensitive to the fraction of naive parents.

![Figure 1: Welfare Effects of Replacing the Boston with the DA: Naive and Sophisticated Parents](image)

In all simulations, parents play the DA and then the Boston. In the Boston, every parent is always truth-telling when she is naive, while always best-responding when being sophisticated. The equilibrium is solved with fractions of naive parents from 0% to 100%. Given each equilibrium, all parents behave naively and then sophisticatedly to calculate the welfare effect.

The first subfigure shows average expected utilities in the DA and in the Boston. For sophisticated parents, their utility loss is on average 0.165 which is equivalent to increasing the distance to a school by 91.29%. However, different from the previous experiment, the naive parents on average have a utility gain in the DA, although the gain only amounts to

\[ \text{The seemingly discontinuity from 90\% to 100\% naive parents may be due to computation errors, as one has to solve for a fixed point in the former while there is no need to do so in the latter.} \]
decreasing the distance to a school by 6.74%.

The probability of being better off is about 44% for naive parents, and more importantly, about 35% for them being hurt by the DA (subfigure 2). For sophisticated ones, only 11% are better off, while about 70% are worse off (subfigure 3).

The different results from the two experiments highlight the significance of parents’ overcautiousness. Given the severe overcautiousness, being naive actually helps, as it offsets some of the deviation from best responses.

6 Concluding Remarks

This paper uses a data set from Beijing to answer two questions: (i) whether poor and/or less educated parents are more likely to be naive and (ii) whether the Boston harms naive parents relative to the DA.

Assuming that students’ preferences are private information, I model school choice under the Boston mechanism as a simultaneous game of incomplete information. Due to the lack of strategy-proofness, submitted choice lists are not necessarily true preferences. While allowing parents to make mistakes in the game, I derive a set of dominated strategies and assume they are not played in equilibrium. I group some lists together and characterize choice probabilities. A simulated maximum likelihood method is used for the estimation.

Results reject the hypothesis that everyone is naive and also reject that everyone is sophisticated. Parents are overcautious in the sense that they avoid top ranking the most popular school too often than what they should do if they play best responses. Their income and education offset the overcautiousness partially. However, this is because wealthier/more educated parents’ true preference order is more likely to be a best response, while there is no evidence of them being more sophisticated. These findings are driven by the fact that wealthier/more educated parents have a better outside option.

Poorer parents pay more attention to uncertainty in the game, and this indicates they try hard to find a best response. However, paying more attention to uncertainty does not help
and sometimes worsen their overcautiousness.

Given parents’ behavior, especially their overcautiousness, in the data, when replacing the Boston mechanism by the DA, both naive and sophisticated parents suffer a utility loss which on average roughly amounts to tripling the distance to a school. For naive parents, only 8% of them are better off in the DA, while 71.5% are worse off. The negative effects are larger for sophisticated parents, and the effects decrease with parents’ income and education because of the outside option.

If every parent is either sophisticated or naive and no one is overcautious, switching from the Boston to the DA has mixed effects. For sophisticated parents, their utility loss is equivalent to increasing the distance to a school by 91.29%. Among them, only 11% are better off, while about 70% are worse off. Naive parents on average have a utility gain in the DA, although the gain is only equivalent to decreasing the distance to a school by 6.74%. The probability of being better off is about 44% for them, but about 35% of them are hurt by the DA.

These results suggest that instead of replacing the Boston by the DA, to improve welfare, it may be more effective if one can help parents find best responses. This is especially important for poorer/less educated parents who have no good outside option. How to help parents find best responses is therefore a worthwhile future research topic.
References


A Appendices

A.1 Proofs

Proof of Lemma

(i) Suppose a participating student submits a full list, and she is rejected by all her choices but the last choice ($s^*$). Then in Round $S$, school $s^*$ must have more available seats than students unassigned. If the total seats left at $s^*$ is $\bar{q}_{s^*}$, then the number of students unassigned is $I - \sum_s q_s + \bar{q}_{s^*}$. Thus, the student must be assigned to her last choice.

Suppose a participating student submits a partial list and is rejected by all the schools in her list. After at most $S$ rounds, she is still unassigned. Since $I \geq \sum_s q_s$, at that point, the number of available seats is at least the number of remaining students, and thus she will be assigned to some school.

(ii) Suppose $C$ and $C'$ have the same first $K$ choices. In any realization of the game (any lottery number), if the student is assigned to one of the first $K$ choices when submitting $C$, she will be assigned to that school when submitting $C'$. If she is not assigned to a school in the first $K$ schools when submitting $C$, she will not be assigned to that school if submitting $C'$ instead. This means she has the same probability to be assigned to any of the first $K$ choices when she submits $C$ or $C'$.

(iii) Suppose $C$ and $C'$ have the same first $K' - 1$ choices. School $s$ is listed as $K$th choice in $C$, but as $K'$th choice in $C'$ and $K' < K$. In any realization of the game (any lottery number), if the student is rejected by $s$ when submitting $C'$, she will not be accepted by $s$ if submitting $C$. If she is accepted by $s$ when submitting $C$, in the same realization of the game, school $s$ has more available seats than applying students in Round $K'$. Thus, she will be assigned to $s$ for sure if submitting $C'$. Moreover, there are cases that $s$ is available in Round $K'$ but not in Round $K$. This implies the probability of being assigned to $s$ weakly increases when moving it toward the top of the list. In the same manner, including an otherwise omitted school in the list has the same effect.

(iv) The number of students listing $s$ as first choice is at most $I$. Since a lottery number is used to determine who will be accepted, among those who have the same first choice, everyone have the same probability being accepted by that school. The probability of being accepted by $s$ is at least $q_s/I$ if a student list $s$ as first choice.

Proof of Proposition

To show the existence of a symmetric equilibrium, I make use of the results in Schmeidler (1973) by reconstructing the Bayesian game into a game of complete information with a nonatomic continuum of players. Theorem 1 in Schmeidler (1973) establishes the existence.

Within the current incomplete information setting, each player $i$ is facing $(I - 1)$ players without knowing their types. Given the distribution of $\{u_{i,s}\}_{s=1}^S$ being common knowledge, it is equivalent to say that $i$ is playing against a continuum of players each of whom is of type $\{u_{j,s}\}_{s=1}^S$, $j \neq i$.

More formally, the game of school choice can be re-written into a game of complete information where the set of players $T$ is $R^S$ endowed with a measure $\mu$ such that for any measurable set $\bar{T} \subset T$,

$$\mu(\bar{T}) \equiv (I - 1) \int 1\left(\{u_s\}_{s=1}^S \in \bar{T}\right) dG(X, Z) dF_{\Sigma}(\varepsilon),$$

51
where each player is indexed by \( \{ u_s \}_{s=1}^S \in \mathbb{R}^S \).

With some abuse of notation, now define a strategy \( \sigma \) as a measurable function from \( T \) to \( \Delta (C) \). The payoff to player \( \{ u_s \}_{s=1}^S \) is

\[
V \left[ \sigma \left( \{ u_s \}_{s=1}^S \right), \sigma \right] = \sum_{s=1}^S A_s \left[ \sigma \left( \{ u_s \}_{s=1}^S \right), \sigma \right] \max (u_s, 0),
\]

where, with \( C^m = (C_1^m, \ldots, C_m^m, \ldots, C_{(I-1)}^m) \),

\[
A_s \left[ \sigma \left( \{ u_s \}_{s=1}^S \right), \sigma \right] = \sum_{n=1}^{L^{(I-1)}} \left\{ \Pr \left[ C_m^m \text{ is played under } \sigma \left( \{ \check{u}_s \}_{s=1}^S \right) \right] d\mu \right\} a_s \left( \sigma \left( \{ u_s \}_{s=1}^S \right), C^m \right).
\]

It can be verified that the above notations are equivalent to the original ones given the independent types across parents, and \( V \left[ \sigma \left( \{ u_s \}_{s=1}^S \right), \sigma \right] \) is continuous in \( \sigma \).

To apply Schmeidler’s theorem, one need to show that \( \forall C, C' \in C \) the following set is measurable,

\[
\left\{ \{ u_s \}_{s=1}^S \in R^S | V [C, \sigma] > V [C', \sigma] \right\},
\]

where \( V [C, \sigma] > V [C', \sigma] \) is equivalent to

\[
\sum_{s=1}^S \left\{ A_s [C, \sigma] - A_s [C', \sigma] \right\} \max (u_s, 0)
\]

\[
= \sum_{n=1}^{L^{(I-1)}} \left\{ \int_{\{ \check{u}_s \}_{s=1}^S} \Pr \left[ C_m^m \text{ is played under } \sigma \left( \{ \check{u}_s \}_{s=1}^S \right) \right] d\mu \right\} \times \left[ a_s \left( C, C^m \right) - a_s \left( C', C^m \right) \right] \max (u_s, 0)
\]

\[
> 0,
\]

which is linear in \( \{ u_s \}_{s=1}^S \). The above set is therefore measurable. By Schmeidler’s Theorem 1, an equilibrium always exist.

(i) From \( i \)’s perspective, for any \( j \neq i \), \( \Pr (u_{j, s} > 0 > u_{j, s'}) \), given \( s \) \& \( \forall s' \neq s \) > 0 given the continuous distribution assumption on \( \varepsilon_j \). In this case, since \( j \) only has one acceptable school \( (s) \), given any equilibrium beliefs, the best response for \( j \) is rank \( s \) at top. Therefore, \( \Pr (s \text{ is top ranked by all } j \neq i) > 0 \) which implies that \( A_s (C, \sigma_{-i}) < 1 \) for all \( C \) such that \( s \) is top ranked.

