Strategic “Mistakes”: Implications for Market Design Research*

Georgy Artemov†  Yeon-Koo Che‡  Yinghua He§

April 4, 2017

Abstract

A field data from Australian college admissions shows that a non-negligible fraction of applicants choose strategies (or rank-ordered lists) that are unambiguously dominated, but that the majority of these “mistakes” are payoff irrelevant. In keeping with this result, we develop a theory suggesting that the presence of such mistakes jeopardizes the identification method based on truthful reporting hypothesis under a (seemingly) strategy-proof mechanism, but leaves the method based on weaker stability condition relatively unscathed. Monte Carlo simulation further confirms this point and quantifies the differences between these two methods in the structural estimation of preference parameters and in a hypothetical counterfactual analysis.

JEL Classification Numbers: C70, D47, D61, D63.

Keywords: Strategic mistakes, weakly dominated strategies, robust equilibria, truthful reporting strategy, stable response strategy.

1 Introduction

Strategy-proofness—or making it a dominant strategy for revealing one’s preferences truthfully—is an important desideratum in market design. Not only does strategy-proofness make it straightforward for a participant to act in one’s best interest, thus minimizing the scope for making mistakes; but it also equalizes the playing field, for even an unsophisticated participant is protected from others who may game the system. Further, it aids empirical research by making participants’ choices easy to interpret.

*We are grateful to Xingye Wu, who has provided excellent research assistance for the theory part. We are grateful to the participants of Asia Pacific Industrial Organization Conference for their comments.
†Department of Economics, University of Melbourne, Australia. Email: georgy@gmail.com
‡Department of Economics, Columbia University, USA. Email: yeonkooche@gmail.com.
§Department of Economics, Rice University, USA; Toulouse School of Economics, France. Email: yinghua.he@gmail.com.
However, this view has been challenged by a growing number of authors who find that strategic mistakes are not uncommon even in strategy-proof environments. Laboratory experiments have shown that a significant fraction of subjects do not report their preferences truthfully even in strategy-proof mechanisms such as applicant-proposing deferred acceptance algorithm (DA) and top-trading cycles (TTC). More alarmingly, similar problems occur in a high-stake real-world context. In a study of assignment for Israeli graduate programs in psychology (which uses DA), Hassidim, Romm, and Shorrer (2016) find that about 19% of applicants either did not list scholarship position, or listed scholarship/non-scholarship positions in a wrong order. Since a scholarship position is unambiguously preferred to a non-scholarship position, such behavior constitutes a dominated strategy. In a similar vein, Rees-Jones (2016) report about 17% of 579 US medical seniors surveyed indicated misrepresenting their preferences in the NMRP.

These findings raise questions on prominent mechanisms used widely in practice and their empirical assessment. At the same time, a mere presence of “mistakes” is not enough to draw conclusions on the matters. If mistakes were made only when they would have made little difference in the outcome, then the full rationality hypothesis may be a reasonable proxy for understanding a mechanism. One would thus require a deeper understanding of what the nature of mistakes is and what circumstances led to those mistakes.

Toward this end, we first study a field data collected from Victorian Tertiary Admissions Centre (VTAC), a central clearinghouse that organizes the match for students in college admissions in Victoria, Australia. The semi-centralized mechanism VTAC uses for assigning students to university “courses” (which are similar to college-major pairs in the US) resembles a serial dictatorship with the serial order given by the nation-wide test score, called Equivalent National Tertiary Entrance Rank (ENTER), except for submittable rank-ordered lists (ROLs) being truncated to 12 and a few other features (that will be explained later). As with Hassidim, Romm, and Shorrer (2016), our study exploits the unique feature of the system in which an applicant can apply for a given course as either a (1) “full-fee” position (or FFP) which requires full tuition or (2) a common-wealth supported position (or CSP) which subsidizes 50% of tuition, or both. Since CSP clearly dominates FFP for any course, ranking a course for the latter but not for the former in a ROL that does not fill up the 12 slots—henceforth called a skip—is unambiguously a dominated strategy.

In the sample year of 2007, we find that 1,009 applicants skipped, which comprises 3.6% out of total 27,922 applicants who submitted fewer than 12 courses, 34% of those who listed at least one full-fee course. Although these figures are not directly comparable with those found in Hassidim, Romm, and Shorrer (2016), as will be explained below, they can be viewed as non-negligible. However, the vast majority of these mistakes were not payoff relevant. Correcting the mistakes

---

1Obviously strategyproof mechanisms, as suggested by Li (2017), makes it more transparent for participants to play a dominant strategy. But this more demanding requirement is often difficult to meet in practice (Ashlagi and Gonczarowski (2016)). While the requirement is met in the case of a serial dictatorship (see Li (2017) and Pycia and Troyan (2016)), the VTAC mechanism does not coincide precisely with a serial dictatorship.

2As will be discussed in detail, listing the latter ahead of the former need not be a dominated strategy in our data.
(i.e., listing the omitted CSP course) would have made a difference only for 14 – 201 applicants out of 1,009 who skipped at least one CSP course, with the exact number depending on where they would have listed them in their ROLs (e.g., top of or just ahead of the full fee positions of the courses) and how the courses would have evaluated them. In other words, payoff relevant mistakes comprise only 0.05 – 0.72% out of all students and 0.47 – 6.78% out of those who listed at least one full-fee dual course.

Our rich micro data set is well suited to investigate who made mistakes, whether the mistakes are payoff-relevant and what circumstances led to them. We find that applicants’ academic ability (measured independently of ENTER) is negatively correlated with skips, suggesting that misunderstanding the mechanism may play a role in making mistakes. However, even controlling for the academic ability, ENTER is also negatively correlated with mistakes. This suggests that applicants omit courses they are unlikely to be admitted; furthermore, we find no evidence that omitting courses is a conscious attempt to game the mechanism and receive a better match.

The individual characteristics correlated with payoff-relevant mistakes are very different. There is no correlation with academic ability anymore, and there is a positive, rather than negative, correlation with ENTER. We also establish that those who make payoff-relevant mistakes overall list less-selective courses. In contrast, the payoff-irrelevant mistakes are not correlated with less-selective courses in ROL, which further suggest a different nature of these two types of mistakes.

A unique feature of Victorian mechanism, which requires applicants to submit several ROLs over time, allows us to observe further differences between skips and payoff-relevant mistakes. While the number of applicants who skip increases over time, the number of applicants who make payoff relevant mistakes decreases. We further exploit the fact that we observe ROLs submitted before and after the applicants receive their ENTER. We study applicants’ response to a “shock”: a deviation of realized ENTER from ENTER forecasted based on the ability test. Positive shock, despite making an applicant eligible for a larger set of courses, leads to a reduction in payoff-relevant mistakes. It has no effect on skips.

To the extent that mistakes do occur and some (small) fraction of them are payoff relevant, it is of interest to know how much they affect the overall outcome. Of particular interest for market design research are how mistakes—some of them payoff-relevant—do affect our ability to recover the underlying preferences of participants and perform a counter-factual analysis of a new hypothetical market design. To study these questions, we first develop a theoretical model of applicants’ behavior in a large matching market operated by a DA mechanism (of which a serial dictatorship is a special case). In keeping with the empirical findings, we focus on an equilibrium concept—called robust equilibrium—which allows applicants to make mistakes as long as they become virtually payoff-irrelevant as the market size grows arbitrarily large.

We show that it is a robust equilibrium behavior for all except for a vanishing fraction of applicants to submit ROLs that differ from their true preferences, conditional on applying at all. In this equilibrium, applicants drop courses at the top of their preferences order that they feel they clearly stand no chance of getting and/or courses at the bottom dominated by a course that they
feel they have a clear shot at. Such a behavior is supported as robust equilibrium behavior since as the market grows large the risk of playing such a strategy disappears for a large fraction—in fact all but a vanishing fraction—of applicants. If applicants behave according to our robustness concept, this result implies that the observed ROLs need not reflect the applicants’ true preference orders. This calls into question empirical identification method based on the hypothesis that applicants submit true preferences as their ROLs in a DA mechanism.

We next show that, in any robust equilibrium, as the market grows large, almost all applicants must be playing a stable response strategy, a strategy that guarantees admission into the most preferred course among those that they could have gotten into had they submitted truthful ROLs. This result implies that stable response behavior is a valid identification restriction in a sufficiently large market. While truthful reporting is stable response, a stable response need not involve truthful ROL. Hence, this latter restriction is weaker. The two theoretical results provide the sense in which the identification method based on truthful reporting is vulnerable to the types of mistakes documented in the first part and at the same time the sense in which the identification method based on a weaker stable response strategy is relatively robust to them.

To gain quantitative insights, we perform a Monte Carlo simulation of school assignment model in which student preferences follow a logit model. We assume a serial dictatorship mechanism with a pre-specified serial order as an initial mechanism. Even though this mechanism is strategyproof, in keeping with our empirical findings, we entertain alternative scenarios that vary in the extent and frequencies in which students make mistakes. Specifically, the assumed behavior ranges from truthful reporting (i.e., no mistakes), behavior exhibiting varying degrees of payoff-irrelevant mistakes, to ones exhibiting varying degrees of payoff-relevant mistakes. Under these alternative scenarios, we structurally estimate applicants’ preferences using (i) truthful reporting and (ii) stable response as two alternative identifying restrictions.

The estimation results highlight the bias-variance tradeoff: Estimation based on truthful reporting uses more information on revealed student preferences and has a much lower standard error than the one based on stable response; however, a bias emerges in the former whenever there are some students making mistakes, while the latter is immune to all payoff-irrelevant mistakes. As expected, the truthful reporting restriction introduces downward biases in the estimators of school quality, especially among popular or small schools, because over-subscribed schools are often skipped by many students.

Given the bias-variance tradeoff between the two approaches, the truthful reporting restriction is obviously preferred whenever it is satisfied. Based on the nesting structure of the two restrictions, we then also propose a statistical test similar to the Durbin-Wu-Hausman test. Indeed, our simulation results show that the test has the correct size and reasonable statistical power.

We further quantify the biases in a counterfactual analysis in which we implement a hypothetical affirmative action policy helping disadvantaged students. In terms of predicting the counterfactual matching outcome, estimates from truthful reporting perform worse than those from stable responses when students make mistakes. When we evaluate the welfare effects of the policy, the
truthful reporting restriction under-estimates the benefits to disadvantaged students and the harm on others; stable response however predicts the effects close to true values. In addition, we also evaluate another approach to counterfactual analysis in market design research: holding submitted ROLs as constant across two policies. It produces an even larger bias than the truthful reporting restriction.

The rest of the paper is organized as follows. Section 2 studies the frequency and nature of strategic mistakes from the VTAC college admissions data. Section 3 explores the theoretical implications of the findings for empirical identification methods. Section 4 performs Monte Carlo simulation on the alternative identification methods.

2 Strategic Mistakes in Australian College Admissions

2.1 Institutional Details and Data

We use the data for year 2007 from Victorian Tertiary Admission Centre (VTAC), which is a centralized clearing house for admissions to tertiary courses in Victoria. Applicants are required to rank tertiary courses they want to be considered for; VTAC also collects academic and demographic information about applicants.

The unit of admission in Victoria, a tertiary course, is a combination of a field of study the applicant wants to pursue and a fee structure. A field of study is roughly equivalent to a major in the US universities, although in Victoria applicants use the same process to apply for technical schools and non-degree programs. There are two fee structures, or seat types. Type 1 seats are subsidized by the government. The government covers about one-half of tuition costs and offers a subsidized loan to cover the other half. Students in Type 2 seats are required to pay full tuition. The tuition vary by major and, to lesser degree, by the university; median tuition is approximately AUD9,000 (about USD7,000), but it can be as high as AUD20,000 (USD15,000). Apart from tuition payments, there is no difference between Type 1 and Type 2 seats. Students who start in Type 2 seat may be allowed to transfer to a Type 1 seat later.

The normal duration of a degree program is three years. There were 1899 majors in 2007, 881 offering both Type 1 and Type 2 seats.

The timing of tertiary admission for domestic applicants is the following. Applications and a rank-ordered list (ROL) of courses, are required by the end of September. There is a single fee, independent of the number of courses listed of approximately USD25. Applicants are allowed to list at most 12 courses. Late applications are allowed, until mid-December, for an additional fee of approximately USD75.

Applicants are allowed to revise ROL submitted at the time of the application during two time windows: a one-week window in late October and three-week window in late December. The changes can be made either over the phone, for a fee, or online, for free.

---

3Students who start in Type 2 seat may be allowed to transfer to a Type 1 seat later.
In mid-December, applicants who are current high school students receive their Equivalent National Tertiary Entrance Rank (ENTER), which we will refer to as Score throughout the paper. Score is based on a variety of state-wide exams. For student $i$, $\text{Score}_i$ shows a percentage of applicants with an aggregated exam grades below the grades of $i$. After receiving their Score, students have about one week to revise their preliminary ROLs; after the deadline, the lists become final.

Once ROLs are finalized, courses select applicants and transmit their selection to VTAC. Using applicant’s ROL, VTAC picks the highest-ranked course that selected the applicant, one of each type, and transmit the offer(s) to the applicant. That is, if applicant’s ROL contains both Type 1 and Type 2 courses, such an applicant may receive two offers, one of each type.\footnote{See appendix for additional institutional details related to the process.}

In selecting applicants, courses follow a pre-specified, published set of rules. For the largest category of applicants, the admission is based solely on their Score. We focus on these applicants in this paper and refer to them as V16 applicants, following the code assigned to them by VTAC. These are the current high school students who follow the standard Victorian curriculum.\footnote{There is a variety of other categories of applicants, such as past high school graduates, high school students who follow non-standard curricula or take tertiary courses while in high school. For these applicants, courses are allowed to use the criteria other than Score. As we do not observe these criteria, we exclude these students from our analysis. Individual courses, usually in fine arts, may require V16 applicants to submit a portfolio or to audition. Courses are also allowed to give small bonuses to applicants. We ignore these in our analysis.} We refer to the average of the highest score among all rejected applicants and the lowest score among all accepted applicants as a cutoff of the course.

