Negative Externalities of Off Platform Options and the Efficiency of Centralized Assignment Mechanisms

Adam Kapor
Princeton University

Mohit Karnani
MIT

Christopher Neilson
Princeton University & NBER*

August 24, 2019

Most Recent Version

Abstract

In this paper we study empirically how off-platform options negatively affect the efficiency of centralized assignment systems in the context of higher education. We develop an empirical model of college applications, aftermarket waitlists and matriculation choices using individual-level administrative data from Chile on almost half a million applications, test scores and enrollment decisions at all on and off-platform higher education options. We focus on a policy change in 2012 when a significant number of off-platform higher education options were added to the centralized assignment system in Chile. Following the increase of on-platform options, matriculation in placed slots rose by 8% and dropout rates at the end of the first year of college dropped by 2 points (a 16% drop). We use the estimated model to quantify the negative externalities caused by off-platform options that explicity accounts for the configuration of on and off-platform options and the degree of aftermarket frictions. We show that when students are allowed to express their preferences for a larger variety of options on the platform and have fewer options off the platform, welfare increases substantially, students begin there studies sooner, and fewer students drop out by the end of the first year of study. These benefits are greater for less advantaged students and women. Counterfactual analysis suggests that more desirable options generate larger negative externalities when not on the platform. Taken together, these results indicate that policymakers considering implementing centralized assingment systems should seriously consider off platform options when designing regulation surrounding centralized assignment systems.

PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CITE.

*The authors wish to thank Franco Calle for excellent research assistance.

1 Introduction

Centralized assignment systems are increasingly common in settings all over the world and as of 2018, at least 46 countries use centralized choice and assignment mechanisms for higher education.[1] Centralized matching systems have been shown to be theoretically appealing and successful in many settings (Abdulkadiroğlu et al., 2017). In practice many implementations diverge from ideal hypothesized settings and require additional considerations when implementing market design policy (Roth, 2002, 2007). One common aspect of implementing centralized markets in practice is that in virtually every case there exist many off-platform options that are available to participants of the match. In the context of assigning students to schools, these can include private schools or charter schools that can choose to not participate in the centralized system. In other cases such as higher education, some providers are in fact excluded from the platform as part of regulation, while others choose not to participate. In practice, matches can be later reneged in favor of other off platform options and waitlists can be impractical and inefficient due to aftermarket frictions.

In this paper we study the empirical relevance of the negative externalities generated by off-platform options and quantify aftermarket frictions that contribute to generating them in practice. We develop an empirical model to study and quantify the negative externalities caused by off platform options as a function of the configuration of on and off platform options as well as an empirical estimate of after market frictions. Our empirical application uses data from the centralized assignment system for higher education in Chile, which is one of the world's longest running college assignment mechanisms based on Deffered Acceptance [2]. We take advantage of a recent policy change that significantly increased the number of on platform institutions from 25 to 33, raising the number of slots from approximately 85,000 to 105,000 (increase of 25%). We estimate a model of college applications, after market waitlists and matriculation choices using individual level administrative data on almost half a million applications, test scores and enrollment decisions at all on and off platform higher education options. We show that when students are allowed to express their preferences for a larger variety of options on the platform and have less options off the platform, welfare increases substantially, students begin there studies sooner, and less students drop out by the end of the first year of study. These quantitative results suggest that off platform options generate negative externalities to the efficiency of the assignment system and that these costs can be meaningful economically.

We use the estimated model to further explore what students are affected by the off platform

[1] See worldwide coverage map in Figure A-1) and (Neilson, 2019) for a world wide review of countries that use policies that implement centralized choice and assingment mechanisms.

[2] A unique national entrance exam was first implemented in 1967 and central assignment based on a Deferred Acceptance algorithm has been used for the last 50 years.

options by comparing outcomes for individuals faced with different configurations of on platform options. We find that in the case of Chile, women and more disadvantaged students are the most adversely affected by the inefficiency created by off platform options. This can be partly due to their higher sensitivity to price and lower utility for private off platform options. We then use the estimated model to evaluate how our results would change in counterfactual exercises when different combinations of higher education options are on or off the platform. We find that more desirable options create larger welfare losses when they are not on the platform and available for people to reveal that they prefer them over other on platform options.

Taken together, these results show empirically that considering off platform options can be very important when planning a policy to implement a centralized assignment system. The type of options and the expected after market difficulties can be evaluated to consider actions to mitigate this problem or to incentivize participation of the most important options. Our empirical framework and counterfactual analysis provide a specific metric to evaluate the cost of losing each university on the platform. The model and the empirical strategy can be used to quantify the costs of off platform options in other settings and hopefully a route to informing policy regarding the costs of having off platform options.

2 The Centralized College Admissions System and Its Expansion

2.1 Context and Setting

There are 18 public universities en Chile, out of which 2 were recently created in 2016 and will be excluded in this paper, leaving us with 16 public universities in our data. Among the other dozens of private universities in the country, there are 9 that are considered traditional and are pooled together with public universities in a consortium called the *Chilean Council of Universities* (CRUCh). These 25 (16 public and 9 private) CRUCh universities, which we will interchangeably call G25, participate in the Unique Admission System (SUA), which is a centralized college admission clearinghouse.

This centralized admission system has long history. The first admission process was held over half a century ago, in 1967, where over 30 thousand applicants were tested on a standardized battery of questions, and then were eligible to apply for a slot in one of the eight founding universities in the system. Except for a policy in the 70's that split the largest university of the country into regional branches, the number of institutions in the platform remained relatively invariant until 2012.

The basic structure of the Chilean centralized admission is as follows. High school graduates register to be take the *Prueba de Selección Universitaria* (herein, PSU), which is a set of tests that comprise a verbal and a math mutiple choice test, and optative history and science tests, which are also of multiple choice. Students that do not attend the math test or the verbal test are immediately disqualified from the process. All test scores are standardized so their sample distributions resemble a normal distribution every year, for each test. The target mean of these (truncated, discretized) normal distributions are all set to be 500 points, and their standard deviations are set to 110 points. The minimum score of the test is assigned a score of 150 points, and the maximum corresponds to 850 points.

After receiving their test scores, students that obtained a simple average of at least 450 points between their math and verbal tests are allowed to apply through the centralized admission system. Students with an average between math and verbal scores below 450 cannot apply and may retake the tests in the following year if they want to do so.

Having a simple math-verbal PSU average of 450 does not however guarantee a slot in a university in the system. Students that decide to apply must do so by submitting a ranked list of *program-university* pairs[3], in which they express their preferences for programs in the system. When they submit this ordered list, students have access to the following relevant, public information: the number of slots that each program offers, how the program *weights* the test scores of applicants, their personal weighted score if they were to apply to a given program, and the weighted score of the last admitted student in all previous years for every program in the platform.

After submitting their ranked program-list, which defined their strictly ordered preferences in the platform, students are assigned to higher education programs by following the Deferred Acceptance (DA) algorithm proposed by (**?**), where programs' preferences for applicants are given by they corresponding weighted scores.[4] Students are assigned to their best feasible option, conditional on all the information in the platform, and receive an admission offer from the corresponding university if they are accepted in a program.

It may be the case that students are waitlisted in zero, one, or multiple programs. A student will only be accepted in a program with capacity k if she is one of the top k applicants in terms of weighted score in that given program. If this student ranks below the kth position of applicants in that program, she will be automatically waitlisted in the program. In other words, students are never completely rejected, they are all placed in waitlists if they are not accepted. Once a student is accepted in an option, all the stated preferences ranked below the option of acceptance are discarded. Thus, after discarding post-acceptance options, students observe the same list they submitted to the system filled with a waitlist indicator in the first zero or more options, and with an acceptance indicator in the last option (the rest of the rows are not considered).

[3] Applicants immediately enter a major field of study when they are admitted in a university. See (Bordon and Fu, 2015) for more details.

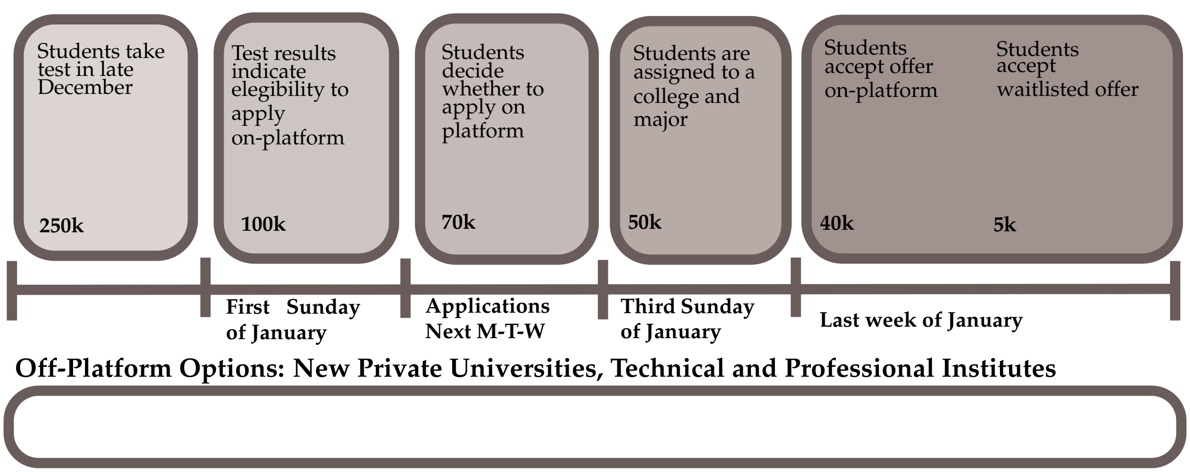
[4] Each program in each university independently submits their weights to the system, so weighted scores capture the preferences of programs for their applicants.

Students that received an acceptance offer then have the chance to enroll in that program. They pay the corresponding matriculation fees and perform costly actions to secure a spot in a program, if they decide to do so. There is no punishment or cost for not enrolling in a program, and students are free to choose if they take or leave the option in which they were allocated.

