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*

March 18, 2019

Abstract

We analyze the application process in the Chilean College Admissions problem. Students can submit up to 10 preferences, but most of the 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 if their admission probabilities are too low. To rationalize this behavior, 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 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 are selective programs and overstates the value of being unassigned and 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 estimation results strongly suggest that “short-list” students should not be interpret as truth-tellers, even in a seemingly strategy-proof environment.

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 specially 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, New York City, 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 the 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 expost 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. For instance, this is the case when the mechanism used is the Deferred-Acceptance algorithm (Gale and Shapley (1962)) and there are no constraints on the length of the ROLs, or when students submit a list that is shorter than the maximum allowed and thus the constraints are 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 students find costly to search for their desired programs (or list them in order) and their believed admission chances are too low. In both cases, assuming that "short-list" students report truthfully would result in biased estimates for preferences. 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, a different 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 preferences of students 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 score. 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% submits 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 very 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 don't 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 theoretically 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 closests papers to ours are Ajayi and Sidibe (2017) and Artemov et al. (2017). Our paper complements them 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 important 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 they make strategic mistakes. However, as it doesn't 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, specially in counterfactual scenarios that involve strategic considerations.

We depart from assuming stability and we model the students' optimal portfolio problem absent of strategic mistakes. We assume students are expected utility maximizers and estimate students' beliefs on admission probabilities and preferences over programs using the assignment mechanism and the information contained in the reported ROLs, as opposed to using enrollment choices or assignment information. In this way, we are able to better understand why students misreport their true preferences even when there are no clear strategic incentives to do so.

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 and 7 we describe the simulations and results. Finally, Section 8 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 the specific major they of their choice instead of going to college first. Thus, from now on we refer to the pair major/university as program. 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 them 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 to the admission requirements, each program announces every year a set of admission weights. Application scores are constructed by the weighted average of students scores and admission weights, thus, 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.

[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 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).

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 rejected him. As a side outcome of this algorithm is obtained a set of cutoffs $\{P_j\}_{j\in M}$, where P_j is the minimum application score among students matched to program j and M is the set of programs. 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

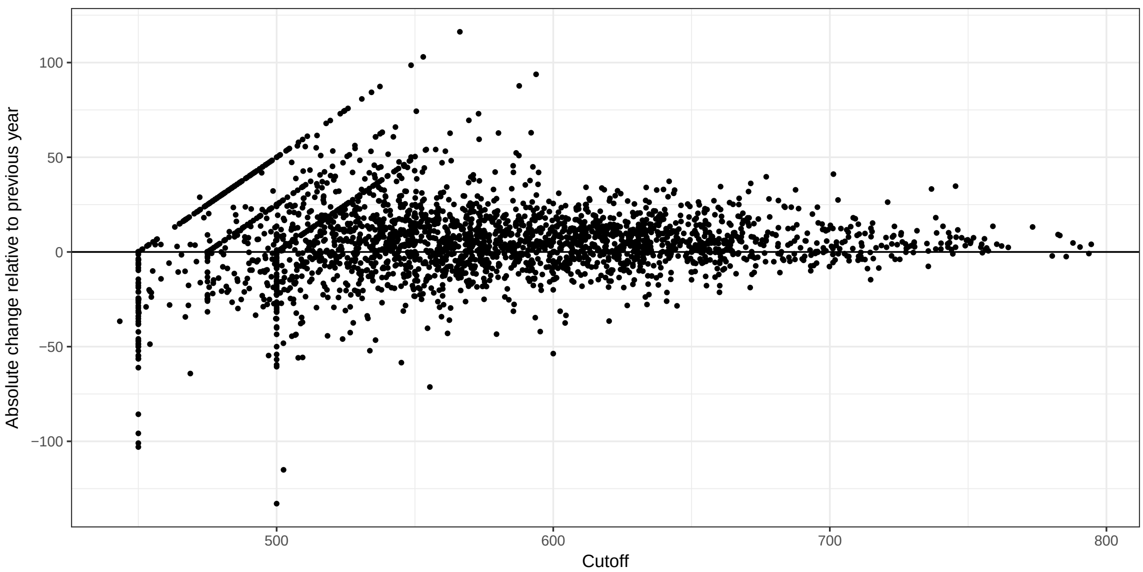
$$i \text{ is assigned to } j \in R_i \Leftrightarrow s_{ij} \geq P_j \text{ and } s_{ij'} < P_{j'}\, \forall j' \in R_i \; st. \; j' \succ_{R_i} j,$$

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 of the mechanism 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

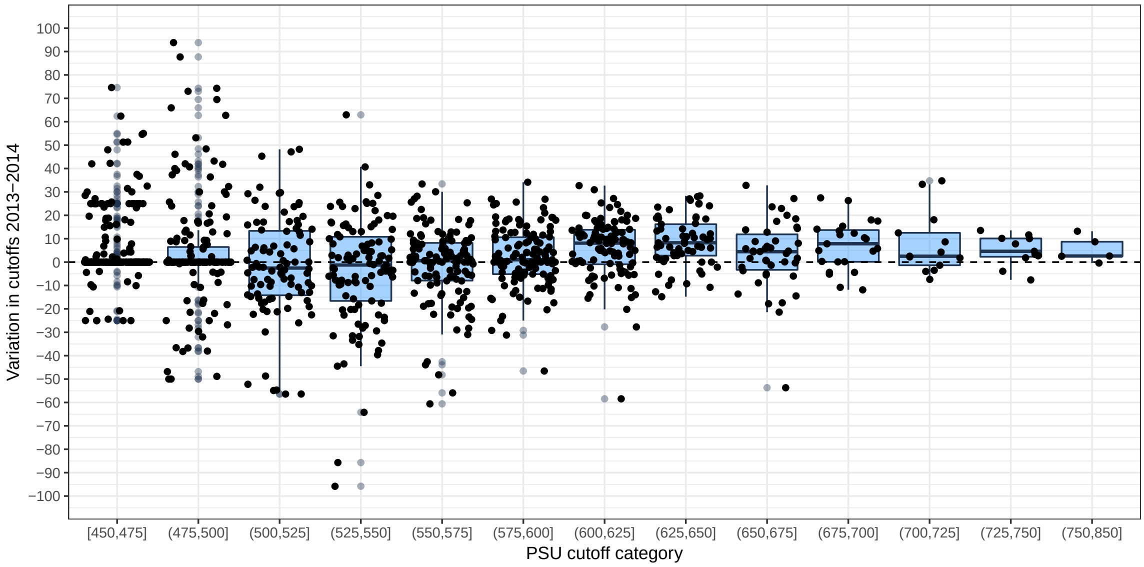
As the Chilean mechanism can be represented by a cutoff structure, the uncertainty that students face on their admission probabilities can be summarized by the variation in cutoffs from year to year. Figures 2.1 and 2.2 show the variation in cutoffs between 2013 and 2014. Each dot shows the variation of a program's cutoff in PSU points with respect to its cutoff level 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 the cutoff does not vary between 2013 and 2014.

