Submitted to *Unknown Journal*

Dynamic College Admissions

TOMAS LARROUCAU
Department of Economics, Arizona State University

IGNACIO RIOS
School of Management, The University of Texas at Dallas

We study the relevance of incorporating dynamic incentives and eliciting private information about students' preferences to improve their welfare and downstream outcomes in centralized assignment mechanisms. Using administrative data and two nationwide surveys, we identify two behavioral channels that largely explain students' dynamic decisions: (i) initial mismatches and (ii) learning. Based on these facts, we build and estimate a structural model of students' college progression in the presence of a centralized admission system, allowing students to learn about their match quality over time and reapply to the system. We use the estimated model to analyze the impact of changing the assignment mechanism and reapplication rules on the efficiency of the system. Our counterfactual results show that policies that provide score bonuses that elicit information on students' cardinal preferences and leverage dynamic incentives can significantly decrease switching and increase students' overall welfare.

KEYWORDS. college admissions, dynamic matching, college retention.

1. INTRODUCTION

According to Kapor et al. (2020a), at least 46 countries use a centralized system to organize their admissions to college, including Turkey, Taiwan, Tunisia, Hungary, and Chile. Although extensive literature analyzes the pros and cons of different mechanisms used to perform the allocation, their effects on policy-relevant downstream outcomes (beyond the initial assignment) is unclear. For instance, policymakers often care about students' retention, which is especially low; only 40% of full-time bachelor's students graduate on time (OECD 2019). This low yield can be particularly severe for developing countries such as Chile, where 30% of students switch, close to 30% drop out, and the overall

Tomas Larroucau: Tomas.Larroucau@asu.edu
Ignacio Rios: iar200000@utdallas.com
We are very grateful for the guidance and helpful comments from Eduardo Azevedo, Hanming Fang, Margaux Luflade, Christopher Neilson, Rakesh Vohra, and seminar participants at UPenn, Western Ontario, TSE, NYU, ASU, UCH, PUC, Exeter, PSU, and the Cowles Summer Conference.

This paper would not have been possible without support from MINEDUC, SUA, DEMRE, and the Vicerrectoría de Tecnologías de la Information (VTI) from the University of Chile. We especially thank Leonor Varas, María Elena Gonzalez, Alejandra Venegas, José Piquer, and Giorgio Parra for their collaboration and support. Any errors or omissions are our own.

on-time graduation rate is 16% (the lowest of all countries in the OECD).[1] As this example illustrates, important downstream outcomes should be taken into account when allocating these resources, and centralized assignment mechanisms may help to improve them.

To understand the effects of centralized mechanisms on outcomes, it is essential to account for features that characterize real-life applications and that are mostly overlooked in the literature. One such feature is that matching markets are typically dynamic. For instance, students can learn over time about their match quality with programs, reapply, and switch from their initial assignment if they are assigned to a more preferred program; and they can also drop out at any point in their college progression.[2] Another feature is that students may have private information that is not elicited in the admissions process and could affect their future outcomes and the higher education system's efficiency. For instance, students' intrinsic motivation or vocation, which would be captured by their cardinal (or the intensity of their) preferences, could affect their persistence in their programs, and thus impact the system's efficiency. Therefore, designing admissions systems that consider the dynamic nature of incentives and elicit information about students' cardinal preferences can be critical for improving students' outcomes and the efficiency of the system.

In this paper, we study how to design matching markets in which agents have dynamic considerations, learn about their match quality through experience, and have private information that may affect their outcomes. Moreover, we evaluate how changes in assignment mechanisms can impact students' welfare and downstream outcomes, including their college grades, on-time graduation rates, and retention. We accomplish this by incorporating dynamic incentives and eliciting information on students' cardinal preferences.

By combining administrative data from the Chilean college admissions system and two nationwide surveys we designed and conducted, we show that two behavioral channels largely explain students' dynamic decisions. The first channel, called the *learning* channel, posits that students learn about their match quality during their college experience, which potentially changes their consumption value and future returns, and thus motivates them to switch or drop out to avoid ex post mismatches. The second channel, called the *initial mismatch* channel, posits that students may enroll in less preferred programs to improve their outside option and later participate again in the admissions process to switch to a more preferred option.

Note that these two channels may have different implications. On the one hand, if learning is limited and thus preferences are persistent over time, it may be desirable to restrict reapplications and force students to internalize the crowd-out externality they generate.[3] On the other hand, if learning explains most students' dynamic decisions, it

[1] Similar concerns arise in school choice, whereby policymakers often care about achieving social mobility, meritocracy, and equal access to opportunities (Tanaka et al., 2020).

[2] In school choice, many systems—including that in NYC (Abdulkadiroğlu et al., 2005a) and (Narita, 2018); Boston (Abdulkadiroğlu et al., 2005b); and Chile (Correa et al., 2021)—have multiple rounds, and students/families can either accept their assignment or reject it and reapply to the system in the next round.

[3] A similar crowd-out externality occurs when students repeatedly take admissions exams (Krishna et al., 2018).

may improve welfare to facilitate switchings and avoid ex post mismatches. Hence, the welfare implications of limiting or encouraging switchings are unclear. [4]

To evaluate the effects of these two channels, we introduce a structural model that captures the application behavior of students, as well as their decisions to enroll, retake the admission tests, reapply, switch, and drop out, allowing students to learn about their unobserved abilities—i.e., match quality—during their academic progression. In particular, we assume that students base their application and enrollment decisions on both (i) the value of studying each program and the continuation value of retaking the admission tests and reapplying to the system and (ii) their labor market prospects. As they progress in college, students observe noisy signals of their unobserved ability from their grades, and they use this information to update their continuation values for each program. Based on this, students decide whether to continue in their current program, reapply to the system, or drop out and choose their outside option. Finally, students face graduation probabilities, and then enter the labor force and receive pecuniary and non pecuniary values from the labor market.

The main challenge in estimating our model is to separately identifying the *learning* and *mismatch* channels. To identify the former, we leverage correlation patterns between students' college grades and their decisions to reapply, switch, or drop out. On the other hand, to identify the latter, we combine two sources of variation: (1) students' beliefs about their admissions probabilities and (2) the persistence of students' preferences and the relation between students' preferred assignment and their outcomes. Specifically, by leveraging the discontinuities generated by admissions cutoffs, we show that there is a positive causal effect of not being assigned to the top-reported preference on the probabilities of reapplying (65% increase) and of switching (58% increase), which supports the existence of the mismatch channel. Overall, our results suggest that learning explains close to half of switching decisions, while mismatches and congestion explain the remaining switches and part of the dropout decisions.

After estimating the structural model, we assess whether changes in the assignment process—either through changes in the reapplication rules or in the assignment mechanism—can affect students' outcomes. We find that penalizing students who switch, as is the case in Turkey; giving a score bonus for all first-year applicants, as is the case in Finland; or allowing students to signal one of their preferences to get a bonus in that specific program, in the spirit of the signaling mechanism in the Economics job market, can significantly reduce switching rates, and at the same time increase students' ex post welfare. We also find that these effects are robust to changes in the fraction of participants who behave strategically, as opposed to other approaches such as constraining the length of application lists.[5]

[4] In Appendix A we present a stylized model and show how a clearinghouse can improve students' outcomes by eliciting cardinal information about their preferences. Moreover, Proposition 1 in Appendix A illustrates how both behavioral channels might affect students' switching behavior and welfare in equilibrium.

[5] Our results motivated the Ministry of Education of Chile to relax the constraints on the length of application lists for the 2023 admissions process.

Our counterfactual experiments stress the importance of correctly balancing the effects of the two behavioral channels: allowing students to learn through experimentation and reducing the crowd-out externality caused by initial mismatches. Overall, our results show that incorporating dynamic incentives and eliciting students' cardinal preferences through changes in the reapplications rules and the assignment mechanisms can significantly affect students' outcomes and their overall welfare. These insights can be informative to improving the design of many matching markets that exhibit similar features. For instance, in organ transplant systems, one of the primary goals is to maximize patient survival (Agarwal et al., 2021). Patients have private information regarding their health, face dynamic considerations such as when to accept or reject an organ, and even learn about organs' qualities over time (Zhang, 2010). In entry-level labor markets, employers may care about turnover, and agents may have private information about their preferences, learn about their match qualities through experience, and face dynamic considerations such as deciding when to enter the labor market, exit, re-enter, and rematch with employers. Our key insight is that market designers should correctly balance the gains from learning through experimentation and the crowd-out externality produced by initial mismatches to improve the efficiency and equity of these markets.

The paper is organized as follows. Section 2 discusses the most closely related literature. Section 3 describes the Chilean college admissions system and provides empirical evidence for the two behavioral channels. Section 4 presents our model, and Section 5 describes our identification strategy. Section 6 describes the estimation approach and its results. Section 7 reports our counterfactual results, and Section 8 concludes.

2. Literature

Our paper combines two strands of the literature: (i) the empirical analysis of assignment mechanisms and (ii) the empirical analysis of college choices under uncertainty.

The first strand focuses on (i) understanding the incentives that centralized assignment mechanisms introduce, (ii) how to use the data generated by these mechanisms to identify and estimate students' preferences/beliefs, and (iii) measuring the welfare effects of changing assignment mechanisms in different settings. Depending on the available data and the incentives students face, researchers have developed various methodologies to identify and estimate students' beliefs and preferences (see Agarwal and Somaini (2019) for a survey). Prior evidence of the effects of changing the assignment mechanism and application rules on students' welfare has yielded mixed results. Researchers have found that mechanisms that elicit the intensity of students' preferences can achieve higher ex ante welfare (Agarwal and Somaini (2018), Calsamiglia et al. (2020), He (2012), among others), but this heavily depends on assumption about students' sophistication (Kapor et al., 2020b), which suggests that the appropriate mechanism depends on the specific setting.

Despite progress in understanding the role of assignment mechanisms and their impact on agents' welfare, the aforementioned studies either consider static settings, assume that preferences do not vary over time, or simply ignore the potential effects of assignment mechanisms on downstream outcomes. [6] Taking a dynamic approach can yield new insights into the classical trade-off between *strategy-proof* mechanisms (such as DA) and mechanisms that elicit the intensity of students' preferences (such as IA). For instance, when students have repeated interactions with the assignment mechanism, ignoring the system's dynamics can lead to biased estimates of the welfare effects of changing the assignment mechanism. The reason is that in static settings, researchers assume that students' indirect utilities are invariant to the counterfactuals. However, if students can have repeated interactions with the assignment mechanism, continuation values might be affected by changes in the mechanism. Moreover, static approaches do not allow researchers to evaluate alternative policies that could enhance welfare, such as modifying reapplication rules, as is the case in Finland and Turkey. Finally, it is crucial to understand the implications of changing assignment mechanisms on students' outcomes, such as students' achievement, persistence, and graduation rates, while allowing for learning and dynamic considerations.

To our knowledge, the only exception to this is Narita (2018), who analyzes theoretically and empirically the welfare performance of dynamic centralized school-choice mechanisms when demand evolves over time. Although the dynamics and learning processes are related, our paper differs substantially, since there are essential differences between school-choice and college admissions systems that affect both the research questions and the identification strategies. In our setting, "switching" costs naturally arise because students incur an opportunity cost when they switch programs and delay graduation. These switching costs are not present in school-choice systems and produce a crowd-out externality that affects the system's efficiency and equity. Given these differences, we focus the impact of changing the assignment mechanism by eliciting preference intensity and modifying reapplication rules on students' welfare and their college outcomes.

The second strand of the literature studies individual education and occupation choices, stressing the role of human capital specificity, uncertainty about match qualities, and how students' choices impact their educational outcomes and labor market returns (see Altonji et al. (2012) and Altonji et al. (2016) for reviews). Almost all papers in this literature focus on decentralized college markets or ignore any rationing mechanism that could play an ROLe in college admissions (an exception is Bordon and Fu (2015)). We use insights from the seminal work by Arcidiacono (2005) and more recent work by Arcidiacono et al. (2016) to model students' learning process and their labor-market outcomes, and augment their methodology by micro-founding the college/major choice process in the presence of a centralized admission system, taking into account students' strategic behavior.

Within this strand of the literature, the closest paper to ours is Bordon and Fu (2015), who analyze the effects of changing the Chilean university system from one in which students choose a college and major at the same time to one in which they choose a

[6] Two recent exceptions are (i) Tanaka et al. (2020), who use a quasi-experimental approach to evaluate the long-run effects of repeated school admission reforms in Japan, and (ii) Agarwal et al. (2021), who structurally evaluate the effect of changing the assignment mechanism of deceased donor kidneys on downstream outcomes.

college first and then a major. [7] Our paper's main difference is that we model the entire application and switching behavior of students and use information from their reported Ranked Ordered List (ROL) over time, grade records, and survey responses to separately identify the persistence of students' preferences from learning. These differences allow us to rely less on the model's particular structure to identify the model primitives. However, we do not consider peer effects in the analysis, and we do not have access to a panel of students' future wages. Our counterfactual experiments also differ in nature. Instead of changing the university system's structure and affecting the *learning* channel, we focus on changes to the assignment mechanism and reapplication rules—which affect the *mismatching* channel—and evaluate the effects of these changes on different outcomes, such as switching and graduation rates.

Our work is complementary to these two strands of the literature, since we provide new insights into the effects of centralized assignment mechanisms from a dynamic perspective. To the best of our knowledge, ours is the first paper that structurally measures the effects of centralized assignment mechanisms and reapplication rules on students' welfare and college outcomes beyond their initial assignment, including achievement, college retention, and on-time graduation rates. Finally, we also contribute to the literature by revisiting the trade-off between eliciting the intensity of students' preferences and guaranteeing *strategy-proofness*, but we do so in a dynamic context that allows for private information about students' preferences and learning about match qualities.

3. College Admissions in Chile

The college admissions process in Chile is semicentralized, with the most selective universities having a centralized system and the remaining institutions conducting their admissions processes independently. This paper's empirical application follows the cohort of 2014 and focuses on the centralized part of the system, known as *Sistema Único de Admisión* (SUA). This part of the system is organized by the *Consejo de Rectores de las Universidades Chilenas* (CRUCH), and its admissions process is operated by the *Departamento de Evaluación, Medición y Registro Educacional* (DEMRE).

To apply to any of the close to 1,500 academic programs offered by the 41 universities that are part of the centralized system, students must undergo a series of standardized tests (*Prueba de Selección Universitaria* or PSU). These tests include Math, Language, and a choice between Science or History, and provides a score for each. Students' performance during high school yields two additional scores: one obtained from the average grade during high school (NEM) and a second that depends on the relative position of the student among his/her cohort (Rank). A distinctive feature of the system is that admission to any program is solely based on these *admission factors*.

After scores are published, students can submit a list with no more than ten programs, ranked in strict order of preference. We refer to these lists as Rank Order Lists

[7] Malamud (2011) also analyzes the trade-offs students face when they specialize early in their college education. The author argues that if the rate of field switching in systems with an early specialization is high, this can be seen as evidence that education provides valuable information on match quality and that match quality has a large impact on education returns.

(ROLs). Notice that students directly apply to a program, i.e., they must list university-major pairs on their ROL. In the remainder of the paper, we refer to these pairs as programs. Also, it is important to note that there is no monetary cost for submitting an application.

On the other side of the market, each program announces its vacancies, the weights on each admission factor, and the set of additional requirements they will consider for applications to be valid. For instance, universities may require a minimum application score or a minimum score in some of the PSU tests, among other less common requirements.[8] Each program's preference list is defined by first filtering all applicants who do not meet these requirements. Students are then ordered based on their application scores, which are computed as the weighted sum of the applicants' scores and the weights predefined by each program.

Considering the vacancies and the preferences of the applicants and programs, DEMRE runs an assignment algorithm to match students to programs. The mechanism used is a variant of the student-proposing Deferred Acceptance algorithm, in which all students tied for the last seat in a program must be admitted. A thorough description of the mechanism can be found in Rios et al. (2021). As a result of the assignment process, each program is associated with a cutoff, such that all students whose weighted score is above it are granted admission, and all students with scores below the cutoff are waitlisted and thus may have to enroll in a lower-ranked preference. This property is known as the *cutoff structure*.

The enrollment process starts right after assignment results are published, and consists of two rounds. In the first round, only assigned students can enroll in their preference of assignment; in the second round, programs with seats left after the first stage can call students on their waitlists and offer them the chance to enroll. Also, at any point, applicants can apply and potentially enroll in a program outside the centralized admissions system, and they also have the option to join the labor force directly. Moreover, students can participate in the admission process as many times as they want, and they can use the scores obtained in the previous year as part of their application.[9]

3.1 *Data*

Our dataset includes (i) administrative data provided by DEMRE and the Ministry of Education (MINEDUC), (ii) two surveys designed and conducted in collaboration with CRUCH and DEMRE aiming to elicit students' preferences and beliefs about admissions probabilities, and (iii) grade records facilitated by CRUCH. Specifically:

- Admissions process: This includes students' socioeconomic characteristics (including self-reported family income, parents' education, and the municipality in which

[8]For instance, limiting the position of a program on a student's ROL or the total number of programs listed from a given university. Some programs, such as music, arts, and acting, may require additional aptitude tests.

[9]To compute the application score, each program uses the weighted average score based on the pool of scores for the current year and the pool of scores for the previous year (if any). The maximum between these two scores is considered as part of the application.

the student lives, among others); scores; applications; final assignment and enrollment decisions, spanning from 2007 to 2020. In addition, this includes data on programs and universities characteristics, including their number of vacancies, weights, tuition, duration, major, and program's location, etc.

- Labor Market: This includes aggregate information about the labor market prospects of each program, spanning from 2014 to 2018. More specifically, we have estimates for average wages (four years after graduation) at the program level and the overall employment probability one year after graduation. We also have data at the major level, including average wage for the first to the fifth year after graduation; five points in the distribution of average wages for the first year and fifth year after graduation (10th, 25th, 50th, 75th, and 90th), employment probabilities for the first and second year after graduation; and the evolution of average wages from the first to the tenth year after graduation.
- Grades: This includes the cumulative GPA for their first three years of college for every student who enrolls in a program that is part of the centralized system in 2014 and 2015. To our knowledge, this is the first paper to use these data.
- Surveys: This includes the results of surveys that we designed and conducted in 2019 and 2020 to gather information on students' preferences for programs and their beliefs about admissions probabilities. These surveys were sent to all students who participated in the PSU tests (more than 150,000 each year) at the end of the application process. We asked students about their top-true preference and their beliefs about their admissions probabilities for each program on their ROL and also for their top-true preference (if not in ROL), among other questions. Moreover, as many students reapply to the centralized system after a year, we have information about students' preferences and their beliefs for a small panel of reapplicants who participated in the survey. To our best knowledge, this is the first time that data on beliefs about admission probabilities and college persistence has been collected for a centralized college admissions system.

Throughout the paper, we focus on a subset of the population to reduce computational complexity. Specifically, we focus on students who graduated from a high school within the Metropolitan region in 2013, participated in the 2014 admissions process (i.e., took the PSU tests), and had an average score between Math and Language above 475.[10]

3.2 *Empirical Facts*

As discussed in Section 1, we posit that students' dynamic decisions are largely explained by two behavioral channels: (i) *mismatching*, whereby students assigned to less

[10]This reduces the number of programs to less than half (435) without major loss, as close to 80% of applications from students living in the Metropolitan region include only programs located in that region. Hence, we treat the Metropolitan region as a market. Finally, we exclude students with average score below 475 (less than 13% of students who can apply) because they do not satisfy loan eligibility requirements.

preferred programs reapply to improve their allocation, and (ii) *learning*, whereby students learn about their match qualities (abilities and preferences) over time and potentially decide to move to other programs. In this section, we provide empirical evidence that supports the existence of these two channels.

One of the main challenges in disentangling these two behavioral channels is that we do not have cardinal information regarding students' preferences, since we only observe their characteristics and ROLs. Moreover, students' reports may not be truthful; some students tend to skip programs for which their admission chances are relatively low (Larroucau and Rios, 2018). Despite this, we claim that reported ROLs still shed light on the intensity of students' preferences. For instance, we know that listing a program in a higher position on the ROL implies a higher preference intensity than programs listed in lower positions (Haeringer and Klijn, 2009). Moreover, not listing a program for which the probability of admission is high enough implies that the ROL programs are preferred (see Larroucau and Rios (2018) for a detailed discussion). Finally, apart from the information we can extract from students' ROLs, adding dynamics can help identify preferences' intensity. For example, students who decide to reapply must have higher intensity in their preferences than students who remain in their program (conditional on observable characteristics and in the absence of learning). Similar information can be inferred from switching and dropout decisions.

3.2.1 *Mismatching* In Table 1, we report the average switching and dropout rates in the first four years, separating by income level—high or low—and gender.[11] First, we observe that close to 23.5% of students switch from the first program they enrolled in, and 23.9% drop out within the first four years. Second, comparing program switching and dropout rates by gender (within an income level), we observe that women are more persistent in their academic progression, since their switching and dropout rates are lower than those for men. On the other hand, comparing these rates by income level (within gender), we observe that low-income students are less likely to switch programs during their academic progression. However, we also observe that low-income students are significantly more likely to drop out.[12] One potential explanation is that low-income students have less flexibility to switch programs and delay graduation due to budget constraints and, at the same time, face a more difficult time in college due to their disadvantageous background, which increases their chances of dropping out. These results suggest that there are significant differences in switching and dropout rates by gender and income, and are similar to those obtained if we focus on switching and dropouts within the first year. Hence, throughout the rest of the paper we focus on the latter for simplicity. This choice is without major loss of generality, since close to 80% of switching takes place in the first two years, and close to 2/3 of these occurs within the first year (see Figure B.4 in Appendix B.3).

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

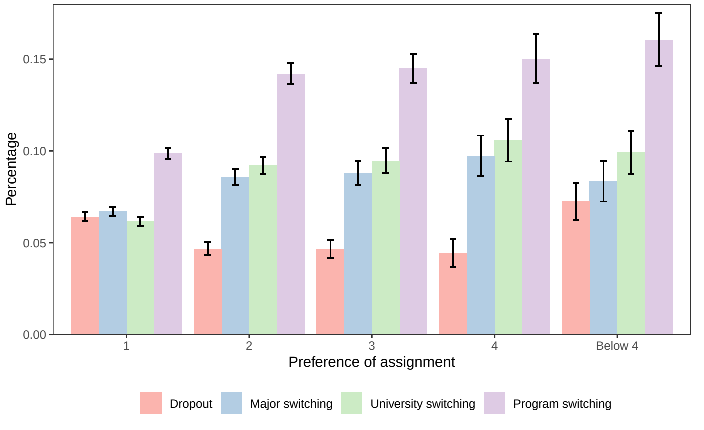
[12] While credit constraints likely play an important role in the drop-out decisions of some students, the large majority of attrition of students from low-income families should be primarily attributed to reasons other than credit constraints (Stinebrickner and Stinebrickner, 2008).

TABLE 1. Switching and Dropout by Gender and Income

	Income	Switches: Program	University	Major	Math type	Dropout
Men	Low	0.235	0.117	0.107	0.044	0.239
		(0.008)	(0.006)	(0.006)	(0.004)	(0.008)
	High	0.258	0.137	0.141	0.051	0.155
		(0.006)	(0.004)	(0.004)	(0.003)	(0.005)
Women	Low	0.182	0.090	0.096	0.046	0.202
		(0.008)	(0.006)	(0.006)	(0.004)	(0.008)
	High	0.226	0.115	0.133	0.068	0.106
		(0.006)	(0.004)	(0.004)	(0.003)	(0.004)
Overall		0.232	0.120	0.126	0.055	0.158
		(0.003)	(0.002)	(0.003)	(0.002)	(0.003)

Note: Standard errors reported in parenthesis.

FIGURE 1. Switchings and dropout



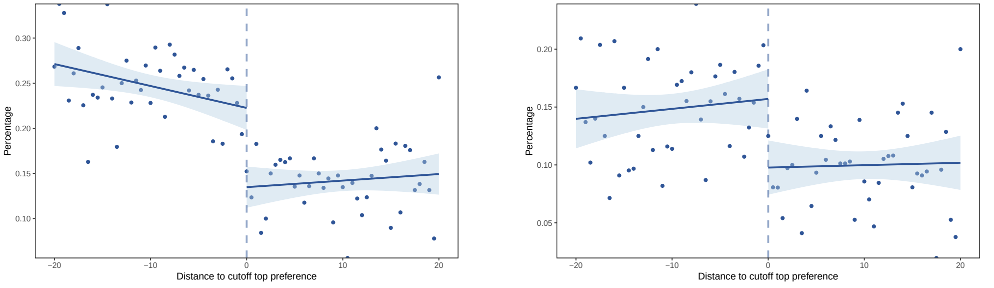
Note: Switching categories do not include stop out.

To assess whether the preference of assignment impacts student outcomes, Figure 1 shows switching and dropout rates (at the end of the first year) conditional on students' preference of assignment. We observe that students assigned to lower reported preferences switch at higher rates compared with students assigned to their top reported preference. Among students assigned to their top reported preference, 9.86% switch programs at the end of their first year, compared with almost 15% of those assigned to their fourth choice. In contrast, we observe no effect of the preference of assignment in first-year dropout rates. These results suggest a strong correlation between the preference of assignment and switching rates. One potential explanation is that there are observable differences between students assigned to lower and higher preferences. For instance, students with low scores are systematically assigned to lower preferences, which generates a positive correlation between assignment preference and switching rates. Similarly, programs listed in lower preferences are more likely to be of lower quality, which incentivizes students to try to switch.

