A Robust Redesign of High School Match

Sam Hwang

Department of Economics, University of Chicago

Introduction

The topic of this project is school choice mechanisms. A school choice mechanism takes rank order lists students submitted as inputs and produce an assignment of students to schools through an algorithm. One such algorithm that is used most in practice is known as the Boston mechanism.

A problem with the Boston mechanism is that students have incentives not to report their true rankings over schools, or in the terminology of economics, it is not "strategy-proof". Under the Boston mechanism, students tend to avoid 1^{st}-ranking their true favorite schools if they believe it is too difficult to get admission from them. Instead they tend to 1^{st}-rank what they believe to be more "safe" choices. Because of this property, the rank order lists submitted by students under the Boston mechanism are not necessarily identical to their true rankings over schools.

In this project I show that even with such rank order lists it is possible to estimate the parameters that determine students' evaluations of schools. My innovation is that contrary to the previous literature, I do not assume students can accurately calculate their admission probabilities. Instead I assume that students can correctly predict to which schools are more difficult to get admission than others.

With the estimates I simulate what would happen if strategy-proof algorithms replace the Boston mechanism. I found that a majority of students would prefer a strategy-proof algorithm to the Boston mechanism.

The Boston Mechanism

The Boston mechanism assign students to schools with the following algorithm:

Round 0. Students submit rank order lists
Round 1. The mechanism assigns students to their 1^{st} choice schools. If a school is overdemanded,

The Boston Mechanism Continued

then those who have priority (e.g., live within walking distance to the school and/or have siblings already attending the school) are assigned first. If there are ties between students who have the same priority, let a random lottery break the tie. Rejected students move on to the next round.
Round k. $k \geq 2$ The mechanism assigns remaining students to their k^{th} choice schools if there are vacancies at the schools. If a school is overdemanded, those who have priorities are assigned first. If there are ties between students who have the same priority, let a random lottery break the ties. Rejected students move on to the next round.
The Termination of the Algorithm The algorithm terminates when all students are assigned or no students have any schools left that they have ranked and still have vacancies.

The Boston Mechanism is Not Strategy-Proof

The Boston mechanism is not strategy-proof, i.e. students may have incentives not to report their true rankings over schools. Here is an example that illustrates why it is not strategy-proof: suppose that there are

- 3 schools, denoted α, β, γ, with one seat each
- 3 students, denoted $1, 2, 3$
  - All students prefer α to β to γ: yet 1 and 2 have strong preferences toward α over β and γ whereas 3 likes α better than β but not as much as 1 and 2 do
  - All students have the same priority to all schools: only random lotteries break ties among them

Assume that 1 and 2 like school α too much not to rank it 1^{st} (and further suppose that they rank β second and γ third) and that 3 knows that. Then 3 can either rank schools truthfully or 1^{st}-rank β. If she chooses the first option, she has $\frac{1}{3}$ chances of being assigned to α, β, and γ, respectively. However, if she chooses the second option, she will be assigned to β with certainty. Therefore, if she likes β enough or does not want the risk of being assigned to her least favorite γ, she might want to $1^{st}-$rank β.

Students' Preferences For Schools

What is the data and what do we want to learn from them? Suppose we have the following data:

- Rank order lists students submitted to the Boston mechanism
- For each school s, some index for the quality of education, denoted by q_s
- For each student i and school s, the length of commute from i's home to s, denoted by c_{is}

We want to estimate how students "evaluate" each school based on q_s, c_{is} and an index of all the other unobservable factors that affect the evaluations, denoted by ϵ_{is}. That is, if we denote s's score for i with p_{is}, then we assume that

$$p_{is} = p(q_s, c_{is}, \epsilon_{is})$$

And it is the function $p(\cdot,\cdot,\cdot)$ that we want to estimate from the data.

A (Weak) Assumption About Students' Ranking Behavior

In estimating the function $p(\cdot,\cdot,\cdot)$, we need to make an assumption about how students choose rank order lists. Previously, economists have assumed that students choose optimal rank order lists based on a perfectly correct prediction about their admission probabilities, i.e., if students believe that there is a 30% chance that they get into a school α if they 1^{st}-rank it and 70% chance for a school β and choose optimal rank order lists based on such beliefs, then after all the rank order lists are submitted and the probabilities are computed, their beliefs turn out to be correct.

My innovation is that I do not assume that students correctly predict the admission probabilities. Instead, I assume that students correctly predict which schools are more difficult to be admitted to than others. Based on such assumption, we can generate inequalities involving the scores of two schools for a student i, p_{is} and $p_{is'}$.

A (Weak) Assumption About Students' Ranking Behavior Continued

If a student i 1^{st}-ranks s that is more difficult to get into than s', then it had to be that she liked s better than s': by way of conditradiction, suppose that she had liked s' better and it was easier to get into s'. Then it would have been better for her to $1^{st}-$rank s' instead of s, which contradicts the fact that i 1^{st}-ranked s. Hence, we establish the following inequality.

$$p_{is} \geq p_{is'}$$

We can find similar inequalities for 2^{nd}-, 3^{rd}-,...ranked schools and for all students.

Estimation and the Results

These inequalities do not provide enough information about the function $P(\cdot,\cdot,\cdot)$ to point-identify $P(\cdot,\cdot,\cdot)$: therefore, the estimates are generally multivalued.

For the estimation, I parameterized $P(\cdot,\cdot,\cdot)$ with 19 parameters. Because I cannot search over the entire 19 dimensional Euclidean space, I restrict the parameter space to be 19 dimensional box $[-25, 25]^{19}$. Then I drew 1.5 million random parameter values from the parameter space and tested whether each of these values explains the inequalities as well as any other, using the technique introduced by Romano, Shaikh and Wolf (2014). **This is where I benefited greatly from RCC: without RCC I could not have tested such a large number of parameter values, which gave the credibility to my estimates.**

I found that for 0.02 percent of 1.5 million candidate values I could not reject the hypotheses that they are not the true parameter values. And for all those parameter estimates, the simulation results indicate that a majority of students would benefit from switching to a strategy-proof mechanism.