Leveraging Uncertainties to Infer Preferences: Robust Analysis of School Choice

Yeon-Koo Che* Dong Woo Hahm† Yinghua He‡

November 14, 2019

Preliminary and Incomplete. Comments are welcome.

Abstract

Recent evidence shows that market participants make mistakes (even) in a strategically straightforward environment but seldom with significant payoff consequences. Uncertainties arising from the use of lotteries or other sources increase payoff consequences of certain mistakes, and force participants to take care to avoid them. Consequently, uncertainties limit the extent to which certain mistakes are made, thus making it possible for one to infer some preference relations reliably. We propose a novel method of exploiting the uncertainties present in a matching environment to systematically and robustly infer student preferences over schools based on students' submitted rank-order lists under the Deferred-Acceptance mechanism (DA). Applicable to all common DA-based school choice systems (e.g., with or without lotteries), our method consists of three steps: (i) tracing out all uncertainties present in the environment, (ii) inferring preferences under each realization of the uncertainties, and then (iii) extending the inferred preferences via the axiom of transitivity. Depending on the nature of uncertainties, the method rationalizes a variety of methods in the literature, ranging from the truth-telling assumption at one extreme (full-support uncertainty) to the stability assumption at the other (limited uncertainty). Further, considering students' payoff-relevant mistakes, we refine our method to strengthen the robustness of the inferred preferences and propose a data-driven testing procedure to balance the tradeoff between robustness and efficiency in preference estimation. We apply our methods to estimate student preferences through a Monte Carlo analysis capturing canonical school choice environment with single tie-breaking lotteries. Finally, we apply our method as well as other existing methods to New York City high school assignment data to explore their implications for preference estimation and counterfactual analysis under a possible policy intervention.

JEL Classification Numbers: C70, D47, D61, D63.
Keywords: School choice, Strategic mistakes, Demand estimation

*Department of Economics, Columbia University, USA. Email: yeonkooche@gmail.com.
†Department of Economics, Columbia University, USA. Email: dongwoo.hahm@columbia.edu.
‡Department of Economics, Rice University, USA. Email: yinghua.he@rice.edu.

1 Introduction

The primary appeal of a strategy-proof mechanism such as Gale and Shapley's Deferred-Acceptance algorithm (henceforth DA) is to make the environment straightforward for the participants by making it a dominant strategy for reporting their preferences truthfully.[1] Strategy-proofness eliminates the need for strategic thinking in application and thus minimizes the scope for mistakes. Further, strategy-proofness aids empirical research by making participants' choices easy to interpret.

Contrary to this conventional wisdom, however, a growing number of studies document that strategic mistakes are not uncommon even in strategy-proof environments. Laboratory and field experiments show that a significant fraction of participants do not report their preferences truthfully,[2] and similar evidence is found even in high-stake real world contexts.[3] Li (2017) provides a theoretical perspective for these findings, suggesting that the strategy-proofness of mechanisms such as DA may not be "obvious" enough for market participants to behave according to the theory.

These recent findings raise serious concerns about our ability to predict or evaluate the outcome of any matching market both theoretically and empirically. First, since dominance is such a basic solution requirement, if it cannot be relied upon as a behavioral assumption, nothing much in the arsenal of economic analytical tools is available for predicting participants' behavior and the allocation outcome resulting from it. Second, the lack of a behavioral prediction could make it all but impossible to interpret the data—such as preferences expressed or the allocation outcome—collected from any matching market including the ones using a strategy-proof mechanism.

Nevertheless, not all is lost. Artemov, Che, and He (2017) (ACH, hereafter) demonstrate empirically that while mistakes are not uncommon in many strategy-proof environments, the super majority of them involve little payoff losses. Namely, mistakes that involve large payoff losses are rather uncommon. Based on this evidence, ACH develop a solution concept, *robust equilibrium*, to accommodate such mistakes with little payoff consequences. Specifically, a robust equilibrium requires that, for any $\epsilon > 0$, all participants act in a way that guarantee them payoffs that are within $\epsilon > 0$ of what they would obtain from a best response (which in the case of strategy-proof mechanism coincides with truthful-reporting) if the economy is sufficiently large (in terms of the number of participants). They apply the solution concept to the DA matching environment in which schools evaluate students via strict rankings (e.g., admission is based on test scores) and students know their priorities before they apply to schools. Given this assumption, they show that, even though students' behavior in a robust equilibrium can depart drastically from truthful reporting and cannot be reliably predicted due to severe multiplicity, the allocational outcome is virtually unique in a large economy, converging to a stable matching. These results lead them to propose stability or its weakening, robust stability, rather than truth-telling, as the identifying assumption for estimating student preferences with school choice data.

However, the strict priority environment assumed by ACH limits the applicability of their proposed method. Essentially, this assumption means that at the time of application students face little uncertain-

[1] See Gale and Shapley (1962), Dubins and Freedman (1981), and Roth (1982).

[2] See Chen and Sönmez (2002), Chen and Kesten (2014), and Rees-Jones and Skowronek (2018), to name just a few.

[3] Hassidim, Romm, and Shorrer (2016) find in their analysis of Israeli postgraduate psychology program admissions that approximately 19% of students either did not list a scholarship position for a program or ranked a non-scholarship position higher than the corresponding scholarship position, against their interests. Rees-Jones (2017) reports that 17% of the 579 surveyed US medical seniors indicate misrepresenting their preferences in the National Resident Matching Program.

ties on the set of schools that will be ex-post feasible to them. In particular, it means that no lotteries are needed to break ties. The assumption, while valid in the environment where admissions are based on exam scores (e.g., in Australian college admissions) or academic performance (e.g., high school admissions in Paris), does not hold in many environments where priorities are coarse and lotteries are therefore needed to break ties.[4] Even when exam scores are used to select students, if students must submit rank-order lists (ROLs) of schools before the exam results are announced, they face some uncertainties.

The limited uncertainties assumed by ACH also means that the robust equilibrium concept places very little discipline on equilibrium strategies, which in turn means that very little use can be made of the rank-order lists submitted by students. Indeed, the ROLs can be used only to invoke (robust) stability, leading to much of the ROL information being "thrown away."[5]

In practice, students face uncertainties because lotteries are used to break ties under coarse priorities or their priorities are not known at the time of application. These uncertainties could expose students to serious payoff risks and force them to act more carefully in a strategy-proof environment. From the researcher's standpoint, the uncertainties therefore present opportunities for inferring student preferences that would otherwise be difficult to learn. To see this, consider an extreme situation called a *full-support uncertainty*: the density of uncertain scores (or lottery draws) of each student is strictly positive for its entire support. An example is when schools have no priorities and Multiple-Tie-Breaking (MTB) lotteries are used in DA. In this case, a student must report her preferences truthfully up to an (almost surely) undersubscribed school, or else she will suffer a non-vanishing payoff loss. Hence, robust equilibrium leads to the so-called **weak-truth-telling** (WTT) behavior: a student reports her top K schools in the truthful order, for some $K \geq 1$.

However, the full-support uncertainty is unlikely to hold in most circumstances. First, schools often have priorities over students, which students know prior to application. Even though the priorities may be coarse, a student could have a very high or low priority at a school, in which case she faces no uncertainty about the (ex-post) feasibility of that school, should she top-rank it, for instance. In particular, suppose a school is clearly out of reach for a student; in that case, she may not bother to apply to that school, regardless of her preferences. Whether she ranks the school or not may not reveal her preference of school in a robust equilibrium. At the same time, preferences for schools at which she faces significant uncertainties (e.g., schools are neither clearly feasible nor clearly infeasible) will be more clearly revealed. Second, the full-support uncertainty fails under the commonly adopted Single Tie-Breaking (STB) lotteries rule, which uses the same lottery draw to break ties at all schools. For example, suppose schools have no priorities over students (so the usual reason for mistakes does not apply), and suppose there are a continuum of students and three schools, a, b, c, each with a continuum of seats. Suppose in equilibrium the schools' (degenerate) cutoffs for lottery draws under DA are $P_a > P_b > P_c$, meaning school $j = a, b, c$ is feasible to a student with a lottery draw ℓ only if $\ell > P_j$. Consider a student whose true preferences are b-a-c, ranked in the descending order. Unlike before, here the student is uncertain about which school

[4] In many school choice programs, schools use a small number of criteria such as sibling attendance and walk zone to prioritize students, which leaves many students to fall into the same priority class, and the resulting ties are broken by lotteries.

[5] A simple intuition is that the environment each student facing becomes largely degenerate in large enough economies, which admits there being so many ways for the student to deviate from her best response (i.e., truth-telling) without suffering serious payoff loss.

would be feasible to her. Nevertheless, no matter her draw, school a is irrelevant for her, since whenever a is feasible to her, b is also feasible to her, and she prefers b over a. Consequently, any of the three ROLs, b-c, b-c-a, and b-a-c, would be optimal. In other words, students need not report their ROLs truthfully. At the same time, however, the uncertainties facing students makes it hopeful that one can extract more meaningful information from the observed ROLs: in this example, no matter which of the three ROLs is observed, we should be able to infer that the student prefers b to both c and a. This is much more than one could learn from invoking stability from a realized assignment: for instance, if the student is assigned c in the data, nothing could be learned. The question is: what preference relations can one infer from the data? The answer depends on, and therefore requires deep understanding about, the structure and nature of uncertainties present in the matching environment.

In this paper, we develop a method that exploits the uncertainties present in a matching environment to extract maximal preferences inferences from the observed ROL data, consistent with the robust equilibrium concept. To this end, we first explore the theoretical implications of robust equilibria on the inferable preferences. Specifically, Theorem 1 establishes that *the proportion of the "stably-assigned" students—i.e., those assigned the favorite feasible schools given the realized uncertainty—converges to one in probability.* Despite the resemblance to the similar result established in ACH, the presence of uncertainties faced by students requires a different proof argument,[6] and more importantly yields different implications for preference inferences. What it means is that, in the limit, the ROL a student submits in a robust equilibrium must be such that for each realization of uncertainties (e.g., lottery draws), she must receive the most preferred school among all feasible schools *given each realization.* Importantly, this means that if one were to simulate all possible realizations of uncertainties present in the matching environment, say by running the DA algorithm with all possible lottery draws, for each student and a given ROL she submitted, one would be able to identify all stable assignments and the set of feasible schools for each realization, and they would constitute all possible set of preference relations one could infer from a robust equilibrium.

To illustrate, recall the above example with STB, and suppose the student (whose true preference order is b-a-c) did submit the ROL of b-c-a (which is a robust equilibrium behavior). Then, in a large enough economy, with cutoffs $P_a > P_b > P_c$, simulation of lotteries yields the following four positive-probability events: (i) the student is assigned b and schools a and c are feasible to her; (ii) the student is assigned b and school c, but not a, is feasible to her; (iii) the student is assigned c and no other school is feasible; and (iv) no school is feasible to the student.[7] By simulating STB, one could "trace out" all these events, and the resulting implication should be: b is revealed preferred to a and c, with no conclusion drawn on the preference between the latter two schools.

The theory therefore leads to the following three-step procedure, called *Transitive Extension, or TE*:

(i) **Tracing out the uncertainties present in the matching environment:** When the uncertainties arise from a clearly-defined lottery procedure, we can simply simulate the lottery and run the matching algorithm numerous times, to trace out all possible outcomes. If the uncertainty arises

[6] The proof of the analogous result in ACH (Theorem 2) rests on the vanishing uncertainty faced by students about their feasible sets in the limit, which does not extend in the current environment. The argument instead focuses on the robust best response to non-vanishing uncertainty.

[7] Event (i) occurs when the student's score is higher than P_a, event (ii) occurs when her score is in between P_b and P_a, event (iii) occurs when her score is between P_c and P_b, and event (iv) occurs when her score is below P_c.

from other sources, such as test scores that are yet to be realized at the time of application, then one may build an empirical model to simulate the distribution of test scores say based on the observables (which may include various performance measures from preceding schools say). One may also include small-sample uncertainties about the empirical distribution of student types by, for example, bootstrapping.

(ii) **Inferring preference relations from realized uncertainties:** For each realization of uncertainties, we identify the assigned school and the schools that are feasible at the same time. Invoking stability, we infer that the student prefers the former over each of the latter schools.

(iii) **Extending preference relations:** The second step produces a collection of preference relations for each student. We then link them via the axiom of transitivity whenever possible. Specifically, an algorithm is developed to recursively perform *transitive closure* operation, which at termination produces the maximal set of preference relations based on transitivity.

The preferences inferred by TE constitute the essential information for estimating student preferences, similar to, but richer than, the usual information on observed choices and choice sets in discrete choice models. In effect, given each realized uncertainty, step (ii) constructs for a student a (hypothetical) choice situation in which she faces a choice set and chooses one option. Step (iii) links all choice situations together for each student.

The TE procedure can be extended to accommodate violations to stability in some realization of uncertainties and thus allow students make payoff-relevant mistakes. This is necessary in the sense that our theory permits violations of stability in any finite economy and that payoff-relevant mistakes, albeit relatively rare, have been empirically documented in a growing literature.

We associate TE with a threshold probability to allow for mistakes. Specifically, a TE with threshold t modifies step (ii) such that, across all realized uncertainties, an assigned school, a, is inferred preferred to another school, b, only if the probability of a being the assignment and, *simultaneously*, b being feasible is above t. Step (iii) is similarly modified. These modifications allow students to make mistakes when certain choice situations are small-probability events.

More importantly, a TE with s threshold allows us to construct a data-driven way to choose the optimal balance between estimation efficiency and robustness to mistakes. We show that student preference relations inferred by a TE with a higher threshold is always a subset of those inferred by a TE with a lower threshold. Moreover, the usual truth-telling assumption leads to the largest possible set of inferred preference relations. This nesting structure leads to a testing procedure that can allow us to choose the optimal method among the TEs with different thresholds and the truth-telling assumption.

We apply this TE procedure and the associated testing procedure to estimate preferences of market participants in a hypothetical Monte Carlo environment to highlight the extent to which the method can make better use of information than the simple stability method or its weakening proposed by ACH and the extent to which it avoids misclassification of preferences that may result from ignoring mistakes made by market participants. Our theoretical results are illustrated in the simulations. We also apply the method to estimate preferences of students to New York City public high schools, and compare the results with those obtained by applying other existing methods.

Literature Review: To be completed.

2 Theoretical Analysis

2.1 Primitives

We begin with Azevedo and Leshno (2016) (denoted as AL) as our modeling benchmark. A (generic) economy consists of a finite set of *schools* $C = \{c_1, ..., c_C\}$ and a set of *students*. Each student has an ex ante type (u, p), where $u = (u_1, ..., u_C) \in [\underline{u}, \overline{u}]^C$ is a vector of von-Neumann Morgenstern utilities of attending schools for some $\underline{u} \leq 0 < \overline{u}$, and $p = (p_1, ..., p_C) \in \mathcal{P} \subset [0,1]^C$ is a vector of scores representing students' "intrinsic" priorities at the schools, with a student with a higher score having a higher priority at a school. The distribution of these scores and the associated support are arbitrary. In particular, we have in mind the case in which the support is finite, meaning a mass of students may be tied at the same score, thus requiring some tie-breaking procedure. We assume that a student with intrinsic priority p realizes scores $s = (s_1, ..., s_C) \in [0,1]^C$ are distributed some distribution G_p that depends on intrinsic priority p. Although only final scores s matter for the assignment, given all students' rank order lists, we still need to keep track of their ex ante types (u, p), because students only know this when they submit their applications. Let η denote the probability measure on (u, s). Its marginal distribution on s_c is assumed to be convex (i.e., forms an interval) and admits a strictly positive density in its interior (so it is atomless).

