# 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 1, 2019

Most Recent Version

**Abstract**

In this paper, we study the empirical relevance of the negative externalities generated by off-platform options and quantify aftermarket frictions that contribution to generating them in practice. The empirical results show that when off-platform options were added in Chile, 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 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 aftermarket frictions. We estimate a model of college applications, aftermarket 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 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.

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 PLEASE DO NOT CITE.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 at least 46 countries have centralized assignment mechanisms for higher education. While theoretically appealing and successful in many settings, in practice many implementations of these systems diverge from ideal hypothesized settings. One common aspect of implementing centralized markets in practice is that in virtually every case there exist many off-platform options that are avaiable 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 as well as choices to not participate. It is known that these off platfrom options can cause negative externalities to the efficiency of the assignment mechanism. In practice, matches can be later reneged in favor of other off platform options and aftermarket frictions can make waitlists impractical and inefficient.

In this paper we study the empirical relevance of the negative externalities generated by off-platform options and quantify aftermarket frictions that contribution 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 worlds longest running college assignment mechanims[1]. We take advantage of a recent policy change that significantly increased the number of on platform institutions from 25 to 33 and raising the number of slots from approximatly 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 effeciency 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 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 adversly affected by the innefficency created by off platform options. This can be due to partly to their higher sensitivity to price and lower utility for private off platform options. We then use

[1] Central assignment based on a Deffered Acceptance algorthm began in the 1970s.

the estimated model to evaluate how our results would change in counterfactual excersizes 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 implment 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 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.

## 2 The Centralized Admission System

### 2.1 Context and Setting

Over last three decades, Chilean higher education has experienced a boom in enrollments. Several structural policy changes that targeted students and institutions have helped to shape what is now a robust and thick educational market. Attending college is nowadays much more of a standard than what it was in the 90's, where higher education was a privilege enjoyed only by a small subset of the Chilean population. Thanks to various vouchers, low-interest credits and scholarships, over a million students are now enrolled in higher education programs, and about a half attend Chilean universities.[2]

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 special ones, which are pooled together with all 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.

[2] Other higher education institutions consider professional institutes (IP) and technical formative centers (CFT).

### 2.2 The Centralization Policy

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

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

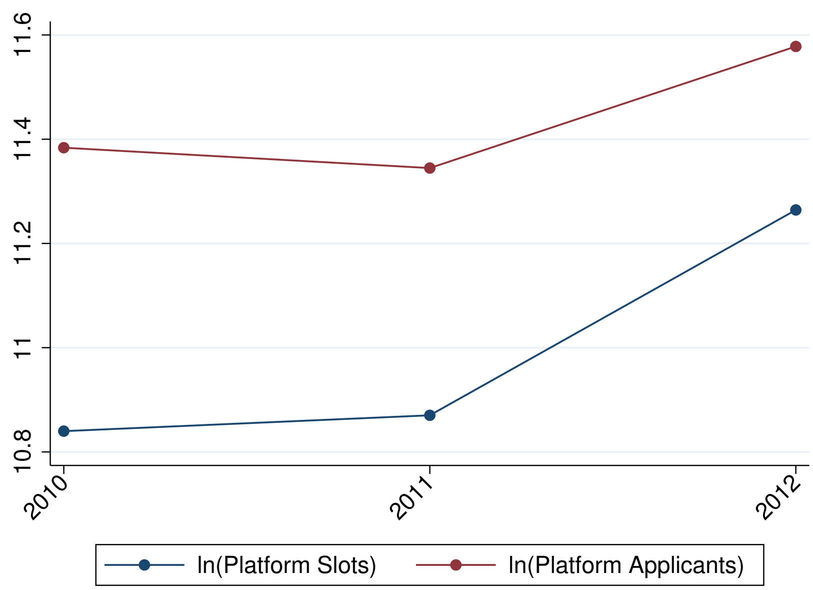
Our data spans all academic years from 2010 to 2012, and we report descriptive statistics for all our most relevant student-level variables for each year in Table 1. With this dataset we are able to observe some meaningful changes in the system after the addition of the G8 universities. Firstly, as commented 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

**Table 1:** Descriptive Statistics by Year

| | ***Year 2010*** | | | ***Year 2011*** | | | ***Year 2012*** | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std.Dev. | Obs | Mean | Std.Dev. | Obs | Mean | Std.Dev. | Obs |
| ***Pre-Application*** | | | | | | | | | |
| Male | 0.473 | 0.50 | 251634 | 0.476 | 0.50 | 250758 | 0.469 | 0.50 | 239368 |
| Private High School | 0.099 | 0.30 | 251634 | 0.100 | 0.30 | 250758 | 0.111 | 0.31 | 239368 |
| Metro Area | 0.646 | 0.48 | 251634 | 0.641 | 0.48 | 250758 | 0.641 | 0.48 | 239368 |
| GPA | 529.526 | 115.51 | 251634 | 531.640 | 110.07 | 250758 | 535.760 | 113.45 | 239368 |
| Math Score | 500.791 | 110.77 | 251634 | 501.075 | 111.27 | 250758 | 503.941 | 110.63 | 239368 |
| Verbal Score | 500.636 | 108.92 | 251634 | 501.035 | 108.34 | 250758 | 504.280 | 109.74 | 239368 |
| ***Post-Application*** | | | | | | | | | |
| Applied Through Platform | 0.349 | 0.48 | 251634 | 0.337 | 0.47 | 250758 | 0.446 | 0.50 | 239368 |
| Admitted in Platform | 0.763 | 0.43 | 87875 | 0.802 | 0.40 | 84512 | 0.877 | 0.33 | 106706 |
| Admitted in First Preference | 0.359 | 0.48 | 87875 | 0.424 | 0.49 | 84512 | 0.451 | 0.50 | 106706 |
| Enrolled in Platform | 0.637 | 0.48 | 67013 | 0.590 | 0.49 | 67803 | 0.624 | 0.48 | 93562 |
| Platform Program Dropout | 0.094 | 0.29 | 42657 | 0.102 | 0.30 | 40011 | 0.085 | 0.28 | 58360 |

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 1, 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%.

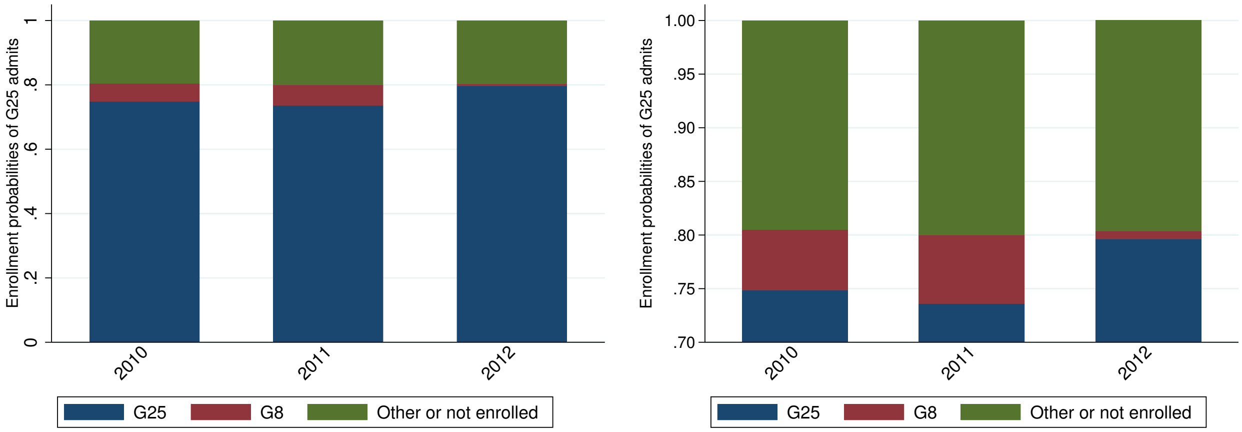
**Figure 1:** Platform log-slots and log-applicants




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 2. This effect is fully driven by the inability to deviate to another G8 option after 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.

