Strategic "Mistakes": Implications for Market Design Research*

Georgy Artemov† Yeon-Koo Che‡ Yinghua He§

July 15, 2017

Abstract

Using a rich data set on Australian college admissions, we show that a non-negligible fraction of applicants choose strategies (or rank-ordered lists) that are unambiguously dominated; however, 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 empirical analyses of market design based on *truthful reporting hypothesis* under a (seemingly) strategy-proof mechanism, but leaves the approaches based on weaker *stable-matching condition* relatively unscathed. Studying the design of matching applicants with colleges, our Monte Carlo simulations further illustrate this point and quantify 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, payoff relevance of mistakes, robust equilibria, truthful reporting strategy, stable response strategy, stable matching.

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 strategyproofness 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

*We are grateful to Xingye Wu, who has provided excellent research assistance for the theory part, and Julien Grenet for his generous help with the Monte Carlo Simulations. We are grateful to the participants of Asia Pacific Industrial Organization Conference and "Econometrics Meets Theory" Conference at NYU 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.

others who may game the system. Further, it aids empirical research by making participants' choices easy to interpret.

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[1] 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 from Victorian Tertiary Admissions Centre (VTAC), a central clearinghouse that organizes the match for applicants in tertiary admissions in Victoria, Australia. The semi-centralized mechanism VTAC uses for assigning applicants to tertiary "courses" (which are similar to college-major-tuition triple 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). Similar to Hassidim, Romm, and Shorrer (2016), our study exploits the unique feature of the system that an applicant can apply for a given college-major pair as either (1) a "full-fee" course (or FF course) which requires full tuition; or (2) a Commonwealth supported course (which we will refer to as "reduced-fee" or RF course) which subsidizes about 50% of tuition; or both. For every college-major pair RF clearly dominates FF course; hence ranking FF but not RF course of the same college-major pair in a ROL that does not fill up the 12 slots—henceforth called a **skip**—is unambiguously a dominated strategy.[2]

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, or 34% out of total 2,963

[1] Obviously 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.

[2] As will be discussed in detail, listing the latter ahead of the former need not be a dominated strategy in our data.

applicants who listed at least one full-fee course. These figures are consistent, although not directly comparable, with those found in Hassidim, Romm, and Shorrer (2016) and can be viewed as non-negligible. However, the vast majority of these mistakes were not payoff relevant. Correcting the mistakes (i.e., listing the omitted reduced-fee course) would have made a difference only for $14-201$ applicants out of 1,009 who skipped at least one RF 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 applicants and $0.47-6.78\%$ out of those who listed at least one full-fee course with a reduced-fee counterpart.[3]

Our rich micro dataset 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.

A unique feature of the Victorian mechanism, which lets applicants to modify their submitted 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 difference between the realized ENTER and the ENTER forecasted based on the ability test. A larger difference, 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 important to understand the implications of mistakes for market design research. Of particular interest is how mistakes—some of them payoff-relevant—affect our ability to recover the underlying preferences of participants and to perform counterfactual analyses of new hypothetical market designs. 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

[3]These figures are made comparable to the findings of Hassidim, Romm, and Shorrer (2016). The percentage of applicants with payoff-relevant mistakes out of those who skip is 1.39% using the lower bound and 19.92% using the upper bound. The corresponding figures obtained by Hassidim, Romm, and Shorrer (2016) are 2.18% and 7.29%, respectively.

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 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.[4]

To gain quantitative insights, we perform a Monte Carlo simulation of college assignment in which applicant's preferences follow a logit model. We assume a serial dictatorship mechanism with a pre-specified serial order. 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 applicants make mistakes. Specifically, the assumed behavior ranges from truthful reporting (i.e., no mistakes), to 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 applicant's 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 applicants 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 college quality, especially among popular or small colleges, because over-subscribed colleges are often skipped by many applicants.

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,

[4]This approach is similar to the one adopted by Deb and Kalai (2015).

we then also adopt 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 applicants. In terms of predicting the counterfactual matching outcome, estimates from truthful reporting perform worse than those from stable responses when applicants make mistakes. When we evaluate the welfare effects of the policy, the truthful reporting restriction under-estimates the benefits to disadvantaged applicants and the harm to 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. In Section 2, we study the frequency and nature of strategic mistakes from the VTAC college admissions data. In Section 3, we explore the theoretical implications of the findings for empirical identification methods. In Section 4, we report Monte Carlo simulations performed 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 the Victorian Tertiary Admission Centre (VTAC), which is a centralized clearinghouse 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 *course*, is a combination of (i) a tertiary institution; (ii) a field of study which the applicant wants to pursue; and (ii) a tuition payments. A tertiary institution may be either a university (including programs not granting bachelor degrees) or a technical school. A field of study is roughly equivalent to a *major* in the US universities. Tuition payments are made in full for full-fee (FF) courses or reduced to about a half for reduced-fee (RF) courses. Tuition payments are set by the government; the median is about AUD9,000 (USD7,000) per year. Applicants in reduced-fee courses are able to take a subsidized loan to cover the rest of the tuition payments. The normal duration of the program is three years. Apart from tuition payments, there is no difference between FF and RF courses. There were 1899 majors in 2007; 881 of them are offered in both options, FF and RF.

Applicants are required to submit their applications in the form of rank-order list (ROL), along with other information, at the end of September. In mid-December, applicants receive their Equivalent National Tertiary Entrance Ranks (ENTER), as a number between zero and 99.95 in 0.05 increments, which we will refer to as *score* throughout the paper. For applicant i, $Score_i$ is i's rank and shows a percentage of applicants with scores below $Score_i$. For most applicants,

ENTER is the sole determinant of the admission. As ROL is initially submitted before ENTER is known, applicants have an opportunity to revise their ROL after the release of ENTER. Offers are released to the applicants in January-February. Applicants have about two weeks to enroll in the course they are offered.

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* (FF/RF), and transmits the offer(s) to the applicant. That is, if applicant's ROL contains both reduced-fee and full-fee courses, such an applicant may receive two offers, one of each type. That feature means that ranking FF course ahead of RF course is not a dominated strategies, as the list of RF courses is treated as separate from the list of FF courses.

In selecting applicants, courses follow a pre-specified, published set of rules. For the largest category of applicants, the admission is based almost exclusively on their scores. We focus on these applicants in the 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. We refer to the median of the highest scores among all rejected applicants and the lowest scores among all accepted applicants as a *cutoff* of the course.[5]

The number of courses that an applicant can list at most 12. The applicants who exhaust the length of their ROLs may be forced to omit some courses that they find desirable; hence we focus only on these who list fewer than 12 courses, who constitute 75% of all applicants.

Out of 27,992 V16 applicants who list fewer than 12 courses, 24,666 have ranked at least one major that offers both RF and FF courses. 2,915 applicants have ranked at least one FF course and all these applicants have also ranked at least one major offering RF and FF courses.

2.2 Skips and Payoff-Relevant Mistakes

If an applicant lists a FF course but does not list the corresponding RF course, we say that applicant "skips" RF course. Even if an applicant skips RF 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, and the course is not feasible for the applicant. 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 will demonstrate 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 the lower and upper bounds of payoff-relevant mistakes. To calculate the lower bound, we assume that skipped RF course is the applicant's least desirable RF course. To calculate the upper bound, we assume that the skipped RF course is the applicant's most desirable course.[6]

In Table 1 we report the number of applicants that make at least one mistake of skipping a

[5] See Appendix A for details on course selection, the other categories of applicants and the definition of a cutoff.

[6] See Appendix A for details on calculating the cutoffs and the bounds.

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, a skip is not a payoff-relevant mistake; and that the fraction of these with payoff-relevant mistakes is less than one percent of all applicants, with either bounds.

Table 1: Skips and Mistakes among V16 Applicants Listing Fewer than 12 Courses

	All	FF listed	Skips	Payoff-relevant mistakes	
				Upper bound	Lower bound
% All	100.00	10.61	3.61	0.72	0.05
% FF listed		100.00	34.05	6.78	0.47
% Skips			100.00	19.92	1.39
Total	27,922	2,963	1,009	201	14

Notes: "All" refers to all V16 applicants who list fewer than 12 courses in the 2007 college admissions. Among them, "FF listed" are the applicants who list at least one full-fee course. "Skips" refers to the applicants who list a full-fee course but do not list the corresponding reduced-fee course at least once. "Payoff-relevant mistakes" refers to the applicants who would have received a different assignment if they had not skipped a reduced-fee course.

2.2.1 Correlation between Applicant's Scores 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 subjectively infeasible for the applicant.[7]

Our leading hypothesis, denoted by H_0, is that (i) applicants may omit subjectively infeasible courses and (ii) there is no systematic pattern in omitting subjectively infeasible courses. In particular, part (ii) means that the likelihood of omitting such a course is independent of the rank of this course in the true preferences of the applicant.

Hypothesis H_0 implies a negative correlation between the probability that an applicant skips a course and applicant's score. The lower the applicant's score, the larger the set of courses with cutoffs above it. As an applicant omits more courses that are subjectively infeasible, she is more likely to omit a reduced-fee course that has a full-fee counterpart in her ROL. That is, she is more likely to be identified as making a skip.

To investigate the relationship between score and skip, we estimate the following empirical model:

$$Skip_i = \alpha + \beta Score_i + Controls_i + \epsilon_i, \tag{1}$$

where $Skip_i = 100$ if applicant i has made at least one skip and zero otherwise.[8] We expect β to be negative.

[7] We formally define the course feasibility later in the theory section. In this section, the subjective feasibility is taken as a perception of an applicant and, therefore, is unobservable to a researcher.

[8] Let $Skip_i \in \{0, 100\}$ is to make the results more readable. If we instead define $Skip_i \in \{0, 1\}$, every estimate will be one percent of the ones reported here.

There are several alternative explanations for the negative relation between $Score_i$ and $Skip_i$:

H_1: Applicants' score is correlated with cognitive abilities. Those with higher cognitive abilities are able to comprehend the mechanism better, reducing the probability of skipping.

H_2: 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 subjectively infeasible from the top of their ROL, so that their feasible courses rank higher.

To test H_1, we will use another measure student ability, different from $score_i$, as a control variable. The important difference between H_2 and H_0 is where the skipped courses are located. H_2 requires that such courses are concentrated at the top of ROL, while H_0 does not impose any such restrictions, as subjectively infeasible courses can be anywhere in the ROL.

We report the results for model (1) in Table 2. The sample of all odd-numbered regressions are V16 applicants who rank fewer than 12 courses; all even-numbered regressions exclude from this sample the applicants who do not rank a FF course. Control variables include school fixed effects and application demographics, e.g., applicants' gender, median income (in logarithm) in the postal code in which the applicant resides, citizenship status, region born, and language spoken at home. The number of FF courses listed may have a "mechanical" effect on the number of skips. If FF course is not listed, then no RF course can be skipped, by definition. Hence, the more are FF courses listed, the larger are the opportunities for skipping. We include eleven dummy variables that correspond to the number of full-fee courses listed by an applicant.

Columns (1) and (2) are baseline regressions showing the negative relation between skips and scores. To account for H_1, 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 months before they sit their final high school exams. The test is similar to the SAT/ACT tests in the U.S., except that 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 different from score. The results are reported in columns (3) and (4) of Table 2. The coefficient on GAT is negative and significant, suggesting that H_1 is valid and the cognitive abilities may play a role in the explanation of mistakes. Yet, after controlling for cognitive ability, $Score_i$ continues be negatively related to $Skip_i$, consistent with H_0.

Regressions (5) and (6) in Table 2 control for high 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 indicator that the applicant attends a school that charges more than AUD11,000 (approx. USD8,000) in fees.[9] Applicants from private schools respond to change in their scores more, but the results for the overall population remain the same.

To address H_2, we exploit the difference in predictions of the location of skipped courses in

[9]We cannot include the dummy variable "School fees" alone, because of the inclusion of high school fixed effects in all regressions.

Table 2: Probability of Skipping a Reduced-Fee Course

	(1)	(2)	(3)	(4)	(5)	(6)
Score	-0.06***	-0.71***	-0.04***	-0.55***	-0.04***	-0.56***
	(0.01)	(0.06)	(0.01)	(0.08)	(0.01)	(0.07)
GAT			-0.05***	-0.35***	-0.04***	-0.33***
			(0.01)	(0.12)	(0.01)	(0.10)
School fees × *Score*					-0.03***	-0.05**
					(0.01)	(0.02)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
N	26325	2766	26325	2766	26325	2766
R^2	0.37	0.29	0.37	0.30	0.36	0.17

Notes: The dependent variable in every regression is *Skip*; *Skip* = 100 if at least one course is skipped, 0 otherwise. The sample includes all V16 applicants who list fewer than 12 courses, with non-missing GAT results and the information on the applicant's high school. Columns (1), (3), and (5) include all such applicants, while (2), (4), and (6) focus on applicants listing at least one full-fee course. Other control variables include gender, postal code median income (in logarithm), citizenship status, region born, language spoken at home, high school fixed effects, and dummy variables for the number of full-fee courses. Standard errors clustered at high school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

applicant's ROL. As mentioned, H_2 predicts that courses will be skipped from the top of a ROL, while H_0 does not place any such restriction.

Consider two applicants, who are identical except that i skips and j does not. H_2 generates the following testable predictions.

(i) Suppose that H_2 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.

(ii) Suppose that H_2 does not hold and i skips courses from anywhere in i' 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.

Based on these predictions, we use the following regression model to test H_2:

$$\textit{Cutoffs top-ranked courses}_i - \textit{Cutoffs bottom-ranked courses}_i = \gamma + \delta \frac{Skip_i}{100} + \zeta Score_i + Controls_i + \epsilon_i. \quad (2)$$

We expect the coefficient δ to be negative if H_2 holds. The results are presented in Table 3. We use three different definitions for *Cutoffs top-ranked courses*$_i$ and *Cutoffs bottom-ranked courses*$_i$. In regressions (1), (4) and (7), we take the difference between the cutoffs of the top-ranked and the bottom-ranked courses for an individual applicant; in regressions (2), (5) and (8), we take the difference between the average of the two highest-ranked 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 FF course, and in (7)–(9), we use the full sample. Finally, in (1)–(3) we only control on score and gender, while in (4)–(9) we control for the length of ROL and use our standard battery of controls.