A variety of school choice settings fall into this model, including the following representative situations:

- Merit-based assignment with known score: In many school choice settings, the priorities are given by some merit scores which are known to students when they apply. Examples include Australian college admissions and Paris high school assignment. This is the case studied by Artemov, Che, and He (2017), and is a special case of our model where G_p is degenerate: i.e., $s = p$.

- Coarse priorities: In many school choice contexts, intrinsic priorities are determined by a handful of criteria, so $\mathcal{P}$ is finite, and these priorities lexicographically dominates the lottery draws. In this case, $\text{Supp}(G_p^c) \cap \text{Supp}(G_{p'}^c) = \emptyset$ for any $p_c \neq p'_c$, where G_p^c is the marginal distribution of students' scores at school c. More specifically, the distribution G_p can be constructed in the following way. First, for each school c, the intrinsic priority score is represented by $q_c \in \{0, \frac{1}{n_c}, ..., \frac{n_c - 1}{n_c}\} =: Q_c$, where n_c is the number of priority groups or intrinsic priorities for school c. The lottery score, ℓ, is either uniform on $[0,1]^C$ in the case of the multiple tie-breaking (MTB) rule or $\ell_1 = ... = \ell_C$ and uniform on $[0,1]$ in the case of the single tie-breaking (STB) rule. Either case, G_p is induced by a composite score $s_c = q_c + \frac{1}{n_c}\ell_c$ for each school c.

- Merit-based assignment with unknown scores: NYC has several specialized high schools, known as "exam" schools, which assign seats based on serial dictatorship in which the serial order is given by students' scores from the SHSAT (Specialized High School Admissions Test). Students submit their ROLs before they take the SHSAT. This can be modeled by interpreting their ex ante score p as their belief about their own abilities. A student with a higher p is more likely to score highly than a student with lower p. Unlike the merit-based assignment without uncertainty, students face uncertainty about their scores at the time of submitting their ROLs. Hence, G_p is typically non-degenerate. Since the SHSAT score is one dimensional, the scores do not have full support, but serial dictatorship guarantees that the cutoff scores are unique given any distribution of ROLs.

A vector u induces ordinal preferences over schools, denoted by an ROL of "acceptable" schools, $\rho(u)$, of length $0 \leq K \leq C$.[8] Assuming that a student has an outside option of zero payoff, the model allows for the possibility that students may find some schools unacceptable. Let $\Theta = [\underline{u}, \overline{u}]^C \times [0,1]^C$ denote the set of student types.

A *continuum economy* consists of the same finite set of schools and a unit mass of students with type $\theta \in \Theta$ and is given by $E = [\eta, S]$, where η is a probability measure over Θ representing the distribution of the student population over types, and S represents the masses of seats $S = (S_1, ..., S_C)$ available at the schools, where $S_j > 0$ and $\sum_{j=1}^{C} S_j < 1$. We assume that η admits continuous density that is positive in the interior of its support (i.e., full support). In the case of serial dictatorship, this assumption holds with a reduced dimensionality of support; students' scores are one-dimensional numbers in $[0,1]$. The atomlessness ensures that indifference either in student preferences or in school priorities arises only for a measure 0 set of students. The full-support assumption means that both students' preferences and schools' priorities are rich (except for the case of serial dictatorship). A *matching* is defined as a mapping $\mu : C \cup \Theta \to 2^{\Theta} \cup (C \cup \Theta)$ satisfying the usual two-sidedness and consistency requirements as well as "open on the right" as defined in AL (see p. 1241). A *stable matching* is also defined in the usual way satisfying individual rationality and no-blocking.[9]

According to AL, a stable matching is characterized via *market-clearing cutoffs*, $P = (P_1, ..., P_C) \in [0,1]^C$, satisfying the demand-supply condition $D_c(P) \leq S_c$ with equality in case of $P_c > 0$ for each $c \in C$, where the demand $D_c(P)$ for school c is given by the measure of students whose favorite school among all feasible ones (i.e., with cutoffs less than her scores) is c. Specifically, given the market-clearing cutoffs P, the associated stable matching assigns those who demand c at P to school c. Given the full-support assumption, Theorem 1-i of AL guarantees a unique market-clearing price P^* and a unique stable matching μ^*. Given the continuous density assumption, $D(\cdot)$ is C^1 and $\partial D(P^*)$ is invertible.

With the continuum economy E serving as a benchmark, we are interested in a sequence of finite random economies approximating E in the limit. Specifically, let $F^k = [\eta^k, S^k]$ be a *k-random economy* that consists of k students each with type θ drawn independently according to η, and the vector $S^k = [k \cdot S]/k$ of capacity per student, where $[x]$ is the vector of integers nearest to x (with a rounding down in case of a tie); η^k is a random measure. A matching is defined in the usual way.

Consider a sequence of k-random economies $\{F^k\}$. A student-proposing DA is used to assign students to schools. We assume that schools act passively and report their priorities and capacities truthfully. We are interested in characterizing an equilibrium behavior of the students in the DA. In one sense, this task is trivial: since the DA is strategy-proof, it is a weakly dominant strategy for each student to rank order all acceptable schools (i.e., with payoff $u_c \geq 0$) according to her true preference order. We call such a strategy the **truthful-reporting strategy** (TRS).[10] We are, however, interested in a more

[8] In the case of a tie, $\rho(u)$ produces a ranking by breaking the tie in some arbitrary (but exogenous) way. Since we shall assume that the distribution of the types is atomless, the tie-breaking becomes immaterial.

[9] Individual rationality requires that no participant (a student or a school) is assigned a partner that is not acceptable. No blocking means that no student-school pair exists such that the student prefers the school over her assignment and the school has either a positive measure of vacant positions or admits a positive measure of students whom the school ranks below that student.

[10] By this definition, TRS does not allow students to rank unacceptable schools, which reduces the multiplicity of equilibria and thus gives TRS a better chance to be the unique equilibrium. However, applying to unacceptable options can happen in real-life matching markets, as documented in He (2017).

robust solution concept allowing for any approximately-optimal behavior. We assume that each student observes her own type θ but not the types of other students; as usual, all students understand as common knowledge the structure of the game. Given this structure, the DA induces a Bayesian game in which the strategy of each student specifies a distribution over ROLs of length no greater than C as a function of her type θ.

In any game (either the limit or the k-random economy), student i's Bayesian strategy is a measurable function $\sigma_i : \Theta \to \Delta(\mathcal{R})$, where $\mathcal{R}$ is the set of all possible ROLs a student may submit. Note that the strategies can be asymmetric; i.e., we do not restrict attention to symmetric equilibria. In any economy (either continuum or finite), a profile of student strategies induces cutoffs $P \in [0,1]^C$ and a distribution of types of students assigned to alternative schools, which we refer to as *an outcome*. We say a school c is *feasible* to a student if her score at c is no less than P_c, and we say that a student *demands* school c if c is feasible and she ranks c in her ROL ahead of any other feasible schools.

We are interested in the following solution concept:

DEFINITION 1. *For a sequence $\{F^k\}_k$ of k-random economies, the associated sequence $\{(\sigma^k_{1\le i\le k})\}_k$ of strategy profiles is said to be a* ***robust equilibrium*** *if, for any $\epsilon > 0$, there exists $K \in \mathbb{N}$ such that for $k > K$, $\{(\sigma^k_{1\le i\le k})\}_k$ is an interim ϵ-Bayes Nash equilibrium—namely, for i, σ^k_i gives student i of each type θ a payoff within ϵ of the highest possible (i.e., supremum) payoff she can get by using any strategy when all the others employ σ^k_{-i}.*

This solution concept is arguably more sensible than the exact Bayesian Nash equilibrium if, for a variety of reasons, students may not play their best response exactly, but they do approximately in the sense of not making mistakes of significant payoff consequences, in a sufficiently large economy.

Indeed, without relaxing a solution concept in some way, one cannot explain the kind of departure from the dominant strategy documented in the literature. To see why, recall that in any finite economy, the cutoffs are not degenerate, so any strategy departing from TRS will result in payoff loss with a positive probability. Hence, to explain a behavior that departs from a dominant strategy, one must relax exact optimality on the students' behavior. At the same time, a solution concept cannot be arbitrary, so some discipline must be placed on the extent to which payoff loss is tolerated. The robust equilibrium concept fulfills these requirements by allowing students to make some mistakes but requiring that the payoff losses from the mistakes disappear as the market grows arbitrarily large.

2.2 Analysis of Robust Equilibria

Robust equilibria require that students make no mistakes with any real payoff consequences as the market gets large. Does this requirement mean that a large fraction of students must report truthfully in a strategy-proof mechanism? We show below that this need not be the case. Specifically, we will construct strategies that are not TRS, yet do not entail significant payoff loss for almost all types of students in a large economy.

The only thing we do differently is to define type $\theta = (u, s^t)$ where $s^t := \sup(\text{Supp}(G_t))$. A **stable response strategy** (SRS) is a strategy whereby a student reports $R \in \Delta(\mathcal{R})$ that would obtain exactly the school that she would in the continuum economy had she reported truthfully. Obviously, TRS is an SRS, but an SRS may deviate from TRS for two reasons. One is due to the intrinsic priorities. Suppose

that a student has a type $s_c^t < P_c^*$ for some school c, meaning in the continuum economy the school c will be out of reach for this student. In this case, simply dropping the school in one's ROL, even though the school may be her most preferred school, would be an SRS. This situation arises for a reason already recognized from Artemov, Che, and He (2017).

Interestingly, an SRS may differ from TRS even when there are no intrinsic priorities. Suppose that assignment is based solely on STB lotteries. And the continuum economy cutoffs under TRS are such that $P_1^* > P_2^* > \cdots > P_C^*$. In this case, if a student has preferences $1 \succ 2 \succ \cdots \succ C$, she would obtain each school with positive probability in the continuum economy. Hence, her unique SRS is TRS. But consider a student with preference 2-1-3-$\cdots$-C, in the descending order. In this case, she will never get assigned to school 1 in the continuum economy under TRS, since whenever her lottery draw ℓ is good enough to be admitted by 1, she will be admitted by 2. In this case, she could simply drop school 1 or insert it anywhere after school 2 in her ROL. That is, she has $(C+1)$ SRSs, only one of which is TRS. In fact, any student with preference differing from 1-2-3-$\cdots$-C, would typically have many SRSs that deviate from TRS.

Definition 2. *For a sequence $\{F^k\}_k$ of k-random economies, the associated sequence $\{(\sigma^k_{1\leq i\leq k})\}_k$ of strategy profiles is said to be **asymptotically stable** if the fraction of students assigned their most preferred feasible schools (given the equilibrium cutoffs and the realized lottery draws or "states" of the world) converges in probability to 1 as $k \to \infty$.*[11]

We call a sequence $\{(\sigma^k_{1\leq i\leq k})\}_k$ of strategy profiles **regular** if there exists some $\gamma > 0$ such that the proportion of students playing TRS is at least γ.

We now state the main theorem:

Theorem 1. *Any regular robust equilibrium is asymptotically stable.*

What this means is that in a large enough economy, for almost all students, for almost all realization of lottery draws, the schools they are assigned are more preferred to those that are feasible given the realized lottery draws (even though they are unknown to the students at the time of their submitting ROLs). Hence, this theorem formally validates the preference relations extracted from our transitive extension procedure that are defined in the next section.

3 Empirical Method: Transitive Extension

Building upon the theoretical results from the previous section, we develop an empirical method for analyzing data on school choice, called Transitive Extension (TE). TE has two advantages over the existing alternatives: (i) it extracts as much information as possible from the data, and (ii) it is robust to students' possible mistakes.

Suppose that we have access to a dataset on a random finite economy F^k in which k students compete for admissions to C schools. Students adopt robust equilibrium strategies, $\sigma^k_{1\leq i\leq k}$. Each school has a finite capacity, and students are assigned through the DA mechanism. Besides the submitted ROLs and

[11] More formally, we require that for any $\epsilon > 0$ there exists $K \in \mathbb{N}$ such that in any k-random economy with $k > K$, with probability of at least $1-\epsilon$, at least a fraction $1-\epsilon$ of all students are assigned their most preferred feasible schools given the equilibrium cutoffs P^k.

the outcome, the researcher observes intrinsic priorities, student characteristics, school attributes, and lottery draws.

The TE method has three steps:

(i) **Tracing out all the uncertainties present in F^k:** From the point view of a student in F^k, there are two sources of uncertainties. First, her own ex-post score is uncertain due to, e.g., the use of lottery to break ties. Second, other students' types are random. To trace out all uncertainties and all possible outcomes for every student, we bootstrap from F^k and create multiple economies of size k, simulate the lottery, and run the matching algorithm numerous times.[12]

(ii) **Extracting preference relations from realized uncertainties:** For each realization of uncertainties, for every student, we identify the assigned school and the schools that are feasible at the same time. Based on Theorem 1, we then invoke stability to infer that the student prefers the former over each of the latter schools.

(iii) **Transitive extension:** The second step produces a collection of preference relations for each student. We then link them via the axiom of transitivity.

In the following, Section 3.1 formalizes the TE procedure; Section 3.2 extends the procedure to allow for mistakes; and Section 3.3 presents a series of statistical tests that can help us choose among different empirical methods.

3.1 Transitive Extension

We consider student i in economy F^k, whose submitted ROL is $R_i = c_{|R_i|}\text{-}c_{|R_i|-1}\text{-}\ldots\text{-}c_2\text{-}c_1$ and whose her intrinsic priorities are p_i.

The collection of i's possible feasible sets is denoted by $\mathcal{F}(R_i, p_i) \subseteq 2^C$ and can be recovered as described above. An element in $\mathcal{F}(R_i, p_i)$ is $\mathbb{F}$, which is i's feasible set under some realized uncertainties.

For any $\mathbb{F} \in \mathcal{F}(R_i, p_i)$, let $A(\mathbb{F}, R_i) \in C \cup \{\emptyset\}$ be the school that i is assigned when the realized feasible set is $\mathbb{F}$. By invoking stability, we infer that i prefers $A(\mathbb{F}, R_i)$ to any other school in $\mathbb{F}$. There can be multiple realized feasible sets that lead to the same assignment, i.e., $A(\mathbb{F}_1, R_i) = A(\mathbb{F}_2, R_i)$ for $\mathbb{F}_1, \mathbb{F}_2 \in \mathcal{F}(R_i, p_i)$. We define $\tilde{\mathcal{I}}^c(R_i, p_i)$, for $c = c_1, \ldots, c_{|R_i|}$, which is the set of schools that are ever feasible to i when c is the assigned school:

$$\tilde{\mathcal{I}}^c(R_i, p_i) = \begin{cases} \emptyset & \text{if } \nexists \mathbb{F} \text{ s.t. } \mathbb{F} \in \mathcal{F}(R_i, p_i) \ \&\ c = A(\mathbb{F}, R_i); \\ \bigcup_{1 \le l \le |\mathcal{F}(R_i,p_i)|:\ \mathbb{F}_l \in \mathcal{F}(R_i,p_i)\ \&\ A(\mathbb{F}_l,R_i)=c} \mathbb{F}_l & \text{otherwise.} \end{cases}$$

Therefore, for school c that is ranked in R_i, $\tilde{\mathcal{I}}^c(R_i, p_i)$ includes c and all the schools that are inferred by stability as less preferred to c.