Now suppose that \( A_s \left( C', \sigma_{-i}^* \right) = 1 \) for some \( C \) such that \( s \) is not top ranked, there must be \( A_s \left( C', \sigma_{-i}^* \right) = 1 \) such that \( s \) is top ranked in \( C' \) (from Lemma 1). Therefore, \( A_s \left( C, \sigma_{-i} \right) < 1 \) for all \( s \) and for all \( C \neq (0, \ldots, 0) \).

Similarly, conditional on being rejected by previous choices, the probability of being accepted by \( s \) is less than one, unless \( s \) is the \( S \)th (the last) choice. Suppose that \( C = (c^1, \ldots, c^K, s, c^{K+1}, \ldots, c^S) \), where \( 1 \leq K \leq (S - 2) \) and \( c^k \neq 0, \forall k = 1, \ldots, K \). Then there is a strictly positive probability that (i) \( q_{c^1} + 1 \) students’ preferences are such that \( c^1 \) is the only acceptable school and (ii) \( q_{c^k} \) students’ preferences are such that \( c^k \) is the only acceptable school, for \( k \in \{2, \ldots, K, s\} \). This is true because \( \sum_{k=1}^K q_{c_k} + q_s + 1 \leq I \) which is implied by \( \sum_{s=1}^S q_s \geq I \) and \( \sum_{s=1}^S q_s - q_{s'} < I \), for
any $s'$. 

For those students, their best response is to rank their only acceptable school first, and therefore, there is a positive probability that a student who submits $C$ is rejected by $s$ conditional on she is rejected by $c_1, \ldots, c^K$ as well.

Since $\sum_{s=1}^{S} A_s (C, \sigma_{-i}^s) = 1$, the above two results implies that $A_s (C, \sigma_{-i}^s) > 0$ for all $s$ and $C \neq (0, 0, 0, 0)$.

(ii) Note that in equilibrium $i$'s value function is $V_i (\sigma_i^+, \sigma_{-i}^-) = \sum_{s=1}^{S} A_s (\sigma_i^+, \sigma_{-i}^-) \max (u_{i,s}, 0)$. From above, $A_s (C, \sigma_{-i}^-) \in (0, 1)$ for all $s$ and for all $C \neq (0, \ldots, 0)$. If $C_i^+ \neq C_i^{**} \neq (0, \ldots, 0)$ are played with positive probability in $\sigma_i^+$ given $\{u_{i,s}\}_{s=1}^{S}, (X_i, Z_i, \varepsilon_i)$, then $V_i (C_i^+, \sigma_{-i}^-) = V_i (C_i^{**}, \sigma_{-i}^-)$, or,

$$
\sum_{s=1}^{S} [A_s (C_i^+, \sigma_{-i}^-) - A_s (C_i^{**}, \sigma_{-i}^-)] \max (u_{i,s}, 0) = 0. 
$$

Since $C_i^+ \neq C_i^{**}$, there is at least 2 schools such that $A_s (C_i^+, \sigma_{-i}^-) \neq A_s (C_i^{**}, \sigma_{-i}^-)$, as a result of part (iv) in Proposition 2.

Since $i$ has at most one unacceptable school and $u_{i,s} \neq u_{i,s'}$ for all $s \neq s'$ with probability one, equation (5) holds ex ante with probability zero. This proves that $i$ plays mixed strategies with probability zero.

(iii) The argument in (ii) implies that if $i$ plays a mixed strategy, $i$ has at least two unacceptable schools with probability one.

Suppose that $C_i^+ = (c_1^+, \ldots, c_S^+)$ is played with positive probability in $\sigma_i^+$, and that $\exists k \in \{1, \ldots, (S - 2)\}$, $u_{i,c_k^+} < 0$ and $u_{i,c_{k+1}^+} > 0$. By applying results in Lemma 1 and (i), one may show that $C_i^+$ must be strictly dominated by excluding the unacceptable school from the list. Therefore, any $C_i^+$ must exclude or include the unacceptable schools at the bottom. \hfill \textsection

**Proof of Proposition 2**

(i) Suppose the first choice in list $C$ is unacceptable, or worse than the outside option. Construct a new list, $C'$, such that the first school is the most preferred school and all other choices in $C'$ are the same as $C$. Then given any realization of the game (any lottery number and any profile of other players lists), if the student is accepted by an acceptable school when submitting $C$, she will be either accepted by the most preferred school or that school. She is weakly better off in any realization. And there must exist cases such that she is matched with the first choice in $C$ when submitting $C$, while she will be matched with the most preferred school when submitted $C'$ instead. Thus, $C$ is strictly dominated by $C'$.

In the same manner, if the first choice in $C$ is the worst school, $C$ is dominated by $C'$ which is the same as $C$ except the first choice in $C'$ is replaced by the most preferred school.

(ii) Since including an otherwise omitted school always weakly increases the probability of being accepted by that school (Lemma 2), adding the acceptable school after the last choice of a partial list, always weakly improves the expected utility. If there are multiple acceptable schools are omitted, adding them to the list from the best to the worse will also weakly improves the expected utility.

(iii) Suppose the submitted list of $i$ is $C = \{c_1, \ldots, c_S\}$ such that $c^K = \hat{s}$ which is worst school, $1 \leq K < S$, such that $u_{i,\hat{s}} = \min_{t=1,\ldots,S} \{u_{i,t}\}$ and $\exists t \in \{K + 1, \ldots, S\}$ s.t., $u_{i,c_t} > 0$. Consider
an alternative list, \( C' = \{ c_1, \ldots, c_{K-1}, s_i, c_{K+1}, \ldots, c_S \} \), where \( u_{i,s} = \max_{t=1,\ldots,S} \{ u_{i,t} \} > 0 \), i.e., replace the worst school with the best one.

Given any realization of the game, if the student is accepted by any school of \( c_1, \ldots, c_{K-1} \) when submitting \( C \), she will be still accepted by that school when submitting \( C' \) instead. By Lemma 2, \( A_{i,s}(C, \sigma_{-i}^*) \geq A_{i,s}(C', \sigma_{-i}^*) \), and the decrease in the probability is distributed to \( s_i, c_{K+1}, \ldots, c_S \) and \( s \) as well. Since \( u_{i,s} = \min_{t=1,\ldots,S} \{ u_{i,t} \} \) and \( u_{i,c^*_i} > 0 \), \( C' \) weakly improves \( i \)'s expected utility. Similar arguments can be made if \( u_{i,s} < 0 \).

(iv) The strict increase can be seen easily in terms of conditional probabilities. \( A_{i,s}(C, \sigma_{-i}^*) \)
eq \begin{align*}
\Pr(i \text{ is rejected by schools ranked above } s \text{ in } C|\sigma_{-i}^*) \\
\ast \Pr(i \text{ is accepted by } s \text{ given } C|\sigma_{-i}^*; i \text{ is rejected by schools ranked above } s)
\end{align*}
If \( A_{i,s}(C, \sigma_{-i}^*) \in (0, 1), \forall s \) and \( \forall C \neq (0, \ldots, 0) \), then the above two terms are both in \((0, 1)\) unless \( s \) is ranked as \( s \)th after all other schools.

Suppose that \( s \) is ranked as \( k \)th choice in \( C \) and ranked as \( k' \)th in \( C' \), where \( 1 \leq k' < k \leq S \) and the 1st to \( (k' - 1) \)th choices are the same in both lists. We want to show that \( A_{i,s}(C', \sigma_{-i}^*) > A_{i,s}(C, \sigma_{-i}^*) \).

From Lemma 2, \( A_{i,s}(C', \sigma_{-i}^*) \geq A_{i,s}(C, \sigma_{-i}^*) \), if \( A_{i,s}(C', \sigma_{-i}^*) = A_{i,s}(C, \sigma_{-i}^*) \), then

\[ \Pr(i \text{ is rejected by schools ranked above } s \text{ in } C|\sigma_{-i}^*) \\
\ast \Pr(i \text{ is accepted by } s \text{ given } C|\sigma_{-i}^*; i \text{ is rejected by schools ranked above } s) \\
= \Pr(i \text{ is rejected by schools ranked above } s \text{ in } C'|\sigma_{-i}^*) \\
\ast \Pr(i \text{ is accepted by } s \text{ given } C'|\sigma_{-i}^*; i \text{ is rejected by schools ranked above } s). \]

We would also have

\[ \Pr(i \text{ is rejected by schools ranked above } s \text{ in } C|\sigma_{-i}^*) < \Pr(i \text{ is rejected by schools ranked above } s \text{ in } C'|\sigma_{-i}^*), \]

because the probability of being rejected by schools ranked between \((k' + 1)\)th to \((k - 1)\)th in \( C \) is positive. Otherwise, it would imply \( A_{i,s}(C, \sigma_{-i}) = 0 \) for some \( C \) and \( s \).

Together, the above equality and inequality imply that

\[ \Pr(i \text{ is accepted by } s \text{ given } C|\sigma_{-i}^*; i \text{ is rejected by schools ranked above } s) > \Pr(i \text{ is accepted by } s \text{ given } C'|\sigma_{-i}^*; i \text{ is rejected by schools ranked above } s) \]

which is impossible because \( s \) ranked higher in \( C' \) than it is in \( C \).

This proves that \( A_{i,s}(C', \sigma_{-i}^*) > A_{i,s}(C, \sigma_{-i}^*) \). Similarly, the same is true if \( s \) is otherwise omitted from the list.