To make a causal claim, we use a regression discontinuity design that exploits the algorithm's cutoff structure to perform the allocation. If we assume that students around

FIGURE 2. Effect of Cutoff Crossing

(a) Reapplication (b) Switching



the cutoff are similar and only differ in their right to enroll in a higher preference, we can estimate the causal effect of interest.[13] In Figure 2, we display binned means of different outcomes as a function of the distance between cutoffs and students' scores in their most preferred listed program.[14] Figure 2a shows that students right below the cutoff are close to 8.7% more likely to reapply the following year, which corresponds to a relative change of close to 62.1%. Figure 2b shows that students below the cutoff are close to 5.8% more likely to switch programs within the centralized system, which corresponds to a relative change of more than 57.9%.[15] These results confirm our previous findings, i.e., that students assigned to lower preferences are more likely to reapply and switch programs the following year.

The previous empirical facts show a causal effect of the preference of assignment on students' persistence in their initial assignment. To show that the mismatch channel partially explains this, we use the 2020 survey on students' preferences and beliefs, in which we find that a significant fraction of students know, before enrolling in their assigned programs, that they will be less likely to remain in that program if they are assigned to a lower reported preference (see the details in Appendix B.3.1). These results cannot solely be explained by students' or programs' characteristics.

3.2.2 *Learning* Students' preferences may change during their first year in college, which could affect their decision to reapply. We analyze students' reapplication at the end of their first year and classify switching into three categories: (i) *Up*, (ii) *Down*, and (iii) *Out*. Students move *Up* (*Down*) if they switch to a program listed above (below) their initial enrollment on their initial ROL. Students move *Out* if they switch to a program not listed on their initial ROL.

[13] A detailed discussion of this analysis and its potential selection issues is provided in Appendix B.2.

[14] In Appendix B.2.1 we report the results of a similar analysis that considers students' top true preferences. The results are relatively the same.

[15] The average fraction of students who reapply is 14.0% and 22.6% for students above and below the cutoff, respectively. The average probability of switching is 10.0% and 15.8% for students above and below the cutoff, respectively.

We find that, among students who switch in their first year, 18.1% move *Down*, 14.8% move *Up*, and 67.1% move *Out*. Moreover, more than half of the latter switches involve more selective programs, i.e., programs with higher admission cutoffs compared with their initial enrollment. These results suggest that both channels explain students' switching significantly. Students who move *Down* or *Out* to less selective programs are likely to have learned about their (poor) match quality (learning channel), and students' who move *Up* or *Out* to more selective programs may be trying to find a better match.

To rule out forced switching, i.e., students who switch because they were expelled due to poor performance, in Table 2 we analyze the effect of first-year grades on reapplication and switching decisions. In all of these models, we control for demographics (gender, income); scores (NEM and the average between Language and Math); and the preference of assignment in the initial year. Columns (1), (3), and (5) include the entire sample, and columns (2), (4), and (6) focus on students with a GPA greater than or equal to 4.0. Since 4.0 is the pass/fail threshold (the scale is from 1.0 to 7.0), by focusing on students with a GPA above 4.0 we rule out the explanation that all students who switch were forced to leave their initial program.

TABLE 2. Effect of Grades on Outcomes

	Reapply SUA		Switch Program		Switch Down or Feasible		Switch Up
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
GPA	−0.905	−0.404	−1.232	−0.300	−1.283	−0.624	0.099
	(0.031)	(0.075)	(0.037)	(0.075)	(0.040)	(0.097)	(0.123)
GPA ≥ 4	No	Yes	No	Yes	No	Yes	No
Observations	13,414	11,120	12,584	10,846	12,584	10,846	12,584

Note: We use data on grades from the cohort that graduated from high school in 2014 and enrolled in 2015 in the program they were assigned to in the centralized system. GPA is measured on a scale of 1 to 7, and failing grades are below 4.0.

We observe that GPA is negatively correlated with the decision to reapply, switch, and in particular switch to a lower preferred program or to a program that was not in the original ROL but was feasible (i.e., admission probability above 0). The latter is confirmed by Figure 3a. In addition, we observe that switches up are not correlated with grades, as shown Figure 3b. Finally, we obtain similar results when we restrict the analysis to students with a GPA above the passing grade. These results suggest that students may learn from their (low) grades and may decide to switch to programs they preferred less, according to their initial ROL.

Our previous results show that students' reported preferences may change during their first year in college. To analyze changes in true preferences, we use the surveys conducted in 2019 and 2020. Specifically, we construct a panel of students that consists of those who participated in both surveys and responded to the same questions (close to 1,300 students), and we compare the top true preference reported in each year. In Figure 4, we plot the fraction of students who changed their top true preference (for programs and also for universities) as a function of their initial preference of assignment (in 2019). First, we observe that on average, close to 65% of the students in the data changed

FIGURE 3. Effect of Grades on Switchings by Type

(a) Down or Feasible (b) Up

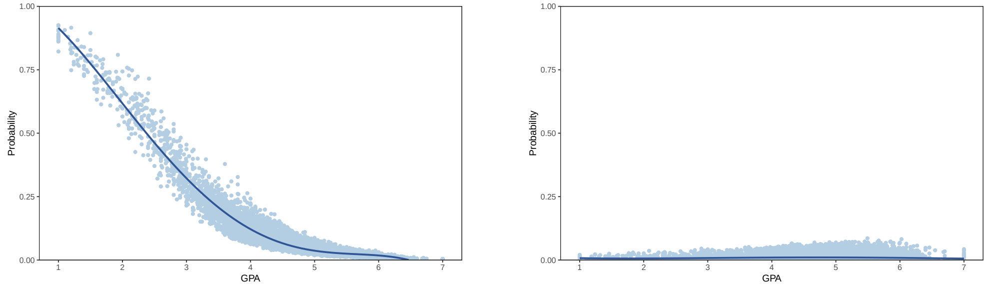
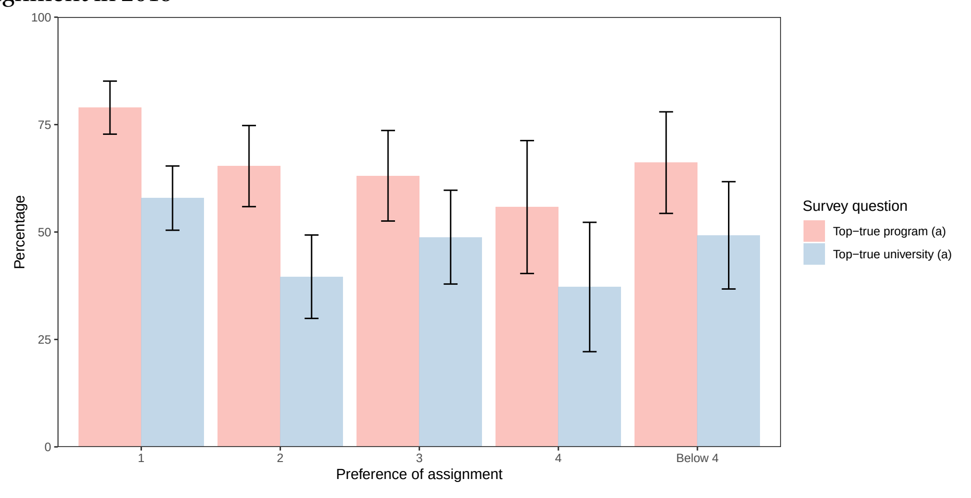


FIGURE 4. Percentage of reapplicants that change their top-true preference, by preference of assignment in 2019



their top true preference after their first year in college. Moreover, close to 50% of students even changed their most preferred university. Second, we observe that students initially assigned to lower preferences are less likely to change their top true preference for programs. This result is consistent with the existence of the *mismatch* and *learning* channels, since students initially assigned to their top-reported preference have a lower probability of being *mismatched*, and thus their reapplication suggests that they learned about their match quality during their first year in college.

4. MODEL

This section describes our model of students' applications, enrollment, and dropout decisions, including learning about their match quality over time. The goal is to have a model that encompasses the empirical evidence described in the previous sections, which allows us to measure how much of students' switching behavior is explained by learning over time vs. initial mismatch, and assess whether students' outcomes can be affected by changing the mechanism and reapplication rules.

Throughout the model, we assume that students learn about their match-qualities with programs and that this information might affect their future returns upon graduation. In this sense, we label unknown match-qualities as unknown abilities to give them a productive meaning. Abilities are assumed to be multidimensional and partially known by students. In particular, students receive signals of their unknown abilities through their college GPA and, based on this information, they update their beliefs. Given their updated beliefs, students choose to (i) continue in their enrolled program, (ii) reapply to the centralized system, expecting to switch programs, or (iii) drop out from the centralized system. Finally, we model labor market returns as a function of the students' major, abilities and observable characteristics, and path through college.[16]

4.1 *Model overview*

For estimation purposes, we consider a three-period model. Periods 1 and 2 correspond to the first and second years of college after graduation from high school. Period 3 starts at the beginning of the third year of college and collapses the later years until graduation from college, with the discounted payoffs received in the labor market. Every period involves several decisions and stages. In period 1, students who graduated from high school make their application decisions, receive their enrollment, choose whether to retake the PSU, obtain their college grades at the end of the first year, and update their beliefs about their unknown abilities. In period 2, students make reapplication decisions, and depending on their assignment and enrollment status, choose between remaining in their current enrollment, switching to their new assigned program, or dropping out. In period 3, students face dropout and graduation probabilities (estimated from the data) and enter the labor market. We describe each of these stages in detail in Appendix C.1.

4.2 *Labor market*

For the labor market stage of the model, we follow Arcidiacono (2004) and Arcidiacono (2005). The labor market is an absorbing state, and utility while in the workforce is given by the present value of lifetime earnings and non pecuniary utility. We further assume that utility is separable over time. In particular, we assume the following specification:

$$V_{ijt}^{w} = \underbrace{\alpha_1^w \left(\alpha_{fm_j} + \alpha_{im_j}\right) + \alpha_2^w A_{ij} + \alpha_3^w \bar{A}_{k_j} + \alpha_4^w A_{ij}^u}_{\text{non pecuniary}} + \underbrace{\alpha_5^w \log\left(E_w\left[\sum_{\tau=0}^{T-t} \beta^\tau P_{m_j\tau}^w w_{ij\tau}\right]\right)}_{\text{pecuniary}}, \tag{1}$$

where the first four terms capture the non pecuniary payoff that individuals perceive from working in a job associated with program j. We allow these payoffs to vary with the

[16] In Appendix C we describe models for enrollment, dropout, and graduation.

student's observed ability in program j, A_{ij}; the average observed ability in college k_j, $\bar{A}_{k_j}$; and unknown ability A^u_{ij}. We also include student i's random coefficient for *major* m_j, α_{im_j}, and *fine major* fixed effects α_{fm_j}.[17] By incorporating random coefficients, we introduce persistence over time on students' unobserved preferences, which can affect both their flow utility and their utility in the work force.[18] The fifth term captures the pecuniary payoff students receive in the work force, with $w_{ij\tau}$ representing the earnings for student i with tenure τ, graduating from program j. In addition, T is the retirement date (which varies by gender), t corresponds to the year (period) in which the student graduates from college and enters the work force, β is a common discount factor, and $P^w_{m_j\tau}$ is the employment probability in *major* m_j for an individual with tenure τ. Notice that student i receives this continuation value only if she graduates from her program. If, instead, student i drops out in period t, we assume that she receives a continuation value given by V_{i0t} that depends only on her observable characteristics X_i. This is formalized in Assumption 1.

ASSUMPTION 1. *If student i graduates from program j in period t, she obtains a continuation value equal to V^w_{ijt}. In contrast, student i receives a continuation value equal to $V_{i0t} = V(X_{i0}, t)$ if she drops out from her program in period t, where X_{i0} includes gender and family income.*

We specify the wage students receive conditional on graduation as a function of their tenure, *major* m_j, observable characteristics Z^w_i (gender), expected grades upon graduation $\bar{G}_{ij}$, and college quality, proxied by the average ability of their classmates $\bar{A}_{k_j}$.[19] More specifically, we assume that the log earnings for student i with tenure τ, graduating from program j in period t, can be written as

$$\log(w_{ij\tau}) = \lambda_{1m_j} + \lambda_2 \bar{A}_{k_j} + \lambda_3 \bar{G}_{ij}\left(A_{ij}, A^u_{ij}\right) + \lambda_4 Z^w_i + \Lambda_{m_j\tau} + \epsilon_{ij\tau}, \tag{2}$$

where $\Lambda_{m_j\tau} = \lambda_{5m_j}\tau + \lambda_{6m_j}\tau^2$ specifies how wages in *major* m_j depend on tenure τ.

4.3 *Academic Progression*

During their academic progression, students receive their flow utility from attending college and observe their grades, which provides them with a signal of their unknown abilities and expected grades upon graduation. As discussed in the previous section, students take into account their ability when computing their pecuniary and non pecuniary labor market returns, and thus the information obtained from their grades is highly valuable. Students may use this information to decide whether to reapply in the next period, continue to be enrolled in the same program, or drop out of college.

[17]We model programs' fixed effects, α_{fe_j}, as the sum of college fixed effect, α_{c_j}, plus a *fine major* fixed effect, α_{fm_j} ("*area carrera genérica*"). Fine majors can take up to 108 categories in our sample. For instance, *Medicine* and *Nursing* are two different *fine majors* from the Health *major*.

[18]In Section 4.3.1 we describe how we model the random coefficients.

[19]We choose to structurally model wages to allow them to change in the counterfactuals. An alternative approach could be to fix wages and incorporate them as observable characteristics in the value function. More details about ability are reported in Section 4.3.2.

4.3.1 *Flow Utility* Let u_{ijt} be the flow utility that student i receives for attending program j at time t,

$$u_{ijt} = \alpha_{fe_j} + \alpha_{im_j} + \alpha_{ik_j} + Z^u_{ij}\alpha - C_{ijt} + \varepsilon_{ijt},$$

where α_{fe_j} is a program fixed effect; α_{im_j} and α_{ik_j} are student i's random coefficients for *major* and *university type*, respectively; and $Z^u_{ij}\alpha$ captures the effect of student and program characteristics that are time invariant,

$$Z^u_{ij}\alpha = \alpha_1 A_{ij} + \alpha_2 \bar{A}_j + \alpha_3 D_{ij} + \alpha_4 \frac{(A_{ij} - \bar{A}_j)}{\bar{\sigma}_j}$$

where D_{ij} is the distance between student i's and program j's municipalities; A_{ij} is student i's observed ability in program j; $\bar{A}_j$ is the average observed ability for students assigned to program j in the previous calendar year (the program's selectivity);[20] and $\bar{\sigma}_j$ is its standard deviation. Finally, C_{ijt} captures the monetary cost for student i to enroll in program j at time t and is given by $C_{ijt} = \alpha_{c0}\left(c_{jt} - \tilde{c}_{ij}\right)$, where c_{jt} is program j's yearly tuition plus enrollment fees and $\tilde{c}_{ij}$ captures the sum of all government-provided scholarships for student i in program j.[21]

We follow Larroucau and Rios (2018) and model the random coefficients as a multivariate regression on a set of students' observable characteristics. In particular,

$$\alpha_{im} = \Delta^m Z^m_i + \chi^m_i, \quad \alpha_{ik} = \Delta^k Z^k_i + \chi^k_i,$$

where Δ^m and Δ^k are matrices of coefficients to be estimated; $\chi^m_i \sim N(0, V^m_\alpha)$ and $\chi^m_i \sim N(0, V^m_\alpha)$ are vectors of idiosyncratic shocks with mean zero and variance-covariance matrices $V^m_\alpha = \sigma^{2m}_\alpha \mathbb{I}$ and $V^k_\alpha = \sigma^{2k}_\alpha \mathbb{I}$, respectively; and Z^m_i and Z^k_i are matrices of observable characteristics, where the former includes students' gender, and the latter includes students' family income type.[22] Finally, ε_{ijt} is an idiosyncratic preference shock that is distributed i.i.d type I extreme value with a scale parameter of one. We specify a location normalization and set the systematic value of the outside option (not enrolling in a program within the centralized system) to be $\bar{u}_{i0t} = 0$.

4.3.2 *Learning* As described in Equation 2, students' labor market returns depend on their grades, which in turn depend on their abilities. We assume that these abilities have

[20] We choose to not model endogenuous peer effects because we lack variation in peer composition over time within programs (see Bordon and Fu (2015) and Allende (2019)). However, this channel is less relevant to our counterfactuals, because we aim to swap students around admission cutoffs, without significantly changing the composition of students within programs.

[21] A large literature analyzes the role of credit constraints in shaping schooling choices (see Lochner and Monge-Naranjo (2012) and Lochner and Monge-Naranjo (2016) for an overview). In our context, previous evidence shows a large effect of loan eligibility on initial college enrollment (Solis, 2017). However, recent evidence suggests that students whose average score exceeds loan eligibility requirements (equal to 475 points) do not seem to be highly sensitive to different prices regarding their college re-enrollment or completion rates (Card and Solis, 2020). Following this evidence, we focus on this sample of students and avoid modeling the potential effects of credit constraints on students' dynamic choices.

[22] We classify students as *low income* if their self-reported family income is below the median of the family income distribution and as *high income* otherwise.

two components, one that is directly observable and known by students (and the econometrician), and another that is unknown and learned from the grades obtained during college. More specifically, we assume that students have beliefs about their abilities and they update them as they observe their grades according to Bayes' rule. To formalize these ideas, we start by modeling students' abilities, then model the grade equation, and finish this section by modeling beliefs and the updating process.

4.3.2.1 Ability. Each student i has an observed subject-specific ability vector $A_i = (A_{is_m}, A_{is_v})$; an unobserved (to the student and the econometrician) subject-specific ability vector $A_i^u = (A_{is_m}^u, A_{is_v}^u)$; and a *major*-specific ability $A_{im_j}^u$ for each *major* m_j. Each component of these ability vectors captures the student's known and unknown math and verbal abilities, indexed by s_m and s_v, respectively. We assume that student i's (un)observed ability in program j is given by the weighted sum of her (un)observed abilities, i.e.,

$$A_{ij} = \sum_{k \in \{s_m, s_v\}} \omega_{jk} A_{ik}, \quad \text{and} \quad A_{ij}^u = A_{im_j}^u + \sum_{k \in \{s_m, s_v\}} \omega_{jk} A_{ik}^u, \tag{3}$$

where ω_{jk} is the admissions weight of factor k in program j. Even though subject-specific components do not vary across programs, there is still variation in students weighted abilities due to the heterogeneity of programs' specific weights, ω. In this sense, although major-specific ability is non-transferable across different majors, subject-specific components are imperfectly transferable, which allows for correlated learning across programs from different majors.

4.3.2.2 Grades. As described above, we assume that students observe their grades at the end of each of the first two periods and, based on these signals, update their beliefs about their unknown abilities. Moreover, we assume that grades depend on the *major* (m_j) of the program in which the student is enrolled, on the known (A_{ij}) and unknown abilities (A_{ij}^u), and on a set of observable characteristics (Z_i^g).[23] Also, to capture the fact that students' initial preferences may affect their performance, we include student i's random coefficients for *major*, α_{im_j}, and *university type*, α_{ik_j}. We assume that the grade equation for the first and second periods is given by

$$G_{ijt} = \gamma_{1m_j} + \gamma_2 A_{ij} + \gamma_3 Z_i^g + \gamma_4 \alpha_{im_j} + \gamma_5 \alpha_{ik_j} + A_{ij}^u + \varepsilon_{ijt}^g, \tag{4}$$

where ε_{ijt}^g is white noise distributed $N(0, \sigma_g^2)$. Therefore, upon receiving her grades, the student can compute a signal of her abilities, a_{ijt}, given by

$$a_{ijt} = G_{ijt} - \left(\gamma_{1m_j} + \gamma_2 A_{ij} + \gamma_3 Z_i^g + \gamma_4 \alpha_{im_j} + \gamma_5 \alpha_{ik_j}\right).$$

4.3.2.3 Beliefs and Updating. We assume that students are rational and update their beliefs using the signals about their unknown abilities that come with their grades, according to Bayes' rule. In particular, we assume that students' initial prior about their unobserved major-specific ability is normally distributed with mean zero and variance

[23]The estimation results described in Section 6 only include gender.

σ_m^2 for all students and majors. Similarly, we assume that students' prior about their unobserved subject-specific abilities is also normally distributed with mean zero and variance σ_s^2 for all students and subjects. Thus, unobserved abilities are centered around observed abilities. We formalize this in Assumption 2.

ASSUMPTION 2. *Students' initial priors on their unobserved major and subject-specific abilities are normally distributed with means zero and variances σ_m^2 and σ_s^2, respectively. These priors are common to all students.*

A direct consequence of this assumption is that the posterior distribution of the overall unknown ability in Equation 3 will also follow a normal distribution. Let $\mu_t(A_{ij}^u)$ and $\sigma_t(A_{ij}^u)$ be the prior mean and standard deviation of A_{ij}^u at the beginning of period t. When it is clear from the context, we will remove the argument and simply write them as μ_{ijt} and σ_{ijt}, respectively. Hence, Assumption 2 implies that

$$\mu_{ij1} = 0, \quad \text{and} \quad \sigma_{ij1}^2 = \sigma_m^2 + \sum_{k \in \{s_m, s_v\}} \omega_{jk}^2 \sigma_s^2.$$

In Proposition 3, we show how to compute the posterior mean and variance of the overall unobserved ability after observing a signal a_{ijt}. We defer the proof to Appendix C.2.

PROPOSITION 3. *Suppose that student i is enrolled in program j in period $t = 1$ and that she observes a signal a_{ijt}. Then, she will update her mean unobserved ability in each program j' according to*

$$\mu_{ij't+1} = E_t\left(A_{ij'}^u | a_{ijt}\right) = \begin{cases} \left(\sigma_{ijt}^2 + \sigma_g^2\right)^{-1} \cdot \left[\sum_{l \in \{s_m, s_v\}} \omega_{j'l}\omega_{jl}\sigma_s^2 a_{ijt}\right] & \text{if } m_{j'} \neq m_j \\ \left(\sigma_{ijt}^2 + \sigma_g^2\right)^{-1} \cdot \left[\sum_{l \in \{s_m, s_v\}} \omega_{j'l}\omega_{jl}\sigma_s^2 a_{ijt} + \sigma_m^2 a_{ijt}\right] & \text{if } m_{j'} = m_j \end{cases}$$

Intuitively, students will learn more about programs similar to the ones they are currently enrolled in, especially for programs that belong to the same major and that place similar weights on admissions scores. It is crucial to notice that, according to our model, only those students who are enrolled in a program observe a signal of their abilities. Hence, we assume that students who are not enrolled do not update their prior.[24]

4.3.3 *Application* Once students get their scores—either the first time they take the exams or after retaking them—they must decide which programs to include on their ROL. We assume that students' application behavior can be classified as one of two types: (i) weak truth-tellers, and (ii) strategic. These types are exogenously given, with students being weak truth-tellers with probability ρ and strategic with probability $1 - \rho$. We assume that weak truth-tellers report their true preferences as long as they exceed the outside option, while strategic students submit an ROL that maximizes their expected value. Following Chade and Smith (2006), we assume that this process can be modeled

[24]We make this assumption because we do not have data on students' grades outside the centralized system.

as an optimal portfolio problem, which can be solved using the Marginal Improvement Algorithm (MIA) (see Appendix C.4). Each student i who applies in period t considers a vector of indirect utilities $\{v_{ijt}\}_{j\in M}$ and a vector of beliefs about admission probabilities $\{p_{ijt}\}_{j\in M}$, and the submitted ROL R_{it} satisfies

$$R_{it} \in \underset{R'\in\mathcal{R},|R'|\leq K}{\text{argmax}} z_{R'(1)} + (1 - p_{R'(1)}) \cdot z_{R'(2)} + \ldots + \prod_{l=1}^{k-1}(1 - p_{R'(l)}) \cdot z_{R'(K)} - c(R'),$$

where $z_{R(k)} = p_{R(k)} \cdot v_{R(k)t}$ represents the expected utility (over the assignment) obtained from the k-th preference in the ROL and $c(R)$ is the cost of submitting the ROL R, which in our case is equal to zero.

This model relies on the assumption that students neglect potential correlations across cutoff distributions. Also, to simplify the analysis, we further assume that students do not include programs in their ROL unless it is strictly profitable, as discussed in Larroucau and Rios (2018). This assumption implies that strategic students will not add programs for which their admission probability is zero. Finally, we assume that students have rational expectations regarding their admission probabilities.[25] These assumptions are formalized in Assumption 4.

ASSUMPTION 4. *Students take the distributions over cutoffs to be independent across programs. In addition, students have rational expectations regarding their admission probabilities, and they include programs in their portfolio only if it is strictly profitable.*

Discussion: The parameter ρ should not be interpreted as a primitive of the model, since we expect it to vary with the counterfactuals. The reason is that in the baseline, it could be payoff equivalent to report an ROL as a weak truth-teller or strategically. However, if we change the assignment mechanism or the reapplication rules, acting as a weak truth-teller may lead to a payoff-relevant strategic mistake. As we do not model the latter, in our counterfactual analysis we consider two scenarios: (i) all students behave strategically[26] and (ii) a fraction $(1 - \rho)$ behaves strategically, where ρ is a lower bound on the level of sophistication.