The table shows that the coefficient on skip is insignificant in any of the regressions, indicating that there is no evidence that applicants eliminate high-cutoff courses from the top of their ROL. Thus, there is no support for H_2 that applicants attempt manipulations.

Table 3: Correlation between Skip and the Difference between Cutoffs of Top- and Bottom-Ranked Courses

	Applicants Listing at Least One FF Course						All applicants		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Skip/100	-1.38	-0.41	0.09	0.06	0.32	0.69	0.31	0.31	0.60
	(1.11)	(0.93)	(1.02)	(1.18)	(1.00)	(1.06)	(0.97)	(0.81)	(0.90)
Score	0.08**	0.08***	0.07**	0.06*	0.07**	0.07**	0.13***	0.11***	0.09***
	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)
Female	-2.70**	-3.15***	-2.65**	-2.55**	-2.92***	-2.30**	-2.66***	-2.45***	-2.07***
	(1.20)	(1.13)	(1.10)	(1.18)	(1.12)	(1.09)	(0.38)	(0.33)	(0.32)
ROL Length				1.17***	0.90***	1.02***	1.04***	0.97***	1.04***
				(0.21)	(0.21)	(0.24)	(0.06)	(0.06)	(0.08)
Other Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
N	2825	2598	2080	2797	2570	2055	26882	23567	17687
R^2	0.18	0.18	0.25	0.21	0.21	0.28	0.06	0.07	0.07

Notes: The dependent variable of all regressions is the difference between the cutoffs of top- and bottom-ranked courses, but it varies in the number of courses we consider. Columns (1), (4), and (7) use the top- and the bottom-ranked courses; columns (2), (5) and (8) use the top two and the bottom two; and columns (3), (6), and (9) use the top three and the bottom three courses. Correspondingly, the samples in (1), (4) and (7) exclude applicants listing less than two reduced-fee courses; (2), (5), and (8) exclude applicants listing less than four reduced-fee courses; (3), (6) and (9) exclude applicants listing less than six reduced-fee courses. The samples in (7) – (9) are based on all V16 applicants listing less than 12 courses and (1) – (6) are based on V16 applicants who list less than 12 courses with at least one full-fee course. Other controls include postal code income (in logarithm), citizenship status, region born, language spoken at home, and dummy variables for the number of FF courses. Standard errors clustered at high school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

2.2.2 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 vary systematically 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 those mistakes are random and none of the coefficients are significant. Due to the sample size, we use the upper bound definition of payoff-relevance: the mistake is considered to be payoff-relevant if an applicant skips reduced-fee course and is eligible for that course.

Our empirical model is

$$Payoff\text{-}relevant\ Mistake_i = \theta + \iota Score_i + Controls_i + \epsilon_i, \tag{3}$$

where $Payoff\text{-}relevant\ Mistake_i = 100$ if i's skip is payoff-relevant, 0 otherwise.

Table 4: Probability of Making Payoff-Relevant Mistakes

			Sub-sample including applicants who			
	All applicants		List Full-Fee Course		Skip	
	(1)	(2)	(3)	(4)	(5)	(6)
Score	0.01***	0.01***	0.16***	0.19***	0.66***	0.61***
	(0.00)	(0.00)	(0.04)	(0.04)	(0.15)	(0.15)
GAT	0.00	0.00	0.00	0.01	0.31	0.28
	(0.01)	(0.01)	(0.06)	(0.06)	(0.19)	(0.19)
School fees × *Score*		0.01		-0.11*		0.22
		(0.01)		(0.06)		(0.27)
Other controls	Yes	Yes	Yes	Yes	Yes	Yes
N	26325	26325	2766	2766	947	947
R^2	0.14	0.14	0.25	0.25	0.48	0.48

Notes: The dependent variable is equal to 100 if an applicant makes at least one payoff-relevant mistake, and 0 otherwise. The sample includes all V16 applicants who list less than 12 courses, with GAT results and the information on the applicant's high school. Columns (1) and (2) include all such applicants; columns (3) and (4) include only those who list at least one full-fee course; (5) and (6) include only those who make at least one skip. Other control variables include gender, postal code income (in logarithm), citizenship status, region born, language spoken at home and dummy variables for the number of full-fee courses. Standard errors clustered at high school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Table 4 we present the results. Note that the two variables that have been significant in regressions for skips, GAT and interaction of score and school fees, are no longer significant. Combining the results for regression equations (1) and (3), reported in Tables 2 and 4 respectively, it appears that while higher-cognitive-ability applicants, measured by GAT, are more successfully to avoid payoff-irrelevant mistakes than others, lower-cognitive-ability applicants do not make more payoff-relevant mistakes. Similarly, attending expensive private school is not correlated with the likelihood of payoff-relevant mistake either. Another notable observation is that score now has a positive and significant coefficient. This may be explained mechanically: applicants with higher scores have more feasible RF courses, hence skipping a RF course is more likely to be payoff-relevant.

2.2.3 Changes in ROL Over Time

To further test our hypothesis that skips are the outcomes of skipping infeasible courses, we use an unusual feature of the Victorian centralized mechanism: a requirement that applicants submit their "preliminary" ROL several months before the deadline of their final ROL and before

applicants learn their scores. If not changed, the preliminary ROL becomes final and is used for the admissions. 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 the mechanism. A 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 necessarily affect the number of payoff-relevant mistakes. With two types of additional information, 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:

$$\Delta(\#Skips_i) = \tau^s + Demeaned_Controls_i + \epsilon_i, \tag{4}$$

$$\Delta(\#Payoff\text{-}relevant\ Mistakes_i) = \tau^m + Demeaned_Controls_i + \epsilon_i, \tag{5}$$

where $\Delta(\#Skips_i)$ is the differences between the number of skips in the final and the preliminary ROLs and $\Delta(\#Payoff\text{-}relevant\ Mistakes_i)$ is the analogous difference for payoff-relevant mistakes. The control variables are been demeaned, and therefore the constants, τ^s and τ^m, capture the average changes over time. According to our hypothesis, the constant τ^s in model (4) could be either positive or negative but the constant τ^m in model (5) must be negative.

In Table 5 we present the results. Odd-numbered regressions are for payoff-relevant mistakes and even-numbered regressions are for skips. Columns (1) and (2) only control for gender and income. As the number of listed FF courses may have a mechanical effect on skips and mistakes, we add these controls in columns (3) and (4). In all specifications 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 (columns 2 and 4) is positively significant.

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 scores, where the expected score is calculated using GAT (see Appendix A.3 for the definition of expected score). We then investigate the change in the number of instances of skips and payoff-relevant mistakes using the following models:

$$\Delta(\#Skips_i) = \phi^s + \chi^s Shock_i + Controls_i + \epsilon_i \tag{6}$$

$$\Delta(\#Payoff\ relevant\ Mistakes_i) = \phi^m + \chi^m Shock_i + Controls_i + \epsilon_i. \tag{7}$$

We report results in Table 6. Regressions (1) and (3) are for payoff-relevant mistakes, and regressions (2) and (4) are for skips. Control variables are gender and income in regressions (1) and (2), with added control for the change in the number of full-fee courses. The results do not

Table 5: Skips and Payoff-Relevant Mistakes: Changes over Time

	#Mistakes (1)	#Skips (2)	#Mistakes (3)	#Skips (4)
Constant	-0.13***	1.12***	-0.19***	0.81***
	(0.00)	(0.00)	(0.01)	(0.02)
Change in # FF courses			8.32***	43.55***
			(1.22)	(2.76)
Female, demeaned	0.21	1.04**	0.10	0.46
	(0.18)	(0.51)	(0.17)	(0.40)
ln(Income), demeaned	0.18	1.94	0.00	1.02
	(0.34)	(1.50)	(0.34)	(1.09)
N	27654	27654	27654	27654
R^2	0.02	0.04	0.13	0.41

Notes: "Mistake" means a payoff-relevant mistake. "FF courses" means full-fee courses. The dependent variable in regression (1) and (3) is the difference in the number of payoff-relevant mistakes between the final ROL and the preliminary (November) ROL. The dependent variable in columns (2) and (4) is the difference in the number of skips between the final ROL and the preliminary ROL. The sample includes all V16 applicants listing less than 12 courses. Female and ln(Income) are demeaned. Standard errors clustered at high school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

change if we include more control variables. 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.

2.2.4 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 the following three empirical models:

$$Enroll_i = \psi^e + \omega^{em} \frac{Payoff\text{-}relevant\ Mistake_i}{100} + \omega^{es} \frac{Skip_i}{100} + Controls_i + \epsilon_i \tag{8}$$

$$Defer_i = \psi^d + \omega^{dm} \frac{Payoff\text{-}relevant\ Mistake_i}{100} + \omega^{ds} \frac{Skip_i}{100} + Controls_i + \epsilon_i \tag{9}$$

$$Reject_i = \psi^r + \omega^{rm} \frac{Payoff\text{-}relevant\ Mistake_i}{100} + \omega^{rs} \frac{Skip_i}{100} + Controls_i + \epsilon_i \tag{10}$$

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

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

	#Mistakes	#Skips	#Mistakes	#Skips
	(1)	(2)	(3)	(4)
Shock to *Score*	-0.02***	-0.02	-0.01**	0.00
	(0.01)	(0.02)	(0.01)	(0.02)
Change in # FF courses			8.32***	43.55***
			(1.21)	(2.76)
Female	0.25	1.08**	0.13	0.46
	(0.18)	(0.52)	(0.17)	(0.40)
ln(Income)	0.16	1.92	-0.01	1.03
	(0.34)	(1.50)	(0.34)	(1.09)
N	27637	27637	27637	27637
R^2	0.02	0.04	0.13	0.41

Notes: "Mistake" means a payoff-relevant mistake. "FF courses" means full-fee courses. The dependent variable in regression (1) and (3) is the difference in the number of payoff-relevant mistakes between the final ROL and the preliminary (November) ROL. The dependent variable in regressions (2) and (4) is the difference in the number of skips between the final ROL and the preliminary ROL. Shock to *Score* is the difference between realized and expected scores. The sample includes all V16 applicants with GAT who list fewer than 12 courses. Other control variables are gender and postal code income (in logarithm). Standard errors clustered at high school level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. *** $p < 0.01$.

Table 7: Strategic Mistakes and Enrollment Decision

	Enroll	Defer	Reject
	(1)	(2)	(3)
Payoff-relevant Mistake/100	-15.75***	13.06***	2.69
	(3.81)	(3.34)	(2.94)
Skip/100	-4.07**	4.76***	-0.69
	(1.99)	(1.57)	(1.73)
Score	0.38***	0.12***	-0.51***
	(0.02)	(0.01)	(0.02)
Other controls	Yes	Yes	Yes
N	23774	23774	23774
R^2	0.17	0.13	0.18

Notes: The dependent variable in column (1) is equal to 100 if an applicant enrolls at the assigned course; in (2) is equal to 1 if an applicant defers enrollment; in (3) is equal to 1 if an applicant rejects the assigned course. Every applicant who makes a payoff-relevant mistake also makes a skip; hence the coefficient on payoff-relevant mistakes is its marginal effect. The sample includes all V16 applicants who list less than 12 courses and who enroll/defer/reject one of the offered courses; that is, it excludes all applicants who received an offer from any irregular process. Control variables include high school fixed effects, gender, postal code income (in logarithm), citizenship status, region born, language spoken at home, being assigned to full-fee course, the number of full-fee courses in ROL, ROL length, and the rank of offered course in ROL. 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 the 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.[10]

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*, or *colleges*, $C = \{c_1, ..., c_C\}$ and a set of *applicants*. Each applicant has a type $\theta = (u, s)$, where $u = (u_1, ..., u_C) \in [\underline{u}, \overline{u}]^C$ is a vector of von-Neumann Morgenstern utilities of attending colleges for some $\underline{u} \leq 0 < \overline{u}$, and $s = (s_1, ..., s_C) \in [0, 1]^C$ is a vector of scores representing the colleges' preferences or applicants priorities at colleges, with an applicant with a higher score having a higher priority at a college. A vector u induces an ordinal preferences over colleges, denoted by a rank-ordered list (ROL) of "acceptable" colleges, $\rho(u)$, of length $0 \leq \ell \leq C$.[11] Assuming that an applicant has an outside option of zero payoff, the model allows for the possibility that applicants may find some colleges unacceptable. Let $\Theta = [\underline{u}, \overline{u}]^C \times [0, 1]^C$ denote the set of applicant types. One special case is **serial dictatorship** in which colleges' preferences for applicants are given by a single score. Australian tertiary admissions can be seen as a case of serial dictatorship. We shall incorporate this case by an additional restriction that the scores of each applicant satisfy $s_1 = ... = s_C$.

A *continuum economy* consists of the same finite set of colleges and a unit mass of applicants with type $\theta \in \Theta$ and is given by $E = [\eta, S]$, where η is a probability measure over Θ representing the distribution of applicant population over types, and masses of seats $S = (S_1, ..., S_C)$ available at the colleges, where $S_i > 0$ and $\sum_{i=1}^{C} S_1 < 1$. We assume that η 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; applicants' scores are one-dimensional number in $[0, 1]$. The altomlessness ensures that indifferences either in applicant preferences or

[10] There is a growing number of work on large matching markets. Pittel (1989), Immorlica and Mahdian (2005), Kojima and Pathak (2009), Lee (2014), Lee and Yariv (2014), Ashlagi, Kanoria, and Leshno (forthcoming), Che and Tercieux (2015a), Che and Tercieux (2015b), Abdulkadiroglu, Che, and Yasuda (2015), Che and Kojima (2010), 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.