Relying on $\mathcal{F}(R_i, p_i)$, $A(\mathbb{F}, R_i)$, $\forall \mathbb{F} \in \mathcal{F}(R_i, p_i)$, and $\tilde{\mathcal{I}}^c(R_i, p_i)$, for $c = c_1, \ldots, c_{|R_i|}$, the TE procedure constructs a set of inferred preferences for student i:

$$\mathcal{I}(R_i, p_i) = \left\{\mathcal{I}^{c_1}(R_i, p_i), \ldots, \mathcal{I}^{c_{|R_i|}}(R_i, p_i)\right\},$$

[12] If the uncertainty arises from other sources, such as test scores that are yet to be realized at the time of application, then one may build an empirical model to simulate the distribution of test scores say based on the observables (which may include various academic performance measures).

where, for school c that is ranked in R_i, $\mathcal{I}^c(R_i, p_i)$ includes c and all the schools that are inferred as less preferred to c. By construction, $\tilde{\mathcal{I}}^c(R_i, p_i) \subseteq \mathcal{I}^c(R_i, p_i)$. In the following, we suppress $\mathcal{I}^c(R_i, p_i)$ and $\tilde{\mathcal{I}}^c(R_i, p_i)$'s dependence on R_i and p_i when there is no confusion.

The TE procedure is as follows:

Step 1 Consider c_1, the bottom-ranked school in R_i. Let $\mathcal{I}^{c_1} = \tilde{\mathcal{I}}^{c_1}$.

Step k $(1 < k \leq |R_i|)$ Consider c_k, the kth-ranked school from the bottom in R_i. If there is no $\mathbb{F}$ such that $\mathbb{F} \in \mathcal{F}(R_i, p_i)$ and $c_k = A(\mathbb{F}, R_i)$, let $\mathcal{I}^{c_k} = \tilde{\mathcal{I}}^{c_k} = \emptyset$; otherwise,

$$\mathcal{I}^{c_k} = \tilde{\mathcal{I}}^{c_k} \bigcup \left(\bigcup_{c_{k'} \text{ ranked below } c_k \text{ in } R_i:\ c_{k'} \in \tilde{\mathcal{I}}^{c_k}} \mathcal{I}^{c_{k'}} \right)$$

The preference relations inferred by TE are encoded in $\mathcal{I}^c$ such that c is more preferred than any school in $\mathcal{I}^c \setminus \{c\}$. The key contribution of TE is in the difference between $\tilde{\mathcal{I}}^c$ and $\mathcal{I}^c$. For school c that is ranked in R_i and is i's assigned school with a positive probability, TE infers that:

(i) c is more preferred than all schools in $\tilde{\mathcal{I}}^c \setminus \{c\}$, which are all the schools ever feasible when c is the assigned school; and

(ii) for any $c' \in \tilde{\mathcal{I}}^c \setminus \{c\}$, by preference transitivity, c is more preferred to all the schools that are less preferable to c'.

For school $c' \notin \tilde{\mathcal{I}}^c \setminus \{c\}$ that is ranked lower than c in R_i and has a positive probability of being i's assigned school, TE does not infer any preference relation between c and c'. This is because the relative ranking between c and c' in R_i may be payoff irrelevant when $c' \notin \tilde{\mathcal{I}}^c \setminus \{c\}$ (i.e., c' is never feasible when c is the assigned school).[13]

An Example of TE. There are five schools, $\{c_0, c_1, c_2, c_3, c_4\}$. Let student i's submitted ROL be $R_i = c_4$-c_3-c_2-c_1. Suppose that there are in total 4 realizations of feasible sets, $|\mathcal{F}(R_i, p_i)| = 4$. More specifically, also shown in Figure 1a,

Feasible set		Assigned school
$\mathbb{F}_1 = \{c_4, c_3\}$	and	$A(\mathbb{F}_1, R_i) = c_4$
$\mathbb{F}_2 = \{c_4, c_1\}$	and	$A(\mathbb{F}_2, R_i) = c_4$
$\mathbb{F}_3 = \{c_2, c_1\}$	and	$A(\mathbb{F}_3, R_i) = c_2$
$\mathbb{F}_4 = \{c_1, c_0\}$	and	$A(\mathbb{F}_4, R_i) = c_1$

The TE procedure goes through four steps:

Step 1: Consider the bottom-ranked school, c_1. $\mathcal{I}^{c_1} = \tilde{\mathcal{I}}^{c_1} = \bigcup_{1 \leq l \leq 4:\ c_1 = A(\mathbb{F}_l, R_i)} \mathbb{F}_l = \{c_1, c_0\}$.

Step 2: Consider the 2nd-ranked school from the bottom, c_2. $\tilde{\mathcal{I}}^{c_2} = \bigcup_{1 \leq l \leq 4:\ c_2 = A(\mathbb{F}_l, R_i)} \mathbb{F}_l = \{c_2, c_1\}$; $\mathcal{I}^{c_2} = \tilde{\mathcal{I}}^{c_2} \bigcup \left(\bigcup_{c_{k'} \text{ ranked below } c_2 \text{ in } R_i:\ c_{k'} \in \tilde{\mathcal{I}}^{c_2}} \mathcal{I}^{c_{k'}} \right) = \tilde{\mathcal{I}}^{c_2} \bigcup \mathcal{I}^{c_1} = \{c_2, c_1, c_0\}$.

[13] For example, c is ranked just above c', but they are never both feasible to i in any lottery realization, although either of them can be feasible in some realizations. In this case, the relative ranking between these two schools is payoff irrelevant.

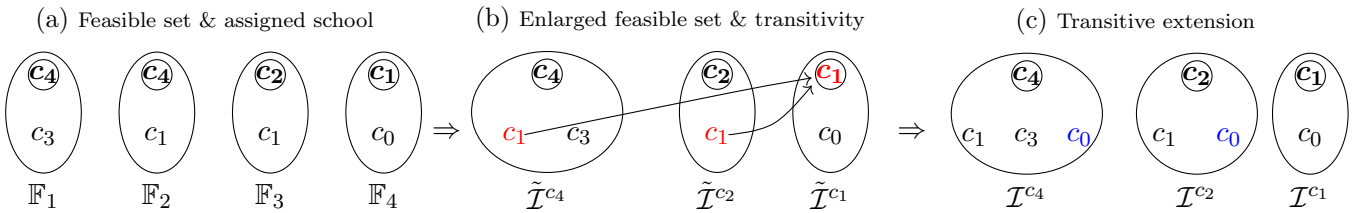


Figure 1: Transitive Extension: An Example

Step 3: Consider the 3rd-ranked school from the bottom, c_3. $\mathcal{I}^{c_3} = \tilde{\mathcal{I}}^{c_3} = \bigcup_{1\leq l\leq 4:\ c_3=A(\mathbb{F}_l,R_i)} \mathbb{F}_l = \emptyset$.

Step 4: Consider the 4th-ranked school from the bottom, c_1. $\tilde{\mathcal{I}}^{c_4} = \bigcup_{1\leq l\leq 4:\ c_4=A(\mathbb{F}_l,R_i)} \mathbb{F}_l = \{c_4, c_3\} \bigcup \{c_4, c_1\} = \{c_4, c_3, c_1\}$; $\mathcal{I}^{c_4} = \tilde{\mathcal{I}}^{c_4} \bigcup \left(\bigcup_{c_{k'} \text{ ranked below } c_4 \text{ in } R_i:\ c_{k'} \in \tilde{\mathcal{I}}^{c_4}} \mathcal{I}^{c_{k'}} \right) = \tilde{\mathcal{I}}^{c_4} \bigcup \mathcal{I}^{c_3} \bigcup \mathcal{I}^{c_1} = \{c_4, c_3, c_1\} \bigcup \emptyset \bigcup \{c_1, c_0\} = \{c_4, c_3, c_1, c_0\}$.

The inferred preferences are encoded in $\mathcal{I}^{c_1}, \ldots, \mathcal{I}^{c_4}$ that are non-empty: $c_1 \succ_i c_0$, $c_2 \succ_i c_1, c_0$, and $c_4 \succ_i c_3, c_1, c_0$. This procedure is illustrated in Figure 1.

3.2 Transitive Extension with Potential Mistakes

As documented extensively in the literature, students make mistakes when playing school choice games. Motivated by our theoretical results, we introduce some "TE threshold" into the TE procedure to accommodate the potential mistakes.

A TE threshold, $t \in [0, 1)$, is to determine if inferred preferences on a given school are reliable; below the threshold, we do not consider the inferred preferences reliable. We now use $\mathcal{I}^{c_k,t}$ to denote the set of schools including c_k and those that are inferred as less preferred to c_k by TE with threshold t. For student i, it works as follows:

Step 1 Consider c_1, the bottom-ranked school in R_i. Let

$$\mathcal{I}^{c_1,t} = \tilde{\mathcal{I}}^{c_1} \setminus \{c \in \tilde{\mathcal{I}}^{c_1} : \Pr(c \ \& \ c_1 \text{ are feasible; all schools ranked above } c_1 \text{ in } R_i \text{ are infeasible}) \leq t\}$$

Step k $(1 < k < |R_i|)$ Consider c_k, the kth-ranked school from the bottom in R_i. If there is no $\mathbb{F}$ such that $\mathbb{F} \in \mathcal{F}(R_i, p_i)$ and $c_k = A(\mathbb{F}, R_i)$, let $\mathcal{I}^{c_k,t} = \tilde{\mathcal{I}}^{c_k} = \emptyset$; otherwise,

$$\mathcal{I}^{c_k,t} = \mathbb{I}^{c_k,t} \bigcup \left(\bigcup_{c_{k'} \text{ ranked in } R_i:\ c_{k'} \in \mathbb{I}^{c_k,t}} \mathcal{I}^{c_{k'},t} \right).$$

where $\mathbb{I}^{c_k,t}$ is the set of schools that are feasible to i together with c_k with probability greater than t when c_k is the assigned school, and is defined by

$$\mathbb{I}^{c_k,t} = \tilde{\mathcal{I}}^{c_k} \setminus \{c \in \tilde{\mathcal{I}}^{c_k} : \Pr(c \ \& \ c_k \text{ are feasible; all schools ranked above } c_k \text{ in } R_i \text{ are infeasible}) \leq t\}$$

...

Step $|R_i|$ Consider the top-ranked school in R_i, $c_{|R_i|}$. If there is no $\mathbb{F}$ such that $\mathbb{F} \in \mathcal{F}(R_i, p_i)$ and $c_{|R_i|} = A(\mathbb{F}, R_i)$, let $\mathcal{I}^{c_{|R_i|}} = \tilde{\mathcal{I}}^{c_{|R_i|}} = \emptyset$; otherwise,

$$\mathcal{I}^{c_{|R_i|},t} = \mathbb{I}^{c_{|R_i|},t} \bigcup \left(\bigcup_{c_{k'} \text{ ranked in } R_i:\ c_{k'} \in \mathbb{I}^{c_{|R_i|},t}} \mathcal{I}^{c_{k'},t} \right).$$

where $\mathbb{I}^{c_{|R_i|},t}$ is the set of schools that are feasible to i together with c_k with probability greater than t when $c_{|R_i|}$ is the assigned school, and is defined by

$$\mathbb{I}^{c_{|R_i|},t} = \tilde{\mathcal{I}}^{c_{|R_i|}} \setminus \{c \in \tilde{\mathcal{I}}^{c_{|R_i|}} : \Pr(c \ \&\ c_{|R_i|} \text{ are feasible}) \leq t\}$$

The above definition implies that $\mathbb{I}^{c_k,t} = \emptyset$ and hence $\mathcal{I}^{c_k,t} = \emptyset$ whenever

$$\Pr(c_k \text{ is the assigned school}) = \Pr(c_k \text{ is feasible; all schools ranked above } c_k \text{ in } R_i \text{ are infeasible}) \leq t,$$

because $\Pr(c \ \&\ c_k$ are feasible; all schools ranked above c_k in R_i are infeasible) is always smaller than or equal to $\Pr(c_k$ is the assigned school), for any c_k and c ranked in R_i.

Student i's preference relations inferred by TE with threshold t are encoded in

$$\mathcal{I}^t(R_i, p_i) = \left\{\mathcal{I}^{c_1,t}(R_i, p_i), \ldots, \mathcal{I}^{c_{|R_i|},t}(R_i, p_i)\right\},$$

where each non-empty $\mathcal{I}^{c,t}$ indicating that c is more preferred to any school in $\mathcal{I}^{c,t} \setminus \{c\}$. For those c such that $\mathcal{I}^{c,t} = \emptyset$, we do not infer the preference relation that c is more preferable than any other schools. The consequences of a TE threshold are twofold: (i) Given a submitted ROL, R_i, if school c_k in R_i only has a small probability of being the assigned school, $\mathcal{I}^{c_k,t}$ is empty (i.e., TE does not infer that c_k being more preferable than other schools); and (ii) when the probability of c_k being i's assigned school is above the threshold, we only consider its relative preferences over schools that are likely to be feasible together with c_k. Thus, the TE threshold allows students to skip applying to schools that are barely within reach and to flip the order of two schools when one of them is barely within reach.

3.3 Choosing among TE with Threshold and Other Methods

We now study how to select the "optimal" TE threshold in the sense that it extracts the maximum information from the dataset while only using reliable information. As shown below, a higher TE threshold corrects more mistakes but extracts less information. Besides, there are several alternative empirical methods that are common in the literature. To aid the selection among different methods, we provide a statistical test.

Let $\mathcal{O}(\text{TE}_t \mid F^k, \sigma^k_{1\leq i\leq k})$ denote the set of *all* preference relations inferred by TE with threshold t given economy F^k and strategies $\sigma^k_{1\leq i\leq k}$. We call an inferred preference relation *misclassified* if it contradicts the student's true preferences. The following proposition shows the trade-off between the amount of information and the reliability of information and thus provides a theoretical foundation that helps us select TE thresholds.

PROPOSITION 1. *For any F^k, $\sigma^k_{1\leq i\leq k}$, and TE thresholds $t_1 < t_2$,*

$$\mathcal{O}(TE_{t_2} \mid F^k, \sigma^k_{1\leq i\leq k}) \subseteq \mathcal{O}(TE_{t_1} \mid F^k, \sigma^k_{1\leq i\leq k}).$$

$\mathcal{O}(TE_{t_1} \mid F^k, \sigma^k_{1\leq i\leq k})$ has a higher number of misclassified preference relations than $\mathcal{O}(TE_{t_2} \mid F^k, \sigma^k_{1\leq i\leq k})$.

Other Estimation Methods In the literature, the following two methods are common:

(i) **Weak Truth-Telling (WTT)**. Given the strategy-proofness of the DA mechanism, one may consider assuming truth-telling. In the absence of outside options, TT implies that every student ranks all available schools. However, the fact that students rarely rank all available schools motivates a weaker version of truth-telling. Following the literature, WTT, a truncated version of truth-telling, entails two assumptions: (a) the observed number of choices ranked in any ROL is exogenous to student preferences and (b) every student ranks her top preferred schools according to her preferences, although she may not rank all schools.

(ii) **Stability**. Stability implies that every student is assigned her favorite feasible school given the ex-post cutoffs in terms of scores $s_{i,c}$ (i.e., given the realized lottery draws).

We similarly define $\mathcal{O}(\text{WTT} \mid F^k, \sigma^k_{1\leq i\leq k})$ and $\mathcal{O}(\text{Stability} \mid F^k, \sigma^k_{1\leq i\leq k})$ as the preference relations inferred by WTT and stability, respectively.