After the enrollment process, where all enrollment costs are sunk, waitlists are processed independently by each institution in a decentralized manner. The commonly adopted approach is to go through the waitlist and inform (through a phone call, for example) each waitlisted applicant that they now have an available slot. Students may accept or decline to enroll if they get through a waitlist of a given program. If they accept, they have to go through the enrollment process again for their new programs. If a student declines to enroll (or does not answer the phone, for example), the corresponding institution moves ahead with the next waitlisted applicant. This process is full of nuisances and frictions: students may be called from multiple waitlisted programs, there might be communication issues (e.g. wrong numbers may be dialed), students may renounce to a waitlisted offer after accepting it (before actually enrolling), and there are no further iterations of this waitlist process if, for example, many admitted students of a program get off other waitlists and decide to enroll elsewhere.

Once all enrollment and waitlist-enrollment decisions have been made and the entry cohort for each program has been finally determined, each program begins their regular academic year in a decentralized fashion. The whole process is repeated in the following year.

Figure 1: Diagram of Application Process On-Platform



Note: Diagram shows the progression of steps for applicants on and off the centralized assignment platform. The numbers of students in each step is for 2011 before the platform was expanded. The baseline is the cohort of students that take the national college entrance exam in 2011.

There have been some mild adjustment in the contents and structure of the tests[5], but the most relevant alteration in the structure of the system was due to a centralization policy in which eight large, off-platform universities were merged into the centralized system, increasing the number of institutions from 25 to 33.

2.2 The Expansion of the Platform

The Unique Admission System (SUA) is a centralized platform dedicated to assign students to all the universities that belong to the Chilean Council of Universities (CRUCh). The CRUCh applies the Deferred Acceptance (DA) algorithm with all applicants and programs on the platform and ensures that the process is fair and transparent. The CRUCh comprises all public universities in the country, and the most prestigious private universities. All non-CRUCh universities were excluded from the SUA and operated under their own decentralized rules until 2012, when the CRUCh allowed these universities to apply to join the SUA. Only the most selective and high-quality universities were allowed into the platform, and they all agreed to follow the rules posed by the SUA. Having the chance of applying to both centralized and decentralized universities, a large fraction of applicants decided to apply through the SUA and at the same time, in a practically costless fashion, apply to off-platform programs. The off-platform programs with higher demands were delivered by the institutions that joined the SUA in 2012. As a result, applicants now had to consider these universities when deciding their optimal applications on the platform. Because of the DA algorithm properties, students that aimed to enroll in these new SUA universities could no longer consider them as "outside options", and had to reveal their true preferences through the admission process. Allegedly, this had a positive impact over students that preferred CRUCh universities, as applications to these programs would be less congested. Thus, we expect that students were more likely to be assigned to their top feasible program.

3 The Data and Stylized Facts

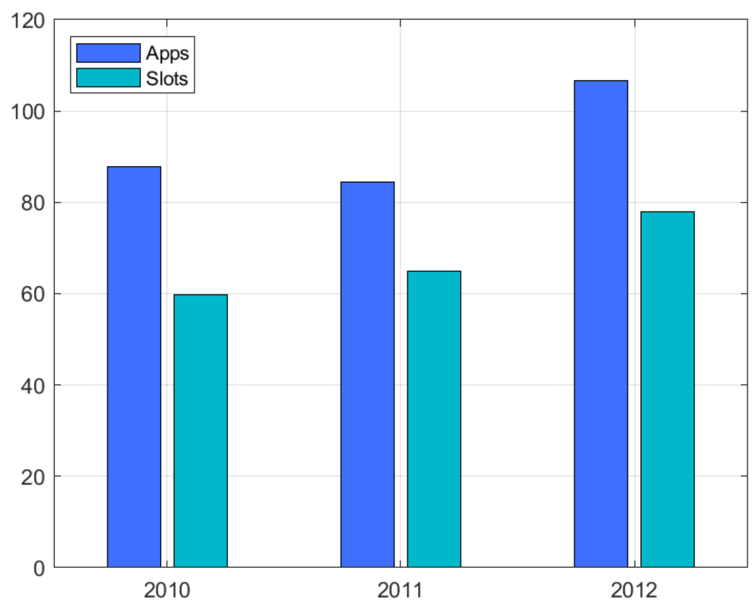
We use administrative records from the Chilean Ministry of Education (MINEDUC), which comprise population-level data on high school records, demographic characteristics of students and higher education enrollment, both in and out of the platform. We matched MINEDUC's data with DEMRE's application records, which includes all platform applicants, their PSU test scores, ordered preferences for programs, application status and rank in each program. Finally, we also engineered some additional features of each program, such as the number of "quantitative" and

[5]In 2004 the name of the admission test changed from *Prueba de Aptitud Académica* (PAA) to *Prueba de Selección Universitaria* (PSU), and so did the evaluated contents. Nevertheless, the same 25 universities remained in the system and the assignment mechanism was practically the same. See (Lafortune et al., 2016) for another policy that restricted the choices that applicants could make.

"qualitative" courses they include, and categorized them in standardized program groups to make them comparable.

Our data spans all academic years from 2010 to 2012, and we report descriptive statistics for all our most relevant student-level variables for each year in Table **??**. With this dataset we are able to observe some meaningful changes in the system after the addition of the G8 universities. Firstly, as commented before, the number of slots in the system increased by almost a half, while the number of applicants incremented just by around a quarter of the pre-policy state. In absence of significant changes in the number of slots offered by each institution, and considering that there is excess demand for them, this suggests that about a half of the students that usually enroll in G8 universities also made use of the platform. This is supported by Figure 2, which depicts the evolution of platform slots and applicants over time. Indeed, in the year of implementation of the policy, excess demand fell from over 50% of vacancies to around 30%.

Figure 2: Platform slots and applicants (in thousands)



Increasing the number of slots in the system naturally implies that the number of applicants that eventually enroll in an on-platform option also increases. This is mechanical, as incorporating new G8 options increases the number of admitted students in the platform and thus increases on-platform enrollment in absolute terms. A less obvious observation is that students that were admitted into a *G25* option increased their enrollment *rate* in G25 institutions after the policy, just as shown in Figure 3. This effect is fully driven by the inability to deviate to another G8 option af-

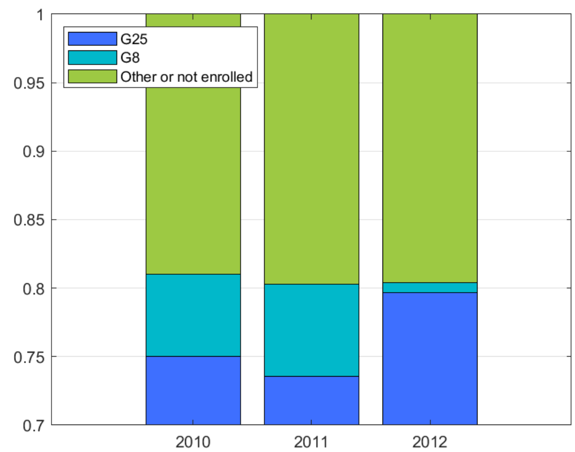
Table 1: Descriptive Statistics by Year

	Mean	Std. Dev.	Obs
Year 2010			
Pre-Application			
Male	0.47	0.50	251634
Private High School	0.10	0.30	251634
Metro Area	0.65	0.48	251634
GPA	529.53	115.51	251634
Math Score	500.79	110.77	251634
Verbal Score	500.64	108.92	251634
Post-Application			
Applied Through Platform	0.35	0.48	251634
Admited in Platform	0.76	0.43	87875
Admited in First Preference	0.36	0.48	87875
Enrolled in Platform	0.64	0.48	67013
Platform Program Dropout	0.09	0.29	42657
Year 2011			
Pre-Application			
Male	0.48	0.50	250758
Private High School	0.10	0.30	250758
Metro Area	0.64	0.48	250758
GPA	531.64	110.07	250758
Math Score	501.08	111.27	250758
Verbal Score	501.04	108.34	250758
Post-Application			
Applied Through Platform	0.34	0.47	250758
Admited in Platform	0.80	0.40	84512
Admited in First Preference	0.42	0.49	84512
Enrolled in Platform	0.59	0.49	67803
Platform Program Dropout	0.10	0.30	40011
Year 2012			
Pre-Application			
Male	0.47	0.50	239368
Private High School	0.11	0.31	239368
Metro Area	0.64	0.48	239368
GPA	535.76	113.45	239368
Math Score	503.94	110.63	239368
Verbal Score	504.28	109.74	239368
Post-Application			
Applied Through Platform	0.45	0.50	239368
Admited in Platform	0.88	0.33	106706
Admited in First Preference	0.45	0.50	106706
Enrolled in Platform	0.62	0.48	93562
Platform Program Dropout	0.08	0.28	58360

Note: This table shows descriptive statistic of the administrative data from DEMRE, the agency that runs the centralized assignment mechanism in Chile.

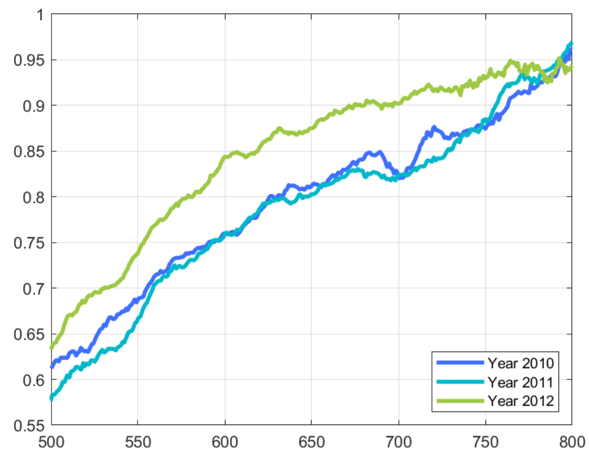
ter being admitted in a G25 program. Thus, students that prefer G8 programs, who must now rank them above other G25 options, allow a better sorting of students who prefer G25 options, which are around 7.5 percentage points ($\sim$ 10%) more likely to enroll into their assigned programs.