[11] In 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.

in college preference arises only for a measure 0 set of applicants.[12] The full-support assumption means that both applicants' and colleges' preferences are rich (except for the case of serial dictatorship). A *matching* is defined as a mapping $\mu : C \cup \Theta \to 2^{\Theta} \cup (C \cup \Theta)$ satisfying the usual two-sidedness and consistency requirements as well as "open on the right" defined in AL (see p. 1241). A *stable matching* is also defined in the usual way satisfying individual rationality and no-blocking.[13]

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 college i is given by the measure of applicants whose favorite college 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 college i. Given the full-support assumption, Theorem 1-i of AL guarantees a unique market clearing price P^* and a unique stable matching μ^*. Given the continuous density assumption, $D(\cdot)$ is C^1 and $\partial D(P^*)$ is invertible.

With the continuum economy E serving as a benchmark, we are interested in a sequence of finite random economies approximating E in the limit. Specifically, let $F^k = [\eta^k, S^k]$ be a *k-random economy*, which consists of k applicants each with type θ drawn independently according to η, and the vector $S^k = [k \cdot S]/k$ of capacity per applicant, 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 an applicant proposing DA being employed to assign applicants to colleges. We assume that colleges are acting passively reporting their preferences and capacities truthfully.[14] We are interested in characterizing an equilibrium behavior of the applicants in the DA. In one sense, this is trivial: since DA is stratetgyproof, it is a weak dominant strategy for each player to rank order all acceptable colleges (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 applicant observes his own type θ but not the types of other applicants; as usual, all applicants understand as common knowledge the structure of the game. Given this, the DA induces a Bayesian game in which the strategy of each applicant specifies a distribution over ROLs of length no greater than C as a function of his type θ.

[12] At the same time, atomlessness rules out the environments where some applicants are ranked the same at some schools and lotteries are used to break ties, such as NYC's high school admissions Abdulkadiroglu, Agarwal, and Pathak (2015).

[13] Individual rationally requires that no participant (an applicant or a college) is assigned a partner that is not acceptable. No blocking means that no applicant-college pair exists such that the applicant prefers the college over her assignment and the college has either a vacant position or admits an applicant it ranks below an applicant in the applicant-college pair.

[14] In the specific context of VTAC mechanism, the common preferences make this an ex post equilibrium strategy. See Che and Koh (2016).

In any game (either the limit or the k-random economy), applicant i's Bayesian strategy is a measurable function $\sigma_i : \Theta \to \Delta(\mathcal{R})$, where $\mathcal{R}$ is the set of all possible ROLs an applicant can submit. Note that the strategies can be asymmetric; i.e., we do not restrict attention to symmetric equilibria. In any economy (either continuum or finite), a profile of strategies by students induce cutoffs $P \in [0,1]^C$. We say a college c is *feasible* to an applicant if his score at c is no less than P_c. And we say an applicant *demands* college c if c is feasible and he ranks c in his ROL ahead of any other feasible colleges.

In the sequel, we are interested in the following solution concept:

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

This solution concept is arguably more sensible than the exact Bayesian Nash equilibrium, if for a variety of reasons 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.

Indeed, without relaxing a solution concept in some way, one cannot explain the kind of departure from the dominant strategy observed in the preceding section. To see this, recall that in the continuum economy, any stable matching, and hence the outcome of DA mechanism, gives rise to cutoffs that are degenerate. This means that the students participating in the DA face no uncertainty with regard to the feasible set of colleges. Hence, one can easily construct a dominated strategy that can do just as well as a dominant strategy—namely the truthful rank-ordering of colleges. For instance, a student can list only one college—the most preferred among those whose cutoffs are below his/her scores—and do just as well as reporting her true ROL. But such a strategy will not be optimal for any finite economy, no matter how large it is. In a finite economy, the cutoffs are not degenerate, so any strategy departing from the truth-telling will result in payoff loss with positive probability. Hence, in order to explain a behavior that departs from a dominant strategy, one must relax exact optimality on the agents' behavior. At the same time, a solution concept cannot be arbitrary, so some discipline must be placed on the extent to which payoff loss is tolerated. The robust equilibrium concept fulfills these requirements: it allows participants to make some mistakes but insists that the payoff loss from the mistakes disappear as the market grows arbitrarily large.

3.2 Analysis of Robust Equilibria

Robust equilibria require that participants should make no mistakes with any real payoff consequences as the market gets large. Does this mean that a large fraction of agents must report

truthfully in a strategy-proof mechanism? We show below that this need not be the case. Specifically, we will construct strategies that deviate from TRS and yet do not entail significant payoff loss in a large economy.

To begin, 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 college given P^* (i.e., he ranks that colleges ahead of all other feasible colleges). The set of SRSs is typically large. For example, suppose $C = \{1, 2, 3\}$, and colleges 2 and 3 are feasible, and an applicant prefers 2 to 3. Then, 7 ROLs—1-2-3, 2-3-1, 2-1-3, 1-2, 2-1, 2-3, 2—constitute his SRSs out of 10 possible ROLs he can choose from. Formally, if an applicant has $\ell \leq C$ feasible colleges, then the number of SRSs is $\sum_{a \leq \ell-1, b \leq C-\ell} \binom{a+b+1}{b} a!b!$. For each type $\theta = (u, s)$ with $\rho(u) \neq \emptyset$ (i.e., with at least one acceptable college), there exists at least one SRS that is untruthful.[15]

For the next result, we construct such a strategy. To begin, let $\hat{r} : \mathcal{R} \times [0,1]^C \to \mathcal{R}$ be a transformation function that maps a preference order $\rho \in \mathcal{R}$ 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 college ahead of all other feasible colleges 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 $\hat{R} : \Theta \to \mathcal{R}$, given by $\hat{R}(u, s) := \hat{r}(\rho(u), s)$, for all $\theta = (u, s)$.[16] 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 δ-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 applicants with types $\theta \in \Theta^\delta$ to play TRS, and for all applicants with types $\theta \notin \Theta^\delta$ to randomize between TRS with probability γ and untruthful SRS $\hat{R}(\theta)$ with probability $1 - \gamma$ in each k-random economy.*

The intuition for Theorem 1 rests on the observation that the uncertainty on cutoffs, and thus feasible set of colleges, vanishes in a sufficiently large economy, which eliminates the payoff-risk of playing non-TRS. Specifically, the sequence of strategy profiles we construct satisfies two properties: (a) prescribes a large fraction of agents to deviate from TRS with a large enough probability and yet (b) it gives rise to cutoffs P^* in the limit continuum economy, namely the cutoffs that would prevail if *all* agents played TRS. That these two properties can be satisfied simultaneously is not trivial and requires some care, since the cutoffs may change as students deviate from TRS. Indeed, the feature that all agents play TRS with some small probability γ is designed to ensure that the same unique stable matching obtains under the prescribed strategies. Given these facts, the well-known limit theorem, due to Glivenko and Cantelli, implies that the (random) cutoffs for any

[15] If an applicant's most favorite college is infeasible, she can drop that college. If is feasible, then she can drop an acceptable college below or add an unacceptable college below, whichever exists.

[16] 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.

large economy generated by the i.i.d. sample of individuals are sufficiently concentrated around P^* under the prescribed strategies that students whose scores are δ away from P^* will suffers very little payoff loss from playing any SRS that deviates (possibly significnatly) from TRS. Indeed, our construction ensures that these are precisely the students who play SRS that deviate from TRS. Since (δ, γ) is arbitrary, the following striking conclusion emerges.

COROLLARY 1. *There exists a robust equilibrium in which each applicant whose TRS is a non-empty list of colleges submits an untruthful ROL with probability arbitrarily close to one.*

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.

If strategic mistakes of the types observed in the preceding section undermine the prediction of TRS, do they also undermine the stability of the outcome? This is an important question on two accounts. If mistakes jeopardize stability in a significant way, then this may call into question the rationale for DA, to the extent that mistakes do occur. If stability remains largely intact despite the presence of mistakes, then they do not raise a fundamental concern.

Aside from the stability prediction, the question is important from an empirical identification perspective. Stability has been an important identification assumption invoked by empirical researchers for preference estimation in a number of contexts, e.g., in decentralized two sided matching (e.g., see Fox (2009) and Chiappori and Salanié (2016) for surveys) as well as centralized matching (e.g., Agarwal (2015)).

Our second theorem shows that strategic mistakes captured by robust equilibrium leaves the stability property of DA largely unscathed.

DEFINITION 2. *For a sequence $\{F^k\}$ of k-random economies with DA matching, the associated sequence $\{(\sigma^k_{1\leq i\leq k})\}_k$ 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 applicants are assigned their most preferred feasible colleges given the equilibrium cutoffs P^k.*

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

THEOREM 2. *Any regular robust equilibrium is asymptotically stable.*

3.3 Discussion

Our setting assumes that colleges rank applicants strictly by some score and that applicants know their scores before playing the college admissions game. This ensures that the probability of an

applicant being accepted by a college is degenerate as the market grows. By imposing this condition, we thus exclude school choice problem in which a lottery is often used to rank applicants after they have submitted their ROLs (see Pathak (2011) for a survey and Abdulkadiroglu, Agarwal, and Pathak (2015) for an study on high school choice in NYC).

The comparison between stability and truthful reporting strategies is also considered in Fack, Grenet, and He (2015). They consider exact Bayesian Nash equilibrium and argue that truthful-reporting strategy may not be the unique equilibrium and may not be equilibrium at all if there is application cost. By doing so, they are able to justify certain deviations from truthful reporting but still have difficulty in accommodating the mistakes in our data. Given that one has ranked a FF course and that there is uncertainty in the cutoffs, it is sub-optimal to skip the corresponding RF course; skipping cannot be justified by application cost either, because the cost of ranking a RF course is plausibly zero in this scenario. In contrast, our Theorem 1 provides a natural explanation for the mistakes.

A more significant difference between the two papers is manifested in Theorem 2. To show that stability can be used as an identifying condition in empirical studies, Fack, Grenet, and He (2015) show the existence of a sequence of finite economies and the associated sequence of Bayesian Nash equilibria leading to asymptotically stable matching outcomes. Their result, however, does not rule out asymptotically unstable equilibrium outcomes even in large market, nor do they identify sufficient conditions for asymptotic stability. Going beyond mere existence, our Theorem 2 shows that all regular robust equilibrium is asymptotically stable, which is also technically more challenging to prove. Theorem 2 thus further justifies the use of stability for identification.

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 $I = 1800$ applicants compete for admission to $C = 12$ colleges. The vector of college capacities is specified as follows:

$$\{S_c\}_{c=1}^{12} = \{150, 75, 150, 150, 75, 150, 150, 75, 150, 150, 75, 150\}.$$

Setting the total capacity of colleges (1500 seats) to be strictly smaller than the number of applicants (1800) ensures that each college has a strictly positive cutoff in equilibrium.

The economy is located in an area within a circle of radius 1 as in Figure C1 (Appendix C) which plots one simulation sample. The colleges (represented by big red dots) are evenly located on a circle of radius 1/2 around the centroid; the applicants (represented by small blue dots) are uniformly distributed across the area. The cartesian distance between applicant i and college c is denoted by $d_{i,c}$.

Applicants are matched with colleges through a serial dictatorship. They are asked to submit a rank-ordered list of colleges, and there is no limit on the number of choices to be ranked. Without loss of generality, colleges have a priority structure such that applicant i is ranked higher by all colleges 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 applicants' preferences over colleges, we adopt a parsimonious random utility model without outside option. Applicant i's utility from being matched with college c is specified as follows:

$$u_{i,c} = \beta_1 \cdot c - d_{i,c} + \beta_2 T_i A_c + \beta_3 Small_c + \sigma \epsilon_{i,c}, \forall i \text{ and } c; \tag{11}$$

where $\beta_1 \cdot c$ is college c's "baseline" quality;[17] $d_{i,c}$ is the distance from applicant i's residence to college c, with a coefficient normalized to -1; $T_i = 1$ or 0 is applicant i' type (e.g., disadvantaged or not, or arts versus sciences); $A_c = 1$ or 0 is college c's type (e.g., known for resources for disadvantaged applicants or art education); $Small_c = 1$ if college c is small, 0 otherwise; and $\epsilon_{i,c}$ is a type-I extreme value; σ, a positive number, allows the variance of utility shocks to be flexible.

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

The coefficients of interest are $(\beta_1, \beta_2, \beta_3, \sigma)$ which are fixed at $(0.3, 2, 0, 1)$ in simulations. By this specification, $Small_c$ does not affect applicant preference. 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 applicant preference profile obtained by randomly drawing $\{d_{i,c}, \epsilon_{i,c}\}_c$ and T_i for all i from the distributions specified above. In all samples, applicant priorities, college capacities, and college types (A_c) are kept constant.

We first simulate the joint distribution of the 12 colleges' admission cutoffs. The simulation lets every applicant submit an ROL ranking all colleges truthfully. After running the serial dictatorship, we calculate the admission cutoffs in each simulation sample. Figure C2 in Appendix C shows the marginal distribution of each college's cutoff from the 1000 samples. Note that colleges with smaller capacities tend to have higher cutoffs. For example, college 11, with 75 seats, in general has the highest cutoff, although the highest baseline quality college is college 12, with 150 seats.

[17] Results do not change when we allow for college fixed effects. We also add a large positive constant to make all $u_{i,c} > 0$, which implies that all schools are acceptable to all applicants.

To generate data on applicant behaviors and matching outcomes for preference estimation, we simulate another 200 samples with new independent draws of $\{d_{i,c}, \epsilon_{i,c}\}_c$ and T_i for all i. 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 applicant strategies.

(i) **STT (Strict Truth-Telling)**: Every applicant submits a rank-ordered list of 12 colleges according to her true preferences. Recall that everyone finds every college acceptable, and therefore STT is equivalent to TRS as in our theoretical model (section 3).