The number of courses that an applicant can list is restricted by 12. Because the applicants who exhaust the length of their ROLs may be forced to omit some courses that they find desirable, we focus only on those who list fewer than 12 courses, which is 75% of all applicants.

Out of 27,992 V16 applicants who list fewer than 12 courses, 24,666 have ranked at least one course that has both Type 1 and Type 2 versions (a dual course). 2,915 applicants have ranked at least one Type 2 course and all of these applicants have also ranked at least one dual course.

2.2 Skips and Payoff-Relevant Mistakes: Descriptive Statistics

If an applicant lists a Type 2 (fee) course, but does not list a corresponding Type 1 (subsidized) course, we say that that applicant “skips” Type 1 course. Even if an applicant skips Type 1 course, the skip may have no effect on the applicants’ assignment for two reasons. First, the applicant’s Score may be below the course cutoff. Second, the applicant may have been assigned to a more desirable course than the one skipped. When the skip leads to a change in applicant’s assignment, we say that the applicant makes a “payoff-relevant mistake”. We argue that most of the skips applicants make are not payoff-relevant mistakes.

As we do not know where in the applicant’s ROL a skipped course should be, we report a lower and upper bounds of payoff-relevant mistakes. The lower bound is calculated assuming that the
skipped course is the least desirable of any listed Type 1 course and the upper bound is calculated assuming that it is the most desirable.\(^6\)

In Table 1 we report the number of applicants that make at least one mistake of skipping a course and the number of applicants for whom skipping a course becomes a payoff-relevant mistake at least once, using both upper and lower bound definitions. The table suggests that for most applicants who skip, it is not a payoff-relevant mistake and that the total fraction of these with payoff-relevant mistakes is less than 1%.

<table>
<thead>
<tr>
<th>Payoff-relevant mistakes</th>
<th>Total</th>
<th>Listing Type 2</th>
<th>Skips</th>
<th>Upper bound</th>
<th>Lower bound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>27,922</td>
<td>2,963</td>
<td>1,009</td>
<td>201</td>
<td>14</td>
</tr>
<tr>
<td>% of V16 &lt; 12</td>
<td>100.00</td>
<td>10.61</td>
<td>3.61</td>
<td>0.72</td>
<td>0.05</td>
</tr>
<tr>
<td>% of Listing Type 2</td>
<td>100.00</td>
<td>34.05</td>
<td>6.78</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>% of Skippers</td>
<td>100.00</td>
<td>19.92</td>
<td>1.39</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## 2.3 Correlation between Applicant’s Score and Skips

Suppose that an applicant expects that course \(c\)’s cutoff will be above the applicant’s \(Score\); hence, the applicant does not expect to be assigned to course \(c\) even if \(c\) is listed in the applicant’s ROL. We call such a course infeasible for the applicant.

Our hypothesis (denoted E0) is that (i) applicants may omit infeasible courses and (ii) there is no systematic pattern in omitting infeasible courses; in particular, the likelihood of omitting an infeasible course is independent of the rank of this course in the true preferences of the applicant. (E0) implies a negative correlation between the probability that an applicant skips a course and applicant’s \(Score\). The lower the \(Score\) of an applicant, the larger the set of infeasible courses. As an applicant omits more infeasible courses, she is more likely to omit a course that has Type 2 counterpart in her ROL. That is, she is more likely to be identified as making a skip.

To investigate this relationship, we consider the following empirical model:

\[
Skip_i = \alpha + \beta Score_i + \gamma X_i^T + \epsilon_i, \tag{1}
\]

where \(Skip_i = 1\) if applicant \(i\) has made at least one skip and zero otherwise. We expect \(\beta\) to be negative.

There are several alternative explanations for the negative relation between \(Score_i\) and \(Skip_i\):

(E1) We explain a negative correlation between \(Skip\) and \(Score\) by a negative relation between the \(Score\) and the number of infeasible courses. However, if the number of acceptable Type 2

\(^6\)See appendix for details on the calculations of cutoffs and bounds as well for alternative definitions and the discussion of robustness.
courses decreased with Score, it would also lead to a negative correlation between Skip and Score. Indeed, suppose that Type 1 courses are omitted at random, irrespective of their infeasibility. This alone implies no correlation between Skip and Score. Suppose, in addition to the random omission, applicants with low scores have larger sets of acceptable Type 2 courses than the applicants with high scores. Then applicants with low scores are more likely to omit Type 1 courses which have the corresponding Type 2 courses in their ROLs, leading to a negative correlation between Skip and Score.

(E2) Applicants’ Score is correlated with cognitive abilities. Applicants with higher cognitive abilities are able to comprehend the mechanism better, thus reducing the probability that they skip.

(E3) Skipping a course is an instance of an applicant’s misguided attempt to gain a better assignment from the mechanism. Specifically, applicants drop courses that are (nearly) infeasible from the top of their ROL, so that their feasible courses rank higher. The important difference between (E3) and (E0) is where the skipped courses are located: (E3) requires that such courses are concentrated at the top of ROL, while (E0) does not impose any such restrictions: infeasible courses can be anywhere in the ROL.

We report the results for model 1 in Table 2. All odd-numbered regressions include V16 applicants who rank fewer than 12 courses and all even-numbered regressions exclude applicants who do not rank a Type 2 course. Control variables $X_i$ include applicants’ gender, median income (in logarithm) of the postal code in which the applicant resides, citizenship status, region born, language spoken at home and school fixed effects.

Regressions (1) and (2) are baseline regressions which establish the negative relation between Skip and Score. To control for explanation (E1), these regressions include eleven dummy variables corresponding to the number of Type 2 courses in ROL, from 1 to 11. The results without these dummies are similar and not reported.\(^7\)

To account for (E2), we include the results of General Achievement Test (GAT). GAT is a test of general knowledge and skills in written communication, mathematics, science and technology, humanities, the arts and social sciences. It is taken by V16 applicants several month before they sit their final high school exams. It does not contribute to applicant’s Score and does not affect the course admission decisions. As such, GAT result is a proxy of applicant’s cognitive ability, which is independent from Score. The results are reported in columns (3) and (4) of Table 2. The coefficient on GAT is negative and significant, suggesting that (E2) is valid and the cognitive abilities may be a part of the explanation of mistakes. Yet, after controlling for academic ability, Score continues be negatively related to Skip, supporting (E0).

Regressions (5) and (6) in Table 2 control for school fees. Victoria has a significant private school system. These schools have both well-resourced career advising services and disproportionately many applicants with higher scores. Thus, we include an interaction of applicant’s Score and an

\(^7\)It is also worth noting that the assumption that applicants with higher scores rank fewer Type 2 courses does not hold true in the data.
indicator that the applicant attends school that charges more than AUD11,000 (approx. USD8,000) in fees. As we see, applicants from private schools respond to change in their Score more, but the results for the overall population remain the same.

Table 2: Probability of Skipping a Type-1 Course

<table>
<thead>
<tr>
<th></th>
<th>Baseline (1)</th>
<th>(E1) and (E2) (3)</th>
<th>School fees (5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>-0.06***</td>
<td>-0.71***</td>
<td>-0.04***</td>
<td>-0.55***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.06)</td>
<td>(0.01)</td>
<td>(0.08)</td>
</tr>
<tr>
<td>GAT</td>
<td></td>
<td>-0.05***</td>
<td>-0.35***</td>
<td>-0.04***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(0.12)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>School fees × Score</td>
<td></td>
<td></td>
<td>-0.03***</td>
<td>-0.05**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>(N)</td>
<td>26325</td>
<td>2766</td>
<td>26325</td>
<td>2766</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.37</td>
<td>0.29</td>
<td>0.37</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.36</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Notes: Columns (1), (3), and (5) include all applicants, while (2), (4), and (6) focus on applicants listing at least one Type-2 course. Standard errors clustered at high school level are in parentheses. * \(p<0.10\), ** \(p<0.05\), *** \(p<0.01\).

To address (E3), we exploit the difference in predictions of the location of skipped courses in applicant’s ROL. As mentioned, (E3) predicts that courses will be skipped from the top of ROL, while our hypothesis does not place any such restriction.

Consider two applicants, who are identical except that \(i\) skips and \(j\) does not.

1. Suppose that (E3) holds. As \(i\) drops high-cutoff courses from the top of \(i\)’s ROL and keeps the bottom of the list the same, we expect that the cutoffs of top-ranked courses in \(i\)’s ROL will be lower than the cutoffs of top-ranked courses in \(j\)’s ROL. The cutoffs for bottom-ranked courses will be the same for both \(i\) and \(j\).

2. Suppose that (E3) does not hold and \(i\) skips courses from anywhere in \(i\)’s ROL. Then both the cutoffs for both top-ranked and bottom-ranked courses in \(i\)’s list will be lower than these in \(j\)’s list.

One possible test of (E3) is to compare the cutoffs of bottom-ranked courses of \(i\) and \(j\) and reject (E3) if \(i\)’s cutoffs are below \(j\)’s.\(^8\) However, such a test relies critically on skipping behavior explained exclusively by (E3). Suppose that (E3) holds for the majority of applicants, while the rest of the skips are explained by (E0). As (E0) holds for some applicants, we would observe that applicants who skip have lower cutoffs of bottom-ranked courses, compared to applicants who do not skip, as in bullet point (2) above. Thus, even a small fraction of applicants who behave according to (E0) would make us reject (E3).

\(^8\)Although this test is not reported, it rejects (E3): the cutoffs of bottom-ranked courses are lower for the applicants who skip.
To account for the possibility of both (E0) and (E3) being correct, we consider the difference between cutoffs of top-ranked and bottom-ranked courses. (E3) predicts that the difference is smaller for those who skip, compared to those who do not; (E0) predicts no difference.

We use the following empirical model:

\[
Cutoffs_{\text{top-ranked courses}}_i - Cutoffs_{\text{bottom-ranked courses}}_i = \gamma + \delta \text{Skip}_i + \zeta \text{Score}_i + \eta \text{X}_i^T + \epsilon_i. \tag{2}
\]

We expect the coefficient \(\delta\) to be negative if (E3) holds. The results are presented in Table (3). We use three different definitions for \(Cutoffs_{\text{top-ranked courses}}_i\) and \(Cutoffs_{\text{bottom-ranked courses}}_i\): in regressions (1), (4) and (7), we take the difference between the cutoffs of the top-ranked course and the bottom-ranked course for an individual applicant; in regressions (2), (5) and (8), we take the difference between the average of the two highest-ranked courses and the average of two lowest-ranked courses; and in (3), (6) and (9), we do the same with three courses. In regressions (1)–(6), we restrict our attention to the applicants who list at least one Type 2 course and in (7)–(9), we use the full sample of V16 applicants. Finally, in (1)–(3) we only control on \(\text{Score}\) and Gender, while in (4)–(9) we control for the length of ROL and use our standard battery of controls. We see that the coefficient on Skip is not significant in any of the regressions: there is no evidence that applicants eliminate high-cutoff courses from the top of their ROL. Thus, there is no evidence that applicants manipulate in the way described in the literature.

Table 3: Effects of Skips on the Difference between Cutoffs of Top- and Bottom-Ranked Courses

<table>
<thead>
<tr>
<th>Applicants Listing Type 2 Course</th>
<th>All applicants</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Skip</td>
<td>-1.38</td>
</tr>
<tr>
<td></td>
<td>(1.11)</td>
</tr>
<tr>
<td>Score</td>
<td>0.08**</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>Female</td>
<td>-2.70**</td>
</tr>
<tr>
<td></td>
<td>(1.20)</td>
</tr>
<tr>
<td>ROL Length</td>
<td>1.17***</td>
</tr>
<tr>
<td></td>
<td>(0.21)</td>
</tr>
<tr>
<td>Other Controls</td>
<td>No</td>
</tr>
<tr>
<td>N</td>
<td>2825</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Notes: The dependent variable of all regressions is the difference between the cutoffs of top- and bottom-ranked courses, but it varies in terms of the number of course we consider. Columns (1), (4), and (7) use top- and bottom-ranked courses; columns (2), (5) and (7) consider top two and bottom two; and columns (3), (6), and (9) take into account top three and bottom three courses. Standard errors clustered at high school level are in parentheses. * \(p < 0.10\), ** \(p < 0.05\), *** \(p < 0.01\).
2.4 Payoff-relevant Mistakes

Our hypothesis that applicants skip courses that they deem infeasible for them does not explain payoff-relevant mistakes. That is, unlike skips, which we expect to systematically vary with Score, payoff-relevant mistakes should be independent of Score. In this section we investigate the characteristics of those who make payoff-relevant mistakes. We expect that these mistakes are random and none of the coefficient are significant. Due to the sample size, we use the upper bound definition of payoff-relevance: applicants skip Type 1 course and are eligible for that course.

Our empirical model is

\[ Payoff-relevant\ Mistake_i = \theta + \iota Score_i + \kappa GAT_i + \lambda X_{it}^T + \epsilon_i \]  

(3)

In Table 4 we present the results. Note that the two variables that were significant in regressions for skips, GAT and interaction of Score and School fees, are no longer significant. Combining the results for regressions (1) and (3), reported in Tables (2) and (4) respectively, it appears as if higher-ability applicants are able to avoid both payoff-relevant and non-payoff-relevant mistakes, while lower-ability applicants only focus on payoff-relevant mistakes and successfully eliminate them. Similarly, the effect of attending expensive private school largely disappears from the regressions. Another notable observation is that Score now has a positive and significant effect. The effect may be explained mechanically: applicants with higher Score are eligible for more courses, hence skipping a Type 1 course is more likely to be payoff-relevant.