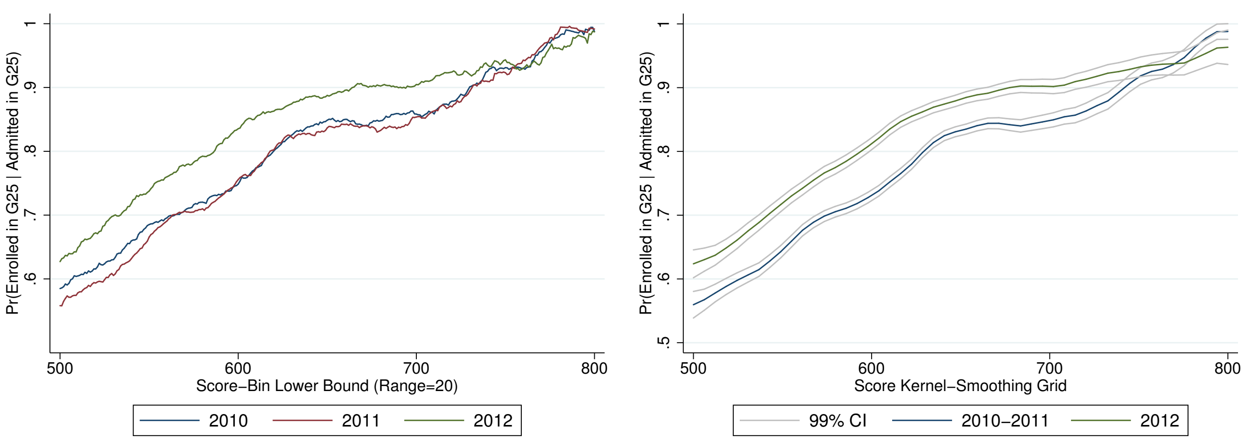
Moreover, we find that this G25 enrollment-likelihood effect is observed across the whole distribution of PSU scores, i.e. G25 admits are more prone to enroll into their admission programs over a very wide range of the test score's support. This is shown in the left panel of Figure 3, where sample probabilities 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

**Figure 2:** Enrollment probabilities for G25 admits




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 3:** Enrollment probability for G25 admits, conditional on scores




We also repeat this exercise conditioning on gender and high school type in Figure 4 and Figure 5, 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.

Increasing enrollment rates fosters efficiency in the system, as waitlists are avoided, along with their potential failures and injustices. Another well-known source of inefficiency is the dropout rate of of the system. 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% re-

**Figure 4:** Enrollment probability for G25 admits, conditional on scores and gender

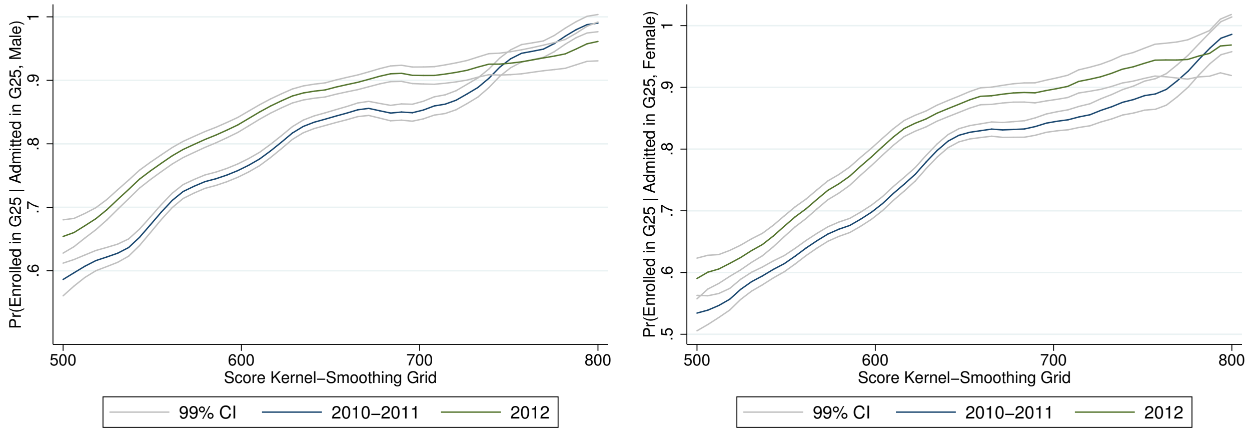
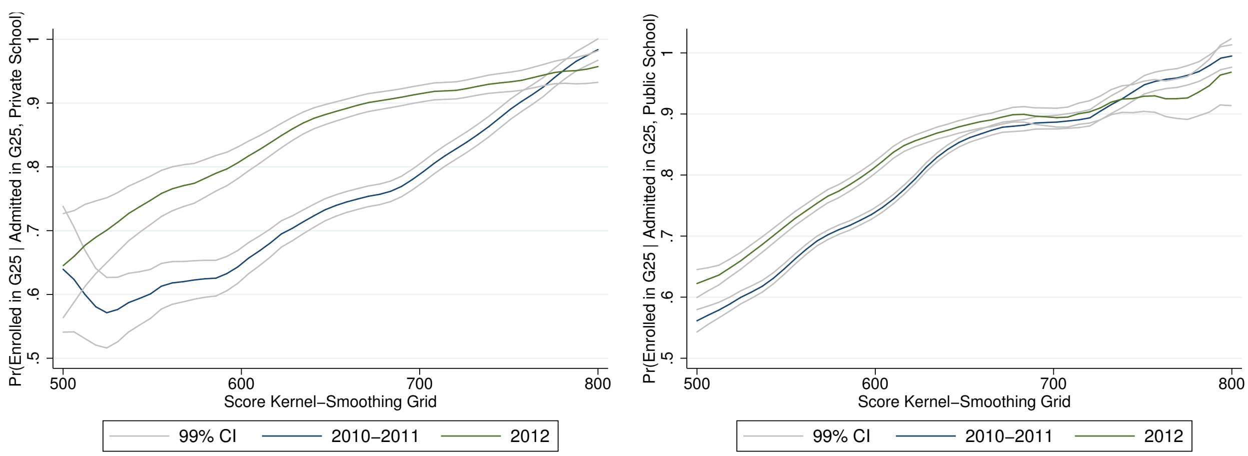



**Figure 5:** Enrollment probability for G25 admits, conditional on scores and school type




duction in overall dropout. Figure Figure 6 replicates Figure 3, but with freshmen dropout rates as relevant outcome variable. In this case, those who benefit the most from the policy appear to be G25 enrollees with mid-range scores between 520 and 700 points, comprising again the vast majority of admitted students.