(ii) **IRR (Payoff Irrelevant Skips)**: A fraction of applicants skip colleges with which they are never matched given the simulated distribution of cutoffs. For a given applicant, a dropped college can have a high (expected) cutoff and thus be "out of reach;" it may also be a college that has a low cutoff, but the applicant is always accepted by one of her more-preferred colleges. To specify the fraction of skippers, we first randomly choose about 21.4 percent of the applicants to be never-skippers who always rank all colleges truthfully. Applicants with $T_i = 1$ are more likely to skip: 95 percent of them are potential skippers (Table C3), compared with 70 percent of those with $T_i = 0$ (Table C4). Among the potential skippers, we consider three scenarios. In **IRR 1**, around one third of them skip all the "never-matched" colleges; **IRR 2** adds another one-third; and **IRR 3** lets all of them skip. When doing so, it is possible that an applicant drops all colleges, because some applicants are never matched. We randomly choose a college for them, so that they submit one-college ROLs.

(iii) **REL (Payoff Relevant Mistakes)**: **In addition to IRR3**, i.e., given all the potential skippers have skipped the never-matched colleges, we now let them make payoff relevant mistakes. That is, they skip some of the colleges 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 colleges with which they are matched with probabilities lower than the threshold. From REL 1 to REL 4, the thresholds are 7.5, 15, 22.5, and 30 percent, respectively. It should be emphasized that the cutoff distribution is not re-computed in any of these DGPs.

To summarize, for each of the 200 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 applicants skip in the simulations. The reported percentages are averaged over the 200 samples.

4.3 Identifying Conditions and Estimation

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

Table 8: Skips and Mistakes in Monte Carlo Simulations (Percentage Points)

	Scenarios (Data Generating Processes)							
	Strict Truth-telling	Payoff Irrelevant Skips			Payoff Relevant Mistakes			
	STT	IRR 1	IRR 2	IRR 3	REL 1	REL 2	REL 3	REL 4
WTT: *Weak Truth-Telling*[a]	100	85	69	53	52	52	52	52
Matched w/ favorite feasible college[b]	100	100	100	100	98	95	93	90
Skippers[c]	0	25	53	79	79	79	79	79
By number of skips:								
Skipping 11 colleges	0	17	37	55	67	70	73	74
Skipping 10 colleges	0	6	11	17	10	8	6	4
Skipping 9 colleges	0	2	4	6	1	1	0	0
Skipping 8 colleges	0	0	0	0	0	0	0	0
Skipping 7 colleges	0	0	0	0	0	0	0	0
Skipping 6 colleges	0	0	0	0	0	0	0	0
Skipping 5 colleges	0	0	0	0	0	0	0	0
Skipping 4 colleges	0	0	0	0	0	0	0	0
Skipping 3 colleges	0	0	0	0	0	0	0	0
Skipping 2 colleges	0	0	0	0	0	0	0	0
Skipping 1 college	0	0	0	0	0	0	0	0
STT: *Strict Truth-telling*	100	75	47	21	21	21	21	21
Reject WTT: Hausman Test[d]	5	10	60	100	97	95	91	87

Notes: This table presents the configurations of the eight data generating process (DGPs). Each entry is a percentage averaged over the 200 simulation samples. Tables C3 and C4 further show the breakdown by T_i. [a]An applicant is "*weak truth-telling*" if she truthfully ranks her top K_i ($1 \leq K_i \leq 12$) preferred colleges, where K_i is the observed number of colleges ranked by i. [b]A college is feasible to an applicant, if the applicant's index is higher than the college's *ex post* admission cutoff. If an applicant is matched with her favorite feasible college, she cannot form a blocking pair with any college. [c]Given that every college is acceptable to all applicants and potentially over-demanded, an applicant is a skipper if she does not rank all colleges. [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 detail in section 4.3.

(i) **WTT** or Weakly Truth-Telling. WTT, which can be considered as a truncated version of TRS or STT, entails two assumptions: (a) the observed number of choices ranked in any ROL is exogenous to applicant preferences and (b) every applicant ranks her top preferred colleges according to her preferences (although one may not rank all colleges). 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 applicants are assigned their favorite feasible college given the *ex post* cutoffs, the model can be estimated by MLE based on a conditional logit model where each applicant's choice set is restricted to the *ex post* feasible colleges and where the matched college is the favorite among all her feasible colleges. If every applicant plays a stable response with respect to the cutoffs, this assumption is satisfied. We call this approach the stability estimator.

(iii) **Robust approach**. When there are payoff-relevant mistakes, some applicants are not matched with their favorite feasible college. As a remedy, we construct a hypothetical set

of feasible colleges for each applicant by inflating the cutoffs of all but her matched college. Given the hypothetical set of feasible colleges, every applicant is more likely to be matched with her favorite feasible college. The estimation then is similar to the stability estimator, except the modified feasible sets.

The formulation of likelihood function and estimation results from the three approaches are discussed in Appendix C. In Table 8, across the eight DGPs, the fraction of skippers increases from zero (in STT) to 79 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 applicants who are weakly truth-telling decreases from 85 percent in IRR 1 to 53 percent in IRR 3, stabilizing around 52 percent in all REL DGPs.

In contrast, stability is always satisfied in STT and IRR 1-3, while the fraction of applicants that can form a blocking pair with some college increases from 2 percent in REL 1 to 10 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, following Fack, Grenet, and He (2015). Under the null hypothesis, both WTT and stability are satisfied, while under the alternative only stability holds. When all applicants, except the 17 percent with the lowest priorities, are matched, WTT 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 a 10-100 percent probability. Notice that we only need 47 percent of applicants (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, β_1, and more details are provided in Appendix C, especially Table C5. Figure 1 plots the distributions of the estimator across the 200 simulation samples, with the true value being $\beta_1 = 0.3$. A consistent estimator should have mean equal to 0.3. Recall that all DGPs use the same 200 samples of simulated applicant preference profiles and that what defers across DGPs is how applicants play the game.

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 WTT 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

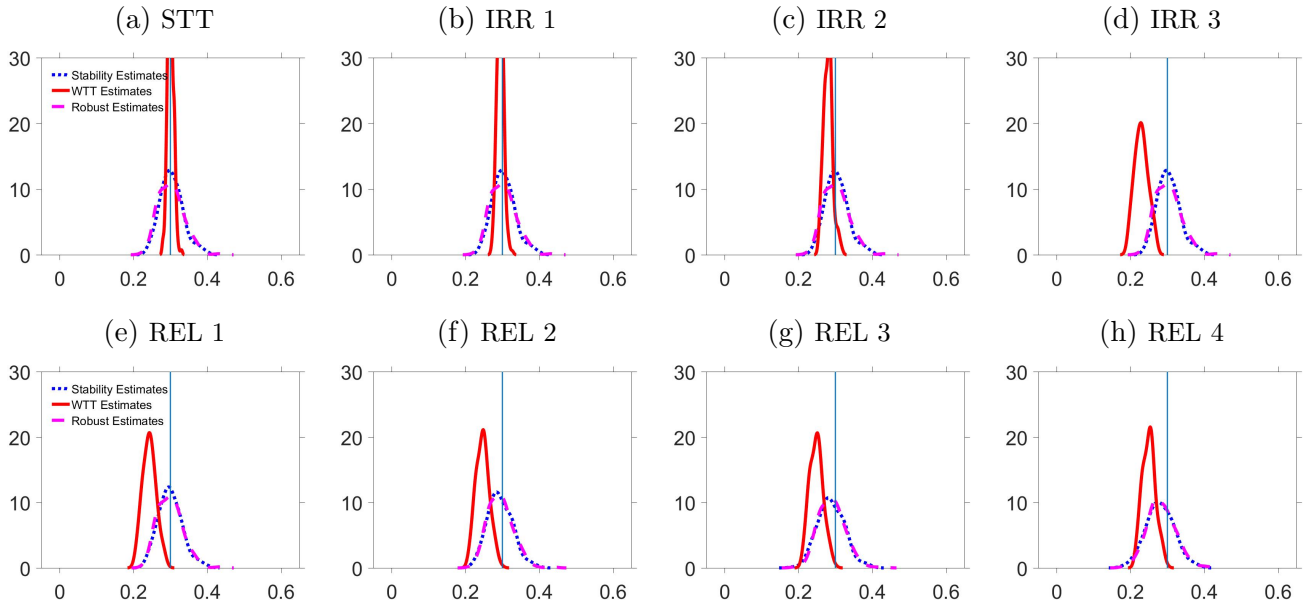


Figure 1: Estimates based on Weak Truth-Telling or Stability (True Value is 0.3)

Notes: The figures focus on the estimates of the quality coefficient (β_1) from two approaches, weakly truth-telling (WTT, red line) and stability (blue dots). The distributions of the estimates across the 200 simulation samples are reported. A consistent estimator should have mean equal to 0.3. Each subfigure uses the 200 estimates from the 200 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 applicant truthfully ranks all colleges; 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.

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 53 percent skippers and 31 percent of applicants violating the truth-telling assumption (IRR 2), the estimates based on WTT from the 200 samples have mean 0.28 (standard deviation 0.01); in contrast, the stability estimates are on average 0.30 (standard deviation 0.03).

The downward bias in the WTT estimator is intuitive. When applicants skip, they omit colleges with which they have almost no chance of being matched. For applicants with low priorities, popular colleges are therefore more likely to be skipped. Whenever a college is skipped, WTT assumes that it is less preferable than all the ranked colleges. Therefore, many applicants are mistakenly assumed to dislike popular colleges, which results in a downward bias in the estimator of β_1. In contrast, this bias is absent in the stability estimator: Whenever a college is skipped by an applicant due to its high expected cutoff, the stability condition does not make this assumption on ranked and skipped colleges.

Figures 1e–h consider the DGPs in which applicants make payoff-relevant mistakes. Neither WTT or stability is satisfied (Table 8), and therefore all estimators are inconsistent. However, the estimators based on stability and the robust approach are still less biased; the means of the

estimates are close to the true value, ranging from 0.28 to 0.30 (see Table C5 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: Applicants 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 colleges 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 colleges.

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

(i) True Preferences (with possible mistakes): We use the true coefficients in utility functions to simulate counterfactual outcomes. Applicants adopt different strategies in each DGP as in our simulation of data (see section 4.2). Specifically, DGP STT requires everyone to submit a truthful 12-college ROL; the potential skippers omit their never-matched colleges in DGPs IRR 1-3; and in DGPs REL 1-4, the skippers additionally omit some colleges with which they have some chance of being matched.

(ii) Submitted ROLs: One assumes that the submitted ROLs are true ordinal preferences and that applicants 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 colleges in true preference order, and therefore applicant preferences can be estimated from the data with WTT as the identifying condition. Under the counterfactual policy, we simulate applicant preferences based on the estimates and let applicants submit truthful 12-college lists.

(iv) Stability Estimates: We estimate applicant preferences from the data with stability as the identifying condition. Under the counterfactual policy, we simulate applicant preferences based on the estimates and let applicants submit truthful 12-college lists.

(v) Robust Estimates: We estimate applicant preferences from the data with the robust approach. Under the counterfactual policy, we simulate applicant preferences based on the estimates and let applicants submit truthful 12-college lists.

When simulating counterfactual outcomes, we use the same 200 simulated samples for estimation. In particular, we use the same simulated $\{\epsilon_{i,c}\}_c$ when constructing preference profiles after preference estimation. Equivalently, we assume that $\{\epsilon_{i,c}\}_c$ are observed and focus on the effects of different estimators of the coefficients in the utility functions.

To summarize, for each of the 200 simulation samples, we conduct 40 different counterfactual analyses: 8 (DGPs: STT, IRR 1-3, and REL 1-4) $\times$ 5 (true preferences and 4 counterfactual approaches: submitted ROLs, truth-telling estimates, stability estimates, and robust estimates).

4.5.1 Performance of the Four Approaches

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

When simulating the outcomes under the counterfactual policy, we assume applicants make mistakes as they do under the current policy. As shown above, matching outcome does not change if applicants make payoff-irrelevant skips, although payoff relevant mistakes would lead to different outcomes.

An informative statistic of a match is the colleges' cutoffs which summarize the joint distribution of applicant priorities and preferences. Figure 2 shows, given each DGP, how the four approaches mis-predict the cutoffs under the counterfactual policy. For each college, indexed from 1 to 12, we calculate the mean of the 200 cutoffs from the 200 simulation samples by using the true preferences and the other four approaches. The sub-figures then depict the mean differences between the predicted cutoffs and the true ones.

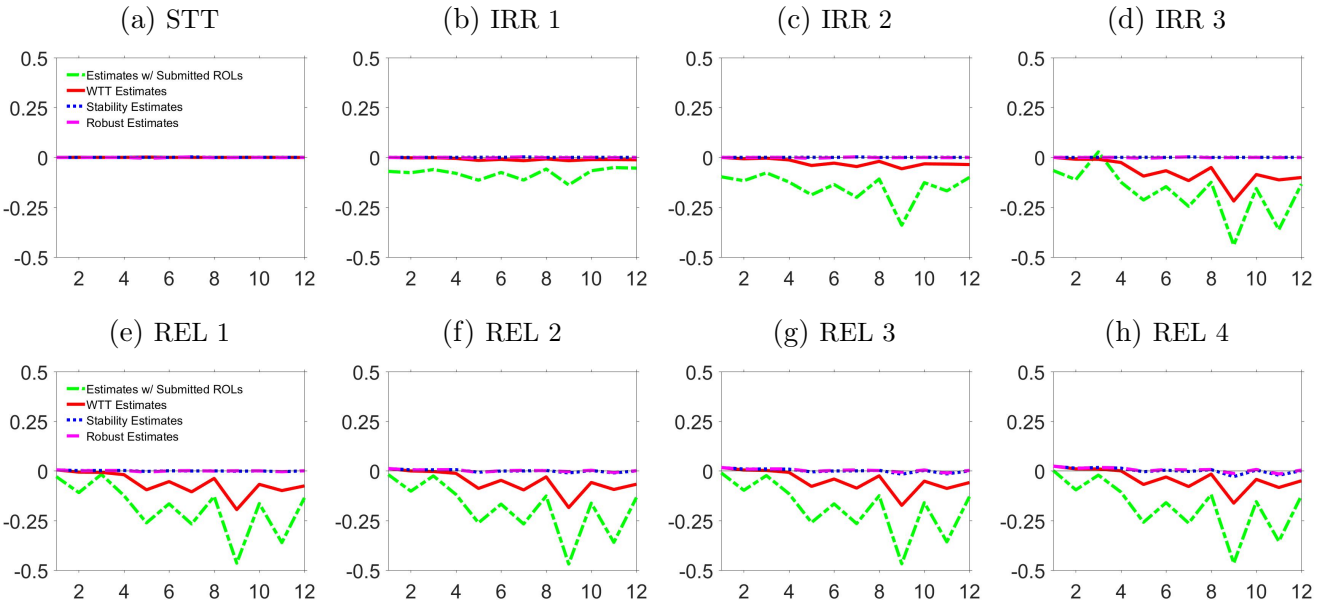


Figure 2: Comparison of the Four Approaches: Biases in Predicted Cutoffs

Notes: The sub-figures present how the predicted cutoffs from each approach differ from the true ones which are simulated based on true preferences with possible mistakes. Each subfigure corresponds to a DGP. Given a DGP, we simulate the colleges' cutoffs following each approach and calculate the mean deviation from the true ones. The X-axis shows the college indices; the Y-axis indicates the deviation of the predicted cutoffs from 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 other three approaches are almost the same.

