Do "Short-List" Students Report Truthfully? Strategic Behavior in the Chilean College Admissions Problem

Tomás Larroucau[1] and Ignacio Rios[2]

[1] *University of Pennsylvania, Department of Economics*
[2] *Stanford, Graduate School of Business*

April 10, 2020

Abstract

We analyze the application process in the Chilean College Admissions problem. Students can submit up to 10 preferences, but most students do not fill their entire application list ("short-list"). Even though students face no incentives to misreport, we find evidence of strategic behavior as students tend to omit programs for which their admission probabilities are too low. To rationalize this behavior, we construct a portfolio problem where students maximize their expected utility given their preferences and beliefs over admission probabilities. We adapt the estimation procedure proposed by Agarwal and Somaini (2018) to solve a large portfolio problem. To simplify this task, we show that it is sufficient to compare a ROL with only a subset of ROLs ("one-shot swaps") to ensure its optimality without running into the curse of dimensionality. To better identify the model, we exploit a unique exogenous variation on the admission weights over time. We find that assuming truth-telling leads to biased results. Specifically, when students only include programs if it is strictly profitable to do so, assuming truth-telling underestimates how preferred selective programs are and overstates the value of being unassigned and the degree of preference heterogeneity in the system. Ignoring the constraint on the length of the list can also result in biased estimates, even if the proportion of constrained ROLs is relatively small. Our estimation results strongly suggest that "short-list" students should not be interpreted as truth-tellers, even in a seemingly strategy-proof environment. Finally, we apply our estimation method to estimate students' preferences for programs and majors in Chile and find strong differences in preferences regarding students' gender and scores.

1 Introduction

In recent years, growing attention has been devoted to understanding the application behavior of students in centralized admission systems. A major question is how to separately identify students' preferences from beliefs on admission chances using only data on application lists and/or enrollment choices. This is especially relevant in settings where the mechanism used is not strategy-proof, such as systems where the Immediate Acceptance mechanism is in place (Agarwal and Somaini (2018), Calsamglia et al. (2018), Kapor et al. (2017), among others), or when the rules/restrictions of the system introduce strategic considerations (Ajayi and Sidibe

(2017), Artemov et al. (2017), Fack et al. (2015)). For instance, a common restriction is to constrain the number of applications that students can submit, as it is the case in Chile, Tunisia, Ghana, among others.

Previous research has exploited well-known properties of the mechanism when trying to identify students' preferences. For instance, a common approach is to assume pairwise stability. In this case, the researcher interprets the enrollment or assignment to be the favorite school among all schools the student is qualified for ex-post (Bordon and Fu (2010), Bucarey (2017), Fack et al. (2015), Artemov et al. (2017), among others). Fack et al. (2015) show that, under certain conditions, stability is a plausible assumption in a large market, as it is an approximate equilibrium outcome of a game of incomplete information. However, when students' information is incomplete, stability is not guaranteed to be an ex-post optimality condition. Although this approach relies mostly on using data on students' enrollment or assignment, Fack et al. (2015) also show that is possible to include information contained in the Rank Order Lists (ROL) using moment inequalities.

Another property of the mechanisms that has been explored for identification is strategy-proofness. When the mechanism used is strategy-proof and students face no other strategic incentives, it is weakly optimal for them to report their true preferences (Haeringer and Klijn (2009)), so the submitted ROLs can be used to identify preferences. This is the case when the mechanism used is the Deferred-Acceptance algorithm (Gale and Shapley (1962)) and there is no constraint on the length of the ROLs, or when students submit a list that is shorter than the maximum allowed and thus the constraint is not binding (Abdulkadiroğlu et al. (2017) and Luflade (2017)). However, this assumption may not always hold. For instance, if students assign zero probability to be admitted to some schools, it is still weakly optimal not to include them in their list. This skipping behavior is also optimal if there are search costs and students believe their admission chances are relatively low. In both cases, assuming that "short-list" students report truthfully would result in biased estimates. One of the main contributions of this paper is to show that this is indeed the case in the Chilean college admissions problem, and that assuming truth-telling of "short-list" students can lead to biased results. Thus, we enrich the recent discussion on whether students are truthful in seemingly strategy proof environments. For example, Fack et al. (2015) analyze the high-school system in Paris, and reject strategy-proofness as an identifying assumption in favor of stability. Shorrer and Sóvágó (2017) study the Hungarian College Admissions process and find that an important fraction of applicants play dominated strategies. In line with the previous studies, Rees-Jones (2017) shows that a significant fraction of residents do not report truthfully in the National Resident Matching Program.

Instead of assuming stability or truth-telling, an alternative approach to estimate student preferences is to model their application behavior. This strand of the literature is very recent, and has mainly focused on school choice settings. Here the researcher interprets the submitted ROL to be the result of an expected utility maximization process given students' beliefs over admission probabilities. Agarwal and Somaini (2018) propose a general methodology to estimate preferences in centralized admission systems where the mechanism can be represented with a cutoff structure. They show that (equilibrium) beliefs can be estimated in a first stage using the data on reports and simulating the assignment process if the researcher is willing to assume a particular structure for beliefs, e.g. rational expectations. After estimating beliefs, they use a likelihood-based approach to estimate preferences in a flexible specification. Kapor et al. (2017) adapt this estimation procedure to incorporate survey data on beliefs and enrollment decisions in the New Haven school choice system. They find that students have biased beliefs over their admission probabilities. In addition, Kapor et al. (2017) use their model to estimate students' preferences and simulate the effects of changing the assignment mechanism in New Haven. Ajayi and Sidibe (2017) analyze the application behavior of students in the centralized high-school system in Ghana. In this case, students can apply to no more than 6 schools, and they do so before knowing their application scores. They model students' beliefs over scores and admission probabilities assuming that students can imperfectly forecast schools' cutoffs using historical data. They propose to jointly estimate beliefs and preferences using an extension of the Marginal Improvement Algorithm (see Chade and Smith (2006)) and the Simulated Method of Moments. Finally, Luflade (2017) analyzes the College Admissions problem in Tunisia. She exploits the sequential implementation of the Tunisian mechanism to identify students' preferences in a first stage. She argues that students who face admission probabilities close to 1 in each round of the mechanism and students who do not completely fill their application list should be interpreted as truth-tellers. Later she estimates students' beliefs over admission probabilities allowing for different levels of sophistication.

In this paper we analyze the Chilean College Admissions problem, where students face a seemingly strategy-proof environment. Even though students are constrained to apply to at most 10 out of 1,400 programs available, only 10% of students submit a ROL for which this constraint is binding. This may suggest that most of the students submit their true preferences. However, we provide evidence that students tend to apply in first preference to programs for which their application score is close to the program's cutoff. Using survey data on students preferences elicited before scores are revealed, we find that students tend to avoid listing programs where their admission probabilities are relatively low. This finding suggests that students who submit short lists shouldn't be interpreted as truth-tellers, and that this behavior is mainly driven by their beliefs on admission probabilities rather than by preference heterogeneity.

Based on this finding, we assume that students do not include programs in their application lists if it is not strictly profitable to do so, and we construct a portfolio problem where students maximize their expected utility of reporting a ROL given their preferences and beliefs over admission probabilities. We adapt the estimation method proposed by Agarwal and Somaini (2018) to solve a large portfolio problem, assuming independence over admission probabilities and rational expectations. A major challenge to implement this methodology is to avoid running into the curse of dimensionality, as the number of potential ROLs grows exponentially with the number of programs. To deal with this, we show that it is sufficient to compare a ROL with only a subset of ROLs ("one-shot swaps") to ensure its optimality, and we incorporate this finding into a Gibbs Sampler estimation algorithm. In addition, we exploit a novel source of variation in the choice environment over admission processes to better identify the model, which is the time variation in admission weights as an exogenous shifter of beliefs on admission probabilities.

We compare the results of this approach against assuming truth-telling of "short-list" students, and we find that assuming the latter leads to biased results. If we assume that students are truthful and ignore the fact that they exclude programs where their marginal benefit is zero, this would result in underestimates of the value of selective programs and overestimates for the value of the outside option. Moreover, assuming truth-telling without taking into account students' beliefs on admission probabilities can lead to overstate the degree of preference heterogeneity in the system. Finally, we show that ignoring the constraint on the length of the list can also result in biased estimates, even if the proportion of constrained ROLs is relatively small.

The closest papers to ours are Ajayi and Sidibe (2017) and Artemov et al. (2017). Our paper complements both from a methodological standpoint, and also in terms of the resulting insights. In Ajayi and Sidibe (2017) students apply before knowing their (unique) application score, so they face two sources of uncertainty: their application score, and the scores and preferences of other students. Hence, students' admission chances are clearly correlated in their setting. In addition, most students in Ghana submit a preference list that contains the maximum number of schools allowed. In the Chilean case, students only face the second source of uncertainty, and most of them report "short-lists". Moreover, Ajayi and Sidibe (2017) use (an extension of) the Marginal Improvement Algorithm (MIA) to approximate heuristically the optimal portfolio problem and simulate data predicted by their model. Later they jointly estimate preferences and beliefs using the Simulated Method of Moments. In contrast, we exploit the richness of our data and the properties of the optimal portfolio to construct identifying restrictions and estimate preferences in a likelihood-based approach. While their estimation procedure is attractive for large scale problems when admission chances are correlated, their identification strategy relies on functional form assumptions on the beliefs formation process and the preference specification. In this sense, one of our main contributions is to adapt the general identification strategy in Agarwal and Somaini (2018) to large-scale portfolio problems. To our knowledge, our method is the first likelihood-based approach to solve large-scale portfolio problems without running into the curse of dimensionality. This methodology can be used to estimate preferences in other large-scale settings of portfolio problems, whenever beliefs on admission probabilities can be estimated in a first stage and assumed to be independent across alternatives.

In the case of Artemov et al. (2017), the authors find a similar pattern in the Australian College Admissions problem, as some students omit programs that are out of their reach. The authors show that assuming truth-telling in this setting can result in biased estimates, and show that stability (similar to Fack et al. (2015)) provides a more robust estimation strategy when students make strategic mistakes. The stability assumption is an attractive alternative for estimating preferences when students do not report truthfully or when they make strategic mistakes. However, as it does not use information on students' beliefs on admission probabilities, the econometrician is unable to extract all the rich information contained in the ROLs for identification. Moreover, understanding students' beliefs becomes important if we want to simulate students' applications, especially in counterfactual scenarios that involve strategic considerations.

The reminder of the paper is organized as follows. In Section 2 we describe the Chilean College Admissions problem, the assignment mechanism and we provide descriptive evidence on the students' application behavior. In Section 3 we present a model of students' portfolio choices. In Section 4 we describe the data and in Section 5 the identification strategy. In Section 6, 7 and 8 we describe the simulations and results. Finally, Section 9 concludes and provides directions for future work.

2 Chilean system

2.1 The Chilean Mechanism

The Chilean university market is semi centralized, with 33 of the most selective universities (close to half) participating in a centralized system run by CRUCH.[1] Students apply directly to

[1] The *Consejo de Rectores de las Universidades Chilenas* (CRUCH) is the institution that gathers these universities and is responsible to drive the admission process, while DEMRE is the organism in charge of applying

the specific major of their choice instead of going to college first. Thus, from now on we refer to the pairs major/university as programs, and we use M to refer to the set of programs. In order to apply to any of the close to 1,400 program in these universities, students undergo a series of standardized tests (*Prueba de Selección Universitaria* or PSU). The PSU tests include Math, Language, Science and History, and each of these tests gives students a standardized score. Students' performance during high-school results in two additional scores, one constructed with the average GPA along high-school (*Notas de Enseñanza Media* or NEM) and the other measuring the relative rank of the student's GPA among his cohort (*Ranking de Notas* or Rank).

After knowing their scores, students can submit a list with no more than 10 programs ranked in strict order of preference[2] at no (monetary) cost. Each program has previously announced its vacancies and a set of admission requirements they will consider for the applications to be valid. However, even if the application to a particular program is not valid, students are still allowed to submit their application list in the system. In addition, each program announces every year a set of admission weights. Application scores are computed as the weighted average of students scores and admission weights. As a result, students' application scores can differ across programs.

Each program's preference list is constructed by ordering admissible students in terms of their application scores. As students can have the same application scores, preferences of programs need not be strict. Considering the preference lists of the applicants and programs, and the vacancies, DEMRE runs an assignment algorithm to match students to programs. The mechanism used is a variant of the student-proposing deferred acceptance algorithm[3] in which tied students in the last seat of a program aren't rejected if vacancies are exceeded. More formally, the allocation rule can be described as follows:

Step 1. Each student proposes to his first choice according to their submitted ROL. Each program rejects any unacceptable student, and if the number of proposals exceeds its vacancies (q), rejects all students whose scores are strictly less than the q-th most preferred student.

Step $k \geq 2$. Any student rejected in step $k-1$ proposes to the next program in their submitted ROL. Each program rejects any unacceptable student, and if the number of proposals exceeds its vacancies (q), rejects all students whose score is strictly less than the q-th most preferred student.

The algorithm terminates either when there are no new proposals or when all rejected students have exhausted their preference lists. The final allocation is obtained by assigning each student to the most preferred program in his ROL that did not reject him. As a side outcome of this, the algorithm generates a set of cutoffs $\{P_j\}_{j\in M}$, where P_j is the minimum application score among students matched to program $j \in M$. Hence, for any student i with ROL R_i and set of scores $\{s_{ij}\}_{j\in M}$, the allocation rule can be described as

$$i \text{ is assigned to } j \Leftrightarrow j \in R_i,\ s_{ij} \geq P_j \text{ and } s_{ij'} < P_{j'}\ \forall j' \in R_i\ st.\ j' \succ_{R_i} j,$$

the selection tests and carrying out the assignment of students to programs.

[2]After scores are announced, they have a period of 5 days to submit their application list, being able to re submit as many times as they want.

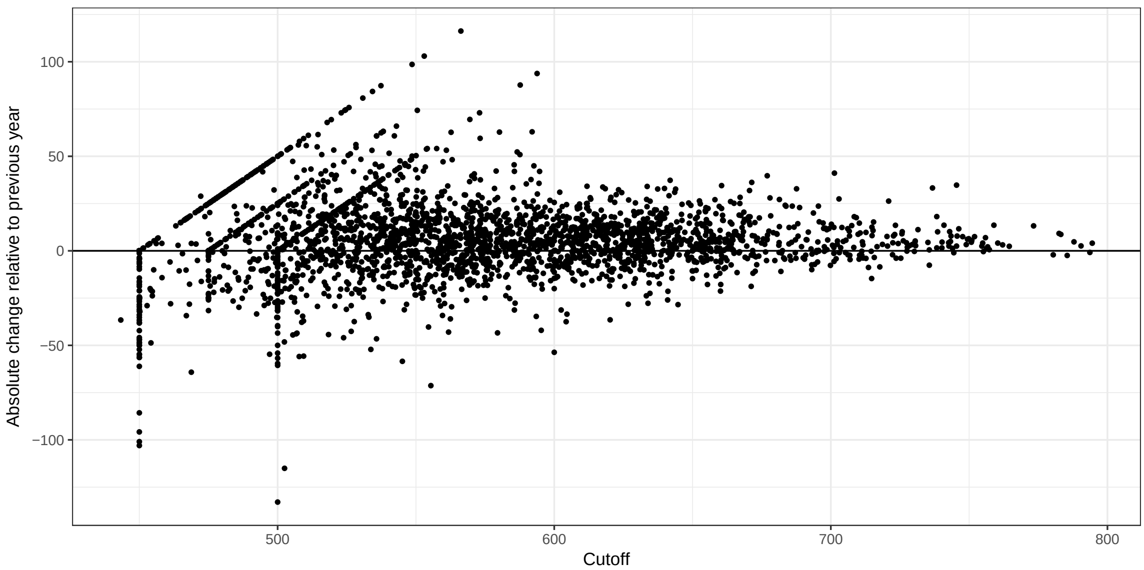
[3]Before 2014 the algorithm used was the university-proposing version. The assignment differences between both implementations of the algorithm are negligible Ríos et al. (2018).

where $\succ_{R_i}$ is a total order induced by R_i over the set $\{j : j \in R_i\}$, such that $j' \succ_{R_i} j$ if and only if program j' is ranked above program j in R_i. This cutoff structure is relevant because it allows to use the framework introduced in Agarwal and Somaini (2018) for identification and estimation. We provide more details in Section 3.2.

2.2 Uncertainty

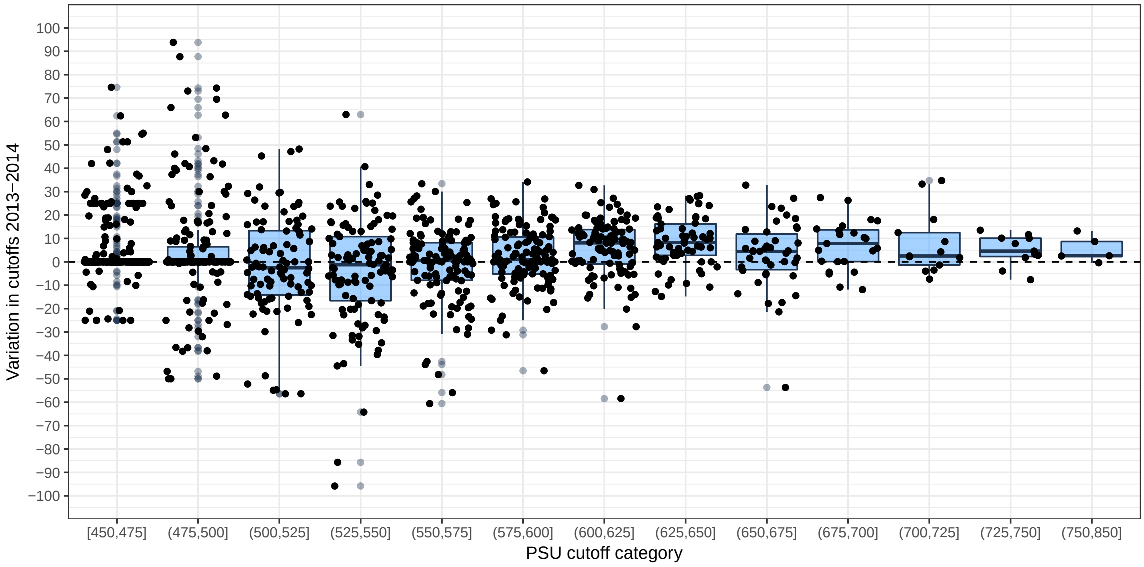
As the Chilean mechanism can be represented using a cutoff structure, the uncertainty that students face on their admission probabilities can be summarized by the variation in cutoffs from year to year. Figures 2.1 and 2.2 show the variation in cutoffs between 2013 and 2014. Each dot shows the variation of a program's cutoff in PSU points with respect to its cutoff in 2013. We observe that less selective programs (with lower cutoffs) had a higher variation in their cutoffs. However, due to minimum score restrictions, there is a mass of low selectivity programs for which their cutoffs do not vary between 2013 and 2014. The uncertainty faced by students is sizable, considering that the standard deviation of PSU test scores is 110 points.

Figure 2.1: Variation in cutoffs 2013-2014



Notes: Scatter plot of absolute variation in cutoffs between 2013 and 2014 with respect to their cutoff in 2014. Each dot represents a program present in both years.

Figure 2.2: Variation in cutoffs 2013-2014 by PSU cutoff category



Notes: Boxplots of absolute variation in cutoffs between 2013 and 2014. Boxplots are computed for different ranges of PSU points.

Given that the cutoff of a given program is by definition the weighted score of its last admitted student, we expect that changes in the number of vacancies would mechanically change the cutoffs if everything stays the same. Variation can also be driven by changes in admission weights, restrictions, and by changes in the population of applicants from year to year. Some of these changes could be anticipated by students, since the number of seats available and the admission requirements of each program are announced before the application process begins.

2.3 Strategic behavior: Selection on Admission Probabilities

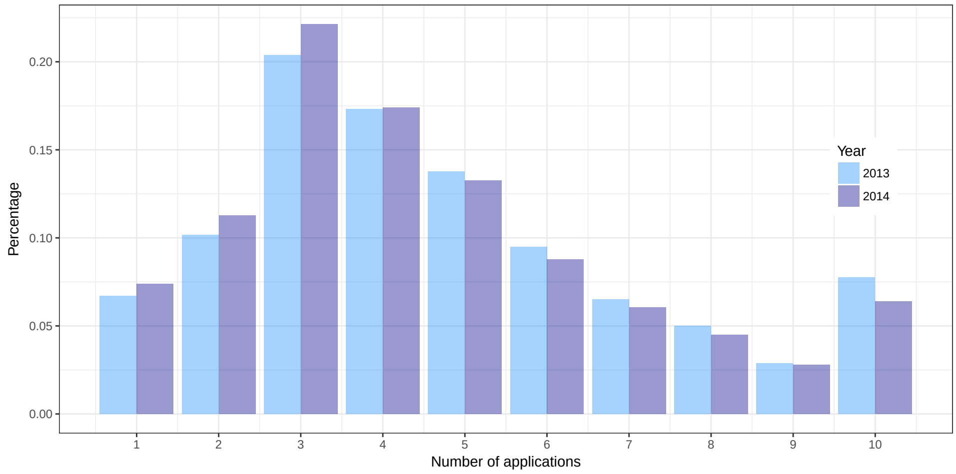
In the previous section we show that students face some degree of uncertainty, as cutoffs are not perfectly correlated across years. A natural question that this raises is whether students take their admission probabilities into account when submitting their applications.

If there were no restrictions on the length of the list, rational students would not need to take their admission probabilities into account in order to choose their (weakly) optimal ROL. As the Chilean mechanism is strategy-proof in the large,[4] (Ríos et al. (2018)) a weakly optimal solution would be to report their true preferences. However, as truth-telling is only weakly-optimal, this rationale does not rule out the possibility that students are misreporting their true preferences, even when restrictions on the length of the list are not binding. We show evidence that this is indeed the case in Chile.

Figure 2.3 shows the distribution of the number of programs considered in the ROL among students who applied to at least one program. We observe that more than 90% of students apply to less than the maximum number of programs allowed (10).

[4]A mechanism is strategy-proof in the large (SP-L) if, for any full-support i.i.d. distribution of students' reports, being truthful is approximately optimal in large markets Azevedo and Budish (2017).

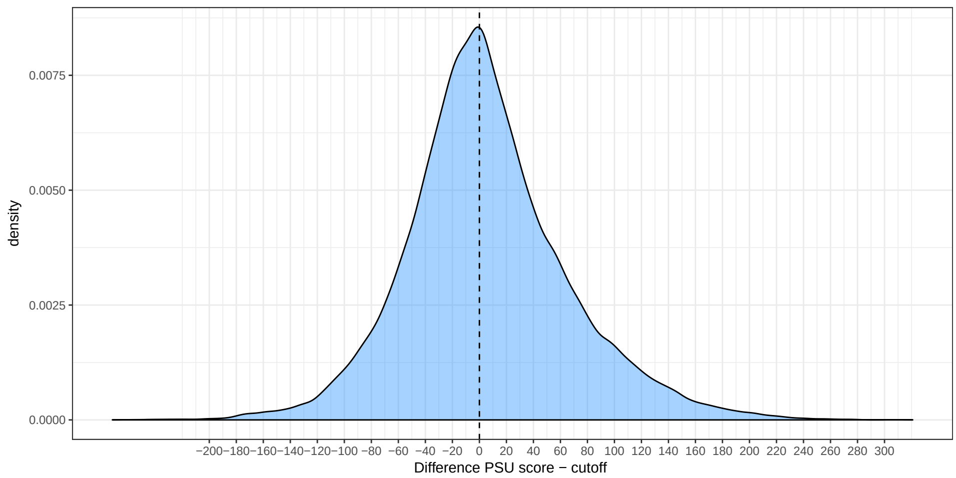
Figure 2.3: Number of applications per student by year



Notes: Distribution of the number of applications per student (length of the ROLs) by year.