We repeat this exercise by conditioning on gender and high school type in Figure 7 and Figure 8, respectively. Most of the efficiency gains 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

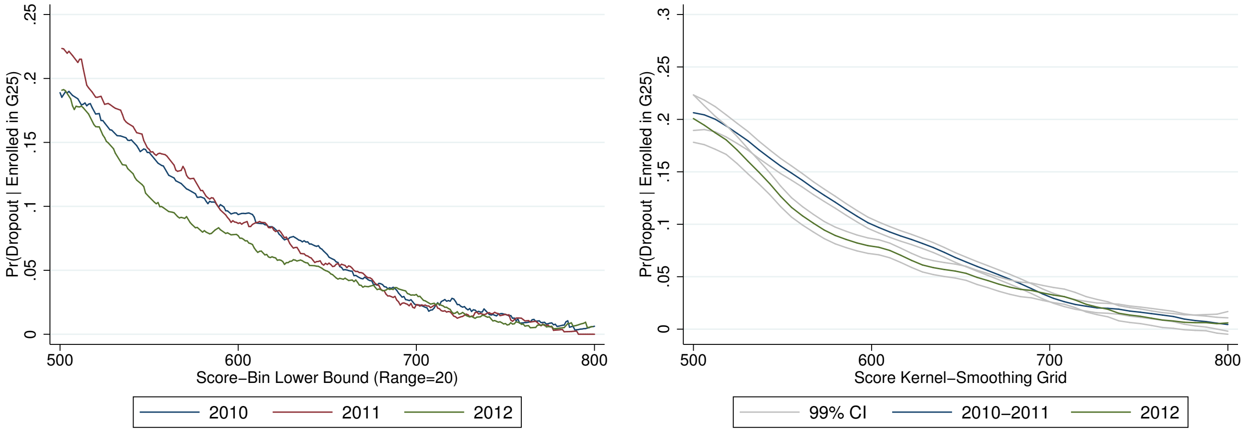
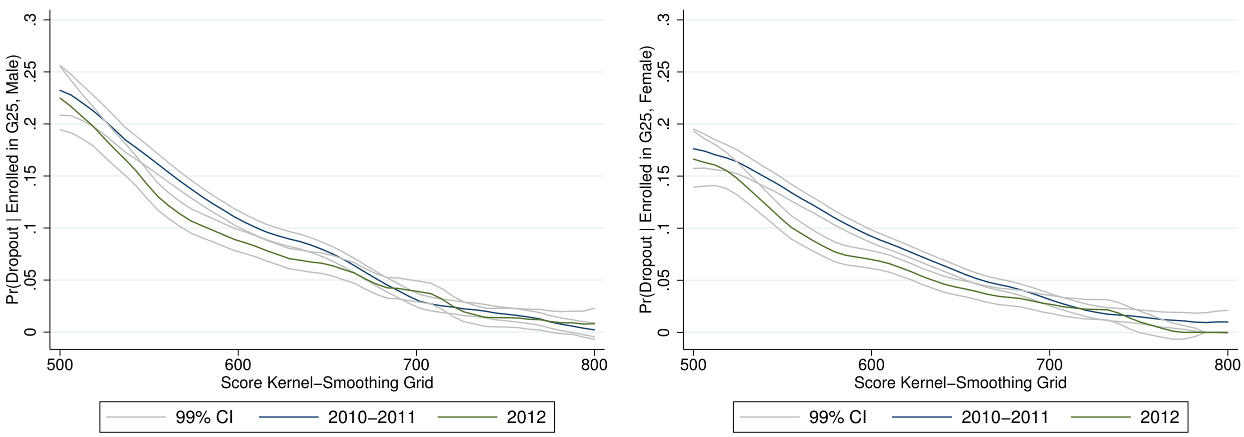



**Figure 7:** Freshmen dropout rate for G25 enrollees, conditional on scores and gender




## 3.1 Reduced Form Evidence

We formalize our the previous descriptive findings by estimating several reduced form specifications. These estimations do not aim to pinpoint the causal impact of the policy on any outcome, but rather provide some suggestive correlations that can be contrasted with our structural model's estimates.

All of our reduced form models are either classic linear regression models or Probit models, and all of them take the following form:

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

where $Y_i$ is some outcome variable for student $i$, $\mathbf{X}_i$ is a vector of control variables including gen-

**Figure 8:** Freshmen dropout rate for G25 enrollees, conditional on scores and school type

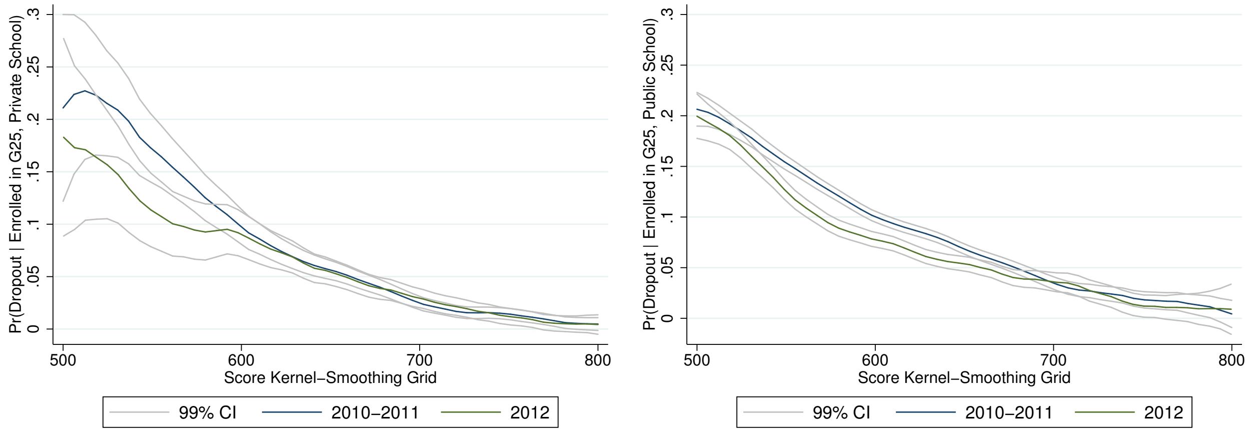



**Table 2:** OLS and Probit (marginal effects) estimates of on-platform admission probabilities

| | (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 |

der, 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.

To account for the fact that admission is conditional on having applied through the platform

**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 |

(see Figure A-2), we also partially correct for sample selection bias by computing a two-step Heckman estimator for both of our previous specifications. Our exclusion variable is an indicator of receiving *BEA*, which facilitates and incentivizes the application process, as in some cases it may yield financial assistance and also increase admission probabilities for some programs with special waitlists. 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.

**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 |

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 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 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.

# 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

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

$u_{i0}^0$ 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 $\pi_j$ 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 $\pi_j$ 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 $\epsilon_{i0}^1$ 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 $j$.[3] They rank the remaining students according to their position on the relevant wait-

[3]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 $\alpha$ summarizes the extent of aftermarket frictions. When $\alpha$ 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:

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

Let

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

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 $J_i^{\text{relevant}}$ 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

We estimate the model via maximum likelihood.

Let

$$v_{ij} \equiv \delta_j + \lambda D_{ij} + \eta_{ij}$$

denote the sum of all utility components except for the idiosyncratic shock $\epsilon_{ij}$.