In Figures 2b–d, corresponding to DGPs IRR 1-3, only the stability and robust 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 applicants tend to omit popular colleges from their lists, both approaches underestimate the demand for these colleges and thus result in under-predicted cutoffs. The bias is even larger for smaller colleges 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 negligible mis-prediction relative to the other two, so do the robust estimates.

(a) Applicants $T_i = 1$ (b) Applicants $T_i = 0$

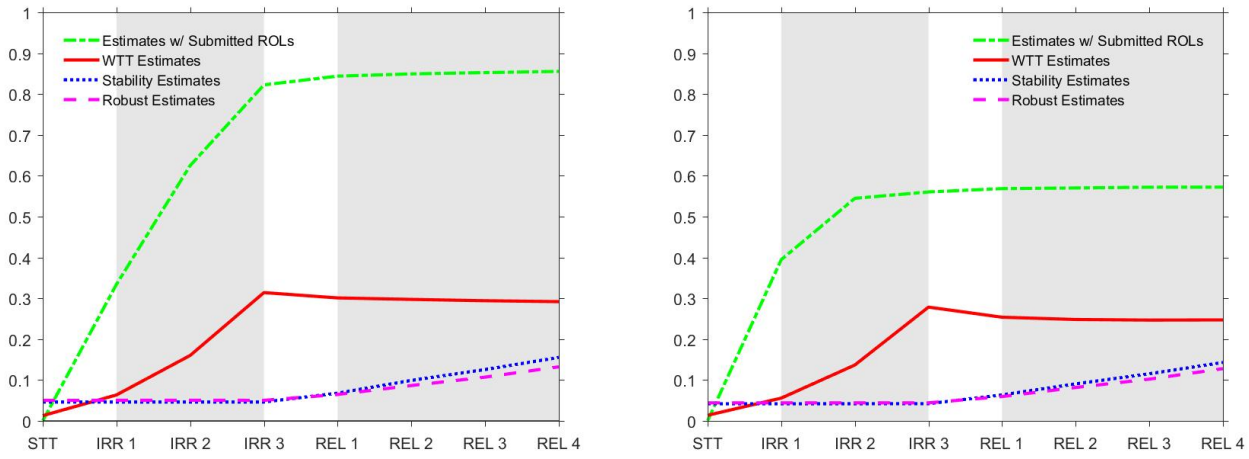


Figure 3: Comparison of the Four Approaches: Mis-predicted Match (Fractions)

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 applicants, $T_i = 1$ and $T_i = 0$.

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

In Figure 3a, among the $T_i = 1$ applicants, 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 8 to 13 percent. The WTT estimates have a lower mis-prediction rate in STT, but under-perform in all other DGPs. The estimates based on submitted ROLs have the highest mis-prediction rates in all DGPs except STT. Lastly, the robust estimates are almost identical to the stability ones in STT and IRR 1-3,

but perform better in REL 1-4. Among the applicants with $T_i = 0$ (subfigure b), the comparison of the four approaches has the same pattern.

(a) Applicants $T_i = 1$: (Fraction Better off)−(Fraction Worse off) (b) Applicants $T_i = 0$: (Fraction Better off)−(Fraction Worse off)

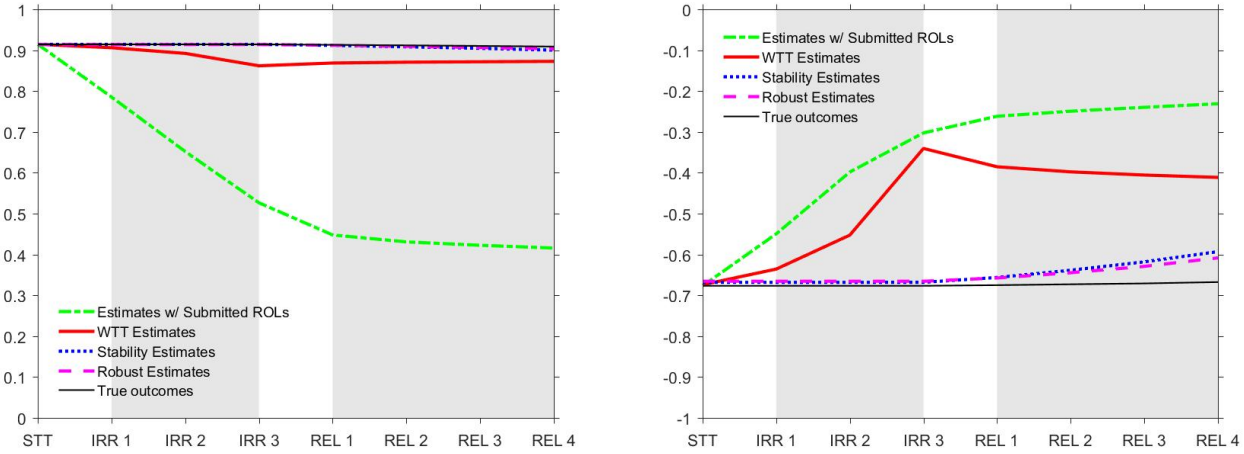


Figure 4: Comparison of the Four Approaches: Mis-predicting Welfare Effects

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 applicants, $T_i = 1$ 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 applicants better off and that of those worse off. The estimated fraction of applicants with $T_i = 1$ being worse off is close to zero in all cases, so is the estimated fraction of applicants with $T_i = 0$ being better off. There are some applicants whose welfare does not change; simulated with true preferences, this fraction is 16 percent among the $T_i = 1$ applicants and 41 percent among the $T_i = 0$ ones. See Tables C6 and C7 in Appendix C for more details.

We now investigate the welfare effects on the $T_i = 1$ applicants 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 applicant under the two policies. If the applicant is matched with a "more-preferred" college 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 applicant preferences in a unique way, applicant's utility associated with a given college is estimated at a different value under each approach. Therefore, measured welfare effects of the counterfactual policy may differ even when an applicant is matched with the same college.

Figure 4 shows the mean difference between the fraction of applicants better off and that of those worse off across the simulation samples.[18] In Figure 4a, among the $T_i = 1$ applicants, the welfare effect predictions based on the stability or robust 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 in DGPs STT and IRR 1; those based on submitted ROLs tend to be biased towards to zero effect when there are more applicants skipping or making mistakes.

[18]There are some applicants whose outcomes do not change. See Tables C6 and C7 in Appendix C for more detailed summary statistics.

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

In summary, estimated welfare effects are biased towards zero for all applicants 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. The robust approach further improves upon the stability estimates when there are payoff-relevant mistakes.

References

Abdulkadiroglu, A., N. Agarwal, and P. A. Pathak (2015): "The Welfare Effects of Coordinated Assignment: Evidence from the NYC HS Match," MIT, unpublished mimeo.

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

Agarwal, N. (2015): "An empirical model of the medical match," *American Economic Review*, 105(7), 1939–1978.

Ashlagi, I., and Y. Gonczarowski (2016): "Stable Mechanisms Are Not Oviously Strategy-proof," .

Ashlagi, I., Y. Kanoria, and J. D. Leshno (forthcoming): "Unbalanced Random Matching Markets," *Journal of Political Economy*.

Azevedo, E. M., and J. W. Hatfield (2012): "Complementarity and Multidimensional Heterogeneity in Matching Markets," mimeo.

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

Che, Y.-K., J. Kim, and F. Kojima (2013): "Stable Matching in Large Economies," mimeo.

Che, Y.-K., and Y. Koh (2016): "Decentralized College Admissions," *JPE*, 124(5), 1295–1338.

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

Che, Y.-K., and O. Tercieux (2015a): "Efficiency and Stability in Large Matching Markets," Columbia University and PSE, Unpublished mimeo.

——— (2015b): "Payoff Equivalence of Efficient Mechanisms in Large Markets," Columbia University and PSE, Unpublished mimeo.

Chiappori, P.-A., and B. Salanié (2016): "The Econometrics of Matching Models," *Journal of Economic Literature*, 54(3), 832–861.

Deb, J., and E. Kalai (2015): "Stability in Large Bayesian Games with Heterogeneous Players," *Journal of Economic Theory*, 157, 1041–1055.

Fack, G., J. Grenet, and Y. He (2015): "Beyond Truth-Telling: Preference Estimation with Centralized School Choice," Paris School of Economics and Rice University, Unpublished mimeo.

Fox, J. T. (2009): "Structural Empirical Work Using Matching Models," *New Palgrave Dictionary of Economics. Online edition.*

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

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

Immorlica, N., and M. Mahdian (2005): "Marriage, Honesty, and Stability," *SODA 2005*, pp. 53–62.

Kojima, F., and P. A. Pathak (2009): "Incentives and Stability in Large Two-Sided Matching Markets," *American Economic Review*, 99, 608–627.

Lee, S. (2014): "Incentive Compatibility of Large Centralized Matching Markets," University of Pennsylvania, Unpublished mimeo.

Lee, S., and L. Yariv (2014): "On the Efficiency of Stable Matchings in Large Markets," University of Pennsylvania, Unpublished mimeo.

Li, S. (2017): "Obviously Strategy-Proof Mechanisms," Harvard, Unpublished mimeo.

Liu, Q., and M. Pycia (2011): "Ordinal Efficiency, Fairness, and Incentives in Large Markets," *mimeo.*

Pathak, P. (2011): "The mechanism design approach to student assignment," *Annual Review of Economics*, 3(1).

Pittel, B. (1989): "The Average Number of Stable Matchings," *SIAM Journal on Discrete Mathematics*, 2, 530–549.

Pycia, M., and P. Troyan (2016): "Obvious Dominance and Random Priority," UCLA, Unpublished mimeo.

Rees-Jones, A. (2016): "Suboptimal Behavior in Strategy-Proof Mechanisms: Evidence from the Residency Match," mimeo.

Sun, Y. (2006): "The exact law of large numbers via Fubini extension and characterization of insurable risks," *Journal of Economic Theory*, 126, 31–69.

A Additional Details on Victorian Tertiary Admission

A.1 Courses

Victorian clearinghouse, Victorian Tertiary Admissions Centre (VTAC), processes applications both for undergraduate and for technical and further education (TAFE) courses. Undergraduate courses include bachelor degrees as well as a variety of diplomas and certificates. An applicant lists any of those courses in a single ROL.

An excerpt of an undergraduate course description from VTAC guide – the main publication of the clearinghouse – is given below:

> Communication
>
> Monash Uni, Berwick: 26111 (CSP), 26112 (Fee), 26113 (Int. Fee)
>
> Title and length: Bachelor of Communication: FT3, PTA.
>
> About the course: [...]
>
> Major studies: [...]
>
> Prerequisites: Units 3 and 4 – a study score of at least 25 in English (any).
>
> Selection mode: CY12: ENTER and two-stage process with a middle-band of approximately 20%. NONY12: Academic record including GPA (see institutional page) and form. See Extra requirements for specifics.
>
> Middle-band: Completing any of classical societies and cultures, geography, history (any), international politics: international studies, international politics: national politics, literature, LOTE (any) or media studies = an aggregate 2 points higher per study, to a maximum 6 points; SEAS.
>
> Extra requirements: [...]

We treat two courses as offering the same program and differing only by the fee if the course code shares the first four digits (which implies that the courses also share the description given above). The course above is offered in three varieties: as a CSP, reduced-fee, course; as a full-fee course; and as a course for international applicants. We are interested in the first two varieties. This course will re-rank 20% of applicants (see selection mode) based on the performance in specific courses listed in the "Middle-band" section. It also re-ranks the affirmative action (SEAS) applicants. Most courses re-rank affirmative action applicants; whether a course re-ranks the applicants based on other criteria varies, and the criteria may be less specific. A small number of courses, such as those in performing arts, require a portfolio, an audition, or an interview. CY12 refers to current year 12 applicants (the focus of our study, category V16 applicants, is part of CY12) and NONY12 refers to non-year 12 applicants.

To determine a cutoff of a course, we select bottom 5% of accepted applicants and top 5% of rejected applicants and then take a median ENTER of applicants in this selection. Usually, the number of rejected applicants is about twice as large as the number of accepted applicants. Thus such a selection over-weights rejected applicants. Furthermore, we do not observe special consideration applicants; the bottom 5% of accepted applicants are more likely to come from this pool. Overall, these applicants often have lower scores than the top 5% rejected. Only cutoffs of

reduced-fee courses are used in the paper, as the number of accepted and rejected applicants for full-fee courses is too small.

A.2 Applicants

There are multiple categories of applicants; we focus on the category "V16", who are the most typical high school graduate in Victoria. They follow the standard state curriculum and do not have any tertiary course credits to claim. Two other categories are also of interest to us: "V14" and "V22". The former are Victorian applicants who follow International Baccalaureate curriculum, while the latter are interstate applicants. These two categories must be evaluated in the same way as V16 by the admission officers. We use them to derive course cutoffs more precisely. We do not use them in the estimations because they miss some control variables that we use. Table A1 gives relative frequencies of these categories of applicants. Among V16 applicants, 27922 fill less than 12 courses; they form our main sample.