In addition to the constraint on the number of programs that can be part of a ROL, some universities add an additional constraint that restricts the position that certain programs can take in a student's report. For instance, the two most selective universities (PUC and UCH) require applicants to apply within the top 4 preferences. This restriction introduces incentives to misreport preferences (Lafortune et al. (2016)), and could explain why students are strategic when choosing where to apply. However, we observe that strategic behavior is present even when these constraints are not binding. Indeed, we observe that students tend to list as their top choice programs for which their scores are close to the cutoff of the current year. In Figure 2.4 we show the distribution of the difference between the weighted score of each student and the cutoff of the program they applied to in first preference. We observe a peak at 0, showing that students tend to list as their top choice, programs for which their weighted score is around the cutoff.

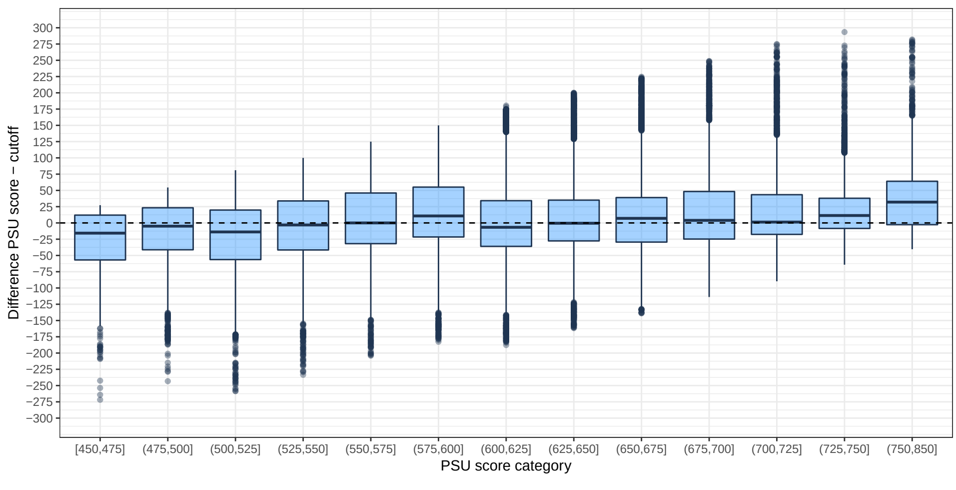
Figure 2.4: Distribution of difference between PSU score and cutoff for first listed preference in 2014



Notes: Empirical (nonparametric) distribution of the difference between each student's PSU score and the cutoff for his/her first listed preference in 2014.

This pattern is still present after controlling for students' weighted scores. Figure 2.5 shows boxplots of the difference in points between the weighted score of the applicant and the cutoff of the program that each student listed in first preference. We observe that the median is around 0 for most categories, showing that students tend to apply in first preference to programs for which their weighted score is around the cutoff.

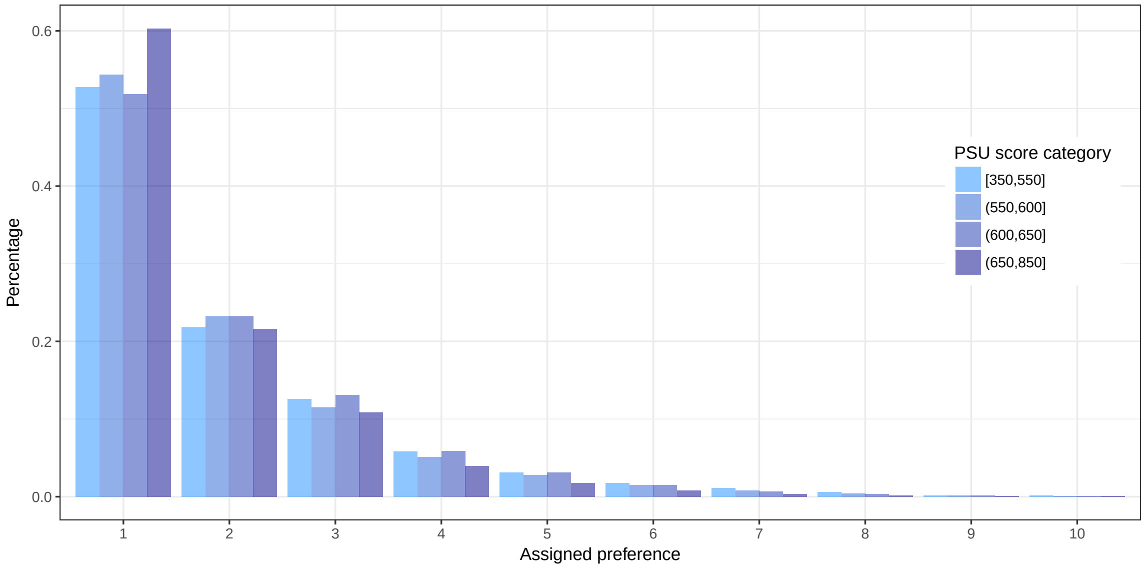
Figure 2.5: Boxplots of difference between PSU score and cutoff for first preference in 2014



Notes: Boxplots of the difference between each student's PSU score and the cutoff for his/her first listed preference in 2014. Each boxplot is computed by different PSU ranges of the cutoff. The solid horizontal lines show the medians for each boxplot and the dashed line is a reference horizontal line at zero.

We also find evidence for this application pattern in the assignment results. Figure 2.6 shows, for different cutoff ranges, the distribution of preference of assignment per student. We observe that the share of students assigned on each preference is roughly the same for different score categories. For instance, we observe that the fraction of students assigned to their top choice is similar regardless of their scores, suggesting that students with lower scores apply in their top preference to programs with lower cutoffs.

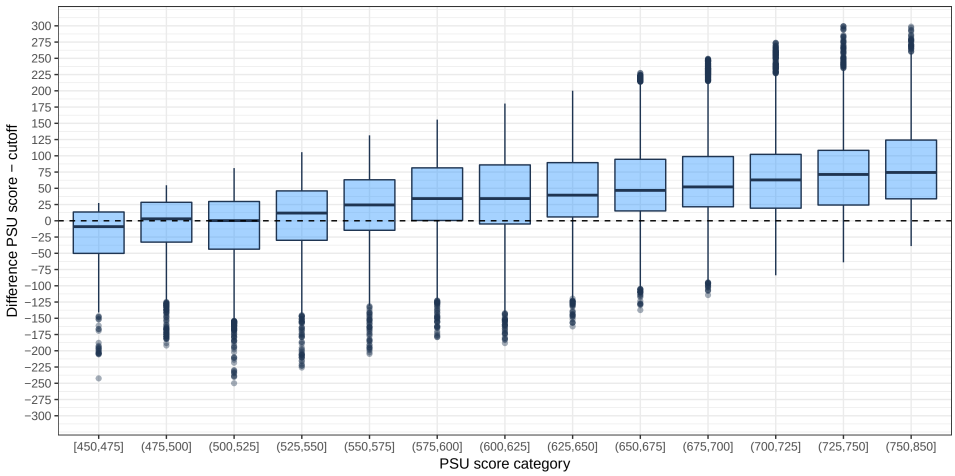
Figure 2.6: Preference of assignment by PSU score category in 2014



Notes: Distribution of the preference in which each student was assigned in 2014. Each distribution is computed by different student's PSU score ranges.

Compared to the application behavior in first preference, we observe in Figure 2.7 that students tend to apply above the cutoffs for their last submitted preference.

Figure 2.7: Boxplots of difference between PSU score and cutoff for last preference in 2014



Notes: Boxplots of the difference between each student's PSU score and the cutoff for his/her last listed preference in 2014. Each boxplot is computed by different PSU ranges of the cutoff. The solid horizontal lines show the medians for each boxplot and the dashed line is a reference horizontal line at zero.

The previous results suggest that students are taking into account their admission probabilities in their application behavior. However, we cannot disentangle how much of this pattern is driven by preference heterogeneity and by beliefs on admission probabilities. For instance, the cutoff can be thought of as a signal of how demanding the program is, and some students may prefer to attend easier ones. To show that this is not the case and that the main force driving this application pattern are beliefs on admission probabilities, we use information from two surveys conducted in 2014 and 2018.

2.3.1 Surveys

Survey on true preferences (pre-applications, 2014). In 2014, CRUCH and DEMRE conducted a survey intended to elicit students' "true" preferences for programs. The survey was sent via email to all participants (roughly 200,000) between October and November, i.e., before the standardized national exam. As a results, all students that responded to the survey (roughly 40,000) did not know their scores at the time of the survey.

Among other questions, students were asked the following:

"The following question is intended to elicit your true preferences over programs [...]" "If you could choose any program to study. In order of priority (the first being the most preferred), what would be the 3 programs chosen by you?"

Due to constraints imposed by DEMRE, the survey only elicited the major of preference but not the most preferred program (pair major/university). This complicates the analysis because there exists large heterogeneity in programs' selectivity across universities. To avoid this problem, we focus on students who reported Medicine as their top preference in the survey. The reason to focus on Medicine is that it is a very selective major, so the cutoffs across universities are close to each other and the minimum cutoff is relatively high. Among the 40,000 students who answered the survey, close to 10% (3,797) reported Medicine as their top preference, and 2,987 of these students ended-up applying to the system. Among these, only 1,360 listed Medicine as their top preference (1,600 in some preference).

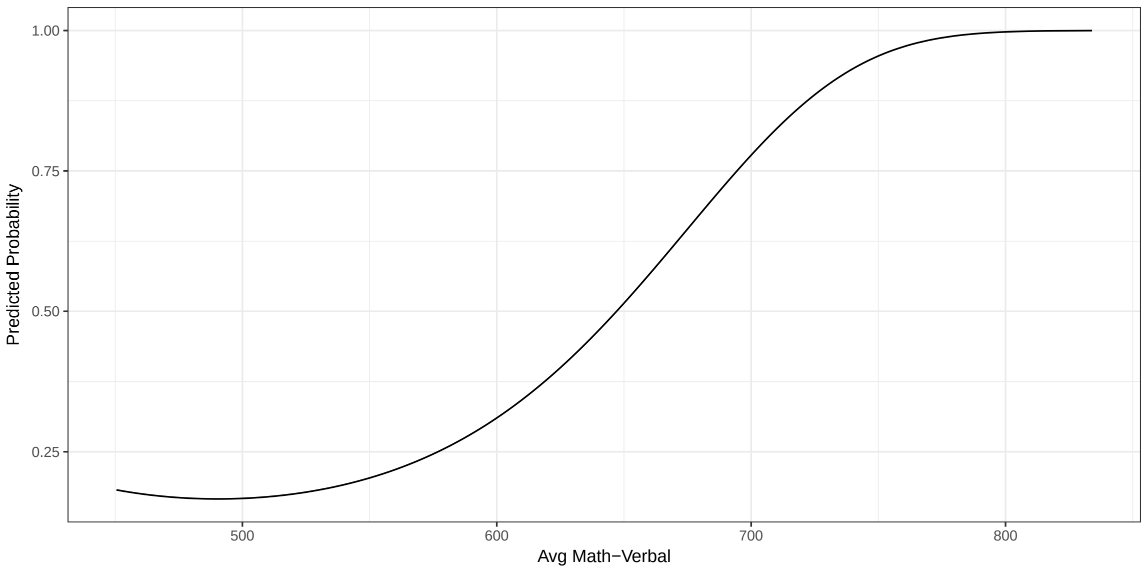
Table 2.1 shows a probit regression where the dependent variable takes value 1 if the student (who listed Medicine in first preference in the survey) applied to Medicine in first preference in the application process. We observe that students with lower scores (thus, with lower chances of been accepted) are more likely to omit Medicine from their top preference. Given how the question was phrased, some students could have interpreted that the report of their first preference in the survey was without considering the tuition of the programs. However, after controlling for family income, the average score of the student is still a statistically significant predictor for omitting or not Medicine in first place. Moreover, looking at the predicted probability of applying to Medicine in first preference, in Figure 2.8, we observe a sharp increase in the range between 600 and 750 points, which is the range where most cutoffs are located.

Table 2.1: Probability of applying to Medicine in first preference, conditional on reporting Medicine in first preference in the survey

	Dependent variable:	
	Apply to Medicine in First Preference	
	(1)	(2)
Math-Verbal[1]	-0.039^{***}	-0.038^{***}
	(0.006)	(0.006)
Math-Verbal squared [2]	0.00004^{***}	0.00004^{***}
	(0.00000)	(0.00001)
Constant	8.640^{***}	8.315^{***}
	(1.863)	(1.875)
Family Income[3]	No	Yes
Observations	2,907	2,919
Notes:		*p<0.1; **p<0.05; ***p<0.01

[1] Average score between Math and Verbal.
[2] Square of average score between Math and Verbal.
[3] Gross (self-reported) family income.

Figure 2.8: Predicted probability for applying to Medicine in first preference



Notes: Predicted probability for applying to Medicine in first preference conditional on reporting Medicine in first preference in the survey. The model does not include family income as a covariate.

These results suggest that students take into account their admission probabilities in their applications, and some of them do not apply to their most preferred programs if their admission probabilities are too low. This result is true even for students who do not face clear strategic incentives given by restrictions in the length of the list in our sample (87% of the students who declared Medicine as their most preferred program in the survey and submitted an application are “short-list” students). This skipping pattern is also described in (Fack et al., 2015) and Artemov et al. (2017).

Survey on true preferences and beliefs (post-applications, 2019). A caveat of the previous survey is that it does not explicitly elicit beliefs on admission probabilities nor true preferences for major-university. To accomplish this, we conducted a new survey right after the application process in 2019. The survey was sent to 154,366 applicants,[5] all of whom knew their scores and already submitted their ROL.[6] Among other questions, students were asked:[7]

"This question aims is to know where you would have applied to in the hypothetical case in which your admission did not depend on your scores. Remember that this is only a hypothetical question and will not have any effect on your application nor in your admission probabilities. If the admissions process did not depend on your PSU score, nor your NEM scores, nor your Ranking scores. What would have been your main program choice?"

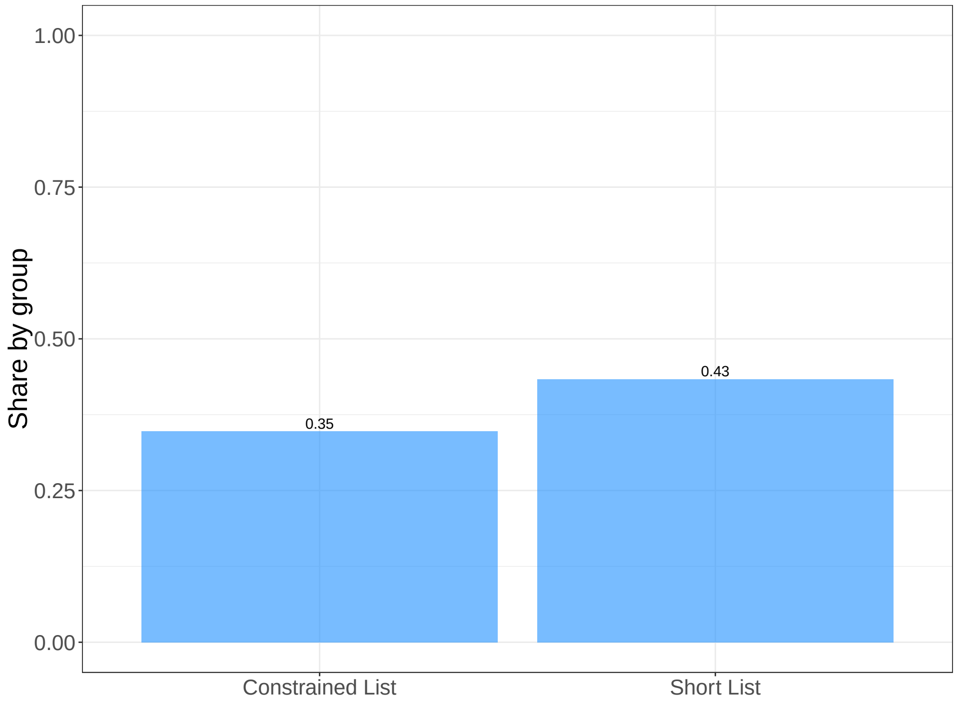
We label as *truth-tellers* all students who answered the same program they listed in their application as their top preference. In Figure 2.9 we show the share of truth-tellers in the survey, separating between those who reported constrained ROLs and those who did not (i.e., reported a short lists). We observe that only 43% of short-list students report their true first preference. On the other hand, for students who list the maximum of programs allowed (constrained lists), the share of truth-tellers is even lower. This suggests that students who report constrained lists are indeed facing more strategic incentives to misreport their true preferences than students who report short-lists.

[5]We received 54,997 incomplete responses and 38,979 full responses.

[6]We describe and analyze this survey in more detail on Larroucau et al. (2019).

[7]We decided to randomize over two versions of this question. The second version included explicitly that in the hypothetical scenario the program was also tuition-free. We do not find significant differences in terms of the share of truth tellers in both versions of the question.

Figure 2.9: Share of survey truth-tellers



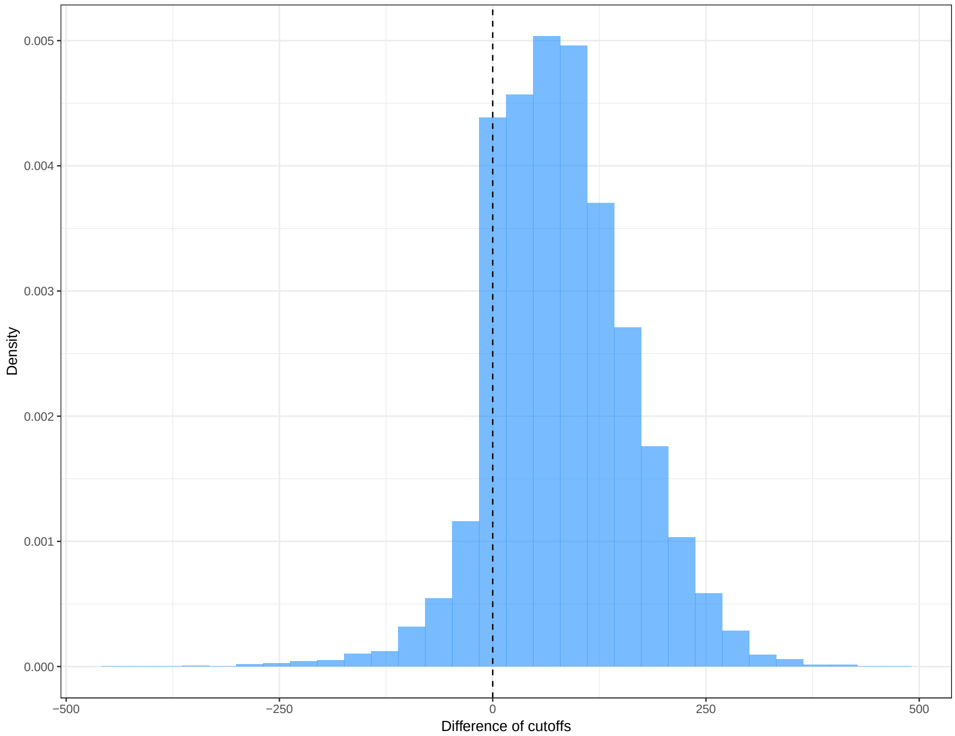
Notes: Share of truth tellers from the sample of students who completed the survey. Results are computed separately for students who submitted short lists (strictly less than 10 programs in their ROLs) and constrained lists (listed 10 programs in their ROLs).

To know whether this skipping behavior is due to beliefs on admission probabilities, we included the following question:

"What do you think is going to be this year's cutoff score in this program?"

This question allows us to obtain a proxy of students' beliefs on their admission probabilities for their first true preference. Figure 2.10 shows the distribution of the difference between the expected cutoff for the first true preference and the expected cutoff for the first listed preference for students who are not truth-tellers. We observe that the distribution is skewed towards positive numbers, which implies that most students who did not include their first true preference in their application list expected a higher cutoff than the cutoff for their first listed preference. This result is again consistent with the hypothesis that students tend to skip programs for which their (believed) admission probabilities are too low.

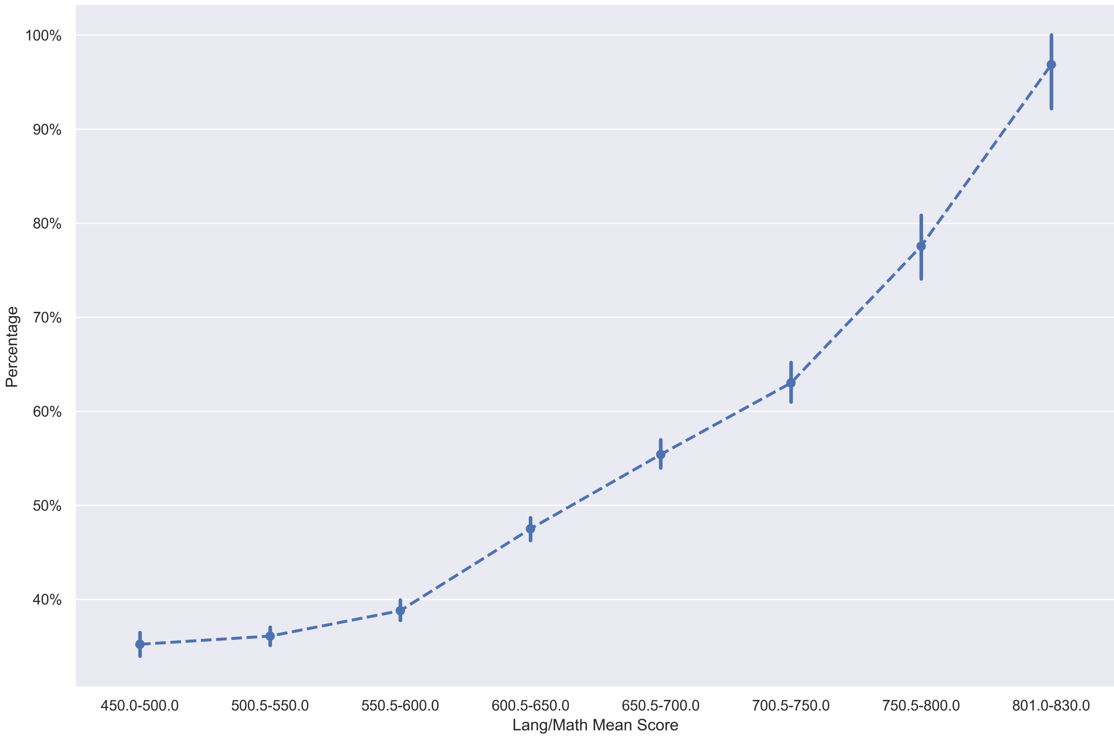
Figure 2.10: Expected cutoffs for first true preference vs first listed preference for students who misreport



Notes: Histogram of the difference between expected cutoff for first true preference and expected cutoff for first listed preference. Results are computed only for the population of students who are not considered as truth tellers in the survey.

Finally, as a robustness check, we calculate the share of truth-tellers by different ranges of average score in Mathematics and Verbal. Figure 2.11 shows that the share of truth-tellers increases for students with higher scores. This result is consistent with the fact that students with higher scores have less constrained choice sets, facing lower incentives to behave strategically and misreport their true preferences.

Figure 2.11: Share of survey truth-tellers by average score



Notes: Share of truth tellers from the sample of students who completed the survey. Results are computed separately for different ranges of the average score in Mathematics and Verbal.

