Essays on Empirical Matching Systems

by

Mohit Karnani

B.Sc., Universidad de Chile (2015)

M.Sc., Universidad de Chile (2017)

Submitted to the Department of Economics and the Statistics and Data Science Center
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ECONOMICS AND STATISTICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2024

Authored by: Mohit Karnani
Department of Economics and the Statistics and Data Science Center
December 15, 2023

Certified by: Esther Duflo
Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics, Thesis Supervisor

Certified by: Parag Pathak
Class of 1922 Professor of Economics, Thesis Supervisor

Accepted by: Isaiah Andrews
Professor of Economics
Chairman, Departmental Committee on Graduate Studies

Essays on Empirical Matching Systems

by

Mohit Karnani

Submitted to the Department of Economics and the Statistics and Data Science Center on December 15, 2023 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY IN ECONOMICS AND STATISTICS

ABSTRACT

This dissertation is a collection of three papers on empirical methods in matching systems. In Chapter 1, I study the estimation of treatment effects in the context of randomized controlled trials conducted with participants in matching systems. I show how conventional methods fail to account for the interference across outcomes induced by matching systems, and therefore yield invalid estimates of causal parameters. I propose a method that solves the interference problem and apply it in two empirical settings. Chapter 2 studies the relevance of the configuration of on- and off-platform options when centralized matching systems operate alongside a decentralized matching process. In these situations, the existence of off-platform options in a decentralized system can affect the outcomes of participants in the centralized system who seek to be matched to on-platform options. We show this by developing and estimating a structural model that considers the interplay between on- and off-platform options in a matching system. Chapter 3 studies the causal effects of different screening and recruiting policies affecting applicants in the Chilean centralized college match. We show how machine learning methods can enhance these screening and recruiting policies.

JEL Classification Codes: D47, C90, I21

Thesis supervisor: Esther Duflo

Title: Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics

Thesis supervisor: Parag Pathak

Title: Class of 1922 Professor of Economics

Acknowledgments

I am profoundly grateful to my advisors -Esther Duflo and Parag Pathak- for their guidance, support and patience over all these years. Their thoughtful ideas and feedback have shaped this dissertation, and have strongly influenced me as a researcher. They are fantastic role models, full of wisdom and kindness, and I am very lucky to have been mentored by them.

I am also deeply indebted to Nikhil Agarwal, Isaiah Andrews and Joshua Angrist for their insightful conversations and discussions. These essays have substantially benefited from the time they generously dedicated to me. I also want to thank Alberto Abadie, David Atkin, Abhijit Banerjee, Dave Donaldson, Anna Mikusheva, Benjamin Olken, Robert Townsend and José Zubizarreta for their comments and ideas; I have learned a lot from them.

This journey was made significantly smoother thanks to the help of MIT's staff, especially Gary King, Julia Martyn-Shah, Shannon May, Beth Milnes and Kim Strampel. I am also grateful to Jerry Hausman, John Castle, and George and Obie Shultz for financial support.

My coauthors are a fantastic group of researchers: Franco Calle, Sebastián Gallegos, Adam Kapor and Christopher Neilson. I owe special gratitude to Chris, who has generously mentored me. He taught me how to work hard while enjoying every milestone along the way.

These years in Cambridge, MA have been full of joy thanks to my friends, including Angie Acquatella, Ari Bronsoler, Salomé Aguilar, María Alejandra Castellanos, Marc de la Barrera, Isabel Di Tella, María Fernanda Escobar, Ardavan Farahvash, Bernardo García, Brindha Kanniah, Carlos Molina, Daniela Paz, Belén Saldías, Clara Sievert, Diana Sverdlin and Jaume Vives. When it comes to bringing me joy, Pooja Harjani deserves a special place on the list, just as she has a special place in my heart.

Finally, this dissertation is dedicated to my family: Anjali, Bhagwan, Harsh and Hriday. Each of them helped shape the person I am today; and for that, I am forever grateful.

Contents

List of Figures

List of Tables

Chapter 1

RCTs with Interference in Matching Systems

Abstract

This paper studies the estimation of treatment effects using data from randomized controlled trials (RCTs) in the context of matching systems, such as centralized school choice systems. If treated units change their inputs (e.g. reported preferences), they can affect the allocation outcomes determined by the matching system in equilibrium. When the outcome of interest is determined in equilibrium (e.g. being matched to a target school), conventional estimators that compare a treatment group against a control group fail to account for violations of the stable unit treatment value assumption (SUTVA) induced by the matching system. Therefore, they yield biased and inconsistent estimates of the average treatment effect on the treated (ATT), even under perfect randomization designs. I propose an estimator that decomposes the observed average difference between the outcomes of the treatment and control units, and consistently estimates the ATT, along with the residual spillover effect. This nonparametric estimator relies on the matching mechanism satisfying a version of the equal treatment of equals (ETE) property. Moreover, the ATT is identified for any counterfactual fraction of treated units, which allows for the estimation of treatment effects under scaled interventions. I illustrate the applicability of this estimator using data on the Chilean centralized school choice system and the Ecuadorian teacher labor market.

1.1 Introduction

About a third of the countries around the world use centralized matching mechanisms to assign students to educational institutions or teachers to their jobs (Neilson, 2021). In these school choice systems, applicants submit rank-ordered lists to a platform that matches them to schools following some predetermined set of rules (Abdulkadiroğlu and Sönmez, 2003). Researchers often conduct large-scale randomized controlled trials (RCTs) where individuals participating in these matching systems are exposed to treatments affecting their application behavior. Examples include informing applicants about nearby high-value-added schools (Ainsworth et al., 2020) or highlighting the non-assignment risk implied by their application (Arteaga et al., 2022).

These RCTs allow researchers to estimate treatment effects on partial equilibrium outcomes using standard methods that compare treated units against control units (Duflo et al., 2007), such as comparing the share of high-value-added schools listed by treatment and control units. However, when the outcome of interest is determined in equilibrium, such as the share of applicants *assigned* to a high-value-added school, conventional estimators fail to account for violations of the stable unit treatment value assumption (SUTVA) induced by the matching system.

This paper studies the estimation of treatment effects when RCTs are conducted with individuals participating in matching systems. We show how standard techniques that compare the equilibrium outcomes of treatment and control units can yield biased and inconsistent estimates of the average treatment effect on the treated (ATT), even under perfect randomization designs. Intuitively, an intervention that induces the treatment group to apply to high-value-added schools not only affects the probability of the treated students of being admitted into these schools, but might also have a *spillover effect* on the outcomes of the control group if schools have fixed capacities and the matching system has to resort to some form of tie-breaking rule.

If this spillover effect changes the outcomes of the control units, then they no longer resemble the counterfactual situation where no treatment took place. Therefore, any direct comparison between the treatment and control outcomes, such as comparing the fraction matched to a high-value-added school in each group, yields an invalid estimate of the causal effect of the treatment on the outcome of interest. This problem has been studied in other contexts, such as evaluating the displacement effects of labor market policies (Crépon et al., 2013).[1]

Indeed, in the presence of interference, the potential outcomes (Neyman, 1990; Rubin, 1974) of a unit are not only a function of the treatment status of that unit, but can also be a function of the treatment status of all other units. For example, in a binary treatment case, instead of having only two potential outcomes per unit, as in the case without interference, the set of potential outcomes with interference scales to 2^{units} per unit. Therefore, any procedure that aims to estimate treatment effects with interference is required to reduce the dimensionality of the set of potential outcomes to make the problem tractable. Our approach achieves this by exploiting properties of matching systems.

Matching systems are interesting to study the problem of interference for at least two reasons. First, matching systems are widely used in practice to allocate scarce resources in high-stakes situations: from school choice systems, to medical residency matches, to cadet branching (Roth and Sotomayor, 1992). Because of this, policymakers and researchers are increasingly more likely to experimentally evaluate interventions in the context of matching systems to inform policy decisions, making the problem of interference more prevalent. Secondly, as the assignment mechanism is determined by an explicit set of rules described by a matching algorithm, researchers can leverage these rules to characterize the interference, without having to rely on design-based solutions. In a sense, matching systems provide information about the data-generating process of the outcomes, and allow researchers to "correct" their estimates using data from an RCT.

We propose an estimator that decomposes the observed average difference between the outcomes of the treatment and control units, and consistently estimates the ATT, along with the residual spillover effect derived from the matching interference. This nonparamet-

[1] Previously proposed solutions to this problem are usually "design based", such as designing a two-stage cluster randomized controlled trial, where clusters are randomly assigned to different treatment intensities (i.e. share of treated units), and then units within each cluster are randomized into their treatment status such that they match the treatment intensity of the cluster. This paper proposes a methodological solution that does not rely on special randomization designs, but is specific to RCTs conducted in matching systems. There are other examples where statistical corrections to interference can be achieved without relying on specific randomization designs, such as interference in networks (Leung, 2020).

ric estimator relies on the matching mechanism satisfying a version of the *equal treatment of equals* property. Moreover, the ATT is identified for any counterfactual fraction of treated units, which allows for the estimation of treatment effects under scaled interventions, targeting a rich set of causal estimands. We also propose an extension to this estimator that accommodates multiple treatment arms and clustered trials.

Finally, we provide two empirical applications for this new estimator. The first application is a simulated experiment in the Chilean centralized school choice system and the second one is a real experiment conducted in the Ecuadorian teacher labor market match. In both cases the implied spillover effects are large, in the latter even to the extent that the estimated ATT cannot be statistically distinguished from zero, i.e. we do not reject that the observed average difference between treatment and control teachers is purely driven by interference in the matching system.

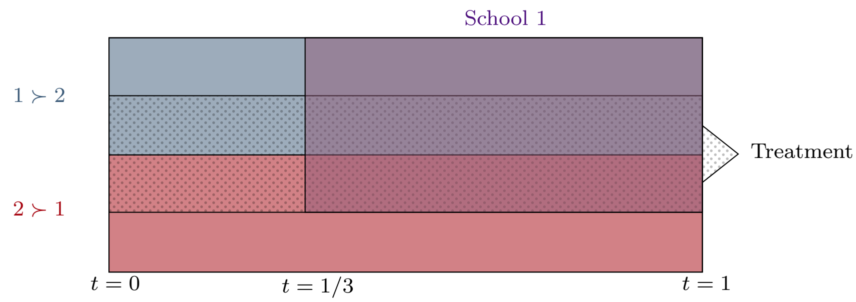
The rest of the paper is organized as follows. In section 1.2 we illustrate the nature of the interference induced by the matching system and formally introduce the general model we study in the paper. Then, section 1.3 shows how conventional estimators are invalid in the presence of matching interference and proposes an estimator that solves this problem. We also prove relevant properties of this estimator in section 1.3, and then propose extensions to it in section 1.4. Finally, section 1.5 shows the results of applying our new estimator in two empirical settings, and section 1.6 concludes.

1.2 The Model

1.2.1 An Illustrative Example

Consider the following simplified model inspired by Azevedo and Leshno (2016). There are two schools, 1 and 2, and a continuum of applicants seeking admission into these schools. Applicants are indexed by a uniform tie-breaker $t \in [0, 1]$. Each school can take up to 50% of the applicants in the market. Suppose that half of the mass of applicants prefer school 1 over school 2 ($1 \succ 2$) and the other half prefers school 2 over school 1 ($2 \succ 1$). Schools' "preferences" give priority to applicants with higher tie-breakers. Let there be a

Figure 1.1: Illustrative Example of Interference



Note: This figure illustrates the simplified model described by Azevedo and Leshno (2016). The original preferences are colored blue for applicants who prefer school 1 over 2 and red for applicants who prefer school 2 over 1. The shaded area in the middle denotes the treatment population and the overlaid purple rectangle indicates the mass of applicants that are admitted into school 1 once the treatment takes place.

matching mechanism that assigns applicants to schools by running the student-proposing deferred acceptance (DA) algorithm (Gale and Shapley, 1962). If applicants report their true preferences to the matching mechanism, everyone is assigned to their most preferred option, implying a tie-breaking cutoff of $\bar{t}_0 = 0$, as there is no excess demand for seats at any school in the very first step of the DA algorithm. Let this be the counterfactual scenario in absence of any treatment.

Suppose we do not observe this counterfactual, and conduct an RCT where half of the applicants are randomly selected to be treated with a behavior-changing intervention. Suppose that the intervention effectively modifies the preferences of the treatment group such that they all prefer school 1 over school 2 ($1 \succ 2$), no matter what their counterfactual preferences are in absence of the treatment. In this situation, we have a mass of $3/4$ of the applicants ranking school 1 over school 2; $1/2$ coming from the treatment group and $1/4$ coming from the control group.

Because of the excess demand for seats in school 1, there is a binding tie-breaker induced by the DA matching mechanism, $\bar{t}_1 = 1/3$. This implies that not everyone ranking $1 \succ 2$ is matched to their top listed option. If the outcome of interest is an indicator for being matched to school 1, we observe $2/3$ of the treatment group being matched to school 1 and $1/3$ of the control group being matched to school 1. This is depicted in Figure 1.1.

Given the observed fractions matched to school 1 in each group, one can (incorrectly) conclude that the average treatment effect of the intervention on the treatment group is an increase of $2/3 - 1/3 = 1/3$ in the probability of admission into school 1. This is incorrect because we know that, in the counterfactual situation described above, half of the treatment group would have been matched to school 1. Therefore, the true average treatment effect on the treated is $2/3 - 1/2 = 1/6$, i.e. only 50% of the observed difference between the treatment and control group outcomes.

This difference arises because of the binding tie-breaker cutoff. Originally, in absence of any intervention, everyone is assigned to their most preferred school, so the unconditional probability of being matched to school 1 is 1/2. However, the treatment not only affects the treatment group by changing their preferences, but it also interferes with the outcomes of the control group by shifting the cutoff score from $\bar{t}_0 = 0$ to $\bar{t}_1 = 1/3$. This constitutes a violation of the stable unit treatment value assumption (SUTVA), due to interference in the matching system (Cox, 1958).

We now extend the underlying idea in this simple model to a more general framework that allows us to study treatment effects with interference in matching systems. The rest of the discussion is still framed as a school choice problem, but the results hold in other more general matching systems.

1.2.2 The Market: Notation and Definitions

Consider a finite set of applicants I and a finite set of schools S with *relative* capacities $q_s \in (0, 1]$, $\forall s \in S$, collected in vector $q \in (0, 1)^{|S|}$. These relative capacities translate into a vector of nq openings, where $n := |I|$. Among these schools, there is an *outside option*, denoted by $\emptyset$, with relative capacity $q_\emptyset = 1$. This null school represents being unmatched.

Applicants have unobserved *types* $\theta^\star$ that lie on a set of types Θ. For example, if individual i's type can be described by preferences for schools $\succ_i$ and test scores t_i, then $\theta_i^\star := (\succ_i, t_i)$.[2]

[2]Types can be more general and incorporate other objects. For example, if applicants also had a priority vector $\rho_i := (\rho_{is})_{s \in S}$ describing the priorities they have at each school, then types would be $\theta_i^\star := (\succ_i, t_i, \rho_i)$. Our general definition of types differs from the usual formulation in the literature studying this class of models (Abdulkadiroğlu et al., 2017). We do this so types contain all relevant variables that can change in an experiment (e.g. a tutoring intervention can change the test scores of treated applicants, even though test scores are usually not included in types). Moreover, in some cases test scores are coarsely measured,

Schools do not have types.

Applicants report their types, observed as the profile $\theta := \times_{i \in I}\theta_i$, to a direct stochastic mechanism $\varphi : \theta \rightarrow [0,1]^{|I| \times |S|}$, which induces a lottery on matchings $\mu(i) \in S,\ \forall i \in I$.[3] The reported types θ need not be the same as the true types $\theta^\star$, i.e. we do not take a stance on the strategy-proofness of φ.[4] We also allow agents to misreport their types due to other behavioral factors, such as heterogeneous levels of sophistication (Pathak and Sönmez, 2008) or subjective beliefs (Kapor et al., 2020).

The entries of matrix $\varphi(\theta)$ are the probabilities of assignment of each applicant i to each school s under profile θ, i.e. $\varphi(\theta)[i,s] := \mathbb{P}[\mu(i) = s|\theta]$.[5] For any profile of reported types θ, each row $\varphi(\theta)[i,:]$ sums to 1, and each column $\varphi(\theta)[:,s]$ adds up to at most nq_s.[6]

We define a *market* as the collection $\mathcal{M} := \{I, S, \theta, q, \varphi\}$. We also define a *large market* as $\lim_{n\rightarrow\infty}\mathcal{M}$. This means that in the large market the options S and relative capacities q_s stay fixed, but the number of applicants of each type grows, i.e. $|I| \rightarrow \infty$ (and we have an infinite mass of each type θ) while $|S|$ is still finite.[7] The vector of total capacities nq also grows.

1.2.3 Experimental Design

In a finite market, individuals I participate in an RCT and are randomly assigned to either a *treatment group* $T \subset I$ or a *control group* $C := I \setminus T$. For example, the treatment group can be a random subset of I that receives information about high-value-added schools in their area, while the control group does not receive any such information. Let the fraction of treated individuals be a constant $r := |T|/|I|$, such that the approximate distribution of

---

and many applicants share the exact same test score. In this situation, test scores resemble priorities and our formulation of type is equivalent to that of the literature.

[3]In situations where types include a continuous variable, such as a continuous test score that determines admissions, the lottery is degenerate and the mechanism is deterministic.

[4]We impose structure on reports, but this is unrelated to the properties of φ.

[5]To be precise about the notion of probability here, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triplet, where $\Omega := S^{|I|}$ is the sample space with possible match outcomes, $\mathcal{F} := 2^\Omega$ is a $\sigma-$algebra containing all valid collections of match events, and $\mathbb{P} : \mathcal{F} \rightarrow [0,1]$ is a probability measure induced by the random mechanism φ.

[6]For deterministic mechanisms, the entries of $\varphi(\theta)$ are either 0 or 1, where each row contains at most one 1-entry and each column contains at most nq_s 1-entries.

[7]This notion of large market is later used in section 1.3 to study the asymptotic properties of the estimators proposed in this paper.

treatment indicators is $d_i \overset{iid}{\sim}$ Bernoulli(r).

Define $\theta_i^d \in \Theta$ to be the **potential report** of applicant i under treatment status $d \in \{0, 1\}$, where $d = 1$ denotes being treated and $d = 0$ being untreated. This notation conveys that each applicant can be in one of two different states. In our running example, these states are with information ($d_i = 1$) or without information ($d_i = 0$). Agents have a corresponding reported type θ_i^d under each state, even though only one is observed (agents cannot be informed and uninformed simultaneously). This definition emulates the idea of potential outcomes in the Neyman–Rubin causal model (Neyman, 1990; Rubin, 1974). Moreover, by utilizing this notation without dependence on the full vector of treatments, we explicitly assume a "consistency rule" on the reported types, which implies that there are no direct spillovers of the treatment on the reported types.

Assumption 1 (No type interference)**.** *Types reported follow*

$$\theta_i = d_i\theta_i^1 + (1 - d_i)\theta_i^0, \ \forall i \in I.$$

This rules out other forms of interference where the treatment status of agent i affects the reported type of another agent j. For example, if i and j are siblings, then θ_j could in principle be a function of d_i (Altmejd et al., 2021; Basse et al., 2019), but our model does not accommodate this spillover. This paper only focuses on spillovers induced by interference in the matching mechanism, and not on other kinds of spillovers that might affect the reported types (e.g. network interference or spatial spillovers). Additionally, Assumption 1 implies perfect compliance with the treatment: informed individuals report their types as if they were informed, and uninformed individuals report their types as if they were not informed. In other words, for each agent i, there are no other reports but the two potential reports θ_i^1 and θ_i^0. Formally, Assumption 1 is a stable unit treatment value assumption regarding the reported types.

Let X_i denote a vector with other characteristics of applicant i that are excluded from θ_i, such as demographics. Suppose applicants' potential reports and characteristics are drawn from the same population distribution F: $\{\theta_i^1, \theta_i^0, X_i\} \sim F$. For notation purposes, subscripts on F denote the random variable drawn from the distribution, e.g. F_{θ^0} is the population distribution of potential types in absence of information.

1.2.4 Target Estimand

Let the outcome of interest be an indicator variable for being matched to a school in some fixed subset $H \subset S$, i.e. $Y_i := \mathbf{1}[\mu(i) \in H|\theta]$. For example, H could be a set of high value added schools targeted by a policymaker, such that Y_i indicates if applicant i was matched to a high value added school or not. Even though this is our preferred outcome variable, the following results directly extend to more general outcomes Y_i that are a linear combination of the match indicators: $Y_i = \sum_{s \in S} w_s \mathbf{1}[\mu(i) = s|\theta]$. For example, w_s could be a measure of value added for school s, and then Y_i would denote the value added of the school assigned to applicant i. In the case we study above, $w_s = \mathbf{1}[s \in H]\ \forall s \in S$.[8]

The target estimand in this model is the **average treatment effect on the treated** (ATT), i.e. the change in the probability of admission into H for the treatment group when the reported types of the treatment group change from the counterfactual profile $\theta^0_T := \times_{i \in T} \theta^0_i$ to the observed profile $\theta^1_T := \times_{i \in T} \theta^1_i$, holding the reported profile of the control group $\theta^0_C := \times_{i \in C} \theta^0_i$ fixed. For simplicity, we collect the two untreated profiles in one term $\theta^0_I := (\theta^0_T, \theta^0_C)$.

Definition 1 (ATT)**.** *The average treatment effect on the treated τ is defined as*

$$\begin{aligned}\tau :=& \mathbb{P}[\mu(i) \in H|i \in T, \theta^1_T, \theta^0_C] - \mathbb{P}[\mu(i) \in H|i \in T, \theta^0_I] \\ =& \mathbb{E}[Y_i|i \in T, \theta^1_T, \theta^0_C] - \mathbb{E}[Y_i|i \in T, \theta^0_I]. \end{aligned} \tag{1.1}$$

The first expectation in (1.1) is conditioned not only on the treatment condition of the treatment group, but also explicitly incorporates the untreated condition of the control group. The second expectation indicates that no one is treated. This extensive notation becomes necessary because of potential SUTVA violations (Rubin, 1980). Indeed, the outcome Y_i can

[8]An essential requirement is that the outcome of interest and target parameters depend only on the matches, rather than other downstream outcomes which could be influenced by the matches, such as test scores. Such outcomes and target parameters are typically studied in short-term experimental evaluations (e.g. how does an information campaign affect the match-outcomes of applicants shortly after), and are the main focus of this paper. We do not consider evaluations that go beyond match-outcomes (e.g. how does an information campaign affect test scores after applicants are matched).

depend on the treatment status of other units $j \neq i$ and their reported types. Importantly, τ also depends on the fixed treatment fraction r, so this target estimand is specifically defined for a given experimental design. We focus on other causal estimands where r may vary in section 1.4.[9]

In a similar fashion, the spillover effect that the treatment group imposes on the control group is the change in the expected outcome of the control group when the reported type profile changes from θ_I^0 to (θ_T^1, θ_C^0), i.e. when the treatment group reports θ_T^1 instead of θ_T^0. We formally define the spillover population parameter in this setting as follows.

Definition 2 (Spillover)**.** *The spillover effect σ on the control group is defined as*

$$\sigma := \mathbb{E}[Y_i | i \in C, \theta_T^1, \theta_C^0] - \mathbb{E}[Y_i | i \in C, \theta_I^0]$$

These two estimands are of separate interest because they are indicative of the impact the intervention has on the treatment group relative to the counterfactual scenario where no treatment takes place, and the displacement it causes in the control group. Even in situations where researchers care about the total effect that an intervention has on the treated relative to the untreated (e.g. the impact of unionization on relative wages (Lewis, 1983) is a classical example), distinguishing between τ and σ can be of importance for policy reasons, such as scaling up an intervention[10] or quantifying and correcting the "harm" that a policy imposes on the untreated units. Given these two estimands of interest, τ and σ, we now study the estimation and identification of these population objects using data coming from an RCT.

1.3 Estimation

1.3.1 Difference in Means

The identification results below are based in the large market approximation using data from an RCT. With this in mind, we first analyze the standard comparison between treatment

[9]In addition, τ depends on the properties of φ and relative capacities q, i.e. on the full structure of the market $\mathcal{M}$, but we assume φ and q to be fixed.

[10]We further develop this idea in the extension provided in subsection 1.4.1.

and control outcomes in absence of interference.

In absence of any interference, when $\sigma = 0$, we can directly estimate τ by computing the sample means of Y_i for individuals in T and C, and subtracting them, as in the standard analysis of RCT outcomes:

$$\widehat{\Delta} := \sum_{i \in T} Y_i/|T| - \sum_{i \in C} Y_i/|C|.$$

Lemma 1. *Without interference, $\widehat{\Delta}$ is an unbiased, consistent estimator of τ.*

Proof. The independent Bernoulli treatment assignments d_i are linearly mapped to reported types θ_i (Assumption 1) drawn from a common distribution. As there is no interference in φ, the reported types are independently mapped to a discrete support and linearly transformed to Y_i, which has finite expectation. Thus, the strong law of large numbers holds.

By the strong law of large numbers, in the large market (as $n \to \infty$) the sample averages converge almost surely to their population expectations, so we have that

$$\begin{aligned} \widehat{\Delta} \to & \mathbb{E}[Y_i | i \in T, \theta_T^1, \theta_C^0] - \mathbb{E}[Y_i | i \in C, \theta_T^1, \theta_C^0] \underbrace{+\sigma}_{=0} \\ = & \mathbb{E}[Y_i | i \in T, \theta_T^1, \theta_C^0] - \underbrace{\mathbb{E}[Y_i | i \in C, \theta_I^0]}_{=\mathbb{E}[Y_i | i \in T, \theta_I^0]} \\ = & \tau, \end{aligned}$$

where we add $\sigma = 0$, implied by no-interference, and $\mathbb{E}[Y_i | i \in C, \theta_I^0] = \mathbb{E}[Y_i | i \in T, \theta_I^0]$ because d_i is randomly assigned. □

Thus, in absence of interference, as the assignment into the treatment is random, and $\sigma = 0$ by assumption, the sample averages converge to the population expectations in (1.1) and $\widehat{\Delta}$ is a valid estimator for τ.