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,

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

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 $L_i^{**} : (j_1, j_2, \ldots, j_K)$, 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

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

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

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

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 $a_{ij} = 1$.[4]

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

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

#### 4.2.1 Preference Estimation

Since the college allocation system uses a Defered Acceptance algorithm to match students to schools, it is natural to assume that preferences are revealed truthfully by students avoiding strategic behavior when they rank their 10 most prefered options. The result in equilibrium is that students are allocated to options they liked the most among all ex-post feasible options. This preference revealing algorithm allows us to estimate preference parameters neatly.

For preferences estimation, we use 2012 data to compute an exploded logit with random coefficients that reflects the theoretical utility maximizing behavior for students. We apply the joint probability to students apps from most prefered to least prefered alternatives among all ex-post waitlist feasible options. In this context, ex-post waitlist feasible options are those where the student applied and could have got into given their scores and after market cutoffs. Since we assume that there are no after-market frictions in 2012, the ex-post waitlist feasible options are equal to Ex-Post feasible options.

[4] 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 $j_0^0$.

#### 4.2.2 *After-Market Friction Estimation*

The estimation procedure for obtaining After-Market friction parameter, required two assumptions: (1) prefference parameters remain constant across years and (2) students have an utility shock when they decide between their placement or other options available during the After Market allocation.

The procedure therefore is to get $\alpha$ that optimizes the likelihood function $\ell_i$ subject to preference parameters estimated with 2012 sample and a utility shock that is independent of the utility shock for preferences that explains why students could be choosing revealed least prefered options than their current placement. We integrate the likelihood across all potential escenarios of callbacks given enrollment realizations during the aftermarket and optimize it via maximum likelihood.

#### 4.2.3 *Description of the Sample*

From the full 2011-2012 sample, we dropped all students that enrolled into non G8 off platform options, observations with missing values in any of the individual level characteristics such as scores in verbal and math and sex, and students with average scores in math and verbal below 450 points.

As can be seen in Table 5 the final number of students in our dataset for 2012 were 162493. They applied to a total of 1334 higher education programs; 1010 from to G25 institutions and 324 from G8 institutions. Whereas in 2011, 168044 students applied to a total of 1140 higher education options, arround 48% were male and 52% female. In both samples, arround 79% of students comming from private schools come also from a big city, meanwhile 64% of students comming from a public high school come from a big city.

In addition, we see that in both samples, students comming from private high schools seem to be going to their most prefered option relative to students from public high schools. Also we see an improvement after the policy change; there is a big increase in the share of students that got into their most prefered option, and a significantly smaller fraction of students going to the outside option. This pattern occurs across all student types after the policy change.

**Table 5:** Summary statistics for 2011-2012 sample

| | 2011 | | | | 2012 | | | |
|---|---|---|---|---|---|---|---|---|
| | Male Private | Male Public | Female Private | Female Public | Male Private | Male Public | Female Private | Female Public |
| Student Characteristics | | | | | | | | |
| Score Math | 647.52 | 556.36 | 620.31 | 531.67 | 643.26 | 556.80 | 618.54 | 534.19 |
| | (94.84) | (78.61) | (85.10) | (70.49) | (87.14) | (77.43) | (80.44) | (69.54) |
| Score Verbal | 647.52 | 556.36 | 620.31 | 531.67 | 643.26 | 556.80 | 618.54 | 534.19 |
| | (94.84) | (78.61) | (85.10) | (70.49) | (87.14) | (77.43) | (80.44) | (69.54) |
| Big City | 0.79 | 0.64 | 0.79 | 0.63 | 0.78 | 0.63 | 0.79 | 0.63 |
| | (0.41) | (0.48) | (0.41) | (0.48) | (0.41) | (0.48) | (0.41) | (0.48) |
| Placement Ranking (%) | | | | | | | | |
| $1^{st}$ | 26.97 | 23.28 | 25.34 | 19.32 | 40.56 | 32.80 | 40.64 | 27.66 |
| $2^{nd}$ | 10.98 | 8.85 | 10.26 | 8.58 | 19.82 | 11.66 | 19.87 | 11.57 |
| $3^{rd}$ | 5.79 | 4.61 | 5.54 | 4.54 | 10.59 | 5.72 | 9.81 | 6.29 |
| $4^{th}$ | 2.17 | 2.37 | 1.94 | 2.41 | 4.44 | 2.70 | 3.91 | 3.09 |
| $5^{th}$ | 1.06 | 1.20 | 0.88 | 1.36 | 2.06 | 1.37 | 1.78 | 1.76 |
| $6^{th}$ | 0.36 | 0.60 | 0.31 | 0.74 | 0.91 | 0.64 | 0.88 | 0.88 |
| $7^{th}$ | 0.20 | 0.32 | 0.16 | 0.35 | 0.36 | 0.29 | 0.47 | 0.53 |
| $8^{th}$ | 0.07 | 0.16 | 0.04 | 0.19 | 0.17 | 0.17 | 0.19 | 0.28 |
| $9^{th}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.12 | 0.08 | 0.10 | 0.12 |
| $10^{th}$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.04 | 0.05 | 0.07 |
| Outside Option | 52.40 | 58.61 | 55.53 | 62.51 | 20.94 | 44.52 | 22.29 | 47.78 |
| N Obs | 12196 | 70567 | 11212 | 74069 | 13030 | 65286 | 11881 | 72296 |
| % | 7.26 | 41.99 | 6.67 | 44.08 | 8.02 | 40.18 | 7.31 | 44.49 |

Note: ...

#### 4.2.4 *Results*

In Table 6 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 $\alpha$ of 0.638.

**Table 6:** Select Preference Estimates for Model with Observable and Unobservable match Effects

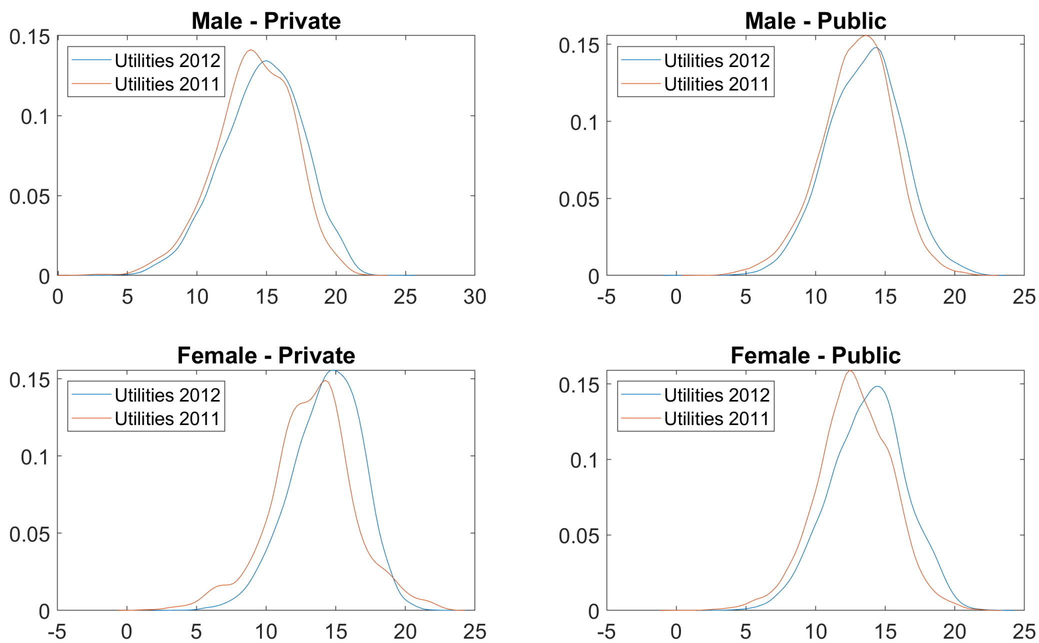
| Prefereces ($\psi^o$) | Male Private | Male Public | Female Private | Female Public |
|---|---|---|---|---|
| Selectivity | 1.475 | 1.525 | 1.330 | 1.444 |
| Selectivity$^2$ | 1.395 | 1.222 | 1.596 | 1.464 |
| Math x STEM | 0.239 | -0.218 | 0.057 | -0.073 |
| Math x Humanities | 0.491 | -0.220 | 0.129 | 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 ($\alpha$) | 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 ($\psi^u$) | | | | |
| STEM ($\sigma$) | 0.916 | | | |
| Humanities ($\sigma$) | 0.441 | | | |
| Humanities vs STEM ($\rho$) | 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