PROPOSITION 2. *For any F^k, $\sigma^k_{1\leq i\leq k}$, and TE thresholds $0 = t_0 < t_1$,*
(i) $\mathcal{O}(TE_{t_1} \mid F^k, \sigma^k_{1\leq i\leq k}) \subseteq \mathcal{O}(TE_{t_0} \mid F^k, \sigma^k_{1\leq i\leq k}) \subseteq \mathcal{O}(WTT \mid F^k, \sigma^k_{1\leq i\leq k})$, and
(ii) the matching given any realization of F^k and ex-post scores is stable if and only if student preference profiles satisfy $\mathcal{O}(TE_{t_0} \mid F^k, \sigma^k_{1\leq i\leq k})$, implying $\mathcal{O}(Stability \mid F^k, \sigma^k_{1\leq i\leq k}) \subseteq \mathcal{O}(TE_{t_0} \mid F^k, \sigma^k_{1\leq i\leq k})$.

Choosing among the empirical methods. Propositions 1 and 2 provide a theoretical foundation for the choice among the methods. Suppose that we have a parametric model and that β is the parameter vector to be estimated. Let β_0 be the true values. Each method leads to an estimator for β. We let them be the "WTT-based" estimator ($\hat{\beta}^{WTT}$), the "stability-based" estimator ($\hat{\beta}^{stability}$), and the TE_t estimator ($\hat{\beta}^{TE_t}$). The following procedure selects the empirical method based on a series statistical tests.

(i) Set a maintained assumption that the estimator based on TE with threshold $\bar{t}$, $\hat{\beta}^{TE_{\bar{t}}}$, is consistent. One can choose $\bar{t}$ high enough to ensure that the information used in the estimation is reliable.

(ii) Specify a set of $N+1$ thresholds, $t_0, t_1, \ldots, t_N$, in a decreasing order, where $t_0 = \bar{t}$ and $t_N = 0$. A selected estimator is determined by going through the following tests:

Test 1 Formulate two hypotheses regarding $\hat{\beta}^{TE_0}$ and $\hat{\beta}^{TE_{t_1}}$—H_0^1: $\hat{\beta}^{TE_0}$ and $\hat{\beta}^{TE_{t_1}}$ are both consistent; H_1^1: only $\hat{\beta}^{TE_{t_0}}$ is consistent. We conduct a Wald test based on the statistic,

$$\left(\hat{\beta}^{TE_{t_0}} - \hat{\beta}^{TE_{t_1}}\right)' \left(V(\hat{\beta}^{TE_{t_0}} - \hat{\beta}^{TE_{t_1}})\right)^{-1} \left(\hat{\beta}^{TE_{t_0}} - \hat{\beta}^{TE_{t_1}}\right).$$

$V(\hat{\beta}^{TE_{t_0}} - \hat{\beta}^{TE_{t_1}})$ is the covariance matrix of $(\hat{\beta}^{TE_{t_0}} - \hat{\beta}^{TE_{t_1}})$. Under the null, the test statistic follows a $\chi^2_{|\beta_0|}$ distribution. Rejecting H_0^1 leads us to accept H_1^1, which is justified by the nesting structure (Proposition 1) and the maintained assumption (in step i). If H_0^1 is rejected, we exit the procedure and choose $\hat{\beta}^{TE_{t_0}}$ as our selected estimator. Otherwise, we continue to Test 2.

Test n $(1 < n < N)$ Similar to Test 1, we test for H_0^n: $\hat{\beta}^{TE_{t_{(n-1)}}}$ and $\hat{\beta}^{TE_{t_n}}$ are both consistent; H_1^n: only $\hat{\beta}^{TE_{t_{(n-1)}}}$ is consistent.[14] If H_0^n is rejected, we exit the procedure and choose $\hat{\beta}^{TE_{t_{(n-1)}}}$ as

[14]This is equivalent to the following joint test. H_0^n: $\hat{\beta}^{TE_{t_0}}$,$\hat{\beta}^{TE_{t_1}}$, …, $\hat{\beta}^{TE_{t_{(n-1)}}}$, and $\hat{\beta}^{TE_{t_n}}$ are all consistent; H_1^n: only $\hat{\beta}^{TE_{t_0}}$,$\hat{\beta}^{TE_{t_1}}$, …, and $\hat{\beta}^{TE_{t_{(n-1)}}}$ are consistent. The nesting structure in Proposition 1 implies that if $\hat{\beta}^{TE_{t_{(n-1)}}}$ is consistent, for any $n' > (n-1)$, $\hat{\beta}^{TE_{t_{n'}}}$ is also consistent.

our selected estimator. Otherwise, we continue to Test $n+1$.

Test N Similar to Test 1, we test for H_0^N: $\hat{\beta}^{TE_{t_{(N-1)}}}$ and $\hat{\beta}^{TE_{t_N}}$ are both consistent; H_1^N: only $\hat{\beta}^{TE_{t_{(N-1)}}}$ is consistent. If H_0^N is rejected, we exit the procedure and choose $\hat{\beta}^{TE_{t_{(N-1)}}}$ as our selected estimator. Otherwise, we continue to Test WTT.

Test WTT Similar to Test 1, we test for H_0^{WTT}: $\hat{\beta}^{TE_{t_N}}$ and $\hat{\beta}^{WTT}$ are both consistent; H_1^{WTT}: only $\hat{\beta}^{TE_{t_N}}$ is consistent. If H_0^{WTT} is rejected, we choose $\hat{\beta}^{TE_{t_N}}$ as our selected estimator; otherwise, $\hat{\beta}^{WTT}$ is selected.

Note that the above procedure does not consider $\hat{\beta}^{stability}$ as equally important, because it can be considered as a special case of $\hat{\beta}^{TE_0}$ and uses less information.

4 Analysis with Monte Carlo Simulations

This section describes the Monte Carlo simulations that we perform to analyze the implications of our theoretical results. A similar simulation setting with strict intrinsic priorities can be found in Fack, Grenet, and He (2019) and Artemov, Che, and He (2017).

4.1 Model Specification

Consider a finite economy in which $k = 1000$ students apply to $C = 12$ schools for admission. The vector of school capacities is specified as follows:

$$\{S_c\}_{c=1}^{12} = \{110, 50, 100, 100, 50, 100, 100, 50, 100, 100, 50, 100\}$$

The total capacity is equal to the total number of students.

The economy is located in an area within a circle of radius 1 as in Figure 2 which plots one simulation sample. The students represented by small blue dots are uniformly distributed in the circle and the schools represented by big red dots are evenly located on a circle of radius 1/2 around the centroid. Denote the Euclidean distance between student i and school c as $d_{i,c}$.

Students are matched with schools through a student-proposing deferred acceptance algorithm with single tie-breaking rule (DA-STB), similar to NYC high school choice. Students submit an ROL of schools that can include any number of available schools. We assume all schools are acceptable to students; hence, student i submits an ROL including all 12 schools if she truthfully reports her preferences.

School priorities over students are coarse. That is, each school has 4 categories of school-specific (intrinsic) priority groups 0, 1, 2 and 3, with a larger number indicating a higher priority. Denote the priority group that student i belongs at school c as $G_{i,c} \in \{0, 1, 2, 3\}$. Therefore, school c prioritizes student i over i' if $G_{i,c} > G_{i',c}$. A student's priority at a school is drawn independently and uniformly from the four groups. Without loss of generality, we assume that all students are eligible/acceptable at each school. Each student knows her priority group at each school at the time of submitting ROL.

To break ties in priorities, we use a single tie-breaking rule in which all students are assigned a random lottery number drawn from $Unif[0, 1]$. Lottery numbers are not known at the time of submitting ROL. Denote the lottery number of student i as l_i. Then, school c prioritizes student i over i' if $G_{i,c} = G_{i,c}$ but $l_i > l_{i'}$. We define a normalized score of student i at school c by $s_{i,c} = \frac{G_{i,c}+l_i}{4} \in [0, 1]$, in which the

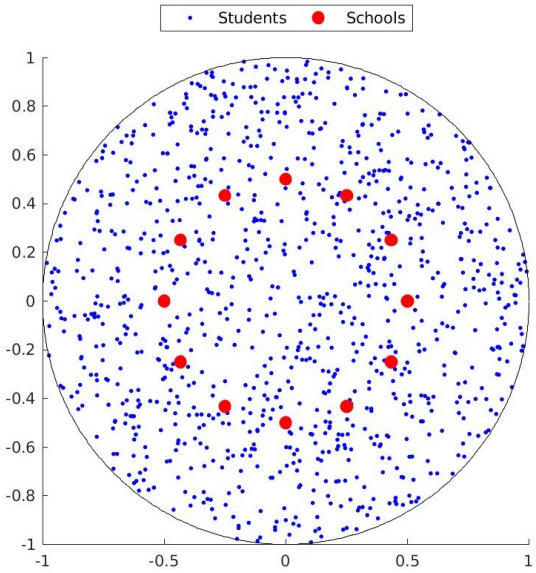


Figure 2: Location of Students and Schools

intrinsic priority lexicographically dominates the lottery draw. In sum, school c prioritizes student i over i' if and only if

$$G_{i,c} > G_{i',c}, \text{ or } G_{i,c} = G_{i',c} \text{ and } l_i > l_{i'};$$

or, equivalently, if and only if $s_{i,c} > s_{i',c}$.

To represent student preferences over schools, we adopt a parsimonious random utility model without an outside option in which the utility functions, are allowed to take any value on the real line. Student i's utility from being matched with school c is specified as follows:

$$u_{i,c} = \beta_1 \times c + \beta_2(T_i \times A_c) + \beta_3 d_{i,c} + \beta_4 Small_c + \epsilon_{i,c}, \quad \forall i, c \tag{1}$$

where $\beta_1 \times c$ is school c's baseline quality, linear in school index; $d_{i,c}$ is the distance from student i's residence to school c; $T_i = 1$ or 0 is student i's type (e.g., disadvantaged or not); $A_c = 1$ or 0 is school c's type (e.g., known for resources for disadvantaged students); $Small_c = 1$ if school c is small, 0 otherwise; and $\epsilon_{i,c}$ is an i.i.d type-I extreme value, implying that the variance of utility shocks is normalized to $\pi^2/6$.

The type of school c, A_c, equals 1 if c is an odd number and otherwise 0. The type of student i T_i is 1 with probability 2/3 among the lowest priority group of school 1 ($G_{i,1} = 0$); and $T_i = 0$ for all students in highest three priority groups ($G_{i,1} \in \{1, 2, 3\}$).

The coefficients of interest are $(\beta_1, \beta_2, \beta_3, \beta_4)$ which are fixed at $(0.3, 2, -1, 0)$ in the simulations. By this specification, schools with larger indices are of higher quality, and $Small_c$ does not play a role in determining student's preference on schools. The purpose of estimation is to recover these coefficients and therefore the distribution of preferences.

4.2 Data Generating Process

Each simulation sample contains an independent preference profile obtained by randomly drawing T_i and $\{G_{i,c}, d_{i,c}, \epsilon_{i,c}\}_c$ for all i from the distributions specified above. In all samples, school capacities and school types (A_c and $Small_c$) are kept constant.

The first set of simulation samples, *cutoff* samples, are used to simulate the joint distribution of the 12 schools' cutoffs (in terms of score, $s_{i,c}$) by letting every student submit an ROL ranking all schools according to her true preferences. To do so, we simulate 250 samples each consisting of 12 schools and 1,000 students. Each sample contains 1,000 sets of independent draws of tie-breaking lotteries l_i. After running the DA algorithm, we calculate the cutoffs in each simulation sample with each draw of lottery. Figure 3 shows the marginal distribution of each school's cutoff from $250 \times 1000 = 250,000$ simulated realizations. Note that schools with smaller capacities tend to have higher cutoffs. For example, school 11 with 50 seats often has the highest cutoff, although school 12 with 100 seats has the highest baseline quality.

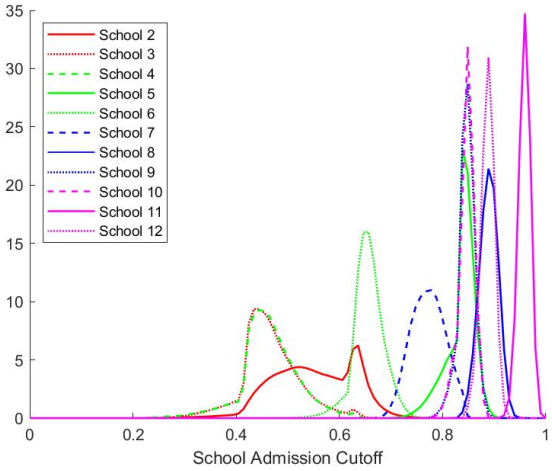


Figure 3: Simulated Cutoff Distribution

To generate data on student behavior and admission outcomes for preference estimation, we simulate another 250 samples, the *estimation* samples, with new independent draws of T_i and $\{G_{i,c}, d_{i,c}, \epsilon_{i,c}\}_c$. For each of the 250 estimation samples, 1,000 independent lotteries are drawn to calculate the admission probabilities which are to be used in the transitive extension algorithm. These samples are used for the estimation, and in each of them, we consider three types of data generating processes (DGPs) with different student behaviors.

(i) **TT (Truth-Telling)** : Every student submits an ROL ranking all 12 schools according to her true preferences.

(ii) **IRR (Payoff Irrelevant Mistakes)**: A fraction of students skip schools with which they are never matched according to the simulated distribution of cutoffs. For a given student, a skipped school can have a high (expected) cutoff and thus be "out of reach." Alternatively, the school may also have a low cutoff, but the student is always accepted by one of her more-preferred schools. To specify the fraction of skippers, we first randomly choose about 25.7% of the students to be never-skippers who always rank all schools truthfully. All other students are potential skippers. Among them, we consider two scenarios. In IRR1, approximately half of the potential skippers skip all the "never-matched" schools; IRR2 makes all of them skip. Students with $T_i = 1$ are more likely to skip than those with $T_i = 0$, as their scores tend to be lower: 95.4 percent of $T_i = 1$ are potential skippers,

compared to 70.1 percent of $T_i = 0$.

(iii) **REL (Payoff Relevant Mistakes)**: In addition to IRR2, i.e., given all the potential skippers have skipped the never-matched schools, we now let them make payoff-relevant mistakes. That is, students skip some of the schools with which they have a small chance of being matched according to the simulated distribution of cutoffs. Recall that the joint distribution of cutoffs is only simulated once under the assumption that everyone is truth-telling. In each of the two DGPs, REL1 and REL2, we specify a threshold and make the skippers omit the schools at which they have an admission probability lower than the threshold. For REL1 and REL2, the thresholds are 10 and 20 percent respectively.

In summary, for each of the 250 estimation samples, we simulate the matching games 5 times: 1 (TT) + 2 (IRR) + 2 (REL). Table 1 summarizes the scenarios under each data generating process (DGP) we consider. Each entry reported is a percentage that is averaged over the 250 estimation samples. The fraction of skippers increases from 0% in TT to 74.3% in IRR2. As a result, weak truth-telling is exactly satisfied only in TT, and the fraction of students reporting preferences consistent with WTT ranges from 29.6% to 66.0% among DGPs IRR1, IRR2, REL1 and REL2.