In empirical practice, the no-interference assumption is likely to hold whenever there is "slack capacity" in S i.e. when the vector of capacities q is such that there are no binding capacity constraints at any school. For example, in a school choice system with heterogeneous preferences and large schools relative to n (i.e. q_s close enough to 1), agents are unlikely to interfere with each other.

Another situation where no-interference is likely to hold is in *small experiments*, i.e. experiments where r is close to 0. In these experiments, the change from profile θ_I^0 to (θ_T^1, θ_C^0) might not significantly change the equilibrium assignments of agents in the control group. Small experiments need not be exclusive to RCTs, but can also be found in *natural experiments*, such as those induced by regression discontinuity designs (Angrist et al., 2022).[11]

Now consider the case where $\sigma \neq 0$, i.e. where there is interference induced by the matching mechanism. $\widehat{\Delta}$ still estimates the second expectation in (1.1) with the sample average of Y_i in the control group. However, unlike in Lemma 1, this sample average is not a valid approximation of the outcomes of the treatment units in absence of the treatment. This implies the following lemma.

Lemma 2. *If $r \neq 0$, and $\sigma \neq 0$, $\widehat{\Delta}$ is a biased, inconsistent estimator of τ.*

Proof. If $r \neq 0$, then the conditioning sets defined by type profiles (θ_T^1, θ_C^0) and (θ_I^0) are different, because the mass of treatment individuals is strictly positive as $n \to \infty$ in the large market. Then, even if the law of large numbers holds,

$$\begin{aligned}
\widehat{\Delta} \to & \mathbb{E}[Y_i | i \in T, \theta_T^1, \theta_C^0] - \mathbb{E}[Y_i | i \in C, \theta_T^1, \theta_C^0] \pm \mathbb{E}[Y_i | i \in T, \theta_I^0] \\
= & \tau + \underbrace{\mathbb{E}[Y_i | i \in T, \theta_I^0] - \mathbb{E}[Y_i | i \in C, \theta_T^1, \theta_C^0]}_{=-\sigma} \\
= & \tau - \sigma \\
\neq & \tau,
\end{aligned}$$

where again $\mathbb{E}[Y_i | i \in C, \theta_I^0] = \mathbb{E}[Y_i | i \in T, \theta_I^0]$ because d_i is randomly assigned. $\square$

As mentioned above, the treatment group affects the assignment probabilities of the control group to schools in H. The treatment shifts the type profile θ_T of the treatment group from θ_T^0 to θ_T^1, which affects the conditioning set of the target population expectation and thus

[11] Theoretically, matching systems can significantly change their equilibrium allocation even upon small changes in the reported type profile. However, in practice, large markets are unlikely to exhibit this behavior. For example, in a large DA market, tiebreaker cutoffs are fixed and small changes in reported types do not change them, so the untreated group remains unaffected.

also the resulting expected outcome for the control group. For example, in subsection 1.2.1 the nature of this interference can be visualized as a change in the equilibrium cutoffs, which not only affects the admission chances of the treatment group, but also those of the control group. This renders the observed sample average of the control group, $\sum_{i \in C} Y_i/|C|$, an invalid estimator of $\mathbb{E}[Y_i|i \in T, \theta_I^0]$.[12]

Moreover, if the treatment induces demand for H such that $\tau \geq 0$, then it is likely that the interference negatively impacts the control group, i.e. $\sigma \leq 0$. This implies that $\widehat{\Delta}$ most likely overestimates τ in empirical practice. In section 1.5, we use data from Ajzenman et al. (2023) to show how this can be the case.

This negative result has a series of implications. First, even when conducting an RCT, the difference-in-means estimator $\widehat{\Delta}$ can be invalid if certain conditions are met: a large experiment in a matching system with interference. Second, it is likely that the bias arising from this interference in empirical applications is *positive*. This means that results from RCTs in matching systems, testing the effectiveness of demand-inducing interventions on applicants' likelihood of being matched to target schools, tend to be *overestimated*.

Without additional assumptions on φ, Lemma 2 poses a practical challenge to experimental studies that change inputs in matching systems to estimate treatment effects on equilibrium outcomes. To overcome this negative result, and correctly identify τ, we impose some structure on the matching mechanism φ.

1.3.2 Structure of φ

For the main results of this paper, we crucially rely on one assumption about φ. We assume that φ satisfies a *modified* version of the *equal treatment of equals* (ETE) property. We first define the standard ETE property.

Definition 3 (Equal treatment of equals). *φ satisfies equal treatment of equals if for any $i, j \in I$ such that $\theta_i = \theta_j$, the assignment probabilities induced by φ are the same for i and*

[12]Note that if the relative mass of the treatment group is negligible in the large market, i.e. if $r = 0$, then $\sigma = 0$ because the conditioning set (θ_T^1, θ_C^0) is the same as θ_I^0. Thus, if $r = 0$, $\widehat{\Delta}$ is a consistent estimator of τ.

j, i.e. given any fixed type profile θ,

$$\theta_i = \theta_j \implies \mathbb{P}[\mu(i) = s|\theta] = \mathbb{P}[\mu(j) = s|\theta], \forall s \in S.$$

Intuitively, ETE induces a "type-stratified RCT" within the matching mechanism (Abdulkadiroğlu et al., 2017), i.e. for any two individuals i, j sharing the same reported type $\theta_i = \theta_j$, their allocation outcomes are drawn from the same distributions $\varphi(\theta)[i, :] = \varphi(\theta)[j, :]$ and any difference in their resulting assignment is as good as random.[13] In other words, ETE implies that assignments are independent of other applicant characteristics X_i after conditioning on θ_i:

$$\{\mu(i) = s\} \perp X_i|\theta_i \quad \forall s \in S.$$

This independence property holds within a matching system satisfying ETE for any given type profile θ and is unrelated to the RCT-independence in the setting we study. The RCT instead shifts the type profile for the treatment group, where d_i is independent of characteristics X_i and potential types: $\{\theta_i^1, \theta_i^0, X_i\} \perp d_i$.

The importance of ETE is straightforward: it allows researchers to directly map reported types to assignment probabilities, without having to control for additional variables that might be correlated with the assignment outcomes.[14] This ignorability condition has proven to be useful in many contexts, such as estimating value added measures for schools (Abdulkadiroğlu et al., 2017) or the effects of minority-integration in school choice systems (Angrist et al., 2022).

Our identification results also rely on mapping reported types to assignment probabilities. However, instead of the standard ETE assumption, we require a *distributional continuity* assumption on φ to hold in the large market. Specifically, we require that when type *distributions* converge, from where reported types are drawn, then their $\varphi-$induced distributions of assignment probabilities also converge.[15]

[13] As mentioned above, if types contain a continuous variable (e.g. a test score that determines allocations), then the mechanism becomes deterministic and there is no source of randomness.

[14] This is usually done by simulation. As individuals sharing a type also share the same probability distribution of assignments, by simulating the match multiple times, researchers can infer the assignment distribution for each type from the observed empirical measure of simulated assignments. This insight motivates our main estimator below.

[15] Recall that each type profile θ is mapped to a bistochastic matrix $\varphi(\theta)$ where the entries correspond

We call this adapted ETE assumption **continuous distributional equal treatment of equals** (CD-ETE).

Assumption 2 (Continuous distributional ETE)**.** *When $n \to \infty$, the mechanism φ satisfies continuous distributional equal treatment of equals: for any converging sequence of type-distributions F_θ^n approaching a type-distribution F_θ, the corresponding sequence of distributions of match probabilities F_φ^n (induced by φ mapping F_θ^n) also converge to the distribution of match probabilities F_φ (induced by φ mapping F_θ) i.e.*

$$\lim_{n\to\infty} \sup |F_\theta^n - F_\theta| = 0 \implies \lim_{n\to\infty} \sup |F_\varphi^n - F_\varphi| = 0.$$

The CD-ETE assumption modifies the standard ETE condition. Under ETE, for any given profile θ, individual reports θ_i can be directly mapped to a distribution of allocation outcomes $\varphi(\theta)[i,:]$. Under CD-ETE, this mapping is extended to sequences of profile distributions, such that if profile distributions F_θ^n converge, then assignment-probability distributions F_φ^n also converge. This CD-ETE assumption can also be interpreted as an assumption of distributional uniform convergence under the sup norm.

This assumption allows us to identify our target estimand τ, without relying on other more stringent requirements like a (random serial priority) cutoff structure (Agarwal and Somaini, 2018) or a fully parametric specification of the model (Kapor et al., 2024).[16] In addition, CD-ETE is satisfied by common matching systems in practice, such as those that implement the DA algorithm. Indeed, in a large DA market, cutoffs are fixed and converging sequences of type distributions imply "smoothly" converging cutoff distributions, which then directly map to matching probabilities based on reported types and the equilibrium cutoffs (Azevedo and Leshno, 2016; Abdulkadiroğlu et al., 2017). This applies not only to the DA algorithm, but also to other mechanisms allowing cutoff representations with smoothly converging cutoffs, such as the random serial dictatorship mechanism (Che and Kojima, 2010).

to the assignment probabilities of each student to each school, modulo their non-assignment risk. Thus, distributions of random types F_θ are mapped to distributions of random matrices F_φ, where the dependence on θ is suppressed to simplify notation.

[16]We do not assume any structure on how d_i affects θ_i, apart from the standard "consistency rule" (Assumption 1), nor on how reported types θ, capacities q or treatment intensity r affect the match.

However, CD-ETE rules out other mechanisms that might exhibit "discontinuous" changes in allocations upon "small" changes in the distribution of types. For example, consider a mechanism that grants priority in a school s to a set of applicants $A \subset I$ if at least half of the population I lists that school on top, and priority to another set of applicants $B \neq A$ if fewer than half of I list s as their first choice. Mechanisms like this can satisfy ETE without satisfying CD-ETE: for converging sequences of type distributions with a limit F_θ in which exactly half of the applicants list s as their top choice, there are two limiting distributions of assignment probabilities. As the limit does not exist, this hypothetical mechanism does not satisfy CD-ETE.[17]

Provided that the mechanism φ satisfies CD-ETE, there exists an estimator that solves the identification problem introduced in Lemma 2. This alternative estimator is described below.

1.3.3 Main Results

As noted above, Lemma 2 follows from $\sum_{i \in C} Y_i/|C|$ not being a valid estimator for $\mathbb{E}[Y_i|\theta_I^0]$ in the presence of interference. We first propose a procedure to consistently estimate the population expectation $\mathbb{E}[Y_i|\theta_I^0]$ using RCT data coming from mechanisms that satisfy CD-ETE. Consider the estimator described with pseudo-code in algorithm 1.

[17] A mechanism satisfying CD-ETE need not satisfy ETE. For example, consider a (dictatorship) mechanism φ in which the individual with index $i = 1$ gets assigned to their top reported choice, while the rest $j \neq 1$ are randomly assigned to their schools, irrespective of their types. This mechanism satisfies CD-ETE because the relative mass of individual $i = 1$ is $1/n$, which decays to 0 as $n \to \infty$, yielding a constant random assignment mechanism in the large market for any converging sequence F_θ^n and limiting distribution F_θ. Nonetheless, this mechanism does not satisfy ETE, as any other individual $j \neq 1$ with type $\theta_j = \theta_1$ will have assignment probabilities $\varphi[j, :] \neq \varphi[1, :]$ (all other agents have equal, interior assignment probabilities, but agent 1 has a degenerate distribution that guarantees placement in their top choice).

Algorithm 1: Consistent estimator of $\mathbb{E}[Y_i|\theta_I^0]$

Data: $\theta_C, n, H,$ (large) B

Function: $\varphi : (\theta, seed) \mapsto \mu \in S^n$ satisfying CD-ETE

Init: $\widehat{\theta}_I^0 \in \Theta^n, Y^0 \in \{0,1\}^{n \times B}$

Result: $\bar{Y}^0$ ▷ (consistent estimator of $\mathbb{E}[Y_i|\theta_I^0]$)

for $b \in \{1, \dots, B\}$ **do**

$\widehat{\theta}_I^0 \leftarrow \texttt{SampleWithReplacement}(\theta_C, n, seed_b)$ ▷ (simulated types)

$\hat{\mu} \leftarrow \varphi(\widehat{\theta}_I^0, seed_b)$ ▷ (simulated matches)

$Y^0[:, b] \leftarrow \mathbf{1}[\hat{\mu} \in H]$ ▷ (simulated outcomes)

$\bar{Y}^0 \leftarrow \sum_{b=1}^B \sum_{i=1}^n Y^0[i, b]/(nB)$ ▷ (simulated sample mean)

This is a simulation-based estimator that relies on 4 steps:

1. Draw n random samples of reported types with replacement from the observed control profile θ_C to create simulated profile $\widehat{\theta}_I^0$.

2. Input these simulated types $\widehat{\theta}_I^0$ into the matching mechanism φ and save the simulated matching outcomes vector $Y^0[:, b]$.

3. Repeat steps (1)-(2) B times, where B is large.

4. Collapse matrix Y^0 by averaging over all components of all vectors $\{Y^0[:, b]\}_{b=1}^B$ to output the scalar $\bar{Y}^0$.

Intuitively, this estimator uniformly samples types with replacement from the empirical distribution induced by the reported control types. As there is no interference in the reported types (Assumption 1), the distribution of reported types $\widehat{\theta}_I^0$ generated by the resampling procedure approaches the population distribution of untreated types θ_I^0 in the large market. Identification of the counterfactual mean then follows from CD-ETE (Assumption 2) and standard results of bootstrap methods.

Lemma 3 (Consistent estimator of counterfactual mean). *If Assumption 1 and Assumption 2 hold, then algorithm 1 yields a simulated sample mean $\bar{Y}^0$ that is a consistent estimator of $\mathbb{E}[Y_i|\theta_I^0]$.*

Proof. As d_i is randomly assigned and Assumption 1 holds, the empirical distribution $F^n_{\theta_C}$ of reported types in the control group θ_C is an independent random sample of size $|C|$ from the population distribution F_{θ^0}.

Then, the random sample $\widehat{\theta}^0_I$ comes from independently drawing n random types with replacement from the empirical measure $F^n_{\theta_C}$. This induces the empirical distribution $\widehat{F}^n_{\theta^0}$.

When $n \to \infty$, by the Glivenko-Cantelli theorem,

$$\lim_{n\to\infty} \sup |\widehat{F}^n_{\theta^0} - F_{\theta^0}| = 0,$$

almost surely.

If φ complies with CD-ETE (Assumption 2), then the plug-in empirical measure $\widehat{F}^n_{\varphi}$, induced by $\varphi(\widehat{\theta}^0_I)$, also converges almost surely to the population distribution F_{φ}:

$$\lim_{n\to\infty} \sup |\widehat{F}^n_{\varphi} - F_{\varphi}| = 0.$$

As φ is a random mechanism, we can draw samples from this consistent estimator of the counterfactual distribution F_{φ} by randomly drawing match-outcomes $\{Y^0[:, b]\}_{b=1}^{B}$ with a large B, as is standard in bootstrap methods (Efron and Tibshirani, 1994). Almost-sure convergence in the CDF of simulated outcomes resulting from the random matches $\widehat{\mu}$ is guaranteed by the Continuous Mapping Theorem because Y_i is a linear function of $\mu(i)$.

Finally, for bounded random variables, if CDFs converge, then their moments also converge. Thus, the sample average $\bar{Y}^0$ converges almost surely to the population expectation $\mathbb{E}[Y_i|\theta^0_I]$, i.e.

$$\lim_{B\to\infty} \lim_{n\to\infty} \sum_{b=1}^{B} \sum_{i=1}^{n} Y^0[i, b]/(nB) = \mathbb{E}[Y_i|\theta^0_I],$$

almost surely.

□

The result in Lemma 3, directly implies identification of τ and σ, as shown below.

Theorem 1 (Consistent estimator of τ and σ)**.** *If Assumption 1 and Assumption 2 hold, then $\hat{\tau} := \sum_{i\in T} Y_i/|T| - \bar{Y}^0$ is a consistent estimator of τ, and $\hat{\sigma} := \sum_{i\in C} Y_i/|C| - \bar{Y}^0$ is a* *consistent estimator of σ, where $\bar{Y}^0$ is the simulated sample mean obtained from algorithm 1.*

Proof. As d_i is randomly assigned, Lemma 3 implies that

$$\bar{Y}_0 \to \mathbb{E}[Y_i|\theta_I^0] = \mathbb{E}[Y_i|i \in T, \theta_I^0] = \mathbb{E}[Y_i|i \in C, \theta_I^0],$$

almost surely.

Then, by the Strong Law of Large Numbers,

$$\begin{aligned}
\hat{\tau} =& \sum_{i \in T} Y_i/|T| - \bar{Y}^0 \\
\to& \mathbb{E}[Y_i|i \in T, \theta_T^1, \theta_C^0] - \mathbb{E}[Y_i|i \in T, \theta_I^0] \\
=& \tau
\end{aligned}$$

and

$$\begin{aligned}
\hat{\sigma} =& \sum_{i \in C} Y_i/|C| - \bar{Y}^0 \\
\to& \mathbb{E}[Y_i|i \in C, \theta_T^1, \theta_C^0] - \mathbb{E}[Y_i|i \in C, \theta_I^0] \\
=& \sigma
\end{aligned}$$

almost surely. □

1.3.4 Inference

We now briefly turn to the matter of statistical inference. First, we focus on constructing confidence intervals for our point estimate. Assumption 1 and Assumption 2 not only imply that the counterfactual expectation $\mathbb{E}[Y_i|\theta_I^0]$ is identified (Lemma 3), but also that the counterfactual distribution of matches $F_\varphi|\theta_I^0$ is identified. By adjusting the procedure described in algorithm 1, under Assumption 1 and Assumption 2, we can also identify the distribution of matches $F_\varphi|(\theta_T^1, \theta_C^0)$. With these two objects, we can construct bootstrap confidence intervals for $\hat{\tau}$. Bootstrap confidence intervals will have correct coverage in large samples provided suitable regularity conditions hold (Horowitz, 2001). A valid procedure is described below:

1. Implement algorithm 1 and average the components of each column in Y^0 to form the set of bootstrapped empirical probabilities $\{Y_b^0\}_{b=1}^B$. As shown in the proof of Lemma 3, $\hat{F}_0(t) := \sum_{b=1}^B \mathbf{1}[Y_b^0 < t]/B$ is a consistent estimator of $F_\varphi|\theta_I^0$ mapped to outcomes Y_i.

2. Implement a variant of algorithm 1 but randomly sampling reported types with replacement from the treatment and control group, instead of only the control group (this variant is described below in algorithm 2). Average the treatment components (rows $i : \hat{\theta}_i \sim \theta_T^1$) of each column in Y^1 to form the set of bootstrapped empirical probabilities $\{Y_b^1\}_{b=1}^B$. Construct $\hat{F}_1(t) := \sum_{b=1}^B \mathbf{1}[Y_b^1 < t]/B$, which is a consistent estimator of $F_\varphi|T, (\theta_T^1, \theta_C^0)$ mapped to outcomes Y_i.

3. Randomly draw samples with replacement from $\hat{F}_1$ to compute the average $\tilde{Y}^1$ and sample from $\hat{F}_0$ to compute the average $\tilde{Y}^0$. Store the difference $\tau_k = \tilde{Y}_k^1 - \tilde{Y}_k^0$ and repeat the process for $k = 1, \ldots, K$, where K is large.

4. As in standard bootstrap applications (Efron and Tibshirani, 1994), use the percentile method to estimate the confidence interval by computing the required quantiles of the empirical distribution of $\{\tau_k\}_{k=1}^K$ for the desired confidence level (e.g. 5th and 95th percentile for a 90% confidence interval).

Similarly, we can use the bootstrapped sample $\{\tau_k\}_{k=1}^K$ to estimate standard errors by computing $\hat{se} := \sqrt{\sum_{k=1}^K (\tau_k - \bar{\tau})^2/(K-1)}$, where $\bar{\tau} = \sum_{k=1}^K \tau_k/K$ is the bootstrap sample mean.

Alternatively, we can directly test the hypothesis of zero treatment effect $H_0 : \tau = 0$ by using the independent samples of empirical assignment probabilities $\{\tilde{Y}_k^1, \tilde{Y}_k^0\}_{k=1}^K$. Formally, we can compute the bootstrap test statistic for equality of means Efron and Tibshirani (1994):

1. Compute and store the $t-$statistic $\hat{t} = \bar{\tau}/\sqrt{(\hat{se}_1^2 + \hat{se}_0^2)/K}$

2. Recenter the samples by computing $\bar{y}^1 = \sum_{k=1}^K \bar{Y}_k^1/K$, $\bar{y}^0 = \sum_{k=1}^K \bar{Y}_k^0/K$ and then generating $y_k^1 := \bar{Y}_k^1 - (\bar{y}^1 - \bar{y}^0)/2$, $y_k^0 := \bar{Y}_k^0 - (\bar{y}^0 - \bar{y}^1)/2$ $\forall k = 1, \ldots, K$. Now both recentered groups share the same sample mean (the pooled sample mean).

3. Form M datasets $\{\mathbf{y}^1, \mathbf{y}^0\}_{m=1}^M$ by drawing K random samples with replacement from each of the recentered samples in each case (so each dataset m has $2K$ observations).

4. Compute the $t-$statistic $\hat{t}_m$ for each of the M datasets.

5. Compute the empirical probability $\hat{p} := \sum_{m=1}^M \mathbf{1}[\hat{t}_m \geq |\hat{t}|]/M$.

These standard nonparametric bootstrap methods are valid in large markets, as CD-ETE (Assumption 2) along with the random assignment of d_i and Assumption 1 imply that we can construct two independent data generating processes $\hat{F}_1$ and $\hat{F}_0$ that approximate the distributions of outcomes Y_i arbitrarily well under profiles (θ_T^1, θ_C^0) and θ_I^0 (i.e. they converge almost surely). As the match-dependent outcomes are bounded and have finite moments, the procedures described above are valid estimators for confidence intervals, standard errors, and $p-$values associated to the null $H_0 : \tau = 0$.

1.4 Extensions

This section introduces extensions that build on algorithm 1. In each case, we provide identification results for different causal parameters under different identification assumptions and research designs.

1.4.1 Scaled Interventions

In many situations, researchers are interested in estimating the impacts of treatments or policies at scale, even though they only observe data coming from a small-scale intervention targeting a limited sample of the population (Allende et al., 2019). Specifically, they target estimands such as

$$\tau(r) := \mathbb{E}[Y_i | i \in T, \theta_T^1, \theta_C^0, r] - \mathbb{E}[Y_i | i \in T, \theta_I^0],$$

which is simply Equation 1.1, but conditioning the first expectation on having a fraction r of the population being treated. A common value of interest is $r = 1$, i.e the ATT $\tau(1)$ under the scenario where a policy is scaled to the whole population. However, other interior values of $r \in (0,1)$ can also be justified, such as when budget constraints do not allow for a full roll out of the intervention or they call for a downscale of the program.

An adaptation of algorithm 1 allows us to consistently estimate $\mathbb{E}[Y_i | i \in T, \theta_T^1, \theta_C^0, r]$ for any fraction $r \in (0,1]$. This is formalized in algorithm 2.

Algorithm 2: Consistent estimator of $\mathbb{E}[Y_i | i \in T, \theta_T^1, \theta_C^0, r]$

Data: $\theta_T, \theta_C, n, r, H,$ (large) B

Function: $\varphi : (\theta, seed) \mapsto \mu \in S^n$ satisfying CD-ETE

Init: $\widehat{\theta}_T^1 \in \Theta^{nr}, \widehat{\theta}_C^0 \in \Theta^{n(1-r)}, Y^0 \in \{0,1\}^{n \times B}$

Result: $\bar{Y}^1$ ▷ (consistent estimator of $\mathbb{E}[Y_i | i \in T, \theta_T^1, \theta_C^0, r]$)

for $b \in \{1, \ldots, B\}$ **do**

- $\widehat{\theta}_T^1 \leftarrow$ `SampleWithReplacement`$(\theta_T, nr, seed_b)$
- $\widehat{\theta}_C^0 \leftarrow$ `SampleWithReplacement`$(\theta_C, n(1-r), seed_b)$
- $\widehat{\theta}_I^r \leftarrow$ `Append`$(\widehat{\theta}_T^1, \widehat{\theta}_C^0)$ ▷ (simulated types)
- $\hat{\mu} \leftarrow \varphi(\widehat{\theta}_I^r, seed_b)$ ▷ (simulated matches)
- $Y^1[:, b] \leftarrow \mathbf{1}[\hat{\mu} \in H][: nr]$ ▷ (simulated outcomes)

$\bar{Y}^1(r) \leftarrow \sum_{b=1}^{B} \sum_{i=1}^{nr} Y^1[i, b]/(nrB)$ ▷ (simulated sample mean)

Intuitively, this variant of algorithm 1 resamples a fraction r of types from the treatment group and a fraction $1-r$ from the control group to form the simulated types $\widehat{\theta}_I^r$. This allows us to compute the estimate $\bar{Y}^1(r)$, which approximates a counterfactual situation where rn units are treated and the remaining are untreated.

Thus, provided that we can compute $\bar{Y}^1(r)$ using algorithm 2 and $\bar{Y}^0$ using algorithm 1, the difference between these two terms yields a consistent estimator of $\tau(r)$ as a function of the treatment fraction r, as long as r is orthogonal to the potential reported types $\{\theta^0, \theta^1\}$.

Corollary 1. *Let $r \in (0,1]$ and $\theta_i^d \perp r$ for $d \in \{0,1\}$. If Assumption 1 and Assumption 2 hold, then $\hat{\tau}(r) := \bar{Y}^1(r) - \bar{Y}^0$ is a consistent estimator of $\tau(r)$, where $\bar{Y}^1(r)$ is defined as the output of algorithm 2 and $\bar{Y}^0$ is the output of algorithm 1.*

Proof. See section A.3. □

The resulting estimates $\hat{\tau}(r)$ for different values of r allow researchers to trace out how their treatment effects are expected to vary under different scale-up or scale-down scenarios vis-à-vis the baseline scenario where no intervention took place, even if they only conducted a single RCT with a fixed r. This can be a policy-relevant tool, especially when it comes to evaluating the expansion of high-cost interventions that might induce applicants to significantly interfere with each other, such as providing free tuition to college applicants (Bucarey, 2018). Having an estimate of $\tau(r)$ allows policymakers to optimize the scale of an intervention (e.g. considering the cost of expanding the intervention, and some welfare metric that depends on $\tau(r)$) and target this optimal $r^\star$ after conducting a smaller experiment.[18]

1.4.2 Multiple Treatment Arms and Clustered Trials

Consider an experiment with multiple treatment arms indexed by $d = 0, \ldots, D$, where $d = 0$ denotes the "pure control" and $D \geq 1$ is the number of treatment arms. In this situation, just as in the standard case, the different treatments can affect the outcomes of the control group. However, these treatments can also interfere among each other in potentially complex ways (e.g. treatments that induce demand for similar schools can interfere more than treatments that promote disjoint sets of schools).