Since now the increase is always strict, it is straightforward to construct the strict dominance for (ii) and (iii).

**Proof of Lemma 3**
For parent $i$, with equilibrium beliefs, $B_i(C, \sigma_{-i}^*) \equiv (A_{i,1}(C, \sigma_{-i}^*), \ldots, A_{i,S}(C, \sigma_{-i}^*)) \in [0,1)^S$, where $A_{i,1}(C, \sigma_{-i}^*) \in [0,1)$ if $C$ has School 1 as 3rd or 4th choice and $A_{i,S}(C, \sigma_{-i}^*) \in (0,1)$ otherwise.

If School 1 is the worst or an unacceptable school for $i$, the lemma satisfies trivially given the results in Proposition 2.

Suppose School 1 is not the worst, and, without loss of generality, assume that $u_{i,2} > u_{i,3} > u_{i,4}$. Since there are at least three acceptable schools, one may transform $i$'s utilities into $\hat{u}_{i,s} \equiv \max (u_{i,s}, 0) - u_{i,3}$. Therefore, $\hat{u}_{i,2} > 0$, $\hat{u}_{i,3} = 0$, $\hat{u}_{i,4} < 0$, and $\hat{u}_{i,4} < \hat{u}_{i,1} = u_{i,1} - u_{i,3}$. Therefore, parent $i$'s problem is equivalent to:

$$\max_{C} \sum_{s=1}^{S} A_{i,s}(C, \sigma_{-i}^*) \hat{u}_{i,s}.$$ 

Denote further that $A_{i,1}(C^k, \sigma_{-i}^*) \hat{u}_{i,1} + V^k(C^k)$ as the value function when School 1 is fixed as $k$th choice and $C^k$ is optimally chosen to maximize the expected utility.

The necessary and sufficient conditions for School 1 not being ranked as top 2 choices are, for $k = 1, 2$,

$$A_{i,1}(C^3, \sigma_{-i}^*) \hat{u}_{i,1} + V^3(C^3) > A_{i,1}(C^k, \sigma_{-i}^*) \hat{u}_{i,1} + V^k(C^k),$$

$$A_{i,1}(C^4, \sigma_{-i}^*) \hat{u}_{i,1} + V^4(C^4) > A_{i,1}(C^k, \sigma_{-i}^*) \hat{u}_{i,1} + V^k(C^k);$$

or equivalently,

$$\hat{u}_{i,1} < \min \left\{ \frac{V^3(C^3) - V^1(C^1)}{A_{i,1}(C^1, \sigma_{-i}^*) - A_{i,1}(C^3, \sigma_{-i}^*)}, \frac{V^3(C^3) - V^2(C^2)}{V^4(C^4) - V^3(C^4)}, \frac{V^3(C^3) - V^2(C^2)}{V^4(C^4) - V^2(C^4)} \right\} \equiv \bar{u}_i.$$

where all the denominators are positive by Lemma 2 and Proposition 1.

It needs to be shown that the minimum operation always returns a non-negative value, or $V^k(C^k) - V^{k'}(C^{k'}) \geq 0$ for $k = 3, 4$ and $k' = 1, 2$. Recall that $A_{i,1}(C^k, \sigma_{-i}^*) \hat{u}_{i,1} + V^k(C^k)$ is defined as the value function when $k$th choice is School 1. Therefore, $V^k(C^k)$ is the part of the expected utility which is from schools other than School 1 while the $k$th choice is taken. Under Assumption ZERO-PROB, by Part (iv) of Proposition 2, moving $s \neq 1$ towards the top strictly increases the probability of being accepted by $s$. Therefore, the arguments similar to those in the proof of Proposition 2’s Part (iii) can be used to verify that $V^k(C^k) - V^{k'}(C^{k'}) \geq 0$ for $k = 3, 4$ and $k' = 1, 2$.

This proves that when $(u_{i,1} - u_{i,s}) < \bar{u}_i$, $i$ does not rank School 1 as top 2, and $\bar{u}_i$ is non-negative and is a function of $i$’s beliefs and preferences but not of $u_{i,1}$. ■
A.2 Data Cleaning and Likelihood Estimation

A.2.1 Data Cleaning and Imputation

The data set that I use in this paper is a subsample of the data set that has been used in Lai, Sadoulet, and de Janvry (2009), Lai (2010), and Lai, Sadoulet, and de Janvry (2011).

Students’ submitted lists are the most important variable. There are two types of "technical errors" among the lists as defined in Lai, Sadoulet, and de Janvry (2009): (i) repeated choice of a school and (ii) applying to schools not accessible to the neighborhood with the assignment system. Among the 914 students, there are 6 cases of type (i) error and 3 of type (ii) error.

The first 6 students submitted lists as follows: (2, 3, 3, 3), (2, 3, 1, 2), (2, 1, 3, 2), (1, 1, 0, 0)–two cases, and (1, 2, 2, 3). I replace these lists by (2, 3, 0, 0), (2, 3, 1, 4), (2, 1, 3, 4), (1, 0, 0, 0), and (1, 2, 3, 4) respectively. The replacement for the first 4 lists is straightforward, as they are payoff-equivalent in any realization of the game. Replacing (1, 2, 2, 3) by (1, 2, 3, 4) is because this student shows a preference of School 3 over School 4. The results do not change in a few cases that I have experimented when (1, 2, 2, 3) is replaced by (1, 2, 0, 0).

The second 3 cases are those who have submitted (2, 3, 1, s')–2 cases and (2, 1, s'', 4), where s' ≠ s'' ∈ {0, 1, 2, 3, 4}. I replace the first list by (2, 3, 1, 4), as they are always payoff equivalent. (2, 1, s'', 4) is either replaced by (2, 1, 0, 0) or (2, 1, 3, 4). (2, 1, 0, 0) is payoff equivalent in the observed play of the game. I also consider (2, 1, 3, 4) as an alternative because the code for School 3 in the application is 15, while the code for s'' is 25. Therefore, it is likely that (2, 1, s'', 4) is submitted or recorded as a typo. I present the results when (2, 1, s'', 4) is replaced by (2, 1, 3, 4).

The main explanatory variables are Distancei,s, Own_Scorei, Parent_Incomei, Parent_Edui, and Awardi.

Distancei,s measures the walking distance between i’s home address and school s, and both addresses are from 1999. I use the Chinese version of Google Maps, http://ditu.google.cn/, to get the walking distance. Students’ home addresses are from the administrative data, and there are 4 students missing home address. Their distances are assigned at the medians.

Own_Scorei is the sum of student i’s scores of Chinese and math in grade 6 which is the final year of elementary school. They scores are from administrative data, but there are 125 missing values. To impute, I follow the 3 steps: (i) I regress these test scores on their test scores from the two semesters of grade 7 controlling middle school and elementary school fixed effects, then I do the out-of-sample prediction. (ii) I run similar regressions but with Parent_Edu as main regressor and then do out-of sample predictions. (iii) The remaining missing values are replaced by the median.

Parent_Incomei is the sum of father’s and mother’s income. There are 108 missing values in father’s income and 100 in mother’s. Some of the missing values in Parent_Incomei are replaced by the households disposable income plus the average difference between Parent_Incomei and the disposable income. I regress their own income on different combinations of their own and their spouse’s education, political affiliations, and ages, the disability status (of either of them), and their spouse’s income, and then do out-of-sample prediction to further impute Parent_Incomei.

Parent_Edui is the average years of schools of parents. There are 49 missing values. I regress Parent_Edu on different subsets of the variables, Parent_Incomei, father’s and mother’s political affiliations, father’s and mother’s job stabilities, and the disability status (of either of them). Then I do out-of-sample predictions to impute Parent_Edui.

Awardi are calculated from 6 questions in the 2002 survey. These questions ask students if they
have received any awards at district level or above in 6 different categories during the six years of elementary study – all-round excellence, excellence in specific subjects, in science and technology contests, in arts and sports, in student leadership, and others. For the responses to each question, it takes one of the values of 0, 1, or missing (which is treated as 0).

A.2.2 Optimization in the Maximum Likelihood Estimation

The objective function, the negative of likelihood function in each case, is minimized mainly using Chris Sims Matlab "csminwel" algorithm which is a quasi-Newton method with BFGS update of the estimated inverse Hessian. Based on my experiments, it is much faster than other Matlab packages.

As discussed in details in Appendix A.3, I use a logit-smoothed accept-reject simulator to get the choice probabilities. Different values of scale factor, \( \lambda \), which determines the degree of smoothing are experimented: \( \lambda = 0.05, 0.025, 0.01, 0.005 \). For a given case, i.e., a given likelihood function, the following procedure is followed to get the estimates.

(i) For any given value of \( \lambda \), to avoid local optimum problem, I use Sims’ algorithm to minimize the objective function four times each of which has randomly chosen starting values. Among the four sets, I find the set of parameters which minimizes the objective function.

(ii) For a given \( \lambda \), I find the set of estimates minimizing the objective function among the 3 sets which are estimated given a different value of \( \lambda \). Using this set of parameter estimates as starting values, I minimize the objective function one more time for each given \( \lambda \). The algorithm used for \( \lambda = 0.01 \) is the "simulannealbnd" algorithm which is a simulated annealing algorithm canned in Matlab. For other values of \( \lambda \), Sims’ algorithm is used again.

(iii) For each \( \lambda \), I now have 5 sets of estimates from which I choose the one minimizing the objective function as the final estimates.

In the paper, I only report results from \( \lambda = 0.01 \), and results are similar for different values of \( \lambda \). All results are available upon request.