The previous evidence suggest that, upon indifference, students tend to left-censor their application lists relative to their true preferences. This motivates the following assumption:

Assumption 1. *A student won't include a program in the portfolio unless it is strictly profitable to do so.*

Given that there are no monetary application costs in our setting, Assumption 1 could be microfunded by including an information acquisition cost due to search frictions. Assumption 1 has two relevant implications:

1. Students won't include programs for which their admission probabilities are 0.
2. If a student includes a program in the list for which his admission probability is exactly 1, then he won't include programs below it.

If Assumption 1 holds for a significant fraction of students, assuming that short-list students are truth tellers would be misleading to understand students' preferences in this setting.

3 Model

Consider a finite set of students N and a finite set of programs M. Each student $i \in N$ is characterized by a vector of indirect utilities u_i and a vector of scores s_i. Each program $j \in M$ is characterized by its number of vacancies q_j and a vector of admission weights ω_j. The application score of student i in program j, s_{ij}, is given by:

$$s_{ij} = \sum_k \omega_j^k s_i^k.$$

We denote by P_j the application score of the last admitted student to program j, and we refer to it as the cutoff.

3.1 Preferences

Let $u_i \sim f_u(u)$ be the vector of indirect utilities of student i. In particular, we assume that the indirect utility of student i for program j can be written as:

$$u_{ij} = u(Z_{ij}, d_{ij}, \xi_j) + \varepsilon_{ij}, \tag{3.1}$$

where Z_{ij} are observable characteristics of student i and program j, d_{ij} is the distance between program j and student i, and ε_{ij} is an idiosyncratic preference shock for which we impose additive separability. We also include an unobserved component ξ_j to capture characteristics that are unobserved by the econometrician.

For identification, we specify a location normalization and a scale normalization. Unless stated differently, we normalize the indirect utility of the outside option to 0,[8] i.e., $u_{i0} = 0$. We can interpret u_{i0} as the value that student i gets for being unassigned in the centralized system. As the Chilean system is semi centralized, this outside option includes the possibility of enrolling in institutions that are not part of the system, but also the possibility of reapplying to programs in the centralized system in a following year. The scale normalization we consider is to set the standard deviation of the unobserved shock ε_{ij} to 1.[9]

3.2 Beliefs: Rational Expectations

As we discussed in Section 2.3, students not only care about the utility they derive from each program, but also about the probability of being admitted in those programs. Estimating these probabilities is a complex task, since students have no information about other students' preferences and scores. However, the cutoff structure of the mechanism (see Section 2.1) allows to summarize all the relevant uncertainty faced by each student in a distribution over cutoffs.

[8] Depending on the simulation exercise that we will perform later, we will also work with an alternative normalization, normalizing the systematic part of the utility of being unassigned to 0.

[9] We also consider an alternative normalization that is common in the school choice literature, setting the coefficient of distance to -1.

As a baseline model, we assume that students have rational expectations about their admission probabilities. In particular, we assume that students know the distribution of indirect utilities as well as the strategies used by other students when submitting their ROLs. Hence, students can infer the conditional distribution of reports. In addition, we assume that students know the distribution of scores, and combining these two sources of information they can infer the distribution of cutoffs. We will further assume that students take these distributions to be independent across programs. These considerations are summarized in Assumption 2.

Assumption 2. *Students have rational expectations over their admission probabilities and take the distributions over cutoffs to be independent across programs.*

Agarwal and Somaini (2018) show that a consistent estimator of these beliefs can be obtained using the following bootstrap procedure:

- For each bootstrap simulation $b = 1, \dots, B$,
  - Sample with replacement a set N^b of N students with their corresponding ROLs and scores.
  - Run the mechanism to obtain the allocation μ^b.
  - Obtain the set of cutoffs $\left\{P_j^b\right\}_{j \in J}$ from the allocation μ^b, i.e. for each $j \in J$,

$$P_j^b = \min\left\{s_{ij} : i \in N^b,\ \mu^b(i) = j\right\}$$

- We can estimate the admission probability of student $i \in N$ in program $j \subset J$ as

$$\hat{p}_{ij} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{\left\{s_{ij} \geq P_j^b\right\}}$$

We estimate these probabilities running $B = 10,000$ bootstrap simulations. The bootstrapped realizations of cutoffs show positive but small correlations among programs that students tend to rank together in their application lists. However, Assumption 2 states that students don't take into account this dependency to form their beliefs over their admission probabilities. Thus, they infer their admission probabilities from the marginal distributions of cutoffs.

3.3 Optimal Portfolio Problem

The portfolio problem for college applications was introduced by Chade and Smith (2006).[10] In this problem, a student must choose a subset S of colleges to which to apply for admission, incurring in an application cost $c(S)$. Formally, let $\mathcal{R}$ be the set of possible ROLs. Consider a fixed student $i \in N$ and let $U : \mathcal{R} \to \mathbb{R}$ be a function that, for each ROL $R \in \mathcal{R}$, returns the expected utility given a set of beliefs on admission probabilities $\{p_j\}_{j \in M}$ and a set of indirect

[10]This problem is a particular case of the simultaneous selection problem presented in Olszewski and Vohra (2016).

utilities $\{u_j\}_{j\in M}$.[11] Then, given Assumption 2 and a ROL $R = \{r_1, \ldots, r_k\}$,

$$U(R) = z_{r_1} + (1 - p_{r_1}) \cdot z_{r_2} + \ldots + \prod_{l=1}^{k-1}(1 - p_{r_l}) \cdot z_{r_k}, \tag{3.2}$$

where $z_j = u_j \cdot p_j$ for each $j \in M$.

The problem faced by student $i \in N$ is to choose a ROL $R \in \mathcal{R}$ without exceeding the maximum number of applications K, in order to maximize his expected utility $U(R)$, given his indirect utilities $\{u_j\}_{j\in M}$, his beliefs over admission probabilities $\{p_j\}_{j\in M}$ and application costs $c(R)$, i.e.

$$R \in \underset{R'\in\mathcal{R}, |R'|\leq K}{\text{argmax}} \; U(R') - c(R'). \tag{3.3}$$

Chade and Smith (2006) show that the optimal portfolio problem is NP-Hard. However, when admission probabilities are independent[12] and the cost of applying to a subset of programs S only depends on its cardinality, i.e. $c_i(S) = c(|S|)$ for some function c, the unconstrained problem is Downward Recursive and the optimal solution is given by a greedy algorithm called Marginal Improvement Algorithm (MIA).

MIA: Marginal Improvement Algorithm (Chade and Smith (2006))

- Initialize $S_0 = \emptyset$
- Select $j_n = \arg\max_{j\in M\setminus S_{n-1}}\{U(S_{n-1} \cup j)\}$
- If $U(S_{n-1} \cup j_n) - U(S_{n-1}) < c\,(S_{n-1} \cup j_n) - c\,(S_{n-1})$, then STOP.
- Set $S_n = S_{n-1} \cup j_n$

MIA recursively adds programs that give the highest marginal improvement to the portfolio, as long as they exceed the marginal cost of adding them. Olszewski and Vohra (2016) show that MIA also returns the optimal ROL when the number of applications is constrained and when $c(S)$ is supermodular.

In our setting, there is no monetary application cost, i.e. $c(R) = 0, \forall R \in \mathcal{R}$. However, Assumption 1 implies that a student will not include a program to the portfolio unless the marginal improvement is strictly greater than 0. In order to account for this, we need to modify MIA's stopping criterion:

MIA + A1: If $U(S_{n-1} \cup j_n) - U(S_{n-1}) \leq c\,(S_{n-1} \cup j_n) - c\,(S_{n-1}) = 0$, then STOP.

Clearly A1 does not affect the optimality of MIA because the value of the portfolio does not change when the marginal improvement of adding a new program is zero.

[11] We omit the dependency on the index i to simplify notation.

[12] Notice that in our case we have assumed in Assumption 2 independence of beliefs on admission probabilities.

3.3.1 One-Shot Swap optimality

The fact that an observed ROL R is optimal provides information about the utilities that are consistent with its optimality. In particular, as students are utility maximizers, the observed ROL R is the one that maximizes student i's expected utility, i.e.

$$U(R) \geq U(R'),\ \forall R' \in \mathcal{R},$$

and when the number of applications is constrained to at most K preferences, an observed ROL R satisfies

$$U(R) \geq U(R'),\ \forall R' \in \bigcup_{l=1}^{K} \mathcal{R}_l, \tag{3.4}$$

where $\mathcal{R}_k$ is the set of ROLs of length k, for $k \in \mathbb{N}^+$.

For a fixed set of beliefs on admission probabilities, Equation 3.4 characterizes the set of utilities $u = \{u_j\}_{j \in M}$ that rationalizes the submitted ROL R to be optimal. However, as the set of possible ROLs grows exponentially, the constraints imposed by Equation 3.4 cannot be used without running into the curse of dimensionality.

Let $\mathcal{P}(R)$ be the set of ROLs R' that can be obtained from making a one-preference permutation of programs within ROL R. We call the ROLs in $\mathcal{P}(R)$ *One-Shot Permutations*(OSP) of ROL R. In addition, let $\mathcal{S}(R)$ be the set of ROLs R' which differ in only one program relative to R, i.e.

$$\mathcal{S}(R) = \left\{R' \in \mathcal{R}_{|R|} :\ \left|R \cap R'\right| = |R| - 1\right\}.$$

We call the ROLs in $\mathcal{S}(R)$ *One-Shot Swaps* (OSS) from ROL R. In a slight abuse of notation, we denote by $\mathcal{S}_j(R)$ the set of one-shot swaps involving program $j \in M \setminus R$, and $\mathcal{S}_{jkl}(R)$ the set of one-shot swaps that replace program r_k with program j, placing the latter in the l-th position of the new ROL.

Notice that any utility maximizing ROL $R = \{r_1, \ldots, r_k\}$ satisfies two conditions:

- $u_{r_1} \geq u_{r_2} \geq \ldots \geq u_{r_k}$, and
- $p_{r_j} > 0$ for each $j = 1, \ldots, k$.

Therefore, we can without loss of generality restrict our attention to those OSS that satisfy these conditions.[13]

Example 3.1. Suppose that $K = 3$, $R = \{ABC\}$, $M = \{A, B, C, D\}$ and that $p_j > 0\ \forall j \in M$. Then,

$$\mathcal{P}(R) = \{ACB, BAC\},$$

and

$$\mathcal{S}(R) = \{ABD, ADB, DAB, ACD, ADC, DAC, BCD, BDC, DBC\}.$$

It is easy to check that the set of constraints

$$U(R) \geq U(R'),\ \forall R' \in \mathcal{P}(R)$$

[13] A ROL R' that does not satisfy these conditions will be weakly dominated by another ROL that is either re-ordering or a subset of the elements of R'.

implies that $u_{r_1} \geq u_{r_2} \geq \ldots, u_{r_{|R|}}$. In Proposition 1 we show that it is sufficient to consider constraints involving one-shot permutations and one-shot swaps in order to ensure the optimality of ROL R.

Proposition 1. Let $R = \{r_1, \ldots, r_k\}$ be a ROL of length at most K, i.e. $k \leq K$. If

$$U(R) \geq U(R'), \; \forall R' \in \mathcal{P}(R) \cup \mathcal{S}(R) \tag{3.5}$$

then

$$U(R) \geq U(R'), \; \forall R' \in \bigcup_{l=1}^{K} \mathcal{R}_l \tag{3.6}$$

Proof. See Appendix A. □

Notice that the cardinality of the set $\mathcal{S}(R)$ is just $|R|^2 \times (M - |R|)$, because we can replace any of the $|R|$ programs in R with one of the $M - |R|$ programs not in R and list that program in $|R|$ different positions. Thus, the number of inequalities grows linearly with the number of programs, reducing considerably the number of inequalities required to characterize preferences.

Our next proposition shows that it is not necessary to consider all possible one-shot swaps. In particular we show that, if a program $j \in M \setminus R$ is such that $p_j > 0$ and $p_j \geq p_{r_k}$ for some program $r_k \in R$, then we can eliminate some one-shot swaps that will be redundant, further reducing the size of the problem.

Proposition 2. Let $R = \{r_1, \ldots, r_k\}$ be a ROL of length at most K, i.e. $k \leq K$, and suppose there exists a program $j \in M \setminus R$ such that $p_j > 0$ and $p_j \geq p_{r_k}$ for some $r_k \in R$. Let $\underline{k} = \arg\min \{p_{r_k}\}$, and let $\bar{k} = \max\{k \; : \; p_{r_k} \leq p_j\}$. Then,

$$U(R) \geq U(R'), \; \forall R' \in \bigcup_{l=1}^{K} \mathcal{R}_l$$

if and only if

$$U(R) \geq U(R'), \; \forall R' \in \mathcal{P}(R) \cup \mathcal{S}_u(R),$$

where $\mathcal{S}_u(R)$ is the set of un-dominated OSS and is given by

$$\mathcal{S}_u(R) = \left(\bigcup_{\substack{l=\underline{k} \text{ st.} \\ p_{R(l)} \leq p_{R(\bar{k})}}}^{\bar{k}} \left(\mathcal{S}_{jll}(R) \cup \bigcup_{k=\bar{k}+1}^{K} \mathcal{S}_{jkl} \right) \right) \cup \left(\bigcup_{k=\bar{k}+1}^{K} \bigcup_{l=\bar{k}+1}^{K} \mathcal{S}_{jkl}(R) \right).$$

Proof. See Appendix A. □

In Section 2.3 we show that students tend to apply in their last preference to programs for which their scores are far above the cutoff of the previous year (see Figure 2.7), and certainly leave out of their ROL programs for which their probability of admission is at least as high. Proposition A.2 guarantees that we do not need to consider one-shot swaps involving those programs, reducing considerably the number of inequalities to consider.

4 Data

Our dataset spans from 2012 to 2014, and includes students' scores, the admission weights for each program, restrictions for applicants, and the final assignment. In addition, we have students' socioeconomic characteristics, including self-reported family income, parents education, municipality where the student lives, among others.

We complement the previous data with information about the characteristics of the universities and programs inside the centralized system, including tuition, duration, major, and programs' location. We also have aggregate information about the labor market prospects of each program, like post-graduation expected income and employment probability.

Tables 4.1 and 4.2 show aggregate statistics about the admission processes from 2012 to 2014.

Table 4.1: Aggregate Statistics Admission Process 2012-2014

	2012	2013	2014
Participants[1]	280,049	280,510	278,736
Applicants	116,336	118,212	119,161
Effective Applicants[2]	106,719	107,550	106,804
Assigned	93,574	95,304	95,568
Universities	33	33	33
Programs	1,335	1,395	1,419
Vacancies[3]	113,231	112,608	110,380

[1] Students who register to take the PSU tests in the current year and/or who can participate in the current admission process using their previous year scores.
[2] Students who submit a ROL with at least one valid application.
[3] Does not include vacancies for the affirmative action track (BEA process).

Table 4.2: Students' Demographics Admission Process 2012-2014

		2012	2013	2014
Applicants		116,336	118,212	119,161
Gender	Female	52.6 %	52.2 %	52.8%
Average Scores	Math/Verbal[1]	574.9	572.5	569.6
	NEM[2]	583.3	581.8	583.6
	Rank[3]	0	604.5	609.6
Income[4]	[0,$288]	40%	36.7%	33%
	[$288,$576]	25.6%	27.2%	28.3%
	[$576,$1,584]	22%	23.1%	24.8%
	>$1,584	12.4%	13%	14%
High-School	Private	18.7%	18.5%	18.5%
	Voucher[5]	52.5%	53.5%	53.5%
	Public	28%	27.3%	27.3%

[1] Score constructed with the average Math score and Verbal score. For students using scores from previous year, we considered the maximum of both averages.
[2] Score constructed with the average grade along high-school.
[3] Score constructed with the relative position of the student among his/her classmates.
[4] Gross Family monthly income in thousands Chilean pesos (nominal). Ranges are constructed by grouping income categories provided by DEMRE. As a reference point, according to CASEN survey, for 2013 the average autonomous monthly family income of the sixth decile was $573,981 Chilean pesos.
[5] Partially Subsidized schools.

5 Identification

As the Chilean mechanism has a cutoff structure and we have constructed beliefs as described in Section 3.2, our model belongs to the general class of models described in Agarwal and Somaini (2018). Under certain conditions, it is possible to obtain non-parametric identification using two sources of variation: (1) including a "special regressor"; and (2) using variation in the choice environment.

The "special regressor" is a covariate that varies between students and programs, that enters additively into the utility function, and that is orthogonal to unobservables. Agarwal and Somaini (2018) propose geographic distance between the school and the student's home as a "special regressor" in their school choice application. In our setting, distance between students' homes and the location of the program they are applying to will also help to identify preferences. However, given that we only observe the municipality of students' homes address and the municipality where programs are located, our measures of distance are rather coarse. Moreover, how important is geographic distance within a city for determining students' choices is unclear. This will likely impact in a lack of identifying variation compared to the school choice setting in which, arguably, distance can be a more important driver for students' (or parents') choices.

To complement the previous source of identification, we use variation in the choice environment faced by students from 2012 to 2014, exploiting a particular feature of the Chilean system: every year, programs can change the weights they assign to each admission factor.[14] As explained before, to apply to a program students undergo a series of standardized tests including Math, Language, and a choice between Science or History, providing a score for each of them. In addition to the PSU scores, students also obtain a score that depends on their average grade along high-school (*Notas de Enseñanza Media* or NEM). Finally, starting in 2013, students receive an additional score that depends on the relative position of the student among his/her cohort (*Ranking de Notas* or Rank). To evaluate the effects of this new factor, CRUCH established that every program had to assign a weight of 10% to the rank score in 2013. In 2014 this restriction was removed, and each program was allowed to choose any weight between 10% and 40%. The inclusion of the rank score in two stages generates variation in the admission weights for each program, which translates into an exogenous variation in the admission probabilities that students faced in those years.

Table 5.1 shows the mean variation of each admission weight between 2012 and 2013, grouped by university. Universities had to choose, for each program, which admissions weights to decrease in order to give a weight of 10% to the rank score. From the 33 universities in the centralized system, more than a third reduced close to 10% the weight for the NEM of their programs. As the rank score is constructed as a function of NEM[15] and both admission scores are highly correlated, we do not expect a big variation on the admission probabilities between years 2012 and 2013. The main source of variation on admission probabilities comes from the change on admission weights from 2013 to 2014.

[14]Weights can vary by program within a given university and also across universities.

[15]Being equal to the NEM score for all students who are bellow the historical average NEM of their High-schools. For more details see Larroucau et al. (2015).

Table 5.1: Variation on Admission weights 2012-2013

Universities	Programs with changes in Rank	Variation on admission weights[1]					
		Rank	GPA	Math	Verbal	History	Science
PUC	41	10	0	-5.1	-3.5	0	-1.3
PUCV	51	10	-8.1	-0.5	-1.3	-0.1	-0.1
UACH	46	10	-4.6	-1	-3.5	-0.5	-0.5
UAH	25	10	-6.2	-0.4	-2.6	-0.8	-0.6
UAI	10	10	-10	0	0	0	2
UANDES	22	10	-10.7	1.4	-0.7	-1.6	0.7
UANT	34	10	0	-4.9	-5.1	0	0
UBB	39	10	-5.1	-1.3	-2.1	2.2	0.9
UCH	48	10	-10	-0.3	0	0.3	0
UCM	25	10	-5	-2.8	-1	-0.6	-0.6
UCN	40	10	-5	-0.1	-4.9	0	0
UCSC	31	10	-0.5	-2.7	-6.3	-0.5	-0.5
UCT	37	10	-6.5	-1.5	-1.9	0.4	0
UDA	24	10	-13.5	3.8	-0.4	0.2	0.2
UDD	36	10	-10	-0.3	0.3	0	0
UDEC	87	10	-10	0	0	0	0
UDP	26	10	-10	-0.2	0	-0.2	0.2
UFRO	38	10	-10	0	0	0	0
UFT	17	10	-3.8	-2.9	-3.8	0.6	0.6
ULA	27	10	-9.3	0.7	0.2	-0.2	0.7
ULS	35	10	-5.6	0.3	-3.9	0	-0.9
UMAG	22	10	-0.9	-3.4	-3.6	-0.7	-3.4
UMAYOR	50	10	-11.2	0.1	0.4	0.7	0.7
UMCE	22	10	-8.9	-1.8	0.7	0.7	2
UNAB	127	10	-10	0	0	0	0
UNAP	29	10	-0.2	-3.4	-5.7	-0.3	-0.3
UPLA	43	10	-4.7	-0.9	-4.4	-0.5	0
USACH	64	10	5.2	-8	-2	-1.8	-3.9
UTA	39	10	0	-4.2	-5.6	0.1	-0.1
UTAL	23	10	-6.5	-0.9	-0.7	-0.4	-1.5
UTEM	27	10	-9.1	-0.7	0	-0.2	0.2
UTFSM	57	10	0	-10	0	0	0
UV	54	10	-10	0	0.1	0	-0.3
Total	1296	10	-6.1	-1.7	-1.7	-0.1	-0.3

[1] Average of absolute difference in programs' admission weights. Variations do not need to add up to 0 along each row, because some programs change from requiring exclusively History or Science to requiring either of them between the two years.

Table 5.2 shows the mean variation of each admission weight between 2013 and 2014 by university. We observe an important difference in the mean variation of the weight assigned to the rank score, ranging from 0 to 30%. On average, there was an increase of 12% in the rank weight and a decrease of 7.2% in the GPA weight.

Table 5.2: Variation on Admission weights 2013-2014

Universities	Programs with changes in Rank	Variation on admission weights[1]					
		Rank	GPA	Math	Verbal	History	Science
PUC	46	7.5	-2.8	-2.4	-0.8	-0.6	-1.1
PUCV	29	4.1	-3.7	-0.3	-0.1	0	0
UACH	50	10	-7.8	-0.4	-0.7	0	-1.1
UAH	21	6.2	-3.3	0.8	-1.9	-1.7	-0.8
UAI	0	0	0	0	0	1.8	1.8
UANDES	8	2	0.2	-0.2	-1.8	0	-0.2
UANT	35	10	-10	0	0	0	0
UBB	40	30	-17.5	-6.1	-4.5	-1.1	-1.5
UCH	50	12.2	-0.5	-3.7	-4.8	-0.5	-2.4
UCM	26	9.4	-0.4	-4.2	-3.1	-0.8	-1.3
UCN	41	5	0	-0.9	-4.1	0	0
UCSC	31	20.3	-12.1	-3.1	-1.8	-2.1	-3.4
UCT	48	30	-12.6	-8.2	-8.8	-0.1	-0.2
UDA	24	8.2	-8.4	0.4	0.2	-0.4	-0.4
UDD	0	0	0	-0.7	0.4	0.3	0.3
UDEC	91	15	-0.7	-7.6	-0.9	-1.4	-4.7
UDP	2	0.4	0	-0.4	0	-1.8	-1.6
UFRO	39	10	0	-3.2	-4.6	-1.5	-1.2
UFT	7	1.6	0	-0.9	-0.5	-0.2	-2.3
ULA	25	30	-20	-4.8	-4.8	-1	-0.2
ULS	28	6.2	-2.6	-1.1	-1.6	-0.4	-0.9
UMAG	22	10	-5.7	-2.3	-1.1	-0.7	-0.2
UMAYOR	26	6.2	-3.2	-2.2	-1.1	0.3	0.3
UMCE	11	7.3	-6.1	-0.9	-0.5	2.5	0
UNAB	38	5.7	-4.6	-1.1	-0.2	-1.8	-0.4
UNAP	31	30	-27.1	-0.3	-0.6	1.5	0.3
UPLA	36	6.8	-6.2	-2.9	1	1.8	1.3
USACH	64	30	-24.4	-0.4	-1.2	0	-3.8
UTA	40	30	-30	1.1	0.4	-1.5	-1.5
UTAL	28	15	-8	-1.6	-5.5	-0.7	0.9
UTEM	0	0	0.6	-0.4	-0.8	0.6	0
UTFSM	59	10	-10	0	0	0	0
UV	51	10.1	-7.1	-1.1	-1.2	-1.1	-2.2
Total	1047	12	-7.2	-2	-1.6	-0.5	-1

[1] Average of absolute difference in programs' admission weights. Variations do not need to add up to 0 along each row, because some programs change from requiring exclusively History or Science to requiring either of them between the two years.