Indeed, there are $2^D - 1$ counterfactual situations of interest, considering all the combinations where each treatment either takes place or not (minus the observed "factual" situation where all treatments take place). Similar to algorithm 1 and algorithm 2, we can simulate all these $2^D - 1$ assignment probabilities, creating a set of estimates $\bar{Y}^a$, where the binary vector $a \in \{0, 1\}^D$ denotes the combination of "active" treatment arms in each estimate. For each estimate, the reported types in the inactive treatment groups are replaced by random samples with replacement from the control group ($d = 0$) and the match is simulated multiple times.

An analogous procedure can be applied to clustered trials, defined by clusters $k = 1, \ldots, K$, where each cluster has a control group and at least one treatment group. In this case, we can define the treatment arms in each cluster with the index $d_k = 0, \ldots, D_k$ and repeat the previous procedure. This yields a set of $2^{\mathbf{D}} - 1$ estimates $\bar{Y}^{\mathbf{a}}$, where $\mathbf{D} := \sum_{k=1}^{K} D_k$ and $\mathbf{a} \in \{0, 1\}^{\mathbf{D}}$ denotes the concatenation of active treatments in each cluster. Importantly,

[18] For example, if $\widehat{\tau}(r)$ is a monetary measure of the ATT (e.g. computed using estimates of value added in future income), and $c(r)$ is the average cost of providing the intervention to a share r of the population, then a policymaker interested in the welfare of the treatment group can solve $r^\star := \arg\max_{r \in [0,1]} nr[\widehat{\tau}(r) - c(r)]$. An analogous estimate of $\widehat{\sigma}(r)$ must be incorporated to account for the welfare of the control group.

for each active treatment arm, the resampling of control units should be done exclusively with units belonging to the cluster of the corresponding treatment group.

For example, consider the design in Ajzenman et al. (2023), where they describe a clustered RCT in the Chilean centralized college application system. They study the effectiveness of humans and chatbots motivating high school students to apply for education degrees when submitting their college applications. In the first cluster ($k = 1$) they targeted high school students who had expressed interest in education degrees with both human and chatbot promoters (so $D_1 = 2$), whereas in the second cluster ($k = 2$) they targeted students interested in social sciences only with chatbots (i.e. $D_2 = 1$). In this case, there are $2^{(2+1)} - 1 = 7$ target parameters that can be estimated with this extension of algorithm 1.[19]

1.5 Empirical Applications

We now show how these tools can be applied in empirical settings with two examples. The first example uses data from the Chilean centralized school choice system and a simulated experiment changing the preferences of applicants in the system. The second example uses data from Ajzenman et al. (2021). This experiment took place in the Ecuadorian centralized teacher labor market, and made teachers more likely to apply for jobs at hard-to-staff schools.

1.5.1 Simulated Experiment: The Chilean School Choice System

Setting

In 2016, Chile initiated the implementation of a centralized school choice system for all PreK-12 public schools. The rollout of the system started with some levels in the Magallanes (southernmost) region in 2016, and ended with all levels across all regions in the country in 2020. In this application, we focus on the 2016 Magallanes-region implementation. Specifically, we simulate an experiment with the 2016 applicants seeking admission to 9th grade[20] in 2017.

[19] To be explicit about the notation above, in this case the vector $\mathbf{a} \in \{0,1\}^3$ represents the active treatment indicators (human-education, bot-education, bot-social science)$\in \{0,1\}^3$.

[20] In Chile, the 9th grade of instruction is the first grade of high school and is called *primero medio*.

In this centralized school choice system, applicants can rank as many schools as they want and then submit this rank-ordered-list (ROL) to an online platform. Only schools that offer their corresponding grade (9th grade in this case) can be listed. The reported types θ are the preferences submitted to the platform. In this specific cohort of 9th grade applicants in 2016, the matching algorithm φ is equivalent to the Student-Proposing Deferred Acceptance (DA) Algorithm (Gale and Shapley, 1962) with a single random tie-breaker, which satisfies CD-ETE.[21] The student-proposing DA algorithm with a single tiebreaker is described below:

Step 0 Generate random tiebreakers $u_i \sim U[0,1]$, such that schools have strict preferences over applicants.

Step 1 All students apply to their favorite school, and all schools *tentatively* accept their most preferred applicants up to their capacity nq_s. All other applicants are rejected.

Step k All students rejected in step $k-1$ apply to their next preferred school. All schools *tentatively* accept their most preferred applicants among the pool of new applicants and previously accepted applicants, up to their capacity nq_s. All other applicants are rejected.

Stop When there are no more rejections, the final allocation is the tentative assignment in the last step.

Data

This implementation of the school choice match was the smallest in the history of the Chilean school choice system: for the 9th grade level there were $|I|$ =1040 applicants that could rank up to $|S|$ =24 schools offering that level. For each applicant, we observe data on the individually submitted types θ (including the realized random tie-breaking numbers) and the resulting DA allocation $\mu(i), \forall i \in I$. We also observe the school capacities $nq_s, \forall s \in S$.

We report summary statistics of this data in Table 1.1. Even though applicants can list up to 24 schools, the average student listed only 3.7 schools, and the longest list had 15

[21]There is a second round held for individuals that remain unmatched after the DA allocation, but we only study the first round of the system. Moreover, for other grade-levels and other years of implementation, the algorithm considers other factors, such as priorities based on siblings or test scores at some selective schools. For more details on how the Chilean matching algorithm works, see Correa et al. (2019).

schools ranked. Moreover, the average school has 53.67 vacant seats, which implies that there is excess supply of vacancies in this system (1288 seats available for 1040 applicants). If schools have heterogeneous qualities and some high-quality schools have slack capacity, inducing demand for these schools could be welfare-enhancing (Ainsworth et al., 2020).

Using a rich set of controls from the Ministry of Education (MINEDUC) and test results from Chile's national 10th-grade math exam (SIMCE), we estimate a fixed-effects regression model with the 2018 (future) SIMCE test scores as a dependent variable:

$$score_i = \sum_{s \in S} \alpha_s \mathbf{1}[\mu(i) = s] + \beta' X_i + \varepsilon_i,$$

where $score_i$ denotes the math test score of student i, X_i contains a set of controls (schooling records, including grades and attendance for the previous 10 years of schooling), and ε_i is assumed to be uncorrelated with the regressors. Under unconfoundedness, the set of fixed effects $\{\alpha_s\}_{s \in S}$ is a value-added measure for schools.

We plot these coefficients in Figure 1.2, along with the fraction of ROLs that list each school. This figure depicts how demand for schools is *negatively* correlated with the estimated value added. One mechanism that could explain this empirical finding is the presence of uninformed applicants, which has both efficiency implications (e.g. if seats at high-value-added schools remain unfilled) and distributive implications (e.g. if low-income families are more prone to being uninformed).[22]

The aforementioned descriptive facts motivate our simulated experiment. Consider an intervention where 20% of the applicants ($r = 0.2$) randomly change their preferences to follow the estimated value-added profile in Figure 1.2, i.e. all students in T list the highest value-added school first, the second highest value-added next and so on. The types of the remaining students are unchanged. Suppose that the outcome of interest Y_i is an indicator equal to 1 if student i is matched with a top-5 VA school.

[22] Even though this correlation is not causal, and there might be other mechanisms explaining the pattern (e.g. students might not apply to high-value-added schools because they offer few seats), we use it to justify the simulated experiment below in the spirit of Allende et al. (2019).

Table 1.1: Summary Statistics

Variable	Mean	St. Dev.	Min.	P25	Median	P75	Max.
Panel A: Schools (N=24)							
Vacancies	53.67	55.78	0	5	31	97	173
Copayment Indicator	0.7917	0.4149	0	1	1	1	1
Coed. Indicator	0.4583	0.5090	0	0	0	1	1
Panel B: Students (N=1,040)							
Female Indicator	0.4769	0.4997	0	0	0	1	1
Low-SES Indicator	0.5221	0.4998	0	0	1	1	1
(Younger) Sibling Indicator	0.0837	0.2770	0	0	0	0	1
Panel C: Applications (N=3,846)*							
Priority Indicator	0.0941	0.2920	0	0	0	0	1
List Length	3.70	1.58	2	3	3	5	15
Panel D: Allocation (N=966)							
1st pref. Indicator	0.6605	0.4738	0	0	1	1	1
2nd pref. Indicator	0.1594	0.3663	0	0	0	0	1
3rd pref. Indicator	0.1046	0.3061	0	0	0	0	1

Note: This table reports summary statistics using data from the Chilean centralized school choice system.
(*List length is collapsed)

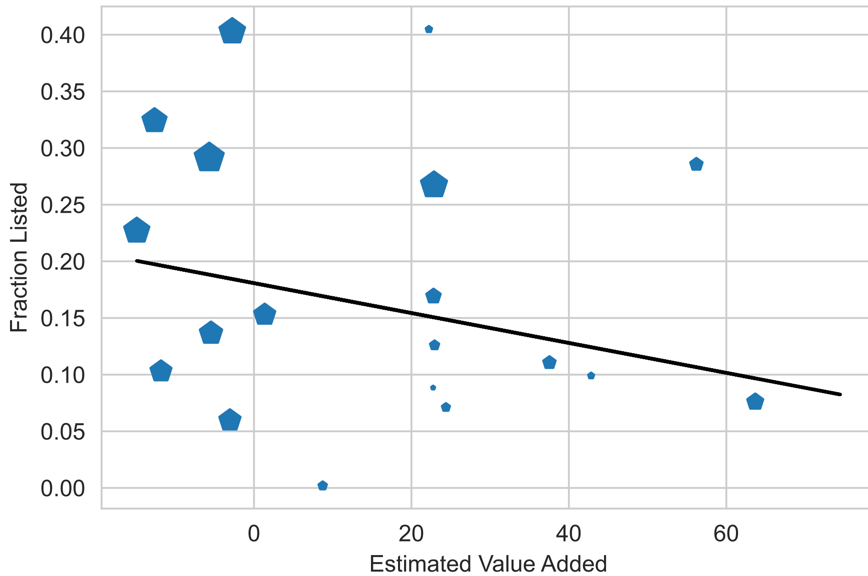
Results

Table 1.2 reports the results of conducting two estimation procedures on the experimentally modified data. The first column reports the estimates from the regression model described by

$$Y_i = \alpha + \beta Treatment_i + \varepsilon_i,$$

where the estimate $\hat{\beta}$ is equivalent to the difference-in-means estimator $\hat{\Delta}$. The second column shows the resulting estimates from applying our consistent estimator described in Theorem 1.

Figure 1.2: Value added is negatively correlated with share listed



Note: This figure plots the estimated value added for each school and the fraction of applications that list each school. Pentagon sizes are proportional to available slots.

Table 1.2: Difference-in-means vs. Consistent estimator

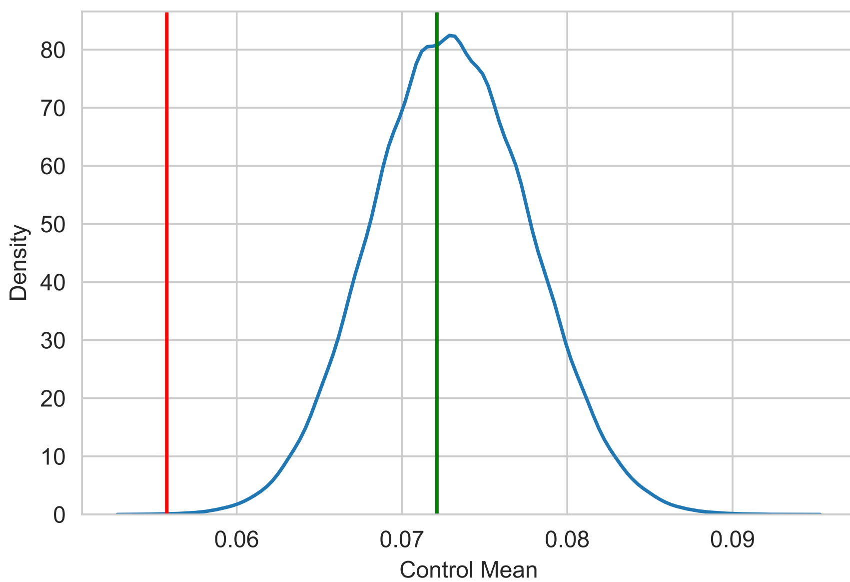
	(1) Mean comparison	(2) Consistent estimator
Treatment Effect	0.0635***	0.0462*
	(0.0174)	(0.0180)
Control Mean	0.0558***	0.0730***
	(0.0123)	(0.0127)
Observations	1040	1040

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Column (1) reports the estimates of a simple linear regression model with the treatment indicator as the independent variable. Column (2) reports the results of the consistent estimator derived from applying algorithm 1 to the simulated data. The $t-$statistic associated with the null hypothesis of equal treatment effects is 0.689 ($p = 0.491$).

The standard regression model estimate implies a 6.4 percentage point increase in the probability of being admitted into a top-5 school. On the contrary, the result of applying

Figure 1.3: Consistent Estimation of the Control Mean

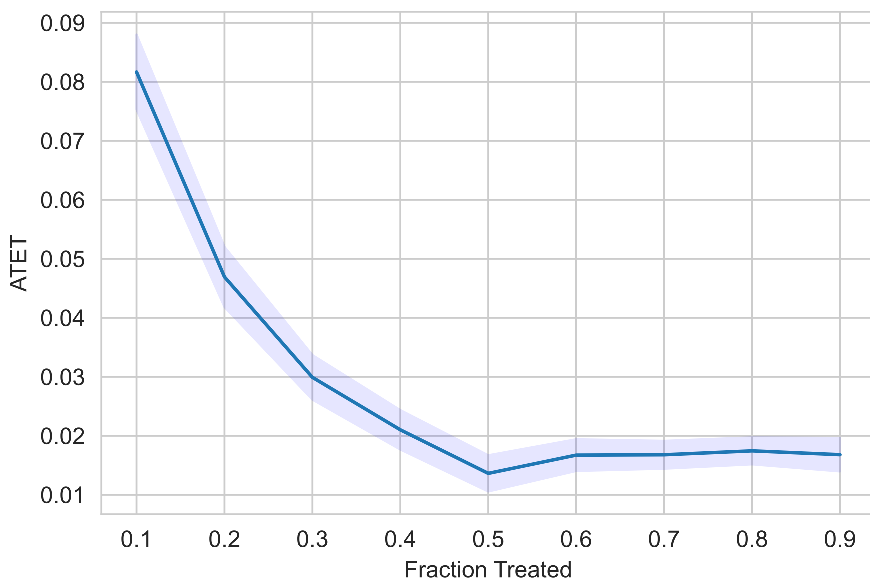


Note: The red line denotes the observed control average of $Y_i, \forall i \in C$ in the experiment, without any corrections. The green line denotes the original control mean of Y_i, in absence of the experimental manipulation, i.e. the true counterfactual mean. The blue function is a smoothed density estimate of the simulated control means derived from applying algorithm 1 to the experimental data.

our consistent estimator to the data suggests an increase of only 4.6 percentage points. The former overestimates the latter by over 37%, albeit the difference in the estimated coefficients is not statistically significant at conventional levels.

Recall that in this simulated experiment, we artificially modified the data. Therefore, we can observe the true "counterfactual" distribution of the data, in absence of any treatment. Figure 1.3 shows how the consistent estimator correctly recenters the control mean by comparing the simulated control means with the true counterfactual probability of assignment into a top-5 school. Indeed, the red line shows the observed control average after the experiment, without using algorithm 1. This observed average significantly underestimates the control mean in absence of the treatment, suggesting that the treatment had spillover effects on the control group that reduced their likelihood of being assigned to a top-5 school. Thus, because the control mean is being underestimated, $\sigma < 0$, the treatment effect is being overestimated, i.e. $\widehat{\Delta} \to \tau - \sigma > \tau$.

Figure 1.4: Consistent Estimates of $\tau(r)$



Note: The blue line plots the point estimates of $\tau(r)$ for different values of r. The shaded regions denote the 95% confidence intervals around the point estimates.

Finally, we apply the extension proposed in subsection 1.4.1 to estimate $\tau(r)$ under different fractions of treatment population $r \in \{0.1, 0.2, \ldots, 0.8, 0.9\}$. The results are plotted in Figure 1.4. These results show a strong decay in $\tau(r)$ for low values of r, with a flatter curve after $r = 0.5$. This could be the case if for high values of r the target schools fill their capacities such that the observed differences between treatment and control units are driven less by τ and more by spillovers σ.

This simulated experiment highlights the empirical relevance of the theoretical results shown above. In this controlled simulation environment, we can observe how recentering the control mean using algorithm 1 to account for SUTVA violations can change the conclusions drawn from an experiment. We now also show how this can be the case in a *real* experiment.

1.5.2 Real Experiment: The Ecuadorian Teacher Labor Market

Setting

The Ecuadorian Ministry of Education uses a centralized matching system to assign teachers to jobs in public schools. The matching system is called *Quiero Ser Maestro*[23] (QSM) and has been in place since 2013. This empirical application focuses on a nationwide experiment conducted in 2019 by the Ministry of Education in collaboration with the Interamerican Development Bank (IDB). More details about this experiment can be found in Ajzenman et al. (2021).

The objective of this experiment was twofold. First, the Ministry identified a set of schools which they classified as "hard to staff", and decided to experimentally evaluate an intervention to better staff these schools. Hard-to-staff schools, which were categorized as such prior to the experiment, tend to be located in low-income areas and exhibit low average test scores when compared to other schools (see Figure A.1 and Figure A.2). Importantly, teaching jobs at these hard-to-staff schools constitute a fixed subset of jobs $H \subset S$ and the Ministry cared about the "staffing indicator" $Y_i = \mathbf{1}[\mu(i) = H]$, which is 1 when a teacher i is matched to a hard-to-staff position.

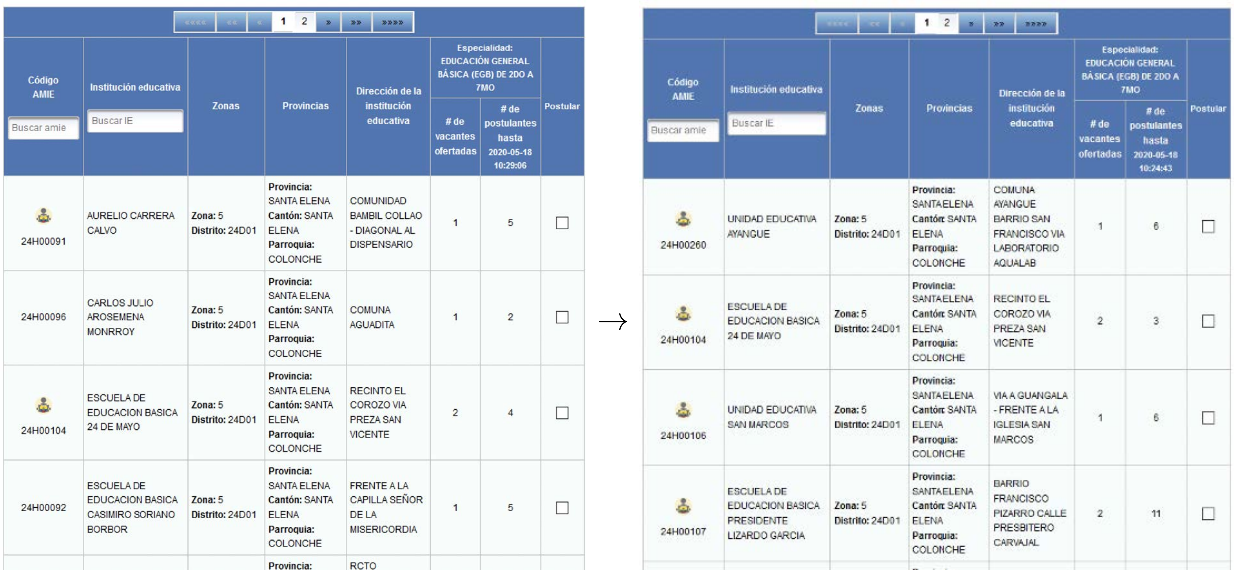
Second, the Ministry and the IDB aimed to evaluate a zero-cost intervention to test for "order effects" in this matching system. Specifically, a third of the teachers applying for jobs were randomly assigned into a treatment group T and saw hard-to-staff schools sorted at the top of the list of options in the application platform ($d_i = 1$). The control group C saw schools listed in alphabetical order ($d_i = 0$). Figure 1.5 depicts how the application platform looked for treatment and control teachers. They intended to test if highlighting hard-to-staff schools at the beginning of the applicants' list of options would translate into a higher likelihood of applying and being matched to these target schools. In other words, they wanted to test how reports θ_i vary, and how staffing outcomes Y_i change as a result.

When teachers apply in the online portal, they first filter schools by *parroquia*[24] and then the system automatically shows all the job openings for their teaching specialty (e.g.

[23] This is "I want to be a teacher" in Spanish.

[24] *Parroquias* (parishes) are geographic administrative units used in Ecuador. They are similar to municipalities in other countries.

Figure 1.5: Screenshots of application platform (Ajzenman et al., 2021)



Código AMIE	Institución educativa	Zonas	Provincias	Dirección de la institución educativa	Especialidad: EDUCACIÓN GENERAL BÁSICA (EGB) DE 2DO A 7MO # de vacantes ofertadas	# de postulantes hasta 2020-05-18 10:29:06	Postular
24H00091	AURELIO CARRERA CALVO	Zona: 5 Distrito: 24D01	Provincia: SANTA ELENA Cantón: SANTA ELENA Parroquia: COLONCHE	COMUNIDAD BAMBIL COLLAO - DIAGONAL AL DISPENSARIO	1	5	☐
24H00096	CARLOS JULIO AROSEMENA MONRROY	Zona: 5 Distrito: 24D01	Provincia: SANTA ELENA Cantón: SANTA ELENA Parroquia: COLONCHE	COMUNA AGUADITA	1	2	☐
24H00104	ESCUELA DE EDUCACION BASICA 24 DE MAYO	Zona: 5 Distrito: 24D01	Provincia: SANTA ELENA Cantón: SANTA ELENA Parroquia: COLONCHE	RECINTO EL COROZO VIA PREZA SAN VICENTE	2	4	☐
24H00092	ESCUELA DE EDUCACION BASICA CASIMIRO SORIANO BORBOR	Zona: 5 Distrito: 24D01	Provincia: SANTA ELENA Cantón: SANTA ELENA Parroquia: COLONCHE	FRENTE A LA CAPILLA SEÑOR DE LA MISERICORDIA	1	5	☐
			Provincia:	RCTO			

Código AMIE	Institución educativa	Zonas	Provincias	Dirección de la institución educativa	Especialidad: EDUCACIÓN GENERAL BÁSICA (EGB) DE 2DO A 7MO # de vacantes ofertadas	# de postulantes hasta 2020-05-18 10:24:43	Postular
24H00260	UNIDAD EDUCATIVA AYANGUE	Zona: 5 Distrito: 24D01	Provincia: SANTAELENA Cantón: SANTA ELENA Parroquia: COLONCHE	COMUNA AYANGUE BARRIO SAN FRANCISCO VIA LABORATORIO AQUALAB	1	6	☐
24H00104	ESCUELA DE EDUCACION BASICA 24 DE MAYO	Zona: 5 Distrito: 24D01	Provincia: SANTAELENA Cantón: SANTA ELENA Parroquia: COLONCHE	RECINTO EL COROZO VIA PREZA SAN VICENTE	2	3	☐
24H00106	UNIDAD EDUCATIVA SAN MARCOS	Zona: 5 Distrito: 24D01	Provincia: SANTAELENA Cantón: SANTA ELENA Parroquia: COLONCHE	VIA A GUANGALA - FRENTE A LA IGLESIA SAN MARCOS	1	6	☐
24H00107	ESCUELA DE EDUCACION BASICA PRESIDENTE LIZARDO GARCIA	Zona: 5 Distrito: 24D01	Provincia: SANTAELENA Cantón: SANTA ELENA Parroquia: COLONCHE	BARRIO FRANCISCO PIZARRO CALLE PRESBITERO CARVAJAL	2	11	☐

(a) Control Screen (b) Treatment Screen

Note: The left panel shows a screenshot of the application platform for a teacher in the control group and the right panel shows a screenshot of the application platform for a teacher in the treatment group. On the left screen schools are sorted alphabetically ("Aurelio...", then "Carlos...", then "Escuela"), while on the right screen the schools at the top of the list are all hard-to-staff schools (highlighted with a circular logo in the first column).

high school math) in the selected parroquia. Some parroquias did not have hard-to-staff schools, while others only had hard-to-staff schools. This implies that teachers who only saw these parroquias did not observe any variation induced by the treatment. The IDB excluded teachers who were not subject to any sorting variation by the treatment due to their parroquia-specialty combination.

In addition, there was a second treatment arm targeting another third of the applicants. This second treatment was not deployed correctly, and thus the IDB also excluded this third of the sample from the analysis. The remaining applicants constitute the "restricted sample" defined by Ajzenman et al. (2021), and the experimental results are valid for this subset of applicants.

Once the application period closes, the QSM platform runs the teacher-proposing DA algorithm to assign teachers to jobs in public schools. Teacher preferences for schools are determined by the rank-ordered lists submitted in the online platform, which are limited to up to 5 jobs listed per applicant. School priorities for teachers are based on a composite score computed for each teacher at each job. Each teacher knows their scores in advance, as well as the number of vacancies for each job (e.g. teacher i knows that school s is hiring 2 high school math teachers and their composite score is 88.8 at that job).