A.2.3 Simulated Maximum Likelihood with Equilibrium Constraints

For the estimation in the case of Bayesian Nash equilibrium, the realized play of the game is used to calculate the equilibrium beliefs. It is also tempting to consider an alternative approach, a (simulated) maximum likelihood with equilibrium constraints as follows:

\[
\max_{\theta, B} \sum_{i=1}^{I} \ln \Pr \left( C_i | X_i, Z_i, B; \theta \right), \quad s.t., \quad \overline{B} \left( C, \sigma_{-i}; B; \theta \right) = B, \quad (6)
\]

where the constraint restricts the beliefs to be a fixed point in equilibrium. \( \overline{B} \) is defined as the "beliefs" implied by \( B \), i.e., \( \overline{B} \left( C, \sigma_{-i}; B; \theta \right) = [A_1 \left( C, \sigma_{-i}; B; \theta \right), \ldots, A_S \left( C, \sigma_{-i}; B; \theta \right)] \), and \( A_s \left( C, \sigma_{-i}; B; \theta \right) = \sum_{n=1}^{L^{(i-1)}} \Pr \left( C_{n,i} \text{ played under } \sigma_{-i} \text{ given } B \text{ and } \theta \right) a_s \left( C, C_{n,i} \right) \). \( \overline{B} \) is also formulated by simulation which is described in Appendix A.5.

\[26\] The package is available at: http://sims.princeton.edu/yftp/optimize/
The problem with this approach is that given $\theta$, there might be multiple equilibria. In other words, the fixed point, $\overline{B}(\cdot, \sigma_{-i} | B; \theta) = B$, may not be unique for given $\theta$, although in many numerical computation procedures there might be a unique solution. I therefore do not consider this approach in the paper. Instead, the method is used to solve the Bayesian Nash equilibrium in counterfactual analyses.
A.3 Heterogeneous information on Students’ Characteristics

This appendix considers the case when some parents know more than others about other students/parents’ characteristics. Therefore, \((X_i, Z_i)\) is no longer private information.

**Proposition 3** Consider the following scenario:

(i) Every parent has the same ability to process information;
(ii) Parent \(i\) also knows the realization of \(X^i \equiv \{\tilde{X}_i, \ldots, \tilde{X}_{i,F}\}\), where \(\tilde{X}_i \equiv (X_i, Z_i)\), \(F\) is fixed and \(X^i\) may be different across parents.

Given the number of schools, as the number of parents becomes larger and the quotas grow at the same rate, the beliefs converge to a common belief, \(B_i (C) \rightarrow B (C), \forall i, \forall C \in C\).

**Proof of Proposition 3**

A student’s decision is to choose one of the \(L\) possible lists. Fix the order of all the lists, and let \(d_i = (d_{i,1}, \ldots, d_{i,L})^T\), and \(d_{i,l} = 1\) if the \(l\)th list is chosen by student \(i\) and \(d_{i,l} = 0\) otherwise. Thus, \(\sum_{l=1}^{L} d_{i,l} = 1\).

Without loss of generality, consider student 1’s decision and suppose \(X^1 = \{\tilde{X}_2, \ldots, \tilde{X}_{F+1}\}\). Her perceived probability of other students’ choices is a function of her information set – \(\tilde{X}_1\), and the distributions of \(\varepsilon_i, i > 1\).

Let \((\pi_{i,1}, \ldots, \pi_{i,L})\) be student 1’s belief about the probability that each list is being chosen by student \(i\). Given the continuous distribution of \(\varepsilon\), \(E (d_{i,l}) = \pi_{i,l} \in (0, 1)\) and \(\sum_{l=1}^{L} \pi_{i,l} = 1\) for all \(l\) and \(i\).

For \(i = 2, \ldots, F+1\), the beliefs are a function of the realization of \(X^1\),

\[
(\pi_{i,1}, \ldots, \pi_{i,L}) = (\pi_{i,1} (X^1), \ldots, \pi_{i,L} (X^1)).
\]

Given that \(\{\tilde{X}_i\}_{i=1}^I\) are i.i.d. across students, then \(\forall i = F + 2, \ldots, I\),

\[
(\pi_{i,1}, \ldots, \pi_{i,L}) \equiv (\bar{\pi}_1, \ldots, \bar{\pi}_L),
\]

which is not a function of \(X^1\).

Consider a vector of random variables, \(N^I \equiv \sum_{i=2}^{I} d_i = (N^I_1, N^I_2, \ldots, N^I_L) \in \mathbb{N}^L\), which are the numbers of students submitting each list, i.e.,

\[
n^I_l = \sum_{i=2}^{I} d_{i,l} \sum_{l=1}^{L} N^I_l = I - 1, N^I_l \geq 0 \text{ and } N^I_l \leq I - 1.
\]

In any realization of the game, \(N^I\) is a sufficient statistics to calculate the probability of being accepted by each school for Student 1, given the anonymity of the mechanism. Therefore, in the
following I focus on the distribution of $\mathbf{N}^I$. Two definitions are also introduced:

\[ \mu^I = \frac{1}{I-1} \left( \sum_{i=2}^{F+1} \pi_{i,1} + (I - F - 1) \pi_1, ..., \sum_{i=2}^{F+1} \pi_{i,L} + (I - F - 1) \pi_L \right), \]

\[ Q^I = \frac{1}{I-1} \left[ \begin{array}{cccc}
\sum_{i=2}^{F+1} \pi_{i,1} (1 - \pi_{i,1}) + (I - F - 1) \pi_1 (1 - \pi_1) & ... & \sum_{i=2}^{F+1} \pi_{i,1} \pi_{i,L} + (I - F - 1) \pi_1 \pi_L \\
\sum_{i=2}^{F+1} \pi_{i,1} \pi_{i,2} & ... & \sum_{i=2}^{F+1} \pi_{i,2} \pi_{i,L} \\
\sum_{i=2}^{F+1} \pi_{i,L} (1 - \pi_{i,L}) & ... & \sum_{i=2}^{F+1} \pi_{i,L} (1 - \pi_L)
\end{array} \right], \]

Consider the number of parents grows, i.e., $I \to \infty$,

\[ \lim_{I \to \infty} \mu^I = (\pi_1, ..., \pi_L) \equiv \mu, \]

\[ \lim_{I \to \infty} Q^I = \left[ \begin{array}{ccc}
\pi_1 (1 - \pi_1) & ... & \pi_1 \pi_L \\
\pi_1 \pi_2 & ... & \pi_2 \pi_L \\
... & ... & ... \\
\pi_1 \pi_L & ... & \pi_L (1 - \pi_L)
\end{array} \right] \equiv Q, \]

where $Q$ is a finite, positive definite matrix, since it is the covariance matrix for $d_i$, for $i > (F + 1)$.

To use the Multivariate Lindeberg–Feller Central Limit Theorem (see for example, Greene (1999), page 117), the following conditions are checked and are satisfied:

\[ \lim_{I \to \infty} \left[ (I - 1) Q^I \right]^{-1} \text{Var} (d_i) = \lim_{I \to \infty} \left( \sum_{j=2}^I \text{Var} (d_j) \right)^{-1} \text{Var} (d_i) = 0, \forall i = 2, ..., I. \]

Therefore,

\[ \sqrt{I-1} \left( \frac{\mathbf{N}^I}{I-1} - \mu^I \right) \overset{d}{\to} N(0, Q), \text{ as } I \to \infty. \]

Moreover, $\lim_{I \to \infty} \sqrt{I-1} (\mu^I - \mu) = 0$, and thus,

\[ \sqrt{I-1} \left( \frac{\mathbf{N}^I}{I-1} - \mu \right) \overset{d}{\to} N(0, Q), \text{ as } I \to \infty. \quad (7) \]

Similarly, when $\mathbf{X}_i$ is private information, with $\widetilde{\mathbf{N}}^I$ as the counterpart of $\mathbf{N}^I$, by the Multivariate Lindberg–Levy Central Limit Theorem, one can show that

\[ \sqrt{I-1} \left( \frac{\widetilde{\mathbf{N}}^I}{I-1} - \mu \right) \overset{d}{\to} N(0, Q), \text{ as } I \to \infty. \quad (8) \]
One need to prove that the sequences of random variables, $\sqrt{T-1} \left( \frac{N^I}{T-1} - \mu \right)$ and $\sqrt{T-1} \left( \frac{\tilde{N}^I}{T-1} - \mu \right)$, would lead to Student 1 having the same beliefs when $I$ grows. Namely, given $n^I$ as an any realization of $N^I$ and $\tilde{N}^I$,

$$
\lim_{I \to \infty} \left[ \Pr \left( N^I = n^I \right) - \Pr \left( \tilde{N}^I = n^I \right) \right] = 0,
$$

which is true because of the convergence in (7) and (8), and because

$$
\lim_{I \to \infty} \Pr \left( N^I = n^I \right)
= \lim_{I \to \infty} \Pr \left[ \sqrt{T-1} \left( \frac{N^I}{T-1} - \mu \right) = \sqrt{T-1} \left( \frac{n^I}{T-1} - \mu \right) \right]
= \lim_{I \to \infty} \Pr \left[ \sqrt{T-1} \left( \frac{N^I}{T-1} - \mu \right) \in \text{Ball} \left( \sqrt{T-1} \left( \frac{n^I}{T-1} - \mu \right), \frac{1}{2\sqrt{T-1}} \right) \right]
= \Phi_Q \left[ \sqrt{T-1} \left( \frac{N^I}{T-1} - \mu \right) \in \text{Ball} \left( \sqrt{T-1} \left( \frac{n^I}{T-1} - \mu \right), \frac{1}{2\sqrt{T-1}} \right) \right]
= \lim_{I \to \infty} \Pr \left( \tilde{N}^I = n^I \right),
$$

where $\text{Ball} \left( \sqrt{T-1} \left( \frac{n^I}{T-1} - \mu \right), \frac{1}{2\sqrt{T-1}} \right)$ is an open ball centered at $\sqrt{T-1} \left( \frac{n^I}{T-1} - \mu \right)$ with a radius of $\frac{1}{2\sqrt{T-1}}$, and $\Phi_Q$ is the distribution function for $N(0, Q)$. The second-to-last equation comes from the definition of convergence in distribution (see for example Bhattacharya (1977) in a multidimensional setting).