#### 4.2.5 *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 9.

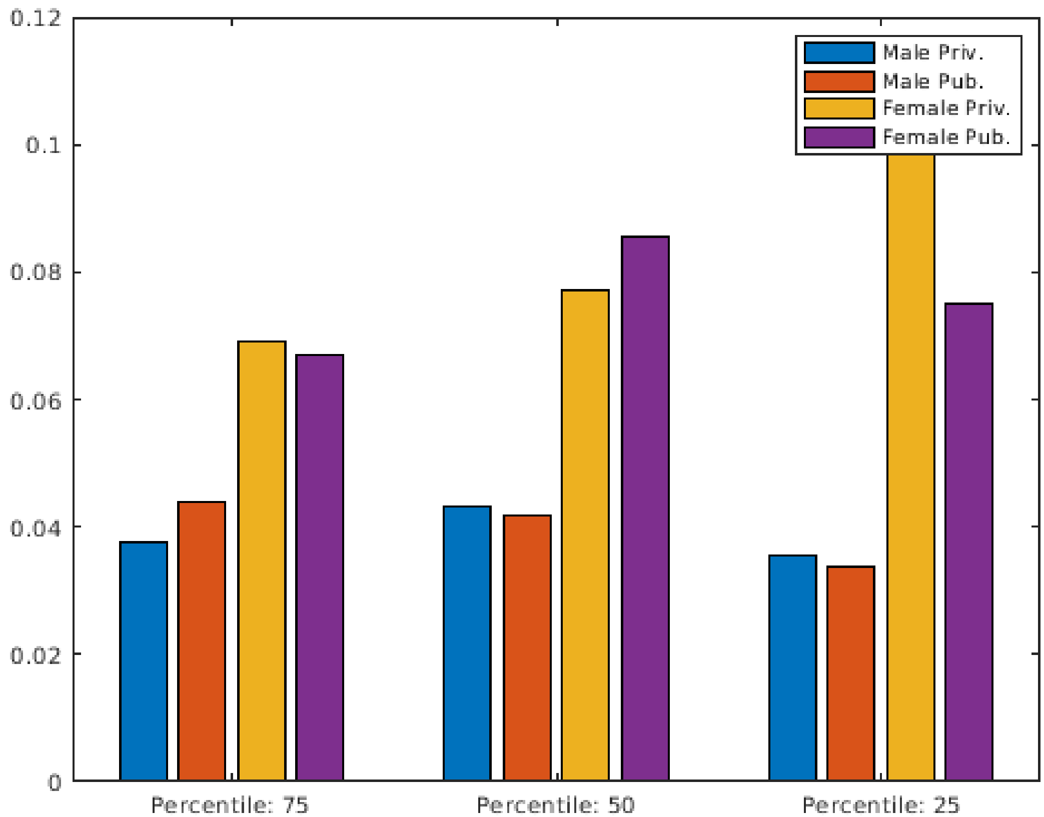
**Figure 9:** Welfare distribution by student type




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.

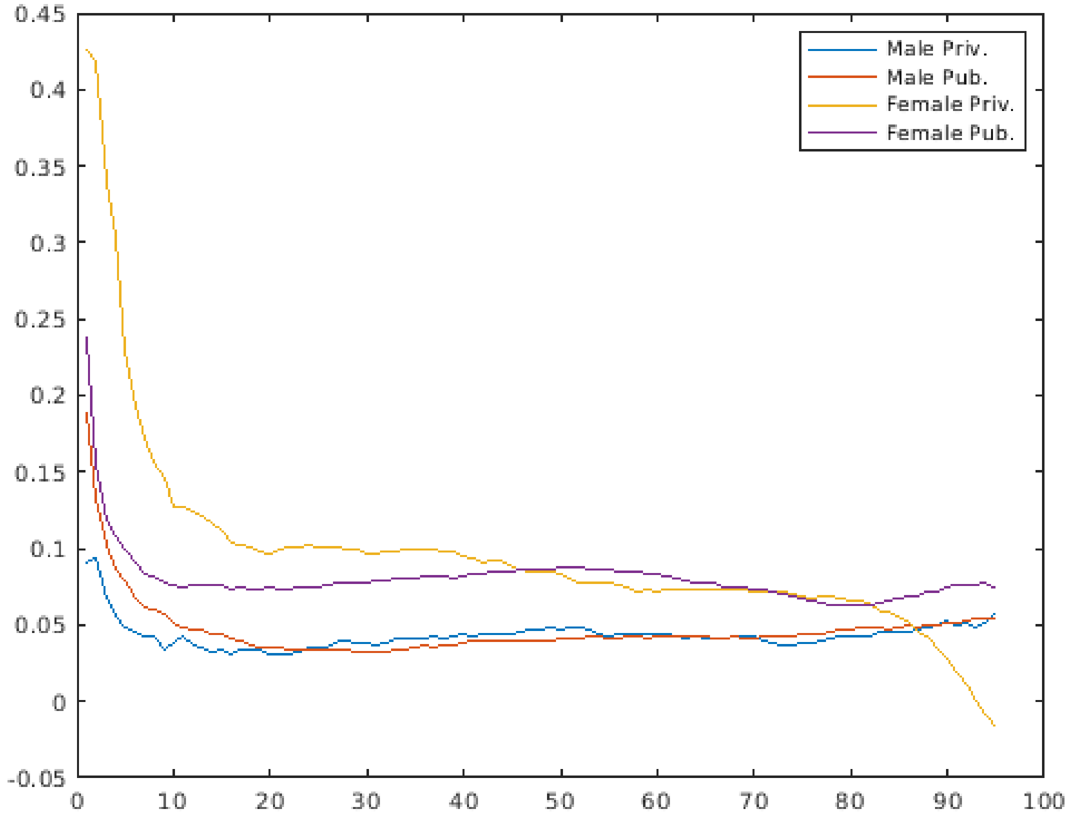
Figures 10 and 11 also evidence the welfare gain. As can be seen, the percentage increase occurs at all percentiles of the utility distribution and for all student types. However, the biggest winners of the policy change appear to be female students, in particular, those that came from private schools 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 10:** 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 11:** Percentage increase in welfare by percentiles across all utility distribution