Tiebreakers in this system give priority to teachers who ranked an option higher. In the event a tie persists, a (tentative) winner is picked at random. A more detailed description of the algorithm is presented in Elacqua et al. (2021), and summarized below:

Step 0 Generate tiebreakers giving priority to teachers that ranked jobs higher and a (dominated) random component. For example, a valid tiebreaker is

$$tiebreaker_{is} = \epsilon \times (6 - rank_{is} + u_i) \quad u_i \sim U[0, 1],$$

where $rank_{is} \in \{1, 2, 3, 4, 5\}$ is the position where teacher i ranked school s and $\epsilon > 0$ is a small enough scaling factor such that the original order induced by test scores is unaltered when adding $score_{is} + tiebreaker_{is}$, modulo the set of tied scores. After incorporating this tiebreaker, all preferences in the system are strict.

Step 1 All teachers apply to their favorite job, and all jobs *tentatively* accept their top-scoring applicants up to their capacity nq_s. All other applicants are rejected.

Step k All teachers rejected in step $k-1$ apply to their next preferred job. All jobs *tentatively* accept their top-scoring applicants among the pool of new applicants and previously accepted applicants, up to their capacity nq_s. All other applicants are rejected.

Stop When there are no more rejections, the final allocation is the tentative assignment in the last step.

Because of this specific tie-breaking rule and the constraint on the length of the rank-ordered lists, this variant of the DA mechanism is not strategy-proof for teachers. However, this mechanism satisfies CD-ETE, as it has a cutoff representation where cutoffs converge in the large market. In fact, the QSM mechanism is a variant of the Cambridge mechanism described in Agarwal and Somaini (2018).

Data

In 2019, there were $|I| = 27,235$ qualified teachers[25] that applied for a teaching job at public schools through the QSM platform. There were 9,074 teachers assigned to a *failed* treatment arm, and they were excluded from the analysis by the IDB. Out of the 18,161 remaining teachers, 12,379 were not subject to any treatment variation because of their parroquia-specialty combination. Thus, the restricted sample as defined by Ajzenman et al. (2021) contains only 5,782 applicants.[26]

Table 1.3 shows descriptive statistics for baseline characteristics under different sample specifications. In the unrestricted sample, about 72% of applicants are women, 51% are single and 10% belong to ethnic minorities. On average, they score about 66/100 points

[25] The Ministry of Education does not allow low-scoring teachers to apply for jobs in public schools. Qualified teachers are those who cleared the screening process comprised by a psychological test and a knowledge test developed by the Ministry. According to Ajzenman et al. (2021): "only 27% of the 129,114 candidates who registered for the teacher selection process passed the eligibility phase".

[26] There is a minor difference in the number of applicants in this analysis when compared to Ajzenman et al. (2021). This is due to 28 applicants not having an ID number in the data provided by the Ministry. In our analysis, we include these 28 applicants, as opposed to Ajzenman et al. (2021), where they are dropped. We do this to replicate the results of the matching process as closely as possible. With our full sample, we are able to correctly assign 99.8% of applicants to their assigned jobs by running the teacher-proposing DA algorithm. The difference is likely due to issues in the raw input data (e.g. some applicants list the same job more than once in their rank-ordered lists).

and have around 3.4 years of experience. The restricted sample is slightly different: there are 5pp more females, they 3pp more likely to be single, they score about 1.3 more points on average, they have about 0.4 more years of experience and they are 0.7pp more likely to belong to an ethnic minority. In all cases, the treatment and control groups are balanced in their baseline characteristics.

These applicants chose from $|S| = 10,067$ jobs when applying through the online platform. On average, jobs offered 1.44 openings, the median being 1 opening. The 75th percentile was 2 openings, the 95th percentile was 3 openings, the 99th percentile was 6 openings, and the maximum was 16 openings. As most jobs offer few vacancies, and the total number of vacancies is smaller than the number of applicants n, it is unlikely that SUTVA holds in this setting.

Results

The first set of results describes the effects of the treatment on the reported types θ. Specifically, we focus on the length of the reported lists and the presence of hard-to-staff schools in them. For these "partial equilibrium" outcomes, we consider Assumption 1 holds and report linear regression estimates.

Table 1.4 reports the results of estimating the parameters of the following regression model:

$$L_i = \alpha + \beta Treatment_i + \gamma' X_i + \varepsilon_i, \tag{1.2}$$

where L_i is either the length of the list submitted to the system, the share of hard-to-staff schools listed, an indicator for listing a hard-to-staff school or an indicator for listing one as a first choice. The vector X_i contains the same set of controls variables as Ajzenman et al. (2021): composite score, indicators for being female, being single, belonging to an ethnic minority, indicators for years of experience and the applicant's district.

Column (1) in Table 1.4 shows no significant difference in the length of the lists reported by treatment and control teachers. However, columns (2)-(4) indicate a significant change in the composition of the lists submitted by the treatment group. Treated teachers increase

Table 1.3: Baseline characteristics and sample balance tests

Variable	(1) Mean (sd)	(2) Control (sd)	(3) Treatment (sd)	(4) Difference (p-value)
Panel A: Full Sample				
Female	0.715	0.714	0.718	0.004
	(0.451)	(0.452)	(0.450)	(0.480)
Single	0.513	0.511	0.517	0.006
	(0.500)	(0.500)	(0.500)	(0.325)
Comp. Score	66.033	66.042	66.016	-0.026
	(9.914)	(9.954)	(9.833)	(0.840)
Experience	3.375	3.390	3.346	-0.044
	(3.278)	(3.291)	(3.253)	(0.297)
Minority	0.098	0.097	0.101	0.004
	(0.298)	(0.296)	(0.301)	(0.278)
Observations	27,235	18,176	9,059	27,235
Panel B: Excluding Failed Treatment Arm				
Female	0.716	0.715	0.718	0.003
	(0.451)	(0.452)	(0.450)	(0.648)
Single	0.512	0.507	0.517	0.010
	(0.500)	(0.500)	(0.500)	(0.167)
Comp. Score	66.016	66.016	66.016	-0.000
	(9.910)	(9.987)	(9.833)	(0.998)
Experience	3.359	3.371	3.346	-0.025
	(3.268)	(3.284)	(3.253)	(0.600)
Minority	0.099	0.096	0.101	0.005
	(0.298)	(0.295)	(0.301)	(0.252)
Observations	18,161	9,102	9,059	18,161
Panel C: Final Restricted Sample				
Female	0.766	0.768	0.765	-0.003
	(0.423)	(0.422)	(0.424)	(0.785)
Single	0.540	0.532	0.549	0.017
	(0.498)	(0.499)	(0.498)	(0.204)
Comp. Score	67.385	67.416	67.353	-0.064
	(9.393)	(9.467)	(9.319)	(0.797)
Experience	3.737	3.739	3.734	-0.005
	(3.275)	(3.275)	(3.276)	(0.954)
Minority	0.106	0.111	0.102	-0.010
	(0.308)	(0.314)	(0.302)	(0.241)
Observations	5,782	2,918	2,864	5,782

Note: This table shows summary statistics for baseline characteristics in different subsamples. Female is an indicator equal to 1 if the applicant is female. Single is an indicator equal to 1 if the applicant is single. Comp. Score is the base composite score of the applicant, excluding application-specific bonuses. Experience is the years of experience of the applicant. Minority is an indicator equal to 1 if the applicant belongs to an ethnic minority. Panel A contains all applicants in the data. Panel B excludes applicants that were assigned to the failed treatment arm. In addition, Panel C also excludes applicants that were not subject to treatment variation because of their parroquia-specialty combination. The final column reports differences in means between the treatment and control groups, along with the associated p-value when testing the null hypothesis of zero difference in means.

Table 1.4: Treatment effects on application behavior

Outcome:	(1) List Length	(2) Share of HTS	(3) Ranked HTS	(4) First is HTS
Treatment	-0.005	0.013	0.015	0.051
	(0.011)	(0.006)	(0.008)	(0.012)
Observations	5765	5765	5765	5765
Control Mean	4.908	0.431	0.865	0.402
p -value	0.65	0.044	0.048	<0.001

Robust standard errors, clustered at the district level, in parentheses.
Treatment p -values reported in the last row.

their share of hard-to-staff schools listed by 1.3 percentage points (a 3% increase) and are 1.5 percentage points more likely to list at least one hard-to-staff school (a 1.7% increase). Moreover, treated applicants are 5.1 percentage points more likely to list a hard-to-staff school as their first choice (a 12.7% increase). Results for other samples are comparable to those in Table 1.4 and are reported in the Appendix (Table A.1 and Table A.2).

It is important to notice that 86.5% of the control group ranked at least one hard-to-staff school, as shown in column (3) of Table 1.4. If hard-to-staff schools are "small", and both experimental groups have high application rates to these schools, treatment applicants might crowd out control applicants from hard-to-staff schools when the treatment induces additional demand for these schools.

We now turn to the estimation of the treatment effect that the intervention had on the matching outcome Y_i. Table 1.5 shows the estimated treatment effects from computing the difference-in-means estimator $\widehat{\Delta}$ and the consistent estimator $\hat{\tau}$ introduced in Theorem 1. Column (1) shows the results of computing the treatment effect using a standard linear regression (Equation 1.2). This estimate implies that teachers who saw hard-to-staff schools at the top of their list were 3.4 percentage points more likely to be assigned to a hard-to-staff school than the control group. The point estimate is significantly different from 0 and accounts for a 12.8% increase in the odds of being assigned to a hard-to-staff school when compared to the control mean.

However, in column (2), we account for SUTVA violations using algorithm 1 to adjust the control mean. Our consistent estimator suggests that the impact of the policy is significantly

Table 1.5: Treatment effects on match outcome

Outcome:	(1) Matched to HTS (Inconsistent)	(2) Matched to HTS (Consistent)
Treatment	0.034	-0.009
	(0.012)	(0.012)
Observations	5765	5765
Control Mean	.27	.313
p -value	.005	.448

Robust standard errors, clustered at the district level, in parentheses.
Treatment (bootstrapped) p -values reported in the last row.
The $t-$statistic associated with the null hypothesis of equal coefficients is 2.553 ($p = 0.011$).

different from $\widehat{\Delta}$, and cannot be statistically distinguished from zero. Indeed, the simulated control mean is about 4 percentage points higher than the observed control mean in the sample. This implies that the observed difference in means, which can be decomposed as $\hat{\tau} - \hat{\sigma}$, is mostly driven by spillovers of the treatment group on the control group, i.e. treated teachers were likely to displace control teachers in the experiment.

1.6 Conclusion

In this paper, we study the estimation of treatment effects in an RCT setting where the experimental units participate in a matching system. In this setting, the no-interference assumption is unlikely to hold when measuring match-dependant outcomes, such as being matched to a hard-to-staff school in the Ecuadorian teacher labor market. This is because treated units might affect the equilibrium placement of control units, as each individual match depends on the system-wide profile of reported types. For example, if an intervention induces applicants to list hard-to-staff schools when applying for a job, this additional demand for the target schools might displace control units that would have listed hard-to-staff schools in the absence of any treatment.

Interference poses a challenge to the estimation of causal treatment effects, even if the data comes from a well-executed RCT. We first show how conventional estimators that compare the average outcomes for treatment units against control units are biased and inconsistent whenever there is interference in a large experiment (i.e. when the fraction of treated units $r \not\to 0$). Indeed, this observed difference not only accounts for the ATT, but also contains the spillover effect, which we define as the "treatment effect" on the control group. Moreover, in empirical settings where the treatment induces demand for a set of options and the outcome depends on being matched to these options, the estimates obtained from the standard difference-in-means approach are likely to overestimate the true ATT. We show how this is the case in two empirical applications.

To solve the interference issue, we propose a nonparametric, simulation-based estimator that consistently estimates the target causal parameter under a modified version of the equal treatment of equals (ETE) condition. This condition on the matching algorithm, which we call continuous distributional equal treatment of equals (CD-ETE), allows researchers to decompose the observed average difference between treatment and control units into an estimator for the ATT, and a residual estimator for the spillover effect. We show the validity of this estimator when SUTVA holds at the report level (Assumption 1), without requiring it to hold for equilibrium outcomes.

Moreover, we propose extensions to this estimator, such that we can estimate treatment effects under different counterfactual intensities of treatment (different fractions of treatment shares r) and accommodate RCTs with multiple treatment arms and clusters. The former is a useful tool for policy recommendations, such as estimating treatment effects under an expansion or contraction of the intervention, and the latter fits popular RCT designs that are implemented in practice.

We apply our estimator using data from the Chilean school choice system, and the Ecuadorian teacher labor market match. In both cases, as expected, we find lower point estimates for the ATT parameter when correcting for the interference induced in the matching system. In the case of Ecuador, we cannot distinguish our estimated treatment effect from a statistical zero, implying that the observed difference between the treatment and control groups in that experiment is likely driven by control units being displaced by treatment units. This result highlights the importance of our contribution, as it pertains to experiments that are conducted for policy evaluations in settings with high-stake outcomes (e.g. jobs of teachers in a whole country).

Finally, in this paper, we specifically focus on RCTs conducted in matching systems that comply with CD-ETE. Future research can explore quasi-experimental settings where no RCT takes place, such as estimating treatment effects with interference using a difference-in-difference research design, or studying local interference where a change in reported types occurs in a regression discontinuity setting. More research related to estimating treatment effects on other downstream outcomes (e.g. future test scores) that can be a function of match outcomes is also warranted. Additionally, further research is needed to study the behavior of matching systems that either do not comply with CD-ETE (e.g. that comply with ETE, but not CD-ETE), or are part of a specific subset of mechanisms that provide additional structure that researchers can exploit to improve on our results (e.g. Agarwal and Somaini (2018) show how simulations of RSP+C mechanisms, which comply with CD-ETE, can be used to construct estimators of match probabilities that are also asymptotically normal).

Chapter 2

Aftermarket Frictions and the Cost of Off-Platform Options in Centralized Assignment Mechanisms

[1] **Abstract**

We study the welfare and human capital impacts of colleges' (non)-participation in Chile's centralized higher education platform, leveraging administrative data and two policy changes: the introduction of a large scholarship program, and the inclusion of additional institutions which raised the number of on-platform slots by approximately 40%. We first show that the expansion of the platform raised on-time graduation rates. We then develop and estimate a model of college applications, offers, waitlists, matriculation, and graduation. When the platform expands, welfare increases, and welfare, enrollment, and graduation rates are less sensitive to off-platform frictions. Gains are larger for students from lower-SES backgrounds.

2.1 Introduction

Providing equitable access to higher education is an important policy objective shared by countries all over the world. One way that education systems support this goal is with centralized application and assignment platforms that minimize application costs and provide

[1] Coauthored with Adam Kapor and Christopher Neilson.

transparent rules regarding access.[2] Centralized mechanisms are theoretically appealing and have been empirically successful in many settings (Abdulkadiroğlu et al., 2017).

In practice, however, the setting may depart from the theoretical ideal in ways that matter for the efficiency and fairness of students' assignments. One such departure is that in virtually every practical implementation there exist many off-platform options that are available to participants of the match. In primary and secondary education, these include private schools or charter schools that do not participate in the centralized system. In other cases, such as higher education, some providers may be excluded from the platform by regulation, while others may choose not to participate. When off-platform options exist, applicants may renege on their assigned matches in favor of programs that did not participate in the centralized process. In turn, these decisions lead to the use of waitlists and aftermarkets, which may be inefficient due to the presence of congestion and matching frictions, and can be inequitable if some students are better able to navigate this partially-decentralized process, negating some of the benefits of the match.

In this paper, we study the empirical relevance of the configuration of on- and off-platform options for students' welfare and for persistence and graduation in higher-education programs. We document the importance of negative externalities generated by off-platform options and quantify a measure of aftermarket frictions that contribute to generating them in practice. Our empirical application uses data from the centralized assignment system for higher education in Chile, which has one of the world's longest running college assignment mechanisms based on the deferred-acceptance algorithm.[3] We take advantage of a recent policy change that increased the number of on-platform institutions from 25 to 33, raising the number of available slots by approximately 40%. We first present an analysis of the policy which shows that when these options are included on the centralized platform, students start college sooner, are less likely to drop out, and are more likely to graduate within seven years. Importantly, these effects are larger for students from lower SES backgrounds, suggesting that the design of platforms can have effects on both efficiency and equity.

[2] As of 2020, at least 46 countries use centralized choice and assignment mechanisms for at least part of their higher education system. See Neilson (2021) for a review of countries that have implemented centralized choice and assignment mechanisms.

[3] A common national entrance exam was first implemented in 1967, and centralized assignment based on a Deferred Acceptance algorithm has been used for at least the last 45 years.

Next, we develop an empirical model to obtain an estimate of aftermarket frictions and to quantify the negative impacts caused by off-platform options as a function of these frictions and the configuration of on- and off-platform options. We estimate a model of college applications, aftermarket waitlists, off-platform offers, matriculation choices, and near-on-time graduation outcomes using individual-level administrative data on almost half a million on-platform applications, test scores, enrollment decisions, and student records at all on- and off-platform higher education options, spanning the years 2010-2012 and exploiting the policy change that expanded the platform. In addition, to recover price sensitivities, our estimation procedure leverages the introduction of a scholarship program with arbitrary eligibility cutoffs, which provides exogenous price variation across options among similar applicants.

We find that when students are allowed to express their preferences for a larger variety of options on the platform, welfare increases substantially, as does the share of students graduating on time. According to our estimates, the welfare gains from platform expansion are roughly 0.263m CLP or U.S. $650 per exam taker. The welfare impact of platform expansion, in turn, is over eight times as large as the further gains from removing all remaining matching frictions, and a quarter as large as the welfare impact of expanding the platform and making all on-platform programs free, a much more expensive policy change. Enrollment gains from platform expansion are more than 80% of those of platform expansion and removing all frictions in waitlists.[4] These quantitative results suggest that off-platform options generate negative impacts on the efficiency of the assignment system and that these costs can be economically meaningful.

We use the estimated model to further explore which students are affected by the off-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 pattern may be partly due to their higher sensitivity to price and lower utility for private off-platform options. We then use the model to evaluate how our results would change in

[4] 0.263m CLP represents welfare differences between a baseline scenario and a counterfactual in which the platform expansion did not occur. The baseline scenario gives 0.031m CLP lower welfare than a frictionless benchmark, while the absence of platform expansion would give 2.94m CLP lower welfare on average than this benchmark. Details are given in Table 2.5. The exchange rate was approximately 500 CLP per USD in 2012.

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.

Intuitively, when a desirable program is not on the platform, it can cause some students who would have placed in that program to instead receive a placement in a different program which is available on the platform. These students may then decline that placement in favor of the off-platform program, creating vacancies, which in turn lead to increased reliance on waitlists which may be subject to frictions. Moreover, the absence of a particular program may distort the placements of other students, even if the students whose placements are affected would never enroll in that program. These students may also be less satisfied and more likely to decline their placement.

Taken together, our results show empirically that the existence of off-platform options affects the equity and efficiency of centralized assignment systems. Our empirical framework and counterfactual analysis allow us to quantify the welfare effects of adding universities to the platform, and provide tools for evaluating the costs of off-platform options in other settings.

This paper builds on and contributes to the empirical literature on the design of assignment and matching procedures for education markets. Abdulkadiroğlu et al. (2017) estimate the welfare impacts of the introduction of a centralized match in New York City schools. Several papers estimate welfare impacts of changes in school assignment mechanisms (Agarwal and Somaini, 2018; Calsamiglia et al., 2018; Kapor et al., 2020). Aue et al. (2020) empirically investigate a merger of school districts. We contribute by quantifying the impacts of a novel aspect of the design of the market—which options are on-platform—and by linking it to real outcomes, such as dropout/graduation rates, in addition to revealed-preference welfare measures. Methodologically, we build on the Gibbs-sampler estimation procedure introduced by McCulloch and Rossi (1994) (see also Rossi et al. (1996)) and applied to rank-ordered school choice data by Abdulkadiroğlu et al. (2017). We extend this procedure to accommodate—in addition to rank-ordered application data—ex-post enrollment decisions in an aftermarket in which individuals' choice sets are unobserved. Our procedure constructs person-specific subsets of the set of programs at which placement chances are nontrivial, and assumes that students truthfully report preferences over this subset; it is related to the stability-based approach of Fack et al. (2019).

Our question is particularly related to issues surrounding "common enrollment" –i.e. school choice policies in which all available schools participate in a single centralized assignment process. Ekmekci and Yenmez (2019) prove that, in the absence of frictions, full participation by all schools or programs is best for students, but programs have incentives to deviate from the match and "poach" students in the aftermarket. Andersson et al. (2018) consider a setting in which private-school and public-school matches take place sequentially. The theoretical literature abstracts from frictions and communication failures in the aftermarket. Our goal is to quantify the impacts of platform expansion in the presence of the frictions that exist in the market, motivating the use of empirics.

More broadly, we contribute to a literature on problems that may arise in decentralized or imperfectly centralized matching markets. These include (lack of) market thickness, "congestion" in decentralized markets, and the inefficient timing or sequencing of transactions (Agarwal et al., 2019; Roth and Xing, 1994; Niederle and Roth, 2009). Our notion of aftermarket frictions captures the idea of congestion: a program has a limited time to process its waitlist, and may fail to contact some students to whom it wishes to extend offers, such as when a student fails to answer his/her phone. However, our model of aftermarket frictions does not accommodate frictions related to exploding offers, which were rare during our sample period.[5]

Our paper adds to a literature that relates choice behavior to outcomes in assignment mechanisms. In contemporaneous work, Agarwal et al. (2020) provide nonparametric identification results for preferences and outcomes in assignment markets. They observe that, in addition to an "assignment shifter" such as discontinuities in admissions offers, an additional source of variation in choices is needed which is excluded from outcomes. In our setting, year-to-year variation in programs' cutoffs plays this role.[6] Our approach to estimation is

[5]Programs may have incentives to make offers with short deadlines, either prior to the match or prior to waitlist movement, in order to capture some students who face uncertainty. Anecdotally, in the years prior to our sample period, off-platform programs made offers which required a large non-refundable deposit which was due after the initial match but before on-platform waitlists cleared. This practice was prohibited by the consumer protection law of Bill 19.955 in 2004, which required that such deposits be refundable as long as the academic program had not yet begun.

[6]Alternative approaches include distance as an excluded preference shifter in school choice (Walters,

closest to Geweke et al. (2003) and Agarwal et al. (2020), who jointly estimate preferences and outcomes (mortality, life-years) using a Gibbs sampler, in hospitals and deceased-donor kidney assignment procedures, respectively. In contemporaneous work using data from the Chilean higher education system, Larroucau and Rios (2020) estimates a dynamic model of preferences, learning about ability, and outcomes such as switching and dropout after enrolling in college.[7] Other papers that combine preference estimation with "outcomes" such as health, human capital, or labor-market impacts include Hull (2018), Walters (2018), and van Dijk (2019).

2.2 Context and Data

2.2.1 Administrative Data Sources

Our administrative data come from three sources. The Ministry of Education of Chile (MINEDUC) provides data for each combination of campus, institution, and major, which we refer to as a *program*. The data provided by MINEDUC assigns each program to a standardized category of broad area and field or major of specialization. MINEDUC also provides panel data on individual-level enrollment and financial aid allocated to each student.

The second source is the *Consejo Nacional de Educación* (CNED) which is the regulatory agency that provides accreditation to higher education programs. This agency publicly reports program information such as accreditation status, posted tuition and student body characteristics.

The third source of data is the agency that runs the centralized application and assignment mechanism (DEMRE) for participating universities. This agency also administers the national college entrance exam, the *Prueba de Selección Universitaria* (herein, PSU). The college entrance exam is a set of multiple-choice tests that comprise a verbal and math component, as well as optional history and science tests. All test scores are standardized so

2018), variation in the set of other units available in a housing allocation mechanism (van Dijk, 2019), and variation in the distribution of future offers in a dynamic decision problem (Agarwal et al., 2020).

[7]Other research estimating preferences for college and major in the context of Chile includes Bucarey (2018), which studies equilibrium effects of a reform in Chile which made college free in 2016. Larroucau and Rios (2020) ask how learning and dynamics can affect the efficiency of the assignment mechanism.

that the sample distribution of each test in each year resembles a normal distribution with a mean of 500 points and a standard deviation of 110 points. The minimum score of the test is assigned a score of 150 points, and the maximum corresponds to 850 points. High school GPA is also transformed to be on the same scale. DEMRE provides masked individual level data on students who took the PSU test including their gender, high school, approximate geographic location, GPA, and test score results.

The agency also provides student-level data on rank-ordered applications, the assignment associated with the initial application, and reported matriculation from the institutions. Importantly, unique identifiers allow us to cleanly link individuals across datasets. The study focuses on the years 2010, 2011 and 2012. Descriptive statistics of the data are presented in Table 2.1.

2.2.2 Chilean Higher Education in Context

Growth and Consolidation of Higher Education

Over the last three decades, the Chilean higher education system expanded dramatically. This rapid growth in tertiary enrollment in Chile was spurred by a combination of a growing middle class and policies such as government backed student loans and scholarships. Growth in demand led to an expansion in the number of programs at newer private institutions (Ferreyra et al. (2017)). In 1989 there were 25 (16 public and 9 private non-profit) universities in Chile, which we will call the G25. These universities enrolled a total of 112,000, 215,000 and 310,000 students in the years 1990, 2000 and 2010, respectively. The decade after 2010 saw a period of consolidation with smaller growth in enrollment, with total matriculation at G25 universities reaching 366,000 in 2019. Since the 1970s the G25 universities have participated in a centralized clearinghouse for processing college applications and admissions. The emergence of newer universities established after 1990 led to an increasing share of enrollment off the centralized platform. Non-G25 universities represented 68% of total enrollment in 2010. In addition to universities, there also exist professional institutes and technical formative centers.