Table 4: Probability of Payoff-Relevant Mistakes

<table>
<thead>
<tr>
<th></th>
<th>V16 Ranking &lt;12 Courses (1)</th>
<th>Listing Type 2 Courses (3)</th>
<th>With Skips (5)</th>
<th>With Skips (6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Score</strong></td>
<td>0.01*** (0.00)</td>
<td>0.16*** (0.04)</td>
<td>0.66*** (0.15)</td>
<td>0.61*** (0.15)</td>
</tr>
<tr>
<td><strong>GAT</strong></td>
<td>0.00 (0.01)</td>
<td>0.00 (0.06)</td>
<td>0.31 (0.19)</td>
<td>0.28 (0.19)</td>
</tr>
<tr>
<td>School fees × Score</td>
<td>0.01 (0.01)</td>
<td>-0.11* (0.06)</td>
<td>0.22 (0.06)</td>
<td>0.22 (0.06)</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td>26325</td>
<td>2766</td>
<td>947</td>
<td>947</td>
</tr>
<tr>
<td><strong>R^2</strong></td>
<td>0.14</td>
<td>0.25</td>
<td>0.48</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Notes: Standard errors clustered at high school level are in parentheses. * p < 0.10, ** p < 0.05, *** p < 0.01.

2.5 Changes in ROL Over Time

To further test our hypothesis that skips are the outcomes of skipping infeasible courses at random, we use an unusual feature of the Victorian centralized mechanism: a requirement that applicants submit their “preliminary” ROL several month before their final ROL will be submitted and before
applicants learn their Score. If not changed, the preliminary ROL becomes final and used for the allocation. As a small effort is needed to change the ROL, we treat preliminary ROL as the best estimate of a final ROL that an applicant would submit, given the information the applicant has at the time.

There are two types of information that an applicant may obtain between submissions of the preliminary and the final ROLs. First, the applicant learns about the courses and about the mechanism. Better understanding of the mechanism may lead to decrease in the number of mistakes (both skips and payoff-relevant ones). Second, the applicant learns his or her Score. As the applicant learns Score, it becomes more clear to him or her which courses are infeasible. That may increase the number of skips, but will not affect the number of payoff-relevant mistakes. With two effects combined, we may see either an increase or a decrease in skips from the preliminary to the final ROL, but we must see the decrease in mistakes. We test this conjecture using two empirical models:

\[
\Delta(\text{Skips}_i) = \tau^s + \nu^s \mathbf{X}_i^T + \epsilon_i \quad (4)
\]

\[
\Delta(\text{Payoff-relevant Mistakes}_i) = \tau^m + \nu^m \mathbf{X}_i^T + \epsilon_i, \quad (5)
\]

where \(\Delta(\text{Skips}_i)\) is the differences between the number of skips in the final and the preliminary ROLs and \(\Delta(\text{Payoff-relevant Mistakes}_i)\) is the analogous difference in payoff-relevant mistakes. According to our hypothesis, the constant \(\tau^s\) in regression 4 could be either positive or negative, but the constant \(\tau^m\) in regression 5 must be negative.

In Table 5 we present the results. Odd-numbered regression are for payoff-relevant mistakes and even-numbered regressions are for skips. Regressions (1) and (2) only control for gender and income. As the number of added Type 2 courses may have a mechanical effect on skips and mistakes, we add these controls in regressions (3) and (4). In all regressions for payoff-relevant mistakes (1 and 3), the constant is negative and significant, implying that the number of payoff-relevant mistakes decreases, as we predict. In contrast, the effect of the revision of ROL on skips (regressions 2 and 4) is positive. Note that our hypothesis does not make any prediction on skips and we report these regression only for the comparison to the regressions for payoff-relevant mistakes.

Finally, we investigate the response of applicants to an unexpectedly high or low Score. Our main independent variable is a Shock, which is defined as the difference between realized and expected Score, where the expected Score is calculated using GAT (see appendix for the definition of expected Score). We then investigate the change in the number of instances of skips and payoff-relevant mistakes using two empirical models:

\[
\Delta(\text{Skips}_i) = \phi^s + \chi^s \text{Shock}_i + \psi^m \mathbf{X}_i^T \quad (6)
\]

\[
\Delta(\text{Payoff relevant Mistakes}_i) = \phi^m + \chi^m \text{Shock}_i + \psi^m \mathbf{X}_i^T. \quad (7)
\]
Table 5: Skips and Payoff-Relevant Mistakes: Changes over Time

<table>
<thead>
<tr>
<th></th>
<th>Payoff-relevant mistakes</th>
<th>Skips</th>
<th>Payoff-relevant mistakes</th>
<th>Skips</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.25**</td>
<td>0.54*</td>
<td>-0.25***</td>
<td>0.55**</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.29)</td>
<td>(0.09)</td>
<td>(0.23)</td>
</tr>
<tr>
<td>Change in # Type 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8.32***</td>
<td>43.55***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.22)</td>
<td>(2.76)</td>
</tr>
<tr>
<td>N</td>
<td>27676</td>
<td>27676</td>
<td>27676</td>
<td>27676</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.02</td>
<td>0.04</td>
<td>0.13</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Notes: Standard errors clustered at high school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We report results in Table 6. Regressions (1) and (3) are for payoff-relevant mistakes, and regressions (2) and (4) are for skips. $X^T$ are gender and income in regressions (1) and (2), with added control for the change in the number of Type 2 courses. The results do not change if we include a more complete list of controls. Note that, mechanically, if applicants do not change their ROLs, a positive shock increase the probability that a skip becomes a payoff-relevant mistake. The table shows that the number of payoff-relevant mistakes decreases with a positive shock, implying that the skips which could potentially become payoff-relevant mistakes are eliminated by the applicants. At the same, there is no significant effect on the number of skips, suggesting that applicants focus on the payoff-relevant part of their ROL following a shock.

Table 6: Effects of Shocks to Applicants Scores on Payoff-Relevant Mistakes and Skips

<table>
<thead>
<tr>
<th></th>
<th>Mistakes (1)</th>
<th>Skips (2)</th>
<th>Mistakes (3)</th>
<th>Skips (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shock to Score</td>
<td>-0.02***</td>
<td>-0.02</td>
<td>-0.01**</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Change in # Type 2</td>
<td></td>
<td></td>
<td>8.32***</td>
<td>43.55***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.21)</td>
<td>(2.76)</td>
</tr>
<tr>
<td>N</td>
<td>27637</td>
<td>27637</td>
<td>27637</td>
<td>27637</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.02</td>
<td>0.04</td>
<td>0.13</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Notes: “Mistake” means payoff-relevant mistake. Standard errors clustered at high school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. 
2.6 Enrollment Decisions

We conclude our empirical analysis by investigating the consequences of skips and payoff-relevant mistakes on the applicants. In Table 7 we report the results for three empirical models:

\[ \text{Enroll}_i = \omega^{em} \text{Payoff-relevant Mistake}_i + \omega^{es} \text{Skip}_i + \beta X^T_i + \epsilon_i \]  
\[ \text{Deferr}_i = \omega^{dm} \text{Payoff-relevant Mistake}_i + \omega^{ds} \text{Skip}_i + \beta X^T_i + \epsilon_i \]  
\[ \text{Reject}_i = \omega^{rm} \text{Payoff-relevant Mistake}_i + \omega^{rs} \text{Skip}_i + \beta X^T_i + \epsilon_i \]

We observe that making a payoff relevant mistake significantly decreases the probability of enrolling into the course and significantly increases the probability of deferring. Skips have similar, but much smaller, effect.

<table>
<thead>
<tr>
<th></th>
<th>Enroll (1)</th>
<th>Deferr (2)</th>
<th>Reject (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payoff-relevant mistake</td>
<td>-15.75***</td>
<td>13.06***</td>
<td>2.69</td>
</tr>
<tr>
<td></td>
<td>(3.81)</td>
<td>(3.34)</td>
<td>(2.94)</td>
</tr>
<tr>
<td>Skip</td>
<td>-4.07**</td>
<td>4.76***</td>
<td>-0.69</td>
</tr>
<tr>
<td></td>
<td>(1.99)</td>
<td>(1.57)</td>
<td>(1.73)</td>
</tr>
<tr>
<td>Score</td>
<td>0.38***</td>
<td>0.12***</td>
<td>-0.51***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>N</td>
<td>23774</td>
<td>23774</td>
<td>23774</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.17</td>
<td>0.13</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Notes: Standard errors clustered at high school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3 Theoretical Implications of Strategic Mistakes

The preceding analysis suggests that applicants tend to make mistakes but only a small percentage of them are payoff relevant. In this section, we explore the implications of these findings for identification methods that are commonly employed in the empirical studies of school assignment. Specifically we consider a large matching market operated by the Gale and Shapley’s deferred acceptance algorithm (Gale and Shapley (1962)), and adopt an equilibrium concept that permits participants to make mistakes as long as they become virtually payoff-irrelevant as the market size grows arbitrarily large.\(^9\)

3.1 Primitives

We begin with Azevedo and Leshno (2016) (in short AL) as our modeling benchmark. A (generic) economy consists of a finite set of courses $C = \{c_1, ..., c_C\}$ and a set of students. Each student has a type $\theta = (u, s)$, where $u = (u_1, ..., u_C) \in \left[u, \bar{u}\right]^C$ is a vector of von-Neumann Morgenstern utilities of attending courses for some $u \leq 0 < \bar{u}$, and $s = (s_1, ..., s_C) \in [0, 1]^C$ is a vector of scores representing the courses’ preferences or students’ priorities at courses, with a student with a higher score having a higher priority at a course. A vector $u$ induces an ordinal preferences over courses, denoted by a rank-ordered list (ROL) of “acceptable” courses, $\rho(u)$, of length $0 \leq \ell \leq C$. Assuming that a student has an outside option of zero payoff, the model allows for the possibility that students may find some courses unacceptable. Let $\Theta = \left[u, \bar{u}\right]^C \times [0, 1]^C$ denote the set of student types. One special case is serial dictatorship in which colleges’ preferences for students are given by a single score. Australian college admissions can be seen as a case of serial dictatorship. We shall incorporate this case by an additional restriction that the scores of each student satisfy $s_1 = ... = s_C$.

A continuum economy consists of the same finite set of courses and a unit mass of students with type $\theta \in \Theta$ and is given by $E = (\eta, S)$, where $\eta$ is a probability measure over $\Theta$ representing the distribution of student population over types, and masses of seats $S = (S_1, ..., S_C)$ available at the courses, where $S_i > 0$ and $\sum_{i=1}^C S_i < 1$. We assume that $\eta$ admits continuous density which 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 number in $[0, 1]$. The altomlessness ensures that indifferences either in student preferences or in course preference arises only for a measure 0 set of students. The full-support assumption means that both students and courses’ preferences are rich (except for the case of serial dictatorship). A matching is defined as a mapping $\mu : C \cup \Theta \rightarrow 2^\Theta \cup (C \cup \Theta)$ satisfying the usual two-sidedness and consistency requirements as well as “open on the right” defined in AL (see p. 1241). A stable matching is also defined in the usual way satisfying individual rationality and no-blocking.

According to AL, a stable matching is characterized via market-clearing cutoffs, $P = (P_1, ..., P_C) \in [0, 1]^C$, satisfying demand-supply condition: $D_i(P) \leq S_i$, with equality in case of $P_i > 0$, for each $i \in C$, where the demand $D_i(P)$ for course $i$ is given by the measure of students whose favorite course among all feasible ones (i.e., with cutoffs less than his scores) is $i$. Specifically, given the market-clearing cutoffs $P$, the associated stable matching assigns those who demand $i$ at $P$ to course $i$. Given the full-support assumption, Theorem 1-i of AL guarantees a unique market clear-

---

Liu and Pycia (2011), Azevedo and Leshno (2016), Azevedo and Hatfield (2012) and Che, Kim, and Kojima (2013). The current work differs largely from these papers because of the solution concept that we adopt and the issue we focus on here.

---

10In 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.

11Individual rationally requires that no participant (a student or a course) is assigned a partner that is not acceptable. No blocking means that no student-course pair exist such that the student prefers the course over her assignment and the course has either a vacant position or admits a student it ranks below that student.
ing price $P^*$ and a unique stable matching $\mu^*$. 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, which consists of $k$ students each with type $\theta$ drawn independently according to $\eta$, 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). A matching is defined in the usual way.

Consider a sequence of $k$-random economies $\{F^k\}$. We consider a student proposing DA being employed to assign students to courses. We assume that courses are acting passively reporting their preferences and capacities truthfully. We are interested in characterizing an equilibrium behavior of the students in the DA. In one sense, this is trivial: since DA is strategic-proof, it is a weak dominant strategy for each player to rank order all acceptable courses (i.e., with payoff $u_c \geq 0$) according to true preference order. We call such a strategy truthful-reporting strategy (TRS). We are, however, interested in a more robust solution concept allowing for any approximately-optimal behavior. We assume that each student observes his own type $\theta$ but not the types of other students; as usual, all students understand as common knowledge the structure of the game. Given this, 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 his type $\theta$.

In any game (either the limit or the $k$-random economy), student $i$’s Bayesian strategy is a measurable function $\sigma_i : \Theta \rightarrow \Delta(\mathcal{R})$, where $\mathcal{R}$ is the set of all possible ROLs a student can submit. Note that the strategies can be asymmetric; i.e., we do not restrict attention to symmetric equilibria. Specifically, we are interested in the following solution concept:

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

This solution concept is arguably more sensible than the exact Bayesian Nash equilibrium, if for a variety of reasons market participants 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. We shall identify a large class of behavior as being consistent with robust equilibrium. To this end, it is useful to define some terminologies. Fix a DA matching (in either finite or continuum economy) with equilibrium cutoffs $P \in [0, 1]^C$. We say a course $c$ is feasible to a student if his score at $c$ is no less than $P_c$. And we say a student demands course $c$ if $c$ is feasible and he ranks $c$ in his ROL ahead of any other feasible courses.