Table A1: Categories of Applicants

	All	CY12			NONY12
		V16	V14 and V22	Other CY12	
Total	74,704	37,266	4,103	9,275	24,060
% of Total	100.00	49.89	5.49	12.42	32.21

Applicants have easy access to the following information: Clearly-in ENTER, Fringe ENTER, % of Offers Below Clearly-in ENTER (all three are available for Round 1 and Final Round), as well as Final Number of Offers (CY12 and Total). Clearly-in ENTER refers to the cutoff above which every eligible applicant must be admitted; Fringe ENTER refers to 5% percentile of the scores of admitted applicants. Note that this cutoff statistic does not distinguish between CY12 and NONY12 applicants; for that reason, we do not use any of these cutoffs to determine payoff-relevance of mistakes.

Even if an applicant skips a feasible RF course (that is, applicant's ENTER is above the course cutoff), the skipping mistake may not be payoff relevant: the applicant may have been admitted to a course that the applicant prefers. Thus, to determine whether the mistake is payoff-relevant, we need to complete the ROL of an applicant by adding back the skipped RF course. We report the lower and upper bounds on the number of payoff relevant mistakes.

For the lower bound, we assume that skipped RF course is the least desirable among acceptable RF courses. Specifically, a skip is considered payoff-relevant if (i) an applicant does not receive an offer from any RF course; (ii) receives an offer from an FF course; (iii) does not list the RF course corresponding to the FF course being offered (the course with the same code except for the last digit); and (iv) the RF course is feasible for this applicant. For the upper bound, we assume that skipped RF course is the most desirable. Specifically, a skip is considered payoff-relevant if (i) an

applicant lists FF course; (ii) does not list a corresponding RF course; and (iii) the RF course is feasible.

The summary statistics for all applicants, applicants with skips and applicants with payoff relevant mistakes are presented in table A2.

Table A2: Summary statistics for applicants by their mistake status

	All	w/Skip	w/Mistake
ENTER	65.77	61.84	78.21
GAT	61.94	59.03	66.44
Female	0.57	0.53	0.59
ln(income)	0.0024	0.0036	-0.0098
Citizen	0.98	0.95	0.94
Perm. resident	0.02	0.04	0.05
Born in			
Australia	0.90	0.85	0.86
Southern and Central Asia	0.01	0.04	0.06
Language spoken at home			
English	0.91	0.86	0.88
Eastern Asian Languages	0.02	0.04	0.06
Number of FF courses in ROL	0.22	2.40	2.93
Attends high school with tuition fees >AUD9000	0.16	0.27	0.34
Total	27,922	1,009	201

Notes: "All" refers to all V16 applicants who list fewer than 12 courses. "Mistake" refers to a payoff-relevant mistake. Numbers for citizenship status, country born and language spoken at home do not sum up to 100 as some entries have been omitted.

A.3 Expected ENTER

We predict ENTER using the following model:

$$\begin{aligned} ENTER_i = & a_0 \\ & + a_{11}\, GAT1_i + a_{12}\, GAT1_i^2 + a_{13}\, GAT1_i^3 \\ & + a_{21}\, GAT2_i + a_{22}\, GAT2_i^2 + a_{23}\, GAT2_i^3 \\ & + a_{31}\, GAT3_i + a_{32}\, GAT3_i^2 + a_{33}\, GAT1_i^3 + \epsilon_i, \end{aligned}$$

where $GAT1, GAT2$ and $GAT3$ are the results of three parts of GAT test (written communication; mathematics, science and technology; humanities, the arts and social sciences).

B 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 $\hat{\rho}(u)$, without loss we can work in terms of the "projected" ordinal type (ρ, s). Also recall that the mechanism depends only on an applicant's ROL and her score. Hence, for the current proof, we shall abuse the notation and call $\theta := (\rho, s)$ an applicant's type, redefine the type space $\Theta := \mathcal{R} \times [0,1]^C$ (the projection of the original types), and let η 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 $\theta \notin \Theta^\delta$ to randomize between TRS with probability γ 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 applicants 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 college an applicant with type $\theta = (\rho, s)$ demands given cutoffs P^* (i.e., her most preferred feasible college given P^*). Since $\hat{r}$ ranks her favorite feasible college ahead of all other feasible colleges, 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 η has full support and since the prescribed strategy has every type θ play TRS with positive probability, the induced measure $\hat{\eta}$ must also have full support.[19] 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 $\theta \notin \Theta^\delta$ has the same set of feasible colleges when the cutoffs are P as when they are P^*. This means that for any such type (ρ, 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 applicants 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 applicants follow the prescribed strategies. Let $\hat{F}^k = [\hat{\eta}^k, S]$ be the resulting "induced" k-random

[19] 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 applicants are coordinating via asymmetric strategies so that exactly γ fraction of each type plays TRS.

economy. By construction, $\hat{F}^k$ consists of k independent draws of applicants 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 $\epsilon'(\overline{u} - \underline{u}) \leq \epsilon$. 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 ϵ-Bayesian Nash equilibrium for each k-random economy for $k > K$.

First, for any type $\theta \in \Theta^\delta$ the prescribed strategy, namely TRS, is trivially optimal given the strategyproofness of DA. Hence, consider an applicant with any type $\theta \notin \Theta^\delta$, and suppose that all other applicants employ the prescribed strategies. Now condition on event $\mathcal{E}^k$. Recall that the set of feasible colleges is the same for type $\theta \notin \Theta^\delta$ 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 $\theta \notin \Theta^\delta$.

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

$$\epsilon'(\overline{u} - \underline{u}) < \epsilon.$$

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

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

where the set $\{c' \in C : s_{i,c'} \geq P_{c'}\}$ is the set of feasible colleges for applicant 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 $\bar{D}^k(P) := \mathbb{E}_{(R_1,...,R_k)}\left[D^k(P)\right]$, where the randomness is taken over the random draws of the types of the applicants and the random reported ROLs according to mixed strategy profile $\left(\sigma_i^k\right)_{1\leq i\leq k}$.

As a preliminary step, we establish a series of claims.

Claim 3. *Fix any P. Then, for any $\alpha > 0$,*

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

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

$$\Pr\{|D_c^k(P) - \bar{D}_c^k(P)| > \alpha\} \leq e^{-2k\alpha^2},$$

since for each $c \in C$, $|D_c^k(P)(R_1,...,R_k) - D_c^k(P)(R_1',...,R_k')| \leq 1/k$ whenever $(R_1,...,R_k)$ and $(R_1',...,R_k')$ differ only in one component.

It then follows that

$$\begin{aligned}
&\Pr\left[\left\|D^k(P) - \bar{D}^k(P)\right\| > \sqrt{|C|}\alpha\right] \\
&\leq \Pr\left[\exists\ c \in C \text{ s.t. } \left|D_c^k(P) - \bar{D}_c^k(P)\right| > \alpha\right] \\
&\leq |C| \cdot e^{-2k\alpha^2}.
\end{aligned}$$

□

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

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

We want to find $\delta > 0$ (which may depend on ε and P) s.t.

$$\left\|\bar{D}^k(P') - \bar{D}^k(P)\right\| < \varepsilon$$

for all $P' \in [0,1]^C$ with $\|P' - P\| < \delta$ and all k.

Define

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

We can find $\delta > 0$ s.t. $\eta\left(\Theta_{P,P'}\right) < \varepsilon/\sqrt{|C|}$ for all P' s.t. $\|P' - P\| < \delta$. This can be guaranteed if we assume the measure η to be absolutely continuous w.r.t. Lebesgue measure.

Then we have

$$
\begin{aligned}
&\left\|\bar{D}^k\left(P'\right)-\bar{D}^k\left(P\right)\right\| \\
&=\sqrt{\sum_{c\in C}\left|\mathbb{E}\left[D_c^k\left(P'\right)-D_c^k\left(P\right)\right]\right|^2} \\
&=\sqrt{\sum_{c\in C}\left|\mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k}\left(\begin{array}{c} I\left\{c\in \arg\max_{R_i^k}\left\{c'\in C: s_{i,c'}\geq P'_{c'}\right\}\right\} \\ -I\left\{c\in \arg\max_{R_i^k}\left\{c'\in C: s_{i,c'}\geq P_{c'}\right\}\right\}\end{array}\right)\right]\right|^2} \\
&\leq\sqrt{\sum_{c\in C}\left[\mathbb{E}\left|\frac{1}{k}\sum_{i=1}^{k}\left(\begin{array}{c} I\left\{c\in \arg\max_{R_i^k}\left\{c'\in C: s_{i,c'}\geq P'_{c'}\right\}\right\} \\ -I\left\{c\in \arg\max_{R_i^k}\left\{c'\in C: s_{i,c'}\geq P_{c'}\right\}\right\}\end{array}\right)\right|\right]^2} \\
&\leq\frac{1}{k}\sqrt{\sum_{c\in C}\left(\mathbb{E}\sum_{i=1}^{k}\left|\begin{array}{c} I\left\{c\in \arg\max_{R_i^k}\left\{c'\in C: s_{i,c'}\geq P'_{c'}\right\}\right\} \\ -I\left\{c\in \arg\max_{R_i^k}\left\{c'\in C: s_{i,c'}\geq P_{c'}\right\}\right\}\end{array}\right|\right)^2} \\
&\leq\frac{1}{k}\sqrt{\sum_{c\in C}\left(\mathbb{E}\sum_{i=1}^{k}I\left\{\theta_i\in\Theta_{P,P'}\right\}\right)^2}\leq\frac{1}{k}\sqrt{\sum_{c\in C}\left(k\cdot\eta\left(\Theta_{P,P'}\right)\right)^2} \\
&=\sqrt{|C|}\eta\left(\Theta_{P,P'}\right)<\varepsilon,
\end{aligned}
$$

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 indentical when $\theta_i\notin\Theta_{P,P'}$. □

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

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

Claim 6. *For any $\epsilon'>0$, there exists a subsequence $\left\{\bar{D}^{k_\ell}\right\}_{\ell=1}^{\infty}$ such that $\lim_{\ell\to\infty}\Pr\{\sup_P||D^{k_\ell}(P)-\bar{D}(P)||>\epsilon'\}=0$.*

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 $i_j=0,...,i_j^*$ such that $||\bar{D}(P_{\mathbf{i}+1})-\bar{D}(P_{\mathbf{i}}^-)||<\epsilon'/2$ for all $\mathbf{i}=(i_1,...,i_C)$, where $\mathbf{i}+1:=(i_1+1,...,i_C+1)$. Let m be the number of such intervals. Using the argument of Glivenko-Cantelli, one can show that for any P there exists $\mathbf{i}$ such that

$$||D^{k_\ell}(P)-\bar{D}(P)||\leq\max\{||D^{k_\ell}(P_{\mathbf{i}})-\bar{D}(P_{\mathbf{i}})||,||D^{k_\ell}(P_{\mathbf{i}+1})-\bar{D}(P_{\mathbf{i}+1})||\}+\epsilon'/2.$$

Suppose event $||D^{k_\ell}(P)-\bar{D}(P)||>\epsilon'$ occurs for some P. Then there must exist $\mathbf{i}$ such that $||D^{k_\ell}(P_{\mathbf{i}})-\bar{D}(P_{\mathbf{i}})||\geq\epsilon'/2$. Since $\bar{D}^{k_\ell}(\cdot)$ converges to $\bar{D}(\cdot)$ in sup norm by Claim 5, there exists K' such that for all $\ell>K'$, $\sup_P||\bar{D}^{k_\ell}(P)-\bar{D}(P)||<\epsilon'/4$. Hence, for $\ell>K'$ and for $\mathbf{i}$, we must have

$$||D^{k_\ell}(P_{\mathbf{i}})-\bar{D}^{k_\ell}(P_{\mathbf{i}})||\geq\epsilon'/4.$$

Combining the arguments so far, we conclude:

$$\begin{aligned}
&\Pr\{\sup_P ||D^{k_\ell}(P) - \bar{D}(P)|| > \epsilon'\} \\
=&\Pr\{\exists P \text{ s.t. } ||D^{k_\ell}(P) - \bar{D}(P)|| > \epsilon'\} \\
\leq&\Pr\{\exists \mathbf{i} \text{ s.t. } ||D^{k_\ell}(P_\mathbf{i}) - \bar{D}(P_\mathbf{i})|| > \epsilon'/2\} \\
\leq&\Pr\{\exists \mathbf{i} \text{ s.t. } ||D^{k_\ell}(P_\mathbf{i}) - \bar{D}^{k_\ell}(P_\mathbf{i})|| > \epsilon'/4\} \\
=&\Pr\{\cup_\mathbf{i}\{||D^{k_\ell}(P_\mathbf{i}) - \bar{D}^{k_\ell}(P_\mathbf{i})|| > \epsilon'/4\}\} \\
\leq&\sum_\mathbf{i} \Pr\{||D^{k_\ell}(P_\mathbf{i}) - \bar{D}^{k_\ell}(P_\mathbf{i})|| > \epsilon'/4\} \\
\leq&\sum_\mathbf{i} e^{-k_\ell \epsilon'^2/8} \\
=&me^{-k_\ell \epsilon'^2/8} \to 0 \text{ as } \ell \to \infty,
\end{aligned}$$

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 $\left\{\left(\sigma_i^k\right)_{1\leq i\leq k}\right\}_k$ are not asymptotically stable. Then by definition, there exists $\varepsilon > 0$ and a subsequence of finite economies $\left\{F^{k_j}\right\}_j$ such that

$$\Pr\left(\text{The fraction of applicants playing SRS against } P^{k_j} \geq 1-\varepsilon\right) < 1-\varepsilon. \qquad (*)$$

By Claim 6, we know that there exists a sub-subsequence $D^{k_{j_l}}$ that converges to $\bar{D}$ uniformly in probability. Given the regularity of the strategies employed by the applicants 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 $P^{k_{j_l}}$ converges to $\bar{P}$ in probability, where $\bar{P}$ is a deterministic cutoff s.t. $\bar{D}\left(\bar{P}\right) = S$.