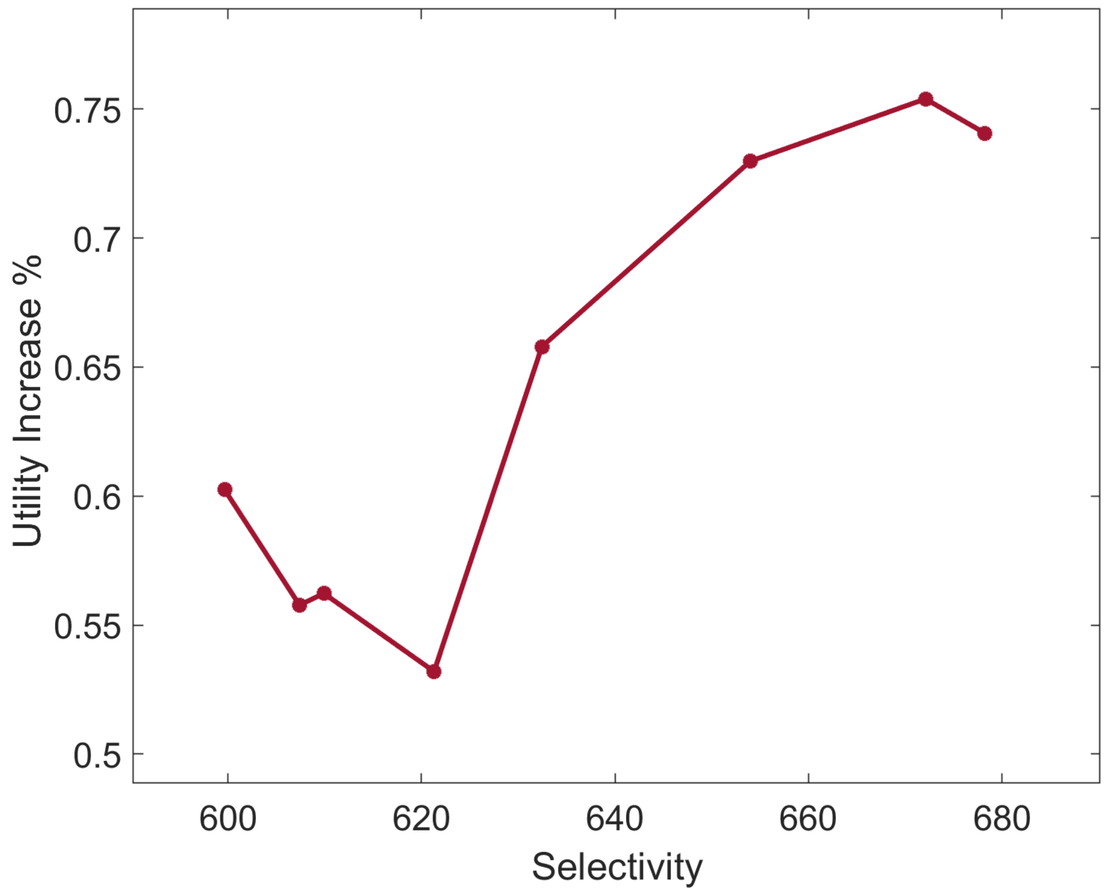



Note: Here we computed the percentage growth for the different levels of utility across all the percentile distribution. The x axis is the utility percentile where we make the comparison, and the y-axis correspond to the percentage increase in the utility in 2012.

# 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 12 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 12:** 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.