By definition, given the information $\mathcal{X}^1$, the beliefs of Student 1 are, $\forall s$,

$$
A_{1,s} \left( C, \sigma_{-i}, \mathcal{X}^1 \right) = \sum_{n=1}^{L(I-1)} \Pr \left( C_{-i}^m \text{ played under } \sigma_{-i} | \mathcal{X}^1 \right) a_s \left( C, C_{-i}^m \right),
$$

which can be re-written as:

$$
A_{1,s} \left( C, \sigma_{-i}, \mathcal{X}^1 \right) = \sum_{\forall n^I} \Pr \left( N^I = n^I \text{ under } \sigma_{-i} | \mathcal{X}^1 \right) \tilde{a}_s \left( C, n^I \right),
$$

where $\tilde{a}_s \left( C, n^I \right)$ is the probability that Student 1 is accepted by $s$ while others’ submitted lists are such that $n^I$ is realized. By the result in (9), as $I \to \infty$,

$$
\left[ A_{1,s} \left( C, \sigma_{-i}, \mathcal{X}^1 \right) - A_{1,s} \left( C, \sigma_{-i} \right) \right] \to 0,
$$

where $A_{1,s} \left( C, \sigma_{-i} \right)$ is the one when $\tilde{X}_i$ is private information.

Since this can be proved this for any other student, the beliefs converge: $B_i \left( C \right) \to B \left( C \right)$, $\forall i$, $\forall C \in \mathcal{C}$. ■

**Corollary** Under the same conditions at in Proposition [3] and that $\tilde{X}_i$ is now common knowledge, the beliefs converge to a common belief, $B_i \left( C \right) \to B \left( C \right)$, $\forall i$, $\forall C \in \mathcal{C}$.
In this corollary, the difference between any two students, $i$ and $j$, is their information about their opponents, the realizations of $\tilde{X}_i$ and $\tilde{X}_j$. However, the difference between $\tilde{X}_i$ and $\tilde{X}_j$ is very limited, since $X_{-i} = (X_{-i,j}, X_j)$ and $X_{-j} = (X_{-i,j}, X_i)$ where $(-i,j)$ denotes the students other than $i$ and $j$. By the same argument in Proposition 3, the beliefs converge.
A.4 Characterization of Choice Probabilities

A.4.1 General Case with All-Positive Beliefs

Given \( S = 4 \), I group the lists into 15 groups, \( g_n, n = 1, \ldots, 15 \): There are one group with \( (0, 0, 0, 0) \), four groups with only one school being ranked, six groups with two schools being ranked, and four groups with all schools are ranked. The main criteria of grouping are the number of schools included in the list and also the identities of them. The choice probabilities now should be interpreted as the probabilities of choosing a group, \( g_n \), or choosing any list within that group, \( C_i \in g_n \).

Also recall that \( m_{K,l}, l \geq K \), is the probability that a \( l \)-school list is submitted while only \( K \) schools are acceptable.

**Characterization of Choice Probabilities** When \( S = 4 \), the conditional probability of \( i \) choosing a group \( g_i \), \( \Pr (C_i \in g_n \mid X_i, Z_i; \theta) \), is

(i) if \( C_i \in g_1 = \{(0, 0, 0, 0)\} \), \( \Pr (u_{i,s} < 0, \text{ for all } s \mid X_i, Z_i; \theta) \);

(ii) if \( C_i \in g_n = \{(c^1, 0, 0, 0)\} \) \( (n = 2, \ldots, 5 \text{ given the identity of } c^1) \),

\[
m_{1,1} * \Pr (u_{i,c^1} > 0 > u_{i,s}, \text{ for } s \neq c^1 \mid X_i, Z_i; \theta)
\]

(iii) if \( C_i \in g_n = \{(c^1, c^2, 0, 0), (c^2, c^1, 0, 0)\} \) \( (n = 6, \ldots, 11 \text{ given the identities of } c^1 \text{ and } c^2) \),

\[
m_{2,2} * \Pr (u_{i,c^1}, u_{i,c^2} > 0 > u_{i,s}, \text{ for } s \neq c^1, c^2 \mid X_i, Z_i; \theta)
+ m_{1,2} * \Pr (u_{i,c^1} > 0 > u_{i,c^2} > u_{i,s}, \text{ for } s \neq c^1, c^2 \mid X_i, Z_i; \theta)
\]

(iv) if \( C_i \in g_n = \{ \text{full lists s.t. the 4th is always } c^4 \} \), \( (n = 12, \ldots, 15 \text{ given the identity of } c^4) \),

\[
\Pr (u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > 0; u_{i,c^4} = \min \{u_{i,s}\}_{s=1}^4 \mid X_i, Z_i; \theta)
+ m_{1,4} * \left[ \Pr (u_{i,c^1} > 0 > \max (u_{i,c^2}, u_{i,c^3}) > \min (u_{i,c^2}, u_{i,c^3}) > u_{i,c^4} \mid X_i, Z_i; \theta) + \Pr (u_{i,c^1} > 0 > \max (u_{i,c^1}, u_{i,c^3}) > \min (u_{i,c^1}, u_{i,c^3}) > u_{i,c^4} \mid X_i, Z_i; \theta) \right]
\]

\[
+ m_{2,4} * \left[ \Pr (u_{i,c^1}, u_{i,c^2} > 0 > u_{i,c^3} > u_{i,c^4} \mid X_i, Z_i; \theta) + \Pr (u_{i,c^1}, u_{i,c^3} > 0 > u_{i,c^2} > u_{i,c^4} \mid X_i, Z_i; \theta) \right]
\]

A.4.2 General Case with Some Possibly Zero Elements in Beliefs

Putting together Proposition 2, Assumptions UNACCEPTABLES, MIXING, A-ZERO-PROB, and A-ZERO-PROB-MIXING, I re-group the lists and characterize the choice probabilities as follows:

**Characterization of Choice Probabilities** When \( S = 4 \), the conditional probability of \( i \) choosing a group \( \bar{g}_n \), \( n = 1, \ldots, 18 \), \( \Pr (C_i \in \bar{g}_n \mid X_i, Z_i; \bar{\theta}) \), is

(i) if \( C_i \in \bar{g}_1 = \{(0, 0, 0, 0)\} \), \( \Pr (u_{i,s} < 0, \text{ for all } s \mid X_i, Z_i; \bar{\theta}) \);

(ii) if \( C_i \in \bar{g}_n = \{(c^1, 0, 0, 0)\} \) \( (n = 2, \ldots, 5 \text{ given the identity of } c^1) \),

\[
m_{1,1} * \Pr (u_{i,c^1} > 0 > u_{i,s}, \text{ for } s \neq c^1 \mid X_i, Z_i; \theta)
\]
(iii) if \( C_i \in \bar{g}_n = \{(c^1, c^2, 0, 0), (c^2, c^1, 0, 0)\} \), s.t., \( c^1 \) or \( c^2 = 1 \), \( n = 6, 7, 8 \) given the identities of \( c^1, c^2 \),

\[
m_{2,2} * Pr(u_{i,c^1}, u_{i,c^2} > 0 > u_{i,s}, \text{for } s \neq c^1, c^2 | X_i, Z_i; \overline{\theta})
+ m_{1,2} * \left[ \Pr(u_{i,c^1} > 0 > u_{i,c^2} > u_{i,s}, \text{for } s \neq c^1, c^2 | X_i, Z_i; \theta) \right]
+ \left[ \Pr(u_{i,c^2} > 0 > u_{i,c^1} > u_{i,s}, \text{for } s \neq c^1, c^2 | X_i, Z_i; \theta) \right]
\]

(iv) if \( C_i \in \bar{g}_n = \{(c^1, c^2, 0, 0), (c^2, c^1, 0, 0)\} \), s.t., \( c^1, c^2 \neq 1 \), \( n = 9, 10, 11 \) given the identities of \( c^1, c^2 \),

\[
m_{2,2} * Pr(u_{i,c^1}, u_{i,c^2} > 0 > u_{i,s}, \text{for } s \neq c^1, c^2 | X_i, Z_i; \overline{\theta})
+ m_{1,2} * \left[ \Pr(u_{i,c^1} > 0 > u_{i,c^2} > u_{i,s}, \text{for } s \neq c^1, c^2 | X_i, Z_i; \theta) \right]
+ \left[ \Pr(u_{i,c^2} > 0 > u_{i,c^1} > u_{i,s}, \text{for } s \neq c^1, c^2 | X_i, Z_i; \theta) \right]
\]