Rise of G8 private universities

Although G25 enrollment increased during the 1990s and 2000s, most of the growth in enrollment occurred at newer private universities outside of the G25. Private universities outside of the G25 enrolled 20,000, 100,000 and 320,000 students in 1990, 2000 and 2010, respectively. By 2019, matriculation had reached 350,000, representing 27% of all college enrollment. A group of eight of the largest and more selective private universities not only saw their enrollment grow but also their share of higher scoring students, especially from private schools. We refer to this group as the G8 throughout the paper. This group is heterogeneous in the location of their campuses and the strengths and specialties of their institutions but had become a close substitute for many traditional programs in the G25. By 2010, the G8 universities had 32% of total G8 + G25 enrollment. While the two most selective institutions belong to the G25, some G8 institutions are much more selective than most G25 institutions, with considerable overlap of selectivity among them.[8]

Financial Aid

A distinctive feature of the structure of financial aid in Chile is that the eligibility rules are a clear function of student and program characteristics known before applying. The average of students' math and verbal test scores determine one dimension of eligibility. The second dimension is a publicly known SES index. Students with scores above a test score cutoff and SES below an SES cutoff were eligible for low-interest government-backed student loans and scholarships, which they could use at any eligible program, including all G25 and G8 programs during our study period. These scholarships provided varying amounts of funding as a function of the student's SES. Government-backed loans covered the remainder up to a program-specific reference tuition. Importantly, this funding was not tied to whether the program participated in the centralized assignment platform. Moreover, eligibility for financial aid is determined before students apply to programs, and follows students to programs. While few general-use scholarships were being provided in 2010, government-backed student loans were used widely and have been shown to significantly

[8] Table B.1 lists each institution in the G25 and G8 and presents statistics regarding the distribution of student test scores at each institution.

alleviate credit constraints and facilitate college attendance when comparing students at the margin of loan eligibility Solis (2017). The vast majority of students who are eligible to apply to programs on the centralized platform are eligible for student loans, and all options in the G25 and G8 were eligible to receive both loans and scholarships[9].

The BVP Scholarship

In 2011 a significant new scholarship policy called the *Beca Vocación de Profesor* (BVP) was introduced with the goal of recruiting teachers with high exam scores. This scholarship covered the full tuition bill for students scoring at least 600 point average on math and verbal admissions exams, a value one standard deviation above the mean, if they enrolled at eligible teaching programs. In return, it imposed a test-score floor, prohibiting participating programs from admitting students with mean math and verbal test scores below 500 points, a value equal to the mean among test-takers. Gallegos et al. (2022) describe the policy and find large impacts on enrollment decisions via regression discontinuity and difference-in-difference designs. In 2010, teaching was the most popular major in Chile. This policy therefore shifted choices for a significant portion of students by effectively eliminating tuition at a subset of options for some students and drastically limiting access to programs for other students. We use this program as a source of match-level price variation in order to estimate willingness to pay for programs.[10]

2.2.3 Institutions Surrounding College Applications and Enrollment

Students Take Tests

Each year, students interested in potentially applying to universities must register to take the national college entrance exam in mid December. This test is free for over 90% of high school graduates, the majority of whom take the test. In 2011, 67% of all current high school graduates took the test, representing 79% of all test takers that year (graduates from

[9]Figure B.1 shows a timeline indicating the major policies promoting access to higher education before and after our sample period. Universities also offer some financial aid options. More information on institution-specific aid can be found in Table B.2.

[10]More details about the policy and the price variation it generates are provided in chapter 3.

previous years may take the test as well). Test results are made available to students in early January. Students are eligible to apply through the centralized admissions system if they obtain a simple average of at least 450 points between their math and verbal tests (450 is half of a standard deviation below the mean of each test). Students with an average math and verbal score below 450 cannot apply, but may retake the tests in the following year if they want to do so. Approximately 250,000 students took the college entrance exam in 2010, 2011 and 2012. In each of these years approximately 10% of test takers were graduates from private high schools which are not subsidized. Panel B of Table 2.1 shows that, among G25 applicants, the fraction of "current cohort" applicants enrolled in high school at the time of applications did not change significantly within the period studied, going from 59% to 62% between 2010 and 2011, and then to 58% in 2012.[11]

Programs Report Capacities and Admissions Rules

Each program on the centralized platform reports to the mechanism a set of weights on subject test scores and high school GPA. Programs choose their weights, subject to constraints, to express preferences for their applicants, who will be ranked according to the weighted average of their scores induced by these weights.[12] Programs also report the desired number of slots to be provided to the mechanism. In 2011, There were approximately 1000 programs among the G25 universities, which together accounted for 67,000 slots. The G8 universities offered 350 additional programs that accounted for an additional 25,000 slots.

Students Report Ranked Ordered Lists

Eligible students who decide to apply to universities on the centralized platform must do so within a short window of time (approximately a week) after receiving their scores. Applica-

[11]Prior to 2012, PSU scores were valid for a single admissions cycle. In 2012 and following years, PSU scores were valid for two admissions cycles. Therefore, students in 2012 had the option to wait and reuse their test scores in the following year. This policy change increased the attractiveness of dropping out so as to switch programs, as well as the value of the outside option of not enrolling in 2012, for students applying in the 2012 cycle. As we find that the expansion of the platform led to *increased* enrollment and graduation rates, our results are unlikely to be driven by this policy change.

[12]These weights typically vary depending on the type of coursework the program offers, with more weight on math and science when programs have more STEM coursework, and less weight on math and science when the program provides more qualitative coursework.

tions consist of a rank ordered list of *programs*, where a program is a narrow field of study (or major) at a specific campus and university. Of the 130,526 students who were eligible to apply in 2011, 63% submitted a rank-ordered list.[13] Table 2.1 and Figure 2.2 present more details regarding the number of test takers, eligible applicants, submitted applications and final assignments.

Throughout the application process, students have access to the following public information: the number of slots that each program offers, how the program *weighs* the test scores of applicants, their eligibility to apply to a given program, their personal weighted score if they were to apply to a given program, and the weighted score of the last admitted student in previous years for every program on the platform. Program-specific eligibility requirements may include minimum scores and minimum average weighted scores depending on each program.[14] Figure 2.1 shows changes in the score of the lowest-scoring admitted student at each G25 program. Overall, cutoffs have considerable persistence. This is especially true when we compare 2010 to 2011, where the correlation between cutoffs is 0.96 (see left panel). Nonetheless, the right panel in this figure shows that there is non-negligible movement in the cutoffs from 2011 to 2012, especially for lower selectivity options. More specifically, while the cutoffs rarely fluctuated over 25 points between 2010 and 2011 (2% of the time), in 2012, as new G8 options were added, cutoffs fell for many G25 programs, especially those with low selectivity. Of G25 programs with 2011 cutoffs under 600pts, 30% saw drops of over 25pts, and 9% of these saw drops of over 50pts. Higher selectivity options were much less affected although small negative changes were common.

Students Are Assigned Seats and Waitlists Are Formed

After submitting their ordered lists, students are assigned to higher education programs following the college-proposing deferred acceptance (DA) algorithm of Gale and Shapley

[13] The maximum number of ranked options increased from 8 to 10 during the period under study. Table 2.1 shows very few students utilized the 8 or 10 slots, and very rarely were students assigned to these low-ranked options (approximately 1%).

[14] The two most selective universities have an additional requirement that makes programs ranked below the fourth place ineligible. See Lafortune et al. (2018) for a description of how this feature can potentially affect student applications.

(1962). This process is discussed in detail in Rios et al. (2021).[15] Programs' preferences for applicants are given by their corresponding weighted scores after filtering out students that do not meet the stated program-specific requirements. We have confirmed that the college-proposing DA algorithm coincides exactly with the observed allocations in the years 2010-2012. 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 into a program.

Applicants may be waitlisted in zero, one, or multiple programs. Students are automatically placed on waitlists at all programs that they preferred to their assigned option, according to their submitted list. When a student is assigned to an option, the student's applications to programs that the student ranked lower to the assigned option are discarded.

Enrollment Decisions On and Off Platform Are Made

Students that receive an acceptance offer have the chance to enroll in that program. If they decide to do so, they pay the corresponding matriculation fees to secure a spot in the program. There is no punishment or cost for not enrolling in a program. After the initial enrollment process ends, waitlists are processed independently by each institution in a decentralized manner.

In addition to the options offered on the centralized platform, students can also apply directly to any number of off-platform university programs as well as a variety of less-selective technical and professional institutes. The decentralized admissions process has varied deadlines and potentially different application requirements, but the vast majority require the college entrance exam. While not coordinated, admissions processes at universities tend to track the timeline of G25 universities with a lag, so that most off-platform offers are finalized after students and programs learn on-platform match assignments. Most of the broader non-university higher education system has rolling admissions until the beginning of classes.

[15] The Chilean process differs from the textbook deferred-acceptance algorithm in its treatment of students with identical scores. If two or more students have identical scores at a program, and the program would otherwise be forced to strictly rank them in order not to propose to more students than its capacity in some round of the DA algorithm, in the Chilean process it proposes to all such students. Thus, in cases of ties, it is possible for programs to exceed their capacities.

Summary of Application, Enrollment and Aftermarket

Figure 2.2 describes the timing of the admissions process, the aftermarket and enrollment. Students take the PSU in December and receive their test results in early January. Given information on test scores, students can calculate the financial aid and loan packages that are available to them at each program. Equipped with this information, applicants have approximately one week to submit a rank-ordered list. Programs provide weights that describe their priorities, their desired number of slots, and, if they choose to do so, a number of extra slots to deal with offers being declined. Applications are processed using a DA algorithm, and assignments are communicated to students. At this point, the aftermarket begins: students decide to accept or reject offers, and programs begin calling waitlisted applicants. Most off-platform enrollment decisions occur at this time as well. Once all enrollment and waitlist-enrollment decisions have been made and the incoming cohort for each program has been determined, each program begins its regular academic year.

2.2.4 Waitlists and Evidence of Aftermarket Frictions

In this subsection we document the overall prevalence of waitlists and show evidence of aftermarket frictions. We see that the system takes steps to reduce the scope of waitlists. In particular, to partially accommodate the possibility of declined offers, the mechanism elicits from each program two capacity measures: a "true" slot count and a number of "extra" seats. The program's capacity in the DA algorithm is the sum of these numbers. Thus, programs may supply excess slots in anticipation of some students declining their offers. An on-platform program may contact students on its waitlist only in the event that enrollment would otherwise fall below its "true" capacity. Therefore, programs which use "extra" seats reduce their reliance on waitlists but face the risk of more acceptances than their "true" capacity.

In practice, programs choose fewer excess slots than needed to achieve full enrollment via initial offers. Figure 2.3 shows that despite the presence of excess slots, students have a positive probability of receiving waitlisted offers in the aftermarket. Moreover, ex post some programs exceed their "true" capacities while others undershoot. This pattern may be explained by financial constraints that put an upper bound on "worst-case" enrollment. The resulting effect is that on average, enrollment is 12% lower than the desired seats originally posted, despite the use of excess slots. Of the programs that had excess demand beyond the desired and extra slots (88%), 70% of these ended up matriculating students from their waitlists. Overall in both 2010 and 2011, approximately 4000 students matriculated through waitlists, representing 8% of all the matriculation on the centralized platform in those years.[16] Figure 2.3 describes the distribution of "true" and "excess" seats as well as programs' yield. We observe heterogeneity in the use of excess seats, with some programs offering none and some offering double their true capacity. Importantly, unlike in the U.S. context in which people apply to universities, the typical program is small and hence faces nontrivial "sampling" uncertainty in the number of accepted offers.

If a program contacts students on its waitlist, the typical approach is to go through the waitlist in order and inform (through a phone call) each waitlisted applicant that they now have an available slot. Students may accept or decline any waitlist offers that they receive. 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 frictions: students may be called by multiple waitlisted programs; there may be communication issues (e.g. wrong numbers may be dialed); students may renege on a waitlisted offer after verbally accepting it but before formally enrolling; the waitlist process operates in real time and terminates at a fixed date, potentially before the market "clears".

Because the use of excess seats means that some programs do not contact their waitlists, one might expect a discontinuity in enrollment chances on average at programs' cutoffs. However, in the absence of frictions one would expect no discontinuity in enrollment probabilities at the initial cutoff among those programs that do contact their waitlists.[17] In Figure 2.4 we present a case study that shows a clear discontinuity in admissions, and then a waitlist which exhibits "gaps".

[16]These numbers do not include students who are admitted off the waitlist through a small government program called *Beca Excellencia Academica* (BEA) that provides additional slots. These waitlist matriculations account for an additional 400 students who get in off the waitlist.

[17]If a college-proposing deferred acceptance procedure "pauses" at the initial assignment, some students decline offers, and then the procedure resumes with programs that are underenrolled proposing to students at the top of their waitlists, there should be no discontinuity in enrollment probabilities among programs that make waitlist offers. We provide a formal argument in section B.3.

The waitlist process is not explicitly regulated by the platform beyond the limit on total slots. Hence it is difficult to get direct data on the way that waitlists are processed. To understand how this process works, we conducted interviews with a handful of officials who administer the recruiting process and, in particular, that supervise the processing of waitlists. Two transcripts are presented below as an example. One administrator who works at a highly-selective program indicated that at their program they don't always go to the waitlist, but when they do, they provide callers three times more numbers than they need to recruit, expecting many to not answer and some to decline. In addition, administrators indicated that they typically expect to conduct multiple rounds of calls, as some students who accept verbally over the phone might not appear to matriculate the next day. The entire process is done quickly, with short deadlines for students to respond to offers, as programs scramble to sign up students before they commit to other options.

> *Each university clears their waitlist with call-centers... we informed them that they got off the waitlist and asked them if they would like to enroll. If they said yes, we would ask them to come early next morning. If they did not arrive, we would try to contact them again. If someone did not want to enroll or did not pick up the phone, we would call the next one... If two students were called and both decided to enroll we would let both of them in... for a single slot in the waitlist, we would call 3 students and then potentially discard some... it is not a rule, it is discretionary.*
> *-Admissions Officer*

> *When it comes to waitlists, DEMRE does nothing: each university clears their waitlist with call-centers... we used to call students and ask them if they had enrolled in some other place. Regardless of the answer, we informed them that they got off the waitlist and asked them if they would like to enroll. If they said yes, we would ask them to come early next morning. If they did not arrive, we would try to contact them again. If someone did not want to enroll or did not pick up the phone, we would call the next one... If two students were called and both decided to enroll we would let both of them in... for a single slot in the*

waitlist, we would call 3 students and then potentially discard some... it is not a rule, it is discretionary... If we were to fill 10 slots and the first 10 people we called said "yes", we would still call 15, but if some said "no" we would go even further down and keep calling. In terms of logistics, we usually had like 3 rounds where we called waitlisted applicants until we filled the list... sometimes, people did not have money to enroll again, so they lost their seats... If 15 people showed up for 10 waitlist slots that we had to fill, we enrolled all 15, otherwise they could file a complaint with the Ministry of Education and we could get sued. That's why, when we had to call waitlisted applicants, there is someone with a high rank that gives you the list of whom to call. She told me to call the first 5, and if they did not pick up by noon I had to inform her... In extreme cases, when we did not fill the slots, we would grant enrollment to some low-rank students that begged for admission, as most of the other students that ranked above them did not show any interest in enrolling. I even know of some universities that eventually give up and allow waitlist enrollments on a first-come-first-serve basis.

- Admissions Officer

While this qualitative interview evidence is not necessarily representative of the experiences at all programs, "gaps" in waitlist enrollment and discontinuities in matriculation probabilities at programs' cutoffs are typical of programs that admit students from waitlists, suggesting that the significant aftermarket frictions described here exist more broadly.

2.3 The Expansion of the Platform

When the G8 universities joined the centralized platform, the number of options available to students increased by over 30% and the number of slots increased by almost 50%.[18] This was an unparalleled change in the supply side of the platform.[19]

[18]We did not see any systematic changes in the number of seats within G25 programs in anticipation of platform expansion. Between 2011 and 2012, roughly 40% of programs kept the same number of true seats + extra seats, 20% decreased capacities, and 40% increased capacities. These changes are similar to those that that took place between 2010 and 2011. A possible explanation is that programs failed to anticipate platform-expansion-induced changes in demand when posting extra seats.

[19]Other preceding policy changes, such as making the PSU tests free for applicants, had important impacts on the number of students applying through the platform, but no other policy had a similar impact on the

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 the G8 options means that G8 placements and enrollment in G8 programs are now counted as on-platform placements and matriculations. A less immediate consequence is that students that were admitted into G25 options increased their enrollment *rate* in G25 institutions after the policy. As depicted in Figure 2.5 and summarized in Table 2.1, when compared to 2011, students placed in G25 programs were around 7 percentage points ($\sim$ 10%) more likely to enroll in their assigned programs in 2012. This effect is driven by students' ability to express preferences for G8 programs and their inability to enroll in G8 programs if assigned to a G25 option, unless they move off a waitlist ($\sim$ 1% of G25 admits). Prior to 2012, students who had been admitted to G25 programs could decline their on-platform offers in favor of an off-platform offer from a G8 program.

We find that students who are initially placed in G25 programs are more likely to enroll in their initial placement over a wide range of PSU scores and program selectivities. This fact is shown in Figure 2.6, where sample probabilities are plotted in 70-point bins for all the years in our data. Overall, we observe a significant average increase in enrollment rates for G25 admits with scores below 750 points. As test scores are adjusted to resemble a normal distribution with a mean of 500 and a standard deviation of 110 points (see Table 2.1), 750 points is approximately the 99th percentile of the score distribution. Thus, the policy increased the enrollment yield for G25 admits across the score distribution.

Figure 2.7 shows the results of this last exercise broken out by gender and high school type. In all cases, there are significant differences between enrollment rates in 2012 and previous years, but the impacts on enrollment probabilities are larger for low-scoring private school applicants. Intuitively, the enrollment rate of private-school students should be more affected by the policy if they were more likely to renege on their platform offers and enroll in private, off-platform institutions. In section 2.5 we show evidence of this behavior, estimating higher average valuations for G8 options from private school students as well as a higher probability that off-platform G8 options extend offers to private school students conditional

number of options from which students could choose. Other policies that expanded access to higher education in Chile are summarized in Figure B.1.

on test scores.

The observed increase in conditional enrollment rates should foster efficiency in the system. Chains of vacancies left by students who renege on their offers may be filled in the aftermarket. When fewer students renege, there is less work for the aftermarket to do and hence the presence of frictions may matter less for students' outcomes.[20]

If students are more likely to leave programs that they consider less desirable, then an additional measure of inefficiency is the rate at which students drop out of the system once enrolled. If match quality increases, we should expect to see fewer students dropping out over time and more students graduating. We investigate these outcomes in the following event study. Table 2.2 shows estimated changes, controlling for test scores and student-type fixed effects, from 2010 to 2015 in platform admission rates, enrollment rates conditional on G25 admission, and 1-year dropout and 7-year graduation rates conditional on G25 enrollment. The coefficients $\{\hat{\beta}_t\}_{t=2010}^{2015}$ are OLS estimates from the following specification:

$$Y_{ist} = \alpha + \sum_{t=2010}^{2015} \beta_t \mathbf{1}[cohort_{is} = t], \qquad s = \{1, 2, 3, 4\},$$

where Y_{ist} denotes the outcome (admission, enrollment, dropout, graduation) of student i, of sex-school type s (1 $\rightarrow$Private-Male, 2 $\rightarrow$Public-Male, 3 $\rightarrow$Private-Female, 4 $\rightarrow$Public-Female), in application-cohort t. The year 2011 is excluded, so that all outcomes are relative to this year. The coefficients β_t correspond to the conditional average differences explained by the indicators $\mathbf{1}[year = t]$, which equal 1 for application year/cohort t and 0 otherwise. The estimates are reported separately for each outcome, year and student type, and 95% confidence intervals are based on heteroskedasticity-robust standard errors.

We find that platform admission rates jump by about 9 percentage points. Enrollment rates increase by about 7 percentage points, and freshman dropout rates fall by roughly

[20] To quantify the externalities on other applicants induced by students' decisions to decline on-platform placements, one might ask the following (infeasible) counterfactual question: if applicants that were to ex-post renege on their assignments were ex-ante excluded from the platform, what would happen to the matches of other applicants? Figure B.2 depicts this counterfactual exercise in which students who receive and ex-post decline on-platform placements are removed from the match ex-ante, for each year. Prior to 2011, removing such students would cause at least 27% of students to receive a placement that they ranked ahead of what they received in the data. This fraction of match-improvements falls to 20% in 2012 following the expansion of the platform.

1.1 percentage point in 2012. These averages mask substantial heterogeneity: private school students increase their admission and (G25) enrollment probabilities more than public school students, but the latter, especially public school women, exhibit larger decreases in their (G25) dropout rates.[21] We find an increase of about 2.5 percentage points, on average, in seven-year graduation conditional on G25 enrollment.

In Table B.3 and Table B.4 we show that these results are robust under alternative specifications. We focus on the G25 programs here in order to isolate the effects of platform expansion on match quality within a fixed set of programs. Our results indicate that match quality within these programs may have improved as a result of platform expansion. Considering all G33 programs, we find similar patterns in enrollment and graduation (applications and admissions offers are not observed pre-2012 at G8 programs). An event study indicates that overall enrollment among test-takers increased by roughly 1.1 percentage points between 2011 and 2012, with G33 graduation increasing by 2.4 percentage points.

2.4 Model

2.4.1 Theoretical Model

In order to estimate the welfare impacts of the policy change and assess which programs' participation decisions had the largest impacts, we estimate a model of students' on-platform applications, aftermarket frictions, enrollment decisions, and human capital outcomes. Our goal is to provide a tractable framework that uses variation in students' choices around the policy change to identify key frictions, and their impacts, in the partially decentralized market.

Our model has four stages, which we describe in detail below:

1. Students submit on-platform applications.

2. The DA procedure runs, and students receive initial placements and waitlist positions.

[21] The 1.1 point reduction in freshman dropout rates accounts for over a 10% fall in overall dropout by the end of the first year of college. Public-school students and low-scoring private school students, especially women, mostly drive the reduction in first-year dropout rates. Retention rates are stable for high-scoring students.

3. The aftermarket takes place. Students receive off-platform and waitlist offers and make final enrollment decisions.

4. Production of human capital takes place. Students drop out or graduate from programs.

A market $t \in T = \{2010, 2011, 2012\}$ is an application cohort consisting of N_t students and a set of available programs $j \in J_t$. Within each cohort t, each student $i = 1, \ldots, N_t$ belongs to one observable group $g \in G$. If student i of group $g(i)$ in cohort $t(i)$ attends program j, he receives utility

$$u_{ij} = \delta_{jg(i)} + z_i \lambda^z_{g(i)} + w_{ij} \lambda^w_{g(i)} + x_j \eta^x_i + p_{ij} \lambda^p_{g(i)} + \epsilon_{ij}, \tag{2.1}$$

where δ_{jg} is a program-level mean utility term, z_i is a vector of student-level observables with coefficients λ^z_g which shift the value of all "inside options", w_{ij} are observed match-level terms with group-specific coefficients λ^w_g, and x_j are program characteristics for which students have a vector of unobserved multivariate-normally-distributed random tastes

$$\eta^x_i \sim N(0, \Sigma^{g(i)}).$$

The terms $\epsilon_{ij} \sim N(0, 1)$ are iid match-level preference shocks. p_{ij} is a match-level net price after accounting for government-provided and institution-specific scholarships available to i at program j in i's market $t(i)$, and is multiplied by group-specific coefficients $\lambda^p_{g(i)}$.

The set of inside options J_t consists of all G8 and G25 programs that operated in year t, and is partitioned into on- and off-platform programs. Let $J^{\text{on}}_t \subseteq J_t$ denote the set of on-platform programs in market t, and $J^{\text{off}}_t = J_t \setminus J^{\text{on}}_t$ the set of off-platform programs. In addition, there is an outside option, $J = 0$, whose value is given by the maximum of two components:

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

These outside option components are independently normally distributed. We have:

$$u_{i0}^0 \sim N(0, \sigma^2_{0,0,g_i})$$
$$u_{i0}^1 \sim N(z_i \gamma_{g(i)}, \sigma^2_{0,1,g_i}).$$

The first component, u_{i0}^0, is known at the time of applications, and represents the value of the best nonselective or noncollege alternative that is known before applications are due, such as entering the labor force. In contrast, the second outside option component, u_{i0}^1, is learned during the aftermarket, after the initial match takes place. This shock rationalizes the decision to apply to G33 programs but then decline all offers.

Programs outside the G25+G8 institutions were not on the platform during our sample period, and form part of the outside option. Non-G33 offers that realize after the on-platform match takes place belong to the second outside option component u_{i0}^1. Hence the second outside option consists of selective non-G33 programs, as well as shocks to (e.g.) entering the labor force that may realize after applications are due.

We impose the location normalization $\mathbb{E}(u_{i0}^0 = 0)$. However, we allow the mean of the second outside option to vary with all individual characteristics that enter utility, including year-by-group effects, as the quality of the best non-G33 offer may depend on test scores and other observables and may vary over time as the set of non-G33 options evolves. Fixing the variance of $\epsilon_{ij} \sim N(0,1)$ normalizes the scale of utility. Because our model includes program-by-group effects δ_g which subsume mean effects of program-level unobservables, the random coefficients η_i^x are mean zero without loss. The covariance matrix of random coefficients is unrestricted.

In practice, the groups are $G \equiv \{\text{male}, \text{female}\} \times \{\text{public/voucher school}, \text{private school}\}$, where the type of high school that the student attended is a proxy for SES in our context. Importantly, all preference parameters, including program effects δ and random-coefficient covariance matrices Σ, differ arbitrarily for each of four types $g \in G$. Thus low- and high-SES students need not agree on a vertical ranking of quality.

Individual-level variables z_i include a constant, i's math and verbal test scores, year indicators, indicators for urban location and current high-school enrollment (as opposed to older applicants applying for a second time), and government-provided scholarship amount. The scholarship amount, scholarship_i, is a known function of a (publicly-known) household SES index, may be used at any G33 institution and hence at any program in $J_{t(i)}$, and can be treated as a shifter of all inside options relative to the outside option because it does not vary across programs within a person.

Program characteristics x_j consist of measures of STEM and humanities course content. Observed match terms consist of a full set of interactions between individuals' math and humanities test scores and the STEM and humanities course content of each program, as well as an indicator for program and student in the same region of Chile—a coarse proxy for proximity between the student's home and the program—and interactions between an indicator for education major and a third-degree polynomial in student average test scores.[22]

To attend program j, student i in market $t(i)$ would pay

$$p_{ij} = \max\{0, \text{listprice}_{j,t(i)} - (\text{scholarship}_i + \text{discount}_{ij})\},$$

where listprice_{jt} is the program's publicly posted price in year t and discount_{ij} captures additional match-specific discounts and subsidies. While a variety of match-specific discounts existed (see Table B.2), we focus on the match-specific discounts provided to qualifying (high-scoring) students at programs j which participate in the BVP program in year $t(i)$, if $t(i)$ is a year in which BVP is available. We define

$$discount_{ij} = \max\{0, (\text{listprice}_{j,t(i)} - \text{scholarship}_i)\mathbf{1}(BVP_{j,t(i)})\},$$

where $BVP_{j,t} = 1$ if program j participated in BVP in year t, or did not formally participate but provided an equivalent scholarship from its own funds.[23]

Program fixed effects absorb variation in listprice_j, and we include scholarship_i in z_i, so that λ_g^p is isolating the effect of price variation caused by the introduction of the BVP

[22] These interaction terms help us isolate "regression-discontinuity" variation in applicants' choices induced by the BVP policy. The next section provides a discussion of the research design.