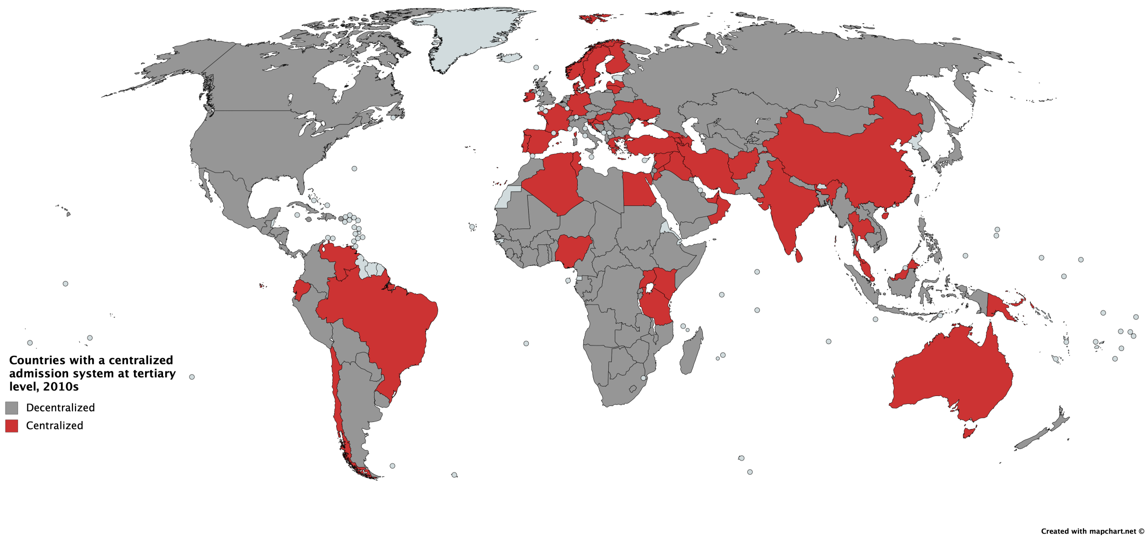
# 6 Concluding Remarks

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

# Figures and Tables

**Figure A-1:** Use of Centralized Assignment Systems 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.

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

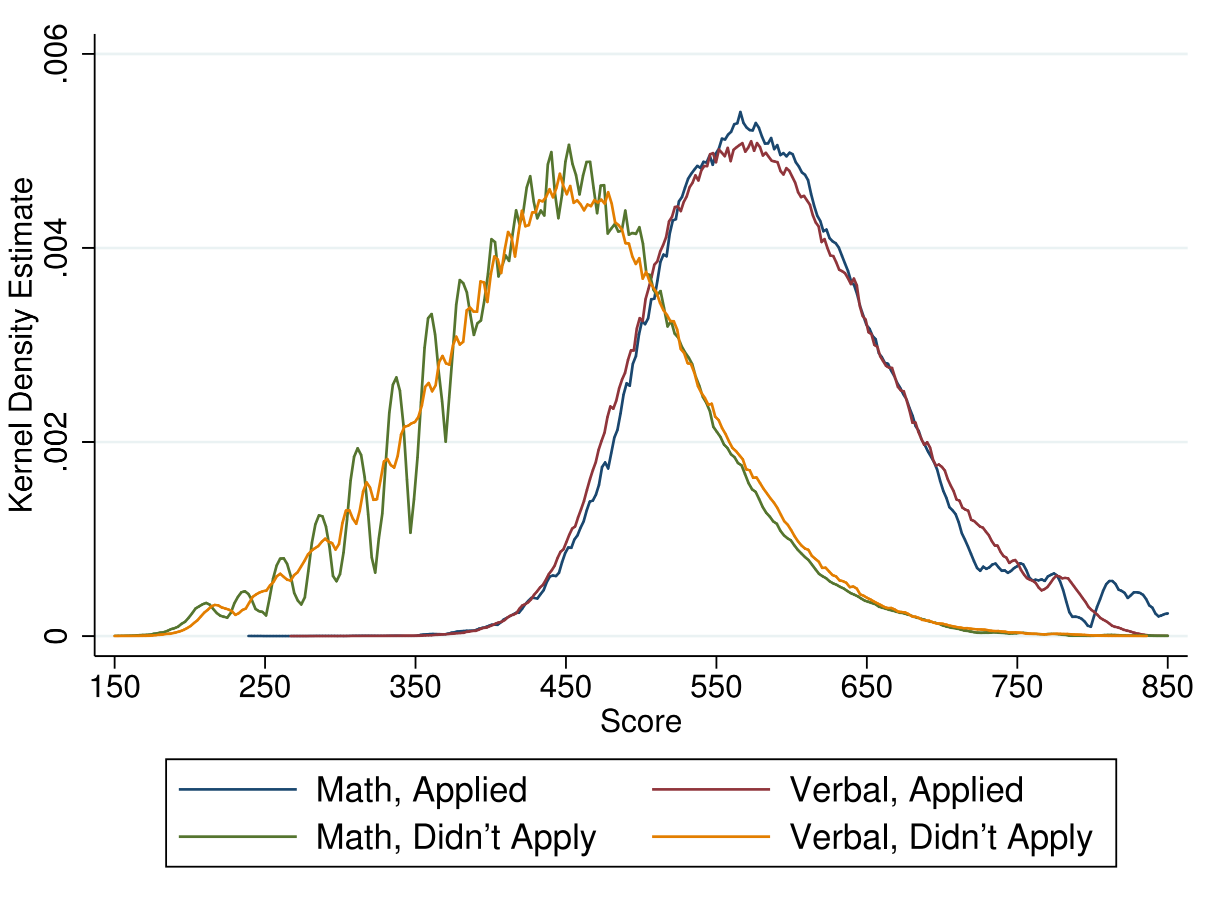



**Figure A-3:** Area share by score quartile with simulated data

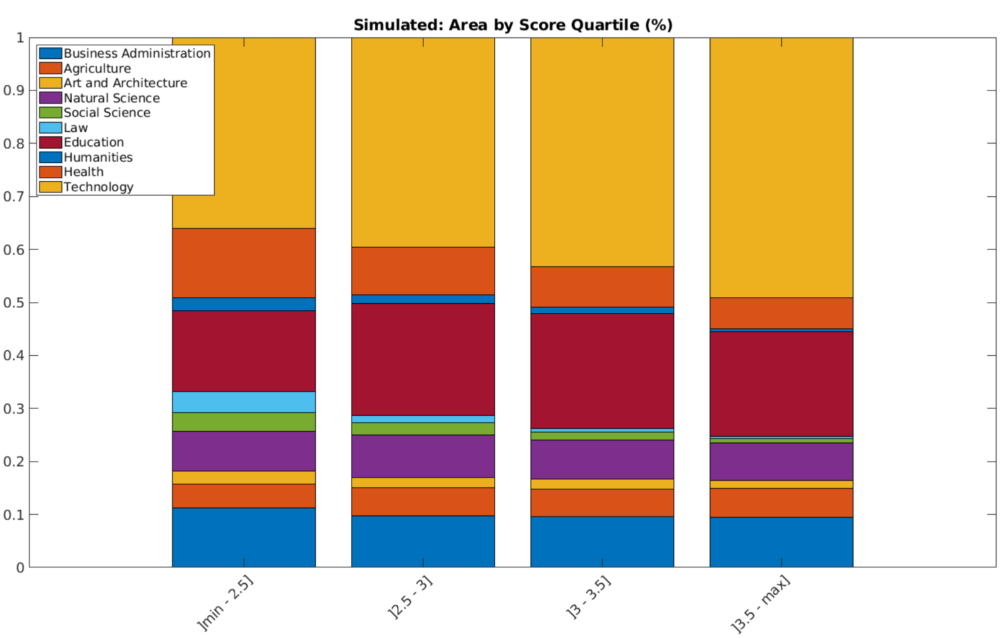


**Figure A-4:** Area share by score quartile with true data

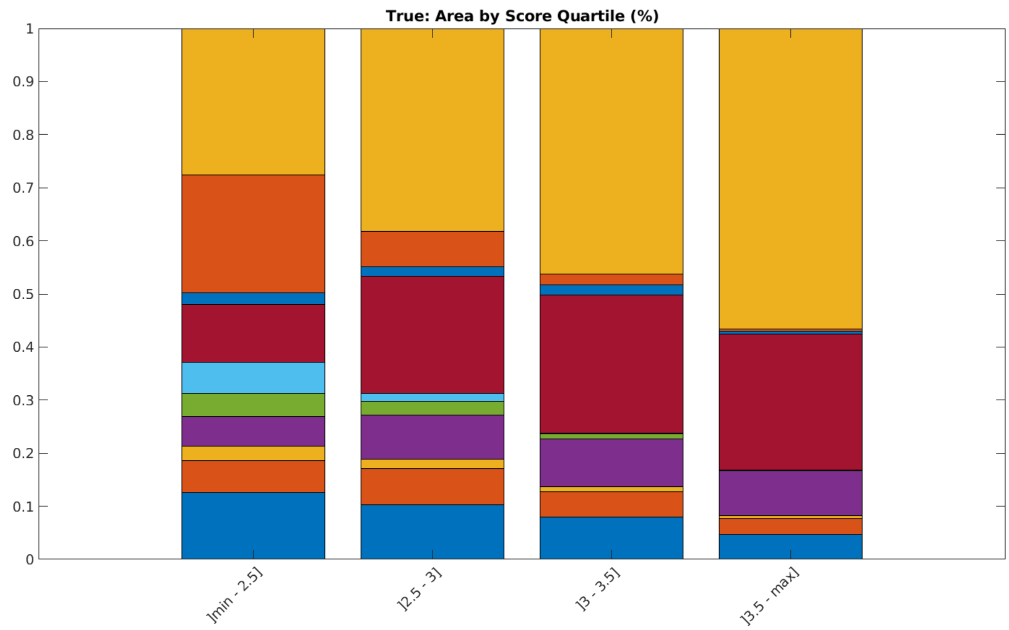