\footnote{In the specific context of VTAC mechanism, the common preferences make this an ex post equilibrium strategy. See Che and Koh (2016).}
3.2 Analysis of Robust Equilibria

Now recall $P^*$ (the unique market clearing cutoffs for the limit continuum economy). We define stable-response strategy (SRS) to be any strategy that demands his most preferred feasible course given $P^*$ (i.e., he ranks those courses ahead of all other feasible courses). The set of SRSs is typically large. For example, suppose $C = \{1, 2, 3\}$, and courses 2 and 3 are feasible, and a student prefers 2 to 3. Then, 7 ROLs—1-2-3, 2-3-1, 1-2-3, 2-1-2, 1-2-3, 2—constitute his SRSs out of 10 possible ROLs he can choose from. Formally, if a student has $\ell \leq C$ feasible courses, then the number of SRSs is $\sum_{a, b}^{a \leq \ell-1, b \leq C-\ell} (a+b+1)a!b!$. For each type $\theta = (u, s)$ with $\rho(u) \neq \emptyset$ (i.e., with at least one acceptable course), there exists at least one SRS that is untruthful.\(^{13}\)

For the next result, we construct such a strategy. To begin, let $\hat{r} : R \times [0, 1]^C \to R$ be a transformation function that maps a preference order $\rho \in R$ to an ROL with the property that:

(i) $\hat{r}(\rho, s) \neq \rho$ for all $\rho \neq \emptyset$ (i.e., untruthful), and $\hat{r}(\emptyset, s) = \emptyset$, and (ii) $\hat{r}(\rho, s)$ ranks the most preferred feasible course ahead of all other feasible courses for each $\rho \neq \emptyset$ (where feasibility is defined given $P^*$). The existence of such a strategy is established above. We then define an SRS $\hat{R} : \Theta \to R$, given by $\hat{R}(u, s) := \hat{r}(\rho(u), s)$, for all $\theta = (u, s)$.\(^{14}\)

Let

$$\Theta^\delta := \{(u, s) \in \Theta | \exists i \text{ s.t. } |s_i - P^*_i| \leq \delta\}$$

be the set of types that have a score that is $\delta$-close to its market clearing cutoff for the continuum economy.

**Theorem 1.** Fix any arbitrarily small $(\delta, \gamma) \in (0, 1)^2$. It is a robust equilibrium for all students with types $\theta \in \Theta^\delta$ to play TRS, and for all students with types $\theta \notin \Theta^\delta$ to randomize between TRS with probability $\gamma$ and untruthful SRS $\hat{R}(\theta)$ with probability $1 - \gamma$ in each $k$-random economy.

Since $(\delta, \gamma)$ is arbitrary, we conclude:

**Corollary 1.** There exists a robust equilibrium in which each student submits an untruthful ROL with probability arbitrarily close to one conditional on submitting a non-empty ROL.

To the extent that a robust equilibrium is a reasonable solution concept, the result implies that we should not be surprised to observe a non-negligible fraction of market participants making “mistakes”—more precisely dominated strategies—even in a strategy-proof environment. It also calls into question any empirical method relying on TRS—any particular strategy for that matter—as an identifying restriction.

We next present a result that justifies the outcome-based method of estimation.

---

\(^{13}\)If a student’s most favorite course is infeasible, she can drop that course. If is feasible, then she can drop an acceptable course below or add an unacceptable course below, whichever exists.

\(^{14}\)Note this SRS is constructed via the transformation function $\hat{r}$. In principle, an SRS can be defined without such a transformation function, although this particular construction simplifies the proof below.
Definition 2. For a sequence \( \{F^k\} \) of \( k \)-random economies with DA matching, the associated sequence \( \{\sigma^k_{1\leq i \leq k}\}\) of strategy profiles is said to be an asymptotically stable if, for any \( \epsilon > 0 \), there exists \( K \in \mathbb{N} \) such that for \( k > K \), with probability at least of \( 1 - \epsilon \), at least a fraction \( 1 - \epsilon \) of all students are assigned their most preferred feasible courses given the equilibrium cutoffs \( P^k \).

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

Theorem 2. Any regular robust equilibrium is asymptotically stable.

4 Analysis with Monte Carlo Simulations

This section provides details on the Monte Carlo simulations that we perform to assess the implications of our theoretical results. Section 4.1 specifies the model, section 4.2 describes the data generating processes, section 4.3 presents the estimation and testing procedures, section 4.4 discusses the estimation results, and, finally, section 4.5 presents counterfactual analyses based on the estimators.

4.1 Model Specification

We consider an economy in which \( I = 400 \) students compete for admission to \( S = 12 \) schools. The vector of school capacities is specified as follows:

\[
\{q_s\}_{s=1}^{12} = \{20, 20, 20, 40, 20, 40, 20, 40, 20, 40, 20, 40\}.
\]

Setting the total capacity of schools (340 seats) to be strictly smaller than the number of students (400) ensures that each school has a strictly positive cutoff in equilibrium. Therefore, omitting any school from one’s application can be considered as a “mistake” that can potentially lead to suboptimal outcomes.

The economy is located in an area within a circle of radius 1 as in Figure B1 (Appendix B) which plots one simulation sample. The schools (represented by big red dots) are evenly located on a circle of radius 1/2 around the centroid; the students (represented by small blue dots) are uniformly distributed across the area. The cartesian distance between student \( i \) and school \( s \) is denoted by \( d_{i,s} \).

Students are matched with schools through a serial dictatorship. They are asked to submit a rank-ordered list of schools, and there is no limit on the number of choices to be ranked. Without loss of generality, schools have a priority structure such that student \( i \) is ranked higher by all schools than those with \( i' < i \). One may consider the order is determined by some exam scores that are private information at the time of submitting ROL.

To represent students’ preferences over schools, we adopt a parsimonious random utility model. Student \( i \)'s utility from being matched with school \( s \) is specified as follows:

\[
u_{i,s} = \beta_1 s - d_{i,s} + \beta_2 T_i A_s + \sigma_{i,s}, \forall i \text{ and } s; \tag{11}\]
where $\beta_1 \times s$ is school $s$’s quality; $d_{i,s}$ is the distance from student $i$’s residence to school $s$, with a coefficient normalized to $-1$; $T_i = 1$ or $0$ is student $i$’s type (e.g., disadvantaged or not, or arts versus sciences); $A_s = 1$ or $0$ is school $s$’s type (e.g., known for resources for disadvantaged students or art education); and $\epsilon_{i,s}$ is a type-I extreme value.

The type of school $s$, $A_s$, is $1$ if $s$ is an odd number; otherwise, $A_s = 0$. The type of student $i$, $T_i$, is $1$ with a probability $2/3$ among the lower ranked students ($i \leq 200$); $T_i = 0$ for all $i > 200$. This way, we may consider those with $T_i = 1$ as the disadvantaged.

The coefficients of interest are $(\beta_1, \beta_2, \sigma)$ which are fixed at $(0.2, 1, 1)$ in simulations. The purpose of estimation is to recover these coefficients and therefore the distribution of preferences.

### 4.2 Data Generating Processes

Each simulation sample is an independent student preference profile obtained by randomly drawing $\{d_{i,s}, \epsilon_{i,s}\}_s$ and $T_i$ for all $i$ from the distributions specified above. In all samples, student priorities, school capacities, and school types ($A_s$) are kept constant.

We first simulate the joint distribution of the $12$ schools’ admission cutoffs. The simulation lets every student submit an ROL ranking all schools truthfully. After running the serial dictatorship, we calculate the admission cutoffs in each simulation sample. Figure B2 in Appendix B shows the marginal distribution of each school’s cutoff from the 300 samples. Note that schools with smaller capacities tend to have higher cutoffs. For example, school 11, with 20 seats, in general has the highest cutoff, although the highest-quality school is school 12, with 40 seats.

To generate data on student behaviors and matching outcomes for preference estimation, we simulate another 300 samples with the same draws of $T_i$ and $\{d_{i,s}\}_s$ but new independent draws of $\{\epsilon_{i,s}\}_s$. These samples are used for the following estimation and counterfactual analysis, and, in each of them, we consider three types of data generating processes (DGPs) with different student strategies.

1. **STT (Strict Truth-Telling)**: Every student submits a rank-ordered list of 12 schools according to her true preferences. Recall that everyone finds every school acceptable, and therefore STT is equivalent to TRS as in section 3.

2. **IRR (Payoff Irrelevant Skips)**: A fraction of students skip schools with which they are never matched given the simulated distribution of cutoffs. For a given student, a dropped school can have a high (expected) cutoff and thus be “out of reach;” it may also be a school that has a low cutoff, but the student is accepted by one of her more-preferred schools almost surely. To specify the fraction of skippers, we first randomly choose about 18 percent of the students to be never-skippers who always rank all schools truthfully. Students with $T_i = 1$ are more likely to skip: 97 percent of them are potential skippers (Table B1), compared with 75 percent of those with $T_i = 0$ (Table B2). Among the potential skippers, we consider three

\[15\] Results do not change when we allow for school fixed effects.
scenarios. In IRR 1, around one third of them skip all the “never-matched” schools; IRR 2 adds another one-third; and IRR 3 lets all of them skip. When doing so, it is possible that a student drops all schools, because some students are never matched. We randomly choose a school for them, so that they submit one-school ROLs.

(iii) REL (Payoff Relevant Mistakes): In addition to IRR3, we now let the potential skippers make payoff relevant mistakes. That is, they skip some of the schools that they have some chance of being matched with given the joint distribution of cutoffs. Recall that the joint distribution of cutoffs is only computed once under the assumption that everyone is strictly truth-telling. In each of the four DGPs, REL 1-4, we specify a threshold matching probability, and the potential skippers drop the schools with which they are matched with probabilities lower than the threshold. From REL 1 to REL 4, the thresholds are 5, 10, 15, and 20 percent. It should be emphasized that the cutoff distribution is not re-computed in any of these DGPs.

To summarize, for each of the 300 simulation samples, we simulate the matching game 8 times: 1 (STT or strict truth-telling) + 3 (IRR, or payoff-irrelevant skips) + 4 (REL, or payoff relevant mistakes). Tables 8 shows how students skip in the simulations. The reported percentages are averaged over the 300 samples.

4.3 Identifying Conditions and Estimation

With the simulated data at hand, the random utility model described by equation (11) is estimated under two different identifying conditions:

(i) WTT or Weakly Truth-Telling. WTT, which can be considered as a truncated version of TRS or STT, entails two assumptions: (a) the 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 one may not rank all schools). The ROLs specify a rank-ordered logit model that can be estimated by Maximum Likelihood Estimation (MLE). We refer to this approach as the WTT estimator.

(ii) Stability. Under the assumption that students are assigned their favorite feasible school given the ex post cutoffs, the model can be estimated by MLE based on a conditional logit model where each student’s choice set is restricted to the ex post feasible schools and where the matched school is the favorite among all her feasible schools. We call this approach the stability estimator.

The formulation of likelihood function and estimation results from both approaches are discussed in Appendix B. In Table 8, across the eight DGPs, the fraction of skippers increases from zero (in STT) to 82 percent in IRR 3 and remains at the same level in REL 1-4. The WTT assumption is completely satisfied only in STT, and the fraction of students who are weakly truth-telling
Table 8: Skips and Mistakes in Monte Carlo Simulations (Percentage Points)

<table>
<thead>
<tr>
<th>Scenarios (Data Generating Processes)</th>
<th>Strict Truth-telling</th>
<th>Payoff Irrelevant Skips</th>
<th>Payoff Relevant Mistakes</th>
</tr>
</thead>
<tbody>
<tr>
<td>WTT: Weak Truth-Telling(^a)</td>
<td>STT</td>
<td>IRR 1</td>
<td>IRR 2</td>
</tr>
<tr>
<td>Matched w/ favorite feasible school(^b)</td>
<td>100</td>
<td>90</td>
<td>76</td>
</tr>
<tr>
<td>Skippers(^c)</td>
<td></td>
<td>24</td>
<td>53</td>
</tr>
<tr>
<td>By number of skips:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>14</td>
<td>31</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>STT: Strict Truth-telling</td>
<td>100</td>
<td>76</td>
<td>47</td>
</tr>
<tr>
<td>Reject Truth-telling: Hausman Test(^d)</td>
<td>5</td>
<td>8</td>
<td>57</td>
</tr>
</tbody>
</table>

Notes: This table presents the configurations of the eight data generating processes (DGPs). Each entry is a percentage averaged over the 300 simulation samples. Tables B1 and B2 further show the breakdown by \(T_i\). \(^a\) A student is "weak truth-telling" if she truthfully ranks her top \(K_i\) (1 ≤ \(K_i\) ≤ 12) preferred schools. \(^b\) A school is feasible to a student, if the student’s index is higher than the school’s \textit{ex post} admission cutoff. If a student is matched with her favorite feasible school, she cannot form a blocking pair with any school. \(^c\) Given that every school is acceptable to all students and potentially over-demanded, a student is a skipper if she does not rank all schools. \(^d\) In each DGP, this reports the percentage of samples that the WTT (weakly truth-telling) assumption is rejected at 5% level in favor of the stability assumption. The test is based on the Durbin-Wu-Hausman test and discussed in details in section 4.3.

decreases from 90 percent in IRR 1 to 62 percent in IRR 3, stabilizing at 61 percent in all REL DGPs.

In contrast, stability is always satisfied in STT and IRR 1-3, while the fraction of students that can form a blocking pair with some school increases from 4 percent in REL 1 to 15 percent in REL 4.

To test WTT against stability, we construct a Durbin-Wu-Hausman-type test statistic from the estimates of the WTT and stability approaches. Under the null hypothesis, both WTT and stability are satisfied, while under the alternative only stability holds. When all students, except the 15 percent with the lowest priorities, are matched, weakly truth-telling implies stability, but not the reverse. Therefore, if WTT is satisfied, the estimator based on WTT is consistent and efficient, while the stability estimator is consistent but inefficient.