We assume that the variation in admission weights does not affect the overall distribution of indirect utilities between those years (conditional on observable characteristics). In this way we rule out the possibility that students could have forecast this change and decided to postpone their admission decisions, or the possibility that students have preferences that depend on the admission weights (for instance, a program choosing a high rank score could be associated with a high equity concern, which could be valued by students). The variation on the admission weights changes the weighted scores and shifts the admission probabilities faced by similar students, allowing us to identify their preferences by looking at the variation on the submitted ROLs between those years, or more precisely, by looking at the variation on the implied lotteries over assignments.

Formally, following Agarwal and Somaini (2018), let $C(R|t, Z, d, \xi)$ be the set of indirect utility vectors u that rationalizes ROL R to be the optimal ROL in the choice environment t, conditional on the vector of observable characteristics Z, the vector of distances d and vector of unobserved characteristics ξ. For the remaining analysis we fix Z, d and ξ, and to simplify notation drop them from the conditional set. We can write the likelihood of observing a ROL R as

$$\mathbb{P}(R|t) = \mathbb{P}\left(R = \underset{R' \in \mathcal{R}}{\operatorname{argmax}}\, U(R')\big|t; f_u\right) = \int \mathbb{1}\{u \in C(R|t)\} f_u(u) du. \tag{5.1}$$

The variation in the choice environment allows us to identify the distribution $f_u(u)$ by looking at the difference in likelihood of reporting R:

$$\mathbb{P}(R|t+1) - \mathbb{P}(R|t) = \int (\mathbb{1}\{u \in C(R|t+1)\} - \mathbb{1}\{u \in C(R|t)\}) f_u(u) du. \tag{5.2}$$

Changes in admission weights, will change the admission probabilities of students with similar observable characteristics, changing the set of indirect utilities for which reporting the ROL R is optimal. Equation 5.2 shows that if we have enough variation we will trace out the conditional distribution of indirect utilities and point identify the parameters. Notice, though, that both Equations 5.1 and 5.2 implicitly assume that given a vector of indirect utilities u, there is a unique optimal ROL R. However, uniqueness does not hold if students face degenerate admission probabilities for some programs. In the next section we describe how to adapt the identification argument to that case.

5.1 Degenerate admission probabilities

There are at least two reasons why students may have degenerate beliefs on their admission probabilities. First, several programs impose admissibility requirements, such as minimum weighted scores, minimum average between math and verbal scores, among many others. These requirements must be satisfied by the student to be eligible, so students that do not satisfy them have probability zero of being assigned to those programs. Nevertheless, DEMRE still allows students to apply to programs for which they do not meet all the requirements and notifies them during the application process that they are not eligible for those programs. Also, as discussed in Section 2.2, the amount of uncertainty that some students face is moderate, so students with very low or high scores can anticipate very accurately that their admission chances are 0 and 1 respectively for some programs.

The degeneracy of beliefs on admission probabilities raises a potential identification concern, as there could be multiple equilibria due to *multipliciy of best responses*. For instance, if a student reports optimally a short-list ROL, it would be payoff equivalent to report the same ROL but adding a program for which she faces an admission probability of zero. Also, if a student faces an admission probability of one to some program listed in her ROL, it would be payoff equivalent to add no programs below it on her list. This issue is analyzed in detail in He (2012) for the school choice system in Beijing. As the author explains: "*Multiplicity in best responses implies multiple equilibria and thus creates challenges for empirical analysis because choice probabilities of actions can no longer be characterized.*". In the presence of degenerate beliefs, the model is incapable of predicting a unique distribution of ROLs conditional on preferences, and it is said to be incomplete à la Tamer (2003).

As stated before, Equations 5.1 and 5.2 do not take into account the issues of the multiplicity of best response raised by He (2012). To account for this, let $L(R) \in \Delta^J$ be the lottery over assignment implied by reporting ROL R and let $\omega(R|u,t) \in [0,1]$ be the mixing probability of reporting an optimal ROL R given the indirect utility vector u and the choice environment t. In addition, define the set $\mathcal{R}_L(R)$ as the set of ROLs that imply the same lottery over assignment than reporting ROL R, that is, $\mathcal{R}_L(R) \equiv \{R' \text{ st } L(R') = L(R)\}$. The mixing probabilities must satisfy then the following equation

$$\sum_{R' \in \mathcal{R}_L(R)} \omega(R'|u,t) = 1 \quad \forall u, R, t. \tag{5.3}$$

We can now rewrite Equations 5.1 and 5.2 allowing for multiplicity of best response:

$$\mathbb{P}(R|t) = \int \mathbb{1}\{u \in C(R|t)\}\omega(R|u,t) f_u(u) du, \tag{5.4}$$

and

$$\mathbb{P}(R|t+1) - \mathbb{P}(R|t) = \int (\mathbb{1}\{u \in C(R|t+1)\omega(R|u,t+1)\} - \mathbb{1}\{u \in C(R|t)\omega(R|u,t)\}) f_u(u) du. \tag{5.5}$$

Equation 5.5 illustrates the identification problem: when there is multiplicity of best responses and students are mixing over payoff equivalent ROLs, the variation on the shares of submitted ROLs over time could be driven by variations in the choice sets $C(R)$, but also by changes in the mixing probabilities over time. Therefore, without imposing assumptions on these mixing probabilities, the identification argument cannot rely only on the variation on submitted ROLs.

Even though there is multiplicity of best response in our setting, it is possible to restore completeness in the model if we define the likelihood of students choosing a lottery over assignments instead of choosing a ROL,

$$\mathbb{P}(L|t) = \int \mathbb{1}\{u \in C(L|t)\}\omega(L|u,t) f_u(u) du. \tag{5.6}$$

where $C(L|t)$ is the set of indirect utilities that rationalize the optimality of lottery L in the choice environment t and $\omega(L|u,t)$ is the mixing probability of choosing lottery L given the indirect utility vector u in choice the environment t.

If we restrict attention to distributions of preferences that have no atomic masses, the probability that two different lotteries are payoff equivalent is zero, which implies that $\omega(L'|u,t) = 1 \quad \forall L', u, t$. In other words, all ROLs that are payoff equivalent must imply the same lottery over assignments. Therefore, we can identify the distribution of indirect utilities using variation in the set of lotteries implied by observed ROLs.

Notice that in the presence of zero admission probabilities, we will not be able to non-parametrically identify preferences, as Agarwal and Somaini (2018) show in their setting. For instance, if preferences depend on students' scores (without any parametric restriction), we cannot disentangle whether low score students do not rank selective programs because their beliefs are zero or because they prefer other programs. In this sense, we need to place some parametric restrictions on preferences. However, as the variation we are exploiting in the data is on the admission weights over time and not on the scores themselves, we will get variation over time in the set of lotteries for students with the same set of scores, but this variation will not be over all the support of scores for each program.

6 Simulations

In order to test our hypothesis and show whether assuming truth-telling can lead to biased results, we perform a series of Monte Carlo simulations changing the data generating process (DGP) and the assumptions we use for estimation. For this analysis we assume a simplified version of students' preferences and sub-sample the data to reduce computational problems.

Consider the following specification for students' preferences:

$$u_{ij} = Z_{ij}\beta + \varepsilon_{ij}, \tag{6.1}$$

where $Z_{ij} = [z_{ij1}, ..., z_{ij5}]$ is a 1×5 row vector. The respective covariates are 1, yearly tuition of program j, last year cutoff for program j, weighted score of student i in program j and distance between student i's municipality and programs j's municipality. The distance metric is constructed by taking the geographic distance between the centroids of those municipalities. We choose this specification because it is simple and includes a covariate that exhibits variation at the student and also at the program level[16].

To further simplify the analysis, we sub-sample the data and look at one major between years 2013 and 2014. As the skipping pattern was clearly observed for students who reported Medicine in their first preference, we select all the Medicine programs present in both years, having a total of 22 programs. In order to be consistent with the real application patterns, we select all the students who applied to Medicine in some preference in those years (close to 14,000 students). Table 6.1 shows descriptive statistics for the 22 programs selected in the sample.

[16]In section 8 we consider a richer preference specification and estimate parameters with the reported ROLs.

Table 6.1: Aggregate Statistics Medicine Programs 2013-2014

		2013			2014		
Program ID	City	Tuition*	Cutoff	%Rank	Tuition*	Cutoff	%Rank
1	ANTOFAGASTA	4,377,000	714.5	10	4,478,000	716.7	20
2	CONCEPCION	4,590,000	734.6	10	4,990,000	746.25	30
3	CONCEPCION	4,506,000	745.4	10	4,642,000	758.95	25
4	COQUIMBO	4,001,000	719.2	10	4,148,000	715.7	15
5	SANTIAGO	3,738,640	760.1	10	4,019,000	773.3	40
6	SANTIAGO	5,922,752	729.2	10	6,171,570	725.3	10
7	SANTIAGO	5,728,253	755.55	10	5,809,163	755.15	15
8	SANTIAGO	6,219,000	743.2	10	6,570,000	735.6	10
9	SANTIAGO	5,940,000	702.45	10	6,144,000	701.1	10
10	SAN FELIPE	4,130,000	735.6	10	4,304,000	740.2	20
11	SANTIAGO	4,605,500	774.45	10	4,835,700	783.15	30
12	SANTIAGO	5,385,000	787.85	10	5,487,000	790.45	20
13	SANTIAGO	5,823,000	716.7	10	6,091,000	712.7	10
14	SANTIAGO	6,200,939	703.5	10	6,448,977	736.75	40
15	TALCA	4,578,900	720.2	10	4,647,600	738.3	25
16	TALCA	4,510,000	714.45	10	4,720,000	723.1	20
17	TEMUCO	5,922,752	694.6	10	6,171,570	720.9	30
18	TEMUCO	3,881,000	732.8	10	4,076,000	742.95	20
19	VALDIVIA	3,998,000	737.85	10	4,090,000	745.65	20
20	VALDIVIA	3,998,000	728.55	10	4,090,000	738.6	20
21	VALPARAISO	4,130,000	752.1	10	4,304,000	754.6	20
22	VIÑA DEL MAR	5,146,780	710.7	10	5,352,651	745.45	40

* Yearly tuition in Chilean pesos (nominal).

Notice that there is an important variation between 2013 and 2014 in the admission weight allocated to the rank score, which will help us to better identify the model. Also, most of the cutoffs are above 700 PSU points, which is almost two standard deviations above the median score in the pool of participants in the system.

Table 6.2 shows aggregate statistics for the sample of students that apply to at least one Medicine program in 2013 or 2014.

Table 6.2: Aggregate Statistics Students Applying to Medicine 2013-2014

		2013	2014
Applicants	Total[1]	7,020	7,313
	With 0 Probability[2]	3,338	3,575
	Final Sample[3]	3,682	3,738
Application Scores	Mean	673.31	680.1
	Median	685.6	693.12
	Standard deviation	73.79	72.7

[1] Students who applied to at least one Medicine program.
[2] Students who applied to at least one Medicine program and have admission probabilities of zero for all Medicine programs.
[3] Students who applied to at least one Medicine program and face positive admission probabilities for at least one Medicine program.

We observe that roughly half of the students face admission probabilities equal to 0 for every program in the sample, even though they listed one of these programs in their actual application list. This suggests that Assumption 1 and 2 do not hold for every student, either because some of them report truthfully and/or because their beliefs over admission probabilities are not given by rational expectations.

6.1 Assuming truth-telling

To test whether assuming truth-telling can result in biased estimates (if students report strategically and A1 and A2 hold), we simulate data using the previous specification under three different DGPs:

- DGP1: students do not take into account their admission probabilities and report no more than K of their most preferred programs.
- DGP2: students report strategically strictly less than K programs (short list students), maximizing their expected utility of their portfolios, and A1 and A2 hold.
- DGP3: students report strategically no more than K programs, maximizing their expected utility of their portfolios, and A1 and A2 hold.

For DGP1 and DGP2 we consider the shocks to follow a Type I Extreme Value distribution with location parameter 0 and scale parameter 1. We label the outside option as program 23, and normalize its systematic utility to 0. We force students to include the outside option in their portfolios and we assign an admission probability equal to 1. Therefore, under DGP2 students won't include programs that are bellow the outside option, that is, if their value is less than or equal to ε_{i0}.

Data under DGP1 can be simply generated by ordering the indirect utilities for each student and reporting at most K of the programs for which the payoff is not bellow the outside option. Data under DGP2 can be generated using MIA + A1, making sure that K is not binding. Finally, for DGP3 we normalize the indirect utility of the outside option to 0. For the unobserved shock we choose a multivariate Normal distribution[17] with mean 0 and variance covariance matrix $\sigma^2\mathbb{I}$ and normalize $\sigma = 1$[18]. As with DGP2, data can be generated using MIA + A1.

6.2 Likelihoods

To show that we can recover the parameters assuming truth-telling if we assume DGP1 is the truth, we construct the likelihood of observing a ROL R under truth-telling. As we have chosen the error terms to be Type I Extreme Value, the likelihood has the form of a ranked-ordered (or exploded) logit. Given the Independence of Irrelevant Alternatives assumption, the likelihood can be seen as sequentially choosing the best available program in the choice set, until either no program is above the outside option or we have reached the capacity constraint of K. For simplicity let $u_{ij} \equiv \bar{u}_{ij} + \varepsilon_{ij}$, then

$$\mathbb{P}\left(R_i|DGP1\right) = \frac{\exp(\bar{u}_{iR_i(1)})}{\sum\limits_{j\in J}\exp(\bar{u}_{ij})}\times\frac{\exp(\bar{u}_{iR_i(2)})}{\sum\limits_{j\in J\setminus\{R_i(1)\}}\exp(\bar{u}_{ij})}\times\ldots\times\frac{\exp(\bar{u}_{iR_i(|R|)})}{\sum\limits_{j\in J\setminus\{R_i(1),\ldots,R_i(|R|-1)\}}\exp(\bar{u}_{ij})}\times V_i\left(R_i(|R|)\right) \tag{6.2}$$

where

$$V_i\left(R_i(|R|)\right) \equiv \begin{cases} 1 & \text{if} \quad |R_i| = K \\ \frac{1}{\sum\limits_{j\in J\setminus\{R_i(1),\ldots,R_i(|R|)\}}\exp(\bar{u}_{ij})} & \text{o.w} \end{cases} \tag{6.3}$$

Notice that we can only infer, under DGP1, that programs not listed in the ROL are less preferred than the outside option, if the length of the ROL is strictly less than K.

Under DGP2, students take into account their admission probabilities and solve their optimal portfolio problem. When A1 holds, students will only include programs for which their marginal benefit is strictly greater than 0. This implies that (i) students will not include programs for which their assignment probabilities are 0 and (ii) students will not include programs below a listed program for which their admission probability is 1. Moreover, as we generate ROLs for which the restriction in the length of the list is not binding, the portfolio problem reduces to simply include programs with the highest indirect utilities, conditional on having a positive probability of admission and being more preferred than the outside option. The implies that the likelihood under DGP2 can be written as:

[17] We change the distribution of the shock for this DGP to be consistent with our proposed estimation method.

[18] Notice that we are restricting the variance covariance matrix, thus the model under this specification is over identified.

$$\mathbb{P}\left(R_i|DGP2\right) = \frac{\exp(\bar{u}_{iR_i(1)})}{\sum\limits_{j\in\tilde{J}_i}\exp(\bar{u}_{ij})} \times \frac{\exp(\bar{u}_{iR_i(2)})}{\sum\limits_{j\in\tilde{J}_i\setminus\{R_i(1)\}}\exp(\bar{u}_{ij})} \times \ldots \times \frac{\exp(\bar{u}_{iR_i(|R|)})}{\sum\limits_{j\in\tilde{J}_i\setminus\{R_i(1),\ldots,R_i(|R|-1)\}}\exp(\bar{u}_{ij})} \times V_i\left(R_i(|R|)\right) \tag{6.4}$$

where

$$V_i\left(R_i(|R|)\right) \equiv \begin{cases} 1 & \text{if} \quad p_{iR_i(|R|)} = 1 \\ \frac{1}{\sum\limits_{j\in\tilde{J}_i\setminus\{R_i(1),\ldots,R_i(|R|)\}}\exp(\bar{u}_{ij})} & \text{o.w} \end{cases} \tag{6.5}$$

and $\tilde{J}_i \equiv \{j \in J : p_{ij} > 0\}$. Notice that students will only consider programs for which their believed admission probabilities are strictly positive. Hence we cannot identify how much they like programs for which their admission probabilities are 0.

6.3 Gibbs' Sampler

It is not possible to write down a likelihood in closed form for DGP3. This is not only because we have chosen a different distribution for the unobserved shocks, but more importantly, because under DGP3 we allow students to report constrained ROLs. When the ROL is constrained, students take into account their admission probabilities not only to include programs for which the marginal benefit is bigger than 0, but also to decide which programs to include in the portfolio if they reach the capacity constraint. Under A2, the solution to this problem is given by MIA. Thus, we exploit its structure to characterize the optimal solution and adapt the estimation procedure proposed by Agarwal and Somaini (2018). The challenge is to obtain unbiased estimates without running into the curse of dimensionality.

In the Gibbs' Sampler approach to estimate discrete choice problems (McCulloch and Rossi (1994)), we obtain draws of the parameters from the posterior distribution by constructing a Markov chain of draws starting from an initial set of the parameters. The posterior mean of this sampler is equivalent to the MLE estimator of the parameters β, σ^2 given the priors and the data.

To initialize the sampler we pick, for each student i, a utility vector u_i^0 that is consistent with the observed ROL R_i to be the optimal choice. We then construct the Markov Chain sampling from the conditional posteriors of the parameters and the utility vectors, conditional on the previous draws. In order to pick the initial vector of utilities (Step 0) and to draw from the conditional posterior of the vector of utilities (Step 2), we must be able to characterize the set of indirect utilities $C(R_i)$ that is consistent with R_i being optimal.

Agarwal and Somaini (2018) do this by constructing a matrix that encodes all possible pairwise comparisons between the chosen ROL R_i and any other ROL $R' \in \mathcal{R}$ that student i could have submitted instead. In their application, students cannot rank more than 3 schools out of 13 available in the system, which gives them 1,885 pairwise comparisons. However, in the Chilean College Admissions' problem students can rank up to 10 programs out of more than 1,400 alternatives, which gives more than 10^{24} possible ROLs, making impossible to encode a matrix with all comparisons.

We exploit the fact that, under A1 and A2, the optimal solution to the portfolio problem is given by MIA and Proposition 1 holds. This additional structure enables us to characterize the set $C(R_i)$ without running into the curse of dimensionality. Basically, we construct a low-dimensional matrix $A_i(R_i)$ with a sufficient set of inequalities that the indirect utilities need to satisfy in order for R_i to be the optimal ROL for student i, without the need of comparing every possible ROL:

$$u_i \in C(R_i) \iff A_i u_i \geq 0 \tag{6.6}$$

Case 1: $|R_i| < K$

As we have shown before, if $|R_i| < K$ the optimal portfolio must include the programs with the $|R_i|$-th highest utilities among the ones for which student i has positive admission probabilities, i.e.:

$$u_{ij} \geq u_{ij'} \quad \forall j \in R_i, j' \notin R_i, st : p_{ij'} > 0. \tag{6.7}$$

Also, as the outside option has been normalized to 0 and it is interpreted to be the option value of being unassigned, it must be that

$$u_{ij} \geq u_{i0} \quad \forall j \in R_i. \tag{6.8}$$

Finally, the marginal benefit of including any other program in the portfolio must be less than or equal to 0. As we have not reached the capacity constraint, this can happen either because none of the programs with positive probability exceeds the outside option, or because student i has admission probability of 1 for the last listed program:

$$\text{if } p_{iR_i(|R_i|)} < 1 \Rightarrow u_{i0} \geq u_{ij'} \quad \forall j' \notin R_i, st : p_{ij'} > 0 \tag{6.9}$$

After the student has chosen the programs to include in his portfolio, it is always optimal to order them in decreasing order of utilities (Haeringer and Klijn (2009)), thus we further know that:

$$u_{iR_i(1)} \geq u_{iR_i(2)} \geq ... \geq u_{iR_i(|R_i|)} \tag{6.10}$$

With all these inequalities we can write down the matrix A_i for any student such that $|R_i| < K$. Inequalities given in Equation 6.7 can be represented with a $1 \times M$ row vector with a 1 in the j-th component, a -1 in the j'-th component and zeros otherwise. Similarly, inequalities defined in Equations 6.8 and 6.9 can be represented with $1 \times M$ row vectors, with a 1 in the j-th component and zeros otherwise and with a -1 in the j-th component and zeros otherwise, respectively. Finally, inequalities defined in Equation 6.10 can be represented with $|R_i| - 1$ row vectors of dimension $1 \times M$, with a 1 in the $R_i(k)$-th component, a -1 in the $R_i(k+1)$-th component and zeros otherwise, for $k = 1, ..|R_i| - 1$.

Case 2: $|R_i| = K$

In this case, the sets of inequalities given by Equations 6.7 and 6.9 will not necessarily hold. However, Proposition 1 establishes that it suffices to characterize $C(R_i)$ using only inequalities given by comparing R_i to *One-Shot Swaps* of R_i. We can express these inequalities as:

$$U_i(R_i) \geq U_i(R') \Leftrightarrow u_i \Pi_{iR_i} \geq u_i \Pi_{iR'} \Leftrightarrow (\Pi_{iR_i} - \Pi_{iR'})\, u_i \geq 0,$$

where Π_{iR} is an $(1 \times M)$ vector with the j-th component being the admission probability of student i to program j conditional on reporting ROL R. We construct the matrix A_i by stacking the row vectors $(\Pi_{iR_i} - \Pi_{iR'})$ and the vectors encoding Equations 6.8 and 6.10.

The key difference with Agarwal and Somaini (2018)'s Gibbs Sampler algorithm is that Proposition 1 allows us to avoid the curse of dimensionality that their implementation would entail in our setting.

Gibbs sampler:

Consider the following specification for students' preferences:

$$u_{ij} = Z_{ij}\beta + \varepsilon_{ij} \tag{6.11}$$

where $\varepsilon_{ij} \sim N\left(0, \sigma^2\right)$, and $Z_{ij} = [z_{ij1}, ..., z_{ijK}]$ is a $1 \times K$ row vector of covariates. The system can be stacked in order to represent the vector of utilities u_i as:

$$u_i = Z_i\beta + \varepsilon_i \tag{6.12}$$

where Z_i is an $M \times K$ matrix of covariates and ε_i is an $M \times 1$ vector of shocks. Consider also the following prior for β :

$$\beta \sim N(\bar{\beta}, A^{-1}) \tag{6.13}$$

Step 0 Start with initial values for $u^0 = \{u_i^0\}_{i=0}^N$ such that $u_i^0 \in C(R_i) \quad \forall i = 1, ..N$, i.e, select u_i^0 to be a solution to the following problem:

$$A_i u_i \geq \epsilon \tag{6.14}$$

with ϵ a small positive number.