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

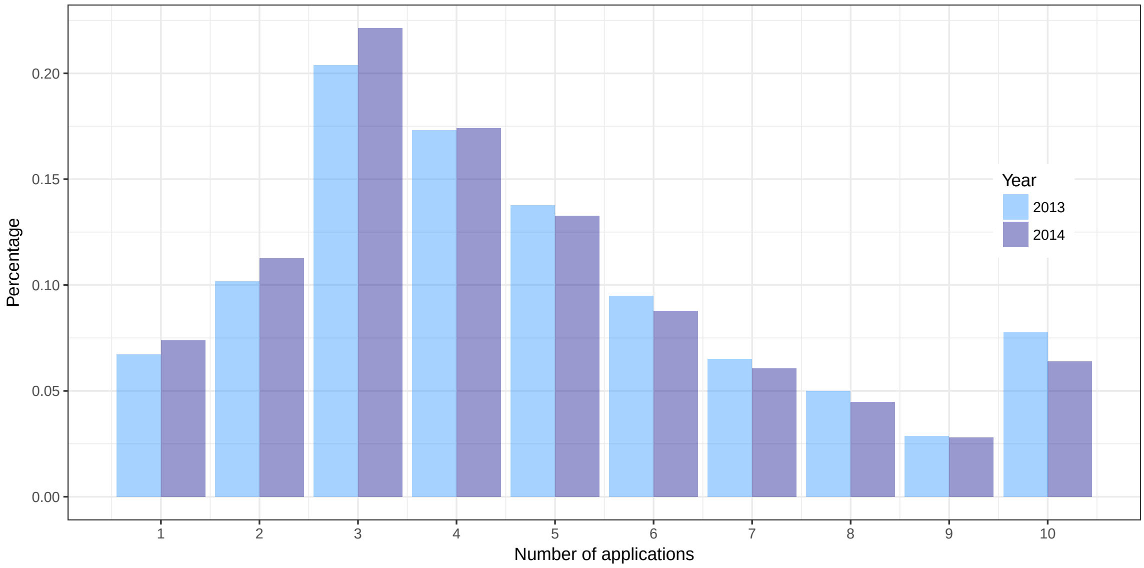
Given that students face some uncertainty on their admission, a natural question is whether they take this into account or just submit their true preference order without considering their admission probabilities.

If there where no restrictions on the length of the list, rational students wouldn't 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 just weakly-optimal, this rationale doesn't 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 in the Chilean College Admissions' Problem this is indeed the case.

Figure 2.3 shows the distribution of the number of programs considered in the ROL among students who apply to at least one program in the system. 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).

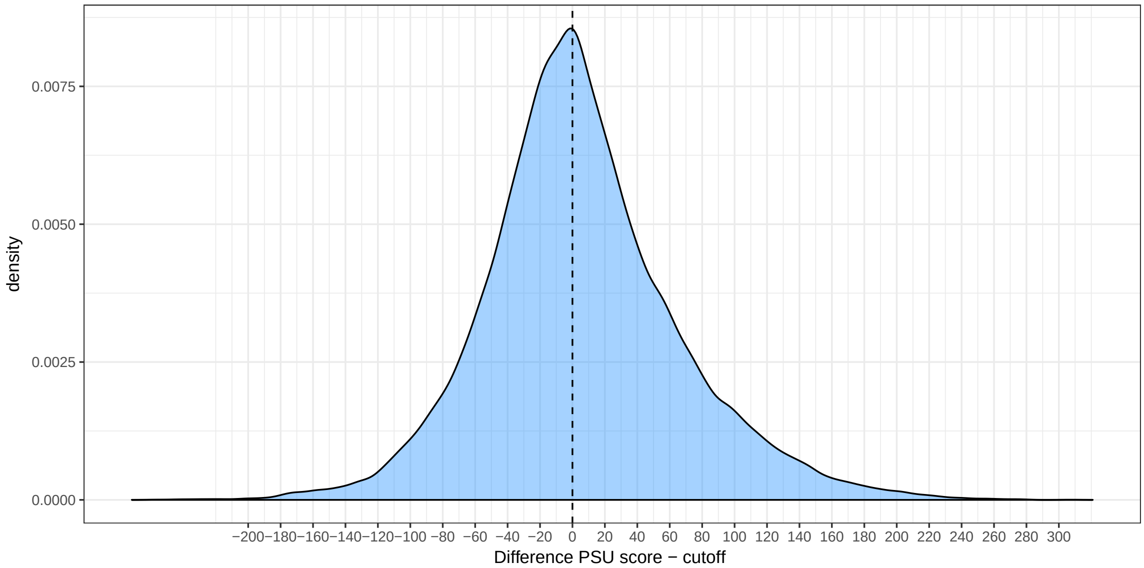
Figure 2.3: Number of applications per student by year



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 apply in first preference to programs for which their scores are close to the cutoff of the current year. Figure 2.4 shows 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 apply in first preference to programs for which their weighted score is around the cutoff.

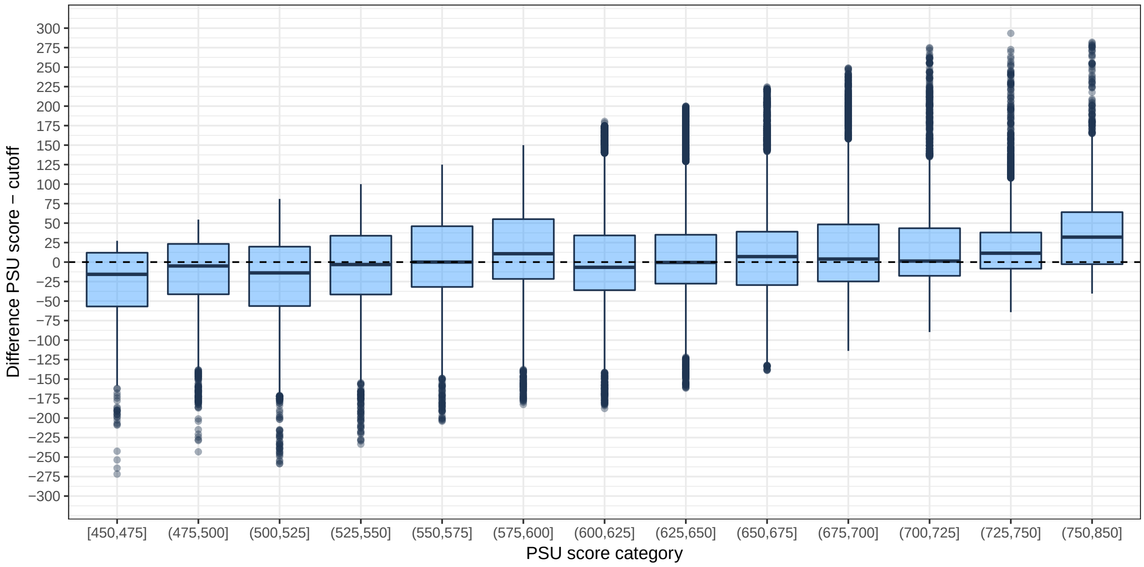
Figure 2.4: Distribution of difference between PSU score and cutoff for first listed preference in 2014



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 persistent after controlling by 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 applied to 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.