Define

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

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

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

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

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

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

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

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

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

By WLLN, there exists $\hat{L}$ s.t. $\hat{L}$ i.i.d. Bernouli 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 $(1-\varepsilon)^{1/2} k_{j_l} > \hat{L}$.

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

$$\Pr\left(\text{The fraction of applicants playing SRS against } P^{k_{j_l}} \geq 1-\varepsilon\right) \geq 1-\varepsilon,$$

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

First, notice that in economy $F^{k_{j_l}}$, an applicant 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 applicant i and some type $\theta \in \hat{\Theta}$ s.t.

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

Then deviating to TRS will give this applicant i with type θ at least a gain of

$$\begin{aligned}
&\delta \cdot \Pr\left(\sigma_i^{k_{j_l}}(\theta) \text{ does not play SRS against } P^{k_{j_l}}\right) \\
&\geq \delta \cdot \Pr\left(\begin{array}{c}\sigma_i^{k_{j_l}}(\theta) \text{ does not play SRS against } \bar{P} \\ \text{and event } A^{k_{j_l}}\end{array}\right) \\
&\geq \delta\left[\Pr\left(A^{k_{j_l}}\right) - \Pr\left(\sigma_i^{k_{j_l}}(\theta) \text{ plays SRS against } \bar{P}\right)\right] \\
&\geq \delta\left[(1-\varepsilon)^{1/6}-(1-\varepsilon)^{1/3}\right],
\end{aligned}$$

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

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

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

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

Then we have

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

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

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

$$
\begin{aligned}
&\Pr\left(\begin{array}{c}\text{The fraction of applicants with } \theta \in \hat{\Theta} \\ \text{playing SRS against } P^{k_{j_l}} \geq (1-\varepsilon)^{1/2}\end{array} \middle| \eta^{k_{j_l}}\left(\hat{\Theta}\right) \geq (1-\varepsilon)^{1/2}\right) \\
&\geq \Pr\left(\begin{array}{c}\text{The fraction of applicants with } \theta \in \hat{\Theta} \\ \text{playing SRS against } \bar{P} \geq (1-\varepsilon)^{1/2}, \\ \text{and event } A^{k_{j_l}}\end{array} \middle| \eta^{k_{j_l}}\left(\hat{\Theta}\right) \geq (1-\varepsilon)^{1/2}\right) \\
&\geq \Pr\left(\begin{array}{c}\text{The fraction of applicants with } \theta \in \hat{\Theta} \\ \text{playing SRS against } \bar{P} \geq (1-\varepsilon)^{1/2}\end{array} \middle| \eta^{k_{j_l}}\left(\hat{\Theta}\right) \geq (1-\varepsilon)^{1/2}\right) \\
&- \Pr\left(\bar{A}^{k_{j_l}} \middle| \eta^{k_{j_l}}\left(\hat{\Theta}\right) \geq (1-\varepsilon)^{1/2}\right) \\
&\geq (1-\varepsilon)^{1/3} - \frac{\Pr\left(\bar{A}^{k_{j_l}}\right)}{\Pr\left(\eta^{k_{j_l}}\left(\hat{\Theta}\right) \geq (1-\varepsilon)^{1/2}\right)} \\
&\geq (1-\varepsilon)^{1/3} - \frac{(1-\varepsilon)^{1/2}\left[(1-\varepsilon)^{1/3} - (1-\varepsilon)^{1/2}\right]}{(1-\varepsilon)^{1/2}} \\
&= (1-\varepsilon)^{1/2},
\end{aligned}
$$

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

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

in economy $F^{k_{j_l}}$

$$
\begin{aligned}
&\Pr\left(\text{The fraction of applicants playing SRS against } P^{k_{j_l}} \geq 1-\varepsilon\right) \\
&\geq \Pr\left(\begin{array}{c}\text{The fraction of applicants with } \theta \in \hat{\Theta} \\ \text{playing SRS against } P^{k_{j_l}} \geq (1-\varepsilon)^{1/2} \\ \text{and } \eta^{k_{j_l}}\left(\hat{\Theta}\right) \geq (1-\varepsilon)^{1/2}\end{array}\right) \\
&= \Pr\left(\eta^{k_{j_l}}\left(\hat{\Theta}\right) \geq (1-\varepsilon)^{1/2}\right) \cdot \Pr\left(\begin{array}{c}\text{The fraction of applicants with } \theta \in \hat{\Theta} \\ \text{playing SRS against } P^{k_{j_l}} \geq (1-\varepsilon)^{1/2}\end{array} \middle| \, \eta^{k_{j_l}}\left(\hat{\Theta}\right) \geq (1-\varepsilon)^{1/2}\right) \\
&\geq (1-\varepsilon)^{1/2} \cdot (1-\varepsilon)^{1/2} \\
&= 1-\varepsilon,
\end{aligned}
$$

which contradicts $(*)$. □

C Monte Carlo Simulations

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

It further presents some additional details of the Monte Carlo simulations. Figure C1 describes the simulated spatial distribution of applicants and colleges in one sample; Figure C2 depicts the marginal distribution of each college's cutoffs across 200 simulations under the assumption that every applicant always truthfully ranks all colleges.

Furthermore, Tables C3 and C4 describe the skipping and mistakes for applicants with $T_i = 1$ and $T_i = 0$, respectively. Table C5 shows the mean and standard deviation of the estimates of each coefficient from different approaches; and Tables C6 and C7 present more detailed estimation results of the welfare effects.

C.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:

$$
\begin{aligned}
u_{i,c} &= \beta_1 c - d_{i,c} + \beta_2 T_i A_c + \beta_3 Small_c + \sigma\epsilon_{i,c} \\
&\equiv \sigma V_{i,c} + \sigma\epsilon_{i,c}, \forall i \text{ and } c;
\end{aligned}
$$

we also define $\mathbf{X}_i = (\{d_{i,c}, A_c, Small_c\}_c, T_i)$ to denote the observable applicant characteristics and college attributes; $L_i = (l_i^1, \ldots, l_i^{K_i})$ represents the rank-ordered list that i submits; and $\boldsymbol{\theta}$ is the vector of coefficients, $\boldsymbol{\theta} = (\beta_1, \beta_2, \beta_3, \sigma)$.

Weakly Truth-Telling. We start with formalizing the estimation under the weakly truth-telling assumption. If each applicant i submits $K_i \equiv |L_i|$ ($\leq C$) choices, under the assumption that applicants are weakly truth-telling, $L_i = (l_i^1, \ldots, l_i^{K_i})$ ranks truthfully i's top K_i preferred colleges.

The probability of applicant i submitting L_i is:

$$\begin{aligned}&\Pr\left(i \text{ submits } L_i \mid \mathbf{X}_i; \boldsymbol{\theta}\right)\\ =&\Pr\left(L_i = (l_i^1, \ldots, l_i^{K_i}) \mid \mathbf{X}_i; \boldsymbol{\theta}; K_i\right) \times \Pr\left(i \text{ submits a ROL of length } K_i \mid \mathbf{X}_i; \boldsymbol{\theta}\right).\end{aligned}$$

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

$$\begin{aligned}\Pr\left(L_i = (l_i^1, \ldots, l_i^{K_i}) \mid \mathbf{X}_i; \boldsymbol{\theta}; K_i\right) =& \Pr\left(u_{i,l_i^1} > \cdots > u_{i,l_i^{K_i}} > u_{i,c'} \; \forall \, c' \in \mathcal{C} \setminus L_i \mid \mathbf{X}_i; \boldsymbol{\theta}; K_i\right)\\ =& \prod_{s \in L_i} \left(\frac{\exp(V_{i,c})}{\sum_{c' \not\succ_{L_i} c} \exp(V_{i,c'})} \right)\end{aligned}$$

where $c' \not\succ_{L_i} c$ indicates that c' is not ranked before c in L_i, which includes c itself and the colleges 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 college (l_i^1) being the most preferred; another for the second-ranked college (l_i^2) being preferred to all colleges except the one ranked first, and so on.

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

$$\ln L_{TT}\big(\boldsymbol{\theta} \mid \mathbf{X}, |\mathbf{L}|\big) = \sum_{i=1}^{I} \sum_{c \in L_i} V_{i,c} - \sum_{i=1}^{I} \sum_{c \in L_i} \ln \Big(\sum_{c' \not\succ_{L_i} c} \exp(V_{i,c'}) \Big),$$

where $|\mathbf{L}|$ is the length of all ROLs.

Stability. We now assume that the matching is stable and explore how we can identify and estimate applicant preferences. Suppose that the matching is μ, which leads to a vector of cutoffs $P(\mu)$. With information on how colleges rank applicants, we can find a set of colleges that are *ex post* feasible to i, $\mathcal{C}(i, P(\mu))$.

The conditions specified by the stability of μ imply the likelihood of applicant i choosing $\mu(i)$ in $\mathcal{C}(i, P(\mu))$:

$$\Pr\left(\mu(i) = \underset{c \in \mathcal{C}(i, P(\mu))}{\arg\max} \, \overline{u}_{i,c} | \mathbf{X}_i, P(\mu); \boldsymbol{\theta} \right).$$

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

$$\ln L_{ST}\left(\boldsymbol{\theta} \mid \mathbf{X}, P(\mu)\right) = \sum_{i=1}^{I} V_{i,\mu(i)} - \sum_{i=1}^{I} \ln \Big(\sum_{c' \in \mathcal{C}_i} \exp(V_{i,c'}) \Big).$$

A key assumption of this approach is that the feasible set $\mathcal{C}(i, P(\mu))$ 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.

The robust approach is the same as the stability estimator, except that the feasible set of each applicant, $\mathcal{C}(i, P(\mu))$, is modified to be $\mathcal{C}(i, P^i(\mu))$, where $P^i(\mu)$ is such that $P^i_s(\mu) = P_s(\mu) + \delta$ if $s \neq \mu(i)$ and $P^i_s(\mu) = P_s(\mu)$ if $s = \mu(i)$. In the results we present here, we choose $\delta = 50/1800$. Recall that the 1800 applicants' scores are uniformly distributed in [0,1].

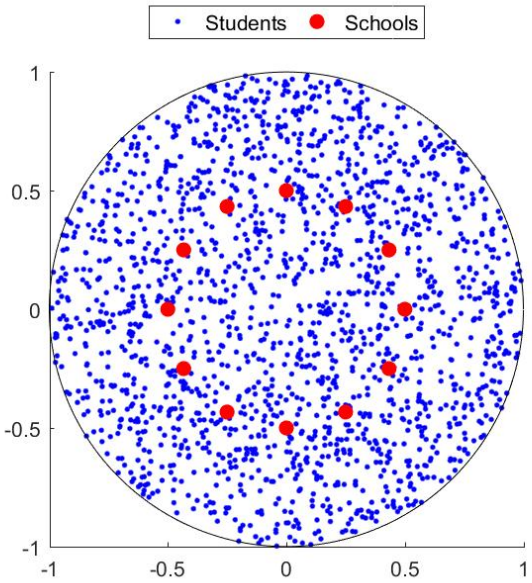


Figure C1: Monte Carlo Simulations: Spatial Distribution of Applicants and Colleges

Notes: This figure shows the spatial configuration of the area considered in the Monte Carlo simulations with 1800 applicants and 12 colleges. The area is within a circle of radius 1. The blue and red circles show the locations of applicants and colleges, respectively, in one simulation sample. Across samples, the colleges' locations are fixed, while applicants' locations are uniformly drawn within the circle.

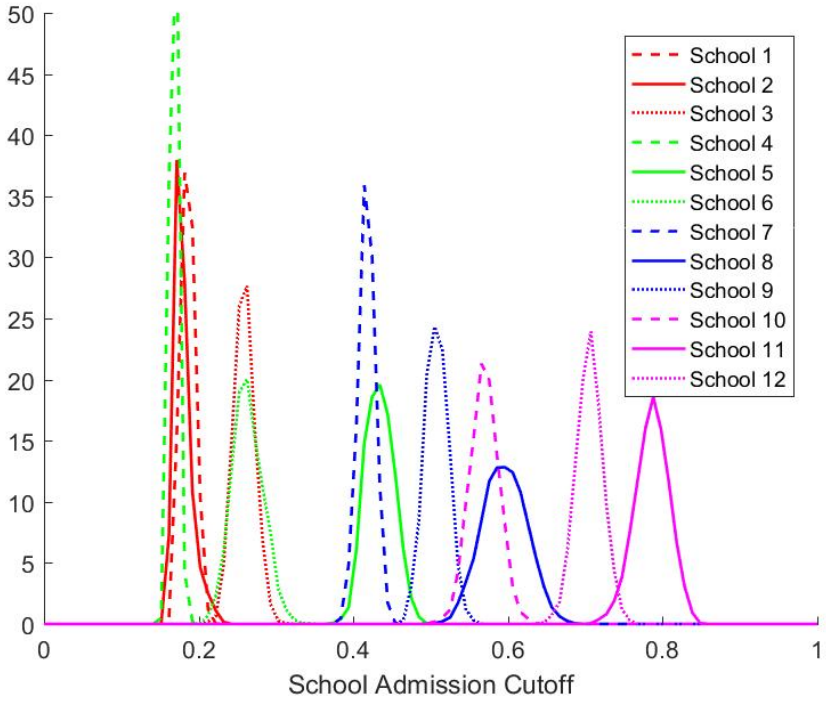


Figure C2: Simulated Distribution of Cutoffs when Everyone is Truth-telling

Notes: Assuming everyone is strictly truth-telling, we calculate the cutoffs of all colleges in each simulation sample. The figure shows the marginal distribution of each college'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 solid line indicates a small college with 75, instead of 150, seats. The simulation samples for cutoffs use independent draws of $\{d_{i,c}, \epsilon_{i,c}\}_c$ and T_i.