Table 8 shows both the size and the power of this test. When the null is true (e.g., in STT), it rejects the null at the desired rate, 5 percent. When the null is not true (in IRR 1-3), it rejects the null with an 8-100 percent probability. Notice that we only need 38 percent of students (as in IRR 3) violating the WTT condition to achieve the 100 percent rejection rate of the null hypothesis.

In REL 1-4, both WTT and stability are violated, while the latter is violated to a lesser extent.
The test is no longer valid, although it still rejects the null at high rates.

### 4.4 Estimation Results

The discussion below focuses on the estimates of the quality coefficient, $\beta_1$, and more details are provided in Appendix B, especially Table B3. Figure 1 plots the distributions of the estimator across the 300 simulation samples, with the true value being $\beta_1 = 0.2$. A consistent estimator should have mean equal to 0.2. Recall that all DGPs use the same 300 samples of simulated student preference profiles and that what defers across DGPs is how students play the game.

![Figure 1: Estimates based on Weak Truth-Telling or Stability (True Value is 0.2)](image)

Notes: The figures focus on the estimates of the quality coefficient ($\beta_1$) from two approaches, weakly truth-telling (WTT, red line) and stability (blue dots). The distributions of the estimates across the 300 simulation samples are reported. A consistent estimator should have mean equal to 0.2. Each subfigure uses the 300 estimates from the 300 simulation samples given a DGP and reports an estimated density based on a normal kernel function. Note that STT as a DGP means that every student truthfully rank all schools; IRR 1-3 only include payoff irrelevant skips, while REL 1-4 have payoff relevant mistakes. See Table 8 for more details on the eight DGPs.

The first noticeable pattern in the figure is that the variance of the estimator is always smaller when we use WTT. Intuitively, this is because weak truth-telling leads to more information being used for estimation.

Figure 1a presents the best-case scenario for the WTT condition in terms of estimation. That is, WTT is satisfied and we use the maximum possible information (i.e., the complete ordinal preferences). As expected, both WTT and stability lead to consistent estimators.

The results from the data containing payoff-irrelevant skips are summarized in Figure 1b–d. As expected, the stability estimates (blue dots) are invariant to payoff-irrelevant skips and stay the same as that those the STT case. In contrast, the WTT estimates (red lines) are sensitive to the fraction of skippers. Even when there are only 53 percent skippers and 24 percent of students...
violating the truth-telling assumption (REL 2), the estimates based on WTT from the 300 samples have mean 0.18 (standard deviation 0.02); in contrast, the stability estimates are on average 0.21 (standard deviation 0.05).

The downward bias in the WTT estimator is intuitive. When students skip, they omit schools with which they have almost no chance of being matched. For students with low priorities, popular schools are therefore more likely to be skipped. Whenever a school is skipped, WTT assumes that it is less preferable than all the ranked schools. Therefore, many students are mistakenly assumed to dislike popular schools, which results in a downward bias in the estimator of $\beta_1$. In contrast, this bias is absent in the stability estimator: Whenever a school is skipped by a student due to its high expected cutoff, the stability condition does not make this assumption on ranked and skipped schools.

Figures 1e–h consider the DGPs in which students make payoff-relevant mistakes. Neither WTT or stability is satisfied (Table 8), and therefore both estimators are inconsistent. However, the estimators based on stability are still less biased; the means of the estimates are close to the true value, ranging from 0.18 to 0.20 (see Table B3 for more details).

### 4.5 Counterfactual Analysis

Making policy recommendations based on counterfactual analysis is one of the main objectives of market design research. In the following, we illustrate how the estimation approaches lead to mis-predicted counterfactual outcomes, while the estimation based on stability yields the most robust results.

We consider the following counterfactual policy: Students with $T_i = 1$ are given priority over those with $T_i = 0$, while within each type they are still ranked according to their indices. That is, given $T_i = T_{i'}$, $i$ is ranked higher by all schools than $i'$ if and only if $i > i'$. One may consider this as an affirmative action policy if $T_i = 1$ indicates $i$ being disadvantaged. The matching mechanism is still the serial dictatorship in which everyone can rank all schools.

The effects of the counterfactual policy are evaluated by the following four approaches.

(i) **True Preferences:** We use the true coefficients in utility functions to simulate counterfactual outcomes. Students submit truthful 12-school ROLs. This outcome is taken as a benchmark.

(ii) **Submitted ROLs:** One assumes that the submitted ROLs are true ordinal preferences and that students submit the same ROLs even when the existing policy is replaced by the counterfactual.

(iii) **WTT Estimates:** One assumes that the submitted ROLs represent top preferred schools in true preference order, and therefore student preferences can be estimated from the data with WTT as the identifying condition. Under the counterfactual policy, we simulate student preferences based on the estimates and let students submit truthful 12-school lists.
Stability Estimates: We estimate student preferences from the data with stability as the identifying condition. Under the counterfactual policy, we simulate student preferences based on the estimates and let students submit truthful 12-school lists.

When simulating counterfactual outcomes, we use the same 300 simulated samples for estimation. In particular, we use the same simulated \( \{\epsilon_{i,s}\}_s \) when constructing preference profiles after preference estimation. Equivalently, we assume that \( \{\epsilon_{i,s}\}_s \) are observed and focus on the effects of different estimators of the coefficients in the utility functions.

To summarize, for each of the 300 simulation samples, we conduct 32 different counterfactual analyses: 8 (DGPs) \( \times \) 4 (counterfactual approaches: true preferences, submitted ROLs, truth-telling estimates, and stability estimates).

4.5.1 Performance of the Three Approaches

Using the counterfactual analysis with true preferences as our benchmark, we compare how the last three approaches perform from two perspectives, predicting the policy’s effect on matching outcomes and welfare.

When simulating the outcomes under the counterfactual policy, we assume students rank all schools truthfully when submitting ROLs. As shown above, matching outcome does not change if students make payoff-irrelevant skips, although payoff relevant mistakes would lead to different outcomes.

An informative statistic of a match is the schools’ cutoffs which summarize the joint distribution of student priorities and preferences. Figure 2 shows, given each DGP, how the three approaches mis-predict the cutoffs under the counterfactual policy. For each school, indexed from 1 to 12, we calculate the mean of the 300 cutoffs from the 300 simulation samples by using the true preferences and the other three approaches. The sub-figures then depict the mean differences between the predicted cutoffs and the true ones.

In Figure 2a, the DGP is STT, and thus the submitted ROLs coincide with true ordinal preferences. Consequently, the predicted cutoffs from this approach are the true ones. The WTT and stability estimates are almost the same.

In Figures 2b–d, corresponding to DGPs IRR 1-3, only the stability estimates are consistent, and indeed they have the smallest mis-prediction relative to the other two. Both of the estimates based on WTT and submitted ROLs have mis-predictions increasing from IRR 1 to IRR 3, and those based on submitted ROLs make larger errors. Since students tend to omit popular schools from their lists, both approaches underestimate the demand for these schools and thus result in under-predicted cutoffs. The bias is even larger for smaller schools (odd-numbered) because they tend to be skipped more often.

When the DGPs contain payoff-relevant mistakes (REL 1-4), none of the approaches is consistent (Figures 2e–h). However, the stability estimates seem to have the smallest mis-prediction relative to the other two.
Figure 2: Comparison of the Three Approaches: Biases in Predicted Cutoffs

Notes: The sub-figures presents how the predicted cutoffs from each approach differ from the true ones which are simulated based on true preferences. Each subfigure corresponds to a DGP. Given a DGP, we simulate the schools’ cutoffs following each approach and calculate the mean deviation from the true ones.

Figure 3: Comparison of the Three Approaches: Mis-predicted Match (Fractions)

Notes: The sub-figures show how each approach to counterfactual analysis mis-predicts matching outcomes under the counterfactual policy. Given a DGP, we simulate a matching outcome and compare them to the true one which is calculated with true preferences. The sub-figures present the average rates of mis-prediction for the two groups of students, $T_i = 1$ and $T_i = 0$.

Figure 3 further shows how each of the three approaches mis-predicts individual outcomes. Because the counterfactual policy is intended to help students with $T_i = 1$, we look at these two groups, $T_i$ equals to 1 or 0, separately.

In Figure 3a, among the $T_i = 1$ students, the stability estimates incorrectly predict the match
of 9 percent of them, whenever stability is satisfied (in DGPs TT and IRR 1-3). Among REL 1-4, the fraction of mis-prediction based on stability increases from 13 to 21 percent. The WTT estimates have a lower mis-prediction rate in TT and IRR 1, but under-perform in all other DGPs. Lastly, the estimates based on submitted ROLs have the highest mis-prediction rates in all DGPs except TT. Among those with $T_i = 0$ (subfigure b), the comparison of the three approaches has the same pattern.

(a) Students $T_i = 1$: (Fraction Better off) – (Fraction Worse off)     (b) Students $T_i = 0$: (Fraction Better off) – (Fraction Worse off)

Figure 4: Comparison of the Three Approaches: Mis-predicting Welfare Effects

Notes: The sub-figures show how each approach to counterfactual analysis mis-predicts the welfare effects of the counterfactual policy for the two groups of students, $T_i = 1$ and $T_i = 0$. Given a DGP, we simulate matching outcomes, calculate welfare effects, and compare them to the true one which is calculated with true preferences. Welfare effects are measured by the difference between the fraction of students better off and that of those worse off. The estimated fraction of students with $T_i = 1$ being worse off is close to zero in all cases, so is the estimated fraction of students with $T_i = 0$ being better off. There are some students whose welfare does not change; simulated with true preferences, this fraction is 16 percent among the $T_i = 1$ students and 41 percent among the $T_i = 0$ ones. See Tables B4 and B5 in Appendix B for more details.

We now investigate the welfare effects on the $T_i = 1$ students and others when the current policy is replaced by the counterfactual one. Given a simulation sample and a DGP, we compare the outcomes of each student under the two policies. If the student is matched with a “more-preferred” school according to the estimated preferences, she is better off; she is worse off if she is matched with a “less-preferred” one. Because each approach to counterfactual analysis estimates student preferences in a unique way, measured welfare effects may differ even when a student is matched with the same school.

Figure 4 shows the mean difference between the fraction of students better off and that of those worse off across the simulation samples.\textsuperscript{16} In Figure 4a, among the $T_i = 1$ students, the welfare effect predictions based on the stability estimates are almost identical to the true value, even in the cases with payoff-relevant mistakes. In contrast, the WTT estimates are close to the true value

\textsuperscript{16}There are some students whose outcomes do not change. See Tables B4 and B5 in Appendix B for more detailed summary statistics.
in DGPs STT and IRR 1; those based on submitted ROLs tend to be biased towards to zero effect when there are more students skipping or making mistakes.

The results for students with $T_i = 0$ are collected in Figure 4b. The general patterns remain the same, although the stability estimates are more biased than in Figure 4a.

In summary, estimated welfare effects are biased towards zero for all students when we assume WTT or take submitted ROLs as true preferences. The stability estimates, however, are very close to the true value; even when there are some payoff-relevant mistakes, the estimates are much less biased than the other two.
References


A Proofs from Section 3

Proof of Theorem 1. Recall the cardinal type \( \theta = (u, s) \) induces an ordinal preference type \( (\rho(u), s) \). Recall that the strategy \( \hat{R}(\theta) \) depends on \( u \) only through \( \rho(u) \), without loss we can work in terms of the “projected” ordinal type \( (\rho, s) \). Also recall that the mechanism depends only on a student’s ROL and her score. Hence, for the current proof, we shall abuse the notation and call \( \theta := (\rho, s) \) a student’s type, redefine the type space \( \Theta := \mathcal{R} \times [0, 1]^C \) (the projection of the original types), and let \( \eta \) be the measure of the projected types (which is induced by the original measure on \( (u, s) \)). The continuum economy \( E = [\eta, S] \) is redefined in this way. Likewise the \( k \)-random economies \( F^k = [\eta^k, S] \) are similarly redefined. Given this reformulation, it suffices to show that it is a robust equilibrium for each type \( \theta \in \Theta^\delta \) to adopt TRS and for each type \( \theta \not\in \Theta^\delta \) to randomize between TRS with probability \( \gamma \) and \( \hat{r}(\theta) \) with probability \( 1 - \gamma \).

We first make the following preliminary observations.

Claim 1. Given the continuum economy \( E = [\eta, S] \), let \( \hat{\eta} \) denote the measure of “reported” types when the students follow the prescribed strategies, and let \( \hat{E} = [\hat{\eta}, S] \) denote the “induced” continuum economy under that strategies. Then, \( \hat{E} \) has the unique stable matching identical to that under \( E \), characterized by the identical cutoffs \( P^* \). The demand under that economy \( \hat{D}(\cdot) \) is \( C^1 \) in the neighborhood of \( P^* \) and has \( \partial \hat{D}(P^*) = \partial D(P^*) \), which is invertible.

Proof. Let \( D^{(\rho,s)}(P^*) \) be the course a student with type \( \theta = (\rho, s) \) demands given cutoffs \( P^* \) (i.e., her most preferred feasible course given \( P^* \)). Since \( \hat{r} \) ranks her favorite feasible course ahead of all other feasible courses, it must be that \( D^{(\hat{r}(\rho,s),s)}(P^*) = D^{(\rho,s)}(P^*) \) for each type \( (u, s) \). It then follows that \( \hat{D}(P^*) = D(P^*) \), where \( \hat{D}(P) \) is the demand at the continuum economy \( \hat{E} = [\hat{\eta}, S] \). Hence, \( P^* \) also characterizes a stable matching in \( \hat{E} \). Further, since \( \eta \) has full support and since the prescribed strategy has every type \( \theta \) play TRS with positive probability, the induced measure \( \hat{\eta} \) must also have full support.\(^{17}\) By Theorem 1-i of AL, then the cutoffs \( P^* \) characterize a unique stable matching at economy \( \hat{E} \). Finally, observe that, for any \( P \) with \( ||P - P^*|| < \delta \), each type \( \theta \not\in \Theta^\delta \) has the same set of feasible courses when the cutoffs are \( P \) as when they are \( P^* \). This means that for any such type \( (\rho, s) \) and for any \( P \) in the set, \( D^{(\hat{r}(\rho,s),s)}(P) = D^{(\rho,s)}(P) \). The last statement thus follows.