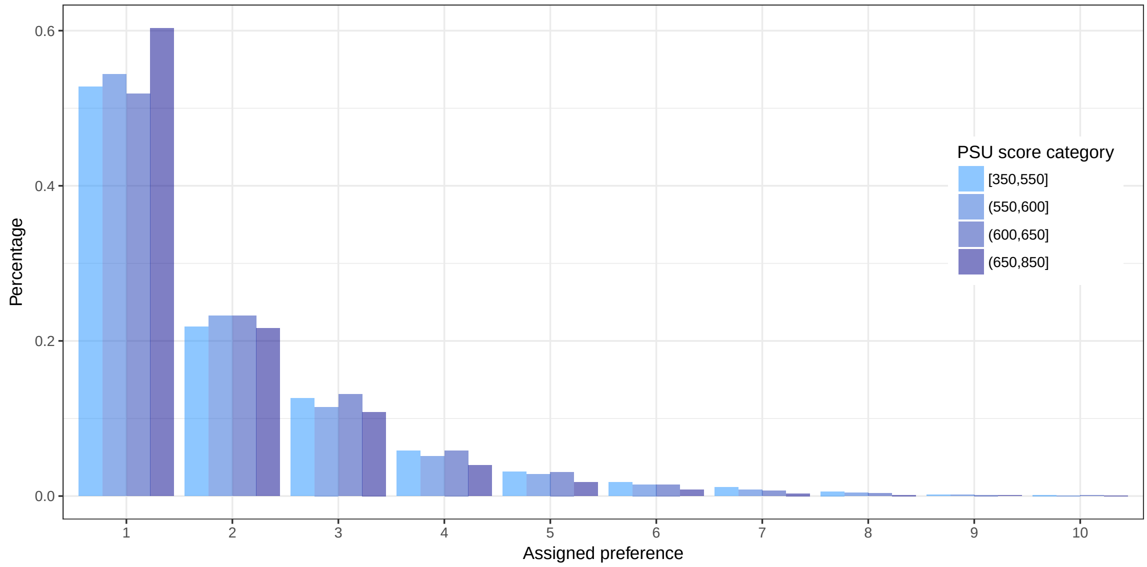
Figure 2.5: Boxplots of difference between PSU score and cutoff for first preference in 2014



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 that result 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.

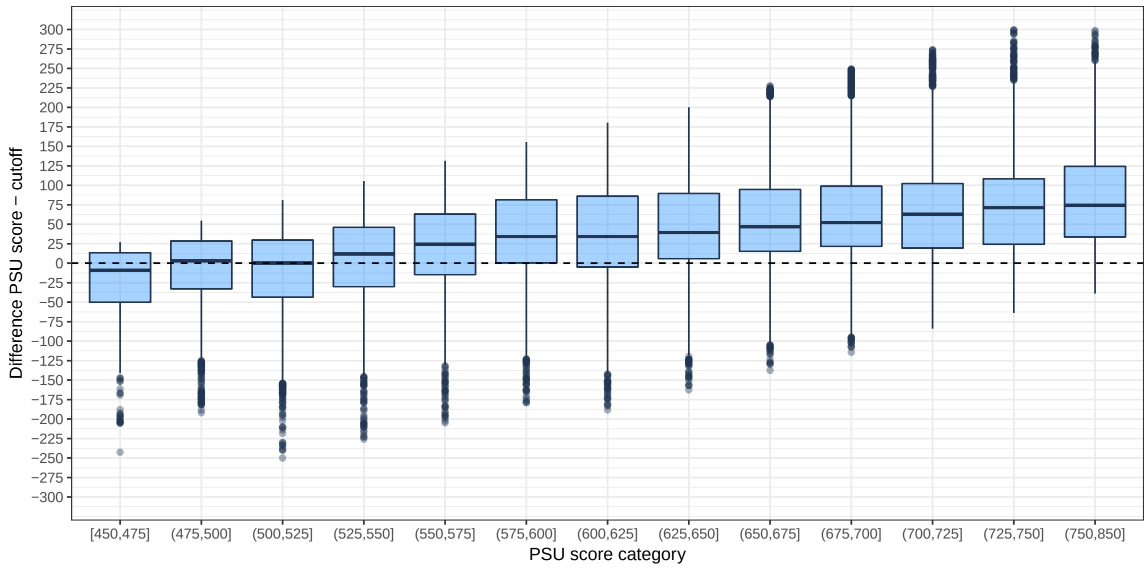
Figure 2.6: Preference of assignment by PSU score category in 2014



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, with this information, we can not disentangle how much of this pattern is driven by preference heterogeneity and how much is driven by beliefs over admission probabilities. Even if students do not take into account their admission probabilities, it could be the case that they prefer programs for which their weighted scores are not too far from the cutoff. For instance, the cutoff could be a signal of the difficulty and students could prefer to avoid more difficult programs. However, if this is the main force driving this application pattern, this would imply that preferences for programs in the Chilean market are extremely heterogeneous. To show that the main force driving this selection pattern are beliefs in admission probabilities, we will use a survey conducted in 2014 to all participants in the centralized system.

2.3.1 Survey

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 (200,000) between October and November, i.e, before the standardized national exam. Hence the students (around 40,000) didn't know their scores when answering it.

In the survey, students were asked the following question:

"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 restrictions imposed by DEMRE, the survey design specified the major of the program but not the name of the university. Therefore, we only know whether a student wants to study a specific major, but not at which university. This complicates the analysis because there exists a large heterogeneity in program's selectivity across universities. To avoid this problem we focus on students who reported Medicine to be their top preference in the survey. The reason to focus in 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 in their first preference, and 2,987 of these students finally applied to the system. Among these, only 1,360 submitted a list with Medicine in 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 in their first 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.

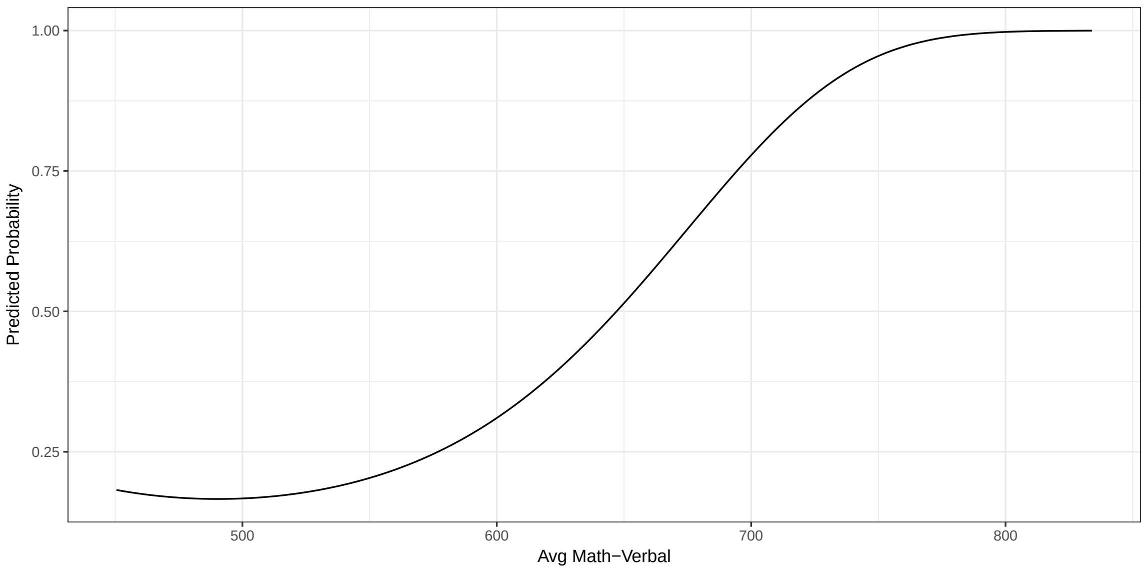
Table 2.1: Probability of applying to Medicine in first preference, conditional on reporting Medicine in first preference in the survey

	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 are taking 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 skipping pattern is also described in Artemov et al. (2017). This result is true even for students who do not face clear strategic incentives given by restrictions in the length of the list (87% of the students who declared Medicine as their most preferred program in the survey and submitted an application are “short-list” students). This implies that when indifferent between doing so or not, students do tend to left-censor their application lists relative to their unconstrained preference order. 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 there is no monetary application cost in our setting, Assumption 1 could be micro-funded by including an information acquisition cost due to search frictions. However, even if the application cost is 0, Assumption 1 will cover at least two scenarios:

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, assuming that short-list students are truth tellers would be misleading to understand student 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:

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

The application score of the last admitted student to program j is the cutoff of program j and will be denoted by P_j.