5. IDENTIFICATION

In this section we describe our identification strategy and how we use the data described in Section 3.1 to this end.

Labor Market. As discussed in Section 3.1, we only have information about wages aggregated at the program and major levels. We identify wage equation parameters (λ) by exploiting variation across programs on students' average wages and their correlation

[25]This is a common assumption in the literature (Agarwal and Somaini, 2018, Larroucau and Rios, 2018). Larroucau et al. (2021) analyze in detail students' subjective beliefs. Although subjective beliefs are biased, beliefs are centered around rational expectation beliefs.

[26]This would hold if information policies that give precise information about admission probabilities to students were implemented.

with students' and programs' characteristics.[27] Non pecuniary labor market parameters (α^w) are identified by the correlation between student observable characteristics, their reported preferences, and graduation probabilities. As we do not have information on wages for students who dropped out of college, we model the value functions of dropping out as a function of students' observable characteristics. Intuitively, these value functions' parameters are identified by the share of students who dropped out conditional on their observable characteristics, including gender and income level.

Flow utility. The identification challenge of separately identifying the parameters that govern unobserved preferences from those related to the learning process is that both channels affect students' choices over time and are unobserved by the econometrician. However, due to the rational expectations assumption and the assumption on common prior beliefs about students' unknown abilities, students' initial application decisions are informative of their unobserved preferences, because students have not received any signal about their unknown abilities when they submit their initial applications. Hence, we can identify the flow utility parameters using students' initial choices and the correlation between students' characteristics and the characteristics of the programs they list and enroll in. In particular, to identify the major and university-type specific parameters (i.e., the parameters that govern the distributions of $\alpha_{im_j}, \alpha_{ik_j}$ in 4.3.1), we leverage the heterogeneity in terms of major and college types within students' ROLs.[28] Then, we use these values as moments to be matched in the estimation procedure. To identify the cost parameter, α_{c0}, we follow the strategy proposed by Kapor et al. (2020a) and exploit a discontinuous change in tuition generated by the scholarship *Beca Vocación de Profesor*.[29] Finally, as is standard practice, we normalize the logit shocks' scale to one, the mean utility of the outside option to zero, and further consider a discount factor β equal to 0.9.

Grades and Learning. According to Equation 4, grades are functions of observed characteristics, students' unobserved preferences for majors/colleges, students' unknown abilities, and the signal's noise. To identify the effect of unobserved preferences, we use the correlation between grades and students' preferences and their assignment, and we also use the correlation between students' application composition—the share of different majors and share of different college types—and grades. Intuitively, if students' unobserved preferences for majors positively affect their college grades, we would expect that students whose ROLs imply a high preference for a particular major—i.e., having a high share of programs that belong to the same major—should also have higher first-year grades than other students. On the other hand, to separate the impact of students' learning about their unknown abilities from the grade noise, we compare the law

[27] We fix wage growth parameters to the rates computed using SIES data.

[28] For each student, we compute the fraction of preferences belonging to each major and university type and then compute the average across students for each major and college type.

[29] Under this scholarship, students with an average score higher than 600 points can enroll an Education program without paying yearly tuition. This change in tuition generates a discontinuity in enrollment in education programs around this cutoff (see Figure D.1 in Appendix D.1), which we exploit for identification (see Kapor et al. (2020a) and Gallegos et al. (2019) for more details on the effect of this scholarship on students' enrollment).

of motion between students' first-year grades and second-year grades for switchers and non-switchers (Arcidiacono et al., 2016), and the correlation between students' first-year grades and the change in students' ROL composition for majors and college types. Formally, consider the following equation that defines student i's posterior unknown ability for program j:

$$\mu_{ij't+1} = \frac{\left(\omega_{s_{j'}}\omega_{s_j} + (1-\omega_{s_{j'}})(1-\omega_{s_j})\right)\sigma_s^2 a_{ijt} + 1_{\{m_{j'}=m_j\}}\sigma_m^2 a_{ijt}}{\sigma_g^2 + \sigma_m^2 + \left(\omega_{s_j}^2 + (1-\omega_{s_j})^2\right)\sigma_s^2},$$

where ω_{s_j} and $\omega_{s_{j'}}$ are the weights that programs j and j' use for math; a_{ijt} is the signal that student i receives from her grades in program j at time t; and σ_m^2, σ_s^2, and σ_g^2 are the variances of the major unknown ability, subject unknown ability, and the grade noise, which are the parameters of interest we want to identify. On the left-hand side of the equation, $\mu_{ij't+1}$ is the unknown ability of student i in program j' at time $t+1$. The posterior unknown ability affects students' switching and dropout decisions and their reapplications. Intuitively, if students' grades have a very low correlation with their outcomes, most of the signal is noise (high σ_g^2). On the other hand, if there is a high (negative) correlation between students' first-year grades (signals) and their switching and reapplication choices—particularly for changing majors or math types—the signal is highly informative about the unknown abilities for major (high σ_m^2) and subjects (high σ_s^2), respectively.[30] [31]

Application. We separately identify students' beliefs about admissions probabilities from their preferences by assuming rational expectations and exploiting distance as a special regressor (Agarwal and Somaini, 2018). To estimate the probability that students are either truth-tellers or strategic (see Section 4.3.3), we use the results of the survey on students' true preferences and the ROLs submitted to construct moments that allow us to identify this parameter. In particular, we use the share of students' applications for which their top-reported choice has zero admission probability. Finally, we add additional identifying information from students who reapply to college.We use the panel of repeated respondents on the 2019 and 2020 surveys and compute the share of reapplicants who report a different top-true preference for programs, majors, and college types. Since we have direct information on top-true preferences, the variation in students' responses gives us an additional information source that helps us identify students' learning.[32]

[30]The value of the signal is also affected by the effect of grades on wages (λ_3) and by the effect of the unknown ability on the non pecuniary work utility (α_4^w). These parameters directly affect switching and dropout probabilities but do not affect the signal's scale in the grade equation of the first period.

[31]The underlying identification assumption is that students' past signals (which are a function of their grades) are a sufficient statistic about how their unknown abilities affect their choices.

[32]We do not have grade information for these cohorts. Thus we can not construct correlations between students' true preferences for programs and their college grades.

Counterfactual outcomes. To identify the distribution of outcomes in the counterfactual, we leverage the variations given over initial assignments by the RDDs shown in Section 3.2. Since the first-order effect of our counterfactuals is to swap students around admission cutoffs, these variations can accurately predict those counterfactual outcomes. We then use the structure of the model to predict outcomes away from the cutoffs and to account for potential equilibrium effects that may change students' application and their initial assignment. Intuitively, these variations help us identify how strong is the initial mismatching channel and relate it to the parameters that govern unobserved persistent heterogeneity and first-time enrollment cost. In Section 6.1 we discuss how we include these variations as moments in the estimation procedure.

6. ESTIMATION

To perform the estimation, we draw a random sample of 4,000 students from the population described in Section 3.1. Moreover, we group majors into four *broad majors*—Science (Science, Farming, and Technology); Social Sciences (Social Sciences, Art and Architecture, and Law); Education and Humanities (Education and Humanities); and Health (Health)—to reduce the number of parameters to be estimated, and we consider three types of college: CRUCH-Public, CRUCH-Private and Non-CRUCH. Finally, to further facilitate the estimation, we classify programs into two types depending on their admission weights: (i) math intensive, which includes programs for which the weight on math is higher than that on verbal; and (ii) verbal intensive in the converse case. In a slight abuse of notation, we denote by s_j the type of program j, and we say that $s_j = s_m(s_v)$ if program j is math (verbal) intensive. Then, instead of considering the weights of each program, we use the average math weight among all programs that belong to the same type. As a result, the unknown ability of student i in program j becomes $A^u_{ij} = A^u_{im_j} + \bar{\omega}_{s_j} A^u_{is_m} + (1 - \bar{\omega}_{s_j}) A^u_{is_v}$, where $\bar{\omega}_{s_j}$ is the average weight on math for programs of type $s_j \in \{s_m, s_v\}$.

6.1 *Estimation Procedure*

We estimate the model parameters, θ, via Indirect Inference (II). The idea behind II is to choose a statistical model that yields a rich description of the data patterns (Bruins et al. (2018)), which allows us to identify the model parameters. This statistical model—also known as the *auxiliary* model—is estimated on both the data and on simulated data from the structural model. The II estimator minimizes an objective function that compares the distance between the estimated data parameters and the parameters estimated from the simulated data.[33] In this sense, the Simulated Method of Moments is a particular case of II, in which the *auxiliary* model is just a vector of moments. In Online Appendix E, we formally introduce the estimator, describe the estimation algorithm, and discuss the auxiliary models considered. Table 3 summarizes each set of moment conditions and their target parameters.

[33] This is known as the Wald approach to II. Other criterion functions can also be used for estimation.

TABLE 3. Estimation moments

Moment description	Targeted parameters
Share of students who retake the PSU	C^{psu}
Share of students who dropout by gender and income level	$\{\alpha_d\}_d, \alpha^w, C^e, \sigma_s^2$
Grade auxiliary models' coefficients	γ, σ_g^2
Wage auxiliary models' coefficients	λ
Switchings and dropout auxiliary models' coefficients	$\sigma_g^2, \sigma_m^2, \sigma_s^2, \alpha_4^w$
RDD auxiliary models' coefficients	$V_{\alpha^m}, V_{\alpha^k}, C^e$
Share of students who reapply	
Share of students who switch programs	$\sigma_m^2, \sigma_s^2, V_{\alpha^m}, V_{\alpha^k}, C^e$
Share of students who switch majors	σ_m^2, V_{α^m}
Share of students who switch majors within math-types	σ_m^2, V_{α^m}
Share of students who switch math-types within majors	σ_s^2
Share of students who switch college-types	V_{α^k}
Share of students who dropout at the end of the first year of college	α^w
Share of students who choose the outside option every year	α^w
Share of students who start college in the second year	
Share of students who remain in the same program after two years	
Share of top-reported preferences by program	$\{\alpha_{fe}\}_j$
Share of students whose top-reported preference is their top-true preference in R_1	ρ
Share of students whose top-reported preference is their top-true preference in R_2	ρ
Share of students whose top-reported preference has zero admission probability	ρ
Share of students with a positive risk of being unassigned given R_1	ρ
Share of ROLs R_1 with length 10	ρ
Share of ROLs R_2 with length 10	ρ
Share of students assigned to their top-true preference in the first period	ρ
Share of students who apply in the first year	
Share of students who apply in the second year	
Share of reapplications that change in their top-true preference	$\sigma_m^2, \sigma_s^2, V_{\alpha^m}, V_{\alpha^k}$
Shares of majors within R_1	V_{α^m}
Shares of college-types within R_1	V_{α^k}
Shares of majors within R_2	V_{α^m}
Shares of college-types within R_2	V_{α^k}
Norm of the difference between the vectors of college-type shares for students who reapply	V_{α^k}
Norm of the difference between the vectors of major shares for students who reapply	σ_m^2, V_{α^m}
Norm of the difference between the vectors of ω shares for students who reapply	$\sigma_s^2, V_{\alpha^m}, V_{\alpha^k}$
Correlation between first-year grades and the norm of the difference between the vectors of major shares for students who reapply	σ_m^2, σ_g^2
Correlation between first-year grades and the norm of the difference between the vectors of ω shares for students who reapply	σ_s^2, σ_g^2
Share of applications by major and college-type, grouped by gender in R_1	Δ^m, Δ^k
Share of applications by major and college-type, grouped by gender in R_2	Δ^m, Δ^k
Share reapplications from top-reported preferences	
Share reapplications from top-true preferences	
Mean of tuition for top-reported preferences, grouped by students' scores and income groups	$\{\alpha_c\}_c$
Mean of observed ability for top-reported preferences	α_1
Mean of average observed ability at the college level for top-reported preferences	α_2
Mean of distance for top-reported preferences	α_3
Mean of relative observed ability position for topreported preferences	α_4
Mean and variance of $\log\left(\frac{s_{lt+1}}{s_{lt}}\right)$ for positive PSU scores	$\{\alpha_l\}_l, \sigma_{psu}$
Mean and variance of $\log\left(\frac{s_{lt+1}}{\bar{s}_t}\right)$ for PSU scores wit zero value in the first year	$\{\alpha_{0l}\}_l, \sigma_{psu}$

6.2 *Results*

Table 4 shows the estimated parameters. We observe that the estimated share of students who apply strategically is 0.74. Thus, a significant fraction of students behave as weak truth-tellers. We also observe that the correlation and persistence of students' preferences by major are relatively high ($\sigma_\alpha^{2m} = 15.69$), considering that the variance of students' idiosyncratic preference shocks is normalized to $\pi^2/6$. Additionally, we observe that the prior variances for subject-specific abilities and major-specific abilities are given by $\sigma_s^2 = 0.48$ and $\sigma_m^2 = 0.34$, and the grade noise variance is given by $\sigma_g^2 = 0.08$, which implies a signal-to-noise-ratio of 0.91.[34] This implies that the signal carries significant information about students' unknown abilities and their match qualities with

[34]We compute the signal-to-noise-ratio as the share of the variance that is attributable to latent ability as opposed to noise, i.e, $SNR = \frac{\sigma_m^2+\sigma_s^2}{\sigma_g^2+\sigma_m^2+\sigma_s^2}$.

TABLE 4. Estimation Results - Parameters

Parameters	Values	Std
Application behavior and Dropout		
Share of strategic ROLs $(1-\rho)$	0.74	[0.022]
Cost of retaking PSU (C^{psu})	4.46	[0.219]
Dropout flow-utility for females $(\alpha^{dropout}_{female})$	19	[1.262]
Dropout flow-utility for males $(\alpha^{dropout}_{male})$	41.8	[1.756]
Dropout flow-utility for low-income $(\alpha^{dropout}_{low-income})$	15.8	[0.83]
First-time enrollment cost (C^{e})	32.16	[0.944]
Flow-utility and Priors		
Tuition (α_c)	-0.14	[0.049]
Relative position (α_4)	-0.28	[0.022]
Distance (α_3)	-1.09	[0.056]
Student observed ability (α_1)	12.92	[0.86]
Program observed ability (α_2)	4.65	[0.26]
Gender effect by major (Δ^m)	(-4.93 -2.46 3.28 1.48)	([0.363] [0.171] [0.256] [0.237])
Variance major random coefficient (σ^{2m}_{α})	15.69	[0.913]
Income effect by college (Δ^k)	(-0.11 -0.12 9.06)	([0.215] , [0.218] , [0.449])
Variance college random coefficient (σ^{2k}_{α})	0.43	[0.075]
Major prior variance (σ^2_m)	0.34	[0.032]
Subject prior variance (σ^2_s)	0.48	[0.103]
Grade equations		
Constant by major (γ_{1m_j})	(3.91 4.32 3.81 3.43)	([0.105] [0.229] [0.14] [0.208])
Student observed ability (γ_2)	0.52	[0.053]
Gender effect (γ_3)	0.36	[0.052]
Random coefficient effect on grades (major) (γ_4)	0.05	[0.015]
Grade shock variance (σ^2_g)	0.08	[0.04]
Evolution of scores		
Std. of ν (σ_{psu})	0.1	[0.007]
Mean prop. change $(\{\alpha_l\}_l)$	(1.06 1.07 1.05 1.02)	([0.004] [0.007] [0.006] [0.001])
Mean prop. change from zero score $(\{\alpha_{0l}\}_l)$	(1.07 1.08)	([0.024] [0.021])
Non-pecuniary work utility		
Major random coefficient (α^w_1)	8.72	[0.363]
Student observed ability (α^w_2)	71.58	[2.688]
College observed ability (α^w_3)	-1.86	[0.592]
Non-pecuniary work value of unknown ability (α^w_4)	178.57	[6.852]
Pecuniary work utility parameter (α^w_5)	75.95	[5.247]
Wage parameters		
Constant by major (λ_{1m_j})	(1.78 1.17 1.07 1.63)	([0.073] , [0.083] , [0.1] , [0.059])
College observed ability (λ_2)	0.03	[0.011]
Grades (λ_3)	0.13	[0.017]
Gender effects (λ_4)	-0.19	[0.094]
Wage shock variance (σ^2_w)	0.68	[0.08]
Wage growth		
Linear term by major (λ_{5m_j})	(0.11 0.18 0.14 0.24)	(-)
Quadratic term by major (λ_{6m_j})	(0 -0.01 -0.01 -0.02)	(-)

Note: The order of majors is Social Sciences, Science, Education and Humanities, and Health. The order of colleges is CRUCH-Public, CRUCH-Private, and Non-CRUCH. Standard deviations are computed via boostrap. Programs' fixed effects are available upon request.

programs. [35] However, the magnitudes of the prior variances should not be interpreted in isolation, because the signal's value is affected by the importance of the unknown ability in the non pecuniary work utility plus the effect of students' grades on their future wages. Thus, we analyze the importance of students' learning regarding their effects on outcomes in the counterfactual experiments.

To highlight the most relevant identifying variations, Table 5a shows the correlation between students' switching and their grades, which is key for identifying the effects of

[35] Notice that although students are significantly more likely to switch majors when receiving low grades than to switch math types, this does not necessarily imply that students' signals are more informative about major-specific abilities than subject-specific abilities. The reason is that dropout decisions are also negatively correlated with grades which implies that the signal carries significant information about ability components that are imperfectly transferable across different majors compared with non transferable ability components.

students' learning on outcomes. As before, we observe that most of the correlation patterns are well matched: Switching *Up* is almost uncorrelated with grades, while switching *Out* to ex ante feasible programs is negatively correlated. However, we underestimate the correlation between grades and dropouts. Table 5b shows the estimated causal effects of the RDD models. This variation is critical for identifying the role of initial mismatches and correctly predicting switching rates in the counterfactuals. We observe that we closely match these moments although we tend to over-predict the level of re applications.[36]

TABLE 5. Goodness of Fit

(a) Correlation between grades and outcomes

	Model	Data
Dropout	-0.055	-0.086
Switching programs	-0.152	-0.148
Switching broad majors	-0.092	-0.075
Switching majors	-0.172	-0.107
Switching math type	-0.079	-0.044
Switching Up	-0.008	0.002
Switching Down	-0.029	-0.032
Switching Out feasible	-0.084	-0.089
Switching Out unfeasible	-0.032	-0.011

(b) Causal effect RDDs

	Model	Data
RDD switch program 1 (level)	0.205	0.1622
RDD switch program 1 (coeff.)	-0.07	-0.0478
RDD reapplications 1 (level)	0.488	0.2261
RDD reapplications 1 (coeff.)	-0.104	-0.0840

Note: The order of colleges is CRUCH-Public, CRUCH-Private, and Non-CRUCH.

7. COUNTERFACTUALS

We now present our counterfactual analysis. As discussed in Section 1, our counterfactuals aim to evaluate whether different policies oriented to elicit cardinal preferences may help to improve students' outcomes and the system's efficiency. To accomplish this, we implement two families of counterfactuals: (i) modifying the assignment mechanism and (ii) modifying reapplication rules. We evaluate these policies in terms of different outcomes, including switching, dropout rates, reapplications, on-time graduation, etc. Moreover, for these counterfactuals, we add two measures of students' welfare: ex ante and ex post. The difference between these two measures is given by evaluating welfare before and after learning about match quality.[37] Both welfare measures are translated into millions of Chilean pesos per year of enrollment as of 2014.

Before proceeding to our primary counterfactual analysis, we analyze the extent to which students' switching and dropout decisions are explained by the two posited behavioral channels (see Appendix F.1 for details). On the one hand, by eliminating the systematic learning channel, we find that students' reapplications, switching, and dropout

[36] In Appendix E we compare all the moments and coefficients predicted with their data counterparts.

[37] Ex post utilities are computed at the end of period two, adding the discounted value function of period three—i.e., after students have made all of their choices in the model.

would decrease by 11%, 52%, and 18% from their baseline values, respectively. Moreover, eliminating learning decreases the value of being assigned to the centralized system relative to the outside option. On the other hand, if we assign every student to their top choice—thus eliminating initial mismatches—we would increase students' retention and reduce switchings by 74%. These results suggest that eliminating initial mismatches is a sensitive approach to reduce switching and increase retention rates, and in turn improving the system's yield.

7.1 *Assignment Mechanisms*

We evaluate the effects of eliciting the intensity of students' preferences by changing the assignment mechanism. In particular, we evaluate two mechanisms:

1. Constrained Deferred Acceptance (CDA): Change the constraint in the length of the ROLs, K. We evaluate $K \in \{1, 2, 3\}$, since most students submit an ROL with length less than or equal to 3.

2. Choice-Augmented Deferred Acceptance with score bonus (CADA): Students can signal one program on their submitted ROL, and receive a bonus φ in their scores related to their high school GPA. We implement this mechanism only for first-period applicants, and therefore students who apply in the second period do not receive the bonus.[38]

Both mechanisms elicit the intensity of students' preferences, since they introduce opportunity costs that students must take into account when submitting their applications. In the case of CDA, constraining the length of applicants' lists limits students in including other programs on their ROLs, and thus they must account for the opportunity cost of including each program. In the case of CADA, students can signal only one program, and thus they must carefully decide which program to target to get the bonus. However, notice that eliciting the intensity of students' preferences may not necessarily lead to higher retention. On the one hand, if eliciting this information decreases initial mismatches, we would expect to reduce inefficient switching. On the other hand, if the assignment mechanism also elicits the intensity of preferences of students who reapply to the system and these change considerably due to learning, we would see an increase in efficient switching due to a higher value of reapplication. In this sense, we expect that under CADA—which provides a score bonus only in the first period—switching would decrease more than in the case in which the score bonus is applied in both periods because the policy also gives a comparative advantage to first-period applicants, which increases switching costs through the higher equilibrium cutoff scores produced by the bonus.

[38] See Abdulkadiroğlu et al. (2015) for details. We choose to implement CADA only in the first period to avoid solving for the continuation values under this mechanism, which would add a high computational burden to the model. To implement this mechanism, we need to specify how to find the optimal ROL for each student, given their preferences and beliefs. Algorithm 3 in Appendix F describes a procedure to accomplish this.

In Table 6 we report the results of these counterfactuals. The first column includes results of the baseline model. The next three columns report the results of constrained DA considering values $K \in \{1,2,3\}$ in decreasing order, and the last three columns report the results of CADA with score bonus $\varphi \in \{10\%, 20\%, 30\%\}$.[39] First, we observe that CDA increases the fraction of reapplicants if K is sufficiently low. This result is intuitive, since reducing the maximum size of the ROLs increases the risk of being unassigned, which increases the incentive to reapply the next year. On the other hand, we observe that limiting the size of the ROL is not very effective for decreasing the overall number of switches and dropouts. Finally, we observe that ex ante and ex post welfare decrease when $K = 1$.

On the other hand, we observe that CADA effectively assigns more students to their top-true preference in the first period, which decreases initial mismatches and students' switches. Also, we observe that CADA increases the fraction of students who apply in the first period and increases the fraction of students who remain enrolled in their programs. As a result, this mechanism leads to higher persistence in programs measured by the share of students who enrolled in the same program in the second year. Furthermore, we observe that CADA considerably increases students' ex post welfare compared with both the baseline and CDA.

Finally, we observe that the overall impact of the bonus relative to the baseline is non monotonic, since welfare increases when $\varphi = 10\%$ and $\varphi = 20\%$ but it then decreases when $\varphi = 30\%$. These results suggest that the gains from learning and having the option of switching could exceed the negative externality imposed by students who switch and displace other students who may have stronger preferences for those programs (ex ante welfare losses). Moreover, these findings confirm that substantially reducing switching could also be inefficient for the system if we do not account for the gains from learning. Overall, these results suggest that CADA with a low score bonus could be a sensible policy to reduce switches and increase students' welfare.

7.2 *Reapplication Rules*

Another policy to reduce the incentive to switch is to provide bonuses to students applying for the first time to the system or to penalize students who reapply and try to switch programs; these policies have been implemented in Finland and Turkey, respectively. To our knowledge, neither of these policies has been analyzed in terms of their impact on students' outcomes. To analyze this, we consider the following two families of policies:

(i) Turkish reapplication rule: Applicants receive a penalty ψ in scores related to their high school GPA if they are currently enrolled in the centralized system.

(ii) Finnish reapplication rule: Applicants receive a bonus φ in scores related to their high school GPA the first time they submit an ROL to the centralized system.

[39] In Appendix F.3, we show supporting evidence that a significant fraction of students would change their application lists strategically when facing a binding constraint on the length of application lists.