Figure 3: Enrollment probabilities for G25 admits



Moreover, we find that this G25 enrollment-likelihood effect is observed across the whole distribution of PSU scores, i.e. G25 admits are more prone to enroll into their admission programs over a very wide range of the test score's support. This is shown in the left panel of Figure 4, where sample probabilities are plotted in 20-point bins for all the years in our data. Its right panel depicts local polynomial smoothed probabilities with 99% confidence intervals. Overall, we observe a significant average increase in enrollment rates for G25 admits with scores below 760 points. It is worth noting that, as test scores are adjusted to resemble a normal distribution with a mean of 500 and a standard deviation of 110 points, 760 points is above the 99th percentile of the score distribution. Thus, the policy increased the enrollment yield for the vast majority G25 admits.

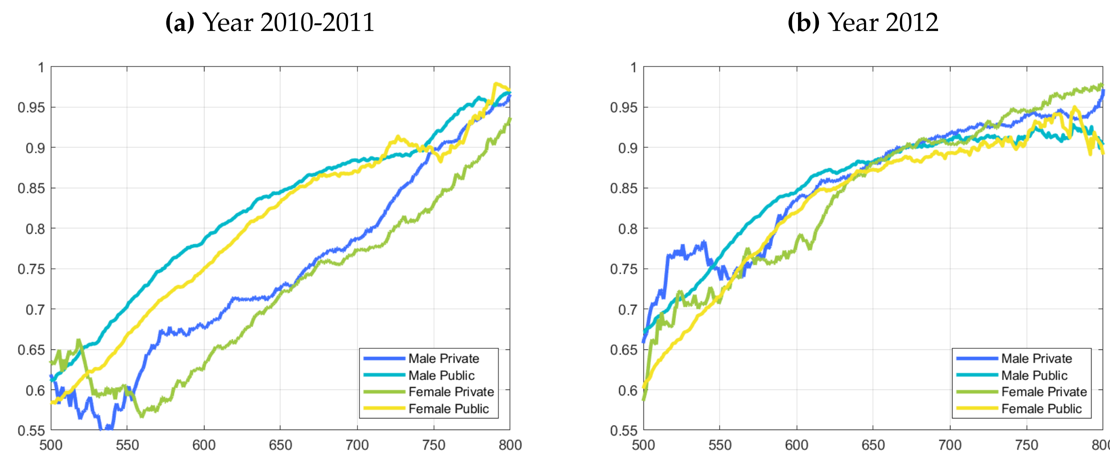
Figure 4: Enrollment Probability for On-Platform Assignments



Note: This figure shows the probability that a student assigned to an option on the platform, accepts and enrolls in that option. The lines show conditional means within 0.2σ standardized test score bins (plus and minus 0.1σ points). The probability of enrollment increases substatially for assignments that ocurr in 2012 relative to the previous two years.

We also repeat this exercise conditioning on gender and high school type in Figure 6 and Figure 5a, respectively. In all cases, there are significant differences between enrollment rates in 2012 and previous years, with larger impacts on mid-range-score and private school applicants.

Figure 5: Enrollment probability, conditional on scores, gender, and type of school.

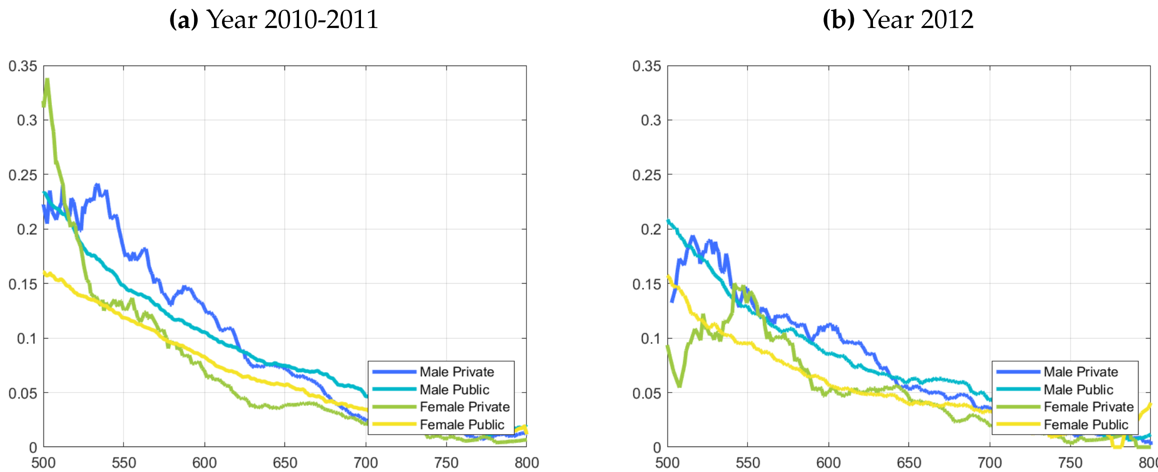


Note: This figure shows the probability that a student assigned to an option on the platform, accepts and enrolls in that option. The lines show conditional means within 0.2σ standardized test score bins (plus and minus 0.1σ points). The left panel shows the probability for four distint types oif students prior the policy change that increased on-platform options. The right panel shows the same probability for each type of student in 2012 when the platform was expanded. The probability of enrollment increases substatially for assignments that ocurr in 2012, especially for public school students.

Increasing enrollment rates fosters efficiency in the system, as waitlists are avoided, along with their potential inefficiencies. Another measure of inefficiency is the rate at which students dropout of the system once enrolled. If the quality of the matches generated goes up, we might expect to see less students dropping out over time. When exploring the evolution of freshmen dropout rates in the system, we find a significant reduction of about 1.5 percentage points, which account for almost a 15% reduction in overall dropout by the end of the first year of college.

?? shows how the probability of dropout varies over time and by gender, scores and type of school. It can be seen that the majority in the increases student retainment appear to be driven by female and public-high-school G25 enrollees. The reduction in dropout rates for males is slightly lower, and there is no significant change in dropout rates for private-high-school freshmen.

Figure 6: Freshmen dropout rate for G25 enrollees, conditional on scores and gender



Note: This figure shows the probability that a student that enrolled in an assigned option dropped out of school by the beginning of their second year of studies. Low scoring public school students seem to benefit more and drop out less after the platform in enlarged.

We build upon the previous descriptive findings by estimating descriptive probability models. These estimations do not aim to pinpoint the causal impact of the policy on any outcome, but rather provide more descriptive correlations. All of our stastistical models are either linear probability models or nonlinear Probit models of the form:

$$Y_i = \alpha + \beta \times \mathbf{1}[AppYear_i = 2012] + \mathbf{B}\mathbf{X}_i + \varepsilon_i, \quad (1)$$

where Y_i is some outcome variable for student i, $\mathbf{X}_i$ is a vector of control variables including gender, high school type, GPA, test scores and student's region. In the case of our linear models, which we estimate by OLS, $\hat{\beta}$ recovers the difference in the conditional means Y_i when comparing year 2012 with other years.

Our first outcome of interest is the probability of being admitted in an on-platform option. The first two columns of Table 2 report the coefficients for the OLS and Probit-MLE estimates. These suggest that admission probabilities rose in around 12-13 percentage points, which would translate to a jump of over 15% in the probability of being admitted in a platform option.

An additional excersize is to attempt to account for the fact that admission is conditional on having applied through the platform. This introduces a potential participation bias which we attempt to partially correct by computing a simple two-step Heckman estimator for both of our previous specifications. Our exclusion variable in this case is an indicator for receiving *BEA*, which is a financial aid status that increases the value of participating on the platform as it may yield financial assistance and increases admission probabilities for some programs with special wait-

Table 2: OLS and Probit (marginal effects) estimates of on-platform admission probabilities

	(1)	(2)	(3)	(4)	(5)	(6)
Year 2012	0.12339***	0.13103***	0.12462***	0.15234***	0.09438***	0.09358***
	(0.00136)	(0.00149)	(0.00141)	(0.00198)	(0.00184)	(0.00181)
Male	0.04022***	0.04063***	0.04152***	0.04945***	0.06894***	0.06871***
	(0.00145)	(0.00143)	(0.00145)	(0.00166)	(0.00189)	(0.00188)
Private HS	-0.05787***	-0.06467***	-0.03804***	-0.03033***	-0.05945***	-0.05906***
	(0.00192)	(0.00214)	(0.00216)	(0.00287)	(0.00267)	(0.00265)
GPA	0.00064***	0.00063***	0.00058***	0.00061***	0.00082***	0.00081***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Verbal Score	0.00036***	0.00041***	0.00021***	0.00020***	0.00042***	0.00040***
	(0.00001)	(0.00001)	(0.00001)	(0.00002)	(0.00002)	(0.00002)
Math Score	0.00093***	0.00105***	0.00075***	0.00079***	0.00095***	0.00095***
	(0.00001)	(0.00001)	(0.00001)	(0.00002)	(0.00002)	(0.00002)
Observations	279093	279093	741760	741760	279093	279093

lists. The results of our sample-selection-bias corrected are in columns (3) and (4) of Table 2. Our estimates now imply a similar increase of 12-15 percentage points.

In columns (5) and (6) we replicate the estimation procedure, but with Y_i being an indicator that equals one when student i is admitted in her first preference, i.e. the program she ranked at the top of the program list she submitted to the platform. Our results suggest that the conditional probability of being accepted in a first pick increased by 9 percentage points, which corresponds to an increase of at least 22% in the likelihood of this event.

We then turn to estimate models that only consider students admitted through the platform. In the first two columns of Table 3 we take as outcome variable the probability of enrolling in the program assigned by the platform. The next four columns repeat the exercise, but conditioning on being admitted in a G25 program. The difference between columns (3)-(4) and (5)-(6) is that the former ones preserve the same outcome variable (enrollment on platform assignment) on this new subsample (G25 admits), while the latter ones study the probability of enrolling in *any* G25 program, conditional on being admitted on a G25 option. The difference is only due to waitlists or special admission processes, which are very limited in the Chilean system.

Platform enrollment rates rose by about 5 percentage points, i.e. by over 7.5% when considering the full sample. When focusing on those students that were admitted in G25 options, we observe an increase of about 7.5 percentage points in the odds of enrolling in their assigned option, and an increase of a bit over 6 percentage points in the probability of enrolling in any G25 university.