Claim 2. Let \( \hat{P}^k \) be the (random) cutoffs characterizing the DA assignment in the \( F^k \) when the students follow the prescribed strategies. Then, for any \( \delta, \epsilon > 0 \), there exists \( K \in \mathbb{N} \) such that for all \( k > K \),

\[
\Pr\{|||\hat{P}^k - P^*|| < \delta\} \geq 1 - \epsilon.
\]

Proof. Let \( \hat{\eta}^k \) be the measure of “stated” types \( (\hat{r}(\rho), s) \) under \( k \)-random economy \( F^k \) when the students follow the prescribed strategies. Let \( \hat{F}^k = [\hat{\eta}^k, S] \) be the resulting “induced” \( k \)-random

\(^{17}\)Throughout, we implicitly assume that a law of large numbers applies. This is justified by focusing on an appropriate probability space as in Sun (2006). Or more easily, we can assume that the students are coordinating via asymmetric strategies so that exactly \( \gamma \) fraction of each type plays TRS.
economy. By construction, \( \hat{F}^k \) consists of \( k \) independent draws of students according to measure \( \hat{\eta} \), so it is simply a \( k \)-random economy of \( \hat{E} \). Since by Claim 1, \( \hat{\eta} \) has full support, \( \hat{D}(\cdot) \) is \( C^1 \) in the neighborhood of \( P^* \) and \( \partial \hat{D}(P^*) \) is invertible, by Proposition 3-2 of AL, for each \( \epsilon' > 0 \), there exists \( K \in \mathbb{N} \) such that for all \( k > K \), cutoffs \( \hat{P}^k \) of any stable matching of \( \hat{E}^k \)—and hence the DA outcome of \( F^k \) under the prescribed strategies—satisfy

\[
\Pr\{||\hat{P}^k - P^*|| < \delta\} \geq 1 - \epsilon'.
\]

We are now in a position to prove Theorem 1. Fix any \( \epsilon > 0 \). Take any \( \epsilon' > 0 \) such that \( \epsilon'(\bar{u} - \underline{u}) \leq \epsilon \). By Claim 2, there exists \( K \in \mathbb{N} \) such that for all \( k > K \), \( \Pr\{||\hat{P}^k - P^*|| < \delta\} \geq 1 - \epsilon' \), where \( \hat{P}^k \) are the cutoffs associated with the DA matching in \( F^k \) under the prescribed strategies. Let \( \mathcal{E}^k \) denote this event. We now show that the prescribed strategy profile forms an interim \( \epsilon \)-Bayesian Nash equilibrium for each \( k \)-random economy for \( k > K \).

First, for any type \( \theta \in \Theta^k \) the prescribed strategy, namely TRS, is trivially optimal given the strategyproofness of DA. Hence, consider a student with any type \( \theta \notin \Theta^k \), and suppose that all other students employ the prescribed strategies. Now condition on event \( \mathcal{E}^k \). Recall that the set of feasible courses is the same for type \( \theta \in \Theta^k \) when the cutoffs are \( \hat{P}^k \) as when they are \( P^* \), provided that \( ||\hat{P}^k - P^*|| < \delta \). Hence, given event \( \mathcal{E}^k \), strategy \( \hat{r}(\theta) \) is a best response—and hence the prescribed mixed strategy—attains the maximum payoff for type \( \theta \notin \Theta^k \).

Of course, the event \( \mathcal{E}^k \) may not occur, but its probability is no greater than \( \epsilon' \) for \( k > K \), and the maximum payoff loss in that case from failing to play her best response is \( \bar{u} - \underline{u} \). Hence, the payoff loss she incurs by playing the prescribed mixed strategy is at most

\[
\epsilon'(\bar{u} - \underline{u}) < \epsilon.
\]

This proves that the sequence of strategy profiles for the sequence \( \{F^k\} \) of \( k \)-random economies forms a robust equilibrium.

Proof of Theorem 2. 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 \( i \), and (random) per capita demand

\[
D^k(P) := \left( \frac{1}{k} \sum_{i=1}^{k} I \left\{ c \in \arg \max_{c' \text{ w.r.t. } R_i^k} \left\{ \{ c' \in C : s_{i,c'} \geq P_{c'} \} \right\} \right\} \right)_{c \in C},
\]

where the set \( \{ c' \in C : s_{i,c'} \geq P_{c'} \} \) is the set of feasible courses for student \( i \) with respect to the cutoff \( P \) and \( I\{\cdot\} \) is an indicator function equal to 1 if \( \{\cdot\} \) holds and 0 otherwise. (Note that the random ROLs \( R_i \)'s are suppressed as arguments of \( D^k(P) \) for notational ease.) Let \( P^k \) be the (random) cutoffs, satisfying \( D^k(P^k) = S^k \).

Let \( \hat{D}^k(P) := \mathbb{E}_{(R_1,\ldots,R_k)}[D^k(P)] \), where the randomness is taken over the random draws of the types of the students and the random reported ROLs according to mixed strategy profile \( (\sigma^k_{1\leq i\leq k}) \).

As a preliminary step, we establish a series of claims.
Claim 3. Fix any $P$. Then, for any $\alpha > 0$,
\[
\Pr \left[ \| D^k(P) - \bar{D}^k(P) \| > \sqrt{|C|} \alpha \right] \leq |C| \cdot e^{-2k\alpha^2}.
\]

Proof. First by McDiarmid’s theorem, for each $c \in C$,
\[
\Pr \{| D^k_c(P) - \bar{D}^k_c(P) | > \alpha \} \leq e^{-2k\alpha^2},
\]
since for each $c \in C$, $| D^k_c(P)(R_1, ..., R_k) - D^k_c(P')(R'_1, ..., R'_k) | \leq 1/k$ whenever $(R_1, ..., R_k)$ and $(R'_1, ..., R'_k)$ differ only in one component.

It then follows that
\[
\begin{align*}
\Pr \left[ \| D^k(P) - \bar{D}^k(P) \| > \sqrt{|C|} \alpha \right] & \leq \Pr \left[ \exists c \in C \text{ s.t. } | D^k_c(P) - \bar{D}^k_c(P) | > \alpha \right] \\
& \leq |C| \cdot e^{-2k\alpha^2}.
\end{align*}
\]

Claim 4. The sequence of functions $\{ D^k(\cdot) \}_k$ are equicontinuous (in the class of normalized demand functions across all $k = 1, ..$).

Proof. Fix $\varepsilon > 0$ and $P \in [0, 1]^C$.

I want to find $\delta > 0$ (which may depend on $\varepsilon$ and $P$) s.t.
\[
\| \bar{D}^k(P') - \bar{D}^k(P) \| < \varepsilon
\]
for all $P' \in [0, 1]^C$ with $\| P' - P \| < \delta$ and all $k$.

Define
\[
\Theta_{P,P'} := \left\{ (u,s) \in \Theta : \exists c \in C \text{ s.t. } s_c \text{ is weakly greater than one and only one of } P_c \text{ and } P'_c \right\}.
\]

We can find $\delta > 0$ s.t. $\eta(\Theta_{P,P'}) < \varepsilon / \sqrt{|C|}$ for all $P'$ s.t. $\| P' - P \| < \delta$. This can be guaranteed if we assume the measure $\eta$ to be absolutely continuous w.r.t. Lebesgue measure.
Then we have
\[
\|\bar{D}^k(P') - \bar{D}^k(P)\| = \sqrt{\sum_{c \in C} |\mathbb{E} [D^k_c(P') - D^k_c(P)]|^2}
\]
\[
= \sqrt{\sum_{c \in C} \mathbb{E} \left[ \frac{1}{k} \sum_{i = 1}^k \left( I \{c \in \arg \max_{R_i^c} \{c' \in C : s_{i,c'} \geq P_{c'} \} \} - I \{c \in \arg \max_{R_i^c} \{c' \in C : s_{i,c'} \geq P_{c'} \} \} \right)^2 \right]}
\]
\[
\leq \frac{1}{k} \sqrt{\sum_{c \in C} \left( \mathbb{E} \sum_{i = 1}^k \left( I \{c \in \arg \max_{R_i^c} \{c' \in C : s_{i,c'} \geq P_{c'} \} \} - I \{c \in \arg \max_{R_i^c} \{c' \in C : s_{i,c'} \geq P_{c'} \} \} \right)^2 \right) \}
\]
\[
\leq \frac{1}{k} \sqrt{\sum_{c \in C} \left( \mathbb{E} \sum_{i = 1}^k I \{\theta_{i} \in \Theta_{P,P'} \} \right)^2} \leq \frac{1}{k} \sqrt{\sum_{c \in C} (k \cdot \eta (\Theta_{P,P'})^2)
\]
\[
= \sqrt{|C| \eta (\Theta_{P,P'})} < \varepsilon,
\]
where the first inequality follows Jensen, and the third inequality is because the two sets \(\{c' \in C : s_{i,c'} \geq P_{c'} \}\) and \(\{c' \in C : s_{i,c'} \geq P_{c'} \}\) are identical when \(\theta_{i} \notin \Theta_{P,P'}\). \(\square\)

**Claim 5.** The sequence of functions \(\{\bar{D}^k\}_{k=1}^{\infty}\) has a subsequence that converges uniformly to some continuous function \(\bar{D}\).

**Proof.** Because the sequence of functions \(\{D^k\}_{k=1}^{\infty}\) defined on a compact set \([0,1]^C\) are uniformly bounded and equicontinuous (by Claim 4), by Arzela-Ascoli theorem, we can find a subsequence \(\{\bar{D}^k\}_{k=1}^{\infty}\) uniformly convergent to some continuous function \(\bar{D}\). \(\square\)

**Claim 6.** For any \(\varepsilon' > 0\), there exists a subsequence \(\{\bar{D}^{k_i}\}_{i=1}^{\infty}\) such that \(\lim_{i \to \infty} \Pr \{ \sup_{P} \|D^{k_i}(P) - \bar{D}(P)\| > \varepsilon' \} = 0\).

**Proof.** Using the argument in the proof of Glivenko-Cantelli, we can partition the space of \(P\)'s into finite intervals \(\Pi_{i_1,..,i_C} [P_{i_j}, P_{i_j+1}]\), where \(i_j = 0, \ldots, i_j^*\) such that \(\|\bar{D}(P_{i+1}) - \bar{D}(P_i^-)\| < \varepsilon'/2\) for all \(i = (i_1, \ldots, i_C)\), where \(i + 1 := (i_1 + 1, \ldots, i_C + 1)\). Let \(m\) be the number of such intervals. Using the argument of Glivenko-Cantelli, one can show that for any \(P\) there exists \(i\) such that
\[
\|D^{k_i}(P) - \bar{D}(P)\| \leq \max\{\|D^{k_i}(P_1) - \bar{D}(P_1)\|, \|D^{k_i}(P_{i+1}) - \bar{D}(P_{i+1})\|\} + \varepsilon'/2.
\]
Suppose event \(\|D^{k_i}(P) - \bar{D}(P)\| > \varepsilon'\) occurs for some \(P\). Then there must exist \(i\) such that \(\|D^{k_i}(P_1) - \bar{D}(P_1)\| \geq \varepsilon' / 2\). Since \(\bar{D}^{k_i}(\cdot)\) converges to \(\bar{D}(\cdot)\) in sup norm by Claim 5, there exists \(K'\) such that for all \(\ell > K'\), \(\sup_P \|D^{k_i}(P) - \bar{D}(P)\| < \varepsilon'/4\). Hence, for \(\ell > K'\) and for \(i\), we must have
\[
\|D^{k_i}(P_1) - \bar{D}^{k_i}(P_1)\| \geq \varepsilon'/4.
\]

33
Combining the arguments so far, we conclude:

\[
Pr\{\sup_P \| D^k(P) - \bar{D}(P) \| > \epsilon' \} \\
= Pr\{\exists P \text{ s.t. } \| D^k(P) - \bar{D}(P) \| > \epsilon' \} \\
\leq Pr\{\exists i \text{ s.t. } \| D^k(P_i) - \bar{D}(P_i) \| > \epsilon'/2 \} \\
\leq Pr\{\exists i \text{ s.t. } \| D^k(P_i) - \bar{D}^k(P_i) \| > \epsilon'/4 \} \\
= Pr\{\cup_i \{\| D^k(P_i) - \bar{D}^k(P_i) \| > \epsilon'/4 \} \} \\
\leq \sum_i Pr\{\| D^k(P_i) - \bar{D}^k(P_i) \| > \epsilon'/4 \} \\
\leq \sum_i e^{-k_i \epsilon'^2/8} \\
= me^{-k_\ell \epsilon'^2/8} \to 0 \text{ as } \ell \to \infty,
\]

where the penultimate inequality follows from Claim 3.

Now we are in a position to prove Theorem 2.

Suppose to the contrary that the sequence of strategy profiles \( \{\sigma^k_i\}_{1 \leq i \leq k} \) are not asymptotically stable. Then by definition, there exists \( \varepsilon > 0 \) and a subsequence of finite economies \( \{F^k_j\}_j \) such that

\[
Pr\left( \text{The fraction of students playing SRS against } P^k_i \geq 1 - \varepsilon \right) < 1 - \varepsilon. \tag{\ast}
\]

By Claim 6, we know that there exists a sub-subsequence \( D^{k_{ji}} \) that converges to \( \bar{D} \) uniformly in probability. Given the regularity of the strategies employed by the students along with the full support assumption, \( \bar{D} \) is \( C^1 \) and \( \partial \bar{D} \) is invertible. Hence (using an argument by AL), we know that \( P^{k_{ji}} \) converges to \( \bar{P} \) in probability, where \( \bar{P} \) is a deterministic cutoff s.t. \( \bar{D}(\bar{P}) = S \).