[23] Every G33 program in the education field provided such a scholarship by 2012. One G8 institution did not participate in BVP but chose to provide an identical scholarship program on its own, presumably to avoid constraints on its ability to admit low-scoring students that would have bound if it had participated.

scholarship, at certain programs, for those group-g students who qualify.

As an outcome, we consider a policy-relevant near-"on-time" graduation measure, graduation within seven years from the program in which a student enrolls. In the event student i of group $g(i)$ in cohort $t(i)$ enrolls in program j, he graduates if his potential human capital h_{ij} is greater than zero. h_{ij} is distributed according to

$$h_{ij} = \overline{\beta}_{j,g(i)} + z_i\beta^z_{g(i)} + w_{ij}\beta^w_{g(i)} + p_{ij}\beta^p_{g(i)} + \nu_{ij},$$

where $\overline{\beta}_{jg}$ are program effects, not necessarily equal to those that enter utility. The error term ν_{ij} is distributed independently across individuals, jointly normally with $(\epsilon_{ij}, x_{j,t(i)}\eta^x_i)$, so that the conditional distribution is given by:

$$\nu_{ij}|\epsilon_{ij}, \eta^x_i \sim N(\rho_{g(i)}(x_{j,t(i)}\eta^x_i + \epsilon_{ij}), 1)$$

for $\rho_g \in \mathbb{R}$. Importantly, our specification allows every term that enters preferences to enter the outcome production function. This specification allows, but does not impose, perfect alignment between preferences and production.[24] We allow match effects on observed determinants of preferences w_{ijt}, p_{ijt} as well as on unobserved determinants of preferences via the correlation terms ρ. These match effects are of interest because the human capital impacts of policies that improve students' welfare will depend on the degree of alignment between preferences and human capital production.

In the first stage of the game, students learn their preferences for all programs except u^1_{i0}, then submit rank-ordered application lists over on-platform programs to a centralized mechanism. Programs rank students according to an index of four test scores and high school GPA, with program-specific weights, which we denote $index_{ij}$.[25] Each program has a fixed number of slots. A college-proposing deferred acceptance procedure runs, producing initial

[24]This would occur when $\beta_{jg} = \delta_{jg}$, $\beta^z_g = \rho_g\lambda^z_g$, $\lambda^w_g = \rho_g\beta^w_g$, $\beta^p_g = \rho_g\lambda^p_g$, and $\rho_g > 0$. If these equalities hold, by grouping and rearranging terms we could rewrite our model as: $h_{ij} = u_{ij} + \tilde{\nu}_{ij}$, where $\tilde{\nu}_{ij} \sim N(0, 1)$ is a shock that is not predictable at the time of applications and enrollment decisions.

[25]In addition to the index formula, some programs have eligibility rules, such as a minimum score on a subset of the exams. In the DA algorithm in practice and in our simulations, applicants who are not eligible are dropped from the program's preference list.

placements.[26]

In addition to its assigned students, each program maintains a waitlist. All students who were eligible to apply to program j and applied to j but were not placed in j or in a program they preferred to j are waitlisted at j. Students may be on multiple waitlists. At the end of the procedure, students learn their initial placements and waitlist status.

We now consider the aftermarket, which we model as a college-proposing DA procedure with a friction. At the beginning of this stage, students learn their second outside option, u^1_{i0}.[27] Students receive offers from off-platform programs and from on-platform programs at which they are waitlisted, and may decline or provisionally accept them.[28] At the end of the process, students enroll in the program they most prefer among programs that have made them an aftermarket offer, their original match, and their outside options.

Off-platform programs $j \in J^{\text{off}}_t$ rank students according to $index_{ij}$—the formula they ultimately adopt when they join the platform—and have fixed capacities. On-platform programs j give maximum priority to students who received an initial placement at j, guaranteeing that a student who receives an initial placement at j can keep that placement if he desires to do so. They rank the remaining students according to their position on the relevant waitlists. If a student is not waitlisted at on-platform program j, he/she is not acceptable to j in the aftermarket.

Let

$$a^*_{ij} = v_{ij}\alpha_{g(i)} + \aleph_{ij}$$

measure program j's ability to contact student i in the aftermarket, where g denotes student

[26]Programs' $index_{ij}$ formulas admit the possibility of ties. In the Chilean process, in practice as well as in our simulations, if in round t of the DA algorithm a program's final proposal would be to some student i with score $index_{ij}$, it proposes to all students i' such that $index_{i'j} = index_{ij}$.

[27]Formally, in the first stage of the aftermarket DA procedure, the second outside option makes a proposal to each student. This offer provides utility u^1_{i0}.

[28]In the aftermarket DA, we assume that off-platform programs k drop students from their preference lists who prefer the first outside option, i.e. for whom $u^0_{i0} > u_{ik}$. That is, students must have been willing to apply to k ex-ante in order for k to propose. This does not affect the final allocation, but greatly reduces the number of iterations required. We have also estimated a model in which off-platform programs do not propose to students who prefer their initial placement. An interpretation of this alternate model is that students must apply to off-platform programs after learning their initial assignments. The results are unchanged.

type, α_g is a vector of type-specific coefficients, the covariates

$$v_{ij} = (1\{j \in G25 \cap \mathcal{J}^{\text{on}}_{t(i)}\}, 1\{j \in G8 \cap \mathcal{J}^{\text{on}}_{t(i)}\}, 1\{j \in G8 \cap \mathcal{J}^{\text{off}}_{t(i)}\}, \text{sameregion}_{ij})$$

consist of indicators for institution type and platform status,[29] as well as an indicator for program and student in the same region of Chile, and the shock term $\aleph_{ij}$ is distributed according to a standard normal distribution, independently across i and j.

Program j is **able to successfully contact student** i if $a^*_{ij} > 0$, if $j = 0$, or if $j > 0$ is i's assigned program in the match. Thus, outside options and the initial assigned program (if any) are always able to successfully contact i, but other programs require a positive draw to do so. If $a^*_{ij} < 0$ and program j is not an outside option or i's assigned program in the initial match, then i is dropped from j's aftermarket priority ordering.

We say that program j is **ex-post aftermarket-feasible** for student i if $index_{ij}$ is at least as high as the lowest value of $index_{ij}$ among students enrolling in j and, in the event j is on-platform, that i applied to j and was not placed in a program she prefers to j. Thus, if i is placed in j, or is waitlisted at j and has a sufficiently high score, or has a sufficiently high score when j is off-platform, then j is ex-post aftermarket-feasible for i. Outside options and the initial assigned program are always ex-post aftermarket feasible.

Program j is **available** to i if j is able to successfully contact i and, in addition, j is ex-post aftermarket-feasible for i. Let

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

denote the event that j is available to i.

Our assumptions imply that i enrolls at his most preferred program j at which $a_{ij} = 1$. The parameters α summarize the extent of aftermarket frictions in availability. These frictions may vary by program type, student type, and student location, as local students may have an advantage. For instance, given that at least some programs ask students to register in person, it may be easier for local students to do so. When the α parameters are small, programs need to make many calls to fill a given vacancy, and thus are likely to leave

[29] The terms $G25$ and $G8$ denote the set of programs belonging to G25 and G8 institutions, respectively.

gaps when they move down their waitlists.

2.4.2 Research Design

Reports and preferences

To infer preferences from reports, we assume that students truthfully report their preferences over the subset of on-platform programs that is relevant for them: those programs that are within reach and that they like at least as much as their favorite "safety" program, in a sense that we make precise in this section.

For each program, we define score bounds,

$$\overline{\pi}_{jt} > \pi_{jt} > \underline{\pi}_{jt},$$

where the "cutoff" value π_{jt} denotes the minimum value of $index_{ij}$ among students placed in program j in the initial match in year t. Say that a program j is

- **ex-ante clearly infeasible** for student i in market $t(i)$ if $index_{ij} < \underline{\pi}_{j,t(i)}$.
- **ex-ante marginal** for student i if $\underline{\pi}_{j,t(i)} \leq index_{ij} < \overline{\pi}_{j,t(i)}$.
- **ex-ante clearly feasible** for student i if $\overline{\pi}_{j,t(i)} \leq index_{ij}$.

Suppose student i's true preference ordering over J_t satisfies

$$u_{i1} > \ldots > u_{ik} > u_{i0}^0 > u_{ik+1} > \ldots > u_{i|J_t|}.$$

Let $\overline{u}_i^{\text{feas}}$ denote i's highest payoff among clearly feasible options:

$$\overline{u}_i^{\text{feas}} = \max\left\{u_{i0}^0, \max_{\{j \in J_t^{on} : \overline{\pi}_{j,t(i)} \leq index_{ij}\}} u_{ij}\right\}.$$

Let

$$J_i^{\text{relevant}} = \{j \in J_t^{on} : index_{ij} \geq \underline{\pi}_{j,t(i)} \text{ and } u_{ij} \geq \overline{u}_i^{\text{feas}}\}$$

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

Let ℓ_i denote i's report after dropping all programs that are ex-ante clearly infeasible and/or are ranked worse than some clearly-feasible program.

We maintain the following assumption, which states that rank-order lists are truthful within the relevant set:

Assumption 3. *For each person i, ℓ_i consists of all elements of $J_i^{relevant}$ in the true preference order.*

Assumption 3 allows students to omit programs that they disprefer to an ex-ante clearly feasible program, and places no restrictions on how or whether students rank ex-ante clearly infeasible options. It implies the following stability properties:

Property 1. *The initial (on-platform) assignment is the program-proposing stable assignment with respect to student preferences for on-platform programs and the first outside option component u_{i0}^0, with capacities equal to the total number of seats for on-platform programs except in cases of ties.*

Property 2. *The final (post-aftermarket) assignment is the program-proposing stable assignment with respect to student preferences, "true" capacities,*[30] *and the modified program priorities induced by dropping student i from program $j > 0$'s rank-order list in the aftermarket whenever $a_{ij}^* < 0$ and i's initial placement is not j.*

Proof. Say that j is ex-post match-feasible for i if $j \in J_{t(i)}^{\text{on}}$ and $index_{ij} \geq \pi_{j,t(i)}$. Because J_i^{relevant} contains the student's most-preferred match-feasible program whenever this program gives higher utility than u_{i0}^0, each student is necessarily matched to this program if and only if it gives greater utility than u_{i0}^0. The second part then holds by construction. □

Assumption 3 is used in estimation. Because reports are truthful within the applicant's relevant choice set, J_i^{relevant}, we may infer preferences for on-platform programs using standard discrete-choice arguments.

Moreover, although we do not fully specify the mapping from utilities to rank-order lists, Property 1 and Property 2 fully specify the mapping from true ordinal preferences,

[30] Programs must continue to propose to the students to whom they originally matched who have not declined their offer, even if this would lead them to exceed their "true" capacities.

availability realizations, capacities and priorities to enrollment: namely, the mapping induced by running the on-platform and aftermarket program-proposing DA procedures on these inputs. This mapping suffices for counterfactuals.

In practice, we choose a narrow bandwidth, $\overline{\pi}_{jt} - \pi_{jt} = \pi_{jt} - \underline{\pi}_{jt} = 25$ for all j, t. For large bandwidths, list length constraints could prevent students from listing all elements of the feasible set. Given our bandwidth specification, the case that the list length constraint could possibly bind is vanishingly rare in practice. Under our baseline bandwidth specification, the event that a student's relevant application is of maximum length but does not contain an ex-post match-feasible program occurs precisely once in our data. In section B.4 we give details on our choice of bandwidth and provide summary statistics on the relevant set under alternative bandwidths.

Using a subset of rank-ordered preference data in estimation, in addition to the restrictions implied by optimal enrollment decisions, allows us to estimate demand and learn substitution patterns without making the potentially strong assumption that applications are truthful. In particular, our estimation procedure allows students to omit irrelevant programs, consistent with theory and evidence on deferred acceptance procedures. We emphasize that a narrower bandwidth places fewer restrictions on preferences. Our strategy is related to the stability-based approach of Fack et al. (2019), and reduces to it as the score bounds $\overline{\pi}$ and $\underline{\pi}$ approach π. It is also related to Che et al. (2020), which uses an alternate approach to rule out payoff-relevant departures from truthful play.

Stable matching mechanisms, such as the (unconstrained) college-proposing DA algorithm, have optimal reports which are "dropping strategies" that may omit some programs but rank the listed programs truthfully (Kojima and Pathak, 2009), consistent with our assumptions. Constraints on list length may also lead applicants to drop some programs (Haeringer and Klijn, 2009). In principle, truthful reporting of preferences in college-proposing DA is approximately optimal in a large market (Azevedo and Budish, 2019), and is exactly optimal when there is a unique stable matching. In our setting in the years 2010-2012 there is either a unique stable matching or a vanishing share of students who receive different assignments under the college-optimal and student-optimal matches.[31] In

[31]Using the observed rank-order lists, we find a unique stable matching in 2010 and 2011. In 2012 there

practice, however, even if applicants cannot gain from non-truthful reporting, some applicants may omit programs that are out-of-reach or irrelevant (Fack et al., 2019; Artemov et al., 2020; Shorrer and Sóvágó, 2018; Hassidim et al., 2016). Larroucau and Rios (2018) provide evidence from the Chilean match that some students omit programs at which they have very low admissions chances. Our approach is consistent with this literature.

Willingness to pay

In order to obtain welfare estimates in dollars, it is crucial to estimate students' willingness to pay for programs. To do so, we exploit two features of the match-level price variation in our data: discontinuities as a function of students' test scores in the availability and size of program-specific scholarships, subsidies, and discounts, and year-to-year variation in the availability of these sources of program-specific funding.

Our variation comes from the introduction of the nationwide BVP scholarship program in 2011, which made scholarships available to high-scoring students at participating teacher-training programs. Our design is a difference-in-differences design exploiting this policy change, embedded in our structural model, which allows us to estimate a price coefficient jointly with other demand parameters.

BVP provides full scholarships to students with mean math+verbal test scores above 600 at participating programs. The set of participating programs includes all teacher-training programs at public institutions, as well as teaching programs at private institutions which chose to participate. Within the G33, every program that could have adopted BVP did so by 2012, with the exception of the relevant program at Universidad Andres Bello, a private G8 university. However, this university introduced a full institution-funded scholarship for students in the teaching program with scores above 600, exactly as if it had participated in BVP.[32]

are multiple stable matchings but they coincide for nearly all applicants—all but seven students—under the submitted rank-order lists. Moreover, under plausible estimates of "true preferences" we also find that the stable matching is unique with respect to true preferences with high probability, and when it is not unique a single-digit number of students would receive different assignments under the two extremal matches. We provide details in section B.4.

[32] BVP provided government funding for these scholarships, but required programs to restrict the number of seats available to students with scores below 500. Universidad Andres Bello did not impose additional constraints on low-scoring students.

Intuitively, high-scoring students are *eligible* for a full scholarship at teaching programs in the *post* years in which these programs are treated. In a two-way fixed effects specification, the coefficient on $\text{eligible}_i * \text{post}_{j,t(i)}$ would reveal the impact on demand of receiving such a scholarship.

In our empirical specification, programs differ in prior attractiveness to students, and students differ in the size of the discount that BVP provides because they receive government scholarships of varying sizes in the event that they do not qualify for BVP. Prior differences in mean utilities between teaching and non-teaching programs are absorbed by program-by-group effects δ_{jg}. We control for the government scholarship amount by including it as an element of z_i.[33] We include math and verbal test scores as an element of z_i as well. In addition, because teaching programs may be differentially (un)attractive to high-scoring students for preexisting reasons (indeed, the introduction of the BVP scholarship was motivated by a public perception that high-scoring students did not want to pursue teaching careers) we include in w_{ij} an interaction between student test scores and an indicator for the teaching major, as well as interactions between squared and cubed student test scores and this indicator.

In summary, our specification includes linear "controls" for test scores, program indicators which subsume an indicator for the teaching major, and a teaching major-specific polynomial in test scores. Modeling the impact of the "running variable" in this way allows us to better isolate the impact on choice probabilities of the variation induced by discontinuities in institutional aid at the test-score cutoff, in addition to the variation due to the policy change.

In addition to the BVP program, there existed other program-specific grants, discounts, and scholarships, which we describe in Table B.2. We have chosen to focus on BVP because of the variation introduced by the policy change, and because all of the match-specific discounts in teaching depend deterministically on test scores, in contrast to other majors in which some discounts and subsidies are discretionary and may depend on unobservables.

[33] Because students receive government scholarships as a function of their socioeconomic status (SES), and SES may enter preferences for programs through multiple channels, we would not wish to use variation in net price that is induced by SES to recover a demand elasticity.

Aftermarket frictions

Our strategy relies on the fact that the ex-post aftermarket-feasible set is observed. For on-platform programs, we observe the index of the lowest enrolled student. Because a program in college-proposing DA process does not make additional offers when its capacity is filled, this student represents the lowest score to whom it ever extends an offer. Consider the set of students who are waitlisted in j but have $index_{ij}$ greater than this cutoff value. As $Pr(a^*_{ij} > 0)$ approaches 1, the share of such students remaining in their original match decreases monotonically to zero. Thus the share of students with observables v who have ex-post feasible programs that they prefer to their original placement according to their application, but who enroll at their original placement, reveals the extent of frictions conditional on v.

In fact, we observe the value of a_{ij} in many cases. For instance, if i enrolls in an on-platform program j at which she was waitlisted, then $a_{ij} = 1$ and $a^*_{ij} > 0$. If there is another on-platform program, k, which i preferred to the on-platform program at which she enrolled, and which was ex-post feasible for i, then $a_{ik} = 0$ and $a^*_{ik} < 0$.

Our approach to off-platform programs is similar. A complication is that applicants' ranking of off-platform programs is unobserved. We exploit the panel structure of the data to identify the distribution of preferences for these programs. G8 programs' unobserved demand-relevant characteristics are identified from rank-order application data in 2012, when they participate in the platform. Pre-reform data then allows us to estimate frictions for off-platform programs.[34] We allow these frictions to differ by type. Discrimination in favor of high-SES applicants, for example, would enter our estimates as larger frictions for low-SES applicants at off-platform G8 programs.

Human capital production function

In the Chilean college match, otherwise-similar students are assigned to programs discontinuously as a function of exam scores. Many papers conduct regression discontinuity designs in Chile and other matching settings to recover local average treatment effects (LATEs) of

[34] We model off-platform programs as conducting admissions as they would if on platform, but with frictions that may differ from those of the on-platform waitlists, but this is not essential. The model could in principle be extended to allow other characteristics, including the student unobservables that are relevant for preferences and human-capital production, to enter off-platform programs' admissions decisions.

program assignment on student-level outcomes of interest such as graduation among the populations local to each discontinuity.[35] Our model implicitly uses this source of variation. In order to identify the distribution of graduation rates under counterfactuals that shift program assignments, however, an additional "choice shifter" is needed that is otherwise excluded from outcomes (Agarwal et al., 2020). In our paper, year-to-year variation in programs' cutoffs plays this role; an observably identical student faces a different choice set in 2010 than in 2012.

2.4.3 Estimation

We estimate the model using a Gibbs sampler, using the universe of data from 2010-2012. We choose diffuse priors, so that our estimates may be interpreted as approximate maximum-likelihood estimates. The Gibbs sampler is convenient for our setting, which involves a high-dimensional discrete unobservable–the latent choice set of each agent in the aftermarket—determined by realizations of a_{ij}. Our approach here builds on the procedure of McCulloch and Rossi (1994) to allow for partial rank-order data as well as constraints implied by the enrollment decision when there is a latent "availability set" from which agents must choose.

For each applicant, we observe the submitted relevant rank-order list ℓ_i, initial placement $\text{placement}_i \in \{0\} \cup J^{\text{on}}_{t(i)}$, enrollment outcome $enroll_i \in \{0\} \cup J_{t(i)}$, observed graduation outcome $graduate_{i,enroll_i}$ for the program in which i enrolls, and observed student-, program-, and match-level characteristics $\omega \equiv (v, w, x, z, p)$.

We augment our data with utilities, availability indices, and human capital indices. For each market $t \in \{2010, 2011, 2012\}$, we construct $u_i \in \mathbb{R}^{|J_t|}$, $h_{i,enroll_i} \in \mathbb{R}$, and $a^*_i \in \mathbb{R}^{|J_t|}$, representing utility, human capital, and availability, respectively, for all students i. We choose initial values that are consistent with observed applications, enrollment decisions and graduation outcomes. In addition, we augment the data with random coefficients $\eta^u_i \in \mathbb{R}^L$ and outside-option utilities $(u^0_{i0}, u^1_{i0}) \in \mathbb{R}^2$ for each i.

Our sampler iterates through the following sequence of draws from conditional posteriors

[35] For instance, considering students who rank program j just above program k and have scores near the threshold for admission to j, one may obtain the average effect of attending j rather than k by comparing just-admitted to just-rejected students' outcomes.

of the parameters and latent variables:

1. For each market t, for each type $g \in G$, for each $i \in \{1, \dots, N_t\}$ of type g, draw:

$$u_{i0}^0, u_{i0}^1, u_i, a_i^*, h_i, \eta_i^x | \ell_i, \text{enroll}_i, \text{graduate}_i, \alpha_g, \beta_g, \gamma_g, \delta_g, \lambda_g, \overline{\beta}_g, \Sigma_g^{rc}, \rho_g$$

2. For each type $g \in G$, draw:

$$\begin{aligned}
\alpha_g &| \{a_i^*\}_{i \in g} \\
\overline{\beta}_g, \beta_g, \rho_g &| h, \{u_i\}_{i \in g}, \{\eta_i^x\}_{i \in g}, \delta_g, \lambda_g \\
\delta_g, \lambda_g &| h, \{u_i\}_{i \in g}, \{\eta_i^x\}_{i \in g}, \overline{\beta}_g, \beta_g, \rho_g \\
\gamma_g, \sigma_{0,g}^2 &| \{u_{i0}^0, u_{i0}^1\}_{i \in g} \\
\Sigma_g &| \{\eta_i^x\}_{i \in g}
\end{aligned}$$

Conditional draws of linear-index parameters $\alpha, \overline{\beta}, \beta, \eta, \delta, \lambda$ and variance/covariance parameters ρ, Σ^{rc} are standard (e.g. see McCulloch and Rossi (1994) and Agarwal and Somaini (2018)). Building on insights from McCulloch and Rossi (1994) and Agarwal and Somaini (2018), we show that, conditional on the data and on all other latent variables and parameters, each element of u, a^*, u_0^0, u_0^1, and h is distributed according to a truncated Normal distribution.

We provide details in section B.4.

2.5 Results and Counterfactual Simulations

2.5.1 Results

In this section we report selected model estimates. All parameters are estimated separately by student type (male - private school, male - public school, female - private school and female - public school). We focus on estimates of frictions and of selected human capital parameters. A full set of estimates is available in Table B.5, Table B.6 and Table B.7.

Table 2.3 shows aftermarket friction parameters. We find that on-platform frictions are high for all types. For non-local applicants to G25 programs, the probability of successful contact ranges from $\Phi(-1.49) \approx 7\%$ for private-school men to $\Phi(-0.86) \approx 19\%$ for public-school women. Local applicants have somewhat higher chances, ranging from $\Phi(-1.486 + 0.151) \approx 9\%$ for private-school men to to 22% for public-school women. At G8 platforms in 2012, the probability of successful contact is similar across types, ranging from roughly 22% to 27% for non-local students.

In contrast, off-platform admissions chances exhibit large differences by type. For students whose scores are above the cutoff, successful contact rates are 70% for non-local private school men, as compared to 22% for non-local public school women.[36] Thus, conditional on exam scores, students from private schools are substantially more likely to have the option to attend off-platform G8 programs.

In Figure B.3 we show the distribution of program mean utility terms (δ) by type. The results indicate that private-school students systematically exhibit stronger preferences for G8 programs, relative to G25 programs, than do students from public schools. Thus private-school students' greater probability of enrolling in G8 programs arises from stronger preferences as well as greater admissions chances.

Table B.7 shows production function parameters. Students with higher math scores are nontrivially more likely to graduate at all programs, but we find much smaller impacts of verbal scores. In addition, we find positive observable "match" effects on the interaction of STEM coursework and math test scores. For public-school students the symmetric 95% posterior probability intervals do not cover zero; effects are more noisily estimated for private-school students. Moreover, we find positive match effects on unobservables. High values of match utility shocks positively predict on-time graduation, significantly so for public-school students.

[36]Off-platform programs may devote more resources to finding and contacting applicants. One possible benefit of platform expansion, not modeled here, is that programs may be able to save costs on this margin.

2.5.2 Model Fit

Before presenting the main results, we show that the estimated model fits the distribution of scores within each program well within each year, and closely matches the observed impacts of the BVP scholarship as well as an event study of enrollment and graduation rates. Our estimation procedure did not specifically target these moments, but we believe they are important. Matching the scores of enrolled students is relevant in our context because the test scores of students enrolled in a given program is a key measure of the popularity of the program. Matching the "DiD" and "regression discontinuity" impacts of the BVP policy suggests that our model is exploiting variation that is appropriate for estimating willingness to pay.

To simulate applications, placements, and enrollment, we draw from the posterior distribution of parameters at every 200th iteration, after throwing out initial burn-in draws. Unlike in our counterfactual simulations in the following section, we do not condition on the latent utility values that were drawn during the MCMC procedure. Rather, we discard our data on agents' applications, enrollment, and other endogenous outcomes, draw utilities and availability shocks from their distributions conditional on the parameters that have been drawn, and use these values to simulate the initial match, aftermarket, final enrollment and graduation patterns. We provide details in section B.4.

Figure 8 shows mean math and verbal scores of students enrolled in each program. The X-axis shows observed values, while the Y-axis shows mean model-predicted values. For display purposes we omit programs with fewer than 50 seats. Large (>100 seat) programs are shown in red. Despite the large number of programs in the data, the model fits this measure well.