Step 1 Draw $\beta^1 | u^0$ from a $N(\tilde{\beta}, V)$, where

$$V = \left(Z^{*\prime} Z^* + A\right)^{-1}, \quad \tilde{\beta} = V\left(Z^{*\prime} u^* + A\bar{\beta}\right) \tag{6.15}$$

$$Z^* = \begin{bmatrix} Z_1^* \\ ... \\ Z_N^* \end{bmatrix} \tag{6.16}$$

$$Z_i^{*'} = C' Z_i, \quad u_i^* = C' u_i^0 \tag{6.17}$$

$$\left(\sigma^2 \mathbb{I}\right)^{-1} = C'C \tag{6.18}$$

Where C comes from the Cholesky decomposition of the inverse of the variance covariance matrix of ε_i.

Step 2 Iterate over students and schools, drawing $u_i^1|\beta^1, \sigma^2, R_i$. For each school $j = 1, ..M$, draw:

$$u_{ij}^1 | \{u_{ik}^1\}_{k=1}^{j-1}, \{u_{ik}^0\}_{k=j+1}^{J}, \beta^1, \sigma^2 \tag{6.19}$$

from a truncated normal $TN(\mu_{ij}, \sigma_{ij}^2, a_{ij}, b_{ij})$, where

$$\mu_{ij} = \sum_{k=1}^{K} \beta_{jk}^1 z_{ijk} \tag{6.20}$$

$$\sigma_{ij}^2 = \sigma^2 \tag{6.21}$$

The truncation points a_{ij} and b_{ij} must ensure the draw u_{ij}^1 lies in the interior of $C(R_i)$ given the previous draws, so they are the solutions to the following optimization problems:

$$\begin{aligned} a_{ij} = \max_{u_{ij}} \quad & u_{ij} \\ st. \quad & Au \geq 0 \\ & u_{ik} = u_{ik}^1 \quad \forall k = 1, ..j-1 \\ & u_{ik} = u_{ik}^0 \quad \forall k = j+1, ..M \end{aligned}$$

$$\begin{aligned} b_{ij} = \min_{u_{ij}} \quad & u_{ij} \\ st. \quad & Au \geq 0 \\ & u_{ik} = u_{ik}^1 \quad \forall k = 1, ..j-1 \\ & u_{ik} = u_{ik}^0 \quad \forall k = j+1, ..M \end{aligned}$$

Given the structure of the optimization problems, we can obtain analytical expressions for a_{ij} and b_{ij}. The goal is to compute bounds that must be satisfied by u_{ij} conditional on having the vector

$$u_i^{-j} = \left(u_{i1}^t, \ldots, u_{i,j-1}^t, u_{i,j+1}^{t-1}, \ldots, u_{i,|J|}^{t-1}\right)$$

For simplicity, we omit index i, as this problem must be solved for each student independently. Notice that the constraint $Au \geq 0$ can be equivalently written as $A^{-j}u^{-j} \geq -A_j u_j$, where A^{-j} is matrix A without column j, and A_j is matrix A's column j. As the term $A^{-j}u^{-j}$ is fixed and known, we can manipulate this expression to isolate u_j, which allows us to obtain inequalities that the vector of indirect utilities must satisfy to rationalize the observed ROL. After some manipulation we get that

$$a_j = \max_{k \in \{k: A_{kj} > 0\}} \frac{-A_k^{-j} u^{-j}}{A_{kj}}$$
$$b_j = \min_{k \in \{k: A_{kj} < 0\}} \frac{-A_k^{-j} u^{-j}}{A_{kj}}$$

where A_k^{-j} is matrix A^{-j}'s k-th row and A_{kj} is the k-th element of column A_j.

Step 3 Set $u^0 = u^1$ and repeat steps 1-2 to obtain a sequence β^k.

Notice that we do not need to solve the optimization problems for the bounds a_{ij} and b_{ij} for every student i. We only need to do so for students who submit a constrained ROL because the bounds for the unconstrained ROLs can be inferred from the set of inequalities, after conditioning on the previous draws. Moreover, even for constrained ROLs we do not need to solve this problem for both bounds for every program j. For example, it is clear that the bounds for all programs j such that $p_{ij} = 0$ are $(-\infty, +\infty)$ and that the upper bound for the first listed program is $+\infty$, regardless of the realizations of the previous draws.

We describe in Appendix B a multivariate version of the Gibbs' Sampler, where we normalize the coefficient of distance to -1 and allow for an unrestricted variance covariance matrix of the random shock, using an Inverse Wishart prior.

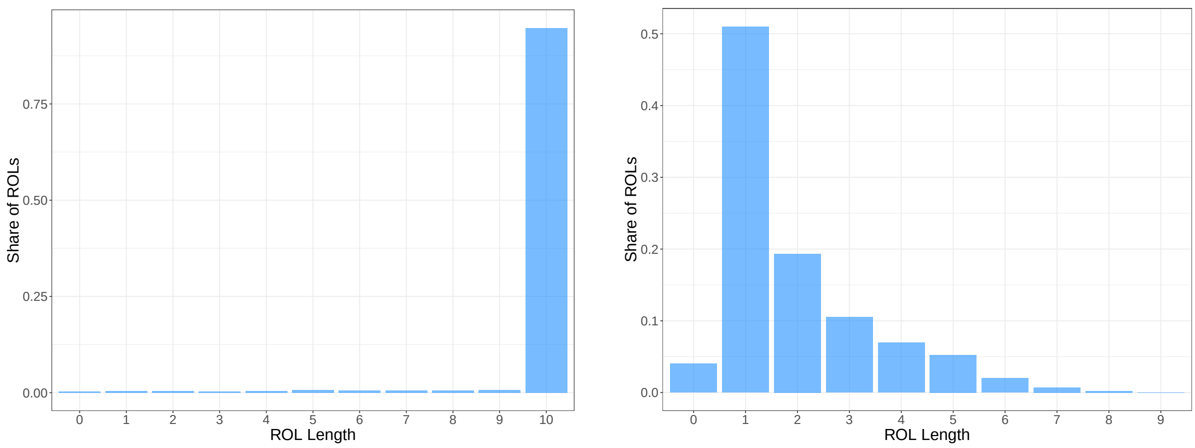
7 Results

To compare the results across simulations and DGPs, we assume that the true underlying parameters are $\beta = (1, -1, 2, 1.5, -1)$. In addition, for most simulations we use $K = 10$, i.e., students can apply to at most 10 programs. For some simulations we choose a smaller K to make the constraint bind for a larger fraction of students.

7.1 DGP1 vs DGP2

We first show some descriptive statistics on the simulated data under both DGPs using one simulation. Figure 7.1 shows the distributions of the length of ROLs under both DGPs:

Figure 7.1: Distribution of Length of ROLs



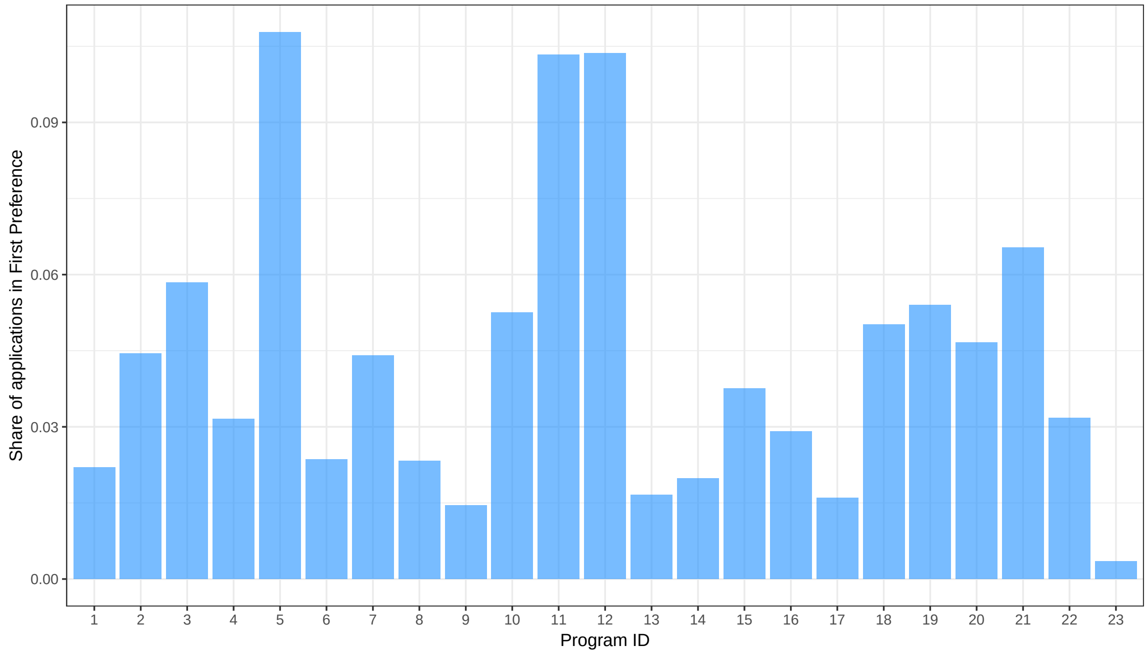
Notes: Distribution of the Length of ROLs under DGP1 (left figure) and DGP2 (right figure).

We observe that under DGP1 most students submit ROLs with 10 programs, and thus the capacity constraint is binding. However, under DGP2 there is an important share of students submitting ROLs of different lengths, with no student applying to more than 9 programs in their ROL. This difference can be explained because students under DGP2 do not consider programs with admission probabilities of 0, which decreases their choice sets compared to DGP1. In addition, some students face programs with admission probabilities exactly equal to 1, which implies that, under DGP2, if they include one of such programs in their ROL they won't include any program with a lower utility.

To see how choices differ in both DGPs, we analyze programs listed in first preference (we label the outside option as program 23). Figure 7.3 shows that under DGP1, programs 5, 11 and 12 are the most preferred programs. However, Figure 7.4 shows that if we were to interpret DGP2 as truth-telling, programs 1, 17, and 9 would be the most preferred ones. This difference is explained because the most preferred programs under DGP1 (the truth) happen to be the most selective ones, that is, programs for which most students face an admission probability of 0, which deters students under DGP2 to include them in their ROLs. In addition, if we were to interpret DGP2 as truth-telling, we would infer that more than 4% of students prefer their outside option to any Medicine program, although less than 1% prefer it.

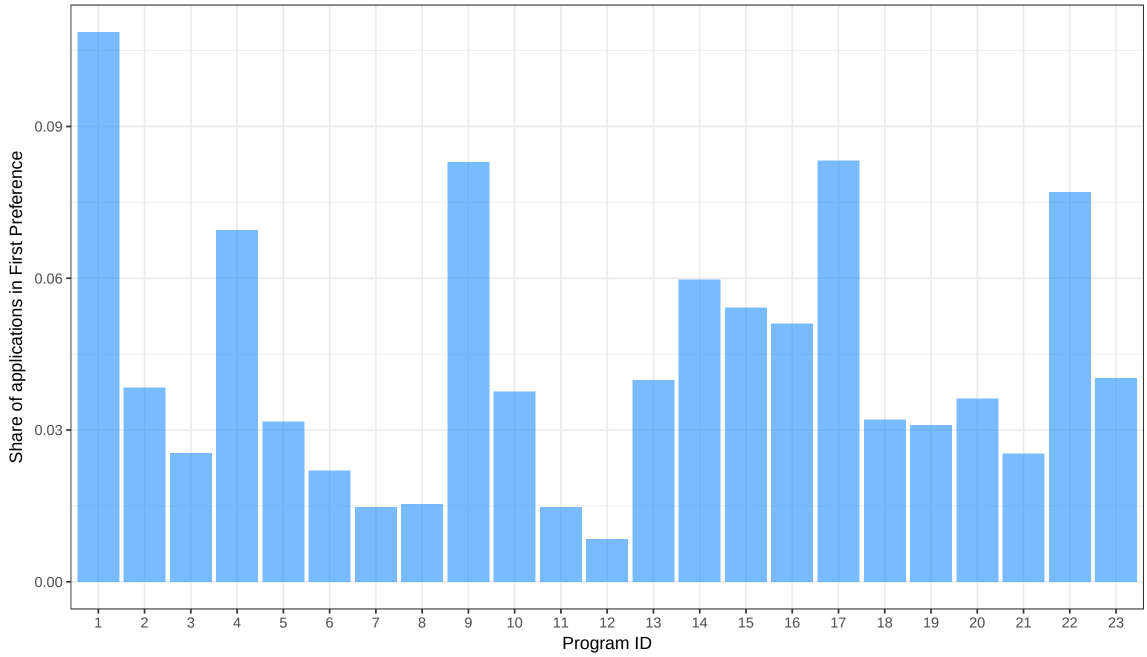
We also observe a fair amount of heterogeneity in first choices, even under DGP1. This is mainly explained by the preference shock, but also because of the heterogeneity induced by the distance covariate.

Figure 7.3: First listed preference under DGP1



Notes: Share of applications in first preference to each program under DGP1. Program 23 is the outside option of being unassigned.

Figure 7.4: First listed preference under DGP2



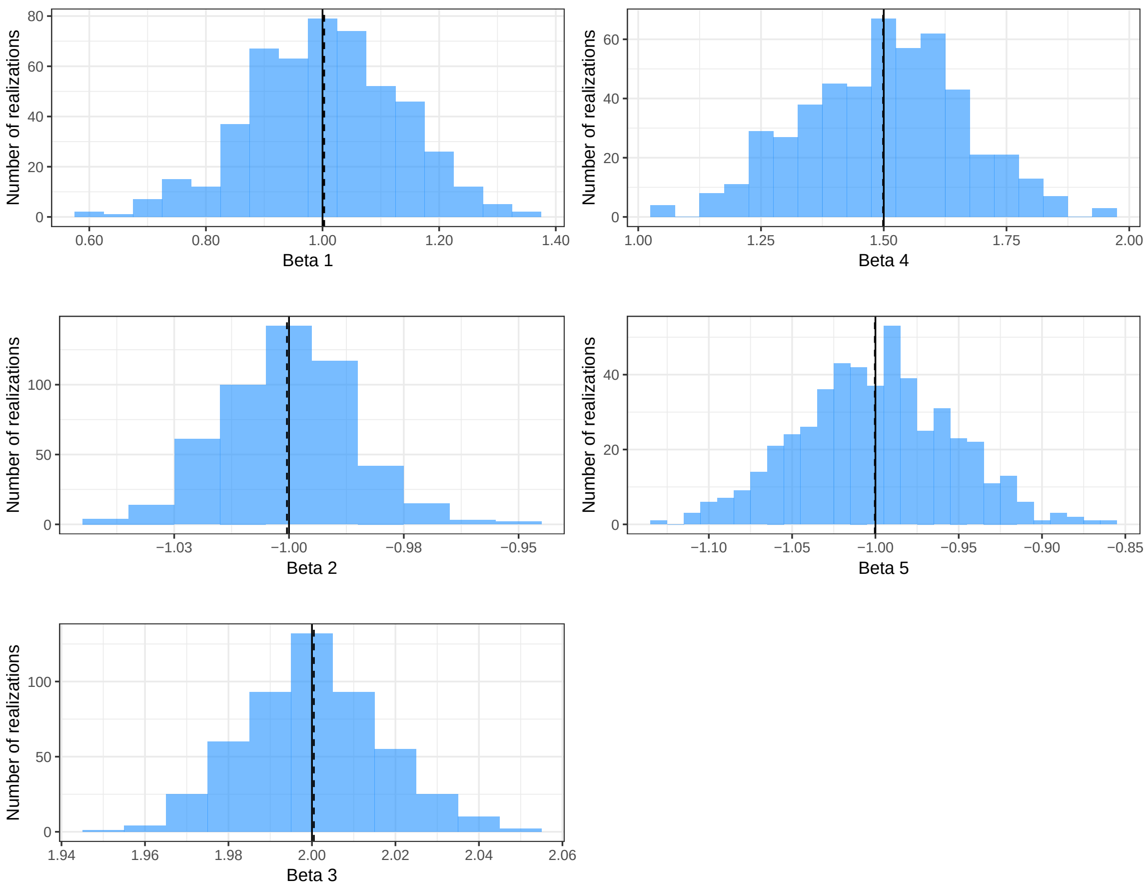
Notes: Share of applications in first preference to each program under DGP2. Program 23 is the outside option of being unassigned.

7.2 Monte Carlo Simulations

7.2.1 DGP1 vs DGP2

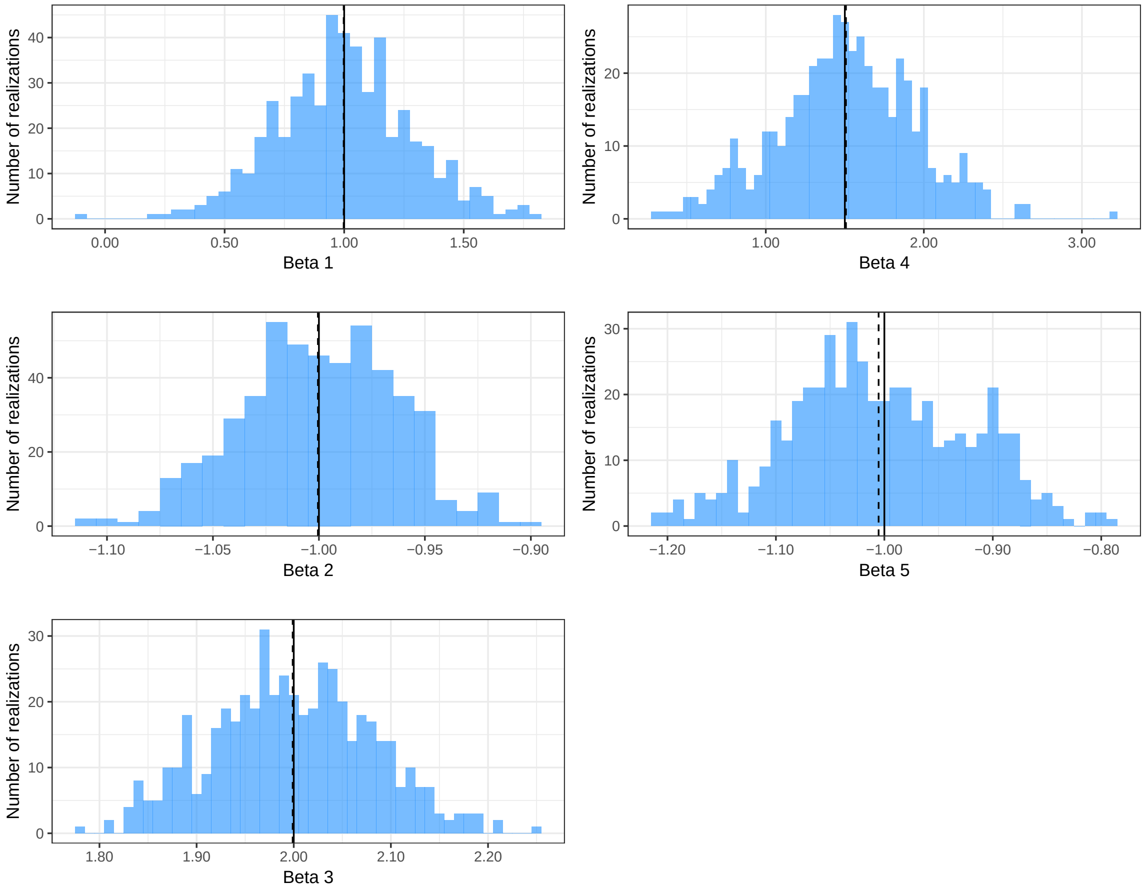
We run 500 Monte Carlo simulations assuming DGP1 and DGP2. For each simulation under DGP1 we estimate the model by Maximum Likelihood assuming DGP1 is the truth. For each simulation under DGP2 we estimate the model assuming DGP2 is the truth. Our goal is to confirm that the likelihoods are correctly specified and that we recover the parameters if the model and the DGPs are consistent. Figures 7.5 and 7.6 show the distribution of the estimators under both DGPs:

Figure 7.5: Monte Carlo under DGP1



Notes: Monte Carlo simulations under DGP1, estimating parameters with Likelihood approach assuming DGP1 is the truth. Solid lines are true parameters, dashed lines are the the corresponding means of the distributions.

Figure 7.6: Monte Carlo under DGP2

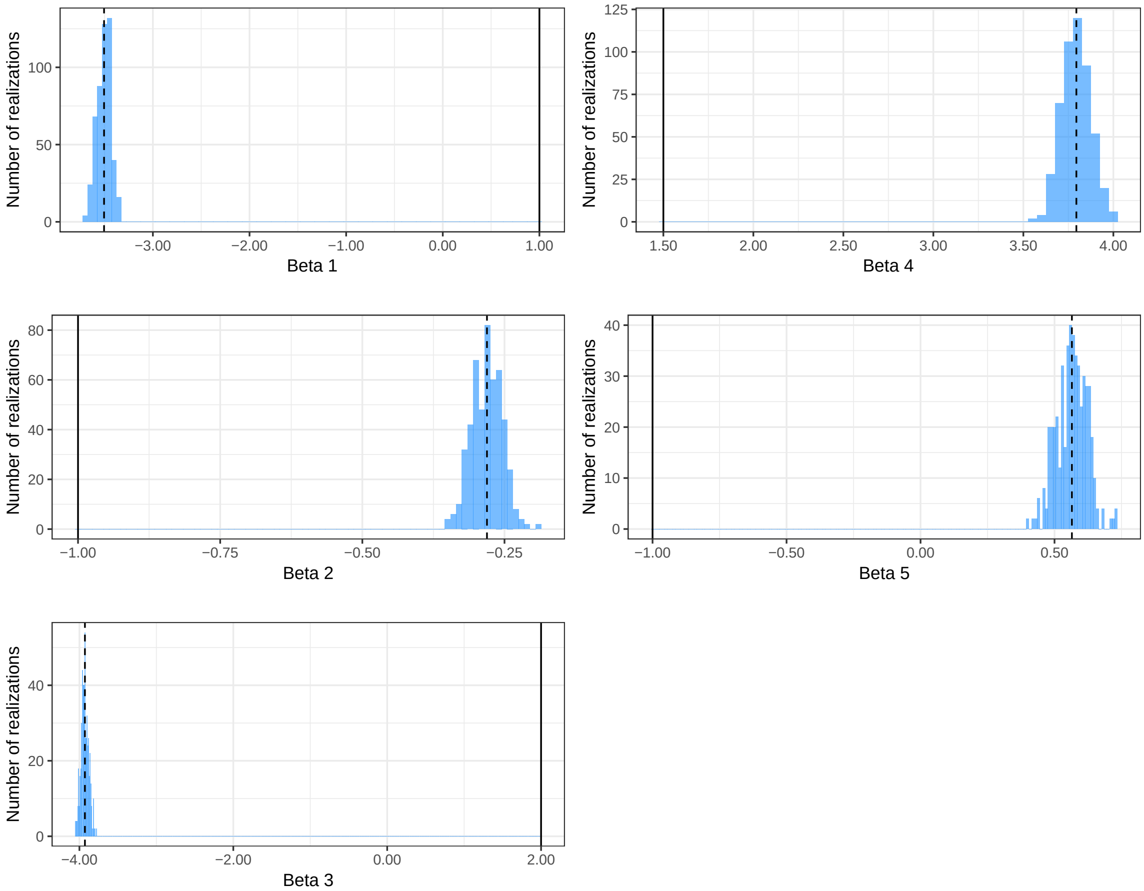


Notes: Monte Carlo simulations under DGP2, estimating parameters with Likelihood approach assuming DGP2 is the truth. Solid lines are true parameters, dashed lines are the the corresponding means of the distributions.

As expected, the distributions of the estimators are centered at the true parameters (solid lines), so the likelihoods are well specified and the model is identified. We observe that estimates given by assuming DGP2 are less precise than those obtained by assuming DGP1. This is because the likelihood under DGP2 includes less information than under DGP1, as the choice sets under DGP2 consider only programs with positive admission probabilities.