Define

\[
\hat{\Theta} := \{(u, s) : |u_c - u_{c'}| > \delta \text{ for all } c \neq c'\} \cap \{(u, s) : |s_c - \bar{P}_c| > \delta\}.
\]

Let’s take \( \delta \) to be small enough s.t. \( \eta\left(\hat{\Theta}\right) > (1 - \varepsilon)^{1/3} \) (this can be done since \( \eta \) is absolutely continuous).

By WLLN, we know that \( \eta^{k_{ji}}\left(\hat{\Theta}\right) \) converges to \( \eta\left(\hat{\Theta}\right) \) in probability, and therefore there exists \( L_1 \) s.t. for all \( l > L_1 \) we have

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

For each economy \( F^{k_{ji}} \), define the event

\[
A^{k_{ji}} := \left\{ \left| P^{k_{ji}}_c - \bar{P}_c \right| < \delta \text{ for all } c \in C \right\}.
\]

Because \( P^{k_{ji}} \) converges to \( \bar{P} \) in probability, there exists \( L_2 \) s.t. for all \( l > L_2 \) we have

\[
Pr\left( A^{k_{ji}} \right) \geq \max\left\{ (1 - \varepsilon)^{1/6}, 1 - (1 - \varepsilon)^{1/2} \left[(1 - \varepsilon)^{1/3} - (1 - \varepsilon)^{1/2}\right] \right\}. \tag{\ast\ast}
\]

34
Because \( \left\{(\sigma^k_i)_{1 \leq i \leq k}\right\}_k \) is a robust equilibrium, there exists \( L_3 \) s.t. for all \( l > L_3 \) the strategy profile \( \left(\sigma^k_i\right)_{i=1}^{k_j} \) is a \( \delta \left[(1 - \varepsilon)^{1/6} - (1 - \varepsilon)^{1/3}\right] \)-BNE for economy \( F^{k_j} \).

By WLLN, there exists \( \hat{L} \) s.t. \( \hat{L} \) i.i.d. Bernoulli random variables with \( p = (1 - \varepsilon)^{1/3} \) 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 \( (1 - \varepsilon)^{1/2} k_j > \hat{L} \).

Now let’s fix an arbitrary \( l > \max\{L_1, L_2, L_3, L_4\} \), and we wish to show that in economy \( F^{k_j} \)

\[
\Pr\left(\text{The fraction of students playing SRS against } P^{k_j} \geq 1 - \varepsilon\right) \geq 1 - \varepsilon,
\]

which would contradicts (*) and complete the proof.

First, notice that in economy \( F^{k_j} \), a student with \( \theta \in \hat{\Theta} \) plays SRS against \( \bar{P} \) with probability no less than \( (1 - \varepsilon)^{1/3} \). To see this, suppose by contrary that there exists some student \( i \) and some type \( \theta \in \hat{\Theta} \) s.t.

\[
\Pr\left(\sigma^k_i (\theta) \text{ plays SRS against } \bar{P}\right) < (1 - \varepsilon)^{1/3}.
\]

Then deviating to TRS will give this student \( i \) with type \( \theta \) at least a gain of

\[
\delta \cdot \Pr\left(\sigma^k_i (\theta) \text{ does not play SRS against } P^{k_j}\right)
\]

\[
\geq \delta \cdot \Pr\left(\sigma^k_i (\theta) \text{ does not play SRS against } \bar{P}\right)
\]

\[
\geq \delta \left[\Pr(A^{k_j}) - \Pr(\sigma^k_i (\theta) \text{ plays SRS against } \bar{P})\right]
\]

\[
\geq \delta \left[(1 - \varepsilon)^{1/6} - (1 - \varepsilon)^{1/3}\right],
\]

which contradicts the construction of \( L_3 \), which implies that the strategy profile \( \left(\sigma^k_i\right)_{i=1}^{k_j} \) is a \( \delta \left[(1 - \varepsilon)^{1/6} - (1 - \varepsilon)^{1/3}\right] \)-BNE for the economy \( F^{k_j} \).

Therefore in economy \( F^{k_j} \), for each student \( i = 1, \ldots, k_j \) and each \( \theta \in \hat{\Theta} \), we have

\[
\Pr\left(\sigma^k_i (\theta) \text{ plays SRS against } \bar{P}\right) \eta^{k_j} \left(\hat{\Theta}\right) \geq (1 - \varepsilon)^{1/2}
\]

\[
= \Pr\left(\sigma^k_i (\theta) \text{ plays SRS against } \bar{P}\right) \geq (1 - \varepsilon)^{1/3},
\]

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

\[
\Pr \left( \text{The fraction of students with } \theta \in \hat{\Theta} \text{ playing SRS against } \bar{P} \geq (1 - \varepsilon)^{1/2} \mid \eta^{k_{hi}} \left( \hat{\Theta} \right) \geq (1 - \varepsilon)^{1/2} \right) \\
\geq \Pr \left( \eta^{k_{hi}} \left( \hat{\Theta} \right) \cdot k_{ji} \text{ i.i.d. Bernoulli random variables with } p = (1 - \varepsilon)^{1/3} \mid \eta^{k_{hi}} \left( \hat{\Theta} \right) \geq (1 - \varepsilon)^{1/2} \right) \\
\geq \Pr \left( \hat{L} \text{ i.i.d. Bernoulli random variables with } p = (1 - \varepsilon)^{1/3} \right.
\quad \text{have a sample mean no less than } (1 - \varepsilon)^{1/2} \\
\left. \eta^{k_{hi}} \left( \hat{\Theta} \right) \geq (1 - \varepsilon)^{1/2} \right)
\geq (1 - \varepsilon)^{1/3},
\]

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

Comparing the finite economy random cutoff \(P^{k_{hi}}\) with the deterministic cutoff \(\bar{P}\), we have

\[
\Pr \left( \text{The fraction of students with } \theta \in \hat{\Theta} \text{ playing SRS against } P^{k_{hi}} \geq (1 - \varepsilon)^{1/2} \mid \eta^{k_{hi}} \left( \hat{\Theta} \right) \geq (1 - \varepsilon)^{1/2} \right) \\
\geq \Pr \left( \text{The fraction of students with } \theta \in \hat{\Theta} \text{ playing SRS against } \bar{P} \geq (1 - \varepsilon)^{1/2}, \text{ and event } A^{k_{hi}} \right) \\
\geq \Pr \left( \text{The fraction of students with } \theta \in \hat{\Theta} \text{ playing SRS against } P \geq (1 - \varepsilon)^{1/2} \right. \\
\quad \left. \eta^{k_{hi}} \left( \hat{\Theta} \right) \geq (1 - \varepsilon)^{1/2} \right) \\
- \Pr \left( \bar{A}^{k_{hi}} \right. \\
\quad \left. \eta^{k_{hi}} \left( \hat{\Theta} \right) \geq (1 - \varepsilon)^{1/2} \right) \\
\geq (1 - \varepsilon)^{1/3} - \frac{\Pr \left( \bar{A}^{k_{hi}} \right.}{\Pr \left( \eta^{k_{hi}} \left( \hat{\Theta} \right) \geq (1 - \varepsilon)^{1/2} \right)} \\
\geq (1 - \varepsilon)^{1/3} - \frac{\Pr \left( \eta^{k_{hi}} \left( \hat{\Theta} \right) \geq (1 - \varepsilon)^{1/2} \right)}{\Pr \left( \eta^{k_{hi}} \left( \hat{\Theta} \right) \geq (1 - \varepsilon)^{1/2} \right)} \\
\geq (1 - \varepsilon)^{1/3} - \frac{(1 - \varepsilon)^{1/2} \left[ (1 - \varepsilon)^{1/3} - (1 - \varepsilon)^{1/2} \right]}{(1 - \varepsilon)^{1/2}} \\
= (1 - \varepsilon)^{1/2},
\]

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

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

\[
\Pr \left( \text{The fraction of students playing SRS against } P^{kjl} \geq 1 - \varepsilon \right)
\geq \Pr \left( \begin{aligned}
\text{The fraction of students with } \theta \in \hat{\Theta} \\
\text{playing SRS against } P^{kjl} \geq (1 - \varepsilon)^{1/2} \\
\text{and } \eta^{kjl} \left( \hat{\Theta} \right) \geq (1 - \varepsilon)^{1/2}
\end{aligned} \right)
\]

\[
= \Pr \left( \eta^{kjl} \left( \hat{\Theta} \right) \geq (1 - \varepsilon)^{1/2} \right) \cdot \Pr \left( \begin{aligned}
\text{The fraction of students with } \theta \in \hat{\Theta} \\
\text{playing SRS against } P^{kjl} \geq (1 - \varepsilon)^{1/2} \\
\text{and } \eta^{kjl} \left( \hat{\Theta} \right) \geq (1 - \varepsilon)^{1/2}
\end{aligned} \right)
\]

\[
\geq (1 - \varepsilon)^{1/2} \cdot (1 - \varepsilon)^{1/2}
\]

\[
= 1 - \varepsilon,
\]

which contradicts (*).

\[ \square \]

## B Monte Carlo Simulations

This appendix describes how we estimate student preferences under the weak truth-telling and the stability condition.

It further presents some additional details of the Monte Carlo simulations. Figure B1 describes the simulated spatial distribution of students and schools in one sample; Figure B2 depicts the marginal distribution of each school’s cutoffs across 300 simulations under the assumption that every student always truthfully ranks all schools.

Furthermore, Tables B1 and B2 describe the skipping and mistakes for students with $T_i = 1$ and $T_i = 0$, respectively. Table B3 shows the mean and standard deviation of the estimates of each coefficient from different approaches; and Tables B4 and B5 present more detailed estimation results of the welfare effects.

### B.1 Estimation

Our formulation of estimation approaches are the same as Fack, Grenet, and He (2015) who also provide more details on assumptions for estimation as well as generalizations to the random utility model.

Recall that the random utility model is specified as follows:

\[
u_{i,s} = \beta_1 s - d_{i,s} + \beta_2 T_i A_s + \sigma \epsilon_{i,s} 
\]

\[\equiv \sigma V_{is} + \sigma \epsilon_{i,s}, \forall i \text{ and } s;\]

we also define $X_i = (\{d_{i,s}, A_s\}_{s}, T_i)$ to denote the observable student characteristics and school attributes; $L_i = (l^i_1, \ldots, l^i_{K_i})$ represents the rank-ordered list that $i$ submits; and $\theta$ is the vector of coefficients, $\theta = (\beta_1, \beta_2, \sigma)$. 

37
Weakly Truth-Telling. We start with formalizing the estimation under the weakly truth-telling assumption. If each student $i$ submits $K_i \equiv |L_i| \leq S$ choices, under the assumption that students are weakly truth-telling, $L_i = (l^1_i, \ldots, l^K_i)$ ranks truthfully $i$’s top $K_i$ preferred schools.

The probability of student $i$ submitting $L_i$ is:

$$\Pr (i \text{ submits } L_i \mid X_i; \theta) = \Pr (L_i = (l^1_i, \ldots, l^K_i) \mid X_i; \theta; K_i) \times \Pr (i \text{ submits a ROL of length } K_i \mid X_i; \theta).$$

Under the assumptions that $K_i$ is orthogonal to $u_{i,s}$ for all $s$ and that $\epsilon_{i,s}$ is a type-I extreme value, we can focus the choice probability conditional $K_i$ and obtain:

$$\Pr (L_i = (l^1_i, \ldots, l^K_i) \mid X_i; \theta; K_i) = \prod_{s \in L_i} \left( \frac{\exp(V_{i,s})}{\sum_{s' \notin L_i} \exp(V_{i,s'})} \right)$$

where $s' \notin L_i$ $s$ indicates that $s'$ is not ranked before $s$ in $L_i$, which includes $s$ itself and the schools not ranked in $L_i$. This rank-ordered (or “exploded”) logit model can be seen as a series of conditional logit models: one for the top-ranked school ($l^1_i$) being the most preferred; another for the second-ranked school ($l^2_i$) being preferred to all schools except the one ranked first, and so on.

The model can be estimated by maximum likelihood estimation (MLE) with the log-likelihood function:

$$\ln L_{TT} (\theta \mid X, |L|) = \sum_{i=1}^{I} \sum_{s \in L_i} V_{i,s} - \sum_{i=1}^{I} \ln \left( \sum_{s' \in S \setminus L_i} \exp(V_{i,s'}) \right),$$

where $|L|$ is the length of all ROLs.

Stability. We now assume that the matching is stable and explore how we can identify and estimate student preferences. Suppose that the matching is $\mu$, which leads to a vector of cutoffs $P(\mu)$. With information on how schools rank students, we can find a set of schools that are ex post feasible to $i$, $S(i, P(\mu))$.

The conditions specified by the stability of $\mu$ imply the likelihood of student $i$ choosing $\mu(i)$ in $S(i, P(\mu))$:

$$\Pr \left( \mu(i) = \arg \max_{s \in S(i, P(\mu))} u_{i,s} | X_i, P(\mu); \theta \right).$$

Given the parametric assumptions on utility functions, the corresponding (conditional) log-likelihood function is:

$$\ln L_{ST} (\theta \mid X, P(\mu)) = \sum_{i=1}^{I} V_{i,\mu(i)} - \sum_{i=1}^{I} \ln \left( \sum_{s' \in S_i} \exp(V_{i,s'}) \right).$$

A key assumption of this approach is that the feasible set $S(i, P(\mu))$ is exogenous to $i$. As shown in Fack, Grenet, and He (2015), it is satisfied when the mechanism is the serial dictatorship and when there are no peer effects.