**Figure A-5:** Outside option shares by type with simulated data

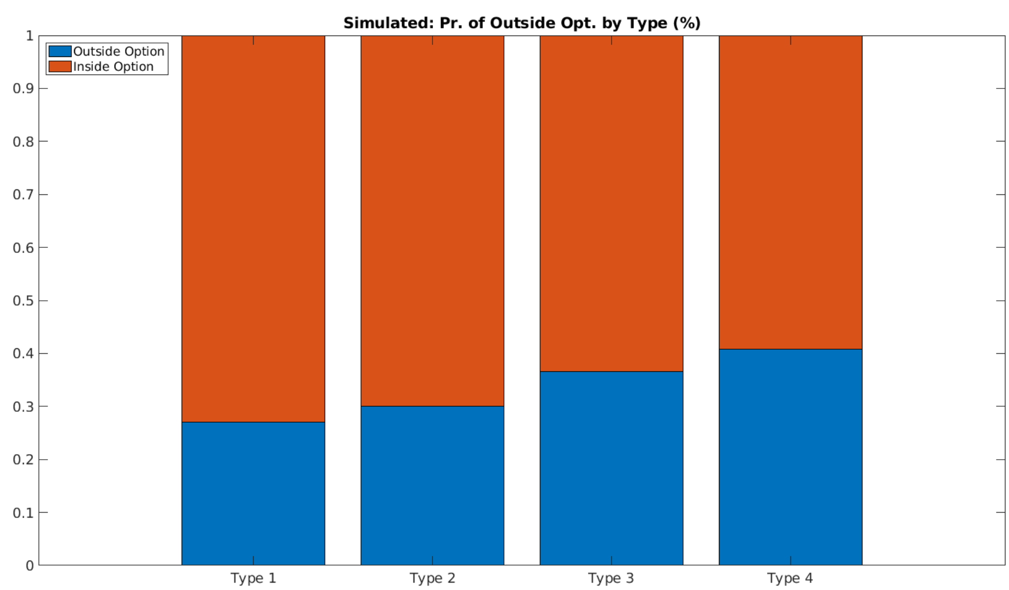


**Figure A-6:** Outside option shares by type with true data

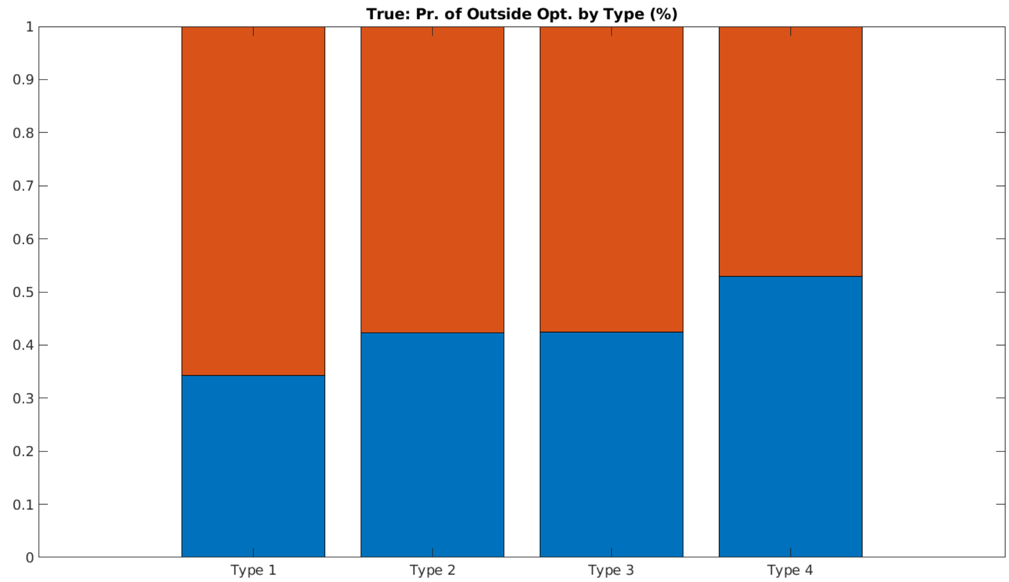



**Figure A-7:** Outside option shares by score quartile with simulated data

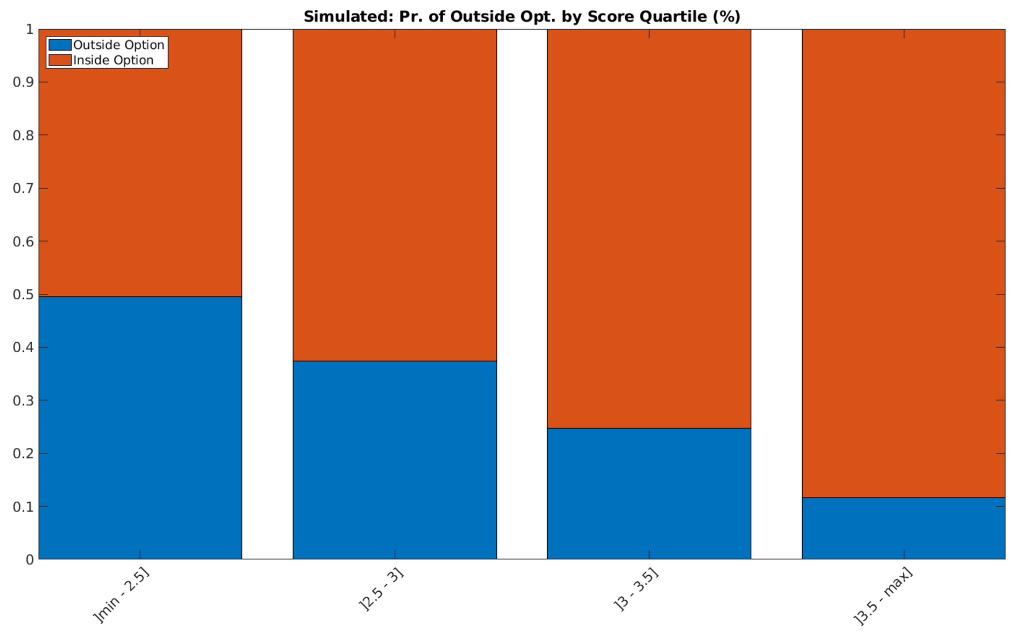


**Figure A-8:** Outside option shares by score quartile with true data

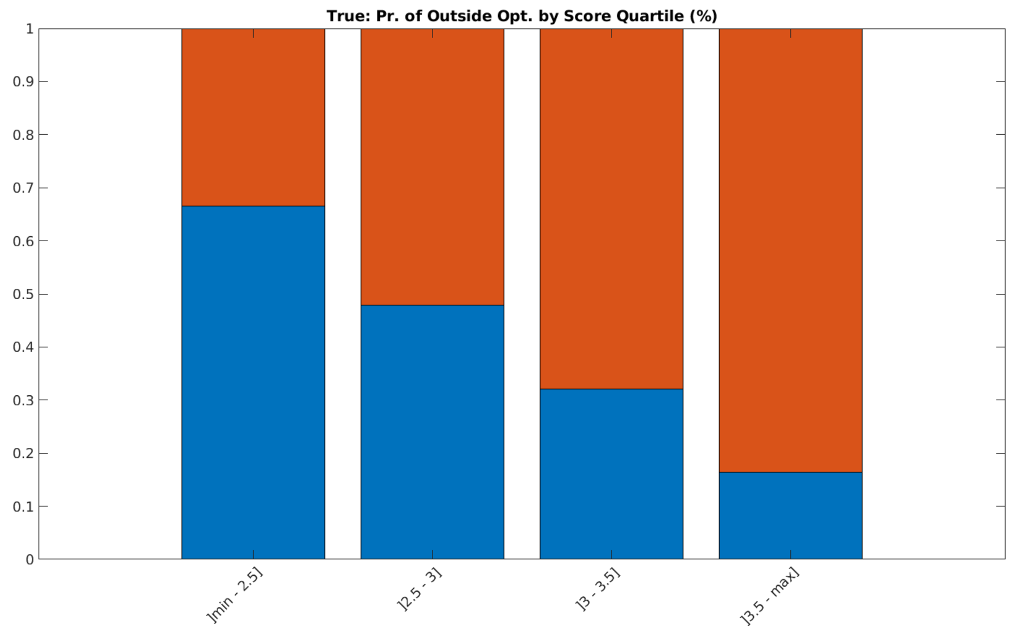



**Figure A-9:** Correlations between true and simulated FL shares