Table 4 shows analogous estimates, but now focusing on freshmen dropout rates of platform admits as outcomes of interest. In the first pair of columns, our outcome variable is simply the

Table 3: OLS and Probit (marginal effects) estimates of on-platform enrollment probabilities

	(1)	(2)	(3)	(4)	(5)	(6)
Year 2012	0.04788***	0.05061***	0.07508***	0.07744***	0.06346***	0.06494***
	(0.00201)	(0.00202)	(0.00218)	(0.00222)	(0.00192)	(0.00200)
Male	0.02523***	0.02554***	0.02892***	0.02923***	0.03451***	0.03500***
	(0.00210)	(0.00210)	(0.00224)	(0.00223)	(0.00199)	(0.00197)
Private HS	-0.03104***	-0.03091***	-0.05727***	-0.05755***	-0.07568***	-0.08004***
	(0.00285)	(0.00296)	(0.00317)	(0.00330)	(0.00274)	(0.00296)
GPA	0.00038***	0.00038***	0.00039***	0.00039***	0.00040***	0.00040***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Verbal Score	0.00047***	0.00047***	0.00043***	0.00043***	0.00035***	0.00036***
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Math Score	0.00108***	0.00110***	0.00103***	0.00105***	0.00112***	0.00120***
	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)	(0.00002)
Observations	228378	228378	199469	199469	199469	199469

first-year dropout probability of any student admitted in the platform. In the last pair of columns, we condition the sample to consider only students that enrolled in a G25 option. These estimates suggest that freshmen dropout rates fell by about 2 percentage points after the policy was implemented. This accounts for at least a 16% drop in the odds of dropping out of a program after the first year of enrollment.

Finally, in our last descriptive exercise we study how the probability of enrolling in a waitlisted program changed with the implementation of the policy. That is, our dependent variable is 1 when a student enrolls in a program in which she was waitlisted in the platform. We do not find any important change in the odds of getting off a waitlist and enrolling in more-desired platform option. Thus, our estimates suggest that this ex-post probability of enrollment stayed constant over time and was not affected by the policy.

Table 4: OLS and Probit (marginal effects) estimates of on-platform dropout probabilities

	(1)	(2)	(3)	(4)
Year 2012	-0.01912***	-0.02016***	-0.01675***	-0.01818***
	(0.00154)	(0.00158)	(0.00165)	(0.00171)
Male	0.02467***	0.02477***	0.02458***	0.02485***
	(0.00165)	(0.00163)	(0.00175)	(0.00173)
Private High School	-0.00134	-0.01065***	0.00387*	-0.00602**
	(0.00186)	(0.00242)	(0.00205)	(0.00275)
GPA	-0.00022***	-0.00022***	-0.00023***	-0.00023***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Verbal Score	-0.00007***	-0.00009***	-0.00008***	-0.00009***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
Math Score	-0.00052***	-0.00059***	-0.00051***	-0.00058***
	(0.00001)	(0.00002)	(0.00002)	(0.00002)
Observations	141028	141028	126621	126621

4 Discrete Choice Model

4.1 Theoretical Model

In order to estimate the welfare impacts of the policy change, and to assess which programs' participation decisions had the largest impacts, we estimate a model of students' demand for on- and off-platform colleges. Our goal is to provide a tractable framework that uses variation in students' choices and enrollments around the policy change to identify key frictions in the partially-decentralized market. Accordingly, we model students' initial on-platform applications, their resulting placements and waitlist statuses, their decisions to accept positions in off-platform programs, the realization of waitlist offers, and students' ultimate matriculation decisions.

Our model has four stages, which we discuss below.

1. Students submit on-platform applications.
2. The DA procedure runs, and students receive initial placements and waitlist positions.
3. The aftermarket takes place. Students receive off-platform and waitlist offers and make final enrollment decisions.
4. Students enroll in programs. Production of human capital takes place.

We now describe the game in detail. A market $t \in T = \{2011, 2012\}$ is an application cohort. Within a market, N students apply to J different educational options. Each student $i \in \{1, ..., N\}$ is characterized by a tuple $(x_i, \eta_i, \varepsilon_i)$, which comprises observable covariates $x_i \in \mathbb{R}^K$, tastes $\eta_i \in \mathbb{R}^L$, and a random idiosyncratic preference-shock $\varepsilon_i \in \mathbb{R}^M$.

Each option $j \in \{1, ..., J_t\}$ is characterized by observable characteristics $x_j \in \mathbb{R}^m$. If student i attends an on-platform program j, he receives utility

$$u_{ij} = \delta_j + \lambda D_{ij} + \eta_{ij} + \epsilon_{ij}, \tag{2}$$

where D_{ij} is distance, $\delta_j = x_j \bar{\beta} + \xi_j$ is a mean utility term, and

$$\eta_{ij} = \sum_{\ell=1}^{M} \sum_{k=1}^{K} x_i^k x_j^m \eta_{\ell,k}^o + \sum_{\ell=1}^{L} x_j^l \eta_{i,\ell}^u \tag{3}$$

is a measure of match quality that depends on observed interactions of student and program characteristics as well as unobserved tastes.

There is also an outside option, $J = 0$, whose value is given by

$$u_{i0} = \max\{u_{i0}^0, u_{i0}^1\},$$

where

$$u_{i0}^0 = \epsilon_{i0}^0, \;\; u_{i0}^1 = x_i \beta^{oo} + \epsilon_{i0}^1.$$

u_{i0}^0 is to be interpreted as the value of the best noncollege alternative, denoted j_0^0 which is known at the time of applications. u_{i0}^1 is the value of the best nonselective or noncollege alternative, denoted j_{i0}^1, that is not known until the beginning of the aftermarket.

Programs are partitioned into on- and off-platform programs. Let $J_t^{\text{on}} \subseteq J_t$ denote the set of on-platform programs in market t, and $J_t^{\text{off}} = J_t \setminus J_t^{\text{on}}$ the set of off-platform programs.

In the first stage of the game, students learn their preferences for all programs except u_{i0}^1, then submit rank-ordered application lists to a centralized mechanism. On-platform programs use eight test scores and high school GPA to generate an index, $sIndex_{ij}$. The programs rank students according to this index. We let r_{ij} denote applicant i's rank on college j's list. A deferred acceptance procedure is used to conduct the initial on-platform match. Each program has a fixed number of slots in this mechanism. Given a realization of preferences, capacities, and students' applications, we let π_j denote the lowest index that was extended an offer by program j. Each student is assigned to his highest-ranked program at which his score is above the cutoff. Moreover, each program maintains a waitlist of length R. The R highest-ranked students who apply to program j and are not admitted to j or programs that they prefer to j are placed on j's waitlist.

We now consider the aftermarket. Off-platform programs $j \in J_t^{\text{off}}$ rank students according to $sIndex_{ij}$. We allow the formula to differ from what they would use if they were on platform. Off-platform programs have cutoffs π_j at which they are indifferent between accepting and declining a student's application given the number of seats that they have and the expected demand and quality of applicants.

Students with $sIndex_{ij} \geq \pi_j$ therefore have the option to enroll in program $j \in J_t^{\text{off}}$. In addition, at this stage each student learns the taste shock ϵ_{i0}^1 for the outside option component u_{i0}^1. Students who prefer an off-platform program or the outside option to their on-platform placement will decline their on-platform placement, potentially leading to vacancies in on-platform programs. In turn, this causes on-platform programs to make offers to waitlisted applicants.

We model this aftermarket as a college-proposing DA procedure with a friction. Off-platform programs and on-platform programs make offers to sudents, who may decline or provisionally accept them. On-platform programs j give maximum priority to students who received an initial placement at j.[6] They rank the remaining students according to their position on the relevant wait-

[6] This guarantees that a student who receives an initial placement at j can keep that placement if he desires to do so.

lists. If a student is not waitlisted at on-platform j, he/she is not acceptable to j in the aftermarket. For each on-platform program $j \in J_t^{on}$ and student i on j's waitlist, let

$$a_{ij} \in \{0, 1\}$$

be in indicator for the event that j is able to successfully contact i. We assume

$$Pr(a_{ij} = 0) \equiv \alpha,$$

independently across i and j. When $a_{ij} = 0$, program j is unable to reach i, in which case i is dropped from j's aftermarket preference ordering.

The parameter α summarizes the extent of aftermarket frictions. When α is large, programs need to make many calls to fill a given vacancy, and thus are likely to leave gaps when they move down their waitlists.

We now describe optimal play. The student-proposing DA algorithm used for the initial on-platform match is strategyproof, and in principle applicants should report rank-order lists corresponding to their true preference rankings. In practice, applicants to centralized mechanisms may omit schools that they perceive as unlikely or irrelevant. [CITE stuff here]

We assume that students truthfully report their preferences over programs at which they have nontrivial admissions chances. In particular, for each program, we define score bounds,

$$\overline{\pi}_j > \pi_j > \underline{\pi}_j.$$

Say that a program j is

- **ex-ante clearly infeasible** for student i if $sIndex_{ij} < \underline{\pi}_j$.
- **ex-ante marginal** for student i if $\underline{\pi}_j \leq sIndex_{ij} < \overline{\pi}_j$.
- **ex-ante clearly feasible** for student i if $\overline{\pi}_j \leq sIndex_{ij}$.
- **ex-post feasible** for student i if $sIndex_{ij} \geq \pi_j$.
- **ex-post waitlist-feasible** for student i if $sIndex_{ij}$ is higher than the lowest score on j's waitlist.

Suppose student i's' true preference ordering over J_t satisfies

$$u_{ij_1} > \ldots > u_{ij_K} > u_{i0}^0 > u_{ij_{K+1}} > \ldots > u_{ij_{J_t}}.$$

Let $\overline{u}_i^{\text{feas}}$ denote i's highest payoff among clearly feasible options, including the outside option component known at the time of applications:

$$\overline{u}_i^{\text{feas}} = \max\left\{u_{i0}^0, \max_{\{j\in J_t^{on}:\overline{\pi}_j\le sIndex_{ij}\}} u_{ij}\right\}.$$

Let

$$J_i^{\text{relevant}} = \{j \in J_t^{on} : sIndex_{ij} \ge \underline{\pi}_j \text{ and } u_{ij} \ge \overline{u}_i^{\text{feas}}\}$$

be the subset of on-platform programs that are not ex-ante clearly infeasible for i and not worse than the best clearly-feasible option.