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:

$$u_{ij} = u(Z_{ij}, d_{ij}, \delta_i^{m_j}, \gamma_i^{k_j}, \xi_j) + \varepsilon_{ij}, \tag{3.1}$$

where Z_{ij} are observable characteristics of student i and program j, d_{ij} is the distance between program j and students' i home, $m(j)$ is the major of the program, $k(j)$ its university 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. We include the unobserved random coefficients $\delta_i^{m(j)}$ and $\gamma_i^{k(j)}$ to capture idiosyncratic tastes for majors and universities.

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[5], 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 an institution outside, 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[6].

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.

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 students can infer the distribution of cutoffs. We will further assume that students take these distributions to be independent across programs. These considerations lead to 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) argue that a consistent estimator of these beliefs can be obtained using the following bootstrap procedure:

- For each bootstrap simulation $b = 1, \ldots, B$,
  - 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

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

[5]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.

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

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)[7]. In this problem, a student must choose once and for all 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 $U : 2^M \to \mathbb{R}$ 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 utilities $\{u_j\}_{j\in M}$.[8] 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.

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

Chade and Smith (2006) show that the portfolio problem is NP-Hard. However, when admission probabilities are independent[9] 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 $S_n = S_{n-1} \cup j_n$

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

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

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

MIA adds recursively programs that give the highest marginal improvement to the portfolio, as long as it exceeds 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 $U(S_{n-1} \cup j_n) - U(S_{n-1}) \leq c\left(S_{n-1} \cup j_n\right) - c\left(S_{n-1}\right) = 0$, 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.

3.3.1 One-Shot Swap optimality

Given the structure of the optimal solution, a question of interest is what can be inferred about preferences by looking at a submitted ROL. 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, an 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 $\mathcal{S}(R)$ 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. 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.[10]

[10] 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'.

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

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

In Proposition 1 we show that it is sufficient to consider constraints involving one-shot swaps from R in order to ensure ROL R is optimal.

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

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

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.

4 Data

We use data on the Chilean College Admissions problem for the years 2012 to 2014. The main data source comes from DEMRE, and includes the scores of students, admission weights for each program, restrictions for applicants and the final assignment. In addition, we have socioeconomic characteristics of students, like self reported family income, parents education, municipality where the student lives, among others[11].

We complement the previous data with data from MIFUTURO, which consists on information about the characteristics of universities and programs inside and outside the centralized system, including tuition, duration, major and the location of the program. 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.

[11] We also have available enrollment data for students inside and outside the system, but we are not using it right now.

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).
[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). Hence, we can obtain non-parametric identification by using two sources of variation:

1. Including a "Special regressor"
2. Using variation in the choice environment

The "special regressor" has to be a covariate that varies in i and j (between students and programs), enters additively into the utility function, and is orthogonal to unobservables. This regressor works as a preference shifter, allowing the researcher to identify preferences if there is no residential sorting on 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' home and the location of the program they are applying will also help to identify preferences. However, given that we just observe the municipality of students' home 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 characteristic of the Chilean system: each year programs declare the weights they will assign to each admission factor. As explained before, in order 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 are also given a score obtained from the average grade along high-school (*Notas de Enseñanza Media* or NEM) and, from year 2013 onward, a second score that depends on the relative position of the student among his/her cohort (*Ranking de Notas* or Rank). Each program defines the weights assigned to each score, the vacancies offered and a set of specific requirements that must be met by applicants to be acceptable, such as minimum application score or minimum tests scores. This information is publicly available for students before they take the standardized national exam.

We observe a large cross-sectional variation in the weights declared by programs: weights vary by program within a given university, but also across universities. Even though weights are quite stable from year to year, we exploit an important time variation between the years 2012 and 2014: the inclusion of the rank score. In order to evaluate the effects of this new factor CRUCH established that for the admission process of 2013 every program in the system had to assign a weight of 10% to the rank score. 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 generated a 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, grouping 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 GPA of their programs. As the rank score is constructed as a function of GPA[12] and both admission scores are highly correlated, we shouldn't expect a big variation for 2012 to 2013 on the admission probabilities. his should mainly affect students with high GPA.

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] Absolute difference in admission weights.

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

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

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] Absolute difference in admission weights.

We assume that the variation in admission weights does not affect the overall distribution of indirect utilities between those years. Here we are ruling out, for example, the possibility that students could 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). It's worth noticing that as the change we are exploiting is in the admission weights and not in the scores, we are able to include scores in the specification of preferences. 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 of the submitted ROLs between those years.

Formally, following Agarwal and Somaini (2018), let $C(R|t)$ be the set of indirect utility vectors u that rationalizes ROL R to be the optimal ROL in choice environment t, conditional on observables. Then variation in the choice environment identifies the distribution $f_u(u)$:

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

Changes in the choice environment, particularly changes in the 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 (C_R in 2013 and C_R in 2014). If we have enough variation we will trace out the conditional distribution of indirect utilities and point identify the parameters.

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's simple and includes a covariate that exhibits variation at the student and also at the program level. Thus, we don't take a stand on whether this specification is a good approximation of students' preferences or not.

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.

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 of 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. To account for this we will include in the final estimation a mixture of students: truth-tellers and strategic (A1 and A2 hold). For the current exercises we drop these students from the sample.

Even though we have selected a sample of programs and students for performing these simulations, the proposed estimation methods are intended to be robust and computationally feasible for the whole sample of programs and students.

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 assign an admission probability of 1. Therefore, under DGP2 students won't include programs that are bellow the outside option, that is, if it's 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[13] with mean 0 and variance covariance matrix $\sigma^2\mathbb{I}$ and normalize $\sigma = 1$[14]. 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 an 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:

Likelihood Assuming DGP1:

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

where

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

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 won't include programs for which their assignment probabilities are 0 and (ii) students won't 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

[13]We change the distribution of the shock for this DGP to be consistent with our proposed estimation method.
[14]Notice that we are restricting the variance covariance matrix, thus the model under this specification is over identified.

probability of admission and being more preferred than the outside option. The implies that the likelihood under DGP2 can be written as:

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

where

$$\tilde{J}_i\equiv\{j\in J:p_{ij}>0\} \tag{6.5}$$

and

$$V_i\left(R_i(|R|)\right)\equiv\begin{cases}1 & \text{if} \quad p_{iR_i(|R|)}=1\\ \frac{1}{\sum\limits_{j\in\tilde{J}_i\backslash\{R_i(1),...,R_i(|R|)\}}\exp(u_{ij})} & \text{o.w}\end{cases} \tag{6.6}$$

Notice that students will only consider programs for which their believed admission probabilities are strictly positive. Hence we can't 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, the portfolio problem won't be as simple as under DGP2. Now 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 it's 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 given by evaluating all possible ROLs.