TABLE 6. Results Counterfactuals - Mechanisms

Outcome	Baseline	Constrained DA $K=3$	Constrained DA $K=2$	Constrained DA $K=1$	CADA $\varphi=10\%$	CADA $\varphi=20\%$	CADA $\varphi=30\%$
Reapplicants [%]	34.27	0.35	1.62	10.01	-11.80	-20.92	-26.32
Program switchings [%]	6.48	-0.40	0.66	20.74	-22.28	-32.84	-39.10
Retakes PSU [%]	21.62	0.44	3.05	16.34	-23.30	-34.70	-40.48
Dropouts [%]	7.90	-0.19	-1.14	-6.83	2.84	3.73	4.56
Dropouts - first year [%]	3.70	-0.54	-1.48	-11.76	11.61	16.88	20.41
Applicants in first period [%]	62.24	0.06	0.33	1.23	1.04	1.73	2.24
Enrolls same program [%]	31.64	-0.13	-0.98	-12.14	7.20	10.61	12.95
Assigned to top true preference [%]	10.46	0.76	2.38	-9.59	16.20	22.60	23.84
Unassigned in first period [%]	44.17	0.33	1.09	9.69	-4.26	-6.30	-7.79
Graduate late [%]	95.04	0.01	0.02	0.41	-0.06	-0.20	-0.18
Difference in Ex Ante Welfare Relative to Baseline (in millions of Chilean pesos)							
Overall	-	0.03	-0.01	-1.14	0.47	0.60	0.54
Difference in Ex Post Welfare Relative to Baseline (in millions of Chilean pesos)							
Overall	-	0.01	-0.08	-1.95	0.62	0.77	0.78

Note: Percentage of change relative to the baseline. Switching and dropout rates are computed with respect to the total sample of participants.

Even though both policies aim to reduce the incentives for switching, they affect students' applications and reapplications in different ways. On the one hand, the Finnish policy directly reduces the incentives to reapply to the system, regardless of the programs students include on their ROLs. As a result, the Finnish policy increases the continuation value of choosing the outside option, and thus increases the fraction of students who wait an extra year to submit their first application. On the other hand, the Turkish policy reduces the incentives to reapply if students where previously enrolled in a program in the system—i.e., it reduces the incentives to apply to programs if they are very likely to switch from them in the future (e.g., programs for which students have low preference intensity). Hence, the Turkish policy may decrease the fraction of students who enroll in the first period in less preferred programs. Despite these differences, we expect that both policies would decrease the frequency of reapplications and switches. In contrast, the welfare effects of these policies is unclear. Students may benefit from these policies, since both the penalty and the bonus help to address the negative externality that switchers generate in the system. However, since under these policies students face more barriers to switching, the benefits of learning become lower, and thus students' welfare may decrease.

In Table 7 we report the results of these counterfactuals. As expected, we observe that the Turkish policy elicits the intensity of students' preferences, assigns more students to their top-true preferences in the first period, and reduces reapplication and switching rates, and the magnitude of the effect is increasing in the magnitude of the penalty. Moreover, we observe that dropout rates for the first year slightly increase as we increase the penalty. A potential explanation for this is that the Turkish policy increases switching costs. Thus, students who receive low signals about their match qualities with their enrolled programs face lower probabilities for switching than the baseline, increasing their incentives to drop out instead. Finally, we observe that welfare increases compared with the baseline as we increase the penalty.

TABLE 7. Results Counterfactuals - Re-Application Rules

		Turkish Rules			Finnish Rules		
Outcome	Baseline	$\psi = 10\%$	$\psi = 20\%$	$\psi = 30\%$	$\varphi = 10\%$	$\varphi = 20\%$	$\varphi = 30\%$
Reapplicants [%]	34.27	-16.81	-29.63	-36.41	-23.84	-34.83	-40.10
Program switchings [%]	6.48	-33.16	-51.53	-63.34	-28.67	-40.31	-46.56
Retakes PSU [%]	21.62	-18.18	-27.79	-32.95	-17.23	-24.34	-25.04
Dropouts [%]	7.90	0.50	0.50	0.32	-0.44	-1.52	-1.96
Dropouts - first year [%]	3.70	4.22	5.70	6.92	0.82	-0.53	-2.29
Applicants in first period [%]	62.24	-0.49	-0.72	-0.79	-7.46	-9.18	-10.45
Enrolls same program [%]	31.64	5.90	9.07	11.17	4.39	5.40	5.46
Assigned to top true preference [%]	10.46	13.39	19.80	21.95	19.51	28.26	29.62
Unassigned in first period [%]	44.17	0.27	0.59	0.72	1.26	2.86	4.27
Graduate late [%]	95.04	-0.22	-0.30	-0.38	-0.13	-0.24	-0.23
Difference in Ex Ante Welfare Relative to Baseline (in millions of Chilean pesos)							
Overall	-	0.52	0.70	0.76	0.48	0.59	0.50
Difference in Ex Post Welfare Relative to Baseline (in millions of Chilean pesos)							
Overall	-	0.45	0.65	0.68	0.37	0.43	0.29

Note: Percentage of change relative to the baseline. Switching and dropout rates are computed with respect to the total sample of participants.

On the other hand, we observe that the Finnish policy has a similar effect on students' outcomes, but the magnitude of the effect varies relative to the Turkish policy. The Finnish policy is better for assigning students to their top-true preference in the first period, which significantly reduces initial mismatches. In addition, the Finnish policy increases the fraction of students unassigned in the first period and the fraction of students who decide to delay their tertiary education entry, even for low values of the subsidy. An explanation for this is that students who know their preferences but do not have good enough scores in the first period are better off waiting a year to retake the exams and improve their application score instead of enrolling in the first year and trying to switch later. These results suggest that both policies can be effective for increasing the system's yield, and thus reducing the congestion externality of initial mismatches. Figure 5a shows a summary of the counterfactual results.[40] Overall, these results suggest that the most desired policy depends on the objective to be addressed. If the goal is to decrease switching and improve the system's yield, the Turkish policy seems to be the best option. In contrast, if the goal is to increase students' welfare, then CADA leads to better outcomes. In summary, CADA and the two reapplication rules can increase students' welfare and the system's yield. However, further constraining the length of application lists does not seem to be an effective policy for these objectives.

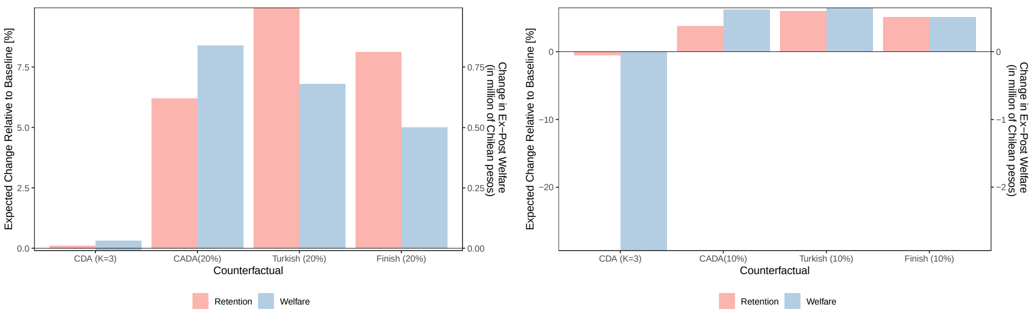
7.3 *Sensitivity to Non Strategic Students.*

It is important to highlight that the counterfactual analyses described above assume that all students are strategic—i.e., their beliefs take into account potential changes in the

[40] For Figures 5a and 5b we compute retention while considering switching and first-year dropouts.

FIGURE 5. Summary of Counterfactuals

(a) Strategic (b) Mixture



distribution of cutoffs induced by the different policies.[41] This assumption is reasonable if precise information about admission probabilities is provided to students. However, many students in practice are not strategic and report their true preferences. For this reason, we conduct the same analysis, assuming that 26% of students are non strategic—similar to the estimation results in the baseline model—as a robustness check. Although we find that the results are directionally the same, the magnitude of the effects changes significantly. Figure 5b shows a summary of the results. In particular, we observe that the overall ex post welfare under CDA decreases as we make the constraint on the length of ROLs more binding. Overall, both reapplication rules and CADA are more robust to deal with students who may not report their preferences strategically compared to constraining the length of their lists, which is a common policy used worldwide to elicit intensity of preferences.

The previous results motivated the Ministry of Education of Chile to relax the constraint on the length of application lists for the 2023 admissions process and propose a policy evaluation for 2024.

8. CONCLUSIONS

In this paper, we analyze the effects of centralized assignment mechanisms on downstream outcomes such as students' decisions to switch or drop out of college. To accomplish this, we study the relevance of eliciting information on participants' cardinal preferences and incorporating their dynamic incentives in the design of the assignment process—features that have been mostly overlooked in the literature.

Using data from the Chilean college admissions system and two nationwide surveys that we designed and conducted, we provide empirical evidence that suggests that two central behavioral channels explain students' dynamic decisions. The first channel, called the *initial mismatch* channel, predicts that students may have incentives

[41]To accomplish this, we follow an approach similar to that of Kapor et al. (2020b). However, our case differs from theirs in that (i) we solve for a stationary distribution in the dynamic application problem, creating a mixture of applicants and reapplicants who participate in the same admission process, and (ii) students need to form believes over a large set of cutoff distributions. Algorithm 2 in Appendix F describes the algorithm to estimate students' equilibrium beliefs over the cutoff distributions.

to switch programs if they were initially assigned to less desired preferences. The second channel, called the *learning* channel, suggests that students may learn about their match qualities during their college progression, and thus may decide to switch to programs in which their match qualities—and their expected outcomes in the labor market—are higher.

Given these findings, we introduce a structural model that captures students' decisions during their academic progression, which allows them to learn about their match quality from their grades. We use the estimated model to evaluate changes in the reapplications rules—by implementing those used in Turkey and Finland—and the assignment mechanism—by adding further constraints on the length of lists and adding the option for students to signal one of the programs in their preferences to obtain a score bonus. Our results show that these reapplication rules and the signaling mechanism are both effective for increasing college retention rates and, at the same time, increasing students' welfare. Moreover, these effects are robust to changes in the fraction of participants who behave strategically, as opposed to other approaches such as constraining the length of the lists. However, lack of sophistication in students' ranking strategies undermines the effectiveness of these policies, which stress the importance of giving students correct information about their admission probabilities and helping them in choosing optimal application lists.

Overall, our results show that incorporating dynamic incentives and eliciting information on participants' cardinal preferences can significantly increase students' welfare and improve downstream outcomes. These insights can be informative for improving the design of many matching markets that exhibit similar features. For instance, in entry-level labor markets, employers care about turnover, and agents may have private information about their preferences, learn about their match qualities through experience, and face dynamic considerations, such as deciding when to enter the market (apply), re-enter (reapply), exit (dropout), or rematch (switch). Our key insight is that market designers should correctly balance the gains from learning through experimentation and the crowd-out externality produced by initial mismatches to improve the efficiency of these markets.

REFERENCES

ABDULKADIROĞLU, ATILA, YEON KOO CHE, AND YOSUKE YASUDA (2015): "Expanding "choice" in school choice," *American Economic Journal: Microeconomics*, 7, 1–42. [26]

ABDULKADIROĞLU, ATILA, PARAG A PATHAK, AND ALVIN E ROTH (2005a): "The New York City high school match," *American Economic Review*, 95, 364–367. [2]

ABDULKADIROĞLU, ATILA, PARAG A PATHAK, ALVIN E ROTH, AND TAYFUN SÖNMEZ (2005b): "The Boston public school match," *American Economic Review*, 95, 368–371. [2]

AGARWAL, NIKHIL, ITAI ASHLAGI, MICHAEL REES, PAULO SOMAINI, AND DANIEL WALDINGER (2021): "Equilibrium Allocations Under Alternative Waitlist Designs: Evidence from Deceased Donor Kidneys," *Econometrica*, 89, 37–76. [4, 5]

AGARWAL, NIKHIL AND PAULO SOMAINI (2018): “Demand Analysis using Strategic Reports: An application to a school choice mechanism,” *Econometrica*, 86, 391–444. [4, 19, 21]

——— (2019): “Revealed Preference Analysis of School Choice Models,” *Annual Review of Economics*, 12. [4]

ALLENDE, CLAUDIA (2019): “Competition under social interactions and the design of education policies,” . [16]

ALTONJI, J. G., P. ARCIDIACONO, AND A. MAUREL (2016): “The Analysis of Field Choice in College and Graduate School. Determinants and Wage Effects,” in *Handbook of the Economics of Education*, Elsevier, vol. 5, 305–396. [5]

ALTONJI, JOSEPH G, ERICA BLOM, AND COSTAS MEGHIR (2012): “Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers,” *Annu. Rev. Econ*, 4, 185–223. [5]

ARCIDIACONO, PETER (2004): “Ability sorting and the returns to college major,” *Journal of Econometrics*, 121, 343–375. [14]

——— (2005): “Affirmative action in higher education: How do admission and financial aid rules affect future earnings?” *Econometrica*, 73, 1477–1524. [5, 14]

ARCIDIACONO, PETER, ESTEBAN AUCEJO, ARNAUD MAUREL, AND TYLER RANSOM (2016): “College attrition and the dynamics of information revelation,” Tech. rep., National Bureau of Economic Research. [5, 21]

BORDON, PAOLA AND CHAO FU (2015): “College-major choice to college-then-major choice,” *Review of Economic Studies*, 82, 1247–1288. [5, 16]

BRUINS, MARIANNE, JAMES A DUFFY, MICHAEL P KEANE, AND ANTHONY A SMITH JR (2018): “Generalized indirect inference for discrete choice models,” *Journal of Econometrics*, 205, 177–203. [22]

CALSAMIGLIA, CATERINA, CHAO FU, AND MAIA GÜELL (2020): “Structural estimation of a model of school choices: The Boston mechanism versus its alternatives,” *Journal of Political Economy*, 128, 642–680. [4]

CARD, DAVID AND ALEX SOLIS (2020): “Measuring the Effect of Student Loans on College Persistence,” *Education Finance and Policy*, 1–60. [16]

CHADE, HECTOR AND LONES SMITH (2006): “Simultaneous Search,” *Econometrica*, 74, 1293–1307. [18]

CORREA, JOSE, RAFAEL EPSTEIN, JUAN ESCOBAR, IGNACIO RIOS, BASTIAN BAHAMONDES, CARLOS BONET, NATALIE EPSTEIN, NICOLAS ARAMAYO, MARTIN CASTILLO, ANDRES CRISTI, AND ET AL. (2021): “School Choice in Chile,” *Operations Research*, 70, 1066–1087. [2]

GALLEGOS, SEBASTIAN, CHRISTOPHER NEILSON, FRANCO CALLE, ET AL. (2019): *Screening and Recruiting Talent At Teacher Colleges Using Pre-College Academic Achievement*, Princeton University, Industrial Relations Section. [20]

HAERINGER, GUILLAUME AND FLIP KLIJN (2009): "Constrained school choice," *Journal of Economic Theory*, 144, 1921–1947. [9]

HE, YINGHUA (2012): "Gaming the Boston School Choice Mechanism in Beijing," *Manuscript, Toulouse School of Economics*, 2012. [4]

KAPOR, ADAM, MOHIT KARNANI, AND CHRISTOPHER NEILSON (2020a): "Aftermarket Frictions and the Cost of Off-Platform Options in Centralized Assignment Mechanisms," . [1, 20]

KAPOR, ADAM J, CHRISTOPHER A NEILSON, AND SETH D ZIMMERMAN (2020b): "Heterogeneous beliefs and school choice mechanisms," *American Economic Review*, 110, 1274–1315. [4, 30]

KRISHNA, KALA, SERGEY LYCHAGIN, AND VERONICA FRISANCHO (2018): "Retaking in High Stakes Exams: Is Less More?" *International Economic Review*, 59, 449–477. [2]

LARROUCAU, TOMÁS, MANUEL MARTÍNEZ, CHRISTOPHER NEILSON, AND IGNACIO RIOS (2021): "Application Mistakes and Information frictions in College Admissions," . [19]

LARROUCAU, TOMAS AND IGNACIO RIOS (2018): "Do "Short-List" Students Report Truthfully? Strategic Behavior in the Chilean College Admissions Problem," . [9, 16, 19]

LOCHNER, LANCE AND ALEXANDER MONGE-NARANJO (2012): "Credit Constraints in Education," *Annual Review of Economics*, 4, 225–256. [16]

——— (2016): "Student loans and repayment: Theory, evidence, and policy," in *Handbook of the Economics of Education*, Elsevier, vol. 5, 397–478. [16]

MALAMUD, OFER (2011): "Discovering one's talent: Learning from academic specialization," *Industrial and Labor Relations Review*, 64, 375–405. [6]

NARITA, YUSUKE (2018): "Match or Mismatch? Learning and Inertia in School Choice," *SSRN Electronic Journal*. [2, 5]

RIOS, IGNACIO, TOMÁS LARROUCAU, GIORGIOGIULIO PARRA, AND ROBERTO COMINETTI (2021): "Improving the Chilean College Admissions System," *Operations Research*, 69, 1186–1205. [7]

SOLIS, ALEX (2017): "Credit Access and College Enrollment," *Journal of Political Economy*, 125, 562–622. [16]

STINEBRICKNER, RALPH AND TODD STINEBRICKNER (2008): "The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study," *American Economic Review*, 98, 2163–84. [9]

TANAKA, MARI, YUSUKE NARITA, AND CHIAKI MORIGUCHI (2020): "Meritocracy and Its Discontent: Long-run Effects of Repeated School Admission Reforms," . [2, 5]

ZHANG, JUANJUAN (2010): "The sound of silence: Observational learning in the US kidney market," *Marketing Science*, 29, 315–335. [4]

Co-editor [Name Surname; will be inserted later] handled this manuscript.

Online Appendix for Dynamic College Admissions

APPENDIX A: MOTIVATING EXAMPLE

We first analyze whether it is—theoretically—possible to increase aggregate students' welfare and increase the system's yield by changing the assignment mechanism and reapplication rules. Furthermore, we provide intuition on how switching behavior can be affected by the assignment mechanism in a dynamic setting.

If students face uncertainty over their admission chances, either because of uncertainty about admission cutoffs or their future application scores, switching can endogenously occur over time. Since students do not know their expost choice sets, they could choose to enroll in a program in the first year and switch in the following year to a more preferred program if their choice set allows them to. Moreover, if students are uncertain about their match quality with programs, and after enrollment learn about their preferences/abilities, they could choose to switch programs or drop out to avoid expost mismatches. Regardless of which mechanism dominates, individual switching and dropouts impose an externality on universities and on other students. Given the sequential nature of colleges' academic progression, when a student switches at the end of the academic year, the resulting vacancy is lost for the next year and, in the absence of a proper transfer system that allows students to switch at different stages of their college progression, this vacancy cannot be reallocated to another student.

To illustrate how switches may arise endogenously, consider a centralized college admissions problem with reapplications and two periods. Let $S = \{A, B\}$ and $C = \{1, 2\}$ be the sets of students and colleges, respectively. We incorporate uncertainty on admissions by assuming that colleges post their first-year vacancies after students submit their applications. For simplicity, we assume that each college offers one seat with probability 1/2 and no seats otherwise in each period. In addition, we assume that the preferences of colleges are

$$B \succ_1 A, \quad A \succ_2 B,$$

i.e., college 1 prefers student B over A, and college 2 prefers A over B. Finally, we assume that colleges care about students' persistence and incur a cost τ per student who does not remain enrolled. This cost captures the idea that colleges make investments in their students and that the vacancy (and corresponding future tuition payments) is lost when students switch.

On the other side of the market, we assume that students are expected utility maximizers, i.e., they submit a preference list that maximizes their expected utility conditional on their preferences and beliefs about admission probabilities. We assume that the utility of student i in college j is given by

$$v^i_j = u^i_j + \xi^i_j,$$

where u_j^i is known exante and such that $u_1^A \succ u_2^A \gg 0$ and $u_2^B \succ u_1^B \gg 0$. ξ_j^i is unknown exante but learned after the first year. We assume that this random component is distributed according to

$$\xi_j^i = \begin{cases} l & \text{with probability } p \\ -l & \text{with probability } (1-p) \end{cases},$$

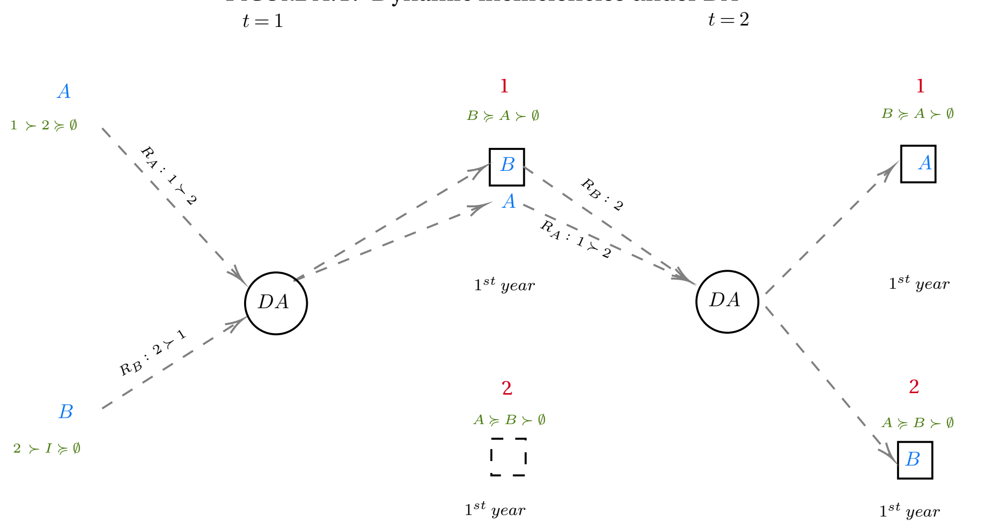
for each student i and college j, and we assume that this distribution is commonly known exante. We make two additional assumptions regarding the component of the utility that is learned: (i) $u_A^1 - l < u_B^1$ ($u_B^2 - l < u_A^1$)—i.e., if students learn that they do not have a high match quality with their current college, they prefer to switch to the other college; and (ii) $u_A^2 - l \succ 0$ ($u_B^1 - l \succ 0$)—i.e., students prefer to enroll in their assigned colleges over the outside option. Then, each student i chooses the ROL R_i^t that maximizes their expected utility in each period t.

Notice that, depending on the mechanism and the reapplication rules, students may submit different ROLs, which in turn may affect their assignment and their academic progression. To illustrate this, we compare the outcomes of two alternative mechanisms: (i) Deferred Acceptance (DA), whereby students can apply to as many colleges as they want; and (ii) DA with no switches (DA-NS); i.e., once students are admitted they cannot reapply and switch to another college.

DA. Under DA, both students apply according to their true preferences, i.e., $R_A^1 = 1 \succ 2$ and $R_B^1 = 2 \succ 1$. Then, when only one college opens a seat, we observe that both students compete for it, and the student the college prefers the most is assigned and the other remains unassigned. However, notice that students are then assigned to their second preference, so they may have incentives to reapply in the second period and switch to their exante top preference, depending on the realized value of the unknown component of their utility. By doing so, students impose a cost on colleges, and impose a crowd-out externality on the other student, since the latter would benefit from getting assigned to their most desired option in the first period. This situation is illustrated in Figure A.1, where we show the case when only college 1 opens a seat.

DA-NS. When no switching is allowed, students still report their true preferences when they apply to the system. However, when they learn that their match quality with their college is poor, they cannot reapply and switch. This reapplication rule introduces a trade-off relative to DA. On the one hand, it reduces switches, which eliminates the cost paid by colleges. In addition, by eliminating competition in the second period, DA-NS increases the probability that unassigned students in the first period will be admitted to their top preference in the second period. On the other hand, preventing students from switching imposes a higher cost if their match quality with their initial college is poor. Hence, it is unclear which mechanism leads to higher aggregate welfare. We formalize this result in Proposition 1 and defer the proof to Appendix A.

FIGURE A.1. Dynamic inefficiencies under DA



PROPOSITION 1. *The difference between the aggregated exante welfare generated by DA-NS relative to DA is given by*

$$\Delta^{DA-NS} = \underbrace{\frac{3\cdot\beta\cdot(1-p)}{8}\cdot\tau}_{\textit{Higher Retention}} + \underbrace{\frac{\beta\cdot(1-p)}{4}\cdot\left[\frac{(u_1^A-u_2^A)+(u_2^B-u_1^B)}{4}\right]}_{\textit{Improvement for first-year unassigned students}} - \underbrace{\frac{3\cdot\beta\cdot(1-p)}{8}\cdot l}_{\textit{Less switches after learning}} .$$

The right-hand side of (1) illustrates this trade-off. The first term captures the lower cost for universities, since switches disappear under DA-NS. The second term captures the increase in students' welfare due to the higher chances of assignment to their top preference. Finally, the last term captures the negative effect of not switching when students learn that their match quality with the college is poor. Then, depending on the relative magnitude of these three components, either reapplication rule may be better.

Notice that switching can endogenously arise due to the two behavioral channels: (i) initial mismatches and (ii) learning. Identifying the prevalence of each channel is an empirical question; it is also our main identification challenge and a relevant question since both channels have different consequences in the design of markets. On the one hand, if students' preferences are persistent over time, it may be desirable to restrict reapplication and force students to internalize the negative externality they impose on other students and colleges. On the other hand, if most of the switches are due to students' learning about their match-quality, it may be welfare-improving to facilitate switching behavior to avoid expost mismatches. Hence, the welfare implications are unclear.