Figure 2.9 shows changes in the probability of enrollment in teaching majors between 2010 and 2011 as a function of mean math and verbal test scores. This variation around the introduction of the BVP program is our key source of price variation.

Our model matches the 6-percentage-point increase at the 600-point cutoff, at which students received a full scholarship in the "post" period. Moveover, it fits the change in enrollment well, away from the cutoff, for students with scores above 500.

Finally, Table 2.4 shows regressions of key outcomes—admission to some G33 program, graduation within seven years from some G33 program, and graduation within seven years conditional on having enrolled in a G33 program—on controls and year indicators among all students eligible to apply to on-platform programs. We show results from the data (first column in each pair) and from model simulations (second column in each pair). We find a close fit. For instance, the data show a 2.4 point increase in overall G33 graduation within seven years among the cohort entering in 2012, relative to the 2011 cohort, conditional on our set of controls. In the model, the corresponding value is 2.5 percentage points.

2.5.3 Impacts of Platform Expansion

Table 2.5 displays the impact of platform expansion on welfare, probability of enrolling in an inside-option program, and probability of near-on-time (seven-year) graduation conditional on enrollment.[37] All counterfactuals are conducted in 2012. We focus on the comparison of model-predicted impacts in 2012, with all inside-option programs on the platform, to an "as-if 2011" counterfactual, in which the population is as in 2012 but the G8 institutions are excluded from the platform. To provide context, we also evaluate the impacts of a "No Frictions" counterfactual in which all inside options are on platform and $a^*_{ij} > 0$ for all students and programs. In this counterfactual, each program's capacity is the maximum of its realized enrollment in 2012 and the number of "true" seats. Thus we do not reduce enrollment in cases in which the program exceeded its "true" capacity. We treat the "No Frictions" counterfactual as a benchmark, and report differences in outcomes, relative to this benchmark, under the other counterfactuals.

Panel A of Table 2.5 shows welfare in units of 1 million Chilean Pesos. We find that a frictionless admissions process would produce mean welfare equivalent to 2.601 million pesos relative to the complete unavailability of all G33 programs. Estimated welfare is larger, in these units, for private-school households because we estimate a price (BVP scholarship) coefficient that is closer to zero for these households relative to other terms; this need not reflect social weights. Relative to this benchmark, the 2012 baseline gives households an

[37]The typical program length is six years. Some medical degrees in Chile have a duration longer than six years but represent a small fraction of students.

average loss equivalent to 0.031 million pesos. The loss from excluding the G8, 0.294 million pesos, is an order of magnitude larger.

Panel B shows impacts on the probability of enrolling in any inside-option program. Private-school students are much more likely to enroll, with roughly 67% attending an inside option, relative to 35-39% of public-school students. We find that excluding the G8 would lead to large drops in enrollment, but that the baseline comes within a percentage point of the frictionless upper bound.

Finally, panel C shows impacts on seven-year graduation rates conditional on enrollment in a G33 program. These are larger for women and, conditional on gender, for private-school students. Excluding the G8 would lead to a third of a percentage point reduction in graduation rates of enrolled students, relative to the case of no aftermarket frictions. In contrast, at baseline graduation rates are similar to those of the frictionless case.

2.5.4 Impacts of Aftermarket Frictions

The results of Table 2.5 suggest that the interaction of frictions and programs' nonparticipation produces welfare losses. We now explore the role of frictions in detail. In Figure 2.10, we plot welfare, enrollment rates, graduation rates conditional on enrollment, and welfare by type as all missed-contact probabilities $pr(a^*_{ij} = 0)$ are multiplied by a factor $(1 - p)$ for $p \in [0, 1]$. We conduct this exercise with all programs on platform, as well as when the G8 is excluded.

Figure 2.11 shows the results of the same exercise differentiating by type of student. The results indicate that welfare increases monotonically as frictions are reduced, both with all programs on-platform and when the G8 is excluded. For students other than private-school men, frictions and platform status interact so that the marginal gains from friction reduction are larger when the G8 is excluded. For male students from private schools, in contrast, impacts of friction reductions are more muted. Intuitively, these students benefit from the lower standards at off-platform programs when women and public-school students are subject to larger frictions, and this benefit counterbalances some of the direct cost of frictions.

2.5.5 Which Programs are Most Important to Include?

Given the estimated parameters, we computed the average welfare loss of removing programs from the platform. We sort programs by selectivity, as measured by mean math+verbal test scores, and divide them into ten equal-sized bins by realized enrollment. We then evaluate the impacts of dropping these programs, one decile at a time, relative to the baseline setting in which all programs are on-platform. We present the results from least to highest selectivity.

Results are shown in Figure 2.12. We show that the utility loss is highest if the programs in the top decile of selectivity are removed. Intuitively, when the most elite programs on the platform are absent, students who would have placed in them instead occupy places in lower-ranked programs, leading to the longest chains of displacement of other students.

Losses are also large, although not as large for the most elite programs, when the least-selective decile is dropped. Including these programs is valuable for a different reason. This decile is the most likely to have vacancies which in turn are less likely to be filled by any student when the programs are off-platform.

Heterogeneous Impacts

We now turn to heterogeneity across and within types. We focus on the main counterfactual of removing the G8 from the platform. Figure 2.13 highlights welfare gains in different dimensions.The first set of bars shows that utility gains from including the G8 are concentrated among students who are not private-school males.

Excluding G8 programs from the platform in 2012 would result in a decrease in welfare of roughly .3m Chilean Pesos for female and public-school male students. For female public-school students, this is roughly 12% of the gap between the 2012 baseline and the complete absence of G33 programs. In contrast, private school male students would experience a smaller loss. Our second set of bars suggests that public school students substantially increase their probability of being matched and enrolling in a higher education degree after the policy: in absence of G8 programs, an additional 4.0% of public school female and 3.4% of public school male applicants would not enroll in any G33 program. In contrast, excluding G8 programs would make roughly 0.5 to 0.8% of private school students choose their outside option. Impacts on graduation rates conditional on enrollment are smaller. Public school students' graduation rates would fall by 0.4 percentage points, while private school female students' graduation rates would fall by twice that amount.

Results on within-type heterogeneity in Figure B.4 show that the distribution of ex-ante welfare shifts to the right for public school students, with much more mass between the utility equivalent of 1.5m and 3m Chilean Pesos. In contrast, private school welfare impacts are heterogeneous, with some students gaining welfare but also more mass close to zero in 2012. Taken together, these estimates suggest that public school students benefit more in terms of the extensive margin of now being able to attend college, while impacts on private school students exhibit greater heterogeneity.

2.5.6 Impacts in Context

Table 2.5 shows impacts of additional counterfactuals relative to the no-frictions benchmark. In "Free G33", all programs are on the platform, and costs are set to zero for all applicants. The average student welfare gain, relative to the 2012 baseline, is roughly .76 million Chilean Pesos.[38] The welfare impact of platform expansion is roughly a quarter as large as the welfare impact of platform expansion together with free college—a much more expensive policy change—would be.[39]

Finally, we compute an assignment that maximizes the sum of students' utilities, as measured in Chilean Pesos, subject to programs' eligibility rules (such as requiring a simple average of 450 points on math and verbal scores) but otherwise ignoring programs' rankings of students. The utility gains from this counterfactual, which holds prices fixed, are slightly smaller than those of providing full scholarships. In addition, this counterfactual, as well as the "free G33" counterfactual, would lead to large gains in enrollment and graduation rates. A "free G33" policy leading to a 5.6 percentage point increase in enrollment relative to the frictionless benchmark, and a 2.6 point increase in the share of students graduating within seven years, while "Max Welfare" would increase enrollment by 1.5 percentage points and

[38] "Free G33" delivers roughly 0.73 million CLP higher welfare per capita than "No Frictions", and the 2012 Baseline scenario delivers 0.03 million CLP lower welfare than "No Frictions".

[39] ("Baseline" - "Exclude G8")/("Free G33"-"Exclude G8")= 0.26.

graduation by 1.7 points.

2.6 Conclusion

This paper studies the empirical relevance of the negative impacts on students that arise in a centralized assignment mechanism when there are off-platform options. When a desirable program is not on the centralized platform, applicants have no ability to communicate to the mechanism how they rank that option relative to other options. Some students may value off-platform options more than the placement that the platform gives them, leading them to decline their placement and creating vacancies in turn. Moreover, the absence of a particular program on the platform may distort the placements of other students, even if the students whose placements are affected would never enroll in the off-platform program. These displaced students be less satisfied with their assignment, and may be more likely to decline their placement, creating further vacancies. These vacancies can lead to an increased reliance on drawing students from waitlists in the aftermarket period.

Aftermarket frictions that generate even small difficulties in processing these waitlists—such as problems contacting or confirming enrollment with applicants —contribute to an assignment that unfairly "skips" some applicants whose scores qualify them for an offer of admission. Depending on the magnitude of the aftermarket frictions and the extent of the use of waitlists, off-platform options may have large impacts on the resulting assignment. To the extent that the quality of the match assigned is associated with real outcomes like retention and on-time graduation rates, off-platform options and aftermarket frictions can have important effects on these outcomes as well.

To study the empirical importance of off-platform options and aftermarket frictions, we use rich administrative data from the higher education system in Chile, one of the longest running centralized assignment systems in the world. We focus on a policy change in 2012 that expanded the supply of slots of the centralized platform by 40%. We first document the impacts on assignments and outcomes. Descriptive analyses and interviews with market participants motivate an empirical model of students' preferences, application decisions, matriculation, and graduation rates in the presence of waitlist and aftermarket frictions. We estimate our model using the universe of students' rank-ordered lists of on-platform options and their enrollment decisions at both on- and off-platform options. We find that the configuration of on- and off-platform options can have meaningful impacts on students' welfare, dropout and graduation in higher education. Counterfactual simulations indicate that platform expansion produced additional student welfare worth roughly 0.263 million Chilean Pesos per test-taker, or roughly $650 per person, and increased enrollment in G33 programs by three percentage points, while raising graduation rates conditional on enrollment.

A post-estimation decomposition shows that the lower-scoring students, women and underprivileged populations benefited the most from having more options on the centralized platform. Programs' absence from the platform redistributes welfare away from public-school students and women toward high-SES private-school men, while reducing total welfare. Counterfactual analysis reveals that welfare is most sensitive to the presence of the most desirable options, as the 10% most selective programs leaving the platform would generate 50% larger losses than removing the median 10% of seats.

We find that aftermarket frictions and off-platform programs interact so that the marginal cost of frictions on student welfare is smaller when all programs are on platform. Moreover, when programs are off platform, match quality decreases, and some students with high scores at waitlisted programs lose their positions to students with lower scores. Because our estimates indicate that scores and idiosyncratic "fit" both contribute to on-time graduation, these two channels lead to lower on-time graduation rates when some programs do not join the platform.

These results show that off-platform options can generate important costs which are relevant to policymakers seeking to implement a centralized assignment system. While we study higher education, the considerations highlighted in this paper are common in many practical settings. One example is urban education markets in developing countries, which typically have a large share of private providers. As more developing countries follow their richer counterparts in implementing centralized systems, policymakers should incorporate the consequences of off-platform options into market design, in the spirit of the broader agenda described in Pathak (2017).

We show that empirical analysis can be helpful to guide policy discussions and quantify key parameters that are needed to evaluate the potential costs of non-participation by different institutions. Our estimates provide a specific metric to evaluate the cost of losing each university on the platform, but our model and empirical strategy also highlight ways to quantify the costs of off-platform options in other settings and provide a route to informing policy regarding the costs of off-platform options.

In this paper we have abstracted from several important aspects of the higher education market when evaluating the benefits of platform expansion. These include the potential benefits of transparency about the process of assignment. One such benefit is that, in a centralized process in which programs rank applicants according to known functions of public information, it may be easier to communicate the rules to applicants. Recent controversies surrounding the admissions process at elite universities in the United States suggest that this margin could be important. We have also ignored the fixed costs of running an admissions office. These costs presumably would be lower when participating in a centralized platform. Finally we have abstracted from supply side considerations related to the incentives that individual providers have to join the platform, and from any effects that platform expansion has on competitive incentives. In our setting, interviews with university administrators suggest that off-platform universities preferred to join the platform, and the binding constraint on their participation was the platform's decision to allow them to enter. However, in other settings, programs may have strong screening motives, or may prefer nonparticipation because of restrictions imposed by the platform on the ranking of applicants and/or the timing of offers. We leave these topics for future research on how best to design markets in practice.

Table 2.1: Sample Descriptive Statistics 2010-2012

	Year 2010		**Year 2011**		**Year 2012**	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Panel A: Test Takers						
Male	0.47	0.50	0.48	0.50	0.47	0.50
Private HS	0.10	0.30	0.10	0.30	0.11	0.31
Metro Area	0.65	0.48	0.64	0.48	0.64	0.48
GPA	530	116	532	110	536	113
Math Score	501	111	501	111	504	111
Verbal Score	501	109	501	108	504	110
Platform App.	0.35	0.48	0.34	0.47	0.45	0.50
Observations	251634		250758		239368	
Panel B: G25 Applicants						
Score at top-ranked program	599	68.7	602	69.0	604	69.17
Ranked 1st Technology (G25)	0.27	0.44	0.28	0.45	0.28	0.45
Ranked 1st Medical Sciences (G25)	0.23	0.42	0.22	0.41	0.25	0.43
Ranked at least 3 programs	0.82	0.38	0.80	0.40	0.84	0.37
Ranked at least 7 programs	0.23	0.42	0.20	0.40	0.25	0.43
Assigned to 8th+ program	$\leq$0.01	-	$\leq$0.01	-	$\leq$0.01	-
Maximum List Length	8		8		10	
Enrolled in HS while applying	0.59	0.49	0.62	0.49	0.58	0.49
Observations	84556		81258		75729	
Panel C: G25 Admits						
G25 Enrollee	0.75	0.43	0.74	0.44	0.80	0.40
G8 Enrollee	0.05	0.22	0.06	0.23	0.01	0.08
Other/Unenrolled	0.20	0.40	0.21	0.41	0.20	0.40
Observations	67013		67803		64662	

Note: This table shows descriptive statistics of the administrative data from DEMRE, the agency that runs the centralized assignment mechanism in Chile. Panel A describes the population of test takers each year in our sample. Panel B presents descriptive statistics for applications that had a G25 option listed as a first preference. Panel C presents descriptive statistics for students admitted into G25 options.

Table 2.2: Event study outcomes by type: Admission, Enrollment, Dropout, Graduation

	Admission	Enrollment	Dropout	Graduation
Year 2010×Male×Private	-0.002	0.005	-0.004	0.007
	(0.006)	(0.007)	(0.005)	(0.010)
Year 2012×Male×Private	0.125***	0.113***	-0.014***	0.037***
	(0.005)	(0.007)	(0.005)	(0.010)
Year 2013×Male×Private	0.134***	0.134***	-0.007	
	(0.005)	(0.007)	(0.005)	
Year 2014×Male×Private	0.145***	0.134***	-0.012**	
	(0.005)	(0.007)	(0.005)	
Year 2015×Male×Private	0.130***	0.140***	-0.003	
	(0.005)	(0.006)	(0.005)	
Year 2010×Male×Public	-0.042***	0.012***	0.001	-0.003
	(0.003)	(0.003)	(0.003)	(0.004)
Year 2012×Male×Public	0.089***	0.056***	-0.017***	0.014***
	(0.002)	(0.003)	(0.003)	(0.004)
Year 2013×Male×Public	0.093***	0.076***	-0.008***	
	(0.002)	(0.003)	(0.003)	
Year 2014×Male×Public	0.097***	0.087***	-0.010***	
	(0.002)	(0.003)	(0.003)	
Year 2015×Male×Public	0.066***	0.087***	-0.006**	
	(0.002)	(0.003)	(0.003)	
Year 2010×Female×Private	-0.017**	0.004	0.008*	-0.023**
	(0.007)	(0.009)	(0.005)	(0.010)
Year 2012×Female×Private	0.121***	0.143***	-0.005	0.027***
	(0.006)	(0.008)	(0.004)	(0.010)
Year 2013×Female×Private	0.145***	0.168***	-0.003	
	(0.005)	(0.008)	(0.004)	
Year 2014×Female×Private	0.148***	0.176***	-0.005	
	(0.005)	(0.008)	(0.004)	
Year 2015×Female×Private	0.135***	0.168***	0.004	
	(0.005)	(0.008)	(0.005)	
Year 2010×Female×Public	-0.043***	0.020***	-0.008***	-0.002
	(0.003)	(0.004)	(0.003)	(0.005)
Year 2012×Female×Public	0.071***	0.061***	-0.027***	0.023***
	(0.003)	(0.004)	(0.003)	(0.005)
Year 2013×Female×Public	0.086***	0.100***	-0.019***	
	(0.003)	(0.004)	(0.003)	
Year 2014×Female×Public	0.098***	0.119***	-0.024***	
	(0.003)	(0.004)	(0.003)	
Year 2015×Female×Public	0.063***	0.116***	-0.018***	
	(0.003)	(0.004)	(0.003)	
Observations	606280	393193	318809	163531

Note: This table shows estimates of the average difference in each outcome, for each type of student, and for each year after 2009. The base year is 2011 and the base type is Female-Public. Admission refers to the probability of being assigned a seat in the platform; Enrollment refers to the probability of enrolling in a platform program conditional on being admitted in a G25 option; Dropout refers to the probability of not being enrolled in any option the year after enrolling in a G25 program; and Graduation refers to the probability of graduating within 7 years of enrolling in a G25 program. The estimating equation includes student covariates (GPA and test scores) and student-type fixed effects. These estimated coefficients are not reported in the table. The results on graduation rates are constrained to years before 2013 because we do not have data after 2019. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 2.3: Selected Estimates

Parameters	Male Private	Male Public	Female Private	Female Public
Aftermarket frictions (α)				
G25	-1.551	-0.88	-1.349	-0.84
	(0.03)	(0.014)	(0.035)	(0.014)
G8 On	-0.827	-0.616	-0.739	-0.595
	(0.042)	(0.037)	(0.045)	(0.032)
G8 Off	0.255	-0.435	0.183	-0.739
	(0.024)	(0.016)	(0.021)	(0.01)
Local	0.232	-0.046	0.187	0.041
	(0.027)	(0.014)	(0.029)	(0.01)

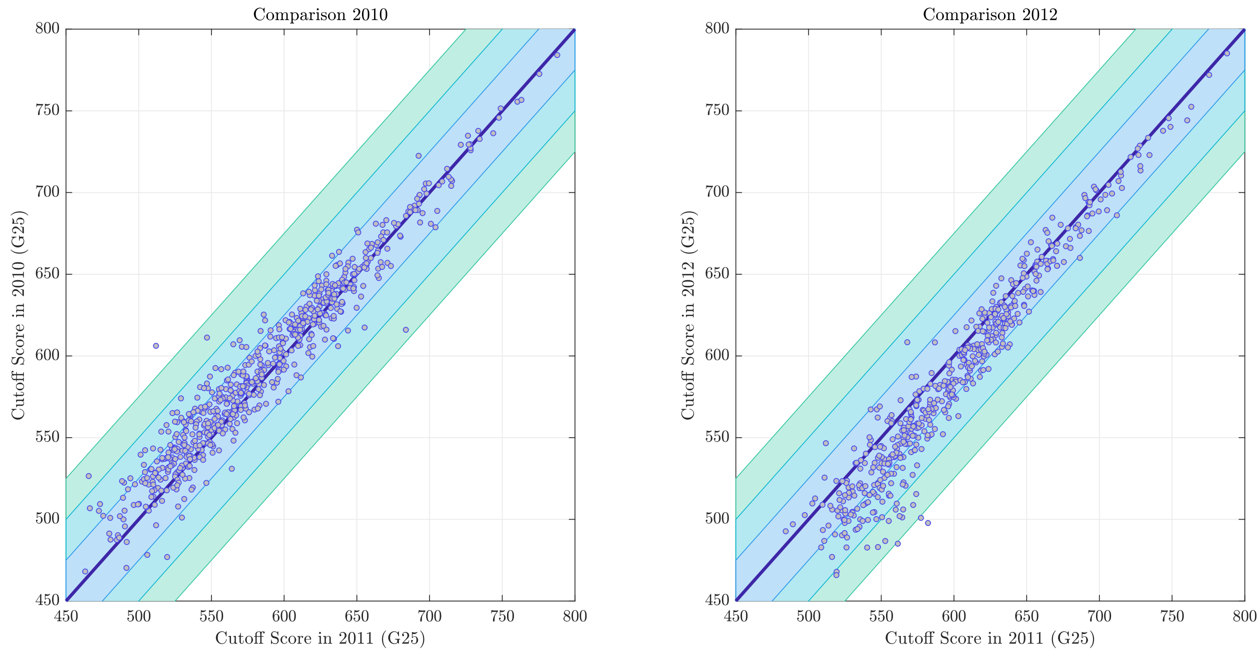
Note: Preference parameters were estimated via Gibbs sampling and include program fixed effects. The number of observations used for the estimation are 484549 and the number of options are 1334 over three years.

Table 2.4: Event study: G33 Enrollment and Graduation

	Data Enroll G33	Model Enroll G33	Data Graduate G33	Model Graduate G33	Data Graduate | Enroll G33	Model Graduate | Enroll G33
				All		
Constant	0.186	0.162	0.067	0.05	0.354	0.359
	(0.0012)	(0.0011)	(0.0011)	(0.0011)	(0.0026)	(0.0035)
Math	0.215	0.201	0.116	0.123	0.015	0.055
	(0.0011)	(0.0013)	(0.001)	(0.0009)	(0.0019)	(0.0021)
Language	0.149	0.132	0.078	0.077	0.024	0.036
	(0.0011)	(0.0017)	(0.001)	(0.0012)	(0.0019)	(0.0023)
GPA	0.055	0.075	0.072	0.056	0.108	0.05
	(0.0008)	(0.0008)	(0.0007)	(0.0009)	(0.0015)	(0.0021)
2010	-0.003	-0.003	-0.007	-0.004	-0.013	-0.01
	(0.0015)	(0.0013)	(0.0013)	(0.0012)	(0.0027)	(0.0028)
2012	0.011	0.022	0.024	0.025	0.026	0.034
	(0.0015)	(0.002)	(0.0013)	(0.0019)	(0.0027)	(0.0034)
Obs.	484549		484549		203596	

Note: this table shows estimates of each outcome for the years 2010-2012. The base year is 2011. We consider the events that a student enrolls in, and graduates within seven years from, some G33 program. "Model" columns: we use the results of 52 simulation draws in which we draw utilities, availability indicators, human-capital indices, and parameters from their estimated posterior joint distribution. In each draw, we simulate the market, then estimate the relevant linear models. We report means and standard deviations of parameter estimates from the relevant models over these draws.

Figure 2.1: Changes in Program Cutoffs Over Time



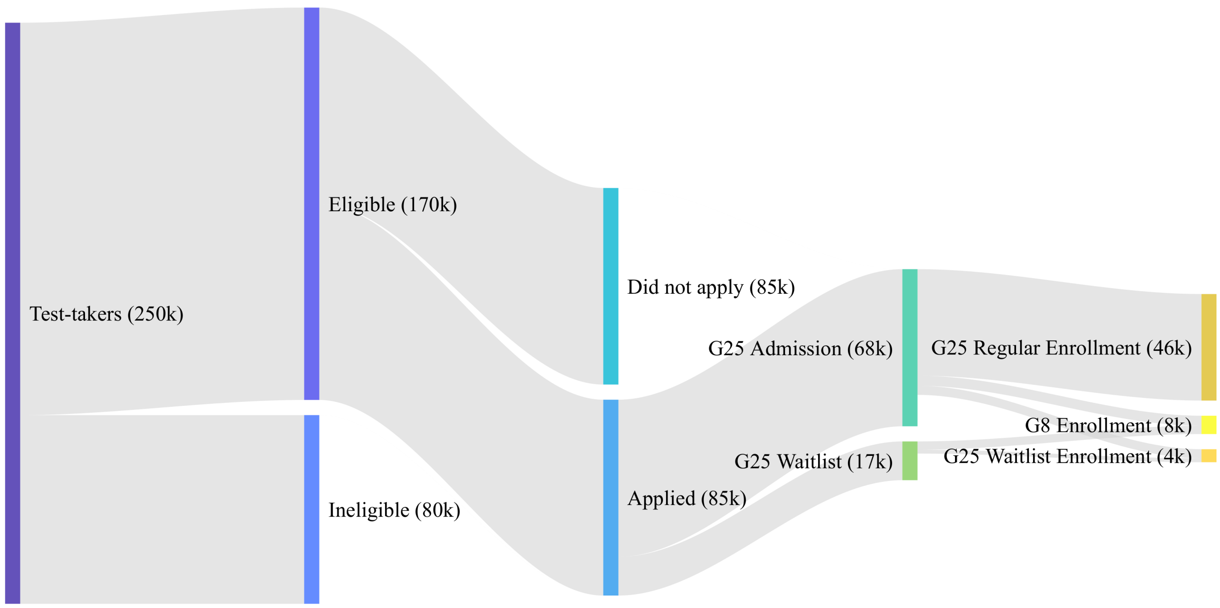
Note: The figure shows program cutoffs in 2011 plotted against those in 2010 (left panel) and again those in 2011 plotted against 2012 cutoffs (right panel). The different colored bands represent ranges of plus and minus 25, 50 and 75 point differences in cutoffs.