(v) if \( C_i \in \bar{g}_n \) = \{ full list s.t. School 1 (denoted as \( c^1 \)) is 1st or 2nd and the 4th is always \( c^4 \),
(\( n = 12, 13, 14 \) given the identities of \( c^1 \)),

\[
Pr(u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > 0; u_{i,c^1} - \min(u_{i,c^2}, u_{i,c^3}) > \overline{u}_i; u_{i,c^4} = \min\{u_{i,s}\}_{s=1}^4 | X_i, Z_i; \overline{\theta})
+ m_{1,4} * \left[ \Pr(u_{i,c^1} > 0; u_{i,c^2} < \min(u_{i,c^2}, u_{i,c^3}) < \max(u_{i,c^2}, u_{i,c^3}) < 0 | X_i, Z_i; \theta) \right]
+ \left[ \Pr(u_{i,c^2} > 0; u_{i,c^3} < u_{i,c^4} < u_{i,c^3} < 0 | X_i, Z_i; \theta) \right]
+ \left[ \Pr(u_{i,c^3} > 0; u_{i,c^2} < u_{i,c^4} < u_{i,c^2} < 0 | X_i, Z_i; \theta) \right]
\]

(vi) if \( C_i \in \bar{g}_n \) = \{ full list \( C \) s.t. School 1 (denoted as \( c^3 \)) is 3rd and the 4th is always \( c^4 \),
(\( n = 15, 16, 17 \) given the identities of \( c^1 \) and \( c^2 \)),

\[
Pr(u_{i,c^1}, u_{i,c^2}, u_{i,c^3} > 0; u_{i,c^1} - \min(u_{i,c^1}, u_{i,c^3}) < \overline{u}_i; u_{i,c^4} < 0 | X_i, Z_i; \overline{\theta})
+ m_{1,4} * \left[ \Pr(u_{i,c^1} > 0; u_{i,c^2} < \min(u_{i,c^2}, u_{i,c^3}) < \max(u_{i,c^2}, u_{i,c^3}) < 0 | X_i, Z_i; \theta) \right]
+ \left[ \Pr(u_{i,c^2} > 0; u_{i,c^3} < u_{i,c^4} < u_{i,c^3} < 0 | X_i, Z_i; \theta) \right]
+ \left[ \Pr(u_{i,c^3} > 0; u_{i,c^2} < u_{i,c^4} < u_{i,c^2} < 0 | X_i, Z_i; \theta) \right]
\]

(vii) if \( C_i \in \bar{g}_{18} \) = \{ full list s.t. School 1 (denoted as \( c^4 \)) is ranked 4th \},

\[
\]
\[ \Pr \left( u_{i,c1}, u_{i,c2}, u_{i,c3} > 0; u_{i,c4} = \min \{ u_{i,s} \}_{s=1}^4 | X_i, Z_i; \bar{\theta} \right) \]

\[ = \Pr \left( u_{i,c3} = \min \{ u_{i,s} \}_{s=1}^4 > 0; u_{i,c4} - \min (u_{i,c1}, u_{i,c2}) < \bar{u}_i; X_i, Z_i; \bar{\theta} \right) \]

\[ + \left[ \Pr \left( u_{i,c2} = \min \{ u_{i,s} \}_{s=1}^4 > 0; u_{i,c4} - \min (u_{i,c1}, u_{i,c3}) < \bar{u}_i; X_i, Z_i; \bar{\theta} \right) \right. \]

\[ + \left. \Pr \left( u_{i,c1} = \min \{ u_{i,s} \}_{s=1}^4 > 0; u_{i,c4} - \min (u_{i,c2}, u_{i,c3}) < \bar{u}_i; X_i, Z_i; \bar{\theta} \right) \right] \]

\[ \times \Pr \left( \text{Ranking School 1 fourth when the worst is acceptable and School 1 is not the worst and is optimally ranked outside top 2.} \right) \]

\[ + m_{1,4} \left[ \Pr \left( u_{i,c1} > 0; u_{i,c4} < \min (u_{i,c1}, u_{i,c3}) < \max (u_{i,c2}, u_{i,c3}) < 0 | X_i, Z_i; \bar{\theta} \right) \right. \]

\[ + \left. \Pr \left( u_{i,c2} > 0; u_{i,c4} < \min (u_{i,c1}, u_{i,c3}) < \max (u_{i,c1}, u_{i,c3}) < 0 | X_i, Z_i; \bar{\theta} \right) \right] \]

\[ + m_{2,4} \left[ \Pr \left( u_{i,c1} > 0; u_{i,c2} > 0; u_{i,c3} > u_{i,c4} | X_i, Z_i; \bar{\theta} \right) \right. \]

\[ + \left. \Pr \left( u_{i,c2} > 0; u_{i,c3} > 0; u_{i,c4} > u_{i,c1} | X_i, Z_i; \bar{\theta} \right) \right] \]

A.4.3 Everyone Is Truth-Telling

Now suppose everyone reports their true preference ranking. To characterize the choice probability, one also need to use Assumptions UNACCEPTABLES and MIXING because of the unacceptable schools. In the following, grouping is not needed.

Characterization of Choice Probabilities Given that researchers observe \((X_i, Z_i)\), the equilibrium beliefs \(B^\ast\), and with \(\theta\) denoting the unknown parameters, in equilibrium, the conditional probability of \(i\) choosing \(C_i\) in equilibrium, \(\Pr (C_i | X_i, Z_i, B^\ast; \theta)\), is:

(i) if \(C_i = (0, 0, 0, 0)\), \(\Pr (u_{i,s} < 0, \text{ for all } s | X_i, Z_i, B^\ast; \theta)\);

(ii) if \(C_i = (c^1, 0, 0, 0)\),

\[ m_{1,1} \times \Pr (u_{i,c1} > 0 > u_{i,s}, \text{ for } s \neq c^1 | X_i, Z_i, B^\ast; \theta) \]

(iii) if \(C_i = (c^1, 0, 0, 0)\),

\[ m_{2,2} \times \Pr (u_{i,c1} > u_{i,c2} > 0 > u_{i,s}, \text{ for } s \neq c^1 \neq c^2 | X_i, Z_i, B^\ast; \theta) \]

\[ + m_{1,2} \times \Pr (u_{i,c1} > 0 > u_{i,c2} > u_{i,s}, \text{ for } s \neq c^1 \neq c^3 | X_i, Z_i, B^\ast; \theta) \]

(iv) if \(C_i = (c^1, c^2, c^3, c^4)\),

\[ \Pr (u_{i,c1} > u_{i,c2} > u_{i,c3} > 0 > u_{i,c4} | X_i, Z_i, B^\ast; \theta) \]

\[ + m_{1,4} \times \Pr (u_{i,c1} > 0 > u_{i,c2} > u_{i,c3} > u_{i,c4} | X_i, Z_i, B^\ast; \theta) \]

\[ + m_{2,4} \times \Pr (u_{i,c1} > u_{i,c2} > 0 > u_{i,c3} > u_{i,c4} | X_i, Z_i, B^\ast; \theta) \]
A.5 The Logit-Smoothed Reject-Accept Simulator and Solving the Equilibrium

This appendix describes the logit-smoothed reject-accept simulator, how to simulate the choice probabilities, and how to find equilibrium beliefs when the parameters in the utility function and parents’ strategies are given.

The simulator is implemented in the following steps, as described in Chapter 5 of Train (2009).

1. Draw a value of the 4-dimensional vector of errors, \( \varepsilon_i^r = L_i^r \eta \), as follows:
   - Draw 4 values from a standard normal density using a random number generator. Stack these values into a vector, and label the vector \( \varepsilon_i^r \). In the paper, randomized Halton sequences are used to reduce variance of the simulator, as suggested in Chapter 9 of Train (2009).
   - Calculate \( \varepsilon_i^r = L_i^r \eta \), where \( L_i^r \) is the Cholesky factor of \( \Sigma \).

2. Repeat Step 1 for \( r = 2, \ldots, 300 \), and calculate \( u_{i,s}^r \) given \((X_i, Z_i)\) and parameters.

3. To calculate logit formula for corresponding events and/or choice probabilities, for example, \( u_{i,4} = \min \{u_{i,s}\}_{s=1}^4 \), I define:

\[
S^r = \frac{1}{1 + \exp \left( u_{i,4}^r - u_{i,1}^r \right) / \lambda} \cdot \frac{1}{1 + \exp \left( u_{i,4}^r - u_{i,2}^r \right) / \lambda} \cdot \frac{1}{1 + \exp \left( u_{i,4}^r - u_{i,3}^r \right) / \lambda},
\]

where \( \lambda > 0 \) is a scale factor and I experiment it with \( \lambda = 0.05, 0.025, 0.01, 0.005 \). Those presented in the paper is from \( \lambda = 0.01 \).

4. The simulated probability of the corresponding event is then \( \tilde{P}_r = \frac{1}{300} \sum_{r=1}^{300} S^r \).

It is easy to verify that \( \tilde{P}_r \) is strictly positive and twice differentiable.