PROOF. Proof of Proposition 1

DA. Under DA, students apply to all schools. Then, in the first period, the expected utility is

$$\frac{1}{4}\cdot 0+\frac{1}{4}\cdot\left(u_1^A+u_2^B\right)+\frac{1}{4}\cdot u_2^A+\frac{1}{4}\cdot u_1^B.$$

In the second period, the value depends on the realized assignment in the first period, all of which happens with probability $1/4$:

1. If $\mu=((A,\emptyset),(B,\emptyset))$, then the second-period expected utility is
$$\frac{1}{4}\cdot 0+\frac{1}{4}\cdot\left(u_1^A+u_2^B\right)+\frac{1}{4}\cdot u_2^A+\frac{1}{4}\cdot u_1^B.$$
2. If $\mu=((A,1),(B,2))$, there are four scenarios depending on the signals observed by the two students. More specifically, let $\xi=\left(\xi^A,\xi^B\right)$ be the signals observed by the end of period 1. Then,
   - If $\xi=(l,l)$, which happens with probability p^2, both students remain enrolled. Then the expected utility in the second period is $u_1^A+u_2^B+2l$.
   - If $\xi=(l,-l)$, which happens with probability $p\cdot(1-p)$, B reapplies and switches with probability 1/2 and remains in 2 otherwise. Then the expected utility in the second period is $u_1^A+l+\frac{1}{2}\cdot\left(u_2^B-l+u_1^B\right)-\frac{\tau}{2}$.
   - If $\xi=(-l,l)$, which happens with probability $p\cdot(1-p)$, A reapplies and switches with probability 1/2 and remains in 1 otherwise. Then the expected utility in the second period is $u_2^B+l+\frac{1}{2}\cdot\left(u_1^A-l+u_2^A\right)-\frac{\tau}{2}$.
   - If $\xi=(-l,-l)$, which happens with probability $(1-p)^2$, both students reapply. Then the expected utility in the second period is $\frac{1}{4}\cdot\left(u_1^A+u_2^B-2l\right)+\frac{1}{4}\cdot\left(u_2^A+u_2^B-l\right)+\frac{1}{4}\cdot\left(u_1^A+u_1^B-l\right)+\frac{1}{4}\cdot\left(u_2^A+u_1^B\right)-\tau$.
3. If $\mu=((A,2),(B,\emptyset))$, only A learns, and thus there are two scenarios:
   - If $\xi^A=l$, which happens with probability p, then A stays and B reapplies. Thus, the expected utility is $u_2^A+l+\frac{1}{2}\cdot u_2^B+\frac{1}{4}\cdot u_1^B$.
   - If $\xi^A=-l$, which happens with probability $1-p$, then A and B reapply. Then the expected utility in the second period is $\frac{1}{4}\cdot\left(u_2^A-l\right)+\frac{1}{4}\cdot\left(u_2^A+u_1^B-l\right)+\frac{1}{4}\cdot\left(u_2^A+u_2^B-l\right)+\frac{1}{4}\cdot\left(u_1^A+u_2^B\right)-\frac{1}{4}\cdot\tau$.
4. If $\mu=((A,\emptyset),(B,1))$, only B learns, and thus there are two scenarios:
   - If $\xi^B=l$, which happens with probability p, then A reapplies and B stays. Thus, the expected utility is $u_1^B+l+\frac{1}{2}\cdot u_1^A+\frac{1}{4}\cdot u_2^A$.
   - If $\xi^B=-l$, which happens with probability $1-p$, then A and B reapply. Then the expected utility in the second period is $\frac{1}{4}\cdot\left(u_1^B-l\right)+\frac{1}{4}\cdot\left(u_1^B+u_1^A-l\right)+\frac{1}{4}\cdot\left(u_1^B+u_2^A-l\right)+\frac{1}{4}\cdot\left(u_1^A+u_2^B\right)-\frac{1}{4}\cdot\tau$.

DA-NS. Under DA-NS, the assignment is performed using DA but students are not allowed to switch in the second period. Then,
$$\frac{1}{4}\cdot 0+\frac{1}{4}\cdot\left(u_1^A+u_2^B\right)+\frac{1}{4}\cdot u_2^A+\frac{1}{4}\cdot u_1^B.$$

In the second period, the value depends on the realized assignment in the first period, all of which happens with probability $1/4$:

1. If $\mu = ((A, \emptyset), (B, \emptyset))$, then the second-period expected utility is

$$\frac{1}{4} \cdot 0 + \frac{1}{4} \cdot \left(u_1^A + u_2^B\right) + \frac{1}{4} \cdot u_2^A + \frac{1}{4} \cdot u_1^B.$$

2. If $\mu = ((A, 1), (B, 2))$, there are four scenarios, as in the case for DA:
   - If $\xi = (l, l)$, which happens with probability p^2, the expected utility in the second period is $u_1^A + u_2^B + 2l$.
   - If $\xi = (l, -l)$, which happens with probability $p \cdot (1-p)$, the expected utility in the second period is $u_1^A + u_2^B$.
   - If $\xi = (-l, l)$, which happens with probability $p \cdot (1-p)$, the expected utility in the second period is $u_1^A + u_2^B$.
   - If $\xi = (-l, -l)$, which happens with probability $(1-p)^2$, the expected utility in the second period is $u_1^A + u_2^B - 2l$.
3. If $\mu = ((A, 2), (B, \emptyset))$, only A learns, and thus there are two scenarios:
   - If $\xi^A = l$, which happens with probability p, the expected utility is $u_2^A + l + \frac{1}{2} \cdot u_2^B + \frac{1}{4} \cdot u_1^B$.
   - If $\xi^A = -l$, which happens with probability $1-p$, the expected utility in the second period is $u_2^A - l + \frac{1}{2} \cdot u_2^B + \frac{1}{4} \cdot u_1^B$.
4. If $\mu = ((A, \emptyset), (B, 1))$, only B learns, and thus there are two scenarios:
   - If $\xi^B = l$, which happens with probability p, the expected utility is $u_1^B + l + \frac{1}{2} \cdot u_1^A + \frac{1}{4} \cdot u_2^A$.
   - If $\xi^B = -l$, which happens with probability $1-p$, the expected utility in the second period is $u_1^B - l + \frac{1}{2} \cdot u_1^A + \frac{1}{4} \cdot u_2^A$.

Then, taking the difference for each scenario, we obtain no differences in the expected utility in the first period. For the second period we obtain that:

1. If $\mu = ((A, \emptyset), (B, \emptyset))$, there is no difference between DA and DA-NS.
2. If $\mu = ((A, 1), (B, 2))$, there are four scenarios as in the case for DA:
   - If $\xi = (l, l)$, then the difference in expected utility is zero.
   - If $\xi = (l, -l)$, then the difference in expected utility is

$$\begin{aligned} & u_1^A + l + \frac{1}{2} \cdot \left(u_2^B - l + u_1^B\right) - \frac{\tau}{2} - \left(u_1^A + u_2^B\right) \\ &= \frac{u_2^B - u_1^B}{2} + \frac{l - \tau}{2}. \end{aligned}$$

- If $\xi = (-l, l)$, then the difference in expected utility is

$$u_2^B + l + \frac{1}{2} \cdot \left(u_1^A - l + u_2^A\right) - \frac{\tau}{2} - \left(u_1^A + u_2^B\right)$$
$$= \frac{u_1^A - u_2^A}{2} + \frac{l - \tau}{2}.$$

- If $\xi = (-l, -l)$, then the difference in expected utility is

$$\frac{1}{4} \cdot \left(u_1^A + u_2^B - 2l\right) + \frac{1}{4} \cdot \left(u_2^A + u_2^B - l\right) + \frac{1}{4} \cdot \left(u_1^A + u_1^B - l\right)$$
$$+ \frac{1}{4} \cdot \left(u_2^A + u_1^B\right) - \tau - \left(u_1^A + u_2^B - 2l\right)$$
$$= \frac{u_1^A - u_2^A}{2} + \frac{u_2^B - u_1^B}{2} + l - \tau.$$

3. If $\mu = ((A, 2), (B, \emptyset))$, only A learns, and thus there are two scenarios:
   - If $\xi^A = l$, then the difference in expected utility is zero.
   - If $\xi^A = -l$, then the difference in expected utility is

$$\frac{1}{4} \cdot \left(u_2^A - l\right) + \frac{1}{4} \cdot \left(u_2^A + u_1^B - l\right) + \frac{1}{4} \cdot \left(u_2^A + u_2^B - l\right) + \frac{1}{4} \cdot \left(u_1^A + u_2^B\right) - \frac{1}{4} \cdot \tau$$
$$- \left(u_2^A - l + \frac{1}{2} \cdot u_2^B + \frac{1}{4} \cdot u_1^B\right)$$
$$= \frac{u_1^A - u_2^A}{4} + \frac{l - \tau}{4}.$$

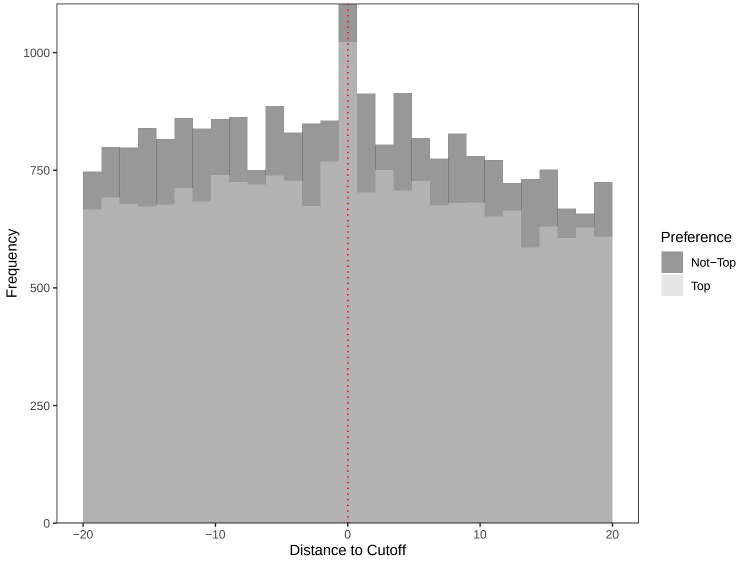
4. If $\mu = ((A, \emptyset), (B, 1))$, only B learns, and thus there are two scenarios:
   - If $\xi^B = l$, then the difference in expected utility is zero.
   - If $\xi^B = -l$, then the difference in expected utility is

$$\frac{1}{4} \cdot \left(u_1^B - l\right) + \frac{1}{4} \cdot \left(u_1^B + u_1^A - l\right) + \frac{1}{4} \cdot \left(u_1^B + u_2^A - l\right) + \frac{1}{4} \cdot \left(u_1^A + u_2^B\right) - \frac{1}{4} \cdot \tau$$
$$- \left(u_1^B - l + \frac{1}{2} \cdot u_1^A + \frac{1}{4} \cdot u_2^A\right)$$
$$= \frac{u_2^B - u_1^B}{4} + \frac{l - \tau}{4}.$$

Finally, multiplying by the corresponding probabilities and adding up terms, we obtain that:

$$DA - DA - NS = \frac{\beta \cdot (1 - p)}{4} \cdot \left(\frac{u_1^A - u_2^A}{4} + \frac{u_2^B - u_1^B}{4} + \frac{3}{2} \cdot (l - \tau)\right).$$

FIGURE A.2. Distribution of preference of assignment around admission cutoffs



Therefore,

$$\Delta^{DA-NS} = DA-NS-DA = \frac{\beta\cdot(1-p)}{4}\cdot\left(\frac{u_2^A-u_1^A}{4}+\frac{u_1^B-u_2^B}{4}+\frac{3}{2}\cdot(\tau-l)\right).$$

□

Notice that these theoretical examples assume that we can find students like A and B in the data—that is, students with similar application scores but different assignment preferences. Figure A.2 shows the distribution of preference of assignment around admission cutoffs. We observe that a significant fraction of students assigned just above admission cutoffs do not rank those programs as their top choices.

APPENDIX B: APPENDIX TO SECTION 3

B.1 *The Chilean Mechanism*

The Chilean mechanism is a variant of the student-proposing Deferred Acceptance algorithm[1] in which students who tie for the last seat of a program are not rejected if vacancies are exceeded. More formally, the allocation rule can be described as follows:

Step 1. Each student proposes to their 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.

[1]Before 2014; the algorithm used was the university-proposing version. The assignment differences between both implementations of the algorithm are negligible Rios et al. (2021).

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

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

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

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.

B.2 *Regression discontinuities*

This section provides causal evidence that the preference of assignment affects different outcomes of interest. We use a regression discontinuity design that exploits the algorithm's cutoff structure to perform the allocation. As a result of the assignment process, each program gets a cutoff such that all students whose weighted score is above it are granted admission, whereas all students with scores below the cutoff are wait-listed and thus may have to enroll in a lower-ranked preference. Hence, if we assume that students around the cutoff are similar and only differ in their right to enroll in a higher preference, we can estimate the causal effect of interest.

Formally, we estimate the effect of being assigned to a higher preference using the following specification:

$$y_{bp} = f_p(d_{bp}) + \delta_p \cdot Z_{bp} + \epsilon_{bp},$$

where y_{bp} is the average outcome of interest for students in bin of distance b applying to preference p; f_p is a smooth function of the distance d_{bp} between the bin's score and the cutoff of their preference p; Z_{bp} is an indicator function equal to 1 if bin b's score is greater than or equal to the cutoff of their p-th preference and 0 otherwise; and ϵ_{bp} is an error term.[2]

Notice that many of the outcomes we consider—e.g., switches and dropouts, among others—rely on students enrolling in the centralized system. If there are significant differences in the enrollment rates between students right above and below the cutoff, then the two samples would not be directly comparable. In that case, there would be a selection problem, and thus we would not be able to point estimate the causal effect of the preference of assignment on the outcomes of interest (Dong, 2017). To show that this is

[2]Similar results are obtained running these models at the student-preference level. We report the results at bin-preference level to match the plots included.

not the case, in Figure B.1b we show the binned means of the probability of enrolling in the centralized system as a function of distance to the cutoff. In addition, the line represents the predicted values obtained from estimating the regression discontinuity model described in (B.2), considering as dependent variable the probability of enrolling in the centralized system. As Figure B.1b and column (1) in Table B.1 show, there are no significant differences in the enrollment probabilities between students above and below the cutoff, so we conclude that the potential selection problem is not a concern in our case.

To assess the causal effect of the preference of assignment on other outcomes, we focus on students who applied to at least two programs in the centralized system, and we restrict the analysis to the top preference of each student for simplicity.[3] In Figure 2 we display binned means of different outcomes as a function of the distance between the cutoffs in their top preference and the students' scores, and in Table B.1 we report the corresponding estimation results.

Figure B.1b shows the probability of enrolling in the student's top preference. As reported in column (2) in Table B.1, exceeding the cutoff increases the probability of enrollment in the top preference by 51.3%. Notice that this is not 100% for two reasons: (i) students whose score exceeds the cutoff may decide not to enroll, and (ii) students whose score was below the cutoff may end up enrolling after being pulled from the waitlist. Figures B.1c and B.1d are discussed in Section 2, and show that being above the cutoff significantly reduces the probability of reapplying and switching programs within the system. These results are confirmed in columns (3) and (4) in Table B.1. Figure B.1f and column (5) in Table B.1 exhibit a similar pattern, as it shows that the probability of major-switching also decreases among students above the cutoff. In contrast, we observe no significant difference in university switching. Finally, in Figure B.1g, we show that there is no effect of exceeding the cutoff on dropout rates.

TABLE B.1. Causal Effect of Crossing the Cutoff in Top Reported Preference

	Enroll (1)	Enroll Top Pref. (2)	Reapp. (3)	Switch (4)	Switch Major (5)	Switch University (6)	Dropout (7)
Z_{ip}	−0.004	0.543	−0.087	−0.058	−0.030	−0.038	0.008
	(0.016)	(0.017)	(0.018)	(0.017)	(0.014)	(0.014)	(0.011)
Observations	5,637	6,512	7,608	5,234	5,234	5,234	5,635

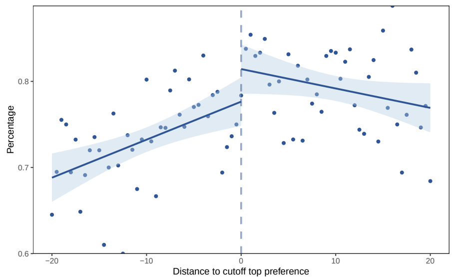
Note: Standard errors in parenthesis.

B.2.1 *Regression Discontinuities with True Preferences* Our previous analysis focuses on the effect of being above or below the cutoff of the top reported preference on different outcomes. Using the 2019 cohort and our nationwide survey, we can perform a similar analysis to estimate the causal effect of being above or below the cutoff of students' top-true preference on their outcomes. In Figures B.2a and B.2b, being above the cutoff significantly reduces the probability of reapplying to the system and being assigned

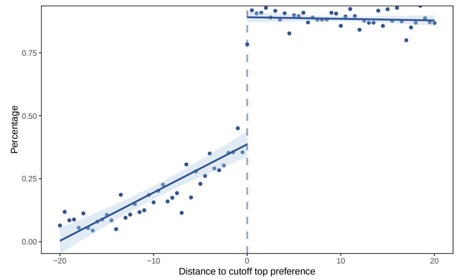
[3]Notice that we could perform the RD analysis for every cutoff, i.e., we could compute for every program the causal effect of being assigned to that program when it is listed as a top reported preference. In this sense, the causal effect we estimate under the current specification is the average of causal effects across all programs that are listed as a top preference.

FIGURE B.1. Effect of Cutoff Crossing

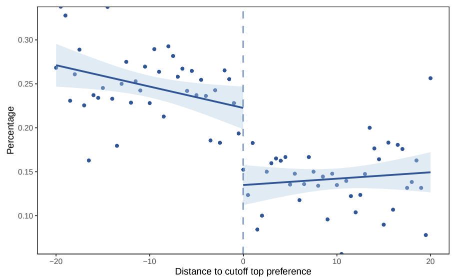
(a) Enrollment



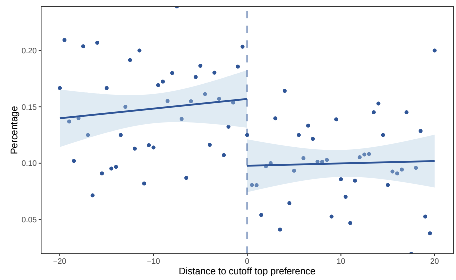
(b) Enrollment Top Pref.



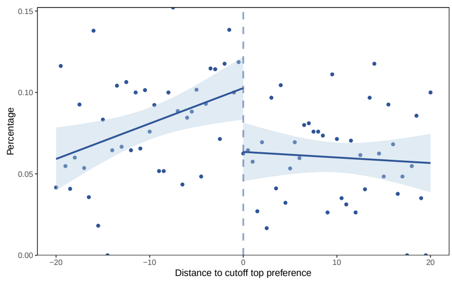
(c) Reapplication



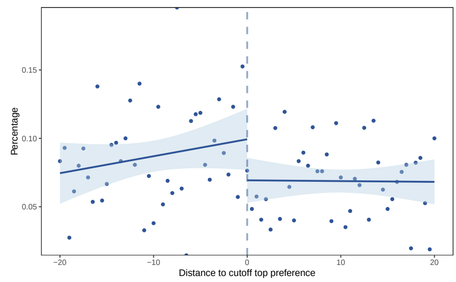
(d) Switching



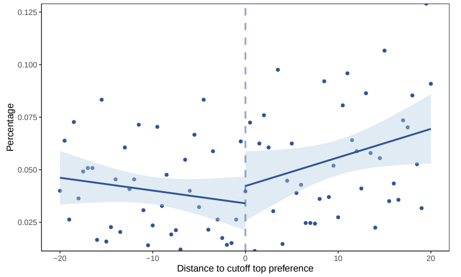
(e) Switching - University



(f) Switching - Major



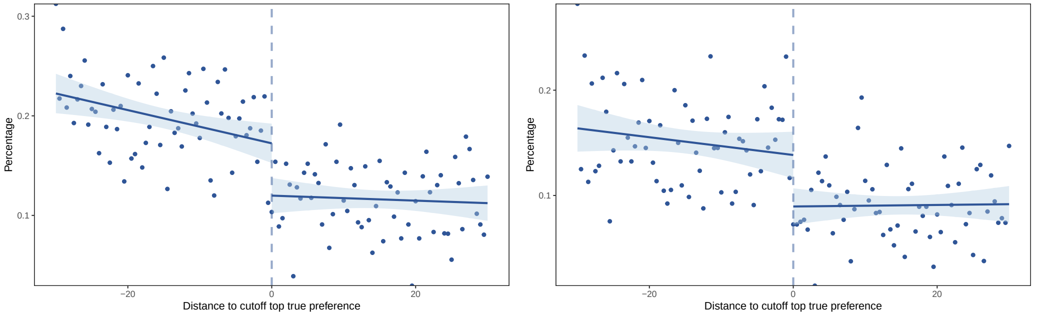
(g) Dropout



to a different program the next year. These results are consistent with those reported in Figure B.1, with the effect on reapplications being slightly smaller and that on switching being somewhat larger in magnitude compared with the analysis above.

FIGURE B.2. Effect of Cutoff Crossing for Top-True Preference

(a) Reapplication (b) Reassignment



B.3 *Additional Evidence*

FIGURE B.3. Students flow across states

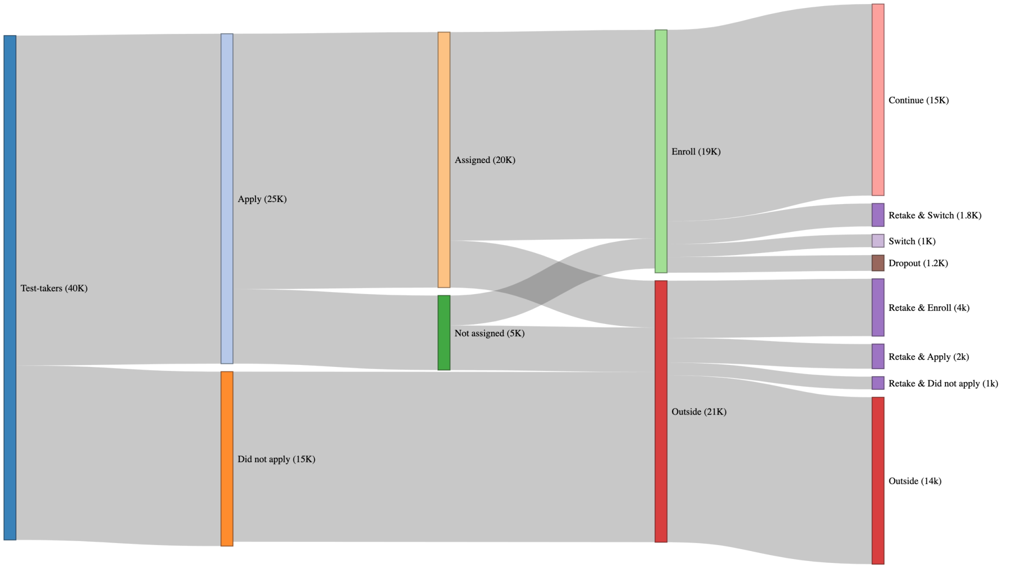
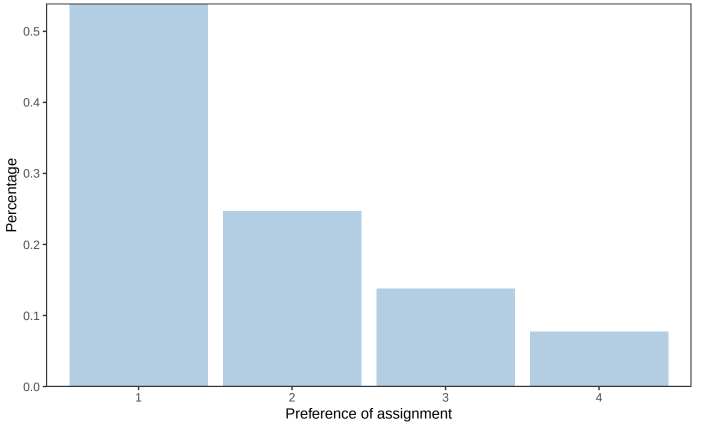
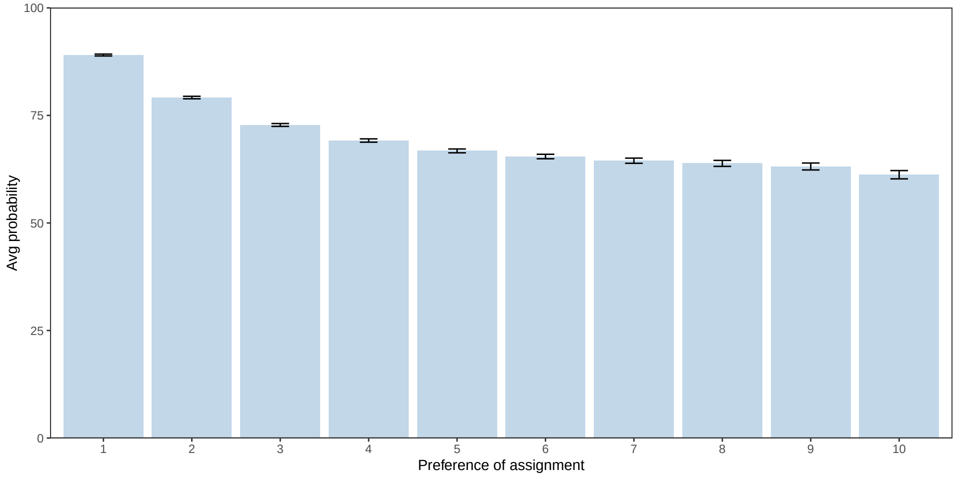


FIGURE B.4. Distribution Year of First Switching



B.3.1 *Perceived persistence and preference of assignment* The regression discontinuity results show a causal effect of preference of assignment on students' persistence with respect to their initial assignment. To show that this is partially explained by the mismatch channel, we use the 2020 survey on students' preferences and beliefs. Figure B.5 shows the average "perceived" probability of remaining enrolled in the same program after 1 year, by preference of enrollment. We observe that there is a significantly lower "perceived" probability of enrollment for lower-ranked preferences. On average, students believe that there is an 85% probability of remaining in the same program after a year for their first reported preference, whereas it is close to 65% for programs ranked below the fourth choice. Figure B.5 also provides evidence of forward-looking behavior (similar to the data patterns observed for students' switching probabilities),[4] which suggests that—in the aggregate—students' subjective beliefs are close to rational expectations beliefs.