Table C3: Skips and Mistakes in Monte Carlo Simulations (Percentage Points): $T_i = 1$ Applicants

	Scenarios (Data Generating Processes)							
	Strict Truth-telling	Payoff Irrelevant Skips			Payoff Relevant Mistakes			
	STT	IRR 1	IRR 2	IRR 3	REL 1	REL 2	REL 3	REL 4
WTT: Weak Truth-Telling[a]	100	73	43	13	14	14	14	14
Matched w/ favorite feasible college[b]	100	100	100	100	96	91	85	80
Skippers[c]	0	31	63	95	95	96	95	95
By number of skips:								
Skipping 11 colleges	0	19	40	62	79	84	87	90
Skipping 10 colleges	0	7	15	22	14	10	8	5
Skipping 9 colleges	0	4	8	11	3	1	0	0
Skipping 8 colleges	0	0	0	0	0	0	0	0
Skipping 7 colleges	0	0	0	0	0	0	0	0
Skipping 6 colleges	0	0	0	0	0	0	0	0
Skipping 5 colleges	0	0	0	0	0	0	0	0
Skipping 4 colleges	0	0	0	0	0	0	0	0
Skipping 3 colleges	0	0	0	0	0	0	0	0
Skipping 2 colleges	0	0	0	0	0	0	0	0
Skipping 1 college	0	0	0	0	0	0	0	0
STT: Strict Truth-Telling	100	70	37	5	5	5	5	5

Notes: This table presents the configurations of the eight data generating process (DGPs), similar to Table 8 but only among the applicants with $T_i = 1$. Each entry is a percentage averaged over the 200 simulation samples. On average, there are 1201 such applicants in each sample. [a] An applicant is "*weak truth-telling*" if she truthfully ranks her top K_i ($1 \leq K_i \leq 12$) preferred colleges (K_i is observed). [b] A college is feasible to an applicant, if the applicant's index is higher than the college's *ex post* admission cutoff. If an applicant is matched with her favorite feasible college, she cannot form a blocking pair with any college. [c] Given that every college is acceptable to all applicants and potentially over-demanded, an applicant is a skipper if she does not rank all colleges. [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 detail in section 4.3.

Table C4: Skips and Mistakes in Monte Carlo Simulations (Percentage Points): $T_i = 0$ Applicants

	Scenarios (Data Generating Processes)							
	Strict Truth-telling	Payoff Irrelevant Skips			Payoff Relevant Mistakes			
	STT	IRR 1	IRR 2	IRR 3	REL 1	REL 2	REL 3	REL 4
WTT: Weak Truth-Telling[a]	100	91	81	72	72	71	71	71
Matched w/ favorite feasible college[b]	100	100	100	100	99	98	97	95
Skippers[c]	0	23	48	70	70	70	70	70
By number of skips:								
Skipping 11 colleges	0	16	35	52	61	64	65	67
Skipping 10 colleges	0	5	10	14	8	6	5	3
Skipping 9 colleges	0	1	3	4	1	0	0	0
Skipping 8 colleges	0	0	0	0	0	0	0	0
Skipping 7 colleges	0	0	0	0	0	0	0	0
Skipping 6 colleges	0	0	0	0	0	0	0	0
Skipping 5 colleges	0	0	0	0	0	0	0	0
Skipping 4 colleges	0	0	0	0	0	0	0	0
Skipping 3 colleges	0	0	0	0	0	0	0	0
Skipping 2 colleges	0	0	0	0	0	0	0	0
Skipping 1 college	0	0	0	0	0	0	0	0
STT: Strict Truth-Telling	100	78	52	30	30	30	30	30

Notes: This table presents the configurations of the eight data generating process (DGPs), similar to Table 8 but only among the applicants with $T_i = 0$. Each entry is a percentage averaged over the 200 simulation samples. On average, there are 599 such applicants in each sample. [a]An applicant is "*weak truth-telling*" if she truthfully ranks her top K_i ($1 \leq K_i \leq 12$) preferred colleges (K_i is observed). [b]A college is feasible to an applicant, if the applicant's index is higher than the college's *ex post* admission cutoff. If an applicant is matched with her favorite feasible college, she cannot form a blocking pair with any college. [c]Given that every college is acceptable to all applicants and potentially over-demanded, an applicant is a skipper if she does not rank all colleges.

Table C5: Estimation with Different Identifying Assumptions: Monte Carlo Results

DGPs	Identifying Condition	Quality ($\beta_1 = 0.3$)		Interaction ($\beta_2 = 1$)		Unobs. ($\sigma = 1$)		Small college ($\beta_3 = 0$)	
		mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
A. Strict Truth-telling (All three approaches are consistent)									
	WTT	0.30	0.01	2.00	0.06	1.00	0.03	0.00	0.02
STT	Stability	0.30	0.03	2.03	0.23	1.01	0.10	0.00	0.08
	Robust	0.30	0.03	2.02	0.25	1.01	0.10	-0.01	0.08
B. Payoff-irrelevant Skips (Only stability and robust are consistent)									
	WTT	0.29	0.01	1.99	0.07	1.07	0.03	-0.04	0.02
IRR 1	Stability	0.30	0.03	2.03	0.23	1.01	0.10	0.00	0.08
	Robust	0.30	0.03	2.02	0.25	1.01	0.10	-0.01	0.08
	WTT	0.28	0.01	1.92	0.10	1.21	0.05	-0.12	0.02
IRR 2	Stability	0.30	0.03	2.03	0.23	1.01	0.10	0.00	0.08
	Robust	0.30	0.03	2.02	0.25	1.01	0.10	-0.01	0.08
	WTT	0.23	0.02	1.50	0.15	1.52	0.11	-0.30	0.05
IRR 3	Stability	0.30	0.03	2.03	0.23	1.01	0.10	0.00	0.08
	Robust	0.30	0.03	2.02	0.25	1.01	0.10	-0.01	0.08
C. Payoff-relevant Mistakes (None is consistent)									
	WTT	0.24	0.02	1.45	0.15	1.45	0.10	-0.27	0.04
REL 1	Stability	0.30	0.03	1.98	0.26	1.02	0.10	-0.02	0.12
	Robust	0.30	0.03	1.99	0.27	1.02	0.10	-0.03	0.12
	WTT	0.25	0.02	1.44	0.15	1.44	0.10	-0.26	0.04
REL 2	Stability	0.29	0.03	1.92	0.32	1.05	0.10	-0.05	0.15
	Robust	0.29	0.03	1.94	0.30	1.03	0.11	-0.05	0.14
	WTT	0.25	0.02	1.45	0.16	1.43	0.10	-0.25	0.04
REL 3	Stability	0.29	0.04	1.88	0.39	1.07	0.11	-0.07	0.17
	Robust	0.29	0.04	1.91	0.34	1.05	0.11	-0.07	0.16
	WTT	0.25	0.02	1.45	0.16	1.42	0.09	-0.25	0.04
REL 4	Stability	0.28	0.04	1.81	0.46	1.10	0.12	-0.09	0.18
	Robust	0.28	0.04	1.85	0.40	1.07	0.12	-0.08	0.17

Notes: This table presents estimates (mean and standard deviation across 200 samples) of the random utility model described in equation (11). The true values are $(\beta_1, \beta_2, \sigma) = (0.3, 2, 1)$, and the coefficient on the small school dummy is zero. It shows results in the eight data generating process (DGPs) with three identifying assumptions, WTT, stability, and the robust approach. WTT assumes that every applicant truthfully ranks her top K_i $(1 < K_i \leq 12)$ colleges, where K_i is the observed number of colleges in i's ROL. Stability implies that every applicant is matched with her favorite feasible college, given the *ex post* cutoffs.

Table C6: Welfare Effects of the Counterfactual Policy on Applicants with $T_i = 1$

DGP	Approach to Counterfactual	Worse Off		Better Off		Indifferent	
		mean	s.d.	mean	s.d.	mean	s.d.
	A. Strict truth-telling						
	Submitted ROLs	0	0	91	1	9	1
	WTT	0	0	91	1	9	1
STT	Stability	0	0	91	1	9	1
	Robust	0	0	91	1	9	1
	Truth	0	0	91	1	9	1
	B. Payoff-irrelevant skips						
	Submitted ROLs	0	0	79	2	21	2
	WTT	0	0	91	1	9	1
IRR 1	Stability	0	0	91	1	9	1
	Robust	0	0	91	1	9	1
	Truth	0	0	91	1	9	1
	Submitted ROLs	0	0	65	2	35	2
	WTT	0	0	89	1	11	1
IRR 2	Stability	0	0	91	1	9	1
	Robust	0	0	91	1	9	1
	Truth	0	0	91	1	9	1
	Submitted ROLs	0	0	53	3	47	3
	WTT	0	0	86	1	14	1
IRR 3	Stability	0	0	91	1	9	1
	Robust	0	0	91	1	9	1
	Truth	0	0	91	1	9	1
	C. Payoff-relevant mistakes						
	Submitted ROLs	0	0	45	3	55	3
	WTT	0	0	87	1	13	1
REL 1	Stability	0	0	91	1	9	1
	Robust	0	0	91	1	9	1
	Truth	0	0	91	1	9	1
	Submitted ROLs	0	0	43	2	57	2
	WTT	0	0	87	1	13	1
REL 2	Stability	0	0	91	2	9	2
	Robust	0	0	91	1	9	1
	Truth	0	0	91	1	9	1
	Submitted ROLs	0	0	42	2	58	2
	WTT	0	0	87	1	13	1
REL 3	Stability	0	0	90	2	10	2
	Robust	0	0	91	2	9	2
	Truth	0	0	91	1	9	1
	Submitted ROLs	0	0	42	2	58	2
	WTT	0	0	87	1	13	1
REL 4	Stability	0	0	90	2	10	2
	Robust	0	0	90	2	10	2
	Truth	0	0	91	1	9	1

Notes: This table presents the estimated effects of the counterfactual policy (giving $T_i = 1$ applicants priority in admission) on applicants with $T_i = 1$. On average, there are 1201 such applicants (standard deviation 14) in each simulation sample. The table shows results in the eight data generating process (DGPs) with five approaches. The one using submitted ROLs assumes that submitted ROLs represent applicant true ordinal preferences; WTT assumes that every applicant truthfully ranks her top K_i ($1 < K_i \leq 12$) preferred colleges (K_i is observed); stability implies that every applicant is matched with her favorite feasible college, given the *ex post* cutoffs; and the robust approach inflates some cutoffs and re-runs the stability estimator. The truth is simulated with the possible mistakes in each DGP. The welfare change of each applicant is calculated in the following way: We first simulate the counterfactual match and investigate if a given applicant is better off, worse off, or indifferent by comparing the two matches according to estimated/assumed/true ordinal preferences. In each simulation sample, we calculate the percentage of different welfare change; the table then reports the mean and standard deviation of the percentages across the 200 simulation samples.

Table C7: Welfare Effects of the Counterfactual Policy on Applicants with $T_i = 0$

DGP	Approach to Counterfactual	Worse Off		Better Off		Indifferent	
		mean	s.d.	mean	s.d.	mean	s.d.
	A. Strict truth-telling						
	Submitted ROLs	68	2	0	0	32	2
	WTT	68	2	0	0	32	2
STT	Stability	67	2	1	0	32	2
	Robust	67	2	1	0	32	2
	Truth	68	2	0	0	32	2
	B. Payoff-irrelevant skips						
	Submitted ROLs	55	2	0	0	45	2
	WTT	65	2	2	0	33	2
IRR 1	Stability	67	2	1	0	32	2
	Robust	67	2	1	0	32	2
	Truth	68	2	0	0	32	2
	Submitted ROLs	40	2	0	0	60	2
	WTT	60	2	5	1	35	2
IRR 2	Stability	67	2	1	0	32	2
	Robust	67	2	1	0	32	2
	Truth	68	2	0	0	32	2
	Submitted ROLs	30	1	0	0	70	1
	WTT	47	2	13	1	40	2
IRR 3	Stability	67	2	1	0	32	2
	Robust	67	2	1	0	32	2
	Truth	68	2	0	0	32	2
	C. Payoff-relevant mistakes						
	Submitted ROLs	26	1	0	0	74	1
	WTT	50	2	11	1	39	2
REL 1	Stability	67	3	1	1	32	2
	Robust	67	3	1	1	32	2
	Truth	68	2	0	0	32	2
	Submitted ROLs	25	1	0	0	75	1
	WTT	51	2	11	1	38	2
REL 2	Stability	66	3	2	2	32	2
	Robust	66	3	2	1	32	2
	Truth	67	2	0	0	32	2
	Submitted ROLs	24	1	0	0	76	2
	WTT	51	2	11	1	38	2
REL 3	Stability	65	4	3	2	33	2
	Robust	65	4	2	2	32	2
	Truth	67	2	0	0	32	2
	Submitted ROLs	23	1	0	0	76	2
	WTT	52	2	11	1	37	2
REL 4	Stability	63	5	4	3	33	3
	Robust	64	4	3	2	33	2
	Truth	67	2	0	0	32	2

Notes: This table presents the estimated effects of the counterfactual policy (giving $T_i = 1$ applicants priority in admission) on applicants with $T_i = 0$. On average, there are 599 such applicants (standard deviation 14) in each simulation sample. The table shows results in the eight data generating process (DGPs) with five approaches. The one using submitted ROLs assumes that submitted ROLs represent applicant true ordinal preferences; WTT assumes that every applicant truthfully ranks her top K_i ($1 < K_i \leq 12$) preferred colleges (K_i is observed); stability implies that every applicant is matched with her favorite feasible college, given the *ex post* cutoffs; and the robust approach inflates some cutoffs and re-runs the stability estimator. The truth is simulated with the possible mistakes in each DGP. The welfare change of each applicant is calculated in the following way: We first simulate the counterfactual match and investigate if a given applicant is better off, worse off, or indifferent by comparing the two matches according to estimated/assumed/true ordinal preferences. In each simulation sample, we calculate the percentage of different welfare change; the table then reports the mean and standard deviation of the percentages across the 200 simulation samples.