With the help of this simulator, the following procedure can be used to solve the Bayesian Nash equilibrium in the Monte Carlo experiment. The basic idea is illustrated in the following mapping:

<table>
<thead>
<tr>
<th>Given the common strategy ( \sigma_i(X_i, Z_i, \varepsilon_i; B) ), start with beliefs ( B(C) ), ( \forall C \in C )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Possible Profiles:</strong></td>
</tr>
<tr>
<td>[ C, C^{-1}_1 ]</td>
</tr>
<tr>
<td>[ C, C^{-2}_1 ]</td>
</tr>
<tr>
<td>( \ldots )</td>
</tr>
<tr>
<td>[ C, C^{-L_{i-1}}_1 ]</td>
</tr>
<tr>
<td>( \ldots )</td>
</tr>
<tr>
<td>( A_s(C, B) = \sum_{n=1}^{L_{i-1}} p(C_{n-1}^i</td>
</tr>
<tr>
<td><strong>Implied Probabilities:</strong> ( \tilde{B}(C, B) = (A_1(C, B), \ldots, A_S(C, B)) )</td>
</tr>
</tbody>
</table>

**Mapping from Beliefs to the Implied Probabilities for Student 1**
Since everyone has the same beliefs, it is suffice to just look at Student 1’s probabilities of being admitted by the schools in her list. The simulation of the implied probabilities has seven steps as following:

1. Draw \( NC (= 20,000) \) profiles of choice lists, \( \{ C_i = (c_{i1}, ..., c_{iS}) \}_{i=2}^I \). Given each profile, \( \{ C_i \}_{i=2}^I \), student 1 tries all \((S!)\) full choice lists. Combine them together, I create \( S! \times NC \) profiles of \( \{ C_i \}_{i=1}^I \). Among 20,000 profiles,
   - a quarter of them are drawn from the distribution of observed lists plus 9 "imputed" observation each of which is one of the 9 lists which are not observed in the data;
   - another quarter are random draws from the \( L (=41) \) possible lists;
   - another quarter are such that a half of the students are randomly fixed at one of the \( L (=41) \) possible lists, while the other half take a random draw from the \( L \) lists;
   - the last quarter are random draws from a distribution of true preference orders which are predicted using the estimates from the Zero_Belief case.

2. Given each profile of lists, \( \{ C_1, \{ C_i \}_{i=2}^I \} \), create a set of random lottery numbers, \( rs = 1 \), and then run the admission process to see which school admits student 1, i.e., get the values for the following indicator functions:
   \[
   1^{rs} \left( \text{Student 1 assigned to } s|C_1, \{ C_i \}_{i=2}^I \right), \ s = 1, ..., S;
   \]

3. Repeat Step 2 with different lottery number draws, \( rs = 2, ..., 1000 \), and calculate the probabilities of Student 1 being admitted by every \( s \) respectively.
   \[
   \widetilde{Pr} \left( \text{Student 1 assigned to } s|C_1, \{ C_i \}_{i=2}^I \right) = \frac{1}{1000} \sum_{rs=1}^{1000} 1^{rs} \left( \text{Student 1 assigned to } s|C_1, \{ C_i \}_{i=2}^I \right), \ s = 1, ..., S.
   \]

4. Repeat Steps 2 and 3 for all \( S! \) profiles lists with \( \{ C_i \}_{i=2}^I \) fixed and Student 1 selecting each of all \( S! \) choice lists.

The above four steps are independent of the belief system and the error terms in the utility functions. Thus they are only simulated once.

5. Simulate the probability of choosing each list by logit-smoothed accept-reject simulator.
   Given the utility functions, simulate \( r = 1, ..., 300 \) draws of \( \{ \eta_i \}_{i=2}^I \). Given the candidate belief, \( B \), the simulated probability of student \( i \) a list \( C_k \) is
   \[
   \widetilde{P} \left( C_k|X_i, \{ z_s \}_{s=1}^S; \theta \right) = \frac{1}{300} \sum_{r=1}^{300} 1^r \left( C_i|X_i, Z_i; \theta; B \right), k = 1, ..., L,
   \]

67
where \( 1^r(C_i|X_i, Z_i; \theta; B) \) is an indicator function of \( C_i \) being choosing as a best response given \( B \). Note that \( 1^r(C_i|X_i, \{ z_s \}_{s=1}^S; \theta) \) may be weighted by the corresponding mixing probabilities.

6. Calculate the average choice probability for the \( L \) choice lists:

\[
\tilde{P}_k = \frac{1}{914} \sum_{i=1}^{914} P_i(C_i|X_i, \{ z_s \}_{s=1}^S; \theta), \quad k = 1, \ldots, L.
\]

7. Calculate the probability of the profiles \( \left\{ \left\{ C_i^{(t)} \right\}_{i=2}^I \right\}_{t=1}^{NC} \) simulated in Step 1 being realized, i.e., if \( \left\{ C_i^{(t)} \right\}_{i=2}^I = (C_2^{(t)}, C_3^{(t)}, \ldots, C_I^{(t)}) \), then

\[
\tilde{P}_T \left( \left\{ C_i^{(t)} \right\}_{i=2}^I \text{ realized} \right) = \frac{1}{K} \prod_{i=2}^L \prod_{k=1}^L \left( \tilde{P}_k \right)^{1(C_i^{(t)}=C_k)}
\]

where \( K \) is a normalization term,

\[
K = \sum_{t=1}^{NC} \tilde{P}_T \left( \left\{ C_i^{(t)} \right\}_{i=2}^I \text{ realized} \right).
\]

8. Calculate the implied probability of Student 1 being admitted by school \( s \) as follows, \( \forall s = 1, \ldots, S \):

\[
\tilde{P}_T \text{ (Student 1 assigned to } s \text{ when submitting } C) = \frac{1}{NC} \sum_{t=1}^{NC} \tilde{P}_T \left( \text{Student 1 assigned to } s | C, \left\{ C_i^{(t)} \right\}_{i=2}^I \right) \times \tilde{P}_T \left( \left\{ C_i^{(t)} \right\}_{i=2}^I \text{ realized} \right).
\]

This is calculated for all \( S! \) possible full lists. All the probabilities together are the simulated implied probabilities, \( B(\cdot, B) \).

9. The equilibrium belief is a fixed point: \( B(\cdot, B) = B \).

Note that the above Steps 6 and 7 can be replaced by the following one step:

6` Calculate the probability of the profiles \( \left\{ \left\{ C_i^{(t)} \right\}_{i=2}^I \right\}_{t=1}^{NC} \) simulated in Step 1 being realized, i.e., if \( \left\{ C_i^{(t)} \right\}_{i=2}^I = (C_2^{(t)}, C_3^{(t)}, \ldots, C_I^{(t)}) \), then

\[
\tilde{P}_T \left( \left\{ C_i^{(t)} \right\}_{i=2}^I \text{ realized} \right) = \frac{1}{K} \prod_{i=2}^I \tilde{P}_i \left( C_i^{(t)}|X_i, \{ z_s \}_{s=1}^S; \theta \right),
\]

68
where \( K \) is a normalization term,

\[
K = \sum_{t=1}^{NC} \Pr \left( \left\{ C^{(t)}_i \right\}_{i=2}^{I} \text{ realized} \right).
\]

The issue with this step in practice is that many \( \Pr \left( \left\{ C^{(t)}_i \right\}_{i=2}^{I} \text{ realized} \right) \) are very likely to be zero, since each individual might have a very low probability choosing a given list. Replacing Step 6’ with Steps 6 and 7 solves this problem while introducing some simulation error. To be more precise, This two procedures converge to random draws from two multinomial normal distributions with the same mean but different variances (difference is bounded), as can be shown using the same arguments in Appendix A.3.
### B Additional Tables

#### Table A-1: Attention on Different Aspects of Uncertainties and School Quality

<table>
<thead>
<tr>
<th></th>
<th>Quota (1)</th>
<th>Prob(Bad School) (2)</th>
<th>Others’ App. (3)</th>
<th>School Quality (4)</th>
<th>Full Sample (5)</th>
<th>Full List (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean(Dep V)</td>
<td>4.232</td>
<td>4.468</td>
<td>2.814</td>
<td>4.151</td>
<td>4.136</td>
<td></td>
</tr>
<tr>
<td>Std Dev(Dep V)</td>
<td>0.863</td>
<td>0.740</td>
<td>1.213</td>
<td>0.460</td>
<td>0.467</td>
<td></td>
</tr>
<tr>
<td>(Parent_Edu_i)</td>
<td>0.016</td>
<td>-0.000</td>
<td>-0.056*</td>
<td>-0.001</td>
<td>-0.004</td>
<td></td>
</tr>
<tr>
<td>(Parent_Inc_i)</td>
<td>-0.084**</td>
<td>-0.003</td>
<td>-0.052</td>
<td>0.043**</td>
<td>0.052**</td>
<td></td>
</tr>
<tr>
<td>(Own_Score_i)</td>
<td>0.053</td>
<td>1.037**</td>
<td>-1.003</td>
<td>0.224</td>
<td>-0.261</td>
<td></td>
</tr>
<tr>
<td>(Awards_i)</td>
<td>0.025</td>
<td>0.048</td>
<td>0.022</td>
<td>0.008</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>(Girl_i)</td>
<td>-0.020</td>
<td>-0.008</td>
<td>0.077</td>
<td>0.026</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td>(Attn_Others_i)</td>
<td>-0.029</td>
<td>0.009</td>
<td>0.063***</td>
<td>0.069***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Attn_Q_i)</td>
<td>0.529***</td>
<td>0.379***</td>
<td>0.610***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attn on Quota</td>
<td>0.339***</td>
<td>-0.080</td>
<td>0.145***</td>
<td>0.161***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prob(Bad School)</td>
<td>0.465***</td>
<td>0.033</td>
<td>0.156***</td>
<td>0.164***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.255</td>
<td>-3.95</td>
<td>6.824</td>
<td>1.198</td>
<td>3.544**</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>457</td>
<td>457</td>
<td>457</td>
<td>676</td>
<td>457</td>
<td>457</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.339</td>
<td>0.344</td>
<td>0.075</td>
<td>0.301</td>
<td>0.299</td>
<td></td>
</tr>
</tbody>
</table>

Results are from OLS regressions, and other variables include fixed effects for elementary schools.

a. The full sample includes every parent whose relevant variables are not missing, and the subsample with full lists are those who submit a full list.