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 can not 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:

$$u_i \in C(R_i) \iff A_i u_i \geq 0 \tag{6.7}$$

Case 1: $|R_i| < K$

As we have shown before, if the $|R_i| < K$, 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.:

$$u_{ij} \geq u_{ij'} \quad \forall j \in R_i, j' \notin R_i, st : p_{ij'} > 0. \tag{6.8}$$

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

$$u_{ij} \geq u_{i0} \quad \forall j \in R_i. \tag{6.9}$$

Finally, the marginal benefit of including any other program in the portfolio must be lees than or equal to 0. As we haven't 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:

$$\text{if } p_{iR_i(|R_i|)} < 1 \Rightarrow u_{i0} \geq u_{ij'} \quad \forall j' \notin R_i, st : p_{ij'} > 0 \tag{6.10}$$

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:

$$u_{iR_i(1)} \geq u_{iR_i(2)} \geq \ldots \geq u_{iR_i(|R_i|)} \tag{6.11}$$

With all these inequalities we can write down the matrix A_i for any student such that $|R_i| < K$. Inequalities given in Equation 6.8 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 6.9 and 6.10 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 6.11 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 6.8 and 6.10 won't 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. Now, following Agarwal and Somaini (2018), we can express these inequalities in the following way:

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

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 6.9 and 6.11.

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:

$$u_{ij} = Z_{ij}\beta + \varepsilon_{ij} \tag{6.12}$$

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:

$$u_i = Z_i\beta + \varepsilon_i \tag{6.13}$$

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 :

$$\beta \sim N(\bar{\beta}, A^{-1}) \tag{6.14}$$

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:

$$A_i u_i \geq \epsilon \tag{6.15}$$

with ϵ a small positive number.

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

$$V = \left(Z^{*'} Z^* + A\right)^{-1}, \quad \tilde{\beta} = V\left(Z^{*'} u^* + A\bar{\beta}\right) \tag{6.16}$$

$$Z^* = \begin{bmatrix} Z_1^* \\ \dots \\ Z_N^* \end{bmatrix} \tag{6.17}$$

$$Z_i^{*'} = C' Z_i, \quad u_i^* = C' u_i^0 \tag{6.18}$$

$$\left(\sigma^2 \mathbb{I}\right)^{-1} = C'C \tag{6.19}$$

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:

$$u_{ij}^1|\{u_{ik}^1\}_{k=1}^{j-1}, \{u_{ik}^0\}_{k=j+1}^{J}, \beta^1, \sigma^2 \tag{6.20}$$

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

$$\mu_{ij} = \sum_{k=1}^{K} \beta_{jk}^1 z_{ijk} \tag{6.21}$$

$$\sigma_{ij}^2 = \sigma^2 \tag{6.22}$$

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. The gain of doing this is that the term $A^{-j}u^{-j}$ is fixed and known, so we can manipulate this expression to isolate u_j and obtain conditions that it must satisfy to obtain a vector of utilities that rationalizes the observed ROL. After some manipulation we obtain that

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

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

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

Notice that we do not need to necessarily 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

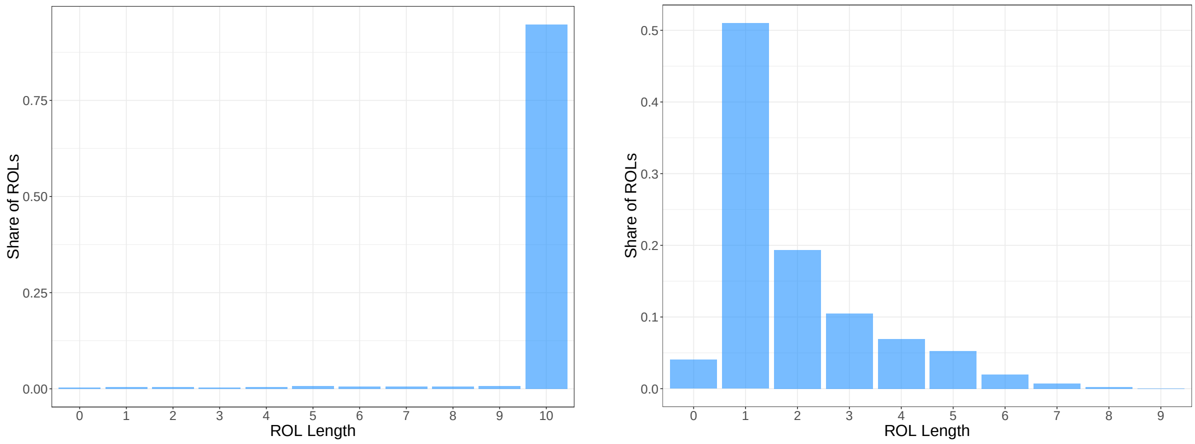
$$\beta = (1, -1, 2, 1.5, -1). \tag{7.1}$$

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



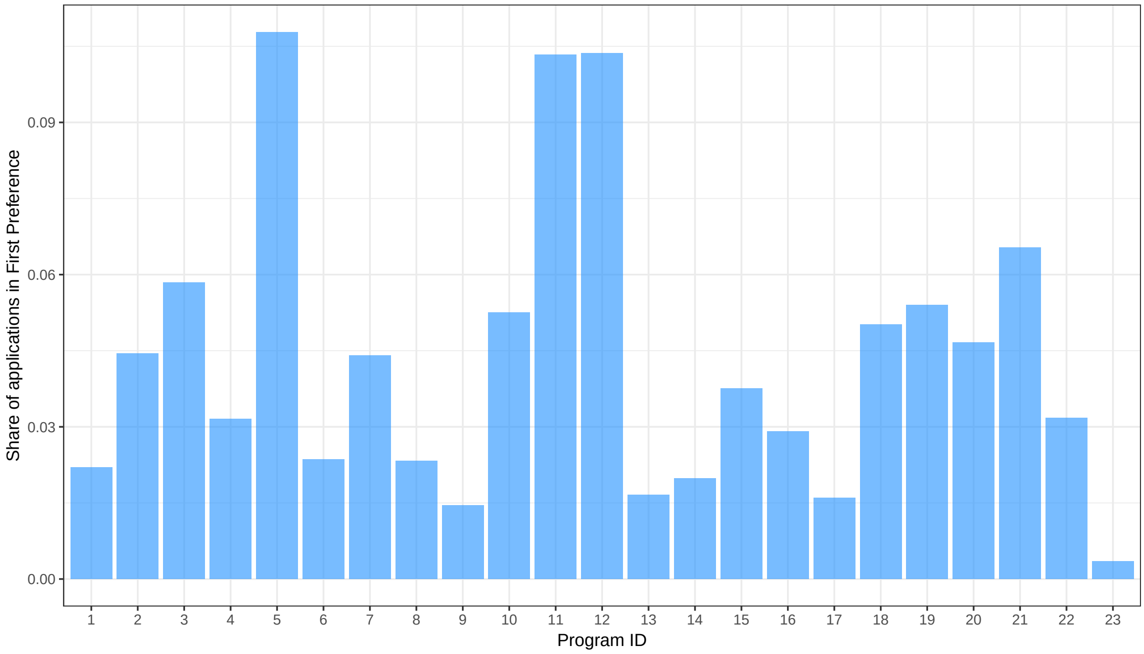
Notes: Distribution of the Length of ROLs under DGP1 (left figure) and DGP2 (right figure).

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.