Table 1: Skips and Mistakes in Monte Carlo Simulations (%)

	Scenarios: Data Generating Processes w/ Different Student Strategies				
	Truth-Telling	Payoff Irrelevant Mistakes		Payoff Relevant Mistakes	
	TT	IRR1	IRR2	REL1	REL2
WTT: *Weak-Truth-Telling*	100	66.0	29.6	39.0	41.9
Matched w/ favorite feasible school	100	100	100	98.6	97.7
Skippers	0	35.9	74.3	74.3	74.3
By number of skips:					
Skipping 11 schools	0	3.2	6.6	39.1	48.1
Skipping 10 schools	0	5.3	11.0	29.0	23.7
Skipping 9 schools	0	6.7	13.9	5.7	2.5
Skipping 8 schools	0	7.0	14.5	0.5	0
Skipping 7 schools	0	6.2	12.9	0	0
Skipping 6 schools	0	4.3	9.0	0	0
Skipping 5 schools	0	2.2	4.5	0	0
Skipping 4 schools	0	0.7	1.5	0	0
Skipping 3 schools	0	0.2	0.3	0	0
Skipping 2 schools	0	0	0	0	0
Skipping 1 schools	0	0	0	0	0
TT: Truth-Telling	100	64.1	25.7	25.7	25.7

4.3 Estimation and Results

The random utility model described by equation (1) is estimated under three methods, WTT, stability, and TE with the following thresholds, $\{0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30\}$. Table 2 presents the estimates (mean and standard deviation across the 250 samples).

We evaluate the performance of the three estimators along three dimensions. The first is the bias-variance tradeoff, focusing on β_2 which measures how students of type $T_i = 1$ value schools of type $A_c = 1$.

The second dimension is the performance of the estimators based on TE relative to the stability-based estimator, which are both robust to (some) strategic mistakes. Lastly, we illustrate how we select the "optimal" estimator based on the statistical tests in Section 3.3.

Table 2: Estimation with Different Identifying Assumptions: Monte Carlo Results

DGPs	Estimation Method	Quality ($\beta_1 = 0.3$)			Interaction ($\beta_2 = 2$)			Distance ($\beta_3 = -1$)			Small school ($\beta_4 = 0$)		
		mean	s.d.	$\sqrt{MSE}$	mean	s.d.	$\sqrt{MSE}$	mean	s.d.	$\sqrt{MSE}$	mean	s.d.	$\sqrt{MSE}$
						A. DGP: Truth-Telling—TT							
TT	WTT	0.30	0.00	0.00	2.00	0.06	0.06	-1.00	0.04	0.04	0.00	0.02	0.02
	TE 0%	0.30	0.01	0.01	2.00	0.14	0.14	-1.00	0.08	0.08	-0.01	0.06	0.06
	TE 15%	0.30	0.01	0.01	2.01	0.19	0.19	-0.99	0.10	0.10	0.00	0.07	0.07
	TE 30%	0.30	0.01	0.01	2.01	0.20	0.20	-0.99	0.11	0.11	0.00	0.08	0.08
	Stability	0.30	0.01	0.01	2.02	0.22	0.22	-0.99	0.12	0.12	-0.01	0.10	0.10
	Selected	**0.30**	**0.01**	**0.01**	**2.00**	**0.11**	**0.11**	**-1.00**	**0.07**	**0.07**	**0.00**	**0.04**	**0.04**
						B. DGP: Payoff-Irrelevant Skips—IRR1, IRR2							
IRR1	WTT	0.21	0.00	0.09	0.82	0.08	1.19	-0.58	0.06	0.43	-0.01	0.02	0.02
	TE 0%	0.30	0.01	0.01	2.00	0.14	0.14	-1.00	0.08	0.08	0.00	0.06	0.06
	TE 15%	0.30	0.01	0.01	2.01	0.18	0.18	-0.99	0.10	0.10	0.00	0.07	0.07
	TE 30%	0.30	0.01	0.01	2.02	0.20	0.20	-0.99	0.11	0.11	0.00	0.08	0.08
	Stability	0.30	0.01	0.01	2.01	0.23	0.23	-0.99	0.13	0.13	-0.01	0.10	0.10
	Selected	**0.30**	**0.01**	**0.01**	**2.01**	**0.16**	**0.16**	**-1.00**	**0.09**	**0.09**	**0.00**	**0.06**	**0.06**
IRR2	WTT	0.12	0.01	0.18	0.05	0.04	1.95	-0.25	0.07	0.76	-0.01	0.01	0.02
	TE 0%	0.30	0.01	0.01	2.00	0.14	0.14	-1.00	0.08	0.08	-0.01	0.06	0.06
	TE 15%	0.30	0.01	0.01	2.01	0.18	0.18	-0.99	0.10	0.10	0.00	0.07	0.07
	TE 30%	0.30	0.01	0.01	2.01	0.20	0.20	-0.99	0.11	0.11	0.00	0.08	0.08
	Stability	0.30	0.01	0.01	2.02	0.22	0.22	-1.00	0.13	0.13	0.01	0.10	0.10
	Selected	**0.30**	**0.01**	**0.01**	**2.00**	**0.16**	**0.16**	**-0.99**	**0.09**	**0.09**	**0.00**	**0.07**	**0.07**
						C. DGP: Payoff-Relevant Mistakes—REL1, REL2							
REL1	WTT	-0.11	0.01	0.41	-0.08	0.03	2.08	-0.16	0.06	0.84	-0.07	0.01	0.07
	TE 0%	0.28	0.02	0.03	1.85	0.24	0.29	-0.90	0.09	0.14	-0.06	0.16	0.17
	TE 15%	0.30	0.02	0.02	1.94	0.24	0.25	-0.96	0.10	0.11	-0.02	0.16	0.16
	TE 30%	0.30	0.02	0.02	1.97	0.24	0.24	-0.97	0.11	0.11	-0.01	0.15	0.15
	Stability	0.30	0.02	0.02	1.96	0.26	0.26	-0.97	0.12	0.13	-0.03	0.18	0.18
	Selected	**0.30**	**0.02**	**0.02**	**1.95**	**0.24**	**0.25**	**-0.96**	**0.10**	**0.11**	**-0.02**	**0.16**	**0.16**
REL2	WTT	-0.13	0.01	0.43	-0.09	0.02	2.09	-0.16	0.06	0.84	-0.13	0.01	0.13
	TE 0%	0.27	0.02	0.04	1.82	0.24	0.30	-0.88	0.09	0.15	-0.05	0.16	0.16
	TE 15%	0.29	0.02	0.03	1.92	0.26	0.27	-0.94	0.11	0.12	-0.02	0.18	0.18
	TE 30%	0.30	0.02	0.02	1.94	0.27	0.27	-0.95	0.11	0.12	-0.02	0.18	0.18
	Stability	0.29	0.02	0.02	1.94	0.28	0.29	-0.96	0.13	0.13	-0.02	0.20	0.20
	Selected	**0.30**	**0.02**	**0.02**	**1.93**	**0.26**	**0.27**	**-0.95**	**0.11**	**0.12**	**-0.03**	**0.18**	**0.18**

Notes: The results are from the 250 Monte Carlo samples.

Bias-Variance Tradeoff Figure 4 plots the distributions of the estimates of β_2 given each DGP; the true value of β_2 is 2.

There are a few notable patterns. First, when the DGP is TT, all estimates are consistent, while the WTT-based estimator having the smallest variance as shown in Panel (a). This is expected, since no students make strategic mistakes violating the WTT or stability assumptions. Furthermore, the WTT-

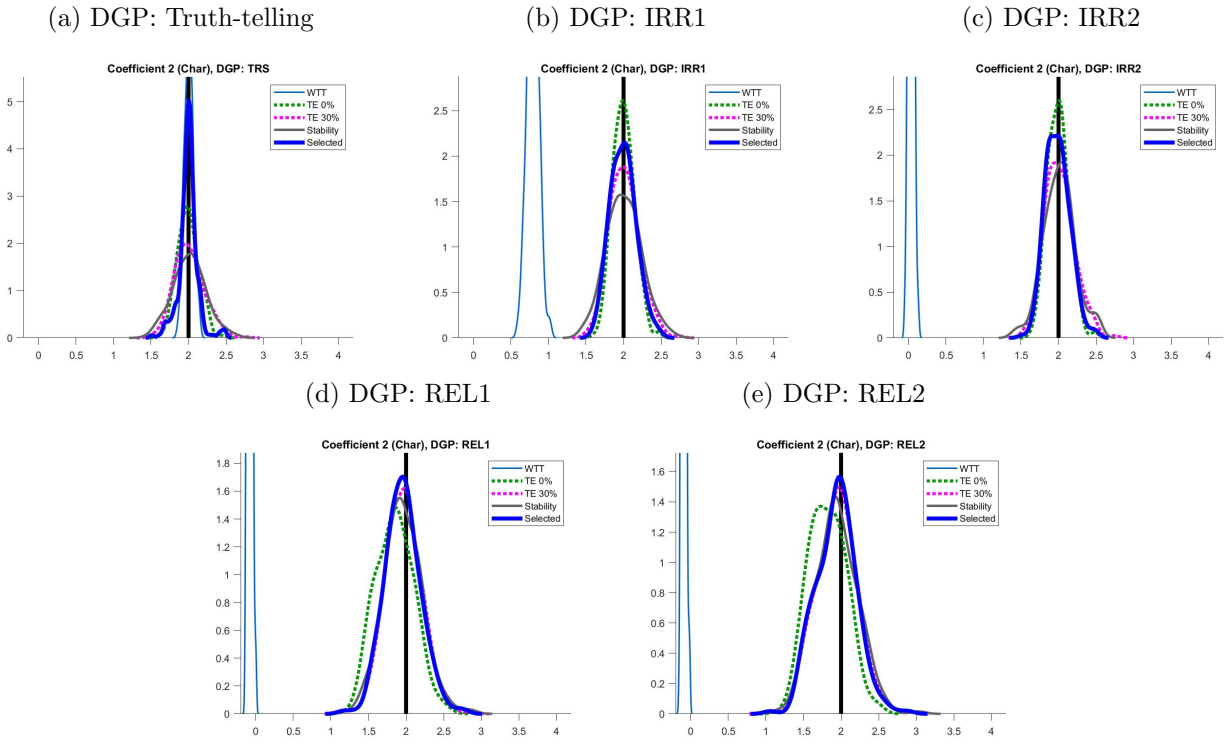


Figure 4: Distribution of Estimates based on WTT, Stability, or Transitive Extension ($\beta_2 = 2$)

based estimator having the smallest variance is not only true under DGP TT but also under other DGPs. Intuitively, this is due to the fact that WTT uses the maximal information that one can infer from observed ROLs.

Second, Panels (b) and (c) show the results from DGPs in which some students make payoff-irrelevant mistakes (IRR1 and IRR2). The WTT-based estimator is susceptible to strategic mistakes. For example, the WTT-based estimates have a mean 0.82 (standard deviation 0.08) in IRR1, and a mean 0.05 (standard deviation 0.04) in IRR2. That is, the WTT-based estimator is no longer consistent and the size of bias grows, as the fraction of mistakes grows from IRR1 to IRR2. On the other hand, the estimators based on TE or stability are robust to payoff-irrelevant mistakes. On average, the estimates from the 250 samples have mean 2.01 (standard deviation 0.19).

Finally, Panels (d) and (e) show the results from DGPs in which some students make payoff-relevant mistakes (REL1 and REL2). Recall that when students make payoff relevant mistakes, neither WTT nor stability is satisfied and thus both the WTT-based and stability-based estimators are inconsistent. However, the bias of the stability-based estimator remains smaller than that of the WTT-based estimator, since the violation of WTT (ranging from 58.1% to 61.0%) is much more severe than the violation of stability (ranging from 1.4% to 2.3%) (see Table 1). As predicted, the estimators based on TE remain consistent, and the magnitude of bias remains very small especially for those which use larger thresholds.

The bias-variance tradeoffs are summarized by the square-root of the mean squared errors in Table 2. In DGP TT, where all estimators are consistent, the WTT-based estimator attains the minimum mean squared error, as it has the minimum variance where biases of all estimators are close to zero. However, in DGPs with payoff-irrelevant mistakes, the mean squared error of the WTT-based estimator is larger than

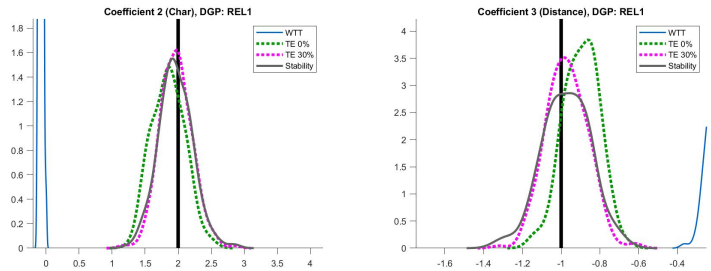
that of any other estimator. This is due to the fact that even though the WTT-based estimator has the smallest variance, it is significantly inconsistent. The mean squared errors of the WTT-based estimator are even larger with DGPs with payoff-relevant mistakes.

Performance of the TE-based estimators. We now compare the relative performance of the estimates based on TE to the stability-based estimator. To the extent that stability is satisfied, both the TE-based estimator and the stability estimator are consistent. However, the procedures by which the two estimators are constructed imply that the estimators based on TE use more information contained in the observed ROLs, and should have a higher precision.

Figure 4 shows the results. The TE-based estimators are in general more precise than the stability estimator, having a higher level of concentration around the true value $\beta_2 = 2$. As reported in Table 2, the TE-based estimators with thresholds, 0, 15, and 30% all have lower standard deviations compared to the stability-based estimator. For example, in IRR1 in which approximately 35.9% students make payoff-irrelevant mistakes, the standard deviation of the stability-based estimator is 0.23, while the TE-based estimators have a standard deviation ranging from 0.14 to 0.20.

We also show how the TE-based estimator perform when the TE threshold increases. First, in DGPs IRR1 and IRR2 where students make payoff-irrelevant mistakes, all TE-based estimators remain consistent. This is because the payoff-irrelevant mistakes by definition are behaviors of students skipping schools with zero admission probability. TE does not infer preferences regarding these schools, regardless of the value of the TE threshold.

(a) REL1 (Fewer Payoff-relevant Mistakes)



(b) REL2 (More Payoff-relevant Mistakes)

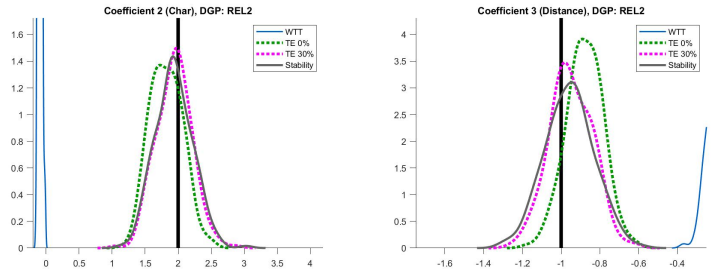


Figure 5: Estimates under REL1 and REL2 ($\beta_1 = 0.3, \beta_3 = -1$)

However, in DGPs REL1 and REL2 where students skip some schools with a positive admission probability and hence make payoff-relevant mistakes, the choice of TE threshold becomes important. Figure 5 shows the estimation results for coefficients $\beta_2 = 2$ and $\beta_3 = -1$. In DGP REL1, where students

drop schools with admission probability less than 10%, biases of the TE-based estimator with threshold 0 are larger than those using higher thresholds. For example, for $\beta_2 = 2$, the bias of TE_0 estimator is 0.15 where those of $TE_{0.15}$ and $TE_{0.30}$ are 0.06 and 0.03, respectively (Table 2). In DGP REL2 where students drop schools with admission probability less than 20%, the bias of TE_0 becomes larger. Furthermore, the bias of the $TE_{0.15}$ estimator also becomes substantial. In both cases, the TE-based estimator with threshold equal to 30% remain unbiased.