In order to show whether assuming truth-telling can lead to biased results in a strategic environment, we estimate the model assuming truth-telling on the simulated data under DGP2. Figure 7.7 shows the distribution of the resulting estimators:

Figure 7.7: Monte Carlo under DGP2, assuming DGP1



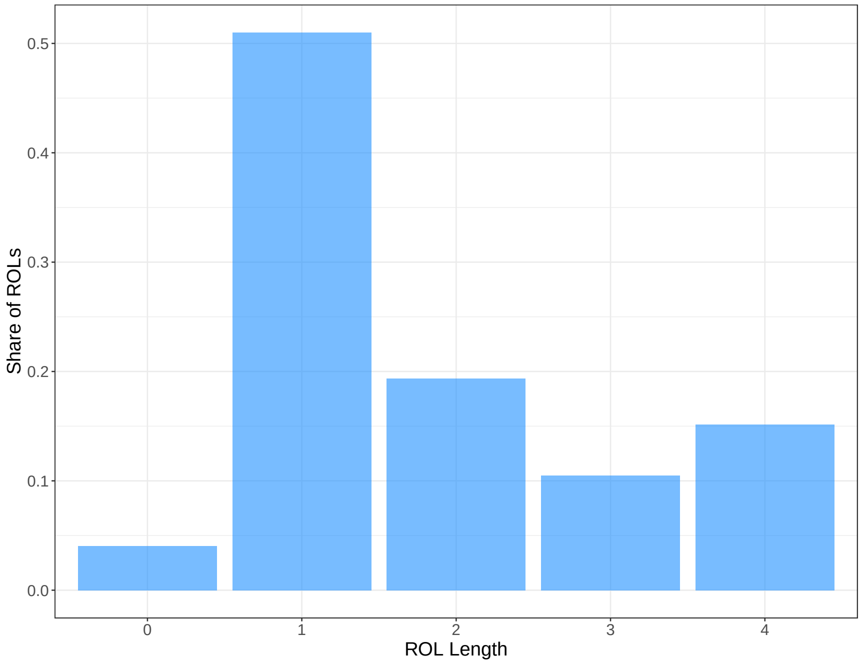
Notes: Monte Carlo simulations under DGP2, estimating parameters with Likelihood approach assuming DGP1 is the truth. Solid lines are true parameters, dashed lines are the the corresponding means of the distributions.

We observe a clear bias in the estimated coefficients. Assuming truth-telling, if data is generated under DGP2, underestimates how preferred are more selective programs compared to other programs (β_4) and compared to the outside option (β_1). These results are in line with the descriptive statistics about the first listed programs. If we include less preference heterogeneity in our specification, or we control by geographic distance, and data is generated by DGP2, we would interpret that preferences for programs are very heterogeneous, without taking into account that this heterogeneity is mainly driven by the heterogeneity in students' choice sets.

7.3 Gibbs Sampler Results

We now simulate data under DGP3, i.e., allowing students to report constrained ROLs. We set $K = 4$ to have a mass ($\sim 20\%$) of students reporting constrained ROLs (similar to what we observe in the actual data) and run 500 Monte Carlo simulations.

Figure 7.8: Distribution of Length of ROLs under DGP3



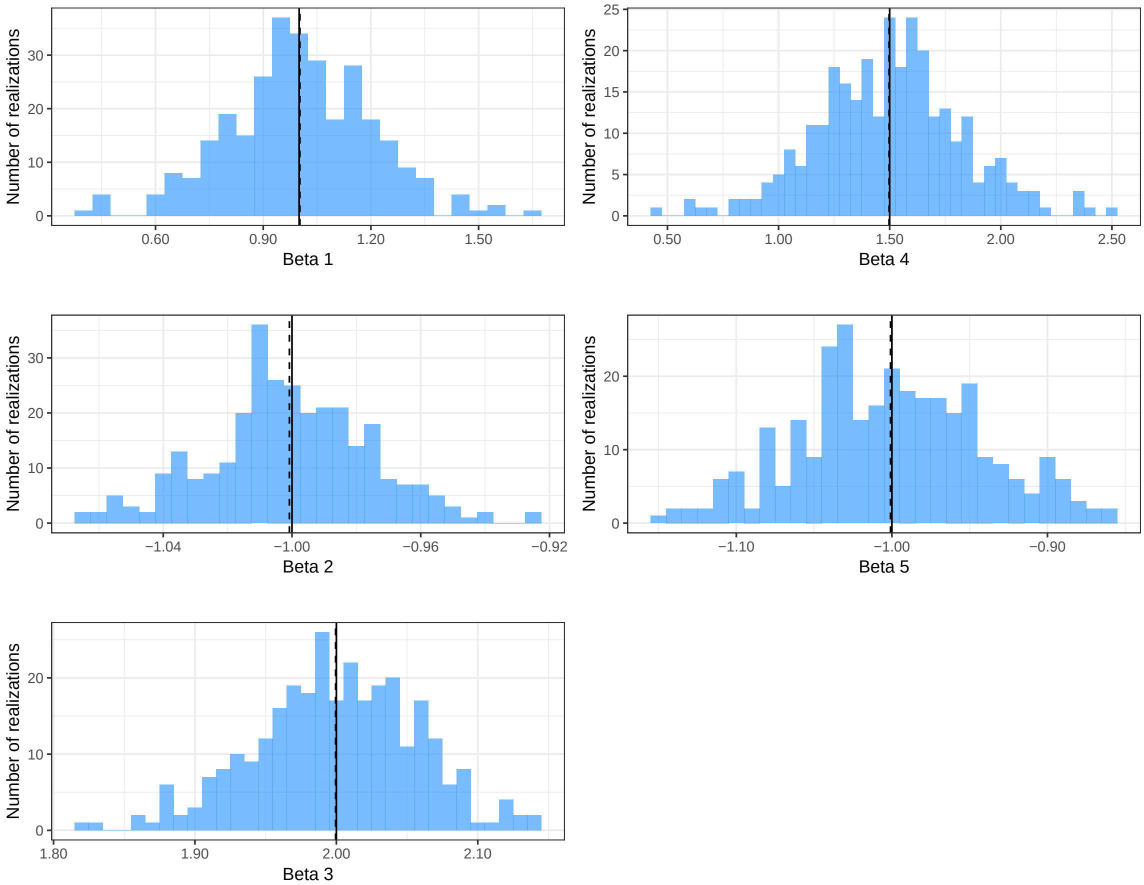
Notes: Distribution of the Length of ROLs under DGP3.

We estimate preferences using the Gibbs Sampler approach discussed in the previous section. We burn-in the first 3,000 iterations and construct the posterior distributions given the data and the priors for the following 2,000 iterations. As standard practice, we choose a diffuse prior

$$p(\beta) \sim N(\bar{\beta}, A^{-1}), \tag{7.1}$$

where $\bar{\beta} = (0, 0, 0, 0)$ and $A^{-1} = 100 \times \mathbb{I}$. The posterior means converge asymptotically to the MLE estimators of the simulated sample. Figure 7.9 shows the distribution of the posterior means for each parameter when data is simulated under DGP3. As expected, there is no bias in estimation.

Figure 7.9: Monte Carlo with Gibbs Sampler under DGP3, assuming DGP3



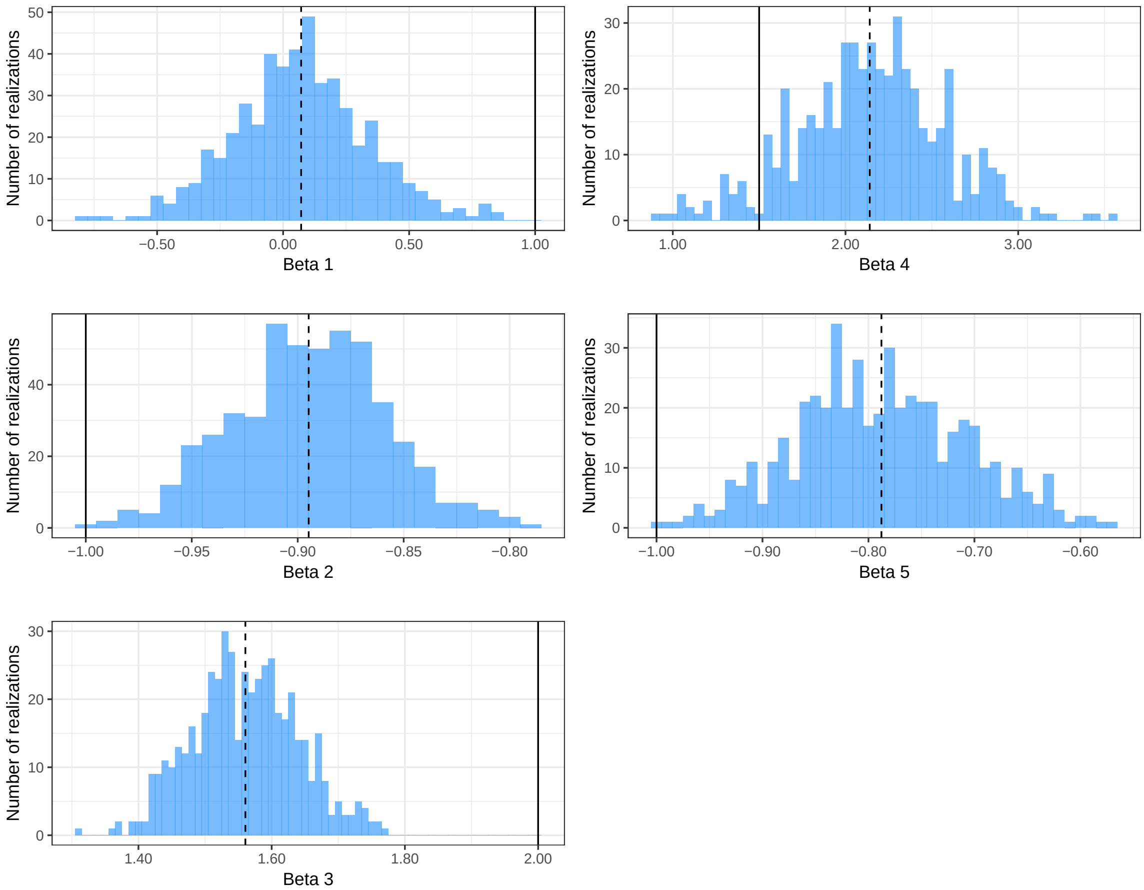
Notes: Monte Carlo simulations under DGP3. Solid lines are true parameters, dashed lines are the means of the distributions of posterior means.

7.3.1 DGP3 vs DGP2

To test the effects of ignoring the constraint in the ROLs, i.e., assuming DGP2 when DGP3 is the truth, we run 500 Monte Carlo simulations under DGP3 with $K = 4$ and estimate the parameters by Maximum Likelihood, assuming DGP2 is the truth. We present the results in Figure 7.10.

We observe that if we ignore the constraint on the length of the list and this constraint binds, estimates are biased. The magnitude of the bias is substantial, even though the fraction of constrained ROLs is close to 20%. In particular, β_1 is downward biased, which can be driven by assuming in the likelihood that programs that are not reported in the list are less preferred than the outside option. As we have pointed out before, even for a constrained ROL, the set of inequalities given by Equations 6.8 and 6.10 hold. This suggests an alternative estimation method, that considers only these inequalities to construct the contribution to the likelihood of constrained ROLs. However, as we are not including information about how preferred are programs not listed in the ROL, estimates will be less precise than the ones obtained by our Gibbs Sampler approach.

Figure 7.10: Monte Carlo under DGP3, assuming DGP2



Notes: Monte Carlo simulations under DGP3, estimating parameters with Likelihood approach assuming DGP2 is the truth. Solid lines are true parameters, dashed lines are the the corresponding means of the distributions.

8 Estimation (in progress)

8.1 Preference specification

We estimate preferences from reported ROLs, enriching the previous model to capture unobserved preferences for majors. In particular, we incorporate a flexible random coefficients model where unobserved heterogeneity depends both on observable characteristics of the students and an unobserved components. Let B be the set of available majors,[19] and let $m(j) \in B$ be the major corresponding to program $j \in M$. We assume that the latent utility of student i for a given program j can be parameterized as:

[19]We refer to majors as the fields of education provided by the International Standard Classification of Education (ISCED) (UNESCO (2012)) that has been adapted to Chile. The modified version of the ISCED fields used in Chile classifies programs into: Farming, Art and Architecture , Science, Social Sciences, Law, Humanities, Education, Technology, Health, Management and Commerce.

$$u_{ij} = \xi_j + \alpha_{i,m(j)} + \beta_d d_{ij} + \frac{\gamma(\bar{s}_i - \bar{P}_j)}{\bar{\sigma}_j} - C_{ij} + \varepsilon_{ij} \tag{8.1}$$

where ξ_j is program j's fixed effect; $\alpha_{i,m(j)}$ is student i's random coefficient for major $m(j)$; d_{ij} is the distance between student i's municipality and program j's municipality; $\bar{s}_i$ is student i's average score,[20] $\bar{P}_j$ is the lowest average score for students assigned in program j in the previous year, and $\bar{\sigma}_j$ is its standard deviation. The coefficient γ captures how much students like a program depending on their ability relative to students assigned in the previous year. On the other hand, C_{ij} captures the monetary cost for student i to enroll in program j, given its yearly tuition t_j. We assume that this cost can be modeled as:

$$C_{ij} = c_0 t_j + c_1 t_j \mathbb{1}_{(\text{low income})} + c_2 t_j \mathbb{1}_{(\bar{s}_i \geq 500)} + c_3 t_j \mathbb{1}_{(\text{low income})} \mathbb{1}_{(\bar{s}_i \geq 500)}. \tag{8.2}$$

Notice that we allow for different price sensitivities depending on the level of income and students' scores. In this way, we capture potential credit constraints that may affect lower income families,[21] as well as potential scholarships or financial aids that high-achieving[22] students may have access to. Finally, $\varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2 \mathbb{I})$ is an idiosyncratic preference shock.

We model the random coefficients as a multivariate regression on a set of students' observable characteristics. Let $\alpha_i \in \mathbb{R}^B$ be the vector of random coefficients of student i. Then, given a vector $z_i \in \mathbb{R}^d$ of observable characteristics, we assume that α_i is given by:

$$\alpha_i = \Delta z_i + \nu_i \sim \text{ iid } N(0, V_\alpha), \quad i = 1, ..., N \tag{8.3}$$

where Δ is a $|B| \times d$ matrix of coefficients, and ν_i is a $|B| \times 1$ vector of idiosyncratic shocks that follow a multivariate normal distribution with zero mean and variance covariance matrix V_α.

In the matrix of covariates z_i we include student i's scores' percentiles on each admission factor and student i's gender. In this way, we allow for unobserved preferences for majors to be different for students with different scores[23] and gender. In addition, notice that we are not allowing for an unrestricted variance covariance matrix of the preference shock. Although this is theoretically possible, the large number of programs in the Chilean setting would lead to a huge number of parameters to be estimated. Instead, we allow for correlations only among programs of the same major by introducing the random coefficients $\alpha_{i,m(j)}$.[24] This reduces considerably the dimensionality of the problem because the matrix V_α has dimensions $|B| \times |B|$ and an unrestricted variance covariance matrix of the preference shock has dimensions $|M| \times |M|$.

[20] The average is taken over the mathematics and verbal PSU tests.

[21] Students are defined to be "low income" students if their self reported family income is bellow the median. Bordon et al. (2015) follow a similar strategy to capture heterogeneous price responsiveness in their model.

[22] See Bucarey (2017) for more details on the choice of this specific threshold.

[23] Unobserved preferences for major could be correlated with students' performance in some admission factors. For instance, students who have high scores in math could be more likely to have a strong preference for STEM programs.

[24] Another alternative would be to consider a block diagonal variance covariance matrix, keeping the number of parameters still manageable. However, estimating restricted variance covariance matrices in a Bayesian setting can be quite challenging. The matrix can no longer be modeled through a Wishart prior and solution methods involve additional simulation steps that would increase the computational time for our proposed Gibbs sampler (Chan and Jeliazkov (2009)). We leave this extension for future research.

8.2 Hierarchical model

To estimate our model, we follow a similar approach than Rossi et al. (2011) and specify a hierarchical Bayes model. The model is built through a series of conditional distributions. The lowest level of the hierarchy is the random utility regression model, conditional on the values of the random coefficients $\{\alpha_i\}$. The higher levels of the hierarchy are given by successive priors on the conditional distribution of the random coefficients. The full model combines the following conditional distributions:

$$u_i | Z_i, \beta, \sigma^2, \alpha_i, R_i \tag{8.4}$$

$$\beta | u_i, Z_i, \sigma^2, \alpha_i \tag{8.5}$$

$$\alpha_i | z_i, u_i, \Delta, V_\alpha \tag{8.6}$$

and the conditional distributions over the hyper-parameters of the random coefficients model:

$$\Delta | \{\alpha_i\}, \bar{\Delta}, A_d, V_\alpha \tag{8.7}$$

$$V_\alpha | \{\alpha_i\}, \Delta, \nu_{ob}, V_{ob} \tag{8.8}$$

where $\bar{\Delta}$, A_d, V_{ob} and ν_{ob} are prior parameters for Δ and V_α (see Appendix C.2 for more details).

Notice that we do not specify a conditional distribution on the variance of the random error, as we fix $\sigma^2 = 1$ as a scale normalization. We choose proper but diffuse priors relative to the likelihood, as it is common practice. We describe the exact form of these distributions and how we adapt the Gibbs sampler to incorporate random coefficients in Appendix C.

8.3 Sample selection

For computational reasons, we divide the country into three geographical zones,[25] and we estimate the model separately for each of them.[26] Specifically, for each zone, we take a random sample of 3,000 students and we restrict their choice set to programs located within their zone. Students who apply to programs outside their zone are dropped from the analysis.[27] This is without major loss, as students tend to apply to programs that are close to their hometowns. For instance, among students from the Central zone, close to 75% applied only to programs within their zone (765 programs). Using this approach, we obtain a good balance between heterogeneity in choices and computational time, making easier to illustrate our methodological results.

[25] North zone (I, II and III regions), Central zone (IV, Metropolitan, V, VI, and VII regions) and South zone (VIII, IX, X, XII, XII).

[26] A similar approach is followed by Bucarey (2017).

[27] Alternatively, we could drop from their ROLs the programs that do not belong to their residence zone and use for estimation the inequalities on indirect utilities implied for the remaining programs.

8.4 Results

8.4.1 Base model

We first estimate preferences for the model without random coefficients, but incorporating the interaction terms of the major fixed effects and the covariates z_i. That is, we assume $\alpha_i = \Delta z_i, \quad i = 1, ..., N$ and use the Gibbs Sampler procedure described in Section 6.3. We burn-in the first 3,000 iterations and construct the posterior distributions given the data and the priors for the following 7,000 iterations. As standard practice, we choose proper but diffuse priors relative to the likelihood. Let $\tilde{\beta} \equiv \text{stack}(\beta, vec(\Delta))$, then the prior is given by

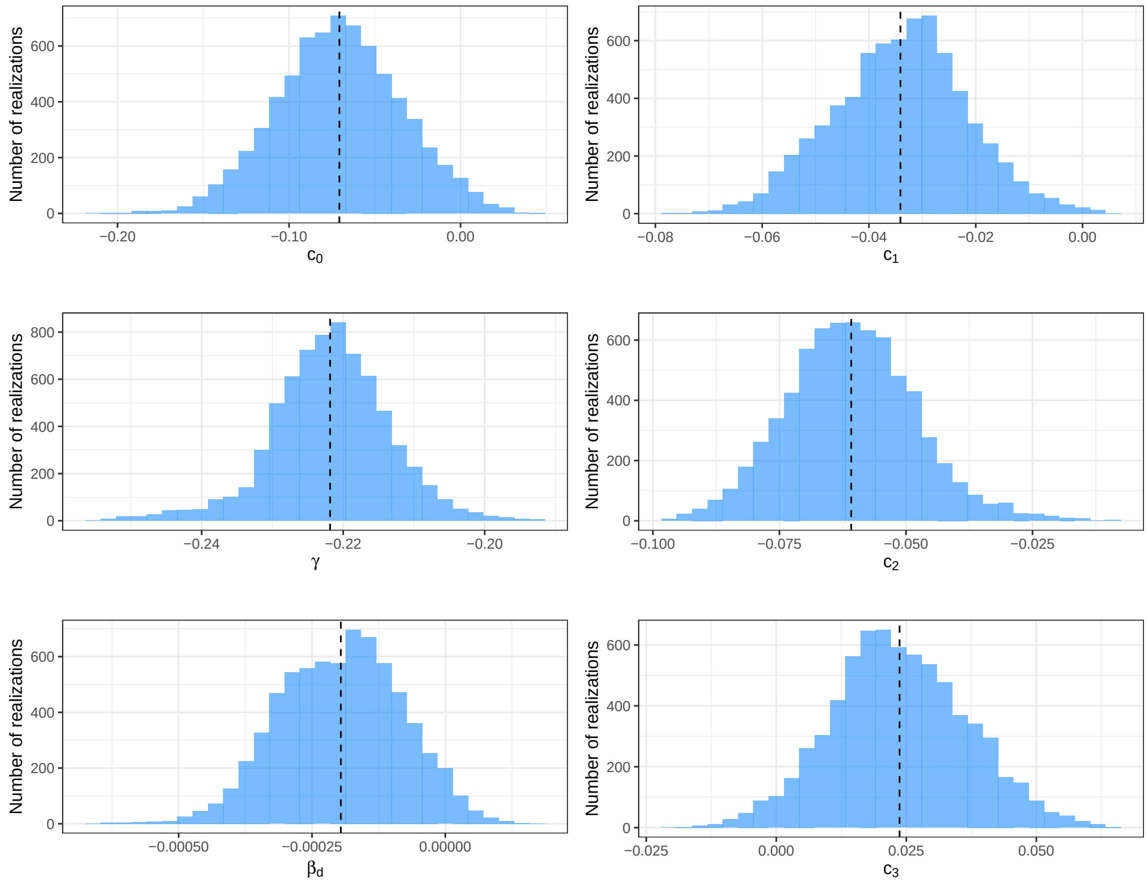
$$p(\tilde{\beta}) \sim N(\bar{\tilde{\beta}}, \tilde{A}^{-1}) \tag{8.9}$$

where $\bar{\tilde{\beta}} = 0, \ \tilde{A}^{-1} = 100 \times \mathbb{I}$.

Figure 8.1 shows the posterior distributions for the parameters of interest for the Metropolitan Region.[28] The dotted line shows the mean of each posterior distribution. Due to our scale normalization, the magnitude of each parameter is in terms of the standard deviation of the preference shock.

[28] Due to the estimation of beliefs in the first stage, the standard errors of the parameters must be corrected.

Figure 8.1: Estimation results with reported ROLs



Notes: Posterior distributions for estimated parameters in the sample. Distance is measured in kilometers and program's tuition is measured in millions of Chilean pesos (nominal).

Interestingly, the term γ, which captures how much students like a program depending on their ability relative to students assigned in the previous year, is found to be negative. A possible interpretation of this result is that students, on average, prefer the most selective programs, regardless of where they stand in the distribution of scores of previously admitted students. To analyze how students' preferences for majors vary with demographics, Table 8.1 shows the posterior mean and standard deviation for the parameters Δ. From these estimates, we can see important gender effects. Women have a higher mean utility than men for studying Health programs, and a lower mean utility for studying Technology programs. Also, the mean utility for majors differs strongly across different score percentiles. As expected, students with a higher score in History receive a positive utility for studying majors that tend to require History, like Social Sciences, Law and Humanities; and students with a high score in Science receive negative mean utilities for studying these majors. These results stress the importance of incorporating students' scores in the utility specification to capture preferences for major. Our identification strategy allows us to incorporate scores as part of preferences because the identifying variation we exploit is the variation on admission weights over time.

