Estimating Preferences in School Choice Mechanisms: Theoretical Foundation and Empirical Approaches

Gabrielle Fack
Universitat Pompeu Fabra

Julien Grenet
Paris School of Economics

Yinghua He
Toulouse School of Economics

Motivation: Growing Popularity of the DA

- **Centralized mechanisms** are widely used to assign students to schools/universities

  $\rightarrow$ Increasing popularity of the **Deferred-Acceptance (DA)** mechanism
- **Large body of theoretical research** but little empirical evidence until very recently
- With the increasing availability of school choice data, **growing number of empirical studies**:
  - estimate preferences from rank-ordered list of choices
  - perform counterfactual policy experiments

  Abdulkadiroglu et al. (2013), Akyol and Krishna (2013), Burgess et al. (2009), Braun et al. (2012), Budish and Cantillon (2012), Carvalho et al. (2014), Hastings et al. (2008)

Motivation: Growing Popularity of the DA

- **Student-Proposing Deferred Acceptance Mechanism:**
  - J schools with a fixed capacity of seats
  - N students who submit rank-ordered lists of schools
  - schools have strict priority ranking of students
  - **Round 1**: Each student proposes to her first choice. Each school *tentatively* accepts the group of applicants with highest priority and rejects the others.
  - **Round 2**: Rejected students apply to their next choice. Each school pools new applicants and those accepted in round 1. It *tentatively* accepts those with highest priority and rejects the others.
  - The process terminates after k **rounds** when no rejections are issued.

Motivation: Truth-Telling?

- Most empirical studies estimate preferences from school choice data under the assumption that individuals are **truth-telling**
  - Students truthfully rank schools
  - Justification: the DA is strategy-proof
- **However:**
  - Under the student-proposing DA, strategy-proofness only implies that truth-telling is a **weakly dominant** strategy
    $\rightarrow$ Multiple equilibria?
  - Many applications of the DA **limit the number of choices** that students can submit, which induces strategic behavior
    Ex: NYC, Paris, Norway, Chile, Finland...
- **Methodological challenge**: How to estimate student preferences without assuming truth-telling?

Results and Contributions

- Idea: focus on the **matching outcome** instead of strategies
- Derive conditions under which the matching outcome is (ex post) asymptotically **stable**
- **Stability $\iff$ everyone is matched with their most preferred school among ex post feasible schools**
  - a school is feasible if its *ex post* cutoff $<$ student's score
- **Empirical approach** based on the matching outcome:
  - use discrete choice model with *personalized* choice sets (ex post feasible schools)
  - Monte Carlo simulations show that this method performs much better than the traditional truth-telling approach

Outline

1. Introduction
2. Model: School Choice as a Bayesian Game
3. Comparison of Empirical Approaches: Monte Carlo Simulations
4. Conclusion

School Choice Model

- Students have strict **preference ordering** of schools (vNM utilities)
- Schools have strict **priority** ranking of students (e.g. GPA score)
- Students are assigned to schools through a **DA mechanism**
- There is **cost of submitting a list** of schools.

  Subsumes:
  - traditional setting: zero cost of ranking all schools
  - constrained DA: infinite cost when limit is reached
  - more generally, cost may increase with size of list
- Bayesian game: **uncertainty** in school admission cutoffs (because of uncertainty in others' preferences) but the distribution is known
- Each student submits the list that **maximizes her expected utility of admission**

Restrictive Identifying Assumption: Truth-Telling

- **Definition**: a student is **truth-telling** under the DA if she submits her top-k most preferred schools
- **Example:** i's preferences over schools are $s_1 \succ s_2 \succ s_3$
  - Truth-telling: $\{s_1, s_2\}$
  - Non-truth-telling: $\{s_2, s_1\}$
  - Non-truth-telling: $\{s_2, s_3\}$ or $\{s_1, s_3\}$
- **Pb 1: When all schools can be listed at no cost, truth-telling is only a weakly dominant strategy**

  Ex: "safe" or "impossible" schools
- **Pb 2: When students cannot list all schools at zero cost, truth-telling is not even a weakly dominant strategy**