Also note that in both REL DGPs, the bias of a TE-based estimator monotonically decreases in the threshold. At the same time, the standard deviation of the estimators monotonically increases. Hence, a natural question that one might raise is how to choose the appropriate method. To answer this question, we apply the tests from Section 3.3.

Choosing among the estimation methods. Recall the main motivation which led us to introduce the TE-based estimator is to "correct" the strategic mistakes that students might make in a school choice environment in order to obtain robust estimates. Without the information on how students make mistakes, we follow the procedure described in Section 3.3 to determine the selected estimator that corrects (almost) all the mistakes and uses the maximum information from the data.

In our Monte Carlo simulation, we consider the following thresholds, $\{0, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30\}$. We impose the maintained assumption that $\hat{\beta}^{TE_{0.3}}$ is consistent. For a given method, whose corresponding estimator is $\hat{\beta}^{method}$, we test the null hypothesis that $\hat{\beta}^{method}$ and $\hat{\beta}^{TE_{0.3}}$ are both consistent, against the alternative that only $\hat{\beta}^{TE_{0.3}}$ is consistent. Across the 250 estimation samples, the fraction of test rejecting the null are reported in Panel A of Table 3. Panel B reports the fraction of a method being selected by the procedure in Section 3.3.

A few patterns are notable. First, when the DGP is TT and thus the null is true for every method, the null is rejected around 5% as expected. As a result, the WTT estimator is most frequently chosen (in 72% of the 250 estimation samples). Second, when students are allowed to make strategic mistakes (in IRR1, IRR2, REL1 and REL2), the WTT-based estimator is always rejected and hence is never selected. Third, when students make payoff-relevant mistakes as in REL1 and REL2, the TE-based estimators with a low threshold are rejected most of the time. For example, in REL1 where students "skip" schools with admission probability smaller than 10%, the TE_0 estimators are rejected in 86% of the 250 estimation samples. As a result, our procedure often selects the TE-based estimators with a sufficiently high threshold.

The performance of the selected estimator is shown in Table 2 and Figure 4. The summary statistics of the selected estimators across the 250 estimation samples are presented in bold face in Table 2. They almost always remain consistent, and the precision is reasonably high among the set of estimators we consider. The square root of the mean-squared error is informative regarding the performance of the selected estimators. In most cases, they attain the lowest $\sqrt{MSE}$ among the set of estimators considered. The qualitatively same results can be seen in Figure 4 where the distribution of the selected estimator is depicted in thick blue line. Across all DGPs, the distributions are centered at the true value of the coefficient, and highly concentrated around the true value.

Table 3: Determining the Selected Estimator: Test Results (at the 5% significance level)

Data Generating Process:		TT	IRR1	IRR2	REL1	REL2
		A. Fraction of Rejecting the Null				
Estimation method	WTT	0.03	1.00	1.00	1.00	1.00
	TE_0	0.04	0.04	0.05	0.86	0.96
	$TE_{0.05}$	0.07	0.06	0.06	0.52	0.83
	$TE_{0.10}$	0.08	0.07	0.09	0.38	0.72
	$TE_{0.15}$	0.07	0.08	0.09	0.30	0.58
	$TE_{0.20}$	0.08	0.08	0.10	0.28	0.49
	$TE_{0.25}$	0.08	0.07	0.10	0.23	0.34
		B. Fraction of Being Selected				
Estimation method	WTT	0.72	0.00	0.00	0.00	0.00
	TE_0	0.02	0.77	0.74	0.08	0.02
	$TE_{0.05}$	0.02	0.02	0.03	0.23	0.08
	$TE_{0.10}$	0.03	0.02	0.02	0.13	0.10
	$TE_{0.15}$	0.04	0.02	0.03	0.09	0.13
	$TE_{0.20}$	0.04	0.04	0.04	0.09	0.14
	$TE_{0.25}$	0.06	0.06	0.05	0.14	0.20
	$TE_{0.30}$	0.08	0.07	0.10	0.23	0.34

Notes: The results are from the 250 estimation samples. For a given method, whose corresponding estimator is $\hat{\beta}^{method}$, we test the null hypothesis that $\hat{\beta}^{method}$ and $\hat{\beta}^{TE_{0.3}}$ are both consistent, against the alternative that only $\hat{\beta}^{TE_{0.3}}$ is consistent. Panel A reports the fraction that the null is rejected. Panel B reports the fraction of a method being selected by the procedure in Section 3.3.

5 Student Preferences in NYC High School Choice (in progress)

The previous sections develop a theoretical framework that allows for students' "mistakes"—deviating from a weakly dominant strategy—and propose a novel estimation method, Transitive Extension (TE), built on the theoretical results. Monte Carlo simulations illustrate that the TE-based estimator is superior to the existing approaches, such as those based on Weak Truth-Telling (WTT) or stability. We now turn to applying the method to a real-world school choice setting, the high school choice in NYC.

We plan to apply the three alternative estimation methods—based on WTT, stability, or TE—and compare and contrast their performances. Next, we will consider a counterfactual scenario and show how different the forecasts are when different estimates are used. We are in the process of acquiring data from the NYC DOE, although our data request has already been approved. Below, we focus on the institutional background and the data we will receive.

5.1 NYC High School Choice

NYC public high school admission system uses the celebrated Student Proposing Deferred Acceptance (SPDA) algorithm built on Gale and Shapley (1962) (Abdulkadiroglu, Pathak, and Roth, 2005, 2009). SPDA is strategy-proof and produces stable matchings.

The procedure collects students' ROLs and has the following steps:

- **Step 1**

  Each student proposes to her first choice. Each program tentatively assigns seats to its proposers

one at a time, following their priority order. The student is rejected if no seats are left at the time of consideration.

- **Step $k \geq 2$**
  Each student who was rejected in the previous step proposes to her next choice in her ROL. Each program considers the students it has tentatively assigned together with its new proposers and tentatively assigns its seats to these students one at a time following the program's priority order. The student is rejected if no seats are left when she is considered.

- The procedure terminates either when there are no new proposals or when all rejected students have exhausted their ROLs.

NYC public high school system consists of two sectors: specialized high schools and regular high schools. There are 9 specialized high schools in NYC: Stuyvesant High School, Brooklyn Technical High School, Bronx High School of Science, High School of American Studies at Lehman College, The Brooklyn Latin School, High School for Mathematics, Science and Engineering at City College, Queens High School for the Sciences at York College, Staten Island Technical High School, and Fiorello H. LaGuardia High School of Music & Art and Performing Arts. Specialized high schools except for Fiorello H. LaGuardia High School use Specialized High School Admission Test (SHSAT) as the sole criterion of admission, which is a required exam for students willing to attend any of the specialized high school. Fiorello H. LaGuardia High School uses audition as the admission criterion. Since specialized high schools use different admission methods from the regular high schools, we do not consider 9 specialized high schools in our analysis.

Next, regular high schools are traditional public schools with mainly six types of programs in terms of admission method: *Unscreened*, *Limited unscreened*, *Screened*, *Audition*, *Educational option* and *Zoned* programs. Multiple programs of different types may be offered by a single school. *Unscreened* programs admit students by a random lottery number attached to each student. *Limited unscreened* programs operate similarly as *Unscreened* programs but give higher priority to students who attended an information session or open houses. *Screened* programs as well as *Audition* programs rank students by individual assortment of criteria. For example, *Screened* programs use several criteria such as final report card grade, statewide standardized test scores and attendance and punctuality. *Audition* programs hold school/program specific auditions to admit students. *Educational option* is a mixture of unscreened and screened programs. They have a purpose of serving students at diverse academic performance levels and divide students into high (highest 16%), middle (68%) and low English Language Arts (ELA) (lowest 16%) levels. 50% of the seats in each group are filled using school-specific criteria similarly as a *Screened* program and the other 50% are filled randomly similarly as an *Unscreened* program. *Zoned* programs give priority or guarantee admission to students who apply and live in the zoned area of the school. (*NYC DOE Introduction to High School Admissions*)

Each program has its own eligibility and priority group criteria. For example, in academic year 2016-17, Young Women's Leadership School in Astoria opened its seats only to female students i.e., being a female student was the only eligibility criterion. Besides, they gave the highest admission priority to continuing eighth graders, and then to Queens students or residents, and then to New York City residents.

The number of priority groups given to students are a lot smaller than the number of applicants to each program. Hence students who apply to programs that do not actively rank students —*Unscreened*,

Limited Unscreened, *Zoned* and the unscreened part of *Educational option*—are often in the same priority group. Hence the ties need to be broken for the SPDA algorithm to run. For this purpose, a random lottery number is drawn and attached to each student which is used to break ties at all programs that require tie-breaking in the same fashion—single tie-breaking (STB) rule. The lottery number is unknown to the student at the moment of application.

The timeline of the admission process is as follows (Corcoran and Levin (2011), Directory of NYC Public High Schools). In October and November, students may apply to specialized high schools for which they should take SHSAT or audition at LaGuardia High School. In December, they are required to submit up to 12 ranked non-specialized high school programs regardless of their application status to specialized high schools. In March, the SPDA algorithms for specialized high schools (SHSAT takers) and non-specialized high schools (all students) are separately run for Round 1. The Department of Education sends each student a letter with their offer from regular schools and offer from specialized high schools, if any. If a student receives offers from both regular and specialized high schools, he/she must choose one. All students who accept either Round 1 offer has a finalized admission decision. If a student didn't apply in Round 1, didn't receive an offer in Round 1 or prefers a program with Round 2 availability over their Round 1 offer, he/she should proceed to Round 2. Round 2 takes place in March and operates with students who applied and programs that have seat availability. Round 2 offers automatically replace Round 1 offers, if any. Students who are not assigned in Rounds 1 and 2 or rejected the assignment go to the administrative round in which students are administratively assigned a school on a case-by-case basis.

5.2 Data and Sample Restriction

The main data that we will use is the administrative data from the New York City Department of Education (NYC DOE) for academic year 2016-2017.[15] There are four sets of data used to construct information on the applicants. First, high school application (HSAP) data includes the rank order lists of students and student information such as English Language Arts (ELA) and math standardized test scores and English-language Learner (ELL) status and information on each round of the process (programs that are ranked, priority rank, priority and eligibility at each program etc.) related to regular high school application. Second, June biographic data includes more comprehensive biographic data of the students including ethnicity, gender, disability status as well as information on attendance and punctuality. Third, standardized test data includes more detailed information on statewide standardized exams. Fourth, zoned DBN data provides information on the zoned school of each student, census tract and school district each student resides in. There exists a unique scrambled student ID variable that enables merging all NYC DOE dataset while personally identifying the student is impossible. Lastly, we use information on each census tract and zip code in NYC fetched from 2016 American Community Survey 5-Year Estimates from US Census database.

School information is constructed using the NYC High School Directory that is published every year before the application process starts. This includes each program's previous year capacity, the number of students who applied in the previous year, eligibility and priority criteria, accountability data such

[15] See Abdulkadiroglu, Agarwal, and Pathak (2017) and Abdulkadiroglu, Pathak, Schellenberg, and Walters (2017) for empirical analyses of the NYC high school choice.

as progress report, graduation rate and college enrollment rate, types of language classes provided etc. Other variables about current 9th graders, such as ethnicity composition and fraction of high performing students, are constructed using the high school application data from the previous year, which is academic year 2015-2016 in our application.

We focus on Round 1 application data of students and schools/programs in Staten Island, which is one of the five boroughs of the New York City: Manhattan, Queens, Brooklyn, Bronx and Staten Island. Staten Island can be considered as a separate economy itself. One can travel to other boroughs of the New York City only via the Staten Island Ferry, or Verrazzano-Narrow Bridge which is the only ground transportation route to other parts of the city. Almost all except for a negligible fraction of middle school aged students who reside in Staten Island are enrolled in middle schools in Staten Island, and applied to high school programs only in Staten Island. We focus on Round 1 application data since almost all students participate in the first round, and a majority of students' school assignments are finalized in Round 1.

References

Abdulkadiroglu, A., N. Agarwal, and P. A. Pathak (2017): "The Welfare Effects of Coordinated Assignment: Evidence from the NYC HS Match," *American Economic Review*, 107(12), 3635–89.

Abdulkadiroglu, A., Y.-K. Che, and Y. Yasuda (2015): "Expanding 'Choice' in School Choice," *American Economic Journal: Microeconomics*, 7, 1–42.

Abdulkadiroglu, A., P. A. Pathak, and A. E. Roth (2005): "The New York City High School Match," *American Economic Review Papers and Proceedings*, 95, 364–367.

——— (2009): "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match," *American Economic Review*, 99(5), 1954–1978.

Abdulkadiroglu, A., P. A. Pathak, J. Schellenberg, and C. R. Walters (2017): "Do parents value school effectiveness?," Discussion paper, National Bureau of Economic Research.

Artemov, G., Y.-K. Che, and Y. He (2017): "Strategic 'Mistakes': Implications for Market Design Research," *WP*.

Azevedo, E. M., and J. D. Leshno (2016): "A supply and demand framework for two-sided matching markets," *Journal of Political Economy*, 124, 1235–1268.

Che, Y.-K., and F. Kojima (2010): "Asymptotic Equivalence of Probabilistic Serial and Random Priority Mechanisms," *Econometrica*, 78(5), 1625–1672.

Chen, Y., and O. Kesten (2014): "From Boston to Chinese Parallel to Deferred Acceptance: Theory and Experiments on a Family of School Choice Mechanisms," Unpublished mimeo.

Chen, Y., and T. Sönmez (2002): "Improving Efficiency of On-campus Housing: An Experimental Study," *American Economic Review*, 92, 1669–1686.

Corcoran, S. P., and H. M. Levin (2011): "School Choice and Competition in the New York City Schools," in *Education Reform in New York City: Ambitious Change in the Nation's Most Complex School System*, ed. by J. O'Day, C. Bitter, and L. M. Gomez. Harvard Education Press.

Dubins, L. E., and D. A. Freedman (1981): "Machiavelli and the Gale-Shapley algorithm," *American Mathematical Monthly*, 88, 485–494.

Fack, G., J. Grenet, and Y. He (2019): "Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions," *American Economics Review*, 109(4), 1486–1529.

Gale, D., and L. S. Shapley (1962): "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 9–15.

Hassidim, A., A. Romm, and R. Shorrer (2016): ""Strategic" Behavior in a Strategy-proof Environment," mimeo.

He, Y. (2017): "Gaming the Boston School Choice Mechanism in Beijing," Toulouse School of Economics and Rice University, Unpublished mimeo.

Li, S. (2017): "Obviously strategy-proof mechanisms," *American Economic Review*, 107(11), 3257–87.

Rees-Jones, A. (2017): "Suboptimal behavior in strategy-proof mechanisms: Evidence from the residency match," *Games and Economic Behavior*.

Rees-Jones, A., and S. Skowronek (2018): "An experimental investigation of preference misrepresentation in the residency match," *Proceedings of the National Academy of Sciences*, 115(45), 11471–11476.

Roth, A. E. (1982): "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 7, 617–628.

Appendix to

Leveraging Uncertainties to Infer Preferences: Robust Analysis of School Choice with Lotteries

Yeon-Koo Che Dong Woo Hahm Yinghua He

November 2019

List of Appendices

Appendix A: Proofs from Section 2 **30**

Appendix B: Proofs from Section 3 **38**

A Proofs from Section 2

Proof of Theorem 1. For any sequence $\{F^k\}$ induced by E, fix any arbitrary regular robust equilibrium $\{(\sigma^k_{1\leq i\leq k})\}_k$. The strategies induce a random ROL, R_i, for each player i.