Table 8.1: Estimates of Δ, base model

Major	Female	History	Science	Verbal	Math	NEM	Rank
Art and Architecture	0.066	0.2	-0.55	0.19	1	0.69	-1.2
	(0.03)	(0.053)	(0.06)	(0.11)	(0.12)	(0.49)	(0.44)
Farming	0.15	0.098	0.93	0.66	-0.49	-0.62	-0.11
	(0.061)	(0.091)	(0.18)	(0.24)	(0.28)	(0.81)	(0.76)
Technology	-0.4	-0.033	0.3	-0.47	0.98	-2.9	1.7
	(0.02)	(0.03)	(0.051)	(0.084)	(0.1)	(0.27)	(0.25)
Management and Commerce	-0.091	0.27	-0.24	-0.8	1	-0.91	0.15
	(0.031)	(0.051)	(0.059)	(0.1)	(0.11)	(0.35)	(0.31)
Education	0.16	0.078	-0.71	0.26	-0.01	-0.53	-0.41
	(0.032)	(0.055)	(0.059)	(0.11)	(0.11)	(0.38)	(0.35)
Science	-0.065	0.032	0.65	-0.3	-0.38	0.28	-1.3
	(0.038)	(0.057)	(0.11)	(0.13)	(0.18)	(0.59)	(0.55)
Health	0.55	-0.25	1.3	-0.57	-2.3	-1.9	0.0069
	(0.026)	(0.035)	(0.065)	(0.11)	(0.13)	(0.37)	(0.34)
Social Sciences	0.065	0.7	-0.66	0.29	-0.25	-1	-0.051
	(0.028)	(0.072)	(0.047)	(0.13)	(0.099)	(0.37)	(0.34)
Law	-0.2	0.68	-0.83	0.84	-0.93	-0.52	0.22
	(0.068)	(0.18)	(0.13)	(0.25)	(0.24)	(0.9)	(0.81)
Humanities	0.11	0.78	-0.5	1	-0.77	0.62	-1.3
	(0.065)	(0.16)	(0.11)	(0.24)	(0.21)	(0.87)	(0.78)

Notes: Estimates are the means of the posterior distributions for the interaction terms between the major fixed effects (rows) and the covariates in z_i (columns). The standrad deviation of each posterior distribution is given in parenthesis.

9 Conclusions

We analyze the application process in the Chilean College Admissions problem, where the majority of students do not fill their entire application lists. We find evidence of strategic behavior, even though students do not face clear strategic incentives to misreport their true preferences. In particular, students tend to omit programs if their admission probabilities are too low. Under the assumption that students do not include programs in their application lists if it is not strictly profitable to do so, we construct a portfolio problem where students maximize their expected utility of reporting a ROL given their preferences and beliefs over admission probabilities.

In order to better identify the model, we exploit an exogenous variation in the admission weights over time that is unique to the Chilean system. Assuming rational expectations and independence of beliefs on admission probabilities, we show that it is sufficient to compare a ROL with only a subset of ROLs ("one-shot swaps") to ensure its optimality. Using this finding we construct a Likelihood-based approach to estimate student preferences, adapting the estimation procedure proposed by Agarwal and Somaini (2018) to solve a large portfolio problem, without running into the curse of dimensionality.

We simulate data on portfolio choices using the Marginal Improvement Algorithm under different DGPs and run Monte Carlo simulations with our proposed estimation methods. We compare our results against assuming truth-telling of "short-list" students and find biased results. If students do not include programs for which their marginal benefit is zero but we assume truth-telling in estimation, we would underestimate how preferred are selective programs and overstate the value of being unassigned. Moreover, assuming truth-telling can lead to overstate the degree of preference heterogeneity in the system. In addition, ignoring the constraint on the length of the list can also result in biased estimates, even if the proportion of constrained ROLs is relatively small.

Our proposed estimation method is computationally feasible for large scale portfolio problems whenever beliefs on admission probabilities can be estimated in a first stage and assumed to be independent across alternatives. Even though we assume strategic behavior of students to generate the data, the estimation procedure is also robust when students do not skip programs if the marginal benefit of including them is zero. Our estimation results strongly suggest that "short-list" students should not be interpret as truth-tellers, even in a seemingly strategy-proof environment.

We apply our estimation method to estimate students' preferences for programs and majors in Chile and find strong differences in preferences regarding students' gender and scores. Students, on average, prefer the most selective programs, regardless of where they stand in the distribution of scores of previously admitted students. Women, conditional on their admission probabilities, tend to prefer more programs in health majors and prefer less programs in technology majors compared to men. Finally, students' scores matter for understanding their preferences for majors, stressing the importance of including them in the utility specification.

References

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017). The Welfare Effects of Coordinated Assignment: The New York City High School Match. *American Economic Review*, 95(2):364–367.

Agarwal, N. and Somaini, P. (2018). Demand analysis using strategic reports: An application to a school choice mechanism. *Econometrica*, 86(2):391–444.

Ajayi, K. and Sidibe, M. (2017). An Empirical Analysis of School Choice under Uncertainty.

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

Azevedo, E. M. and Budish, E. (2017). Strategy-proofness in the Large.

Bordon, P. and Fu, C. (2010). College-major choice to college-then-major choice. *Review of Economic Studies*, 82(4):1247–1288.

Bordon, P., Fu, C., and Gazmuri, A. (2015). Competition and the cannibalization of college quality Growth in College Enrollment in Chile , 2005-2014.

Bucarey, A. (2017). Who Pays for Free College? Crowding Out on Campus.

Calsamglia, C., Fu, C., and Güell, M. (2018). Structural estimation of a model of school choices: The boston mechanism vs. its alternatives.

Chade, H. and Smith, L. (2006). Simultaneous Search. *Econometrica*, 74(5):1293–1307.

Chan, J. C. C. and Jeliazkov, I. (2009). MCMC estimation of restricted covariance matrices. *Journal of Computational and Graphical Statistics*, 18(2):457–480.

Fack, G., Grenet, J., and He, Y. (2015). Beyond Truth-Telling: Preference Estimation with Centralized School Choice. *PSE Working Papers*, 35(295298):1–79.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, 69(1):9–15.

Haeringer, G. and Klijn, F. (2009). Constrained school choice. *Journal of Economic Theory*, 144(144):1921–1947.

He, Y. (2012). Gaming the Boston School Choice Mechanism in Beijing. *Manuscript, Toulouse School of Economics*, 2012(May).

Kapor, A., Neilson, C. A., and Zimmerman, S. (2017). Heterogeneous Beliefs and School Choice Mechanisms.

Lafortune, J., Figueroa, N., and Saenz, A. (2016). Do you like me enough? The impact of restricting preferences ranking in a university matching process.

Larroucau, T., Ríos, I., and Mizala, A. (2015). The effect of including high school grade rankings in the admission process for chilean universities. *Revista de Investigación Educacional Latinoamericana*, 52262(1):95–118.

Larroucau, T., Rios, I., and Neilson, C. (2019). Strategic mistakes in the Chilean College Admissions Problem.

Luflade, M. (2017). The value of information in centralized school choice systems.

McCulloch, R. and Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64(1-2):207–240.

Olszewski, W. and Vohra, R. (2016). Simultaneous selection. *Discrete Applied Mathematics*, 200:161–169.

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

Ríos, I., Larroucau, T., Parra, G., and Cominetti, R. (2018). Improving the Chilean College Admissions System.

Rossi, P. E., McCulloch, R. E., and Allenby, G. M. (2011). Hierarchical Modelling of Consumer Heterogeneity: An Application to Target Marketing. pages 323–349.

Shorrer, R. and Sóvágó, S. (2017). Obvious mistakes in a strategically simple college admissions environment.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *Review of Economic Studies*, 70(1):147–165.

UNESCO, I. (2012). *International standard classification of education: ISCED 2011*. UNESCO Institute for Statistics Montreal.

Appendix

A Proofs

A.1 Proposition 1

In a slight abuse of notation, we denote by $R(k)$ the k-th preference in ROL R, and by $R(j_1, j_2)$ for $1 \leq j_1 < j_2 \leq |R|$ the subset of ROL R that includes preferences from j_1 to j_2, i.e. $R(j_1, j_2) = \{r_{j_1}, r_{j_1+1}, \ldots, r_{j_2}\}$. We will use this notation in the proof of our main Proposition.

Proposition 1. Let $R = \{R(1), \ldots, R(k)\}$ be a ROL of length at most K, i.e. $k \leq K$. If

$$U(R) \geq U(R'),\ \forall R' \in \mathcal{S}(R) \tag{A.1}$$

then

$$U(R) \geq U(R'),\ \forall R' \in \bigcup_{l=1}^{K} \mathcal{R}_l \tag{A.2}$$

Proof. We proceed by induction on the maximum number of programs allowed in a ROL, K.
Basis Notice that Proposition 1 trivially holds for $K = 1$. Thus, we consider as basis $K = 2$. Suppose that ROL $R = AB$ is optimal. Equation A.1 implies that $U(R) \geq U(R'),\ \forall R' \in \mathcal{S}(R) = \{AX, XA, BX, XB :\ X \in M \setminus R\}$. Then, to show the optimality of R it remains to show that $U(R) \geq U(XY)$ for any $X, Y \in M \setminus R$. From Equation A.1 we know that $z_B > z_j\ \forall j \in M \setminus R$, we know that $U(XB) > U(XY)$, and since $U(R) = U(AB) \geq U(XB)$ we conclude that Equation A.2 holds in this case.

Step In the induction step we must show that if Proposition 1 holds for ROLs of length at most $K = k$, then it must also hold for $K = k + 1$. To show this, we first show that if $R \in \mathcal{R}_{k+1}$ satisfies Equation A.1, then there exists a subset $R_k \subset R$ of length k that satisfies Equation A.1 for $K = k$. Then, using the inductive hypothesis we know that R_k is optimal when $K = k$. Since we know that MIA is optimal under Assumption 2, it is enough to show that Equation A.1 guarantees that the marginal benefit of adding program $\tilde{r} = R \setminus R_k$ to ROL R_k leads to the highest marginal improvement. This, combined with the fact that R_k is optimal for $K = k$, implies that R is optimal when $K = k + 1$, and therefore Equation A.2 holds for $K = k + 1$.

Let $R \in \mathcal{R}_{k+1}$ be a ROL satisfying Equation A.1 for $K = k + 1$. Without loss of generality we assume that $|R| = k + 1$, i.e. $u_j > u_0$ and $p_j > 0$ for all $j \in R$, and thus we write $R = \{R(1), \ldots, R(k + 1)\}$.[29] Let R_k be the subset of R that maximizes the expected utility given $K = k$, i.e.

$$U(R_k) = \max_{R' \subset R, |R'| \leq k} U(R').$$

Let $\tilde{r} = R \setminus R_k$ be the program left out, and $\bar{R} = \{j \in M \setminus R :\ p_j > 0,\ u_j > u_0\}$ the set of programs that are not in R and that would lead to a weak improvement of the expected utility if added to a ROL. Since R satisfies Equation A.1, we know that

$$z_{r_{k+1}} \geq z_j,\ \forall j \in \bar{R}, \tag{A.3}$$

[29] The case where $|R| < K$ is straightforward and thus we omit it.

and since $\tilde{r}$ is not in R_k we also know that

$$z_{R_k(k)} \geq z_j,\ \forall j \in \bar{R} \tag{A.4}$$

The first step of the proof is to show that

$$U(R) \geq U(R'),\ \forall R' \in \mathcal{S}(R) \Rightarrow U(R_k) \geq U(R''),\ \forall R'' \in \mathcal{S}(R_k)$$

To find a contradiction suppose that this is not the case, i.e. $U(R) \geq U(R'),\ \forall R' \in \mathcal{S}(R)$ but $\exists R'' \in \mathcal{S}(R_k)$ such that $U(R_k) < U(R'')$. By optimality of R_k among all the subsets of R, it must be the case that $r = R'' \setminus R_k \in \bar{R}$. Equation A.4 implies that r cannot be the last preference of R'' (otherwise we would have $U(R'') < U(R_k)$). Hence, $R''(k) = R_k(k)$ and therefore we have $u_r > u_{R_k(k)}$. Combining this with Equation A.4 we have $p_{R_k(k)} > p_r$. Then, by Lemma 1 we know that

$$U(R'') > U(R_k) \Rightarrow U(R'' \cup \{\tilde{r}\}) > U(R_k \cup \{\tilde{r}\}) = U(R)$$

but since $R'' \cup \{\tilde{r}\} \in \mathcal{S}(R)$ this contradicts the assumption given by Equation A.1 for ROL R with $K = k + 1$.

Now we know that R_k satisfies Equation A.1 with $K = k$, so by inductive hypothesis we know that R_k is optimal among the ROLs of length at most $K = k$. It remains to show that adding $\tilde{r}$ to R_k leads to the maximum marginal benefit. Nevertheless, this is direct from Equation A.1 applied to R, since $U(R) = U(R_k \cup \{\tilde{r}\}) \geq U(R'),\ \forall R' \in \mathcal{S}(R)$ implies that

$$MB(R_k, \tilde{r}) = U(R_k \cup \{\tilde{r}\}) - U(R_k) > U(R_k \cup \{j\}) - U(R_k) = MB(R_k, j),\ \forall j \in \bar{R}.$$

We then know that R_k satisfies Equation A.2 for $K = k$ and that adding $\tilde{r}$ to R_k leads to the maximum marginal improvement. Since MIA is optimal in our setting, we conclude that $R = R_k \cup \tilde{r}$ must be optimal for $K = k + 1$, concluding our proof. □

Lemma 1. Let $R \in \mathcal{R}$ and $R' \in \mathcal{S}(R)$ such that $U(R) \geq U(R')$. Then, for any $r \in M \setminus R \cup R'$,

$$U(R \cup \{r\}) \geq U(R' \cup \{r\}).$$

Proof. Let j_1, j_2 the indexes of the first and last preference where R and R' are different, i.e. $j_1 < j_2$ and $R(k) = R'(k)\ \forall k \in [1, j_1) \cup (j_2, |R|]$. Without loss of generality we assume that $R(j_1 \notin R')$ and thus $R'(j_1) = R(j_1 + 1)$, i.e. the first difference between R and R' is a program in R but not in R'.[30] Now suppose that program r is such that

$$u_{R(l-1)} \geq u_r \geq u_{R(l)},$$

i.e. upon entering ROL R, program r would take the n-th position in the new ROL, i.e. $R \cup \{r\} = (R(1), \ldots, R(l-1), r, R(l), \ldots, R(k))$. We want to show that

$$U(R) \geq U(R') \Rightarrow U(R \cup \{r\}) \geq U(R' \cup \{r\}).$$

We consider three particular cases:

[30] The proof for the converse case where the first difference between R and R' is a program in R' but not in R is analogous.

Case 1: Suppose $n < j_1$. Then $R(l) = R'(l)$, $\forall l = 1, \ldots, n$, and therefore[31]

$$U(R \cup \{r\}) - U(R' \cup \{r\}) = (1 - p_r) \cdot \left(\prod_{l=1}^{n-1} (1 - p_{R(l)}) \right) \cdot \big[U(R(n,k)) - U(R'(n,k)) \big] \geq 0$$

where the inequality follows from $U(R) - U(R') \geq 0$.

Case 2: Suppose $n > j_2$. We know that

$$\begin{aligned}
U(R) - U(R') &= U(R(1, j_1 - 1)) + \left(\prod_{l=1}^{j_1-1} (1 - p_{R(l)}) \right) \cdot \big[z_{R(j_1)} + (1 - p_{R(j_1)}) \cdot U(R(j_1 + 1, j_2 - 1)) \big] \\
&+ (1 - p_{R(j_1)}) \cdot \left(\prod_{l=1}^{j_1-1} (1 - p_{R(l)}) \right) \cdot \left(\prod_{l=j_1+1}^{j_2-1} (1 - p_{R(l)}) \right) \cdot U(R(j_2 + 1, k)) \\
&- U(R(1, j_1 - 1)) - \left(\prod_{l=1}^{j_1-1} (1 - p_{R(l)}) \right) \cdot U(R(j_1 + 1, j_2 - 1)) \\
&- z_{R(j_2)} \cdot \left(\prod_{l=1}^{j_1-1} (1 - p_{R(l)}) \right) \cdot \left(\prod_{l=j_1+1}^{j_2-1} (1 - p_{R(l)}) \right) \\
&- (1 - p_{R(j_2)}) \cdot \left(\prod_{l=1}^{j_1-1} (1 - p_{R(l)}) \right) \cdot \left(\prod_{l=j_1+1}^{j_2-1} (1 - p_{R(l)}) \right) U(R(j_2 + 1, k)) \\
&= \left(\prod_{l=1}^{j_1-1} (1 - p_{R(l)}) \right) \cdot \big[z_{R(j_1)} - p_{R(j_1)} \cdot U(R(j_1 + 1, j_2 - 1)) \big] \\
&+ \left(\prod_{l=1}^{j_1-1} (1 - p_{R(l)}) \right) \cdot \left(\prod_{l=j_1+1}^{j_2-1} (1 - p_{R(l)}) \right) \cdot \big[(p_{R(j_2)} - p_{R(j_1)}) U(R(j_2 + 1, k)) - z_{R(j_2)} \big] \\
&\geq 0
\end{aligned}$$

Similarly, after some algebra we obtain that

$$\begin{aligned}
U(R \cup \{r\}) - U(R' \cup \{r\}) &= \left(\prod_{l=1}^{j_1-1} (1 - p_{R(l)}) \right) \cdot \big[z_{R(j_1)} - p_{R(j_1)} \cdot U(R(j_1 + 1, j_2 - 1)) \big] \\
&+ \left(\prod_{l=1}^{j_1-1} (1 - p_{R(l)}) \right) \cdot \left(\prod_{l=j_1+1}^{j_2-1} (1 - p_{R(l)}) \right) \cdot \big[(p_{R(j_2)} - p_{R(j_1)}) U(R(j_2 + 1, k) \cup \{r\}) - z_{R(j_2)} \big]
\end{aligned}$$

To show that $U(R \cup \{r\}) - U(R' \cup \{r\}) \geq 0$ it is therefore enough to show that

$$(p_{R(j_2)} - p_{R(j_1)}) U(R(j_2 + 1, k) \cup \{r\}) - z_{R(j_2)} \geq (p_{R(j_2)} - p_{R(j_1)}) U(R(j_2 + 1, k)) - z_{R(j_2)}$$

which is equivalent to show that

$$U(R(j_2 + 1, k) \cup \{r\}) \geq U(R(j_2 + 1, k)).$$

---

[31] $U(R(l,k))$ is the utility derived from the subset of ROL R starting in preference l up to the k-th preference. Then,

$$U(R(l,k)) = z_{R(l)} + (1 - p_{R(l)}) z_{R(l+1)} + \ldots + \prod_{l=n}^{k-1} (1 - p_{R(l)}) \cdot z_{R(k)}.$$

Since we assume that r enters ROL R in the n-th position, this is equivalent to show that

$$z_r + (1-p_r)\cdot U(R(n,k)) \geq U(R(n,k))$$

which is direct from the fact that $u_r \geq u_{R(n)} \geq \ldots \geq u_{R(k)}$ and $p_r > 0$. Thus, we conclude that $U(R\cup\{r\}) - U(R'\cup\{r\}) \geq 0$.

Case 3: Suppose $n \in [j_1, j_2]$. Doing some algebra we know that

$$\begin{aligned}
U(R) - U(R') &= z_{R(j_1)} + (1-p_{R(j_1)})\cdot U(R(j_1+1, j_2-1)) \\
&\quad + (1-p_{R(j_1)})\cdot\left(\prod_{l=j_1+1}^{j_2-1}(1-p_{R(l)})\right)\cdot U(R(j_2+1,k)) \\
&\quad - U(R(j_1+1,j_2-1)) - \left(\prod_{l=j_1+1}^{j_2-1}(1-p_{R(l)})\right)\cdot z_{R(j_2)} \\
&\quad - (1-p_{R(j_2)})\cdot\left(\prod_{l=j_1+1}^{j_2-1}(1-p_{R(l)})\right)\cdot U(R(j_2+1,k)) \\
&= z_{R(j_1)} - p_{R(j_1)}\cdot U(R(j_1+1,j_2-1)) \\
&\quad + \left(\prod_{l=j_1+1}^{j_2-1}(1-p_{R(l)})\right)\cdot\big((p_{R(j_2)} - p_{R(j_1)})U(R(j_2+1,k) - z_{R(j_2)})\big) \\
&\geq 0
\end{aligned}$$

Similarly, doing similar algebra we find that

$$\begin{aligned}
U(R\cup\{r\}) - U(R'\cup\{r\}) &= z_{R(j_1)} - p_{R(j_1)}\cdot U(R(j_1+1,j_2-1)) + \\
&\quad + \left(\prod_{l=j_1+1}^{j_2-1}(1-p_{R(l)})\right)\cdot\big((p_{R(j_2)} - p_{R(j_1)})U(R(j_2+1,k) - z_{R(j_2)})\big)
\end{aligned}$$

We notice that if

$$z_{R(j_1)} - p_{R(j_1)}\cdot\left[U(R(j_1+1,n-1)) - \left(\prod_{l=j_1+1}^{n-1}(1-p_{R(l)})\right)\cdot z_r\right] \geq (1-p_r)\cdot\big[z_{R(j_1)} - p_{R(j_1)}\cdot[U(R(j_1+1,n-1))]\big] \tag{A.5}$$

then

$$U(R\cup\{r\}) - U(R'\cup\{r\}) \geq (1-p_r)\cdot\big(U(R) - U(R')\big)$$

and since $U(R) - U(R') \geq 0$ this would imply that $U(R\cup\{r\}) \geq U(R'\cup\{r\})$. It is easy to see that this is indeed the case. In fact, since $p_r > 0$ and $p_{R(l)} < 1$ for $l = 1,\ldots,n$, Equation A.5 is equivalent to show that

$$z_{R(j_1)} - p_{R(j_1)}\cdot U(R(j_1+1,n-1)) - p_{R(j_1)}\cdot u_r\cdot\prod_{l=j_1+1}^{n-1}(1-p_{R(l)}) \geq 0$$

which in turn is equivalent to show that

$$z_{R(j_1)} - p_{R(j_1)}\cdot U(R(j_1+1,n-1)\cup\{r\}) \geq 0.$$

Then,

$$\begin{aligned} z_{R(j_1)} - p_{R(j_1)} \cdot U(R(j_1+1, n-1) \cup \{r\}) &= p_{R(j_1)} \cdot u_{R(j_1)} - p_{R(j_1)} \cdot U(R(j_1+1, n-1) \cup \{r\}) \\ &= p_{R(j_1)} \cdot \left[u_{R(j_1)} - U(R(j_1+1, n-1) \cup \{r\})\right] \end{aligned}$$

and since $u_{R(j_1)} \geq u_{R(j_1+1)} \geq \ldots \geq u_{R(n-1)} \geq u_r$, we conclude that $u_{R(j_1)} \geq U(R(j_1 + 1, n-1) \cup \{r\})$, so we conclude that Equation A.5 holds, concluding the proof.

□

A.2 Proposition A.2

In this section we prove the following proposition.