  Intuition: If a student can only rank 3 schools out of 10, she might prefer to skip her most preferred school in favor of a more feasible one $\rightarrow$ problem for empirical studies in constrained DA settings

Identifying Assumption: Stable Matching

- Truth-telling assumption too restrictive: **Stability?**
- A matching is **stable** when there is no student i who prefers school s to her assignment, and has higher priority than at least one student who is assigned to school s (= no "blocking pairs")
- **Problem:** in the Bayesian Nash eq. of the school choice game, stability is never guaranteed ex post
- However, we show that under fairly general conditions, the matching outcome is **asymptotically stable** :

$$\#\text{ of students per school } \uparrow \Rightarrow \text{ P(blocking pairs)} \to 0$$

- **Intuition:** uncertainty in school admission cutoffs decreases with market size $\Rightarrow$ the probability that a student "misses" her best possible school converges to zero

Identifying Assumption: Stable Matching

- **How to estimate preferences** under the assumption that the matching outcome is asymptotically stable?
- Under **asymptotic stability**:
  - truth-telling may not be satisfied: some omitted schools may be better than some of the ranked schools
  - but almost all students are assigned to their most preferred school among ex post feasible ones
- Preferences can be **point identified** using discrete choice models
- **Key differences with traditional approach** (truth-telling):
  - instead of using unrestricted choice sets, construct personalized setss that only include *ex post* feasible schools
  - instead of using rank-ordered lists, consider only students' assigned school

$\rightarrow$ **Monte Carlo simulations** to compare both approaches

Monte Carlo Simulations: Setup

- **Market size**:
  - N students (in most simulations, N=750)
  - 5 schools with equal capacity ($N/5 = 150$)
- Student i's **utility** from attending school s is:

$$U_{i,s} = \theta_s + \epsilon_{i,s} \quad \forall\, i, s$$

  θ_s: school fixed effects

  $\epsilon_{i,s}$: i.i.d. idiosyncratic draws from type-I Extreme Values

- **Student priorities** are school-specific and uniformly distributed on the unit interval:

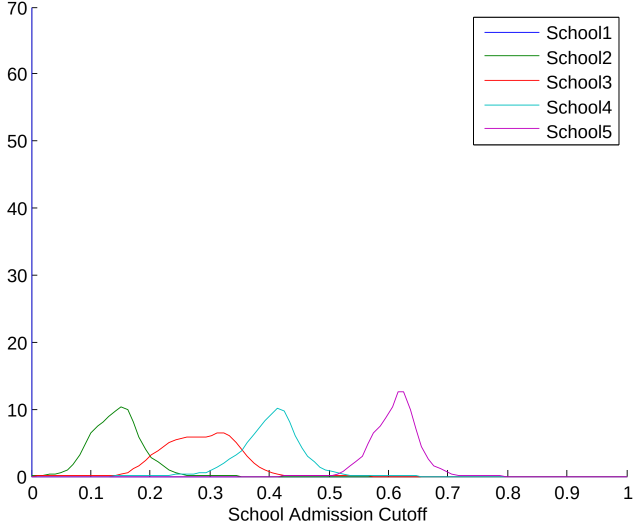
$$e_{i,s} \sim U(0,1)$$

  Priorities are correlated across schools ($\rho = 0.8$) and are known with certainty by students (we relax this assumption later)

Degenerating Cutoff Distribution

- **Benchmark: the unconstrained student-proposing DA**
  - students submit truthful and complete rankings of all 5 schools
  - 300 Monte Carlo samples
- The **distribution of school admission cutoffs**
  - is jointly normal
  - degenerates to mass points as the number of students increase