These ROLs induce a (random) function $M^k := (M_a^{C'k})_{a\in C', C'\subset C}$ that maps $p \in [0,1]$ to

$$M_a^{C'k}(p) := \frac{1}{k}\sum_{i=1}^{k} I\left\{a \succ_{R_i^k} b, \forall b \in C' \setminus \{a\} \text{ and } s_{ia} \geq p\right\},$$

which denotes the empirical (normalized) measure of students who rank a most highly among C' and have scores for school a exceeding p in economy F^k. (Note that we suppress the dependence on the random ROLs R_i for notational ease.) Let $\bar{M}^k := (\bar{M}_a^{C'k})_{a\in C', C'\subset C}$ be the vector, defined by $\bar{M}_a^{C'k}(p) := \mathbb{E}_{(R_1,\ldots,R_k)}\left[M_a^{C'k}(p)\right]$, where the expectation is taken over the random draws of the types of the students and the random reported ROLs according to mixed strategy profile $\left(\sigma_i^k\right)_{1\leq i\leq k}$.

As a preliminary step, we establish a series of claims, culminating in Claim 4, which establishes that along a subsequence, the cutoffs of the finite economies converge in probability to those in the limit economy. The statement proof follows from the last claim, Claim 6.

Claim 1. *Fix any $p \in [0,1]$. Then, there exists $B > 0$, such that for any $\alpha > 0$,*

$$\Pr\left[\left\|M^k(p) - \bar{M}^k(p)\right\| > \sqrt{|B|}\alpha\right] \leq |B| \cdot e^{-2k\alpha^2}.$$

Proof. First by McDiarmid's theorem, for each $C' \subset C, a \in C'$,

$$\Pr\{|M_a^{C'k}(p) - \bar{M}_a^{C'k}(p)\,| > \alpha\} \leq e^{-2k\alpha^2},$$

since for each $a \in C$, $|M_a^{C'k}(p)(R_1,...,R_k) - M_a^{C'k}(p)(R_1',...,R_k')| \leq 1/k$ whenever $(R_1,...,R_k)$ and $(R_1',...,R_k')$ differ only in one component. Let $B := 2^{|C|} \times |C|$.

It then follows that

$$\begin{aligned}
&\Pr\left[\left\|M^k(p) - \bar{M}^k(p)\right\| > \sqrt{|B|}\alpha\right] \\
&\leq \Pr\left[\exists (a, C')a \in C', C' \subset C \text{ s.t. } \left|M_a^{C'k}(p) - \bar{M}_a^{C'k}(p)\right| > \alpha\right] \\
&\leq |B| \cdot e^{-2k\alpha^2}.
\end{aligned}$$

□

Claim 2. *The sequence of functions $\left\{\bar{M}^k(p)\right\}_{k=1}^{\infty}$ has a subsequence that converges to some continuous function $\bar{M}(p)$ uniformly in p. $\bar{M}(p)$ is strictly decreasing in p.*

Proof. For any possible ROL r and intrinsic score vector $\mathbf{q} = (q_c)_{c\in C}$,

$$m_{r,\mathbf{q}}^k = \frac{1}{k}\left[\sum_{i=1}^{k} \Pr\{R_i = r, \tilde{\mathbf{q}} = \mathbf{q}\}\right]$$

be the expected fraction of students who submit a ROL r and have intrinsic priority scores $\mathbf{q}$ in random k economy. Recall $(r, \mathbf{q})$ has a finite support, so the vector $\mathbf{m}^k := (m^k_{r,\mathbf{q}})_{r,\mathbf{q}}$, which lies within a compact set, must have a convergent subsequence. Let $\mathbf{m}^\infty = (m^\infty_{r,\mathbf{q}})_{r,\mathbf{q}}$ be the limit of that subsequence.

Let $R_a^{C'}$ be the set of ROLs that rank a most highly among $C' \subset C$, and let $q_c^-(p) = \inf\{q_c | q_c \leq p \leq q_c + \frac{1}{n_c}\}$ be the marginal priority group for school c if the cutoff for that school were p. Then, we can write each

$$\bar{M}_a^{C'k}(p) = \sum_{r \in R_a^{C'}, \mathbf{q}: q_a \geq p} m^k_{r,\mathbf{q}} + \sum_{r \in R_a^{C'}, \mathbf{q}: q_a = q_a^-(p)} m^k_{r,\mathbf{q}} \left(q_a^-(p) + \frac{1}{n_a} - p \right) n_a$$

since the uniform lottery score means that the fraction of the marginal priority group $q_a^-(p)$ receiving score higher p is $(p - q_a^-(p))n_a$. Clearly, along the converging subsequence, this quantity converges to

$$\bar{M}_a^{C'}(p) := \sum_{r \in R_a^{C'}, \mathbf{q}: q_a \geq p} m^\infty_{r,\mathbf{q}} + \sum_{r \in R_a^{C'}, \mathbf{q}: q_a = q_a^-(p)} m^\infty_{r,\mathbf{q}} \left(q_a^-(p) + \frac{1}{n_a} - p \right) n_a,$$

and the convergence is uniform in p.

The regularity of the robust equilibrium along with the full support assumption ensure that $\sum_{r \in R_a^{C'}, \mathbf{q}: q_a = q_a^-(p)} m^\infty_{r,\mathbf{q}}$ is bounded away from zero, which implies that $\bar{M}_a^{C'}(p)$ is strictly decreasing with a slope bounded away from zero. □

Claim 3. *For any $\epsilon' > 0$, there exists a subsequence $\{M^{k_\ell}\}_{\ell=1}^\infty$ such that $\lim_{\ell \to \infty} \Pr\{\sup_p ||M^{k_\ell}(p) - \bar{M}(p)|| > \epsilon'\} = 0$.*

Proof. Using the argument in the proof of Glivenko-Cantelli, we can partition the space of p into finite intervals $[p_i, p_{i+1}]$, where $i = 0, ..., i_j^*$ such that $||\bar{M}(p_{i+1}) - \bar{M}(p_i)|| < \epsilon'/2$ for all i. Let N be the number of such intervals. Using the argument of Glivenko-Cantelli, one can show that for any p there exists i such that

$$||M^{k_\ell}(p) - \bar{M}(p)|| \leq \max\{||M^{k_\ell}(p_i) - \bar{M}(p_i)||, ||M^{k_\ell}(p_{i+1}) - \bar{M}(p_{i+1})||\} + \epsilon'/2.$$

Suppose event $\{||M^{k_\ell}(p) - \bar{M}(p)|| > \epsilon'\}$ occurs for some p. Then there must exist i such that $||M^{k_\ell}(p_i) - \bar{M}(p_i)|| \geq \epsilon'/2$. Since $\bar{M}^{k_\ell}(\cdot)$ converges to $\bar{M}(\cdot)$ uniformly, by Claim 2, there exists K' such that for all $\ell > K'$, $\sup_p ||\bar{M}^{k_\ell}(p) - \bar{M}(p)|| < \epsilon'/4$. Hence, for $\ell > K'$ and for i, we must have

$$||M^{k_\ell}(p_i) - \bar{M}^{k_\ell}(p_i)|| \geq \epsilon'/4.$$

Combining the arguments so far, we conclude:

$$\begin{aligned}
&\Pr\{\sup_p ||M^{k_\ell}(p) - \bar{M}(p)|| > \epsilon'\} \\
=&\Pr\{\exists p \text{ s.t. } ||M^{k_\ell}(p) - \bar{M}(p)|| > \epsilon'\} \\
\leq&\Pr\{\exists i \text{ s.t. } ||M^{k_\ell}(p_i) - \bar{M}(p_i)|| > \epsilon'/2\} \\
\leq&\Pr\{\exists i \text{ s.t. } ||M^{k_\ell}(p_i) - \bar{M}^{k_\ell}(p_i)|| > \epsilon'/4\} \\
=&\Pr\{\cup_i\{||M^{k_\ell}(p_i) - \bar{M}^{k_\ell}(p_i)|| > \epsilon'/4\}\} \\
\leq&\sum_i \Pr\{||M^{k_\ell}(p_i) - \bar{M}^{k_\ell}(p_i)|| > \epsilon'/4\} \\
\leq&\sum_i e^{-k_\ell \epsilon'^2/8} \\
=&Ne^{-k_\ell \epsilon'^2/8} \to 0 \text{ as } \ell \to \infty,
\end{aligned}$$

where the penultimate inequality follows from Claim 1. □

Claim 4. *There exist a subsequence of random k economies such that the resulting cutoffs of schools along that subsequence of finite economies converge to some (deterministic) cutoffs $\bar{\mathbf{p}} = (\bar{p}_c)_{c\in C}$ in probability.*

Proof. Fix any random k or limit economy. Below we do not index the economy, keeping in mind it applies to any finite or limit economy. The cutoffs of schools are determined recursively as follows (see Che and Kojima (2010) and Abdulkadiroglu, Che, and Yasuda (2015)): Let $C^0 \equiv C$ $p^0 \equiv 1$, and $X_a^0 \equiv 0$ for every $a \in C$. Given $C^0, p^0, \{X_a^0\}_{a\in C}, \ldots, C^{t-1}, p^{t-1}, \{X_a^{t-1}\}_{a\in C}$, and for each $a \in C$ define

$$p_a^t = \inf\left\{p \in [0,1] \,\middle|\, X_a^{t-1} + \left(M_a^{C^{t-1}}(p) - M_a^{C^{t-1}}(p^{t-1})\right) \geq S_a\right\}, \tag{A.1}$$

$$p^t = \max_{a\in C^{t-1}} p_a^t, \tag{A.2}$$

$$C^t = C^{t-1} \setminus \{a \in C^{t-1} | p_a^t = p^t\}, \tag{A.3}$$

$$X_a^t = X_a^{t-1} + M_a^{C^{t-1}}(p^t) - M_a^{C^{t-1}}(p^{t-1}). \tag{A.4}$$

Each recursion step t determines the cutoff of a new school, given the cutoffs $\{p^0, ..., p^{t-1}\}$ of schools determined in the preceding steps. Students with score $s < p^{t-1}$ can never be assigned schools $C \setminus C^{t-1}$. For each school $a \in C^{t-1}$ with remaining capacity, a measure X_a^{t-1} of the seats is claimed by students with scores higher than p^{t-1}, so only measure $S_a - X_a^{t-1}$ of seats can be assigned to students with scores $s < p^{t-1}$. If school a has the next highest cutoff, p^t, then the remaining capacity $S_a - X_a^{t-1}$ must equal the measure of those students who prefer a most among C^{t-1} and have scores in $[p^t, p^{t-1}]$. This, together with the fact that school a has cutoff p^t, implies (A.1) and (A.2). The recursion definition implies (A.3) and (A.4). Note that there are only finite recursion steps, at most the number of schools. The recursive equations uniquely determine the set of cutoffs $\{p^0, ..., p^C\}$. The cutoff for school $a \in C$ is then given by $P_a := \{p^t | p_a^t = p^t\}$.

We are finally in a position to prove the statement. Consider the cutoffs in the random k-economy, $p^k = \{p^{0k}, ..., p^{Ck}\}$ and those $\bar{p} = \{\bar{p}^0, ..., \bar{p}^C\}$ in the limit economy. By Step 2, $M_a^{C'k}(\cdot)$ converges uniformly to $\bar{M}_a^{C'}(\cdot)$ in probability and that $\bar{M}_a^{C'}(\cdot)$ is continuous and strictly decreasing, has a slope bounded below in absolute value. Since the equations (A.1)-(A.4) are continuous in $M_a^{C'}$'s, p^{ik} converges in probability to $\bar{p}^i$, recursively, taking subsequences at each recursion step, if necessary (recall there are finite recursion steps). □

For the next result, we define a few terminologies. In the random k economy, we say a student with θ plays **SRS against p** when she submits a ROL that yields the same assignment as truthful reporting for almost all realizations of STB lottery draws, and she plays **non-SRS against p** when she does not play SRS against **p**

The following preliminary result is useful.

CLAIM 5. *Fix any* **p**, *and let* $q_c^- := \sup\{q_c \in Q_c | q_c \leq p_c\}$ *for each* $c \in C$. *Let* $\xi > 0$ *be such that*[A-1]

$$\xi < \min_{c \in C}\{n_c(p_c - q_c^-) : n_c(p_c - q_c^-) > 0\},$$

$$\xi < \min_{c \in C}\{1 - n_c(p_c - q_c^-) : n_c(p_c - q_c^-) < 1\}$$

and

$$\xi < \min_{a,b \in C}\{|n_a(p_a - q_a^-) - n_b(p_b - q_b^-)| : n_a(p_a - q_a^-) \neq n_b(p_b - q_b^-)\}.$$

Then, when a student with any θ *changes from an SRS against* **p** *to a non-SRS against* **p**, *her assignment worsens in the strict first-order stochastic dominance sense as follows: for each* j, *the probability of receiving* j*-th best or better schools are weakly higher under the former strategy than under the latter strategy, and there exists* j' *such that the assignment probability differs at least by* ξ *for the a* j'*-th best or better schools.*

Proof. Note that a student is assigned a school c (say by "top-ranking" it) either if she has intrinsic priority is $q_c > q_c^-$ ("clear-in") or if $q_c = q_c^-$ ("marginal") but her lottery draw is above the critical threshold $\hat{s}_a = n_c(p_c - q_c^-)$. Fix an SRS and a non-SRS. The former produces the same assignment as truthful reporting with probability one, and assigns a better school than the latter in the sense of first-order stochastic dominance. Further, there must exist a smallest j such that the probability of receiving the j-th best or better school is strictly higher under the former strategy than under the latter strategy. Let a be that j-th best school for this student. It must be that the student must be assigned a with strictly positive probability. This means that the schools that are better than a for this student—and thus would be listed ahead of a under truthful reporting—cannot be a clear-in school, or else she would have been assigned at worst that school with probability one under the SRS. By the same logic, either the all these better schools are infeasible, or the lottery thresholds for these schools must be strictly higher than the lottery threshold for a. In the former case, our choice of ξ ensures that the student is assigned s under the SRS with probability at least ξ. In the latter case, the thresholds for the schools better than a are all no less than ξ, and if a is also a marginal school for

[A-1] Note since C and Q_c's are all finite, it is possible to find such an $\xi > 0$.

this student, they must be at least ξ higher than the threshold for a. Similarly, under the non-SRS strategy, the same observations apply to some other school, say b, that is worse than a. Since the lottery draws are uniform on $[0,1]$, a shift from the SRS to the non-SRS results in the change of assignment from a to b with probability at least ξ.[A-2] In short, the probability of the student being assigned her j-th best or better schools are higher under the SRS than that under the non-SRS at least by ξ. Since the selection of an SRS and non-SRS is arbitrary, the result holds. □

The next result is crucial.

CLAIM 6. . *The fraction of students who play non-SRS in* $\left\{\left(\sigma_i^k\right)_{1\leq i\leq k}\right\}_k$ *converges in probability to zero as* $k\to\infty$.

Proof. Suppose not. Then, there exists $\varepsilon>0$ and a subsequence of finite economies $\left\{F^{k_j}\right\}_j$ such that

$$\Pr\left(\text{The fraction of students playing SRS against } P^{k_j} \text{ is no less than } 1-\varepsilon\right)<1-\varepsilon. \qquad (*)$$

By Claim 4, there exists a sub-subsequence $\mathbf{p}^{k_{j_l}}$ converges to $\bar{\mathbf{p}}$ in probability, where $\bar{\mathbf{p}}$ are the deterministic cutoffs identified in Claim 4.