We assume that i's report consists of all elements of J_i^{relevant} in the true preference order. At the end of the aftermarket, our assumptions imply that i enrolls at his most preferred program (including outside options) that makes him an offer.

4.2 Estimation

Let $v_{ij} \equiv \delta_j + \lambda D_{ij} + \eta_{ij}$ denote the sum of all utility components except for the idiosyncratic shock ϵ_{ij}. Let $lik_i(v_i, L, D)$ be the exploded-logit likelihood formula for preference list L and choice set D under deterministic utilities v: that is,

$$lik_i(v, L, D) = \left(\frac{\exp(v_{ij_1})}{1+\sum_{j\in D}\exp(v_{ij})}\right)\left(\frac{\exp(v_{ij_2})}{1+\sum_{j\in D\setminus\{j_1\}}\exp(v_{ij})}\right)\cdots\left(\frac{\exp(v_{ij_k})}{1+\sum_{j\in D\setminus\{j_1,\dots,j_{k-1}\}}\exp(v_{ij})}\right).$$

Suppose that student i submits list L_i to the initial match. Dropping programs which are ex-ante clearly infeasible for i, we obtain a sublist $L_i^{**} : (j_1, j_2, \dots, j_K)$, for some $K \in \mathbb{N}$. If some program in L_i^{**} is ex-ante clearly feasible for i, let k denote the index of the first such program. Otherwise let $k = K$. Let

$$L_i^* = (j_1, \dots, j_k).$$

The likelihood of i's application list is given by

$$lik_i^{app}(L_i^*, D_i^{on}) = \int lik_i(v_i, L_i^*, D_i^{on})\, dF(v_i|x_i, \theta) \tag{4}$$

where

$$D_i^{on} = \{j \in J_t^{on} : sIndex_{ij} \ge \underline{\pi}_j\}$$

is the set of all on-platform programs that are not clearly infeasible for i.

Let $\mathcal{L}_i$ be the set of ordinal lists over

$$D_i \equiv D_i^{on} \cup \{j \in J_t^{off} : j \text{ ex-post feasible for } i\}$$

Let F_i be the set of waitlist outcomes and ordinal preference lists $(a \in 0, 1^K, L \in \mathcal{L})$ that are consistent with i's behavior. In particular, say that $(a, L) \in F_i$ if L is an ordered list of some subset of $J_i^{on} \cup J_i^{off} \cup j_0^0, j_{i0}^1$, the restriction of L to D_i^{on} is L_i^*, we have a_j=0 whenever j is not ex-post waitlist-feasible for i, and i matriculates at the most-preferred program according to L which is ex-post feasible or is ex-post waitlist feasible and satisfies $a_{ij} = 1$.[7]

Let W_i denote the set of programs that are ex-post waitlist feasible for i at which i is waitlisted. Assume that bandwidths are sufficiently wide that every ex-post waitlist-feasible schools is ex-ante marginal or clearly feasible. In this case $W_i subseteq D_i$.

The full likelihood of i's on-platform application and matriculation outcome is given by

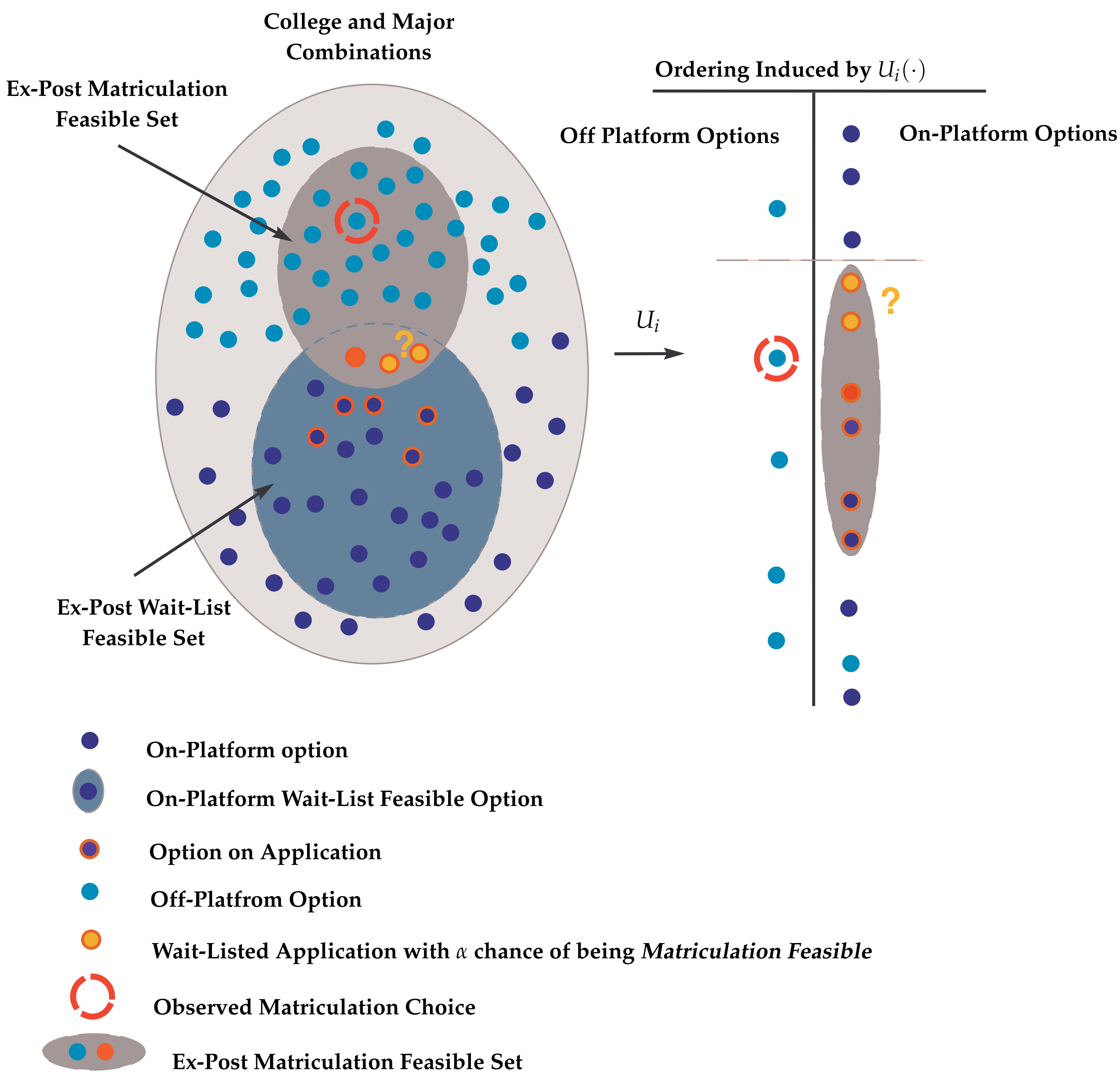
$$\ell_i = \sum_{(a_i, L) \in F_i} Pr(a_i|\alpha) \int lik_i(v_i, L, D_i \cup \{j \in W_i : a_{ij} = 1\} \cup \{j_{i0}^1\}) dF(v_i|x_i, \theta)$$

where

$$Pr(a_i|\alpha) = \prod_{\{j \in W_i \,:\, a_{ij}=1\}} (1-\alpha) \prod_{\{j \in W_i \,:\, a_{ij}=0\}} \alpha.$$

[7]If i enrolls in no program, then either j_{i0}^1 is preferred to the best feasible program, or j_0^0 is. If i received an initial placement in the match, this program must be preferred to j_0^0.

Figure 7: Diagram of Preference Ordering for Applications and Matriculation Choices



4.2.1 Results

In Table 5 we report select estimates for our more parameterized model for the likelihood lik_i which considers fixed effects by institution, by major and match effects between student's scores and course composition for each program, and random coefficients. All coefficients are divided by student type (male - private school, male - public school, female - private school and female -

public school). The result of the optimization for the after market friction reports an α of 0.638.

Table 5: Select Preference Estimates for Model with Observable and Unobservable match Effects