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Table A-2: Deviation from Best-Responding and Truth-Telling Prediction: Partial Lists

<table>
<thead>
<tr>
<th>Rank-Order Lists</th>
<th>Observed Data Percent&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Deviation from the Prediction of Best Responding</th>
<th>Deviation from the Prediction of Truth-Telling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean&lt;sup&gt;b&lt;/sup&gt; (10^-2)</td>
<td>Std. Dev.</td>
<td>Mean&lt;sup&gt;b&lt;/sup&gt; (10^-2)</td>
</tr>
<tr>
<td>(1,4,0,0)</td>
<td>0.00%</td>
<td>-0.06***</td>
<td>0.001</td>
</tr>
<tr>
<td>(1,3,0,0)</td>
<td>0.77%</td>
<td>0.07</td>
<td>0.087</td>
</tr>
<tr>
<td>(1,2,0,0)</td>
<td>1.09%</td>
<td>-0.28</td>
<td>0.104</td>
</tr>
<tr>
<td>(2,1,0,0)</td>
<td>0.66%</td>
<td>0.54**</td>
<td>0.081</td>
</tr>
<tr>
<td>(2,3,0,0)</td>
<td>2.95%</td>
<td>2.63***</td>
<td>0.169</td>
</tr>
<tr>
<td>(2,4,0,0)</td>
<td>0.22%</td>
<td>0.15</td>
<td>0.047</td>
</tr>
<tr>
<td>(3,1,0,0)</td>
<td>0.11%</td>
<td>-0.23**</td>
<td>0.033</td>
</tr>
<tr>
<td>(3,4,0,0)</td>
<td>0.22%</td>
<td>0.10</td>
<td>0.047</td>
</tr>
<tr>
<td>(3,3,0,0)</td>
<td>0.11%</td>
<td>-1.13***</td>
<td>0.034</td>
</tr>
<tr>
<td>(4,1,0,0)</td>
<td>0.00%</td>
<td>0.00***</td>
<td>0.000</td>
</tr>
<tr>
<td>(4,3,0,0)</td>
<td>0.22%</td>
<td>0.21</td>
<td>0.047</td>
</tr>
<tr>
<td>(4,2,0,0)</td>
<td>0.00%</td>
<td>-0.01***</td>
<td>0.000</td>
</tr>
<tr>
<td>(2,0,0,0)</td>
<td>0.11%</td>
<td>0.11</td>
<td>0.033</td>
</tr>
<tr>
<td>(3,0,0,0)</td>
<td>0.66%</td>
<td>0.01</td>
<td>0.081</td>
</tr>
<tr>
<td>(1,0,0,0)</td>
<td>0.22%</td>
<td>-0.17</td>
<td>0.047</td>
</tr>
<tr>
<td>(4,0,0,0)</td>
<td>0.11%</td>
<td>0.03</td>
<td>0.033</td>
</tr>
<tr>
<td>(0,0,0,0)</td>
<td>19.80%</td>
<td>-0.71</td>
<td>0.381</td>
</tr>
</tbody>
</table>

A t-test is run for the null hypothesis that each mean equals zero. *** p<0.01, ** p<0.05, * p<0.1.

a. It is the percentage out of the total of 914 students.
b. The means should be interpreted as percentages points, i.e., -0.06 means -0.06 percentage points.
Table 3: Who Is Truth-Telling More Often: Regression Analysis of Deviations from Truth-Telling

<table>
<thead>
<tr>
<th></th>
<th>Most Under-Used List: (1,3,2,4)</th>
<th>Most Over-Used List: (2,3,1,4)</th>
<th>Most Likely True Preference: (1,2,3,4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Deviation from Truth_Telling</td>
<td>Deviation from Truth_Telling</td>
<td>Deviation from Truth_Telling</td>
</tr>
<tr>
<td></td>
<td>Mean: -0.177; Std Dev: 0.223</td>
<td>Mean: 0.199; Std Dev: 0.408</td>
<td>Mean: -0.113; Std Dev: 0.356</td>
</tr>
</tbody>
</table>

### Parent_Inc_i
- **Mean:** 0.003 0.003 0.002
- **Std Dev:** 0.007 0.009 0.010

### Parent_Edu_i
- **Mean:** 0.008** 0.004 0.003
- **Std Dev:** 0.004 0.006 0.006

### Own_Score_i
- **Mean:** 0.307*** 0.452*** 0.372***
- **Std Dev:** 0.082 0.098 0.093

### Awards_i
- **Mean:** 0.015* 0.013 0.008
- **Std Dev:** 0.009 0.014 0.014

### Own_Score_i
- **Mean:** -0.002 -0.009 -0.009
- **Std Dev:** 0.014 0.016 0.016

### Gain_i
- **Mean:** 0.867* 4.550* 1.680***
- **Std Dev:** 0.512 2.496 0.275

### Attn_U_i
- **Mean:** -0.006 -0.006 0.039**
- **Std Dev:** 0.012 0.012 0.017

### Attn_Q_i
- **Mean:** -0.002 -0.001 -0.028
- **Std Dev:** 0.017 0.017 0.035

### Attn_Others_i
- **Mean:** -0.008 -0.008 0.018
- **Std Dev:** 0.006 0.006 0.012

### pTT=BR
- **Mean:** -0.366** 0.003 -0.019
- **Std Dev:** 0.158 0.028 0.028

### Gain_i
- **Mean:** 0.015 -0.012 -0.005
- **Std Dev:** 0.006 0.007 0.007

### pBR_i;k
- **Mean:** -0.535** -0.524 -1.652***
- **Std Dev:** 0.258 0.519 0.525

<table>
<thead>
<tr>
<th></th>
<th>914</th>
<th>810</th>
<th>810</th>
<th>914</th>
<th>810</th>
<th>810</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs.</td>
<td>914</td>
<td>810</td>
<td>810</td>
<td>914</td>
<td>810</td>
<td>810</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.065</td>
<td>0.083</td>
<td>0.081</td>
<td>0.053</td>
<td>0.067</td>
<td>0.069</td>
</tr>
</tbody>
</table>

Definitions of \( p_{TT=BR} \) and \( Gain_i \) are in Table 11. \( p_{TT=BR} \) is the probability of the list (in the dependent variable) being a best response.

Elementary school fixed effects are included. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
Table A-4: Who Strategizes Better: Regression Analysis of Sophistication Measures

<table>
<thead>
<tr>
<th></th>
<th>Most Likely True Preference Order: (1,2,3,4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent V.: Deviation from Best-Responding Prediction</td>
<td>Mean: 0.0324; Std. Dev: 0.346</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>Coefficient</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent_Inci</td>
<td>0.005</td>
<td>0.017</td>
<td>-0.000</td>
<td>0.021</td>
<td>0.006</td>
<td>0.022</td>
<td>-0.006</td>
<td>0.021</td>
</tr>
<tr>
<td>Parent_Edui</td>
<td>-0.001</td>
<td>0.006</td>
<td>-0.007</td>
<td>0.007</td>
<td>-0.012*</td>
<td>0.007</td>
<td>-0.012*</td>
<td>0.007</td>
</tr>
<tr>
<td>Own_Scorei</td>
<td>0.266*</td>
<td>0.136</td>
<td>0.494**</td>
<td>0.219</td>
<td>0.537***</td>
<td>0.172</td>
<td>0.632***</td>
<td>0.231</td>
</tr>
<tr>
<td>Awardsi</td>
<td>-0.011</td>
<td>0.013</td>
<td>-0.008</td>
<td>0.019</td>
<td>-0.005</td>
<td>0.015</td>
<td>-0.020</td>
<td>0.021</td>
</tr>
<tr>
<td>Girli</td>
<td>-0.004</td>
<td>0.023</td>
<td>-0.004</td>
<td>0.023</td>
<td>-0.005</td>
<td>0.023</td>
<td>-0.009</td>
<td>0.024</td>
</tr>
<tr>
<td>(p_{TT=BR}^i)</td>
<td>-0.184</td>
<td>0.357</td>
<td>-0.106</td>
<td>0.388</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaini</td>
<td>1.822</td>
<td></td>
<td>1.421</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Attn_{U}^i)</td>
<td></td>
<td>-0.012</td>
<td>-0.011</td>
<td></td>
<td>(0.019)</td>
<td>(0.019)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Attn_{Q}^i)</td>
<td></td>
<td>0.053*</td>
<td>0.053*</td>
<td></td>
<td>(0.028)</td>
<td>(0.028)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Attn_{Others}^i)</td>
<td></td>
<td>-0.018*</td>
<td>-0.018*</td>
<td></td>
<td>(0.011)</td>
<td>(0.011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(p_{TT}^i,k)</td>
<td></td>
<td>-1.062***</td>
<td>-1.260***</td>
<td>-1.136***</td>
<td>-1.310***</td>
<td></td>
<td>(0.376)</td>
<td>(0.360)</td>
</tr>
<tr>
<td>Obs.</td>
<td>914</td>
<td></td>
<td>914</td>
<td></td>
<td>914</td>
<td></td>
<td>810</td>
<td></td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.029</td>
<td>0.042</td>
<td>0.044</td>
<td>0.044</td>
<td>0.057</td>
<td>0.057</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Definitions of \(p_{TT=BR}^i\) and \(Gain_i\) can be found in Table 11. \(p_{TT}^i,k\) is the probability that the list (in the dependent variable) is the true preference ranking. All regressions include a constant and elementary school fixed effects. Robust standard errors in parentheses. *** \(p<0.01\), ** \(p<0.05\), * \(p<0.1\).