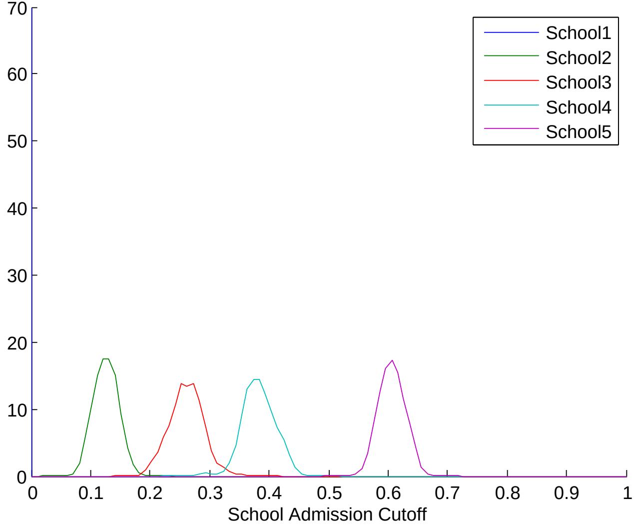
Distribution of Cutoffs: Kernel Density Estimates (300 MC Replications)



N=250

Note: cutoff of school 1 = 0 in all replication samples

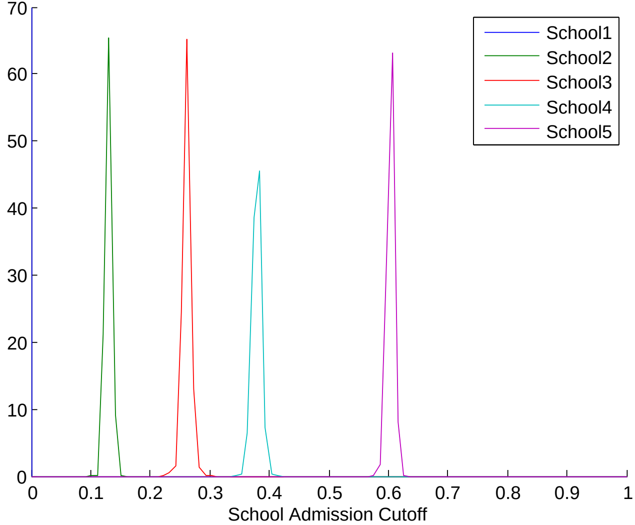
Distribution of Cutoffs: Kernel Density Estimates (300 MC Replications)



N=750

Note: cutoff of school 1 = 0 in all replication samples

Distribution of Cutoffs: Kernel Density Estimates (300 MC Replications)



N=10,000

Note: cutoff of school 1 = 0 in all replication samples

Monte Carlo Simulations: Constrained DA

- **Constrained DA:** students can only submit a list of up to K schools ($K < 5$). In most simulations, $K = 2$
- Given the equilibrium distribution of cutoffs, each student selects the K-**school list that maximizes her expected utility** out of all possible K-school lists that preserve the student's relative ranking of schools
- The **equilibrium distribution of cutoffs** is solved by finding a fixed point

Model Estimation under Alternative Assumptions

- **Simulated Data**: 300 Monte Carlo samples of students' submitted 2-school lists and school admission cutoffs
- The parameters of the model are estimated by maximum likelihood, under **alternative identifying assumptions**
  - **TT** (Truth-Telling): students rank their top-2 schools
  - **AS** (Asymptotic Stability): students are assigned to their preferred school among the *ex post* feasible ones

Model Estimation under Alternative Assumptions (750 Students, 5 Schools, 2 Choices, 300 MC Replications)

- **Model TT (Truth-Telling): students rank their top 2 most preferred schools**

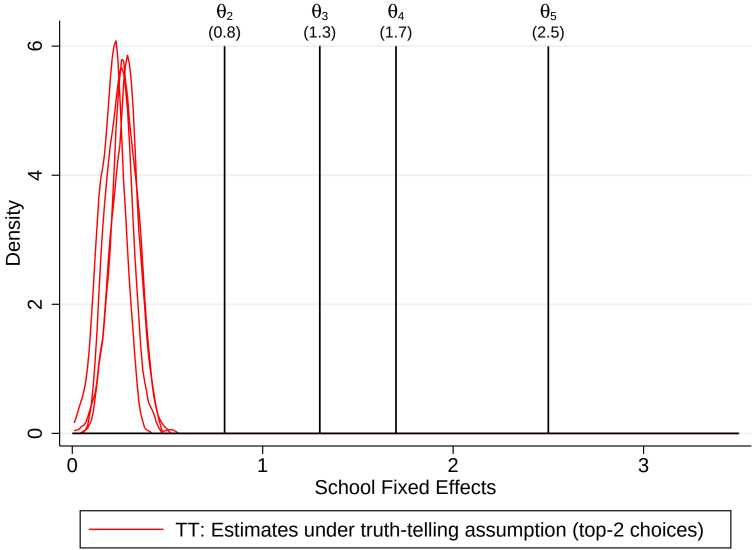
School fixed effects	Parameter value	Mean of MC Estimates	S.D.	MSE
School 2	0.80	**0.28**	0.07	0.280
School 3	1.30	**0.25**	0.07	1.113
School 4	1.70	**0.28**	0.07	2.028
School 5	2.50	**0.20**	0.07	5.272