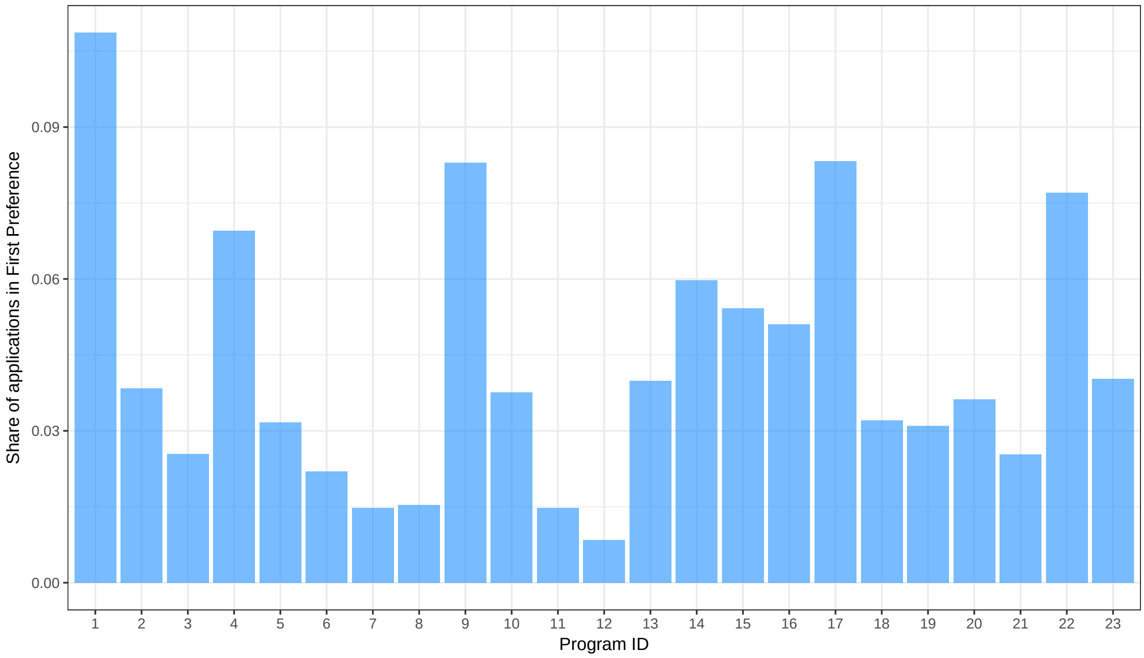
We also observe a fair amount of heterogeneity in first choices, even under DGP1. This is mainly explained by the preference shock, but also because of the heterogeneity induced by the distance covariate.

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



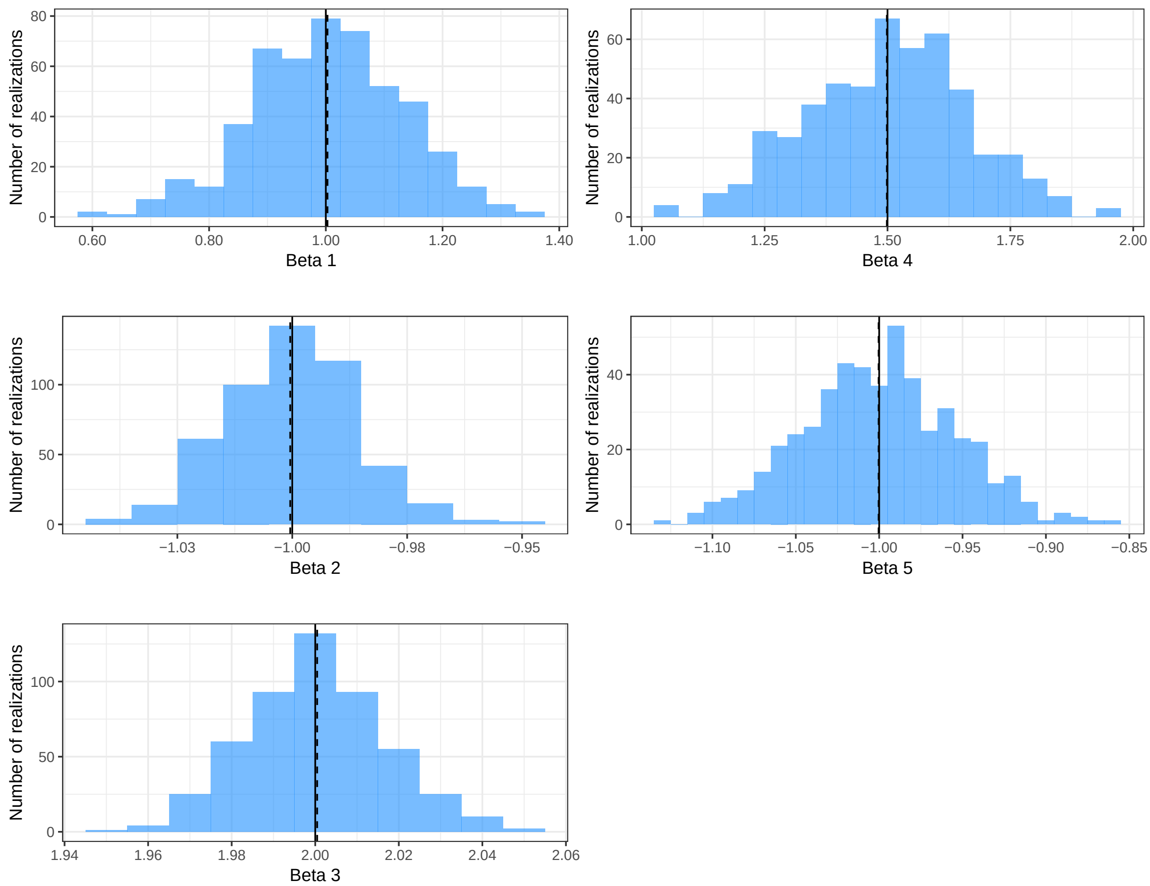
Notes: Share of applications in first preference to each program under DGP2. Program 23 is the outside option of being unassigned.

7.2 Monte Carlo Simulations

7.2.1 DGP1 vs DGP2

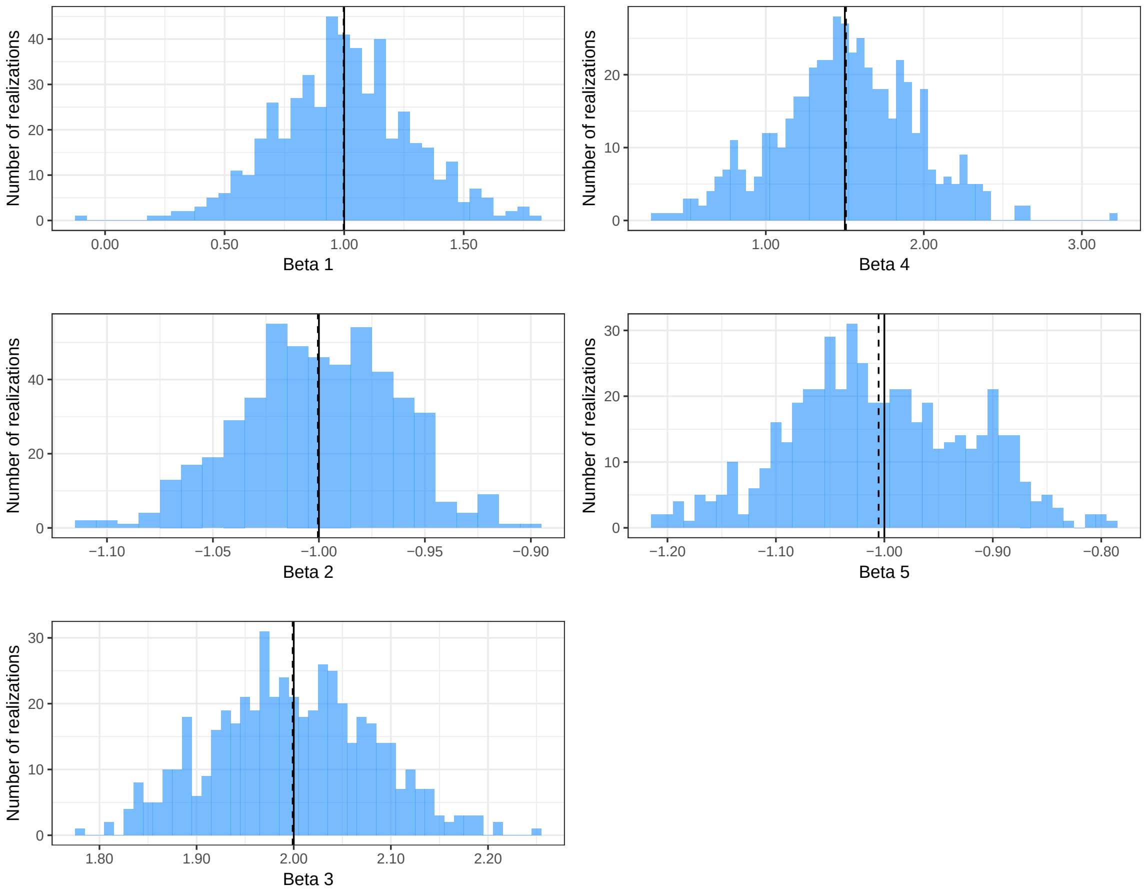
We run 500 Monte Carlo simulations assuming DGP1 and DGP2. For each simulation under DGP1 we estimate the model by Maximum Likelihood assuming DGP1 is the truth. For each simulation under DGP2 we estimate the model assuming DGP2 is the truth. Our goal is to confirm that the likelihoods are correctly 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