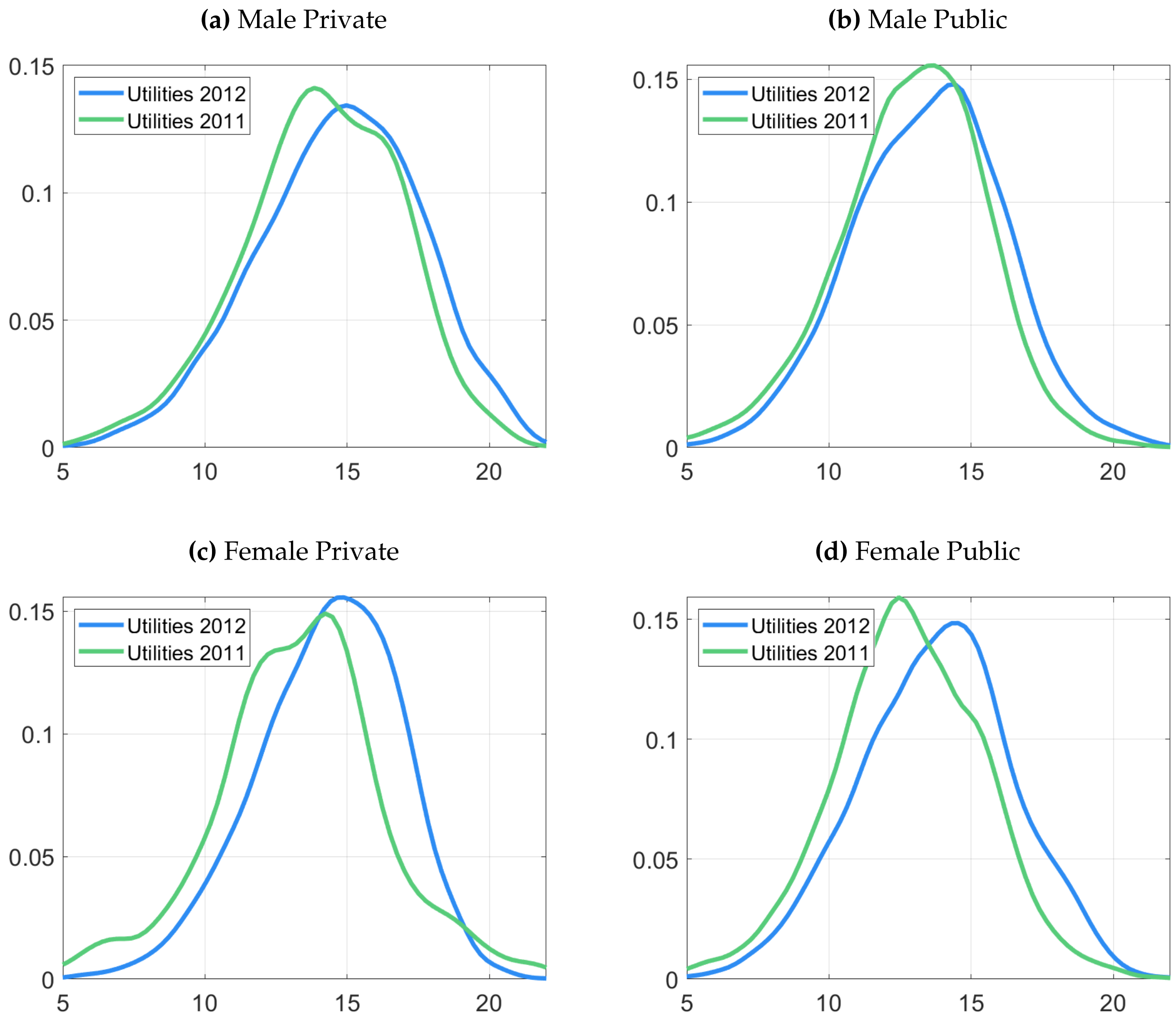
Prefereces (ψ^o)	Male Private	Male Public	Female Private	Female Public
Selectivity	1.475	1.525	1.330	1.444
Selectivity^2	1.395	1.222	1.596	1.464
Math x STEM	0.239	-0.218	0.057	-0.073
Math x Humanities	0.491	-0.220	0.129	0.159
Math x STEM^2	0.218	-0.129	0.040	-0.074
Math x Humanities^2	0.530	-0.106	0.039	0.110
Verbal x STEM	-0.347	-0.295	0.107	0.263
Verbal x Humanities	-0.064	-0.050	-0.010	0.265
Verbal x STEM^2	-0.058	-0.183	-0.008	0.212
Verbal x Humanities^2	-0.083	-0.336	0.061	0.480
Same City	2.653	3.330	2.465	3.251
Large City	0.490	-0.966	-1.196	-1.689
After-Market friction (α)	0.638			
Outside Option (u^0)				
Math	1.230			
Verbal	1.061			
Sex	5.727			
School Type	5.038			
Big City	1.086			
Variance Covariance (ψ^u)				
STEM (σ)	0.916			
Humanities (σ)	0.441			
Humanities vs STEM (ρ)	0.0013			

Note: Preference parameters were estimated via maximum likelihood and consider institution and major fixed effects using 2012 sample. The number of observations used for the regression are 162493 and the number of options are 1334

4.2.2 Welfare Gains of the Policy

To identify welfare gains of the policy we computed the utility value for the option in which each student ended up enrolling given our estimates of preference parameters in both samples. The utility distribution for 2012 sample is our counterfactual scenario after the policy change, while the distribution for 2011 is our benchmark. We derive those calculations by sub samples of student type. Welfare gains are easy to see from Figure 8.

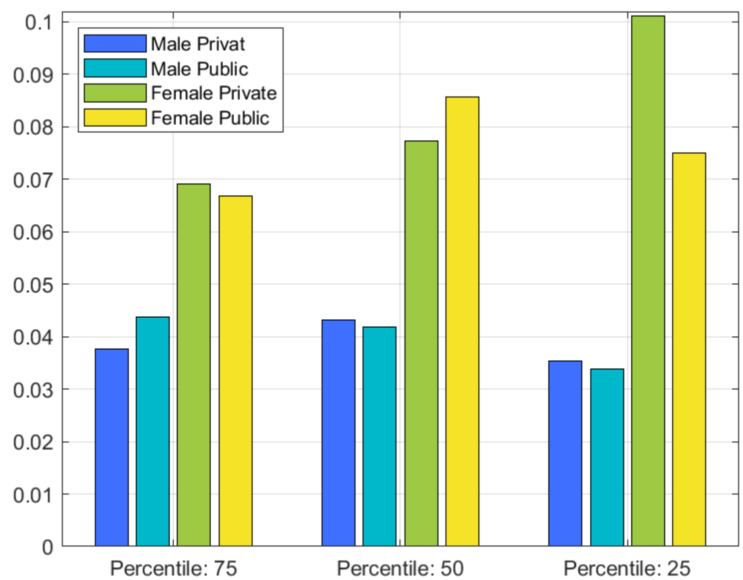
Figure 8: Distributional change in welfare after policy change in 2012



Note: The densities plotted in the figure are the enrollment utilities for students in years 2011 and 2012. All students that went to outside option were removed from this calculation. The x-axis correspond to the utility levels in the distribution and the y-axis are the kernel density estimation.

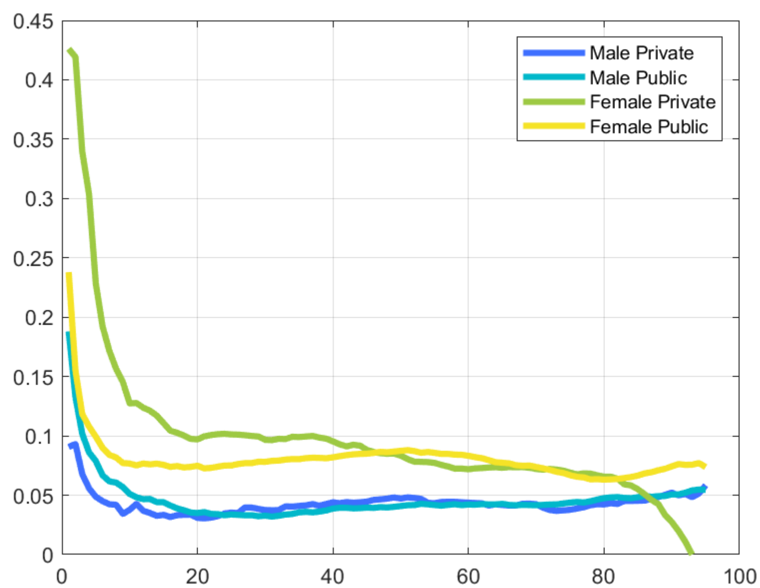
Figures 9 and 10 also evidence the welfare gain. As can be seen, the percentage increase occurs at all percentiles of the utility distribution and for all student types. However, the biggest winners of the policy change appear to be female students, in particular, those that came from private schools and reported low utility levels to begin with. Their utility increase after the policy change in arround 12% if they are in the lowest 25 percentile of the distribution. After them, all other student types seem to be increasing their utility levels around 5%.

Figure 9: Percentage increase in welfare by percentiles (75, 50 and 25)



Note: We divided utilities for both years 2011 and 2012 in percentiles 25, 50 and 75 and computed the percentage increse.

Figure 10: Percentage increase in welfare by percentiles across all utility distribution

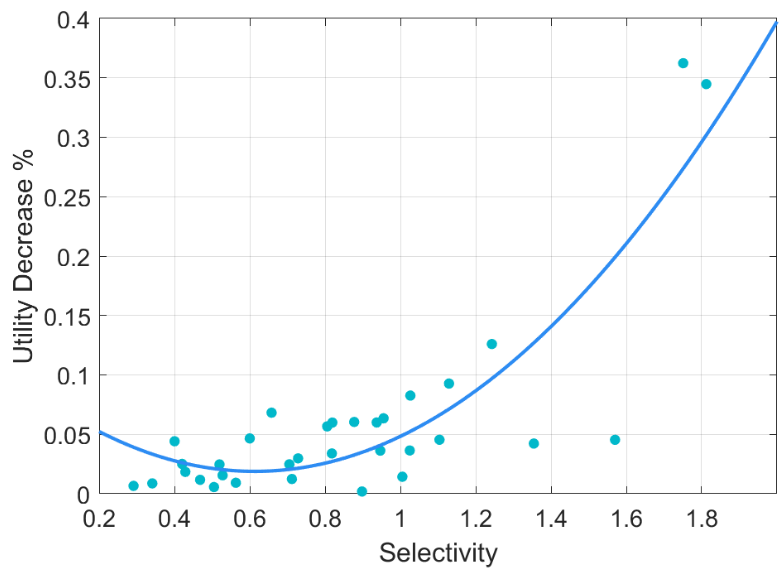


Note: Here we computed the percentage growth for the different levels of utility across all the percentile distribution. The x axis is the utility percentile where we make the comparison, and the y-axis correspond to the percentage increase in the utility in 2012.

5 Counterfactual Simulations

Given estimated parameters we computed what could be the average welfare loss of removing programs ordered by selectivity level in a context of after market frictions ($\alpha = 0.638$). Results are shown in Figure 11 and they suggest that the utility loss is higher if the most selective institutes are removed. However, the relation seems to be non lineal.

Figure 11: Utility loss of removing options ordered by selectivity



Note: Utility loss is calculated as the absolute value of the percentage change of the overall utility before removing the institution versus the overall utility after removing it.

6 Concluding Remarks

This paper studies the negative externalities off plaform options can generate for the efficiency of an assignment mechanism and overall welfare. The empirical results show that when off platform options were added in Chile, matriculation in placed slots rose by 8% and welfare increased by 6%. Dropout rates at the end of the first year of college dropped by 2 points (a 16% drop) showing that these estimated effects had real effects on outcomes policymakers care about. A postestimation decomposition shows that the lower scoring students, women and underpriviledged populations were the ones that most benefited from having more options on the centralized platform. Counterfactual analysis reveals that more desirable options cause bigger negative externalities with the most selective college leaving the platform generates 50% more welfare loss than the average college. These results show that off platform options can generate important costs and that considering these negative externalities can be a critical factor when planning a policy to implment a centralized assignment system. We show that empirical analysis can be helpful to guide policy discussions by quantifying key parameters that are needed to evaluate the potential costs of non participation from different actors. The type of options and the expected after market frictions can be evaluated to consider actions to mitigate this problem or to incentivize participation of the most important options. Our estimates provide a specific metric to evaluate the cost of losing each university on the platform. The model and the empirical strategy also highlights ways to quantify the costs of off platform options in other settings and hopefully a route to informing policy regarding the costs of off platform options.