FIGURE B.5. Average "perceived" probability of remaining enrolled in the same program, by preference of enrollment



The previous evidence does not guarantee that there are match effects between students and programs that are correlated with college persistence. For instance, a similar pattern could be observed if all students agree on their preference rankings over programs, and most of the correlation between reported preferences and college persistence was due to programs' characteristics. To rule this out and give evidence of match effects, we exploit the panel structure of students' ROLs as we observe the perceived persistence probability for every program listed in the ROL. We consider the following specification:

$$P_{ij} = \alpha_i + \alpha_j + X_{ij}\beta + \beta_R R_i(j) + \varepsilon_{ij},$$

[4]Notice that stop-out and dropout probabilities do not exhibit a positive correlation with the preference of assignment. Thus, they cannot drive most of the correlations shown in Figure B.5.

TABLE B.2. Two-way Fixed Effects Regression Results

	Dependent variable: Prob. of Persistence	
	Estimate	Std. Error
Preference 2	-9.891	8.3652e-03
Preference 3	-16.844	8.7019e-03
Preference 4	-21.355	9.3909e-03
Preference 5	-24.831	1.0631e-02
Preference 6	-27.148	1.1805e-02
Preference 7	-29.164	1.3202e-02
Preference 8	-30.329	1.5004e-02
Preference 9	-31.995	1.6876e-02
Preference 10	-34.757	1.9483e-02
Constant	89.181	1.0186e-02
Observations	159,894	
R^2	0.095	
Adjusted R^2	0.095	

where P_{ij} is the perceived persistence probability of student i in program j; α_i is students i's fixed effect; α_j is program j's fixed effect; X_{ij} are student-program characteristics that include a third-degree polynomial of the application score of student i in program j; $R_i(j)$ is the position of program j in ROL R_i; and ε_{ij} is an i.i.d shock. Table B.2 shows the estimation results. The preference of enrollment has a significant and strong effect on the perceived probability of persistence. We conclude that there are match effects in the setting, which exhibit a strong correlation with students' college persistence.

The results reported so far show that (i) there is a clear effect of the preference of assignment on the switching behavior of students, (ii) a significant fraction of students forecast this, and (iii) these results cannot be explained by students' or programs' characteristics solely.

APPENDIX C: APPENDIX FOR SECTION 4

C.1 *Model stages*

Each period involves the following stages and decisions:

Period 1:

(i) *Applications:* Given students' preferences, beliefs about their admissions and enrollment probabilities, prior beliefs about their unknown abilities, and the labor market return of studying each option, students make application decisions to the centralized system. After obsrving their preference shocks' realizations, students choose an ROL that maximizes their expected utility.

(ii) *Assignment:* Once applications have been submitted, a matching algorithm computes students' assignment to each program. In particular, this process is approximated by drawing a set of admissions cutoffs from the joint distribution of cutoffs and assigning students according to the matching algorithm's cutoff structure.

(iii) *Enrollment:* Once the assignment is realized, students face probabilities of enrollment in their assigned program or choosing the outside option.

(iv) *PSU preparation:* At the beginning of students' first year of college, or in the outside option, students choose whether to prepare for and retake the PSU tests. This decision affects their flow utility while in college and can improve the set of potential programs they can enroll in in the second period if they choose to reapply to the system.

(v) *Grades:* At the end of the year, students receive their college grades—which are noisy signals of their unknown abilities—and update their beliefs.

Period 2:

(i) *Reapplications:* At the beginning of period 2, students observe the realization of new preference shocks and PSU scores and, given their updated beliefs about their unknown abilities, decide whether to reapply[5] to the centralized system.

(ii) *Assignment:* Once applications are submitted, a matching algorithm computes students' assignment to each program. In particular, this process is approximated by drawing a set of admission cutoffs from the joint distribution of cutoffs and assigning students according to the matching algorithm's cutoff structure.

(iii) *Enrollment:* Once the assignment is realized, students face exogenous probabilities of enrollment in their assigned program. If students do not enroll in their assigned program, they can choose between staying in their current enrollment or dropping out of college.

(v) *Grades:* At the end of the year, students receive their college grades and update their beliefs regarding their abilities.

Period 3:

(i) *Dropout:* Students face an exogenous sequence of dropout probabilities for every year they are enrolled after their second period.

(ii) *Expected graduation:* Students face a graduation probability for every year they are enrolled after completed the formal duration of their program. Both dropout and graduation probabilities are estimated from the data, depending on programs' and students' observable characteristics.

(iii) *Labor market:* Students who graduate receive lifetime earnings and nonpecuniary payoffs based on their college decisions. Students who do not graduate receive the value function of students who dropped out.

C.2 *Learning*

Proposition 3 allows us to obtain the posterior mean and variance for student i's unknown ability in any program j'. We show how the student's statistical problem can be

[5] Students can also apply for the first time in period 2.

rewritten to make inference about each component in A^u_{ij}. To make inference about $A^u_{im_j}$ we can write

$$\begin{aligned}
&A^u_{ijt} = A^u_{ij} + \varepsilon^g_{ijt} \\
\Leftrightarrow &A^u_{ijt} = A^u_{im_j} + \sum_k \omega_{jk} A^u_{ik} + \varepsilon^g_{ijt} \\
\Leftrightarrow &A^u_{ijt} = \mathbb{E}_{t-1}\left[\sum_k \omega_{jk} A^u_{ik}\right] + \nu_{gti} \\
\Leftrightarrow &A^u_{ijt} - \mathbb{E}_{t-1}\left[\sum_k \omega_{jk} A^u_{ik}\right] = \nu_{gti}
\end{aligned}$$

where

$$\nu_{gti} \sim N\left(A^u_{im_j}, \sigma^2_g + \sum_k \omega^2_{jk}\sigma^2_s\right)$$

where now we treat $A^u_{ijt} - \mathbb{E}_{t-1}\left[\sum_k \omega_{jk} A^u_{ik}\right]$ as the new signal, and we make inference about $A^u_{im_j}$. This statistical problem now fits into the frame of DeGroot (2005) (we can follow similar steps to make inference about each A^u_{il}).

We can now write the posterior mean unknown ability if student i enrolls in program j' in the second period, after receiving the first-period signal a_{ij1} in program j:

$$\begin{aligned}
E_1\left(A^u_{ij'}|a_{ij1}\right) &= E_1\left(A^u_{im_{j'}} + \sum_k \omega_{j'k} A^u_{ik}|a_{ij1}\right) \\
&= E_1\left(A^u_{im_{j'}}|a_{ij1}\right) + \sum_k \omega_{j'k} E_1\left(A^u_{ik}|a_{ij1}\right).
\end{aligned}$$

Notice that if the student switches majors, i.e $m_{j'} \neq m_j$, she learns nothing about her major-specific unknown ability in her new program. This implies that the posterior mean equals the prior; that is,

$$E_1\left(A^u_{im_{j'}}|a_{ij1}\right) = 0.$$

So the posterior mean is given by

$$E_1\left(A^u_{im_{j'}}|a_{ij1}\right) = \begin{cases} 0 & \text{if } m_{j'} \neq m_j \\ \frac{\sigma^2_m a_{ij1}}{\sigma^2_g + \sum_k \omega^2_{jk}\sigma^2_s + \sigma^2_m} & \text{o.w} \end{cases}$$

We now turn to compute the posterior mean for the subject-specific unknown ability, i.e, $E_1\left(A^u_{ik}|a_{ij1}\right)$ $\forall k$. Although the student's subject-specific unknown ability does not vary across programs, given that grades depend on the average ability and average ability varies depending on the program-specific admissions weights ω_j, the amount of subject-specific unknown ability learned by the student will be program-specific.

The subject-specific posterior unknown ability is given by

$$E_1\left(A^u_{il}\middle|a_{ij1}\right)=\frac{\omega_{jl}\sigma_s^2 a_{ij1}}{\sigma_g^2+\sigma_m^2+\sum_k \omega_{jk}^2\sigma_s^2}$$

Finally, we can write the posterior mean for the unknown ability in any program j' by

$$E_1\left(A^u_{ij'}\middle|a_{ij1}\right)=\begin{cases}\frac{\sum_l \omega_{j'l}\omega_{jl}\sigma_s^2 a_{ij1}}{\sigma_g^2+\sigma_m^2+\sum_k \omega_{jk}^2\sigma_s^2} & \text{if } m_{j'}\neq m_j\\ \frac{\sigma_m^2 a_{ij1}}{\sigma_g^2+\sum_k \omega_{jk}^2\sigma_s^2+\sigma_m^2}+\frac{\sum_l \omega_{j'l}\omega_{jl}\sigma_s^2 a_{ij1}}{\sigma_g^2+\sigma_m^2+\sum_k \omega_{jk}^2\sigma_s^2} & \text{o.w}\end{cases}$$

Intuitively, the posterior mean places more weight on the signal for subjects with a higher admissions weight in ω_j. In this sense, student i learns more about her math ability if she enrolls in engineering, which places a high admissions weight on math.

C.3 *Model solution*

In this subsection, we describe the solution of the model via Backward Induction.

In period three, the terminal value function is given by

$$\begin{aligned}V_{ijt}(\mu_{ij2},\tau_{ijt}) = E_t\left[\sum_{t'=\tau_{ijt}+1}^{T_f} P^g_{ijt'}\left(\mathbb{E}_\varepsilon\left[\sum_{t''=0}^{t'-(\tau_{ijt}+1)}\beta^{t''}u_{ij(t+t'')}\right]+\beta^{t'-\tau_{ijt}}\underbrace{V^w_{ij(t+t'-\tau_{ijt})}(\mu_{ij2})}_{\text{Value fcn Labor market}}\right)\right]\\ +E_t\left[\sum_{t'=\tau_{ijt}+1}^{T_f} P^d_{ijt'}\left(\mathbb{E}_\varepsilon\left[\sum_{t''=0}^{t'-(\tau_{ijt}+1)}\beta^{t''}u_{ij(t+t'')}\right]+\beta^{t'-\tau_{ijt}}\underbrace{V_{i0(t+t'-\tau_{ijt})}}_{\text{Value fcn Dropout}}\right)\right],\end{aligned}$$

where μ_{ij2} is the posterior unknown ability of student i in program j after observing the period one signal, and τ_{ijt} is the number of academic years the student has completed in program j at the beginning of period three. In period three, there are no decisions to be made, and the value functions can be collapsed to the period-two value functions of enrolling in program j in the following way:

$$V_{ijt}(\mu_{ij2},\tau_{ijt})=u_{ijt}-\mathbb{1}_{\{(j\neq 0)\cap(\tau_{ijt}=0)\}}C^e+\beta\mathbb{E}_\varepsilon\left[V_{ijt+1}(\mu_{ij2},\tau_{ijt+1})\right],$$

where C^e is a first-time enrollment cost.

The value function in period one is then given by

$$\begin{aligned}V_{ijt}(\mu_{ij1},\tau_{ijt},\vec{s}_{it}) = \max_{d^s_{it}} E_0\Big[u_{ijt}-d^s_{it}C^{psu}-\mathbb{1}_{\{j\neq 0\}}C^e+\\ \beta\int_{a_{ij1}}\int_{\vec{s}_{it+1}}\underbrace{EmaxROL(\tau_{ijt}+1,\vec{s}_{it+1},\mu_{i2}(a_{ij1}))}_{\text{continuation value of reapplications}}\underbrace{d\pi(a_{ij1})}_{\text{signal}}\underbrace{dF(\vec{s}_{it+1}|\vec{s}_{it},d^s_{it})}_{\text{future scores}}\Big].\end{aligned}$$

Notice that in period one, the value function of enrolling in program j considers that (i) the student will update her beliefs about her unknown abilities for every program (μ_{i2}), (ii) in the next period, her scores could change if she retakes the PSU ($d^s_{it} = 1$), and (iii) she will have the option of submitting an optimal application in the second period ($EmaxROL\,(\cdot)$). In Appendix C.5 we derive analytical expressions for the continuation value of reapplications.

Actions. In periods one and two, students can choose to submit an application list. The indirect utility over assignment for student i to program k in period t, given her current enrollment in program j, $v_{ikt}|j$, can be written as

$$v_{ikt}|j = P^e_{it} \cdot V_{ikt} + (1 - P^e_{it}) \cdot \max\{V_{i0t}, V_{ijt}\}$$

Given students' indirect utilities over the assignment and their beliefs about admission probabilities, students choose an application list—depending on their application type—as detailed in Section 4.3.3.

C.4 *MIA*

Chade and Smith (2006) show that the optimal portfolio problem is NP-Hard. However, when admission probabilities are independent[6] 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 the Marginal Improvement Algorithm (MIA).

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

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

MIA recursively adds programs that give the highest marginal improvement to the portfolio, as long as they exceed the marginal cost of adding them. Olszewski and Vohra (2016) show that MIA also returns the optimal ROL when the number of applications is constrained and when $c(S)$ is supermodular. If Assumption 4 does not hold, the strict inequality in MIA's stopping criteria becomes a weak inequality. In this case, if students face degenerate admission probabilities, there could be a multiplicity of best response (He (2012)). We discuss this potential identification threat in Larroucau and Rios (2018). Assumption 4 is a sufficient condition to rule out the multiplicity of best response.

[6]Notice that in our case, in Assumption 4 we have assumed independence of beliefs about admission probabilities.

C.5 *EmaxROL*

In this subsection, we analyze the problem of computing the expected value of reporting an optimal ROL in the centralized system, in which the expectation is taken over next-period preference shocks. Formally, let the utility of being assigned to program j by $u_j = \bar{u}_j + \varepsilon_j$, then define the $EmaxROL$ by

$$EmaxROL := \mathbb{E}_{\varepsilon}\left[U(R_{max}) := \max_{R' \in \mathcal{R}, |R'| \leq K} U(R') - c(R')\right]$$

where, given Assumption 4 and an ROL $R = \{r_1, \ldots, r_k\}$,

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

where $z_j = p_j \cdot u_j = p_j \cdot (\bar{u}_j + \varepsilon_j)$ for each $j \in M$.

Finding a potentially closed-form solution to this problem is relevant because it allows us to characterize the continuation value in any dynamic model in which students can make application decisions over time. However, to the best of our knowledge, this problem has not yet been analyzed in the literature. The following example shows why this problem is different from computing the continuation value in a dynamic discrete choice model, which is usually referred to as an $Emax$ operator or inclusive value.

Example (EmaxROL).

Consider a portfolio problem in which students can submit ROLs of length $K = 1$ and there is no cost of application; i.e, $c(R) = 0, \forall R \in \mathcal{R}$. In this case, the $EmaxROL$ problem simplifies to the expectation—over the preference shocks—of choosing the program that gives the highest expected utility over assignment; that is

$$\begin{aligned} EmaxROL &:= \mathbb{E}_{\varepsilon}\left[U(R_{max}) := \max_{R' \in \mathcal{R}, |R'| \leq K} U(R') - c(R')\right] \\ &= \mathbb{E}_{\varepsilon}\left[\max_{j' \in \mathcal{J}} p_j (\bar{u}_j + \varepsilon_j)\right] \\ &= \mathbb{E}_{\varepsilon}\left[\max_{j' \in \mathcal{J}} p_j \bar{u}_j + p_j \varepsilon_j\right] \end{aligned}$$

Even though in this case the $EmaxROL$ reduces to the expectation of choosing the best alternative in a discrete choice problem, now the preference shocks are weighted by potentially different admission probabilities $\{p_j\}$. This key difference—compared with a traditional discrete choice problem—means that, even if we assume that preference shocks are distributed i.i.d type-I extreme value, the resulting random shocks, $p_j \varepsilon_j$, won't have equal variance. This implies that the inclusive value formulas derived by Rust (1987) do not hold.

The previous example shows that in general, the $EmaxROL$ will not have a closed-form solution, even when preference shocks are distributed i.i.d type-I extreme value.

We next show sufficient conditions under which the $EmaxROL$ can be efficiently approximated.

C.5.1 *Pairwise Stability* Under mild assumptions, Fack et al. (2019) show that the allocation outcome from constrained DA satisfies pairwise stability with respect to students' true preferences. We can exploit this fact for efficiently computing the $EmaxROL$.

When the allocation satisfies pairwise stability, the srtudent's problem reduces to choosing the most preferred program among the programs for which she is expost admissible. That is, given a realization of programs' cutoffs, $\{P_j\}_{\mathcal{J}}$, student i's allocation , $\mu(i|\{P_j\}_{j\in\mathcal{J}})$, satisfies pairwise stability if and only if

$$\mu(i|\varepsilon_i, \{P_j\}_{j\in\mathcal{J}}) = \underset{j\in J_i(\{P_j\}_{j\in\mathcal{J}})}{\operatorname{argmax}} \ \bar{u}_{ij} + \varepsilon_{ij}$$

where

$$J_i(\{P_j\}_{j\in\mathcal{J}}) := \{j \in \mathcal{J} : s_{ij} \geq P_j\} \bigcup \{j = 0\}$$

This implies that we can write the $EmaxROL$ as follows:

$$\begin{aligned} EmaxROL &:= \mathbb{E}_{\varepsilon}\left[U(R_{max}) := \max_{R'\in\mathcal{R}, |R'|\leq K} U(R') - c(R')\right] \\ &= \mathbb{E}_{\{P_j\}_{j\in\mathcal{J}}}\left[\mathbb{E}_{\varepsilon_i}\left[\max_{j\in J_i(\{P_j\}_{j\in\mathcal{J}})} \bar{u}_{ij} + \varepsilon_{ij} \Big| \{P_j\}_{j\in\mathcal{J}}\right]\right], \end{aligned}$$

and when ε_{ij} are distributed i.i.d type-I extreme value, the previous expression reduces to

$$\begin{aligned} EmaxROL &= \mathbb{E}_{\{P_j\}_{j\in\mathcal{J}}}\left[\mathbb{E}_{\varepsilon_i}\left[\max_{j\in J_i(\{P_j\}_{j\in\mathcal{J}})} \bar{u}_{ij} + \varepsilon_{ij} \Big| \{P_j\}_{j\in\mathcal{J}}\right]\right] \\ &= \mathbb{E}_{\{P_j\}_{j\in\mathcal{J}}}\left[\log\left(\sum_{j\in J_i(\{P_j\}_{j\in\mathcal{J}})} \exp\left(\bar{u}_{ij}\right)\right) + \gamma\right], \end{aligned}$$

where γ is the Euler's constant.

If we take the distribution of cutoffs to be invariant to students' individual reports, following an argument similar to that of Agarwal and Somaini (2018); we can estimate in a first stage the distribution of cutoffs $\{\hat{P}_j\}_{j\in\mathcal{J}}$ and then estimate the structural parameters of the model. This implies that we can compute the frequency of the random sets by using the bootstrap realizations of the cutoffs $J_i(\{P_j^{\tilde{b}}\}_{j\in\mathcal{J}})$ just once, where $\tilde{b} = 1, ..., \tilde{B}$ is a random sample of the bootstrap realizations of the cutoffs. We can then approximate the $EmaxROL$ by

$$EmaxROL = \mathbb{E}_{\{P_j\}_{j\in\mathcal{J}}}\left[\log\left(\sum_{j\in J_i(\{P_j\}_{j\in\mathcal{J}})} \exp\left(\bar{u}_{ij}\right)\right) + \gamma\right]$$

$$\approx \frac{\sum_{\tilde{b}\in\tilde{B}} \log\left(\sum_{j\in J_i(\{P_j^{\tilde{b}}\}_{j\in\mathcal{J}})} \exp\left(\bar{u}_{ij}\right)\right) + \gamma}{\tilde{B}}$$

Pairwise stability in the dynamic problem. We can a follow a similar calculation and give an expression for the expected value of reporting an ROL in our dynamic setting. The expected value over assignment to program k, given that student i is currently enrolled in program j, is given by

$$\begin{aligned} v_{ikt} &= P_{it}^e V_{ikt} + (1-P_{it}^e)\max\{V_{i0t}, V_{ijt}\} \\ &= P_{it}^e\left(\bar{V}_{ikt} + \varepsilon_{ikt}\right) + (1-P_{it}^e)\max\{\bar{V}_{i0t}+\varepsilon_{i0t}, \bar{V}_{ijt}+\varepsilon_{ijt}\}. \end{aligned}$$

then we can write

$$\begin{aligned} \mathbb{E}_\varepsilon\left[\max_k v_{ikt}\right] &= \mathbb{E}_\varepsilon\left[\max_k P_{it}^e\left(\bar{V}_{ikt}+\varepsilon_{ikt}\right) + (1-P_{it}^e)\max\{\bar{V}_{i0t}+\varepsilon_{i0t}, \bar{V}_{ijt}+\varepsilon_{ijt}\}\right] \\ &= P_{it}^e \mathbb{E}_\varepsilon\left[\max_k\left(\bar{V}_{ikt}+\varepsilon_{ikt}\right)\right] + (1-P_{it}^e)\,\mathbb{E}_\varepsilon\left[\max\{\bar{V}_{i0t}+\varepsilon_{i0t}, \bar{V}_{ijt}+\varepsilon_{ijt}\}\right] \end{aligned}$$

and we get that

$$EmaxROL(\tau_{ijt}, \vec{s}_{it}, a_{ij1}) \approx$$
$$\frac{\sum_{\tilde{b}\in\tilde{B}} P_i^e \log\left(\sum_{k\in J_i(\{P_k^{\tilde{b}}\}_{k\in\mathcal{J}}, \vec{s}_{it})} \exp\left(\bar{V}_{ikt}\right)\right)}{\tilde{B}} + (1-P_i^e)\log\left(\exp\left(\bar{V}_{ijt}\right) + \exp\left(\bar{V}_{i0t}\right)\right) + \gamma$$

Finally, when students retake the PSU in the first period, $d_{it-1}^s = 1$, we also need to integrate over students' future scores. Under Assumption 5 and using Gauss-Hermite polynomials, we can approximate the integral with stochastic scores over $EmaxROL$ by

$$\int_{\vec{s}_{it}} EmaxROL(\tau_{ijt}, \vec{s}_{it}, a_{ij1})\,\underbrace{dF(\vec{s}_{it}|\vec{s}_{it-1}, d_{it-1}^s)}_{\text{future scores}} \approx$$
$$\frac{1}{\sqrt{\pi}}\sum_{k=1}^{n_w} w_k \left(\frac{\sum_{\tilde{b}\in\tilde{B}} P_i^e \log\left(\sum_{l\in J_i\left(\{P_l^{\tilde{b}}\}_{l\in\mathcal{J}}, \vec{s^k}_{it}\right)} \exp\left(\bar{V}_{til}\right)\right)}{\tilde{B}} \right) +$$
$$(1-P_i^e)\log\left(\exp\left(\bar{V}_{ijt}\right) + \exp\left(\bar{V}_{i0t}\right)\right) + \gamma$$

where

$$\vec{s^k}_{it} = \max\{\vec{s^k}_{it-1}, \tilde{s^k}_{it}\}$$

with

$$\tilde{s^k}_{ilt} = \begin{cases} \alpha_l(1+\sqrt{2}\sigma_{psu}x_k)s_{ilt} & \text{if } s_{ilt} > 0 \\ \alpha_{0l}(1+\sqrt{2}\sigma_{psu}x_k)\bar{s}_{it} & \text{if } s_{ilt} = 0 \end{cases}$$

where n_w is the number of nodes at which we evaluate the integrand and w_k is the k-th integration weight for the k-th integration node x_k, given by the Gauss-Hermite formula. The accuracy of the previous approximation will depend on the number of nodes used to approximate the integral, n_w, and the number of joint draws of the cutoff scores, $\tilde{B}$.

C.6 *Exogenous Models*

C.6.1 *Dropout and Graduation* We assume that the academic progression concludes with students either (i) graduating from their program (after period 2) or (ii) dropping out. We assume that these outcomes are exogenously given so that the probabilities of graduating and dropping out depend only on the student's observable characteristics and on their first-year and second-year choices. This is formalized in Assumption 2.