Proposition 1. Let $R = \{R(1), \ldots, R(k)\}$ be a ROL of length at most K, i.e. $k \leq K$, and suppose there exists a program $j \in M \setminus R$ such that $p_j > 0$ and $p_j \geq p_{R(k)}$ for some $R(k) \in R$. Let $\bar{k} = \arg\min_k \left\{p_{R(k)} \;:\; p_{R(k)} \leq p_j\right\}$. Then, if $\bar{k} \in \{1, \ldots, |R|\}$,

$$U(R) \geq U(R'),\ \forall R' \in \mathcal{S}(R)$$

if and only if

$$U(R) \geq U(R'),\ \forall R' \in \mathcal{S}(R) \setminus \bigcup_{k=\bar{k}}^{|R|} \bigcup_{l=1}^{\bar{k}-1} \mathcal{S}_{jkl}(R).$$

As discussed in Proposition X, we only need to consider inequalities related to one-shot swaps when $|R| = K$. Hence, suppose that $|R| = K$, and let $j \in M \setminus R$ be a program for which $p_j > 0$ and $\bar{k} = \arg\min_k \left\{p_{R(k)} \;:\; p_{R(k)} \leq p_j\right\} \geq 1$. We only care about programs with $p_j > 0$ as those are the only candidates to enter the ROL. The main idea of the proof is to show that the inequalities related to the one shot-swaps in $\bigcup_{k=\bar{k}}^{|R|} \bigcup_{l=1}^{\bar{k}-1} \mathcal{S}_{jkl}(R)$ are implied by inequalities that are already included in the set of inequalities, and thus are redundant. In particular, given $R = \{R(1), \ldots, R(K)\}$, we know that

$$u_{R(1)} \geq u_{R(2)} \geq \ldots \geq u_{R(k)} \geq u_0$$

Our first lemma is a general property of MIA, which states that if a program j dominates another program j' (i.e., $u_j \geq u_{j'}$ and $p_j > p_{j'}$), then if j' enters the optimal ROL R we must have that j also belongs to R.

Lemma 2. Consider two programs j, j' such that $p_j > p_{j'}$ and $u_j \geq u_{j'}$, and let R be the optimal ROL. Then, $j' \in R$ implies that $j \in R$.

Proof. We know that R is the resulting ROL from applying MIA. Then, it is enough to prove that j is included before j' during the execution of MIA. To see this, suppose that at some point in the algorithm we have a ROL $\tilde{R} \subset R$ such that $j \notin \tilde{R}$ and $j' \notin \tilde{R}$ (notice that $\tilde{R}$ may be the empty set). In addition suppose that, conditional on entering $\tilde{R}$, j would be place in the k-th position, while j' would be placed in the k'-th position of the ROL. Since $u_j \geq u_{j'}$ we know that

$k \leq k'$. Let $\tilde{R}_{jk}$ be the resulting ROL that adds j to $\tilde{R}$ in the k-th position, and similarly define $\tilde{R}_{jk'}$ and $\tilde{R}_{j'k'}$. Then, we know that

$$U(\tilde{R}_{jk}) \geq U(\tilde{R}_{jk'}) \geq U(\tilde{R}_{j'k'}),$$

where the first inequality comes from k being the optimal position to place j, and the second inequality comes from $p_j > p_{j'}$ and $u_j \geq u_{j'}$, since

$$p_j u_j + (1 - p_j) U(\tilde{R}(k', |\tilde{R}|)) \geq p_{j'} u_j + (1 - p_{j'}) U(\tilde{R}(k', |\tilde{R}|)) \geq p_{j'} u_{j'} + (1 - p_{j'}) U(\tilde{R}(k', |\tilde{R}|)).$$

Finally, our previous inequality implies that

$$U(\tilde{R}_{jk}) - U(\tilde{R}) \geq U(\tilde{R}_{j'k'}) - U(\tilde{R}),$$

and therefore the marginal improvement of including j is greater than or equal to that of including j' for any ROL $\tilde{R}$, concluding the proof. □

Our next lemma shows that program j will never replace a program in position l if $l < \bar{k}$.

Lemma 3. Let $\tilde{K} = \left\{k \,:\, k < \bar{k}\right\}$. The constraints

$$U(R) \geq U(R'),\ \forall R' \in \bigcup_{k=1}^{K} \bigcup_{l \in \tilde{K}} \mathcal{S}_{jkl}(R).$$

are redundant and thus can be omitted.

Proof. To find a contradiction, suppose that the constraint is not redundant for some $k \in \tilde{K}$. Hence, there exists a ROL $R' \in \mathcal{S}(R)$ that eliminates program k and places program j in some position of the new ROL, such that if this constraint is omitted it would be violated, i.e., $U(R) < U(R')$.

Since $R' \in \mathcal{S}(R)$ and k is removed, we know that $\bar{k} \in R'$. We also know that $u_{R(k)} \geq u_{R(\bar{k})}$ (since $k < \bar{k}$), and we also know that $p_{R(k)} \geq p_{R(\bar{k})}$ (by definition of $\bar{k}$). Hence, by Lemma 2 we have that $\bar{k} \in R'$ implies that $k \in R'$, which leads to a contradiction. □

From our previous lemma we know that all OSS that include j and remove some program $R(l)$ such that $l < \bar{k}$ are redundant. Now we start analyzing programs in preferences $\tilde{K} = \left\{\bar{k}, \ldots, K\right\}$, as these are candidates to be removed.

Let $\tilde{k} = \max\left\{k \,:\, p_{R(k)} \leq p_j\right\}$, i.e., $\tilde{k}$ is the lowest ranking for which the probability is still below program j. Notice that, given these definitions, $\bar{k} \leq \tilde{k}$. Our next lemma shows that constraints involving programs $R(k)$ with $k \in \left\{\bar{k}+1, \ldots, \tilde{k}-1\right\}$ and $p_{R(k)} \geq p_{R(\tilde{k})}$ are also redundant.

Lemma 4. Let $\tilde{k} = \max\left\{k \,:\, p_{R(k)} \leq p_j\right\}$, and let $K' = \left\{k \,:\, p_{R(k)} \geq p_{R(\tilde{k})},\ k \in \tilde{K},\ k < \tilde{k}\right\}$ Then, the constraints

$$U(R) \geq U(R'),\ \forall R' \in \bigcup_{k=1}^{K} \bigcup_{l \in K'} \mathcal{S}_{jkl}(R).$$

are redundant and thus can be omitted.

Proof. To find a contradiction suppose that the statement is not true, i.e., one of these constraints is not redundant. Let $k \in K'$ be the preference of the program that would be removed in that constraint. Since $k < \tilde{k}$ we know that $u_{R(k)} \geq u_{R(\tilde{k})}$, and by Hypothesis we know that $p_{R(k)} \geq p_{R(\tilde{k})}$. Hence, we program $R(k)$ dominates $R(\tilde{k})$, and thus $\tilde{k} \in R'$ implies that $k \in R'$ by Lemma 2. However, this contradicts the fact that program $R(k)$ was removed to form the new ROL R', concluding our proof. □

Our results so far show that all constraints removing programs $R(k)$ with $k < \bar{k}$ or $k \in \left\{\bar{k}+1, \ldots, \tilde{k}-1\right\}$ provided that $p_{R(k)} \geq p_{R(\tilde{k})}$ are redundant. Hence, the constraints that may not be redundant are those removing programs $R(k)$ where $k > \tilde{k}$ or $k \in \left\{\bar{k}+1, \ldots, \tilde{k}-1\right\}$ with $p_{R(k)} < p_{R(\tilde{k})}$. In the next two lemmas we show which constraints that involve removing these programs are also redundant.

Our next lemma deals with the case where $k > \tilde{k}$, and thus, by definition, $p_{R(k)} > p_j$. It shows that, if program $R(k)$ is removed to include program j, then all constraints that involving placing j in a preference above $\tilde{k}$ are redundant.

Lemma 5. Let $\underline{K} = \left\{k \ : \ p_{R(k)} > p_j, \ k > \tilde{k}\right\}$. Then, the constraints

$$U(R) \geq U(R'), \ \forall R' \in \bigcup_{l \in \underline{K}} \bigcup_{k=1}^{\tilde{k}} \mathcal{S}_{jkl}(R).$$

are redundant and thus can be omitted.

Proof. To find a contradiction, suppose this is not true. Then, we may have that j enters the ROL in a position $k \leq \tilde{k}$. This implies that $u_j \geq u_{R(\tilde{k})}$, and by definition of $\tilde{k}$ we know that $p_j \geq p_{R(\tilde{k})}$. Hence, program j dominates program $R(\tilde{k})$, and thus should be added first by MIA. Moreover, this would imply that the ROL R' that replaces $\tilde{k}$ with j would lead to a higher expected utility. However, this is not possible because we still have the constraint $U(R) \geq U(R')$ for $R' \in \mathcal{S}_{j\tilde{k}\tilde{k}}$, and thus we have a contradiction. □

Our last lemma deals with the case where we want to remove program $R(k)$ with $k \in \left\{\bar{k}+1, \ldots, \tilde{k}-1\right\}$ and $p_{R(k)} < p_{R(\tilde{k})}$. This lemma guarantees that all one shot-swaps that add program j in position above $\tilde{k}$ after removing a program $R(k)$ are redundant.

Lemma 6. Let $\bar{K} = \left\{k \ : \ p_{R(k)} < p_{R(\tilde{k})}, \ k \geq \bar{k}\right\}$. The constraints

$$U(R) \geq U(R'), \ \forall R' \in \bigcup l \in \bar{K} \bigcup_{k=1}^{\tilde{k}} \mathcal{S}_{jkl}(R).$$

are redundant and thus can be omitted.

Proof. Without loss of generality, consider a particular $k \in \bar{K}$. To find a contradiction, suppose that the constraint removing $R(k)$ and placing j in a preference above $\tilde{k}$ is not redundant. Hence, the corresponding ROL R' that removes $R(k)$ and places j in some position above $\tilde{k}$ can lead to a utility higher than R, i.e., $U(R') > U(R)$. Since j is placed above $\tilde{k}$, we know that $u_j > u_{R(\tilde{k})}$, and by definition of $\tilde{k}$ we also know that $p_j > p_{R(\tilde{k})}$. Therefore, program j dominates program $R(\tilde{k})$. However, the constraint implied by $\mathcal{S}_{j\tilde{k}\tilde{k}}(R)$ implies that

$$u_{R(\tilde{k})} \cdot p_{R(\tilde{k})} + (1 - p_{R(\tilde{k})}) \cdot U(R(\tilde{k}+1, K)) \geq u_j \cdot p_j + (1 - p_j) \cdot U(R(\tilde{k}+1, K))$$

which contradicts that j dominates $\tilde{k}$, concluding our proof. $\square$

Combining the four previous lemmas we obtain that the set of constraints implied by the set of $\mathcal{S}_o(R)$ can be omitted, where

$$\mathcal{S}_o(R) = \left(\bigcup_{k=1}^{K} \bigcup_{l=1}^{\bar{k}-1} \mathcal{S}_{jkl}(R) \right) \cup \left(\bigcup_{k=1}^{K} \bigcup_{l \in K'} \mathcal{S}_{jkl}(R) \right) \cup \left(\bigcup_{l \in \underline{K}} \bigcup_{k=1}^{\tilde{k}} \mathcal{S}_{jkl}(R) \right) \cup \left(\bigcup_{l \in \bar{K}} \bigcup_{k=1}^{\tilde{k}} \mathcal{S}_{jkl}(R) \right)$$

Hence, we have that

$$U(R) \geq U(R'),\ \forall R' \in \mathcal{S}(R) \setminus \mathcal{S}_o(R)$$

implies that

$$U(R) \geq U(R'),\ \forall R' \in \mathcal{S}(R).$$

Finally, we observe that $\mathcal{S}_u(R) = \mathcal{S}(R) \setminus \mathcal{S}_o(R)$ can be written as

$$\mathcal{S}_u(R) = \left(\bigcup_{\substack{l=\bar{k} \text{ st.} \\ p_{R(l)} < p_{R(\tilde{k})}}}^{\tilde{k}} \mathcal{S}_{jll}(R) \cup \left(\bigcup_{k=\tilde{k}+1}^{K} \mathcal{S}_{jkl} \right) \right) \cup \left(\bigcup_{k=\tilde{k}+1}^{K} \bigcup_{l=\tilde{k}+1}^{K} \mathcal{S}_{jkl}(R) \right)$$

where the index k represents the position where the new program j is placed, and the index l represents the program $R(l)$ that leaves the ROL. If we write $\tilde{k} = 0$ if $p_{R(k)} > p_j$ for all $k \in \{1, \ldots, K\}$, then this expression also works.

B Multivariate Gibbs Sampler

Consider the following specification for students' preferences:

$$u_{ij} = Z_{ij}\beta - d_{ij} + \varepsilon_{ij}, \tag{B.1}$$

where $Z_{ij} = [z_{ij1}, ..., z_{ijK}]$ is a $1 \times K$ row vector of covariates. The system can be stacked in order to represent the vector of utilities u_i as:

$$u_i = Z_i\beta - d_i + \varepsilon_i \tag{B.2}$$

where Z_i is an $M \times K$ matrix of covariates, d_i an $M \times 1$ vector of distances and ε_i is an $M \times 1$ vector of shocks. Consider also the following independent priors for β and Σ:

$$\beta \sim N(\bar{\beta}, A^{-1}) \tag{B.3}$$

$$\Sigma \sim IW(\nu_0, V_0) \tag{B.4}$$

Step 0 Start with initial values Σ^0 and $u^0 = \{u_i^0\}_{i=0}^N$ such that $u_i^0 \in C(R_i) \quad \forall i = 1, ..N$, i.e, select u_i^0 to be a solution to the following problem:

$$A_i u_i \geq \varepsilon \tag{B.5}$$

with ε a small positive number.

Step 1 Draw $\beta^1|u^0, \Sigma^0$ from a $N(\tilde{\beta}, V)$, where

$$V = \left(Z^{*'}Z^* + A\right)^{-1}, \quad \tilde{\beta} = V\left(Z^{*'}u^* + A\bar{\beta}\right) \tag{B.6}$$

$$Z^* = \begin{bmatrix} Z_1^* \\ ... \\ Z_N^* \end{bmatrix} \tag{B.7}$$

$$Z_i^{*'} = C'Z_i, \quad u_i^* = C'u_i^0 \tag{B.8}$$

$$\left(\Sigma^0\right)^{-1} = C'C \tag{B.9}$$

Where C comes from the Cholesky decomposition of $\left(\Sigma^0\right)^{-1}$

Step 2 Draw $\Sigma^1|u^0, \beta^1$ from an $IW(\nu_0 + N, V_0 + S)$

$$S = \sum_{i=1}^{N} \varepsilon_i \varepsilon_i' \tag{B.10}$$

$$\varepsilon_i = u_i^0 - Z_i\beta^1 + D_i \tag{B.11}$$

Step 3 Iterate over students and schools, drawing $u_i^1|\beta^1, \Sigma^1, R_i$. For each school $j = 1, ..M$, draw:

$$u_{ij}^1|\{u_{ik}^1\}_{k=1}^{j-1}, \{u_{ik}^0\}_{k=j+1}^{J}, \beta^1, \Sigma^1 \tag{B.12}$$

from a truncated normal $TN(\mu_{ij}, \sigma_{ij}^2, a_{ij}, b_{ij})$, where

$$\mu_{ij} = \sum_{k=1}^{K} \beta_{jk}^1 z_{ijk} - d_{ij} \tag{B.13}$$

$$\sigma_{ij}^2 = \Sigma_{jj}^1 - \Sigma_{j(-j)}^1 \left[\Sigma_{(-j)(-j)}^1\right]^{-1} \Sigma_{(-j)j}^1 \tag{B.14}$$

The truncation points a_{ij} and b_{ij} must ensure the draw u_{ij}^1 lies in the interior of $C(R_i)$ given the previous draws, so they are the solutions to the following optimization problems:

$$\begin{aligned} a_{ij} = \max_{u_{ij}} \quad & u_{ij} \\ st. \quad & Au \geq 0 \\ & u_{ik} = u_{ik}^1 \quad \forall k = 1, ..j-1 \\ & u_{ik} = u_{ik}^0 \quad \forall k = j+1, ..M \end{aligned}$$

$$\begin{aligned} b_{ij} = \min_{u_{ij}} \quad & u_{ij} \\ st. \quad & Au \geq 0 \\ & u_{ik} = u_{ik}^1 \quad \forall k = 1, ..j-1 \\ & u_{ik} = u_{ik}^0 \quad \forall k = j+1, ..M \end{aligned}$$

We implement all of these linear problems using Gurobi.

Step 4 Set $\Sigma^0 = \Sigma^1$ and $u^0 = u^1$ and repeat steps 1-3 to obtain a sequence (β^k, Σ^k).

C Random Coefficients Gibbs Sampler

C.1 Model

Consider the following specification for students' preferences:

$$u_{ij} = \alpha_{im(j)} + Z_{ij}\beta + \varepsilon_{ij} \tag{C.1}$$

where $\varepsilon_{ij} \sim N\left(0, \sigma^2\right)$, $Z_{ij} = [z_{ij1}, ..., z_{ijK}]$ is a $1 \times K$ row vector of covariates and $\alpha_{im(j)}$ are major random coefficients. The system can be stacked in order to represent the vector of utilities u_i as:

$$u_i = I_m \alpha_i + Z_i \beta + \varepsilon_i \tag{C.2}$$

where Z_i is an $M \times K$ matrix of covariates, ε_i is a $M \times 1$ vector of shocks, α_i is an $|B| \times 1$ vector of major random coefficients such that

$$\alpha_i \equiv \begin{pmatrix} \alpha_{i1} \\ ... \\ \alpha_{i|B|} \end{pmatrix}, \tag{C.3}$$

and I_m is a $M \times |B|$ matrix such that for each row r and column c we have that

$$I_m[r,c] = \begin{cases} 1 & \text{if } c = m(r) \\ 0 & \text{o.w} \end{cases} \tag{C.4}$$

We model the random coefficients as a multivariate regression on a set of demographic variables z_i:

$$\alpha_i = \Delta z_i + \nu_i \sim \text{ iid } N(0, V_\alpha), \quad i = 1, ..., N \tag{C.5}$$

where Δ is a $|B| \times d$ matrix of coefficients, z_i is a $d \times 1$ vector of student's observable characteristics and ν_i is a $|B| \times 1$ student's specific shock.

C.2 Priors

Consider the following priors:

$$\beta \sim N(\bar{\beta}, A^{-1}) \tag{C.6}$$

$$V_\alpha \sim IW(\nu_{bo}, V_{bo}) \tag{C.7}$$

$$\delta = \text{vec}(\Delta) \sim N(\bar{d}, (V_\alpha \otimes A_d^{-1})) \tag{C.8}$$

C.3 Sampler

Step 0 Start with initial values for $\alpha^0 = \{\alpha_i^0\}_{i=0}^N$, Δ^0, V_α^0 and initial values for $u^0 = \{u_i^0\}_{i=0}^N$ such that $u_i^0 \in C(R_i) \quad \forall i = 1, ..N$, i.e, select u_i^0 to be a solution to the following problem:

$$A_i u_i \geq \epsilon \tag{C.9}$$

with ϵ a small positive number.

Step 1 Draw $\beta^1 | \alpha^0, u^0, \sigma^2$ from a $N(\tilde{\beta}, V)$, where

$$V = \left(Z^{*\prime} Z^* + A\right)^{-1}, \quad \tilde{\beta} = V\left(Z^{*\prime} u^* + A\bar{\beta}\right) \tag{C.10}$$

$$Z^* = \begin{bmatrix} Z_1^* \\ ... \\ Z_N^* \end{bmatrix} \tag{C.11}$$

$$Z_i^{*'} = C' Z_i, \quad u_i^* = C' \left(u_i - I_m \alpha_i\right) \tag{C.12}$$

$$\left(\sigma^2 \mathbb{I}\right)^{-1} = C'C \tag{C.13}$$

Where C comes from the Cholesky decomposition of the inverse of the variance covariance matrix of ε_i.

Step 2 Draw $\alpha_i^1 | \beta^1, u_i^0, z_i, \Delta^0, V_\alpha^0, \sigma^2$ from

$$\alpha_i \sim N\left(\bar{b}, \left(I_m^{*'} I_m^* + V_\alpha^{-1}\right)^{-1}\right) \tag{C.14}$$

where

$$\bar{b} = \left(I_m^{*'} I_m^* + V_\alpha^{-1}\right)^{-1} \left[I_m^{*'} I_m^* \hat{\alpha}_i + V_\alpha^{-1} \bar{\alpha}_i\right] \tag{C.15}$$

$$\bar{\alpha}_i = \Delta z_i \tag{C.16}$$

$$\hat{\alpha}_i = \left(I_m^{*'} I_m^*\right)^{-1} I_m^{*'} \tilde{u}_i \tag{C.17}$$

$$I_m^* = C' I_m \tag{C.18}$$

$$\tilde{u}_i = C' \left(u_i - Z\beta\right) \tag{C.19}$$

Step 3 Iterate over students and schools, drawing $u_i^1 | \beta^1, \sigma^2, \alpha^1, R_i$. For each school $j = 1, ..M$, draw:

$$u_{ij}^1 | \{u_{ik}^1\}_{k=1}^{j-1}, \{u_{ik}^0\}_{k=j+1}^{J}, \beta^1, \alpha_i^1, \sigma^2 \tag{C.20}$$

from a truncated normal $TN(\mu_{ij}, \sigma_{ij}^2, a_{ij}, b_{ij})$, where

$$\mu_{ij} = \sum_{k=1}^{K} \beta_{jk} z_{ijk} + \alpha_{im(j)} \tag{C.21}$$

$$\sigma_{ij}^2 = \sigma^2 \tag{C.22}$$

For simplicity we omit index i, as this problem must be solved for each student independently, then the truncation points can be computed by

$$a_j = \max_{k \in \left\{k : A_{kj} > 0\right\}} \frac{-A_k^{-j} u^{-j}}{A_{kj}}$$

$$b_j = \min_{k \in \left\{k : A_{kj} < 0\right\}} \frac{-A_k^{-j} u^{-j}}{A_{kj}}$$

where A_k^{-j} is matrix A^{-j}'s k-th row and A_{kj} is the k-th element of column A_j.

Step 4 Draw $\Delta^1|\alpha^1, \bar{\Delta}, A_d, V_\alpha^0$ from

$$\delta = \text{vec}(\Delta) \sim N\left(\tilde{d}, V_\alpha^{-1} \otimes \left(Z_\alpha' Z_\alpha + A_d\right)^{-1}\right) \tag{C.23}$$

$$\tilde{d} = \text{vec}(\tilde{D}), \tilde{D} = \left(Z_\alpha' Z_\alpha + A_d\right)^{-1}\left(Z_\alpha' Z_\alpha \hat{D} + A_d \bar{D}\right) \tag{C.24}$$

$$\hat{D} = \left(Z_\alpha' Z_\alpha\right)^{-1} Z_\alpha' G \tag{C.25}$$

G is an $N \times |B|$ matrix with each α_i' as a row

Z_α is an $N \times d$ matrix with each z_i' s a row

$\bar{D} = \text{stack}(\bar{d})$ is a $d \times B$ matrix formed column by column from the elements of $\bar{d}$.

Step 5 Draw $V_\alpha^1|\alpha^1, \Delta^1, \nu_{ob}, V_{ob}$ from

$$V_\alpha \sim IW(\nu_{bo} + N, V_{bo} + S) \tag{C.26}$$

where

$$S = \sum_i (\alpha_i - \bar{\alpha}_i)(\alpha_i - \bar{\alpha}_i)', \bar{\alpha}_i = \Delta z_i \tag{C.27}$$

Step 6 Set $u^0 = u^1$, $\alpha^0 = \alpha^1$, $\Delta^0 = \Delta^1$, $V_\alpha^0 = V_\alpha^1$ and repeat steps 1-5 to obtain a sequence β^k, α^k, Δ^k and V_α^k .