Fixed effect of school 1 is normalized to 0.

→ Model performs very poorly

→ Identifying assumption is true for only 55.2% of students

Model Estimation under Alternative Assumptions (750 Students, 5 Schools, 2 Choices, 300 MC Replications)



Model Estimation under Alternative Assumptions
(750 Students, 5 Schools, 2 Choices, 300 MC Replications)

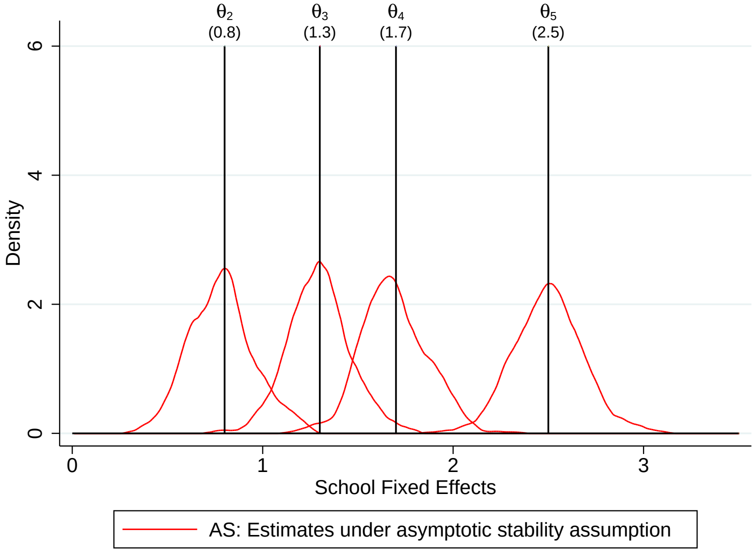
- **Model AS (Asymptotic Stability): students are assigned to their most preferred among the ex post feasible schools**

School fixed effects	Parameter value	Mean of MC Estimates	S.D.	MSE
School 2	0.80	**0.78**	0.17	0.028
School 3	1.30	**1.29**	0.16	0.026
School 4	1.70	**1.69**	0.17	0.029
School 5	2.50	**2.50**	0.18	0.032

→ Model performs well

→ Identifying assumption is true for 99.5% of students

Model Estimation under Alternative Assumptions
(750 Students, 5 Schools, 2 Choices, 300 MC Replications)



Sensitivity Analysis

- We investigate the **sensitivity of the asymptotic stability approach to:**
  - market size (number of students)
  - number of schools that students can rank
  - students' uncertainty about their own priorities

Sensitivity Analysis

School F.E.	True Value	Mean Estimates			
School 2	0.80	0.78	0.70	0.67	0.79
School 3	1.30	1.29	1.14	1.11	1.30
School 4	1.70	1.69	1.56	1.49	1.69
School 5	2.50	2.50	2.42	2.29	2.50
Model Setting:					
Number of students		750	100	750	750
Number of schools		5	5	5	5
Max number of choices		2	2	2	3
Uncertainty in priorities		No	No	Yes	Yes

Conclusion

- **Traditional approach has serious limitations:**
  - with school choice data, truth-telling assumption is too strong
  - we can't just apply discrete choice models
- **Alternative approach: asymptotic stability**
  - in large markets, condition is satisfied under quite general conditions
  - implies that with very high probability, students are assigned to their preferred school among *ex post* feasible ones
  - discrete choice models with personalized choice set can be easily applied
- **Next step:** derive moment inequalities to use information from the full list of submitted choices