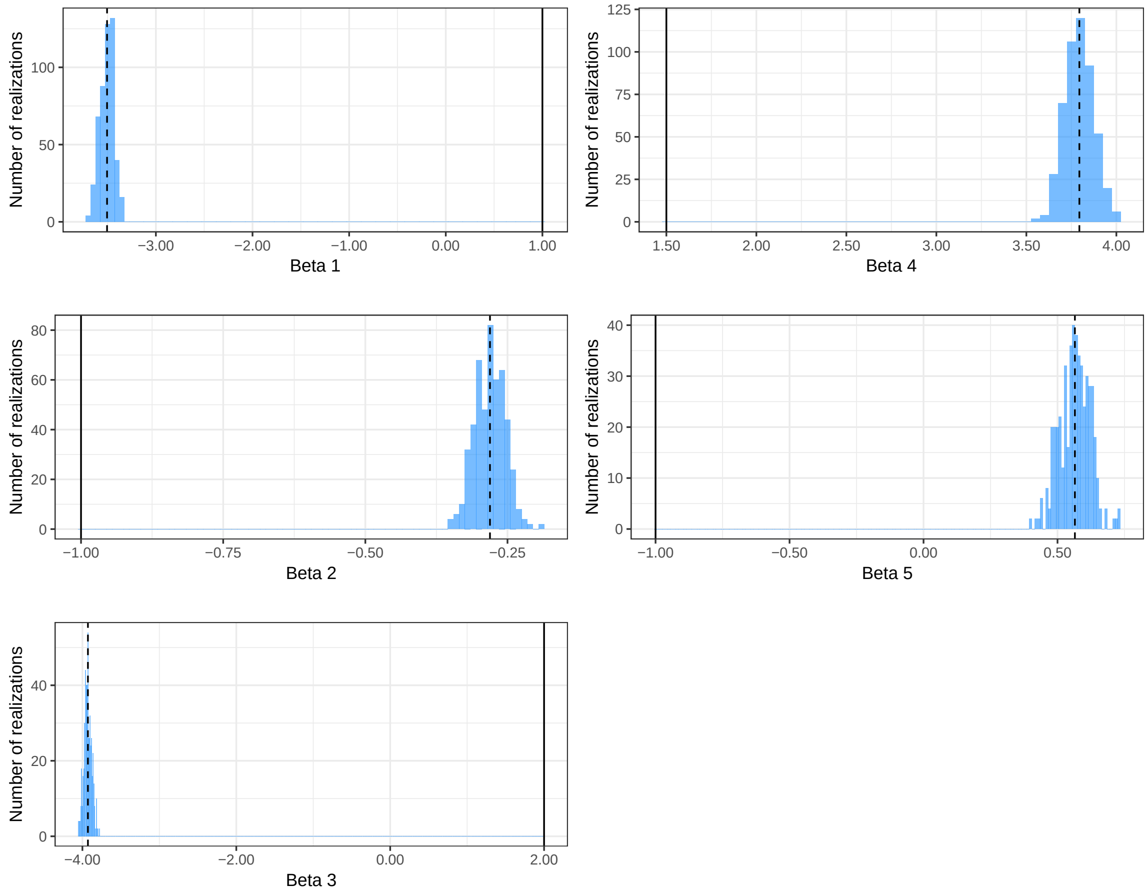


Notes: Monte Carlo simulations under DGP2, estimating parameters with Likelihood approach assuming DGP2 is the truth. Solid lines are true parameters, dashed lines are the the corresponding means of the distributions.

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 efficient 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 give 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



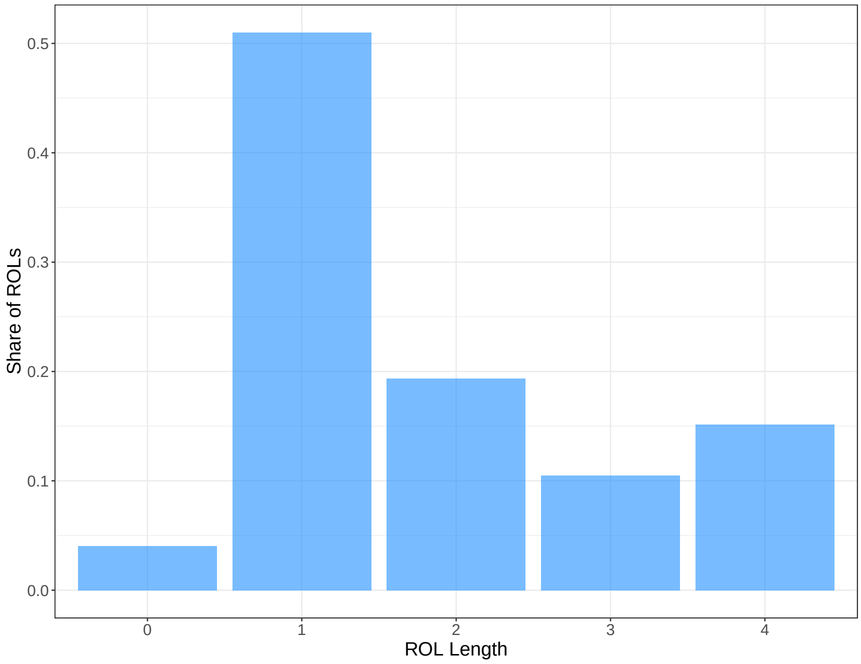
Notes: Monte Carlo simulations under DGP2, estimating parameters with Likelihood approach assuming DGP1 is the truth. Solid lines are true parameters, dashed lines are the the corresponding means of the distributions.

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 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 the choice sets each student faces.