Define

$$\hat{\Theta}:=\left\{(u,s):|u_c-u_{c'}|>\delta \text{ for all } c\neq c'\right\}.$$

We choose $\delta>0$ small enough so that $\delta<\xi/3$, where ξ satisfies the hypothesis of Claim 5 with $\mathbf{p}=\bar{\mathbf{p}}$ and that $\eta(\hat{\Theta})>(1-\varepsilon)^{1/3}$ (this can be done since η is absolutely continuous). The first condition guarantees that, whenever $\mathbf{p}^k$ is within δ distance of $\bar{\mathbf{p}}$, the hypothesis of Claim 5 holds for $\xi=\delta$ for vector $\mathbf{p}^k$, and thus the statement of Claim 5 holds for SRS and non-SRS against $\mathbf{p}^k$ with δ playing the role of ξ. This fact will be used later.

By WLLN, we know that $\eta^{k_{j_l}}(\hat{\Theta})$ converges to $\eta(\hat{\Theta})$ in probability, and therefore there exists L_1 such that for all $l>L_1$ we have

$$\Pr\left(\eta^{k_{j_l}}(\hat{\Theta})\geq(1-\varepsilon)^{1/2}\right)\geq(1-\varepsilon)^{1/2}.$$

For each economy $F^{k_{j_l}}$, define the event

$$A^{k_{j_l}}:=\left\{\left|p_c^{k_{j_l}}-\bar{p}_c\right|<\delta \text{ for all } c\in C\right\}.$$

Because $\mathbf{p}^{k_{j_l}}$ converges to $\bar{\mathbf{p}}$ in probability, there exists L_2 s.t. for all $l>L_2$ we have

$$\Pr\left(A^{k_{j_l}}\right)\geq\max\left\{(1-\varepsilon)^{1/6},1-(1-\varepsilon)^{1/2}\left[(1-\varepsilon)^{1/3}-(1-\varepsilon)^{1/2}\right]\right\}. \qquad (**)$$

Because $\left\{\left(\sigma_i^k\right)_{1\leq i\leq k}\right\}_k$ is a robust equilibrium, there exists L_3 s.t. for all $l>L_3$ the strategy profile $\left(\sigma_i^{k_{j_l}}\right)_{i=1}^{k_{j_l}}$ is a $\delta^2\left[(1-\varepsilon)^{1/6}-(1-\varepsilon)^{1/3}\right]$-BNE for economy $F^{k_{j_l}}$.

[A-2] The student may be assigned both a and b under either strategy with strictly positive probabilities. The statement applies even in this case since the lottery thresholds must be a least ξ apart if both schools are marginal.

By WLLN, there exists $\hat{L}$ s.t. $\hat{L}$ i.i.d. Bernoulli random variables with probability $(1-\varepsilon)^{1/3}$ have a sample mean greater than $(1-\varepsilon)^{1/2}$ with probability no less than $(1-\varepsilon)^{1/3}$. Then we find L_4 s.t. $l > L_4$ implies $(1-\varepsilon)^{1/2} k_{j_l} > \hat{L}$.

Now let's fix an arbitrary $l > \max\{L_1, L_2, L_3, L_4\}$, and we wish to show that in economy $F^{k_{j_l}}$

$$\Pr\left(\text{The fraction of students playing SRS against } \mathbf{p}^{k_{j_l}} \text{ is no less than } 1-\varepsilon\right) \geq 1-\varepsilon,$$

which would contradict $(*)$ and complete the proof.

We first prove that in economy $F^{k_{j_l}}$, a student with $\theta \in \hat{\Theta}$ plays SRS against $\bar{\mathbf{p}}$ with probability no less than $(1-\varepsilon)^{1/3}$. To see this, suppose to the contrary that there exists some student i and some type $\theta \in \hat{\Theta}$ s.t.

$$\Pr\left(\sigma_i^{k_{j_l}}(\theta) \text{ plays SRS against } \bar{\mathbf{p}}\right) < (1-\varepsilon)^{1/3}.$$

Then deviating to truthful reporting will give this student i with type $\theta \in \hat{\Theta}$ at least a gain of

$$\begin{aligned}
&\mathbb{E}\left[i\text{'s gain from deviation} \mid \sigma_i^{k_{j_l}} \text{ plays non-SRS against } \mathbf{p}^{k_{j_l}}\right] \Pr\{\sigma_i^{k_{j_l}} \text{ plays non-SRS against } \mathbf{p}^{k_{j_l}}\} \\
\geq &\mathbb{E}\left[i\text{'s gain from deviation} \mid \{\sigma_i^{k_{j_l}} \text{ plays non-SRS against } \mathbf{p}^{k_{j_l}}\} \wedge A^{k_{j_l}}\right] \cdot \Pr\left(\begin{array}{c}\sigma_i^{k_{j_l}}(\theta) \text{ plays non-SRS against } \bar{\mathbf{p}} \\ \textit{and} \text{ event } A^{k_{j_l}}\end{array}\right) \\
\geq &\delta^2 \cdot \Pr\left(\begin{array}{c}\sigma_i^{k_{j_l}}(\theta) \text{ plays non-SRS against } \mathbf{p}^{k_{j_l}} \\ \textit{and} \text{ event } A^{k_{j_l}}\end{array}\right) \\
\geq &\delta^2 \cdot \Pr\left(\begin{array}{c}\sigma_i^{k_{j_l}}(\theta) \text{ plays non-SRS against } \bar{\mathbf{p}} \\ \textit{and} \text{ event } A^{k_{j_l}}\end{array}\right) \\
\geq &\delta^2 \left[\Pr\left(A^{k_{j_l}}\right) - \Pr\left(\sigma_i^{k_{j_l}}(\theta) \text{ plays SRS against } \bar{\mathbf{p}}\right)\right] \\
\geq &\delta^2 \left[(1-\varepsilon)^{1/6} - (1-\varepsilon)^{1/3}\right],
\end{aligned}$$

where the inequalities are explained as follows. The first inequality holds since the gains from the deviation is nonnegative whether the event $A^{k_{j_l}}$ occurs or not. The second inequality results from two observations: (i) the assignment probability improves in the sense of Claim 5 at least by δ since the hypothesis of Claim 5 holds for δ in place of ξ against $\mathbf{p} = \mathbf{p}^{k_{j_l}}$, which in turn follows from our choice of δ (i.e., set as at most $1/3$ of ξ required for $\mathbf{p} = \bar{\mathbf{p}}$) and the fact that conditional on event $A^{k_{j_l}}$, $\mathbf{p}^{k_{j_l}}$ is within δ of $\bar{\mathbf{p}}$.[A-3] and (ii) the utility gain from an improved assignment is at least δ for type $\theta \in \hat{\Theta}$. The third inequality holds since, given our choice δ, conditional on event $A^{k_{j_l}}$, a non-SRS against $\bar{\mathbf{p}}$ is a non-SRS against $\mathbf{p}^{k_{j_l}}$. The above inequalities contradict $l > L_3$, which implies that the strategy profile $\left(\sigma_i^{k_{j_l}}\right)_{i=1}^{k_{j_l}}$ is a $\delta^2\left[(1-\varepsilon)^{1/6} - (1-\varepsilon)^{1/3}\right]$-BNE for the economy $F^{k_{j_l}}$.

Therefore, in economy $F^{k_{j_l}}$, for each student $i = 1, \ldots, k_{j_l}$ and each $\theta \in \hat{\Theta}$, we have

$$\begin{aligned}
&\Pr\left(\sigma_i^{k_{j_l}}(\theta) \text{ plays SRS against } \bar{\mathbf{p}} \,\middle|\, \eta^{k_{j_l}}(\hat{\Theta}) \geq (1-\varepsilon)^{1/2}\right) \\
&= \Pr\left(\sigma_i^{k_{j_l}}(\theta) \text{ plays SRS against } \bar{\mathbf{p}}\right) \geq (1-\varepsilon)^{1/3}, \qquad (***)
\end{aligned}$$

[A-3] This construction ensures that distinct lottery thresholds under $\mathbf{p}^{k_{j_l}}$ are apart from one another at least by δ, and all interior thresholds are away from 0 and 1 at least by δ.

where the first equality holds because student i's random report according to her mixed strategy is independent of random draws of the students' type.

Then we have

$$\begin{aligned}
&\Pr\left(\begin{array}{c}\text{The fraction of students with } \theta\in\hat{\Theta}\\ \text{playing SRS against } \bar{\mathbf{p}} \text{ is no less than } (1-\varepsilon)^{1/2}\end{array}\ \middle|\ \eta^{k_{j_l}}(\hat{\Theta})\geq(1-\varepsilon)^{1/2}\right)\\
&\geq \Pr\left(\begin{array}{c}\eta^{k_{j_l}}(\hat{\Theta})\cdot k_{j_l} \text{ i.i.d. Bernoulli random variables with probability } (1-\varepsilon)^{1/3}\\ \text{have a sample mean no less than } (1-\varepsilon)^{1/2}\end{array}\ \middle|\ \eta^{k_{j_l}}(\hat{\Theta})\geq(1-\varepsilon)^{1/2}\right)\\
&\geq \Pr\left(\begin{array}{c}\hat{L} \text{ i.i.d. Bernoulli random variables with probability } (1-\varepsilon)^{1/3}\\ \text{have a sample mean no less than } (1-\varepsilon)^{1/2}\end{array}\right)\\
&\geq (1-\varepsilon)^{1/3},
\end{aligned}$$

where the first inequality is because of the inequality (***) and that σ_i's are independent across students conditioning on the event $\eta^{k_{j_l}}(\hat{\Theta})\geq(1-\varepsilon)^{1/2}$, and the second inequality is because $l>L_4$ and $\eta^{k_{j_l}}(\hat{\Theta})\geq(1-\varepsilon)^{1/2}$ imply $\eta^{k_{j_l}}(\hat{\Theta})\cdot k_{j_l}>\hat{L}$.

Comparing the finite economy random cutoff $\mathbf{p}^{k_{j_l}}$ with the deterministic cutoff $\bar{\mathbf{p}}$, we have

$$\begin{aligned}
&\Pr\left(\begin{array}{c}\text{The fraction of students with } \theta\in\hat{\Theta}\\ \text{playing SRS against } \mathbf{p}^{k_{j_l}} \text{ is no less than } (1-\varepsilon)^{1/2}\end{array}\ \middle|\ \eta^{k_{j_l}}(\hat{\Theta}\ )\geq(1-\varepsilon)^{1/2}\right)\\
&\geq \Pr\left(\begin{array}{c}\text{The fraction of students with } \theta\in\hat{\Theta}\\ \text{playing SRS against } \bar{\mathbf{p}} \text{ is no less than } (1-\varepsilon)^{1/2},\\ \text{and event } A^{k_{j_l}}\end{array}\ \middle|\ \eta^{k_{j_l}}(\hat{\Theta})\geq(1-\varepsilon)^{1/2}\right)\\
&\geq \Pr\left(\begin{array}{c}\text{The fraction of students with } \theta\in\hat{\Theta}\\ \text{playing SRS against } \bar{\mathbf{p}} \text{ is no less than } (1-\varepsilon)^{1/2}\end{array}\ \middle|\ \eta^{k_{j_l}}\ (\hat{\Theta})\geq(1-\varepsilon)^{1/2}\right)\\
&-\Pr\left(\bar{A}^{k_{j_l}}\middle|\ \eta^{k_{j_l}}(\hat{\Theta})\geq(1-\varepsilon)^{1/2}\right)\\
&\geq (1-\varepsilon)^{1/3}-\frac{\Pr\left(\bar{A}^{k_{j_l}}\right)}{\Pr\left(\eta^{k_{j_l}}(\hat{\Theta})\geq(1-\varepsilon)^{1/2}\right)}\\
&\geq (1-\varepsilon)^{1/3}-\frac{(1-\varepsilon)^{1/2}\left[(1-\varepsilon)^{1/3}-(1-\varepsilon)^{1/2}\right]}{(1-\varepsilon)^{1/2}}\\
&=(1-\varepsilon)^{1/2},
\end{aligned}$$

where the last inequality is because of (**).

The construction of L_1 implies $\Pr\left(\eta^{k_{j_l}}(\hat{\Theta})\geq(1-\varepsilon)^{1/2}\right)\geq(1-\varepsilon)^{1/2}$, and so finally we have in

economy $F^{k_{j_l}}$

$$
\begin{aligned}
&\Pr\left(\text{The fraction of students playing SRS against } \mathbf{p}^{k_{j_l}} \text{ is no less than } 1-\varepsilon\right) \\
&\geq \Pr\left(\begin{array}{c}\text{At least } (1-\varepsilon)^{1/2} \text{ of students with } \theta \in \hat{\Theta} \text{ play SRS against } \mathbf{p}^{k_{j_l}} \\ \textit{and } \eta^{k_{j_l}}(\hat{\Theta}) \geq (1-\varepsilon)^{1/2}\end{array}\right) \\
&= \Pr\left(\eta^{k_{j_l}}(\hat{\Theta}) \geq (1-\varepsilon)^{1/2}\right) \cdot \Pr\left(\begin{array}{c}\text{At least } (1-\varepsilon)^{1/2} \text{ of students with } \theta \in \hat{\Theta} \\ \text{play SRS against } \mathbf{p}^{k_{j_l}}\end{array} \,\middle|\, \eta^{k_{j_l}}(\hat{\Theta}) \geq (1-\varepsilon)^{1/2}\right) \\
&\geq (1-\varepsilon)^{1/2} \cdot (1-\varepsilon)^{1/2} \\
&= 1-\varepsilon,
\end{aligned}
$$

where the last inequality follows from the above string of inequalities. Therefore, we have obtained a contradiction to $(*)$. □

The statement of Theorem 1 follows from Claim 6.

□

B Proofs from Section 3

Proof of Proposition 1. For any c_l ranked in student i's submitted R_i, TE with threshold t_1 infers that c_l is more preferred than all other schools included in $\mathcal{I}^{c_l,t_1}$. Moreover, $\mathcal{I}^{c_l,t_2} \subseteq \mathcal{I}^{c_l,t_1}$, as evident in the TE procedure whenever $t_1 < t_2$. Therefore, TE with threshold t_1 infers that c_l is more preferred than any school in $\mathcal{I}^{c_l,t_2} \setminus \{c_l\}$. The same argument can be made for any school ranked in R_i. This proves the nesting structure. The second statement is implied by the nesting structure. □

Proof of Proposition 2. The first part of the first nesting structure is from Proposition 1.

To show $\mathcal{O}(\text{TE}_{t_0} \mid F^k, \sigma^k_{1\leq i\leq k}) \subseteq \mathcal{O}(\text{WTT} \mid F^k, \sigma^k_{1\leq i\leq k})$, notice that $\mathcal{I}^{c_k,t_0}$ can never include a school that is ranked above c_k in i's ROL, R_i. That is, TE with threshold 0 never infers that c_k is more preferred to a school that is ranked above c_k in R_i. Therefore, TE with threshold 0 cannot infer more preference relations than WTT.

$\mathcal{O}(\text{Stability} \mid F^k, \sigma^k_{1\leq i\leq k}) \subseteq \mathcal{O}(\text{TE}_{t_0} \mid F^k, \sigma^k_{1\leq i\leq k})$ is true because TE with threshold 0 uses all possible lottery realizations as well as preference transitivity; by contrast, stability relies on one lottery realization and thus cannot use transitivity. □