References

Abdulkadiroğlu, Atila, Nikhil Agarwal, and Parag A Pathak, "The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match," *American Economic Review*, 2017, *107* (12), 3635–89.

Bordon, Paola and Chao Fu, "College-Major Choice to College-Then-Major Choice," *The Review of Economic Studies*, oct 2015, *82* (4), 1247–1288.

Lafortune, Jeanne, Nicolás Figueroa, and Alejandro Saenz, "Do you like me enough? The impact of restricting preferences ranking in a university matching process," *Working Paper*, 2016, (October).

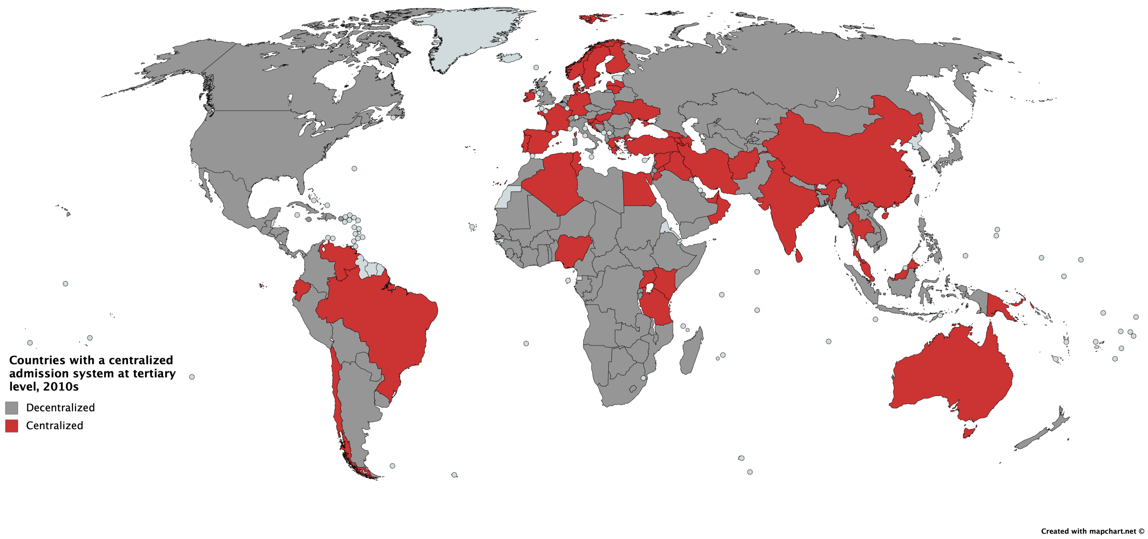
Neilson, Christopher A., "The Rise of Centralized Mechanisms in Education Markets Around the World," Technical Report August 2019.

Roth, Alvin E, "The economist as engineer: Game theory, experimentation, and computation as tools for design economics," *Econometrica*, 2002, *70* (4), 1341–1378.

— , "The Art of Designing Markets," *Harvard Business Review*, 2007, *85* (10), 118.

Appendix

Figure A-1: Use of Centralized Assignment Systems in Higher Education Across the World



Note: A large number of countries currently utilize centralized assignment mechanisms in higher education. Red countries indicate that the country has at least a subset of higher education options that are assigned by a centralized assignment mechanims. Virtually none of these platforms include all the higher education options.

Type of Institution	Number of Students	Average Score
Traditional Universities	53.00	601.2
Private Universities	39.71	550.7
Professional Institutes	21.96	505.2
Technical Institutes	12.85	496.2