ASSUMPTION 2. *Students have rational expectations over their graduation and dropout probabilities. Moreover, we assume that this decision follows a multinomial logit model, i.e.,*

$$P^g_{ij\tau} = \frac{\exp\left(X_{ij\tau}\psi^g\right)}{1+\sum\limits_{a\in\{g,d\}}\exp\left(X_{ij\tau}\psi^a\right)}, \quad \textit{and} \quad P^d_{ij\tau} = \frac{\exp\left(X_{ij\tau}\psi^d\right)}{1+\sum\limits_{a\in\{g,d\}}\exp\left(X_{ij\tau}\psi^a\right)}$$

where $P^g_{ij\tau}$ and $P^d_{ij\tau}$ represent the probabilities that student i decides to graduate and drop out from program j after τ periods enrolled in the program, respectively; $X_{ij\tau}$ is a vector of observable characteristics;[7] *and ψ^g, ψ^d are vectors of parameters that need to be estimated.*

C.7 *Enrollment*

Once students submit their optimal ROLs, they observe a draw from the joint distribution of cutoffs. Let $\vec{\bar{s}}^{\,t} = \left\{\bar{s}^t_j\right\}_{j\in M}$ be the vector of realized cutoffs in period t. Based on the mechanism's cutoff structure, the allocation can easily be obtained by assigning each student to the highest preference for which their application score is greater than or equal to the realized cutoff.

[7] The vector includes a constant, student's gender, a dummy variable that identifies whether the student's family income is below the median of the income distribution, and student's high school GPA.

After the assignment results are released, students decide whether to enroll in their assigned preference, remain enrolled in their current program if they are reapplying, or take the outside option. For simplicity, we do not model this decision and simply assume that students enroll in their preference of assignment with some probability P_{it}^e that depends on their observable characteristics.[8] This is formalized in Assumption 3.

ASSUMPTION 3. *Student i enrolls in her assigned program in period t with probability P_{it}^e, which is given by*

$$P_{it}^e = \frac{\exp\left(X_i^e \psi^e\right)}{1 + \exp\left(X_i^e \psi^e\right)},$$

where X_i^e is a vector of observable characteristics.[9]

If students do not enroll in their new assignment, we allow them to choose the best alternative between remaining in their current program for one more period or choosing the outside option. In Appendix C.3, we show that the solution to the student's problem can be obtained via Backward Induction.

C.8 *Admission Process*

Every time a student decides to (re)apply, we assume that they go through the following steps: (i) PSU tests, (ii) application, and (iii) enrollment.

C.8.1 *PSU Tests* As described in Section 3, the assignment is based on a series of admission factors, which include the PSU tests and two additional scores related to the student's performance during high school. Let $\mathcal{L} = \{1, \ldots, L\}$ be the set of admission factors and let $\vec{s}_{it} = \{s_{itl}\}_{l \in \mathcal{L}}$ be the vector of scores of student i in period t. In addition, let ω_{jtl} be the weight that program j assigns to factor $l \in \mathcal{L}$ in period t. Then, the application score of student i in program j and period t is given by

$$s_{ijt} = \sum_{l \in \mathcal{L}} \omega_{jtl} \cdot s_{itl}.$$

Since students can retake the PSU tests and reapply, we need to model (i) the evolution of their scores and (ii) the evolution of their beliefs about the admission weights that programs will use in the future. To model the former, we assume that the scores of student i in period $t+1$, $\vec{s}_{it+1}$, are exogenously given conditional on their scores in period t, $\vec{s}_{it}$, and the observable characteristics, X_i. To address the latter, we assume that students correctly forecast future weights. This assumption is likely to hold in practice, since admission weights are relatively stable over time. These considerations are formalized in Assumption 4.

[8]Students pay an enrollment cost C^e for the first time they enroll in a program, which captures both administrative and potentially psychological costs of the first-time enrollment process.

[9]The vector includes a constant, student's gender, a dummy variable that identifies if the student's family income is below the median of the income distribution, and student's High-school GPA.

ASSUMPTION 4. *Conditional on retaking the exam, the scores of student i in period $t+1$ are exogenously given and distributed according to*

$$\vec{s}_{it+1} \sim F_{\vec{s}_{it},X_i}(s), \quad \forall i$$

where $F_{\vec{s}_{it},X_i}(s)$ is the distribution of scores conditional on the initial vector of scores $\vec{s}_{it}$ and the observable characteristics X_i. In addition, we assume that students correctly forecast the admission weights $\{\omega_{jtl}\}_{l\in\mathcal{L}}$ used by each program j in each period t.

As a simplifying assumption, we further assume that the evolution of scores in each admission factor is proportional to the student's current scores, as described in Assumption 5.[10]

ASSUMPTION 5. *The scores of student i evolve according to the following process:*

$$\vec{s}_{it+1}|\vec{s}_{it}, X_i \sim \max\{s_{it}, \tilde{s}_{it+1}\}$$

with

$$\tilde{s}_{ilt+1} = \begin{cases} \alpha_l(1+\nu_{it+1})s_{ilt} & \text{if } s_{ilt} > 0 \\ \alpha_{0l}(1+\nu_{it+1})\bar{s}_{it} & \text{if } s_{ilt} = 0 \end{cases} \quad \text{and} \quad \nu_{it+1} \sim N(0, \sigma_{psu}),$$

where s_{itl} is the score of student i in exam l in period t, $\bar{s}_{it}$ is the average Math-Verbal score of student i in period t, and $\{\alpha_l, \alpha_{0l}\}_{l\in\mathcal{L}}$ and σ_{psu} are parameters to be estimated.

Finally, we assume that students must pay a cost for retaking the PSU tests. This cost accounts for the direct cost of taking the exam and the time spent to prepare for it. Since we do not have information regarding preparation time to retake the PSU for this cohort, we assume that this cost is a constant C^{psu}.

APPENDIX D: APPENDIX FOR SECTION 5

D.1 *Additional Evidence*

APPENDIX E: APPENDIX FOR SECTION 6

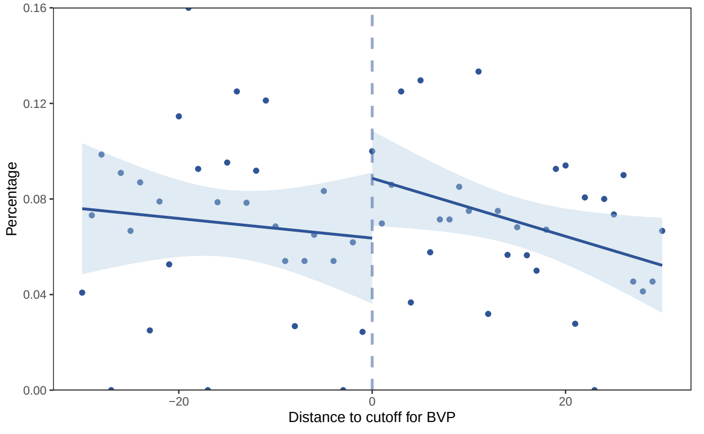
E.1 *Estimator*

We now introduce the estimator, closely following Bruins et al. (2018). Let $y_i := (y_{it}, ..., y_{iQ})$ be a collection of Q outcomes for student i', and let $\mathbf{y} := \{y_i\}_{i=1}^N$ denote the aggregate outcomes of all students $i \in \{1, ..., N\}$. Similarly, let x_i and $\mathbf{x}$ be individual and aggregate students' and programs' characteristics, and η_i and η be individual and aggregate random shocks. Let $\hat{\beta}_n$ be the vector of parameter estimates of the auxiliary model; that is,

$$\hat{\beta} := \underset{\beta}{\text{argmax}}\, \mathcal{L}(\mathbf{y}, \mathbf{x}; \beta) = \underset{\beta}{\text{argmax}}\, \frac{1}{N}\sum_{i=1}^{N} l(y_i, x_i; \beta),$$

[10]This specification captures the fact that students use the maximum application score from both pools of test scores for each program they apply to.

FIGURE D.1. Application to Education around Cutoff for BVP



where $l(\cdot;\beta)$ is the log-likelihood function given the vector β. Let $\eta^b := \{\eta_i^b\}_{i=1}^N$ denote a set of simulated draws for the random shocks of the structural model for simulations $s = 1, ..., S$, where each set of draws is simulated independent of each other. Let $\theta \in \Theta$ be a vector of parameters from the structural model, with $d_\theta \leq d_\beta$. Given the observable characteristics $\mathbf{x}$ and a parameter vector $\theta \in \Theta$, we can use the structural model to simulate data $\mathbf{y}^b(\theta) := \{y_i^b(\theta)\}_{i=1}^N$ and estimate the auxiliary model on each simulated sample:

$$\hat{\beta}^b := \underset{\beta}{\operatorname{argmax}}\, \mathcal{L}(\mathbf{y}^b(\theta), \mathbf{x}; \beta).$$

Let $\bar{\beta}(\theta)$ be the average of these estimates, i.e., $\bar{\beta}(\theta) := \frac{1}{B}\sum_{s=1}^S \hat{\beta}^b(\theta)$. Then, the II estimator minimizes the following function:

$$Q(\theta) := \left(\bar{\beta}(\theta) - \hat{\beta}\right)^T W \left(\bar{\beta}(\theta) - \hat{\beta}\right)$$

where W is a positive-definite weighting matrix.

For a given value of the parameters θ, and given the first-stage estimates—i.e., students' beliefs and enrollment, dropout, graduation, and employment probabilities—computing the objective function $Q(\theta)$ involves solving the model via Backward Induction and then forward simulating outcomes.[11] Solving the model is computationally expensive and especially computing the continuation value terms, since they depend on realization of the random coefficients $\{\alpha_i\}_{i=1}^N$ (which are known to students), which restricts the number of draws of the random coefficients we can use to evaluate the objective function. To reduce the noise due to a small number of draws, we consider a larger

[11] In which we have suppressed the dependency on the first-stage estimators for readability.

number of draws for those shocks that do not affect the backward induction. The current set of estimation results uses 50 draws for shocks that do not affect the backward-induction procedure (preference shocks, enrollment shocks, etc.) and 5 draws for random coefficient shocks per student[12]. In Algorithm 1, Appendix E, we describe in detail how we perform the estimation and discuss some related technical considerations.

E.2 *Auxiliary models and weighting matrix*

We use as an *auxiliary* model a combination of regression models—including data analogs of the grade equations, wage equations, linear probabilities models of graduation, linear probability models of switching and dropout, and RDD models—and a vector of moment conditions. The parameters of the model are identified jointly by the moment conditions generated by the auxiliary model. However, some sets of parameters are directly linked to particular moment conditions. We describe this auxiliary model and the matrix of weights in detail in Appendix E.2.

We describe now in detail the regressions and moment conditions we use in the estimation and the sets of parameters that explain most of each moment's variation.

E.2.0.1 *Grade equations.* The auxiliary model that targets the grade equations' structural parameters (γ) are given by the regression analogs of Equation 4:

$$G_{ij1} = \beta^{\gamma}_{1m_j} + \beta^{\gamma}_2 A_{ij} + \beta^{\gamma}_3 Z^g_i + \beta^{\gamma}_4 \mathbb{1}\{j = R_{1i}(1)\} + \beta^{\gamma}_5 s_{1im_j} + \beta^{\gamma}_6 s_{1ik_j} + \varepsilon^g_{ij1},$$

$$G_{ij2} = \beta^{\gamma}_{7m_j} + \beta^{\gamma}_8 S_{ij2} + \beta^{\gamma}_9 A_{ij} + \beta^{\gamma}_{10} Z^g_i + \beta^{\gamma}_{11} \mathbb{1}\{j = R_{ti}(1)\} + \beta^{\gamma}_{12} s_{2im_j} + \beta^{\gamma}_{13} s_{2ik_j} + \varepsilon^g_{ij2},$$

and

$$G_{ij2} = \beta^{\gamma}_{14} + \beta^{\gamma}_{15} Sw_{ij2} \left(\beta^{\gamma}_{16} + \beta^{\gamma}_{17} Sw_{ij2}\right) G_{ij1} + \varepsilon^{ts}_{ij},$$

where s_{tim_j} and s_{tik_j} are the shares of major m_j and college-type k_j in the ROL of student i associated with the assignment in period t respectively; $\mathbb{1}\{j = R_{ti}(1)\}$ is an indicator function that equals 1 if the student is assigned to her top-reported preference relative to the ROL associated with the assignment in period t; $S_{ij2} = 1$ if the student is in her second academic year in period 2 and $S_{ij2} = 0$ otherwise; and $Sw_{ij2} = 1$ if the student switched programs in period 2 and $Sw_{ij2} = 0$ otherwise.

E.2.0.2 *Wage equation.* The auxiliary models that target the parameters in the wage equation (λ) are given by

$$\log(\bar{w}_{j(\tau=4)}) = \beta^{\lambda}_{1m_j} + \beta^{\lambda}_2 \bar{A}_{k_j} + \beta^{\lambda}_3 \bar{G}_j + \beta^{\lambda}_4 \bar{Z}^w + \epsilon_{j(\tau=4)}$$

and

$$\log(\bar{w}_{m_j\tau}) = \beta^{\lambda}_{5m_j} + \beta^{\lambda}_{6m_j}\tau + \beta^{\lambda}_{7m_j}\tau^2 + \epsilon_{m_j\tau},$$

where τ is tenure after graduating[13].

[12] We have run robustness checks estimating the model with up to 10 realizations of random coefficient shocks per student, and estimation results are relatively similar.

[13] See Section 3.1 for a description of the aggregate data on wages.

E.2.0.3 *Nonpecuniary labor market parameters.* The auxiliary model that targets the parameters that specify the nonpecuniary payoffs in the work force (α^w) is given by the following linear probability model:

$$y_{ij} = \beta_1^w s_{1im_j} + \beta_2^w \mathbb{1}\{j = R_{1i}(1)\} + \beta_3^w A_{ij} + \beta_4^w \bar{A}_{k_j} + \beta_5^w Z_i^g + \varepsilon_{ij}^w.$$

E.2.0.4 *Learning parameters.* The auxiliary models that target the parameters associated with students' learning process (σ_s^2, σ_m^2, σ_g^2, and α_4^w) are given by the following linear probability models of switchings and dropout: for each outcome $O \in \{$switching major, switching math-type, switching program, switching major within math-type, switching math-type within major, switching up, switching down, switching out feasible, switching out unfeasible, dropping out$\}$,

$$O_{ij} = \beta_{1m_j}^o + \beta_2^o A_{ij} + \beta_3^o Z_i^g + \beta_4^o \mathbb{1}\{j = R_{1i}(1)\} + \beta_5^o s_{1im_j} + \beta_6^o s_{1ik_j} + \beta_7^o G_{ij1} + \varepsilon_{ij}^o,$$

where O_{ij} equals 1 if student i enrolled in program j has an outcome O and 0 otherwise.

E.2.0.5 *Unobserved preferences' parameters.* The auxiliary models that target the parameters associated with students' unobserved and persistent preferences (parameters governing α_{im_j}, α_{ik_j}, and C^e) are given by the following regression discontinuity design (RDD) models:[14] for each outcome $O^{RDD} \in \{$switching program, reapplying, switching program, switching up or out unfeasible$\}$,

$$O_{ij}^{RDD} = \beta_1^{o^{RDD}} + \beta_2^{o^{RDD}} \mathbb{1}\{D_{1ci} \geq 0\} + \beta_3^{o^{RDD}} D_{1ci} + \beta_4^{o^{RDD}} \mathbb{1}\{D_{1ci} \geq 0_{ij}\} \times D_{1ci} + \varepsilon_{ij}^{o^{RDD}},$$

where O_{ij}^{RDD} equals 1 if student i enrolled in program j has an outcome O^{RDD} and 0 otherwise; D_{1ci} is the distance to the cutoff of the top-reported preference for student i.

E.2.0.6 *Weighting matrix and standard errors.* We use as a weighting matrix a diagonal matrix. Each element on the diagonal is the inverse of each data moment's variance, which we obtain via a bootstrap procedure. We combine moments from different data sets and sample sizes. We weight up some moments in the weighting matrix that are key to identifying the parameters that affect the learning and initial mismatching channels: the correlations between students' first-year college grades and the different types of switching, the levels and causal effects of the RDD models, and the fraction of students assigned to their top-true preference. In addition, we weight up some moments that affect the baseline values for our counterfactuals' outcomes of interest: moments related to the evolution of scores; broad-major dummies in the wage and grade equations; correlation between grades and wages; shares of retakers, dropouts, and switchers; and top-reported market shares.[15] We do not use the optimal weighting matrix because of the numerical complexities involved in computing the derivatives of the objective function $Q(\theta)$. Therefore, our estimator will be unbiased but not efficient.

[14]We consider only students whose application scores for their top-reported preference are at a distance less than or equal to 30 points.

[15]The exact weighting scheme is available upon request.

E.3 *Technical considerations*

We estimate the model via II for the following reasons:

(i) We only have remote access to data on students' grades, and CRUCH only allowed us to obtain regression results and summary statistics at the aggregate level; this makes it difficult to estimate a likelihood-based estimator. However, II allows us to estimate a rich statistical representation of the data on students' grades and use the estimated parameters to construct moment conditions to estimate the model's structural parameters.

(ii) Parameters that involve the grade equation and wage equation have a clear reduced-form representation in the data.

(iii) Estimating students' preferences in a portfolio setting is computationally challenging for likelihood-based estimation methods (see Larroucau and Rios (2018)). However, given a set of model parameters, simulating data from our structural model is relatively fast because under Assumption 4, we can simulate strategic ROLs efficiently using the MIA.[16]

In addition, the structural model yields a mixture of continuous and discrete outcomes. This feature complicates the estimation procedure for a simulation-based method such as II, because the objective function, $Q(\theta)$, becomes a multidimensional step function that inherits the discontinuities produced in the simulated data.[17] Bruins et al. (2018) propose a solution to overcome these computational difficulties by introducing noise and smoothing to the objective function. They refer to this estimation procedure as "Generalized Indirect Inference" (GII). With the smoothed objective function, the researcher can use a gradient-based optimization method to minimize the objective function, which tends to be faster than gradient-free optimization routines. We choose to avoid this smoothing procedure, and we estimate the objective function and find the global optimum using MIDACO solver (Schlueter et al. (2013)).[18] We choose to do this because the model has close to 260 parameters to be estimated, and the gradient must be computed through numerical simulation. Thus, the evaluation of the gradient would take several minutes. The computational time of this approach could be significantly reduced by parallelizing the numerical approximation of the gradient. However, we have chosen to parallelize the objective function's computation instead and increase the number of draws in the forward-simulation stage to smooth the objective function. Since solving the model and forward-simulating outcomes are completely independent across students, we parallelize the algorithm's outer loop to evaluate $Q(\theta)$.[19]

[16] See Appendix C.4.

[17] For a given realization of the random shocks, measures constructed from discrete outcomes of the model change discontinuously when we change the value of the structural parameters.

[18] MIDACO uses an evolutionary hybrid algorithm based on the Ant Colony Optimization (ACO) metaheuristic (Schlueter et al. (2009)).

[19] The model is coded in RcppArmadillo and parallelized with OpenMP.

Algorithm 1 Computing $Q(\theta)$

Input. Value of the structural parameters θ and first-stage estimates $\hat{p}$, $\hat{P}^e$, $\hat{P}^d$, $\hat{P}^g$, and $\hat{P}^w$.

Output. Value of the objective function $Q(\theta)$.

Step 1. For each student i, program j, and simulation b

Step 1.a. Draw a vector of random coefficients $\alpha_i^{m_{rc}}$,

Step 1.b. Solve the model by Backward Induction,

Step 1.c. For each simulation in N_s and for each date, draw a vector of preference shocks $\varepsilon_i^{m_s,m_{rc}}$, enrollment shocks $\varepsilon_i^{e,m_s,m_{rc}}$, wage shocks $\epsilon_i^{m_s,m_{rc}}$, vector of random cutoff scores $P^{m_s,m_{rc}}$ from the empirical distribution of cutoffs, vector of PSU score shocks $\nu_i^{m_s,m_{rc}}$, vector of unknown abilities $A_i^{u,m_s,m_{rc}}$, and grade shocks $\varepsilon_i^{g,m_s,m_{rc}}$

Step 1.d. Forward simulate the model and obtain a set of outcomes $y_i^{m_s,m_{rc}}$,

Step 2. For each simulation, estimate the *auxiliary* model parameters, $\hat{\beta}^{m_s,m_{rc}}(\theta)$, on the simulated sample

Step 3. Compute $\bar{\beta}(\theta) = \frac{1}{N_{rc} \times N_s} \sum_{m_{rc}} \sum_{m_s} \hat{\beta}^{m_s,m_{rc}}(\theta)$

Step 4. Return $Q(\theta) := \left(\bar{\beta}(\theta) - \hat{\beta}\right)^T W \left(\bar{\beta}(\theta) - \hat{\beta}\right)$

E.4 *Results*

TABLE E.1. Estimation Results - Goodness of Fit (I)

Targets	Model	Data
Share retakers	0.24	0.255
Share dropouts	0.059	0.054
Share dropouts females	0.043	0.046
Means share d dropouts low income	0.068	0.06
Share reapplicants	0.356	0.147
Share program switching	0.057	0.071
Share broad major switching	0.024	0.029
Means share major switching	0.045	0.044
Means share d switch math within major 1	0.002	0.004
Means share d switch major within math 1	0.025	0.028
Means share d switch uni	0.035	0.04
Means share d switch college type	0.021	0.021
Share dropout end of first period	0.028	0.029
Share enrolls first in second period	0.129	0.05
Share first year in second period	0.189	0.102
Share second year in second period	0.277	0.38
Share top true is pref 1ROL 1	0.424	0.424
Share top true is pref 1ROL 2	0.512	0.443
Share ROL length 1 10	0.298	0.063
Share ROL length 2 10	0.304	0.062
Share d applies 1	0.663	0.654
Share d applies 2	0.399	0.223
Share toptrue prefs changed reapps	0.282	0.654
Share reapps from top reported prefs	0.321	0.256
Share reapps from top true prefs	0.037	0.076
Mean tuition of top reported prefs	3.717	3.776
Mean distance of top reported prefs	7.25	10.156
Mean relpos of top reported prefs	-3.186	-1.845
Mean average share math types ROL (year 1)	0.275 0.725	0.387 0.613
Mean average share math types ROL (year 2)	0.327 0.673	0.451 0.549
Mean average share majors ROL (year 1)	0.14 0.022 0.048 0.036 0.109 0.064 0.048 0.005 0.231 0.296	0.124 0.022 0.068 0.05 0.118 0.048 0.07 0.02 0.21 0.247
Mean average share majors ROL (year 2)	0.134 0.029 0.048 0.068 0.126 0.093 0.044 0.009 0.226 0.222	0.107 0.027 0.06 0.043 0.142 0.041 0.092 0.022 0.264 0.194

Note: Majors are Social Sciences, Science Education and Humanities, and Health. College types are CRUCH-Public, CRUCH-Private, and Non-CRUCH.

TABLE E.2. Estimation Results - Goodness of Fit (II)

Targets	Model	Data
Means corr norm broad majors grades 1	-0.208	-0.074
Means corr norm majors grades 1	-0.321	-0.1
Means corr norm math types grades 1	-0.124	0.009
Means share toptrue broad majors changed	0.145	0.213
Means share toptrue majors changed	0.21	0.281
Means share toptrue math types changed	0.089	0.296
Means share toptrue prefs changed from oo	0.28	0.617
Means share toptrue broad majors changed from oo	0.147	0.191
Means share toptrue majors changed from oo	0.19	0.209
Means share toptrue math types changed from oo	0.072	0.31
Means share d switch math type	0.022	0.019
Means means tuition of toptrue pref 1 low income	3.568	4.193
Means means tuition of toptrue pref 1 above median	4.089	4.465
Means means observed ability scores i of topreported pref 1	1.143	1.112
Means means observed ability scores program of topreported pref 1	1.693	1.467
Means share apply topreported with prob zero	0.316	0.299
Means means risk ROL 1	0.311	0.318
Mean average share college types ROL (year 1)	0.372 0.353 0.275	0.341 0.458 0.188
Mean average share college types ROL (year 2)	0.35 0.309 0.341	0.354 0.501 0.145
Norm difference on broad major shares	0.267	0.386
Means norm diff broad major shares from oo	0.17	0.338
Means norm diff math types shares	0.216	0.284
Means norm diff math types shares from oo	0.1	0.268
Mean average dummy math types (year 1)	0.514 0.843	0.359 0.641
Mean average dummy math types (year 2)	0.59 0.816	0.445 0.555
Mean average dummy math types (year 1, females)	0.62 0.791	0.447 0.553

Note: Majors are Social Sciences, Science Education and Humanities, and Health. College types are CRUCH-Public, CRUCH-Private, and Non-CRUCH.