38
Figure B1: Monte Carlo Simulations: Spatial Distribution of Students and Schools

Notes: This figure shows the spatial configuration of the area considered in the Monte Carlo simulations with 400 students and 12 schools. The area is within a circle of radius 1. The blue and red circles show the locations of students and schools, respectively, in one simulation sample. Across samples, the schools' locations are fixed, while students' locations are uniformly drawn within the circle.
Figure B2: Simulated Distribution of Cutoffs when Everyone is Truth-telling

Notes: Assuming everyone is strictly truth-telling, we calculate the cutoffs of all schools in each simulation sample. The figure shows the marginal distribution of each school’s cutoff, in terms of percentile rank (between 0 (lowest) and 1 (highest)). Each curve is an estimated density based on a normal kernel function. A dotted line indicates a small school with 20, instead of 40, seats. The simulation samples for cutoffs use the same draws of $\{d_{i,s}\}$ and $T_i$ for each student as in the samples for game plays, but $\{\epsilon_{i,s}\}$ are different.
Table B1: Skips and Mistakes in Monte Carlo Simulations (Percentage Points): $T_i = 1$ Students

<table>
<thead>
<tr>
<th>Scenarios (Data Generating Processes)</th>
<th>Payoff Irrelevant Skips</th>
<th>Payoff Relevant Mistakes</th>
</tr>
</thead>
<tbody>
<tr>
<td>STT</td>
<td>IRR 1</td>
<td>IRR 2</td>
</tr>
<tr>
<td>WTT: Weak Truth-Telling\textsuperscript{a}</td>
<td>100</td>
<td>77</td>
</tr>
<tr>
<td>Matched w/ favorite feasible school\textsuperscript{b}</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Skippers\textsuperscript{c}</td>
<td>0</td>
<td>26</td>
</tr>
</tbody>
</table>

By number of skips:

<table>
<thead>
<tr>
<th></th>
<th>11</th>
<th>10</th>
<th>26</th>
<th>40</th>
<th>55</th>
<th>61</th>
<th>68</th>
<th>73</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0</td>
<td>4</td>
<td>9</td>
<td>14</td>
<td>27</td>
<td>28</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>4</td>
<td>11</td>
<td>17</td>
<td>13</td>
<td>8</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>6</td>
<td>10</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

STT: Strict Truth-telling

|          | 100 | 74  | 40  | 3   | 3   | 3   | 3   | 3   |

Notes: This table presents the configurations of the eight data generating process (DGP)s, similar to Table 8 but only among the students with $T_i = 1$. Each entry is a percentage averaged over the 300 simulation samples. On average, there are 132 such students in each sample. \textsuperscript{a}A student is “weak truth-telling” if she truthfully ranks her top $K_i$ ($1 \leq K_i \leq 12$) preferred schools. \textsuperscript{b}A school is feasible to a student, if the student’s index is higher than the school’s \textit{ex post} admission cutoff. If a student is matched with her favorite feasible school, she cannot form a blocking pair with any school. \textsuperscript{c}Given that every school is acceptable to all students and potentially over-demanded, a student is a skipper if she does not rank all schools. \textsuperscript{d}In each DGP, this reports the percentage of samples that the WTT (weakly truth-telling) assumption is rejected at 5% level in favor of the stability assumption. The test is based on the Durbin-Wu-Hausman test and discussed in details in section 4.3.
Table B2: Skips and Mistakes in Monte Carlo Simulations (Percentage Points): $T_i = 0$ Students

<table>
<thead>
<tr>
<th>Scenarios (Data Generating Processes)</th>
<th>Payoff Irrelevant Skips</th>
<th>Payoff Relevant Mistakes</th>
</tr>
</thead>
<tbody>
<tr>
<td>STT</td>
<td>IRR 1</td>
<td>IRR 2</td>
</tr>
<tr>
<td>WTT: Weak Truth-Telling$^a$</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td>Matched w/ favorite feasible school$^b$</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Skippers$^c$</td>
<td>0</td>
<td>23</td>
</tr>
</tbody>
</table>

By number of skips:

<table>
<thead>
<tr>
<th></th>
<th>11</th>
<th>10</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>WTT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>STT</td>
<td>0</td>
<td>16</td>
<td>34</td>
<td>49</td>
<td>58</td>
<td>60</td>
<td>63</td>
<td>65</td>
<td>63</td>
<td>63</td>
<td>63</td>
</tr>
<tr>
<td>IRR 1</td>
<td>4</td>
<td>9</td>
<td>14</td>
<td>14</td>
<td>11</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>IRR 2</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IRR 3</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>REL 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>REL 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>REL 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>REL 4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

STT: Strict Truth-telling

|       | 100 | 77  | 50  | 25  | 25  | 25  | 25  | 25  |

Notes: This table presents the configurations of the eight data generating process (DGP)s, similar to Table 8 but only among the students with $T_i = 0$. Each entry is a percentage averaged over the 300 simulation samples. On average, there are 268 such students in each sample. $^a$A student is “weak truth-telling” if she truthfully ranks her top $K_i$ (1 ≤ $K_i$ ≤ 12) preferred schools. $^b$A school is feasible to a student, if the student’s index is higher than the school’s ex post admission cutoff. If a student is matched with her favorite feasible school, she cannot form a blocking pair with any school. $^c$Given that every school is acceptable to all students and potentially over-demanded, a student is a skipper if she does not rank all schools. $^d$In each DGP, this reports the percentage of samples that the WTT (weakly truth-telling) assumption is rejected at 5% level in favor of the stability assumption. The test is based on the Durbin-Wu-Hausman test and discussed in details in section 4.3.
<table>
<thead>
<tr>
<th>DGPs</th>
<th>Identifying Condition</th>
<th>Quality ($\beta_1 = 0.2$)</th>
<th>Interaction ($\beta_2 = 1$)</th>
<th>Unobservable ($\sigma = 1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
</tr>
<tr>
<td>A. <strong>Strict Truth-telling (Both WTT and stability are satisfied)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STT</td>
<td>WTT</td>
<td>0.20</td>
<td>0.01</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.21</td>
<td>0.05</td>
<td>1.06</td>
</tr>
<tr>
<td>B. <strong>Payoff-irrelevant Skips (Only stability is satisfied)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IRR 1</td>
<td>WTT</td>
<td>0.19</td>
<td>0.01</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.21</td>
<td>0.05</td>
<td>1.06</td>
</tr>
<tr>
<td>IRR 2</td>
<td>WTT</td>
<td>0.18</td>
<td>0.02</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.21</td>
<td>0.05</td>
<td>1.06</td>
</tr>
<tr>
<td>IRR 3</td>
<td>WTT</td>
<td>0.13</td>
<td>0.02</td>
<td>-0.18</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.21</td>
<td>0.05</td>
<td>1.06</td>
</tr>
<tr>
<td>C. <strong>Payoff-relevant Mistakes (Neither WTT or stability is satisfied)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REL 1</td>
<td>WTT</td>
<td>0.14</td>
<td>0.02</td>
<td>-0.23</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.20</td>
<td>0.06</td>
<td>0.91</td>
</tr>
<tr>
<td>REL 2</td>
<td>WTT</td>
<td>0.14</td>
<td>0.02</td>
<td>-0.26</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.20</td>
<td>0.06</td>
<td>0.77</td>
</tr>
<tr>
<td>REL 3</td>
<td>WTT</td>
<td>0.15</td>
<td>0.02</td>
<td>-0.23</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.19</td>
<td>0.06</td>
<td>0.63</td>
</tr>
<tr>
<td>REL 4</td>
<td>WTT</td>
<td>0.15</td>
<td>0.02</td>
<td>-0.19</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.18</td>
<td>0.06</td>
<td>0.52</td>
</tr>
</tbody>
</table>

**Notes:** This table presents estimates (mean and standard deviation across 300 samples) of the random utility model described in equation (11). The true values are ($\beta_1, \beta_2, \sigma$) = (0.2, 1, 1). It shows results in the eight data generating process (DGPs) with two identifying assumptions, WTT and stability. WTT assumes that every student truthfully ranks her top $K_i$ (1 < $K_i$ ≤ 12) schools. Stability implies that every student is matched with her favorite feasible school, given the ex post cutoffs.
Table B4: Welfare Effects of the Counterfactual Policy on Students with $T_i = 1$

<table>
<thead>
<tr>
<th>DGP</th>
<th>Approach to Counterfactual</th>
<th>Worse Off</th>
<th>Better Off</th>
<th>Indifferent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
</tr>
<tr>
<td>A. Strict truth-telling</td>
<td>Submitted ROLs</td>
<td>0.00</td>
<td>0.00</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>WTT</td>
<td>0.00</td>
<td>0.00</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.00</td>
<td>0.00</td>
<td>0.84</td>
</tr>
<tr>
<td>B. Payoff-irrelevant skips</td>
<td>Submitted ROLs</td>
<td>0.00</td>
<td>0.00</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>IRR 1</td>
<td>WTT</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.00</td>
<td>0.00</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td>0.00</td>
<td>0.00</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>IRR 2</td>
<td>WTT</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.00</td>
<td>0.00</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td>0.00</td>
<td>0.00</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>IRR 3</td>
<td>WTT</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.00</td>
<td>0.00</td>
<td>0.84</td>
</tr>
<tr>
<td>C. Payoff-relevant mistakes</td>
<td>Submitted ROLs</td>
<td>0.00</td>
<td>0.01</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>REL 1</td>
<td>WTT</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.00</td>
<td>0.01</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td>0.01</td>
<td>0.01</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>REL 2</td>
<td>WTT</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.00</td>
<td>0.01</td>
<td>0.83</td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td>0.02</td>
<td>0.02</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>REL 3</td>
<td>WTT</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.00</td>
<td>0.02</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td>0.02</td>
<td>0.02</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>REL 4</td>
<td>WTT</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.00</td>
<td>0.02</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Notes: This table presents the estimated effects of the counterfactual policy (giving $T_i = 1$ students priority in admission) on students with $T_i = 1$. On average, there are 132 such students (standard deviation 6.42) in each simulation sample. The true values of the welfare effects are in the first row of the table (in bold). The table shows results in the eight data generating process (DGPs) with three approaches. The one using submitted ROLs assumes that submitted ROLs represent student true ordinal preferences; WTT assumes that every student truthfully ranks her top $K_i$ ($1 < K_i \leq 12$) preferred schools; and stability implies that every student is matched with her favorite feasible school, given the ex post cutoffs. The welfare change of each student is calculated in the following way: We first simulate the counterfactual match and investigate if a given student is better off, worse off, or indifferent by comparing the two matches according to estimated/assumed ordinal preferences. In each simulation sample, we calculate the fractions of different welfare change; the table then reports the mean and standard deviation of the fractions across the 300 simulation samples.
Table B5: Welfare Effects of the Counterfactual Policy on Students with $T_i = 0$

<table>
<thead>
<tr>
<th>DGP</th>
<th>Approach to Counterfactual</th>
<th>Worse Off mean</th>
<th>s.d.</th>
<th>Better Off mean</th>
<th>s.d.</th>
<th>Indifferent mean</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>A. Strict truth-telling</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td><strong>0.59</strong> 0.04</td>
<td>0.00</td>
<td><strong>0.00</strong> 0.00</td>
<td>0.41</td>
<td>0.00</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>WTT</td>
<td>0.58 0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.42</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.57 0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>0.41</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>B. Payoff-irrelevant skips</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td>0.51 0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.49</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>IRR 1</td>
<td>WTT</td>
<td>0.56 0.04</td>
<td>0.02</td>
<td>0.00</td>
<td>0.43</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.57 0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>0.41</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td>0.41 0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.59</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>IRR 2</td>
<td>WTT</td>
<td>0.51 0.03</td>
<td>0.04</td>
<td>0.00</td>
<td>0.45</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.57 0.03</td>
<td>0.01</td>
<td>0.00</td>
<td>0.41</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td>0.31 0.03</td>
<td>0.00</td>
<td>0.01</td>
<td>0.68</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>IRR 3</td>
<td>WTT</td>
<td>0.38 0.03</td>
<td>0.11</td>
<td>0.01</td>
<td>0.51</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.57 0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.41</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>C. Payoff-relevant mistakes</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td>0.26 0.03</td>
<td>0.00</td>
<td>0.01</td>
<td>0.73</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>REL 1</td>
<td>WTT</td>
<td>0.40 0.03</td>
<td>0.10</td>
<td>0.01</td>
<td>0.50</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.56 0.03</td>
<td>0.02</td>
<td>0.01</td>
<td>0.42</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td>0.25 0.03</td>
<td>0.00</td>
<td>0.01</td>
<td>0.75</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>REL 2</td>
<td>WTT</td>
<td>0.41 0.03</td>
<td>0.09</td>
<td>0.01</td>
<td>0.50</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.55 0.03</td>
<td>0.03</td>
<td>0.01</td>
<td>0.42</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td>0.24 0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>0.76</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>REL 3</td>
<td>WTT</td>
<td>0.42 0.02</td>
<td>0.09</td>
<td>0.01</td>
<td>0.49</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.53 0.02</td>
<td>0.04</td>
<td>0.01</td>
<td>0.43</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Submitted ROLs</td>
<td>0.23 0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>0.77</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>REL 4</td>
<td>WTT</td>
<td>0.42 0.02</td>
<td>0.09</td>
<td>0.01</td>
<td>0.49</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stability</td>
<td>0.52 0.02</td>
<td>0.05</td>
<td>0.01</td>
<td>0.43</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table presents the estimated effects of the counterfactual policy (giving $T_i = 1$ students priority in admission) on students with $T_i = 0$. On average, there are 268 such students (standard deviation 6.42) in each simulation sample. The true values of the welfare effects are in the first row of the table (in bold). The table shows results in the eight data generating process (DGPs) with three approaches. The one using submitted ROLs assumes that submitted ROLs represent student true ordinal preferences; WTT assumes that every student truthfully ranks her top $K_i$ ($1 < K_i \leq 12$) preferred schools; and stability implies that every student is matched with her favorite feasible school, given the ex post cutoffs. The welfare change of each student is calculated in the following way: We first simulate the counterfactual match and investigate if a given student is better off, worse off, or indifferent by comparing the two matches according to estimated/assumed ordinal preferences. In each simulation sample, we calculate the fractions of different welfare change; the table then reports the mean and standard deviation of the fractions across the 300 simulation samples.