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



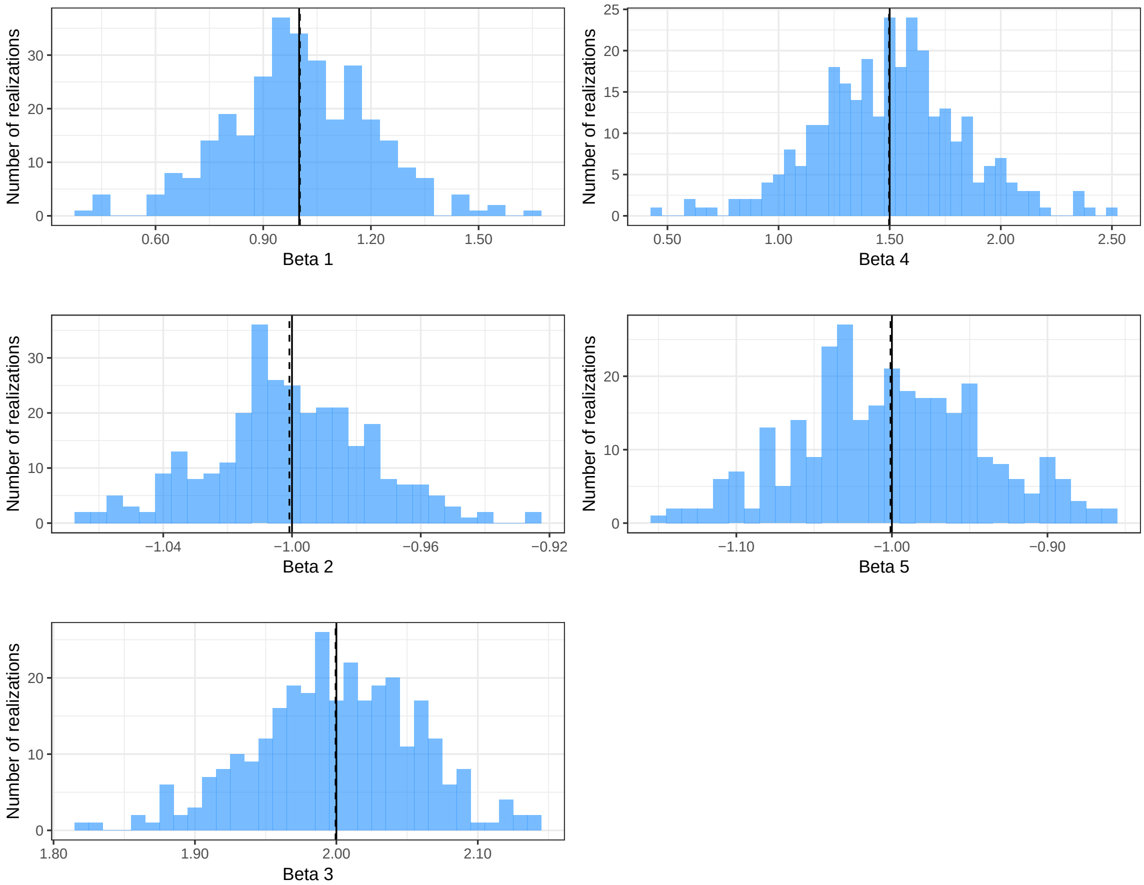
Notes: Distribution of the Length of ROLs under DGP3.

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

$$p(\beta) \sim N(\bar{\beta}, A^{-1}), \tag{7.2}$$

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



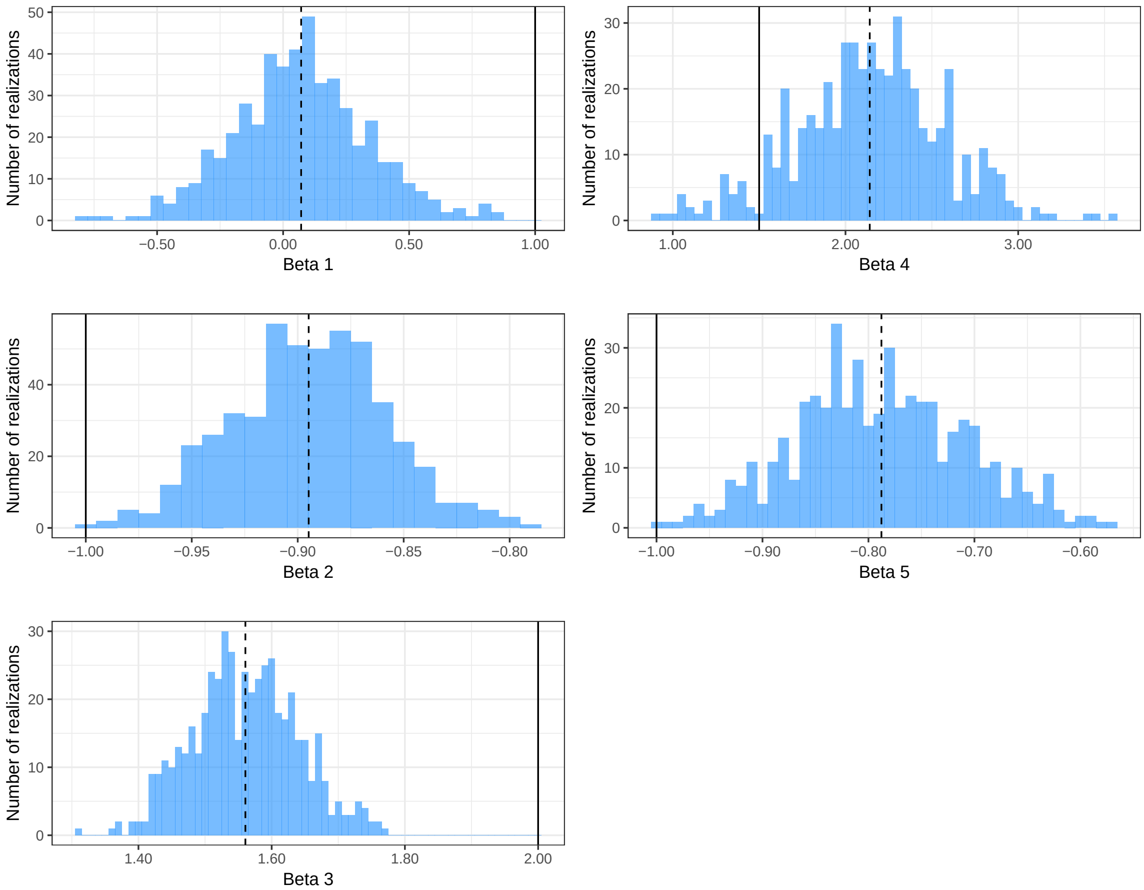
Notes: Monte Carlo simulations under DGP3. Solid lines are true parameters, dashed lines are the means of the distributions of posterior means.

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. Results are presented in Figure 7.10.

We observe that if we ignore the constraint on the length of the list and this constraint binds, estimates will be 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 6.9 and 6.11 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 efficient 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 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 don't include programs in their application lists if it's 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. Moreover, as Agarwal and Somaini (2018) show, it is easy to extend the Gibbs Sampler approach to accommodate mixtures of students with different levels of sophistication.

Our estimation results strongly suggest that "short-list" students should not be interpret as truth-tellers, even in a seemingly strategy-proof environment.

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.

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.

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.

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.

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.

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

Appendix

A 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 $R = \{R(1), \ldots, R(k + 1)\}$.[15] 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

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

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

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

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

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'.[16] 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:

Case 1: Suppose $n < j_1$. Then $R(l) = R'(l),\ \forall l = 1, \ldots, n$, and therefore[17]

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

[16] 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.

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

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

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

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

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

Similarly, after some algebra we obtain that

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

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

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

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,

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

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.

□

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

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

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

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

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

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

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

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

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

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).