TABLE E.3. Estimation Results - Goodness of Fit (III)

Targets	Model	Data
Evolution of Score		
Means scores evolution lang	0.042	0.033
Means scores evolution math	0.01	0.009
Vars scores evolution lang	0.009	0.035
Vars scores evolution math	0.01	0.006
Means scores evolution hist nozero	0.055	0.041
Vars scores evolution hist nozero	0.009	0.008
Means scores evolution cien nozero	0.059	0.058
Vars scores evolution cien nozero	0.009	0.01
Means scores evolution hist zero	0.067	0.068
Vars scores evolution hist zero	0.009	0.011
Means scores evolution cien zero	0.067	0.024
Vars scores evolution cien zero	0.009	0.013
Market Shares and Shares Within ROL		
Shares broad majors within ROL (year 1)	0.354 0.362 0.054 0.231	0.358 0.319 0.09 0.21
Shares broad majors within ROL (year 2)	0.319 0.401 0.053 0.226	0.35 0.264 0.114 0.264
Norm difference on broad major shares	0.267	0.386
Means norm diff major shares	0.421	0.494
Means norm diff major shares from oo	0.213	0.419
Dummies broad majors within ROL (year 1)	0.527 0.537 0.083 0.388	0.496 0.452 0.164 0.274
Dummies broad majors within ROL (year 2)	0.5 0.579 0.09 0.382	0.469 0.39 0.195 0.331
Dummies broad majors within ROL (year 1, women)	0.392 0.544 0.14 0.491	0.502 0.318 0.202 0.376
Market shares broad major (enrollment 1)	0.635 0.056 0.014 0.018 0.008 0.033 0.022 0.021 0.002 0.069 0.121	0.506 0.068 0.014 0.036 0.036 0.058 0.03 0.031 0.01 0.077 0.134
Market shares broad major (enrollment 2)	0.535 0.069 0.022 0.023 0.015 0.04 0.034 0.027 0.003 0.092 0.142	0.529 0.074 0.013 0.035 0.028 0.054 0.028 0.032 0.009 0.074 0.123
Market shares broad major (enrollment 1, females)	0.601 0.056 0.013 0.023 0.009 0.038 0.028 0.041 0.003 0.11 0.079	0.517 0.06 0.015 0.045 0.036 0.068 0.029 0.042 0.012 0.109 0.066
Market shares broad major (enrollment 2, females)	0.428 0.072 0.023 0.036 0.019 0.055 0.042 0.053 0.006 0.156 0.11	0.538 0.065 0.014 0.043 0.027 0.065 0.028 0.044 0.01 0.105 0.061
Market shares by college type (year 1)	0.635 0.129 0.168 0.068	0.518 0.165 0.23 0.087
Market shares by college type (year 2)	0.535 0.163 0.207 0.095	0.556 0.153 0.206 0.084
Market shares by college type (year 1, low-income)	0.794 0.047 0.055 0.105	0.645 0.18 0.14 0.035
Market shares by college type (year 2, low-income)	0.744 0.048 0.05 0.158	0.688 0.164 0.115 0.033

Note: Majors are Social Sciences, Science Education and Humanities, and Health. College types are CRUCH-Public, CRUCH-Private, and Non-CRUCH.

TABLE E.4. Estimation Results - Goodness of Fit (IV)

Targets	Model	Data
Auxiliary Model: Grade Equation 1		
Grade 1 observed ability 1	0.502	0.445
Grade 1 d pref 1	0.002	0.067
Grade 1 female	0.249	0.171
Broad Majors	3.891 4.361 4.211 3.659	4.075 3.696 4.269 4.162
Grade 1 broad major share	0.187	0.196
Grade 1 college major share	-0.051	-0.086
Grade 1 hat sigma g1	0.712	0.681
Auxiliary Model: Grade Equation 2		
Grade 2 observed ability 2	0.439	0.457
Grade 2 d pref 1	-0.004	0.028
Grade 2 female	0.161	0.218
Broad Majors	3.977 4.485 4.345 3.785	4.096 3.769 4.306 4.325
Grade 2 second-year student	0.203	-0.014
Grade 2 broad major share	0.157	0.171
Grade 2 college major share	-0.037	-0.308
Auxiliary Model: Time Series for Grades		
Grades ts no switchers constant	0.44	0.837
Grades ts no switchers slope	0.912	0.813
Grades ts switchers constant	2.734	3.508
Grades ts switchers slope	0.478	0.285

Note: Majors are Social Sciences, Science Education and Humanities, and Health. College types are CRUCH-Public, CRUCH-Private, and Non-CRUCH.

TABLE E.5. Estimation Results - Goodness of Fit (V)

Targets	Model	Data
Auxiliary Model: Wage Equation		
Broad Majors	2.915 2.536 2.251 2.974	2.524 2.691 2.143 2.715
Wage grades 2 wages	0.013	0.013
Wage observed ability college wages	0.082	0.136
Wage female wages	-0.258	-0.187
Wage standard error	0.076	0.067
Auxiliary Model: Wage Growth Equation		
Wage growth broad major dummies	2.6 1.93 1.779 2.193	2.116 2.208 1.752 2.078
Wage growth broad major-specific linear	0.119 0.194 0.138 0.275	0.114 0.176 0.141 0.236
Wage growth broad major-specific quadratic	-0.003 -0.008 -0.007 -0.021	-0.004 -0.009 -0.012 -0.021
Auxiliary Model: Non-Pecuniary Utility Equation		
Work np pref 1	0.044	0.037
Work np observed ability np	0.055	0.173
Work np observed ability college np	0.009	-0.04
Broad Major	0.56 0.531 0.604 0.594	0.442 0.31 0.403 0.519
Work np standard error	0.199	0.227

Note: Majors are Social Sciences, Science Education and Humanities, and Health. College types are CRUCH-Public, CRUCH-Private, and Non-CRUCH.

TABLE E.6. Estimation Results - Goodness of Fit (VI)

Targets	Model	Data
Auxiliary Model: Dropout Equation		
Broad Major	0.592 0.616 0.608 0.608	0.482 0.494 0.494 0.52
Grade coeff.	-0.055	-0.086
Ability	-0.101	-0.037
Top Pref.	0.012	0.022
Female	-0.171	0.015
Percentage Broad Major / College Type	-0.018 -0.008	-0.002 -0.006
Auxiliary Model: Program Switching Equation		
Broad Major	0.916 1.068 1.048 0.925	0.801 0.781 0.821 0.822
Grade coeff.	-0.152	-0.148
Ability	0.043	0.04
Top Pref.	-0.048	-0.029
Female	-0.046	0.008
Percentage Broad Major / College Type	-0.056 0.009	-0.042 0.031
Auxiliary Model: Broad Major Switching Equation		
Broad Major	0.516 0.589 0.65 0.552	0.429 0.436 0.463 0.471
Grade coeff.	-0.092	-0.075
Ability	0.055	0.018
Top Pref.	-0.016	-0.001
Female	0	0.012
Percentage Broad Major / College Type	-0.15 0.05	-0.086 0.016
Auxiliary Model: Major Switching Equation		
Broad Major	0.912 1.085 1.062 0.887	0.6 0.576 0.61 0.611
Grade coeff.	-0.172	-0.106
Ability	0.068	0.03
Top Pref.	-0.035	-0.004
Female	-0.005	0.011
Percentage Broad Major / College Type	-0.071 0.019	-0.084 0.023
Auxiliary Model: Math Type Switching Equation		
Broad Major	0.405 0.504 0.498 0.43	0.254 0.237 0.242 0.265
Grade coeff.	-0.079	-0.044
Ability	0.03	0.005
Top Pref.	-0.017	-0.003
Female	-0.001	0.008
Percentage Broad Major / College Type	-0.023 0.007	-0.029 0.012

Note: Majors are Social Sciences, Science Education and Humanities, and Health. College types are CRUCH-Public, CRUCH-Private, and Non-CRUCH.

TABLE E.7. Estimation Results - Goodness of Fit (VII)

Targets	Model	Data
Auxiliary Model: Major Switching within Math Equation		
Broad Major	-0.001 0.001 0.002 0.025	0.044 0.037 0.042 0.051
Grade coeff.	0.002	-0.009
Ability	-0.003	0
Top Pref.	-0.003	0
Female	-0.004	0.003
Percentage Broad Major / College Type	0.003 -0.002	0.005 0.002
Auxiliary Model: Math Switching within Major Equation		
Broad Major	0.506 0.582 0.566 0.482	0.377 0.363 0.397 0.384
Grade coeff.	-0.091	-0.069
Ability	0.035	0.026
Top Pref.	-0.02	-0.002
Female	-0.008	0.005
Percentage Broad Major / College Type	-0.044 0.01	-0.049 0.013
Auxiliary Model: Switching College Type Equation		
Broad Major	0.346 0.392 0.414 0.366	0.313 0.309 0.329 0.324
Grade coeff.	-0.051	-0.052
Ability	0.022	0.017
Top Pref.	-0.005	-0.004
Female	-0.008	-0.001
Percentage Broad Major / College Type	0.04 -0.127	-0.006 -0.059

Note: Majors are Social Sciences, Science Education and Humanities, and Health. College types are CRUCH-Public, CRUCH-Private, and Non-CRUCH.

TABLE E.8. Estimation Results - Goodness of Fit (VIII)

Targets	Model	Data
Auxiliary Model: Switching Up Equation		
Broad Major	0.103 0.121 0.114 0.115	-0.001 0.005 0.004 -0.005
Grade coeff.	-0.008	0.002
Ability	0.009	0.002
Top Pref.	-0.042	-0.025
Female	-0.016	-0.001
Percentage Broad Major / College Type	-0.035 -0.001	0.003 0.014
Auxiliary Model: Switching Down Equation		
Broad Major	0.223 0.242 0.236 0.212	0.185 0.171 0.185 0.186
Grade coeff.	-0.029	-0.032
Ability	0.009	0.01
Top Pref.	0.005	0.007
Female	-0.004	-0.001
Percentage Broad Major / College Type	-0.052 -0.035	-0.023 -0.005
Auxiliary Model: Switching Out-Feasible Equation		
Broad Major	0.386 0.44 0.416 0.367	0.458 0.444 0.465 0.466
Grade coeff.	-0.084	-0.089
Ability	0.011	0.025
Top Pref.	0.007	-0.003
Female	-0.013	0.009
Percentage Broad Major / College Type	0.046 0.03	-0.016 0.015
Auxiliary Model: Switching Out-Unfeasible Equation		
Broad Major	0.205 0.265 0.282 0.23	0.055 0.056 0.063 0.069
Grade coeff.	-0.032	-0.011
Ability	0.014	0.001
Top Pref.	-0.018	-0.001
Female	-0.014	0.001
Percentage Broad Major / College Type	-0.015 0.015	-0.001 0.008

Note: Majors are Social Sciences, Science Education and Humanities, and Health. College types are CRUCH-Public, CRUCH-Private, and Non-CRUCH.

TABLE E.9. Estimation Results - Goodness of Fit (IX)

Targets	Model	Data
Auxiliary Model: RDD Switching		
Constant	0.205	0.162
Discontinuity	-0.069	-0.045
Slope - Left	-0.001	0.001
Slope - Right	0.002	-0.001
Auxiliary Model: RDD Re-Application		
Constant	0.488	0.226
Discontinuity	-0.104	-0.08
Slope - Left	-0.004	-0.002
Slope - Right	0.005	0.002
Auxiliary Model: RDD Switch Up or Out Unfeasible		
Constant	0.146	0.049
Discontinuity	-0.079	-0.034
Slope - Left	0	0
Slope - Right	0.001	-0.001
Auxiliary Model: RDD Teaching		
Constant	0.035	0.047
Discontinuity	0.033	0.03
Slope - Left	-0.001	-0.003
Slope - Right	-0.001	0.002
Other Moments		
Share of preference of assignment	0.517 0.303 0.092 0.036 0.017 0.009 0.006 0.006 0.005 0.005 0.004	0.41 0.298 0.146 0.079 0.033 0.017 0.009 0.004 0.002 0.001 0.001
Means means share broad majors within ROL 1 enr 1	0.891	0.838
Means means share college type within ROL 1 enr 1	0.857	0.684
Means share toptrue majors changed	0.21	0.281
Means share assigned to top true	0.124	0.211
Means share toptrue majors changed from oo	0.19	0.209
Means norm diff college type shares from oo	0.133	0.4
Means corr norm college types grades 1	0.078	0.035
Market shares broad major (enrollment 1)	0.635 0.056 0.014 0.018 0.008 0.033 0.022 0.021 0.002 0.069 0.121	0.506 0.068 0.014 0.036 0.036 0.058 0.03 0.031 0.01 0.077 0.134
Market shares broad major (enrollment 2)	0.535 0.069 0.022 0.023 0.015 0.04 0.034 0.027 0.003 0.092 0.142	0.529 0.074 0.013 0.035 0.028 0.054 0.028 0.032 0.009 0.074 0.123
Market shares broad major (enrollment 1, females)	0.601 0.056 0.013 0.023 0.009 0.038 0.028 0.041 0.003 0.11 0.079	0.517 0.06 0.015 0.045 0.036 0.068 0.029 0.042 0.012 0.109 0.066
Market shares broad major (enrollment 2, females)	0.428 0.072 0.023 0.036 0.019 0.055 0.042 0.053 0.006 0.156 0.11	0.538 0.065 0.014 0.043 0.027 0.065 0.028 0.044 0.01 0.105 0.061

Note: Majors are Social Sciences, Science Education and Humanities, and Health. College types are CRUCH-Public, CRUCH-Private, and Non-CRUCH.

APPENDIX F: APPENDIX FOR SECTION 7

F.1 *Understanding Behavioral Channels*

To assess to which extent students' switching and dropout decisions are explained by the behavioral channels previously described, namely, initial mismatches and learning, we consider three counterfactuals:

1. No Systematic Learning: Sets the value of the standard deviation of each unknown ability to zero. Hence, there are no unknown abilities.
2. No Mismatch: Assigns each student to their top preference, independent of programs' capacities. As a result, programs' capacities may be exceeded. This counterfactual eliminates initial mismatches, which allows us to isolate the learning channel.

3. No Mismatch or Systematic Learning: Combines the two previous counterfactuals, which allows us to isolate the learning channel from the effects of the idiosyncratic shocks (random learning).

The first column in Table F.1 reports the results of the baseline model, which includes the two main behavioral channels. The next three columns match the three counterfactuals described above. We group the first two columns as *With Mismatches* and the last two columns as *No Mismatches* to highlight the fact that in the latter, the mismatch channel is not present. Notice that in the case with no mismatches the number of seats offered by each program may be exceeded. Finally, each row represents an outcome of interest, including statistics regarding reapplications, switching, dropout, enrollment, and on-time graduation, among others.

TABLE F.1. Results Counterfactuals - Behavioral Channels

	With Mismatches		No Mismatches	
Outcome	Baseline	No Systematic Learning	Baseline	No Systematic Learning
Reapplicants [%]	42.01	37.52	21.73	19.08
Program switchings [%]	5.51	2.63	1.47	0.10
Retakes PSU [%]	23.15	21.64	7.53	7.50
Dropouts [%]	6.66	5.46	13.82	12.20
Dropouts - first year [%]	2.93	1.79	10.43	8.90
Applicants in first period [%]	69.28	65.62	89.70	87.92
Enrolls same program [%]	27.92	29.98	49.94	51.70
Assigned to top true preference [%]	12.19	14.06	100.00	100.00
Unassigned in first period [%]	51.86	54.52	10.30	12.07
Graduate late [%]	94.97	94.56	93.18	93.08

Note: Switching and dropout rates are computed with respect to the total sample of participants.

With Mismatches. We start by focusing on the first two columns. First, we observe that having no learning decreases the number of reapplications, program switches, and dropout rates but increases the number of unassigned students in the first period. By shutting down the learning process, we increase the persistence of students' preferences over time, which translates into lower switching rates. Additionally, without the gains from learning, the value from enrolling in the centralized system drops. Therefore, a higher fraction of students choose the outside option. Finally, we observe that the systematic learning channel explains close to one-half of students' switching behavior.

No Mismatches. We now focus on the case with no mismatches. Recall that in this counterfactual, all students are assigned to their most desired preference, possibly exceeding the vacancies of programs. For this reason, the fraction of students who are unassigned decreases considerably, and thus these results are not directly comparable to those previously described. However, comparing the two columns labeled "Baseline" provides an idea of the benefits of eliminating congestion and initial mismatches. In particular, we observe that the fraction of students who reapply is considerably smaller, and so are the switching rates. The reason is that this counterfactual assigns students to their most desired program, which eliminates congestion and initial mismatches in the assignment, and thus reduces the incentives for students to reapply or switch. On the other hand, we observe an increase in the dropout rates at the end of the first year and within the first 2 years. However, notice that this rate is computed relative to the entire population, so naturally this increases as more students are assigned under this counterfactual.[20] We also observe that eliminating mismatches improves on-time graduation rates and the fraction who graduate from their first enrollment, which are mainly driven by the reduction in switching rates. These results suggest that eliminating initial mismatches is a sensitive approach to reduce switching and increase on-time graduation rates, and thus improve the system's yield.

F.2 *Finding equilibrium beliefs*

Index each counterfactual experiment and the baseline model by τ; then the rational expectations equilibrium cutoff distributions, $\hat{p}\,(\tau)$, can be computed with the following algorithm:

[20]If we compute the dropout rate for those students who are assigned, we observe that the rates are similar for both columns labeled as Baseline $(6.66/(100-51.86)=13.83\%$ vs. $13.82/(100-10.3)=15.4\%)$.

Algorithm 2 Computing $\hat{p}(\tau)$

Input. Structural parameter estimates $\hat{\theta}$, first-stage estimates $\hat{p}$, $\hat{P}^e$, $\hat{P}^d$, $\hat{P}^g$, and $\hat{P}^w$, and tolerance level ϵ_{tol}.

Output. Rational expectations equilibrium cutoff distributions $\hat{p}(\tau)$

Step 1. For each program j

Step 1.a. Solve the model and simulate outcomes given the rules implied by counterfactual τ and the estimated objects $\left(\hat{\theta}, \hat{p}, \hat{P}^e, \hat{P}^d, \hat{P}^g, \hat{P}^w\right)$

Step 1.b. Obtain a set of simulated ROLs and scores $\left(R_1^0, R_2^0, s_1^0, s_2^0\right)$

Step 1.c. For each program j, estimate the mean and standard deviation of the cutoff distributions $\hat{\delta}_j^0 \equiv \left(\hat{\mu}_j^0, \hat{\sigma}_j^0\right)$

Step 2. $\delta_{diff} = 2\epsilon_{tol}$, $k = 1$, $\rho = 0.9$

Step 3. While $\delta_{diff} > \epsilon_{tol}$

Step 3.a. For each student i, solve the model via Backward Induction given τ, the estimated parameters $\left(\hat{\theta}, \hat{P}^e, \hat{P}^d, \hat{P}^g, \hat{P}^w\right)$, and cutoff distributions $\hat{p}^{k-1}$, and obtain the continuation values for each student and state

Step 3.b. Forward simulate first period ROL R_{i1}^k given τ, the estimated parameters $\left(\hat{\theta}, \hat{P}^e, \hat{P}^d, \hat{P}^g, \hat{P}^w\right)$, cutoff distributions $\hat{p}^{k-1}$, and continuation values

Step 3.c. For each program j, estimate the mean and standard deviation of the cutoff distributions $\hat{\delta}_j^0 \equiv \left(\hat{\mu}_j^0, \hat{\sigma}_j^0\right)$

Step 3.d. Given initial first-period applications R_1^k, second-period applications R_2^{k-1}, and students' scores s_1^k and s_2^{k-1}, run the Chilean matching mechanism and obtain an allocation $\mu^k\left(R_1^k, R_2^{k-1}, s_1^k, s_2^{k-1}\right)$

Step 3.e. Given $\mu^k\left(R_1^k, R_2^{k-1}, s_1^k, s_2^{k-1}\right)$, τ, the estimated parameters $\left(\hat{\theta}, \hat{P}^e, \hat{P}^d, \hat{P}^g, \hat{P}^w\right)$, cutoff distributions $\hat{p}^{k-1}$, and continuation values, forward simulate second period ROLs R_{i2}^k

Step 3.f. Given $\left(R_1^k, R_2^{k-1}, s_1^k, s_2^{k-1}\right)$, run the boostrap procedure and estimate the rational expectations cutoff distributions $\tilde{p}^k$ Take a convex combination of the realized cutoffs $\tilde{p}^k$ and $\hat{p}^{k-1}$ (pointwise), i.e, $\hat{p}^k = \rho^k \hat{p}^{k-1} + \left(1-\rho^k\right)\tilde{p}^k$

Step 3.g. Estimate the mean and standard deviation of the cutoff distributions $\hat{\delta}_j^k \equiv \left(\hat{\mu}_j^k, \hat{\sigma}_j^k\right)$

Step 3.h. Compute $\delta_{diff} = ||\hat{\delta}^k - \hat{\delta}^{k-1}||$ $\hat{p}(\tau) = \hat{p}^{k-1}$ $k++$

Algorithm 3 Constrained Deferred Acceptance with signal and bonus ψ

Input. Indirect utilities v, application scores s, cutoff distributions P, and application score bonus ψ

Output. Optimal ROL $R\left(v, s, P, \psi_{\mathcal{T}}\right)$

Step 1. For each program j

Step 1.a. Compute admission probabilities given cutoff distributions P and application scores $\tilde{s}(j) = \{s_1, ...s_{j-1}, \psi_{\mathcal{T}} s_j, s_{j+1}, ...s_J\}$

Step 1.b. Compute and store optimal ROL $R\left(v, \tilde{p}(j)\right)$ using MIA

Step 2. Compute optimal signal

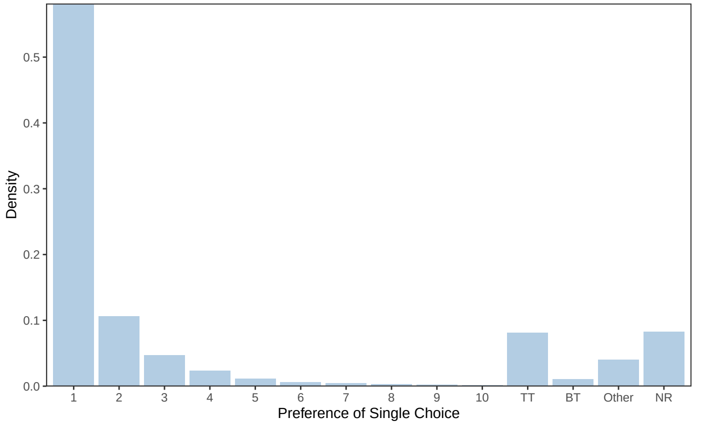
$$s_j^* = \underset{j}{\operatorname{argmax}} \left\{R(v, \tilde{p}_j)\right\}$$

Step 3. Compute optimal ROL $R(v, \tilde{p}_j)$

F.3 *Hypothetical scenario*

In 2022, we conducted a survey to elicit preferences and beliefs, which is similar to the 2019-2021' versions. In this version, we included a question to elicit information about how students would change their application lists if they could apply to a single program—i.e., applying under CDA with $K = 1$. Figure F.1 shows the distribution of the chosen program in the hypothetical scenario relative to their submitted ROL. Labels from *1-10* identify the share of students whose hypothetical program coincides with their k-th reported preference; *TT*, *BT*, and *Other* identify the share of students who report in the hypothetical scenario a program outside their submitted ROL; finally, *NR* identifies the share of students who did not respond to the survey question. We observe that under CDA with $K = 1$, a significant fraction of students would choose a program that is not their top-reported program on their current list (close to 40%). This suggests that a significant fraction of students would react strategically and take into account their admission probabilities when facing a binding constraint on the length of the list, consistent with our modeling assumptions.

FIGURE F.1. Distribution of applications under CDA with $K = 1$



REFERENCES

AGARWAL, NIKHIL AND PAULO SOMAINI (2018): "Demand Analysis using Strategic Reports: An application to a school choice mechanism," *Econometrica*, 86, 391–444. [20]

BRUINS, MARIANNE, JAMES A DUFFY, MICHAEL P KEANE, AND ANTHONY A SMITH JR (2018): "Generalized indirect inference for discrete choice models," *Journal of Econometrics*, 205, 177–203. [24, 28]

CHADE, HECTOR AND LONES SMITH (2006): "Simultaneous Search," *Econometrica*, 74, 1293–1307. [18]

DEGROOT, MORRIS H (2005): *Optimal statistical decisions*, vol. 82, John Wiley & Sons. [16]

DONG, YINGYING (2017): "Regression Discontinuity Designs With Sample Selection," *Journal of Business & Economic Statistics*, 37, 171–186. [8]

FACK, GABRIELLE, JULIEN GRENET, AND YINGHUA HE (2019): "Beyond truth-telling: Preference estimation with centralized school choice and college admissions," *American Economic Review*, 109, 1486–1529. [20]

HE, YINGHUA (2012): "Gaming the Boston School Choice Mechanism in Beijing," *Manuscript, Toulouse School of Economics*, 2012. [18]

LARROUCAU, TOMAS AND IGNACIO RIOS (2018): "Do "Short-List" Students Report Truthfully? Strategic Behavior in the Chilean College Admissions Problem," . [18, 28]

OLSZEWSKI, WOJCIECH AND RAKESH VOHRA (2016): "Simultaneous selection," *Discrete Applied Mathematics*, 200, 161–169. [18]

RIOS, IGNACIO, TOMÁS LARROUCAU, GIORGIOGIULIO PARRA, AND ROBERTO COMINETTI (2021): "Improving the Chilean College Admissions System," *Operations Research*, 69, 1015–1348. [7]

RUST, JOHN (1987): "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica: Journal of the Econometric Society*, 999–1033. [19]

SCHLUETER, M., J.A. EGEA, AND J.R. BANGA (2009): "Extended Ant Colony Optimization for non-convex Mixed Integer Nonlinear Programming," *Computers and Operations Research*, 36, 2217–2229. [28]

SCHLUETER, M., S. ERB, M. GERDTS, S. KEMBLE, AND J.J. RUCKMANN (2013): "MIDACO on MINLP Space Applications," *Advances in Space Research*, 51, 1116–1131. [28]