Figure A-2: Test score distributions for applicants and non-applicants

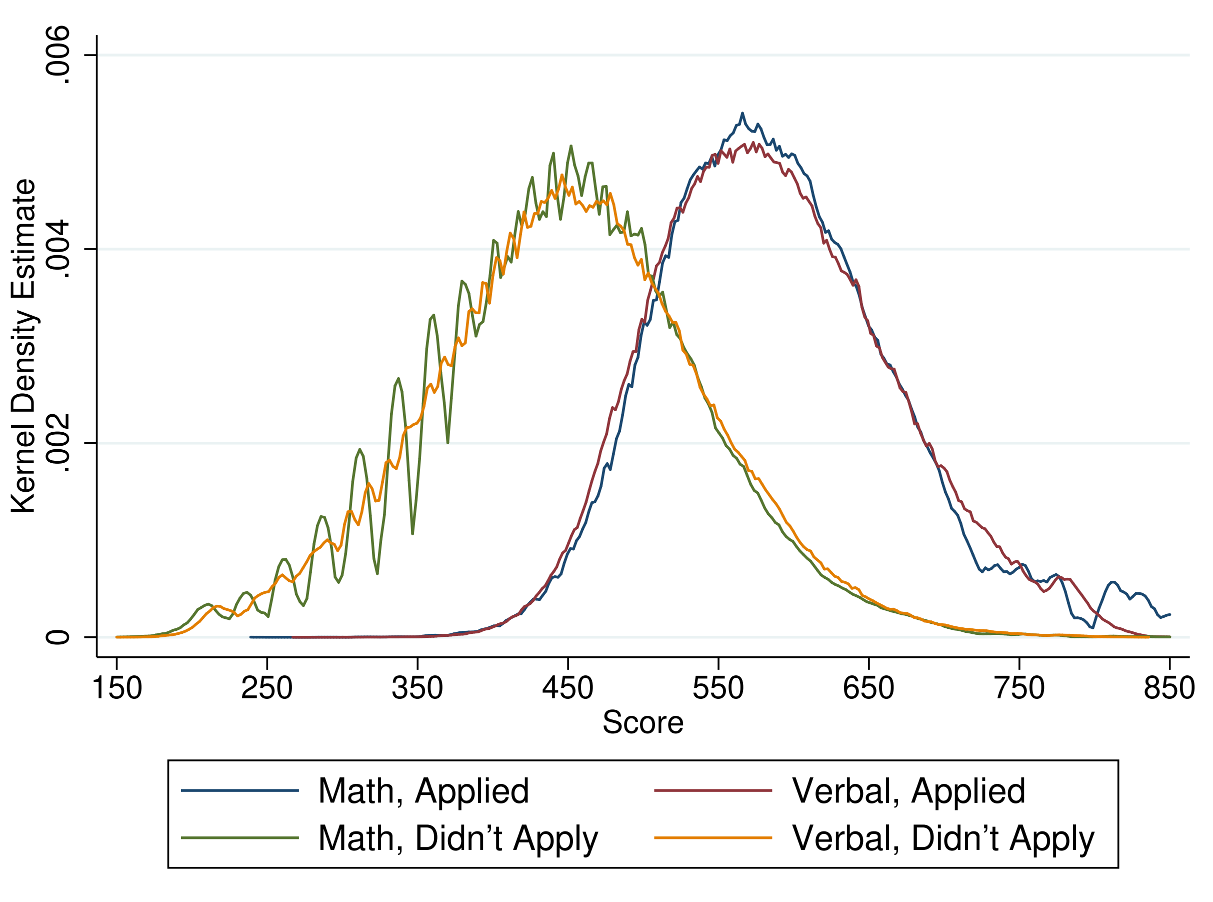


Figure A-3: Area share by score quartile with simulated data

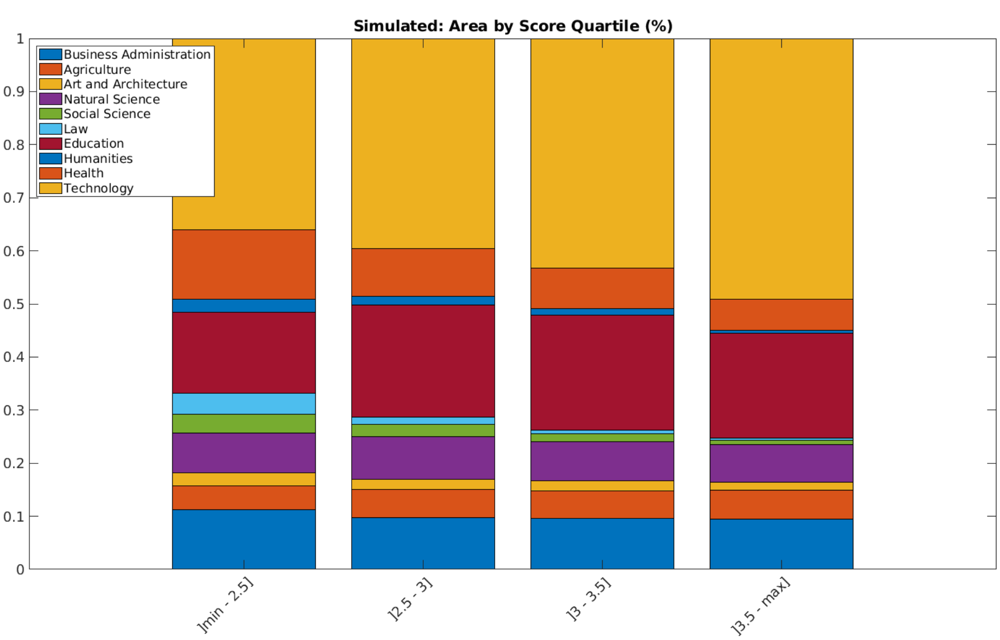


Figure A-4: Area share by score quartile with true data

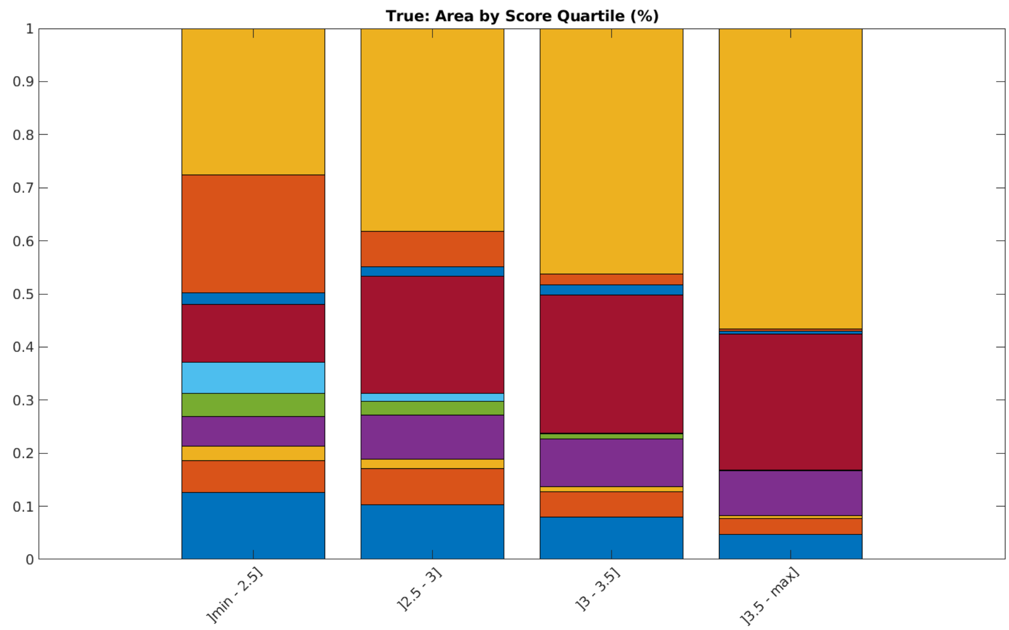


Figure A-5: Outside option shares by type with simulated data

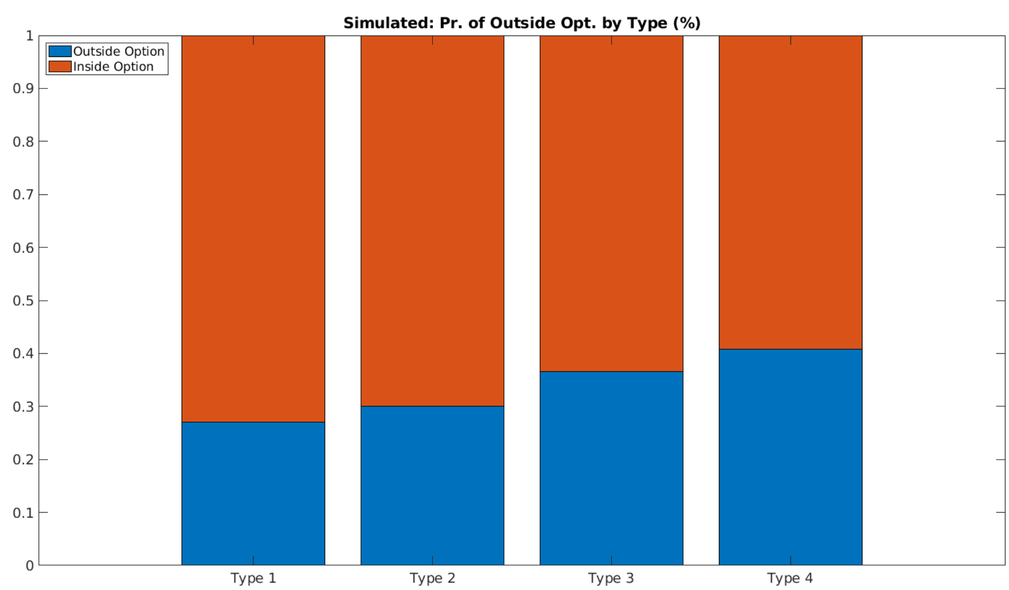


Figure A-6: Outside option shares by type with true data

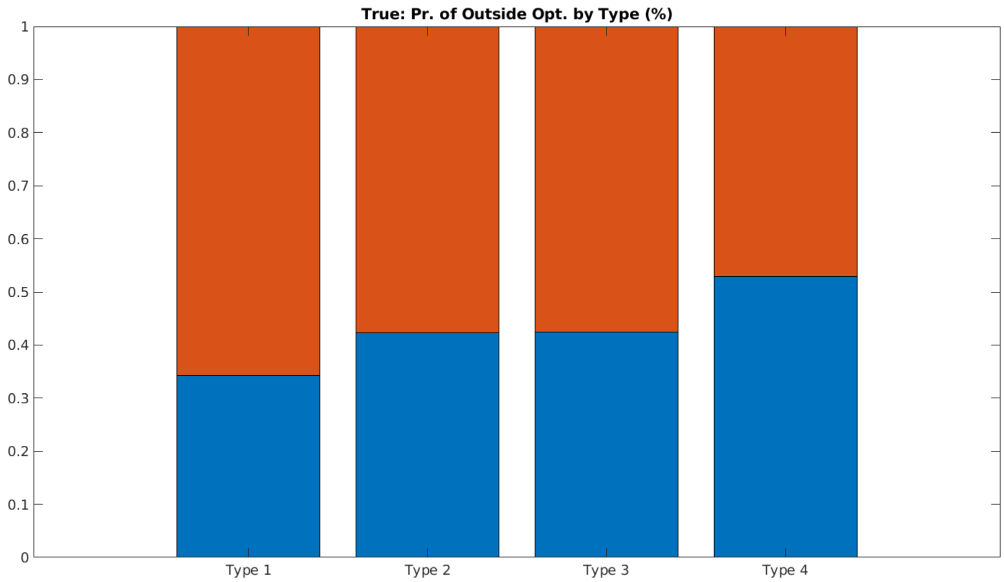


Figure A-7: Outside option shares by score quartile with simulated data

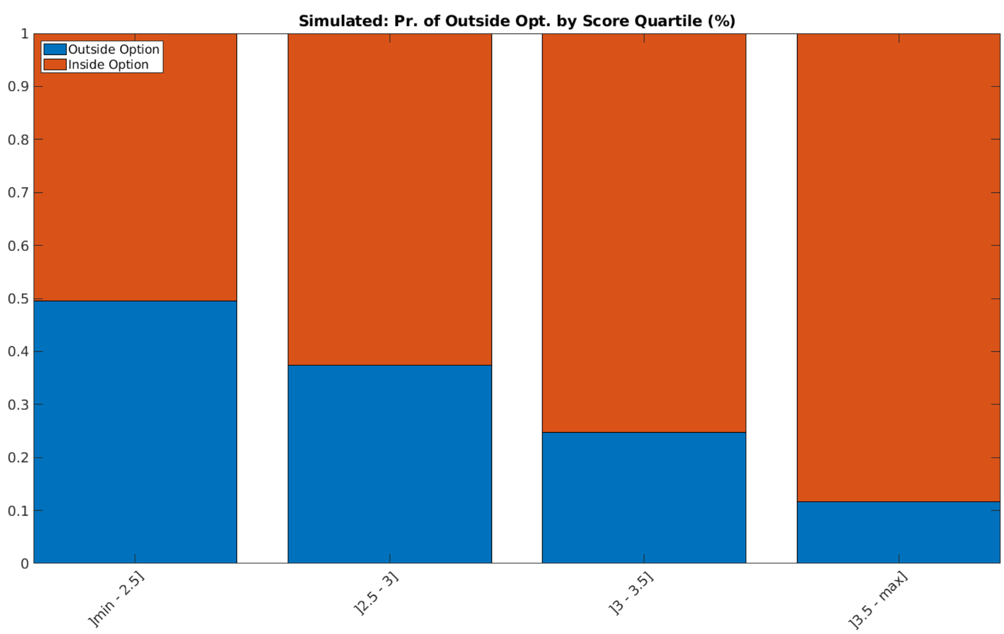


Figure A-8: Outside option shares by score quartile with true data

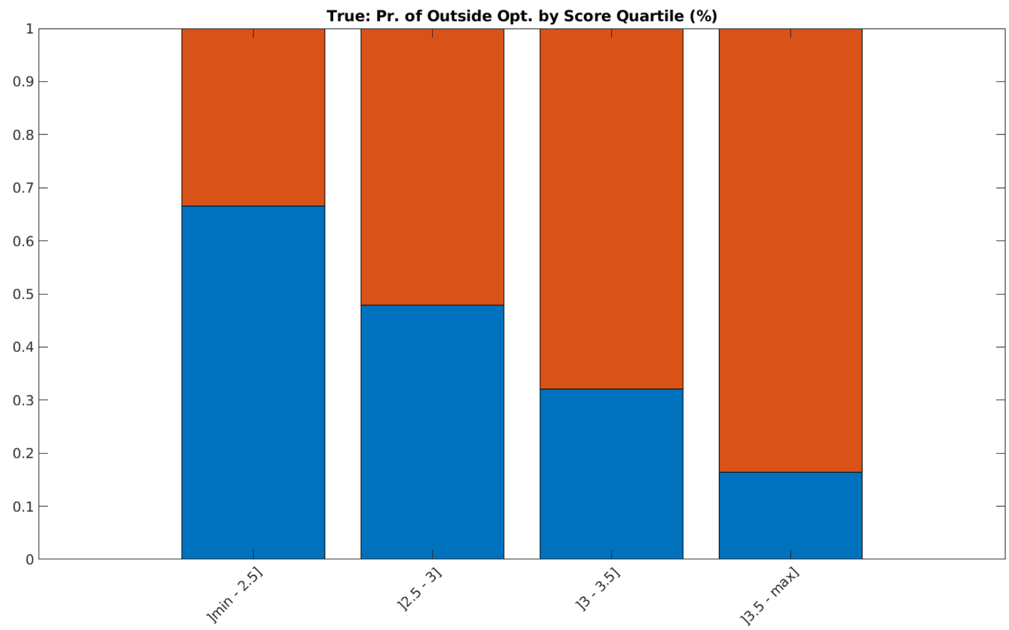


Figure A-9: Correlations between true and simulated FL shares