Table 2.5: Main Counterfactual Results

Counterfactual	All (Avg.)	Male Private	Male Public	Female Private	Female Public
A. Welfare (1m CLP)					
No Frictions (*)	2.601	4.381	2.293	6.211	1.964
	(0.079)	(0.27)	(0.076)	(0.421)	(0.038)
Baseline - *	-0.031	-0.043	-0.034	-0.047	-0.024
	(0.002)	(0.005)	(0.002)	(0.008)	(0.003)
Exclude G8 - *	-0.294	-0.116	-0.303	-0.285	-0.318
	(0.007)	(0.013)	(0.011)	(0.027)	(0.008)
Max Welfare - *	0.639	-0.969	0.321	4.172	0.635
	(0.017)	(0.261)	(0.09)	(0.424)	(0.092)
Free G33 - *	0.73	2.077	0.565	1.736	0.47
	(0.006)	(0.038)	(0.007)	(0.041)	(0.005)
B. Enrollment (pct)					
No Frictions (*)	42.27	67.785	39.674	67.551	35.862
	(0.025)	(0.096)	(0.04)	(0.08)	(0.041)
Baseline - *	-0.621	-0.657	-0.741	-0.548	-0.519
	(0.016)	(0.056)	(0.03)	(0.111)	(0.04)
Exclude G8 - *	-3.891	-1.144	-4.185	-1.315	-4.544
	(0.039)	(0.186)	(0.075)	(0.179)	(0.066)
Max Welfare - *	1.5	-13.916	0.384	12.65	3.455
	(0.067)	(2.427)	(0.71)	(0.734)	(0.742)
Free G33 - *	5.648	9.097	5.518	5.392	5.185
	(0.134)	(0.71)	(0.232)	(0.5)	(0.154)
C. Seven-year Graduation (pct)					
No Frictions (*)	52.213	52.895	41.175	69.228	57.74
	(0.202)	(0.762)	(0.341)	(0.521)	(0.281)
Baseline - *	-0.012	-0.017	-0.077	0.031	-0.041
	(0.047)	(0.13)	(0.084)	(0.134)	(0.034)
Exclude G8 - *	-0.338	-0.273	-0.447	-1.043	-0.413
	(0.14)	(0.29)	(0.096)	(0.702)	(0.148)
Max Welfare - *	1.737	0.016	0.909	4.448	0.767
	(0.558)	(0.739)	(0.317)	(3.704)	(0.277)
Free G33 - *	2.605	7.654	1.97	5.471	0.932
	(1.034)	(4.285)	(0.34)	(3.969)	(0.206)

Note: All counterfactuals conducted using 2012 data. We draw from the posterior joint distribution of parameters and latent utilities (u, u_0). Waitlist processes and realizations of frictions a are simulated according to parameters α. We conduct 26 draws for each counterfactual. "No Frictions": all programs on platform, $a^*_{ij} > 0$ for all i, j. "Baseline": all programs on platform, parameters as estimated. "Exclude G8": G8 programs off platform. "Max Welfare": maximize sum of student utilities subject to eligibility constraints but otherwise ignoring programs' preferences, holding pricing rules fixed. "Free G33": All programs on platform, all programs free for all students.

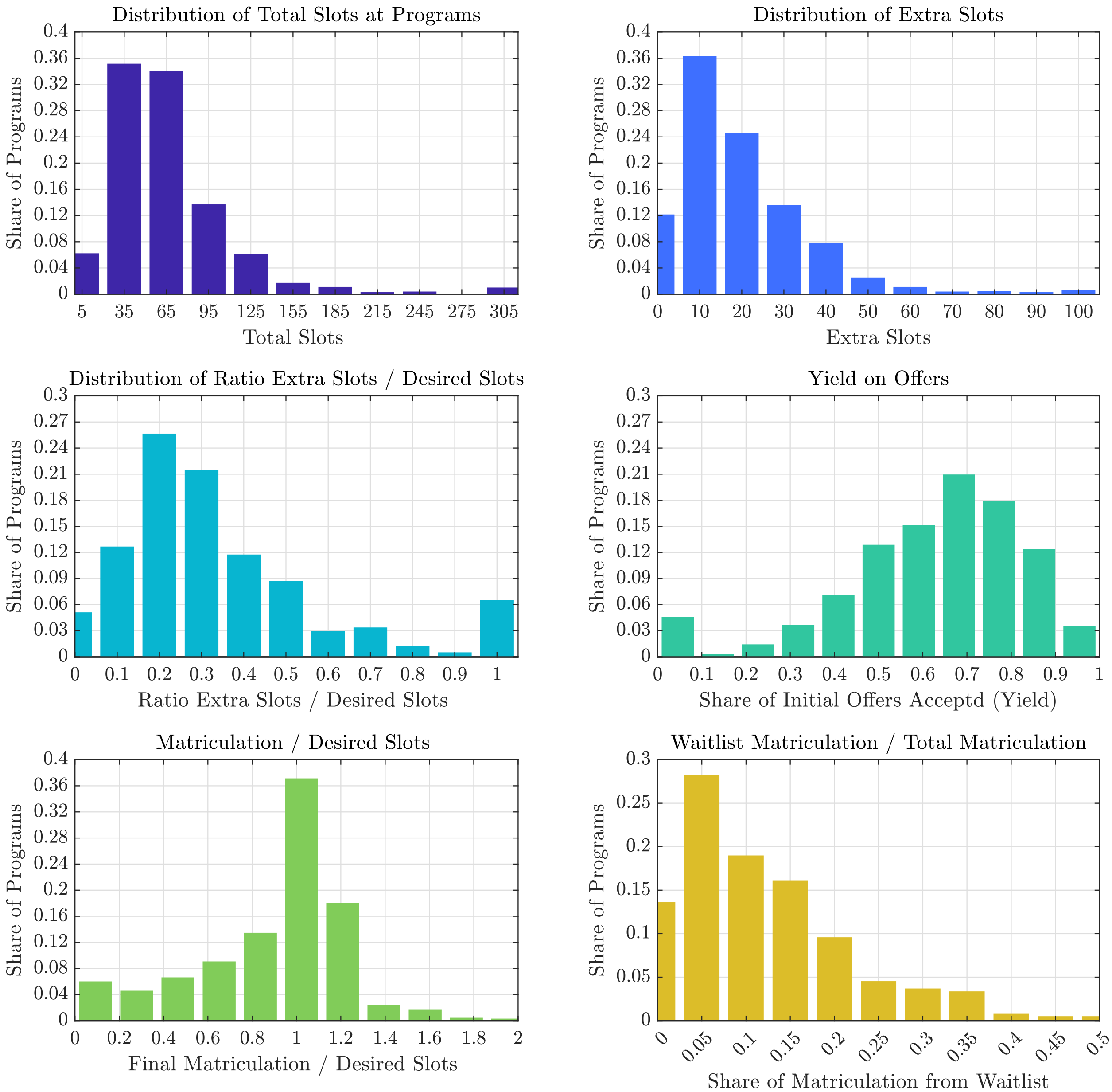
Figure 2.2: Diagram of the on-platform application process



PSU (Dec. 13-14) → Scores (Jan. 3) → Applications (Jan. 3-9) → Offers (Jan. 13) → Enrollment*

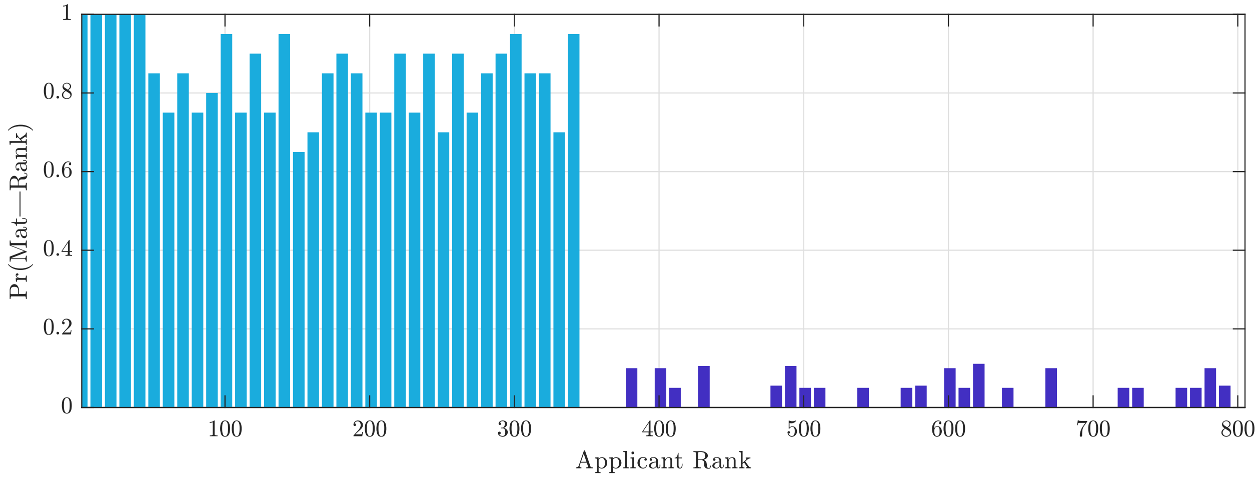
Note: Diagram shows the progression of steps for applicants on the centralized assignment platform, and the flow of the mass of applicants throughout the process. The numbers of students in each step is for 2011, before the platform was expanded. The baseline is the cohort of students that take the national college entrance exam in late 2010, seeking admission in 2011. *Enrollment dates are not mandated by the platform, but universities usually conduct the enrollment process within a week of the date that offers are released.

Figure 2.3: Total Slots, Excess Slots, Program Yield



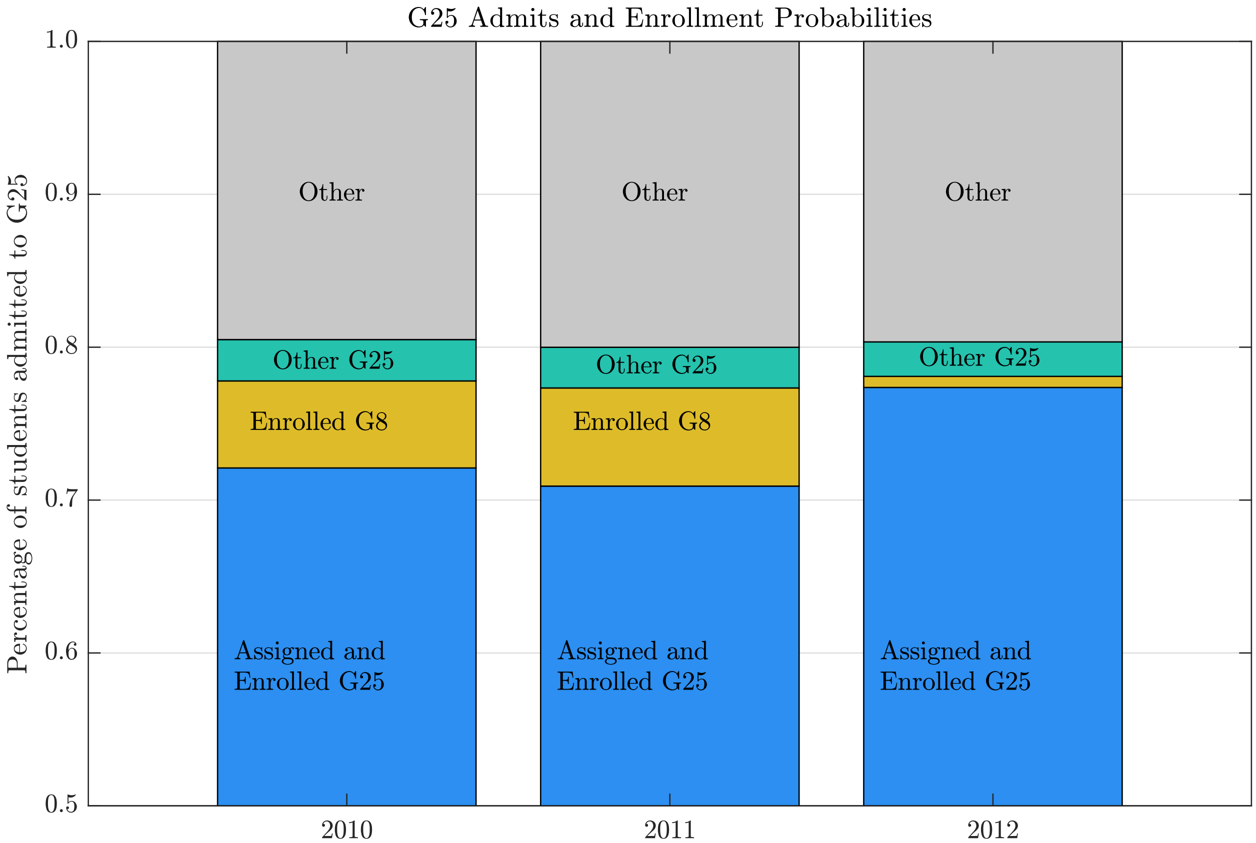
Note: This figure describes the distribution of posted slots, extra slots, yield, matriculation and waitlist matriculation in 2011. The top left panel shows the distribution of total slots, with the highest 2% of programs not show. The top right panel shows the distribution of extra slots posted in expectation of declined offers. The left middle panel presents the distribution of the ratio between extra slots and desired matriculation. The right middle panel presents the distribution of the yield that initial offers have. The bottom left panel shows the ratio between ex-post matriculation and ex-ante desired slots. The bottom right panel shows the number of waitlist matriculated students as a share of total matriculation. 34% of the programs do not have any waitlist matriculation either because it was not needed or not possible because they had no excess demand.

Figure 2.4: Case Study: Enrollment Probability at Economics - University of Chile - 2010/2011



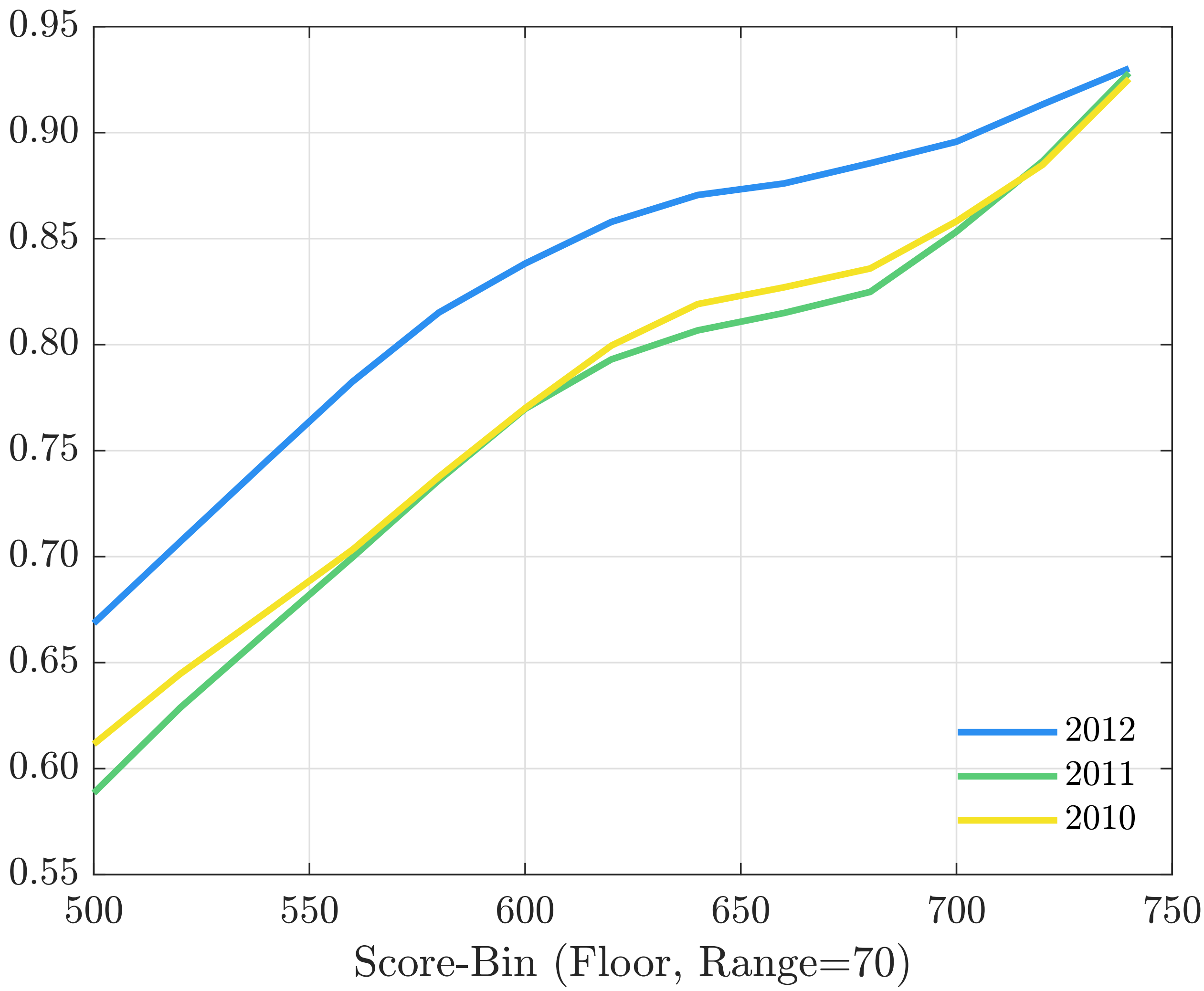
Note: This figure shows the probability of enrolling students who are admitted or waitlisted as a function of their rank. The figure shows Economics at the University of Chile which is a highly selective program with a large class of over 300 slots offered. Two of the authors did their undergraduate training at this program. The x-axis shows the student rank (from 1 being the highest to the last admit). The y-axis shows the probability that students will enroll, shown in bins of 10 students.

Figure 2.5: Enrollment probabilities for G25 admits



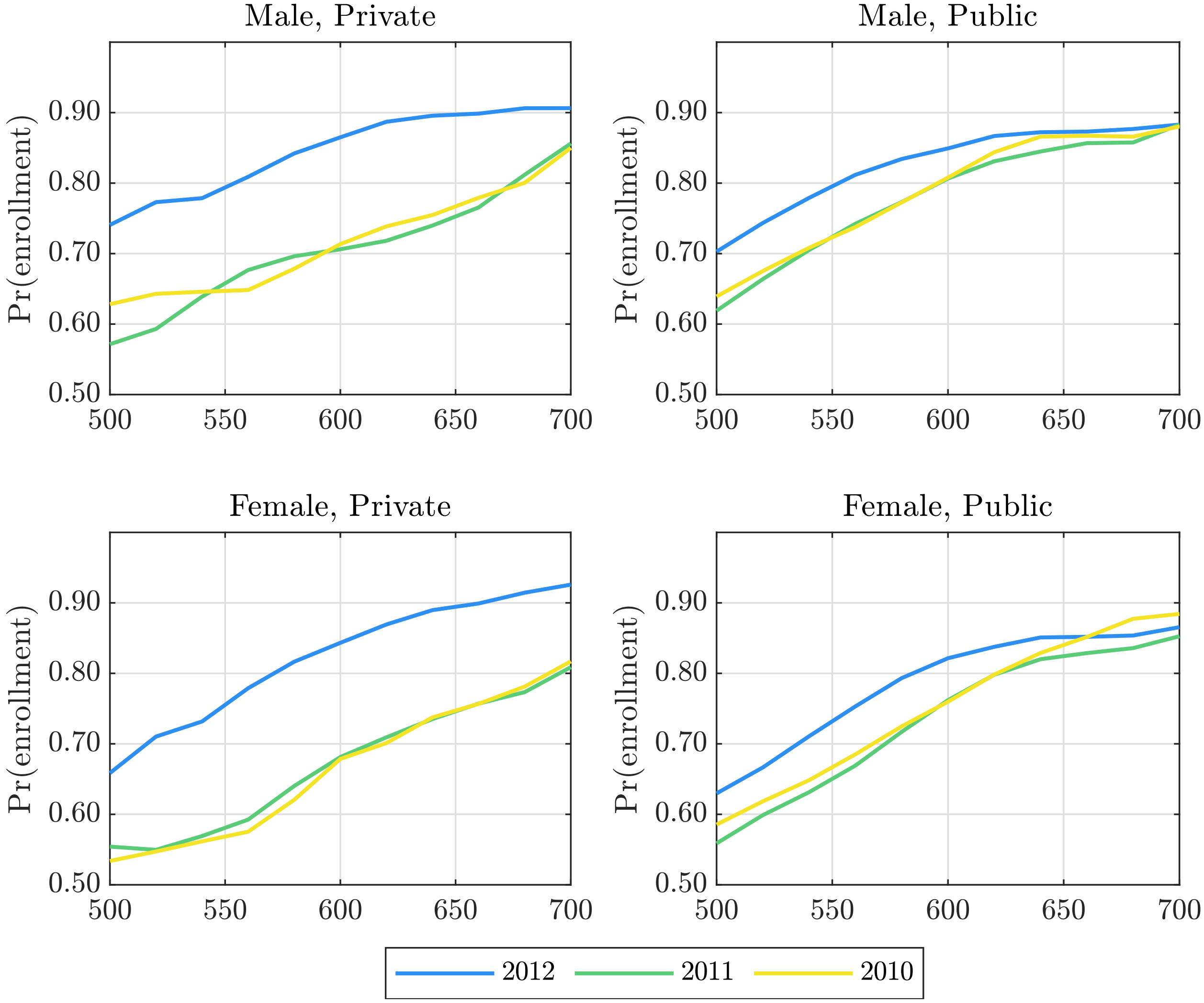
Note: This figure shows enrollment probabilities for students admitted to traditional ("G25") options, by year. The share of such students who enrolled in G25 programs increased, and the share enrolling in G8 programs decreased, in 2012. In 2012, the only way for such students to be admitted to G8 programs was off of waitlists.

Figure 2.6: Enrollment probability, G25



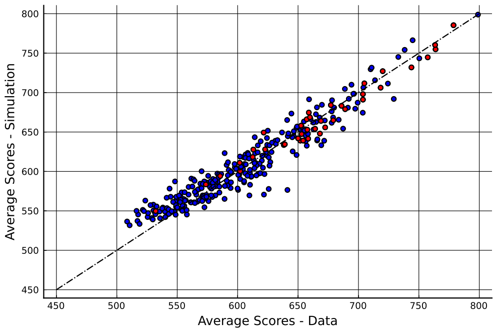
Note: The figure show the probability that a student assigned to an option on the platform, accepts and enrolls in that option. The lines show conditional means within 70 points, and the "floor" of the range is shown in the x-axis (e.g. 600 corresponds to the range [600, 670]).

Figure 2.7: G25 Enrollment Probability By Gender/SES.

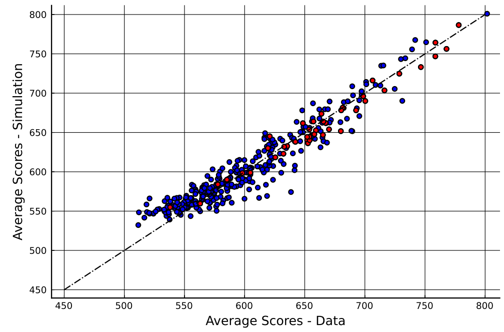


Note: The figures show the probability that a student assigned to an option on the platform, accepts and enrolls in that option. The lines show conditional means within 70 points, and the "floor" of the range is shown in the x-axis (e.g. 600 corresponds to the range [600, 670]).

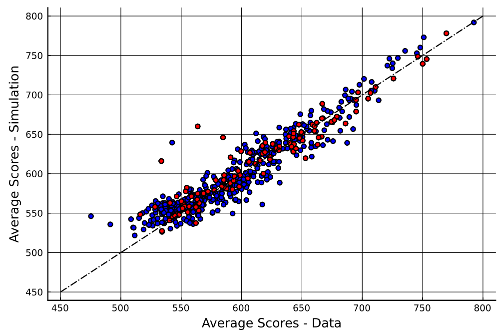
Figure 2.8: Model Fit - Selectivity



(a) Enrolled students, 2010



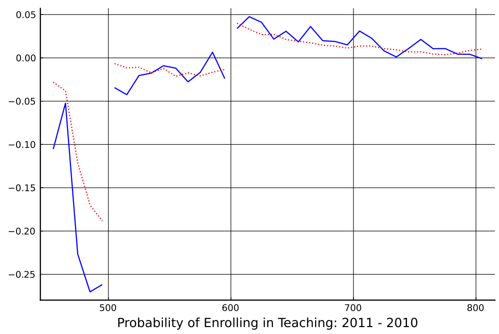
(b) enrolled students, 2011



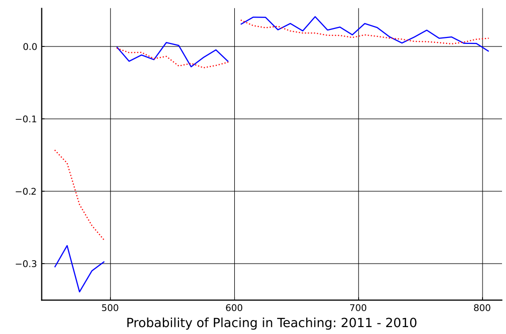
(c) enrolled students, 2012

Note: each panel shows, for each program, the mean math and verbal scores of students enrolled in the program in the data (X-axis) and in simulations (Y-axis). We restrict to programs with at least 50 seats. Red dots denote enrollment greater than 100 students.

Figure 2.9: BVP Impacts



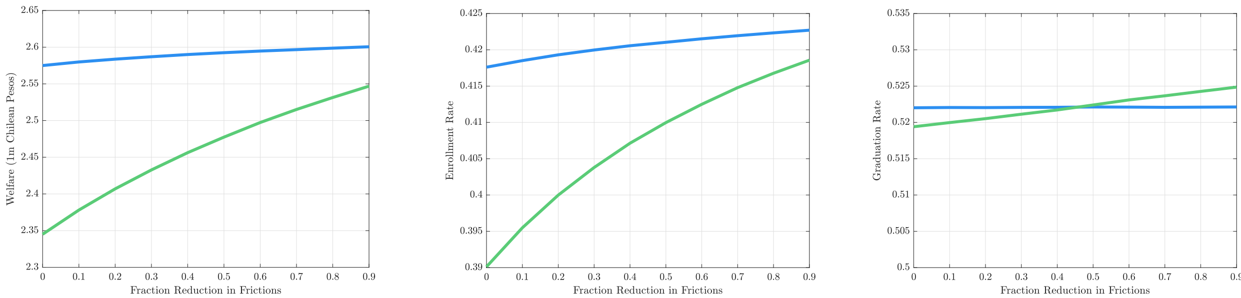
(a) Pr(Enroll in Teaching)



(b) Pr(Place in Teaching)

Note: the left panel shows the difference in the probability of enrollment in teaching majors between 2010 and 2011 as a function of mean math+verbal test scores. The right panel shows differences in the probability of obtaining an initial on-platform placement in a teaching major between 2010 and 2011 as a function of math+verbal test scores. Students above 600 points are eligible for full scholarships, while students with scores below 500 are restricted from entering participating programs.

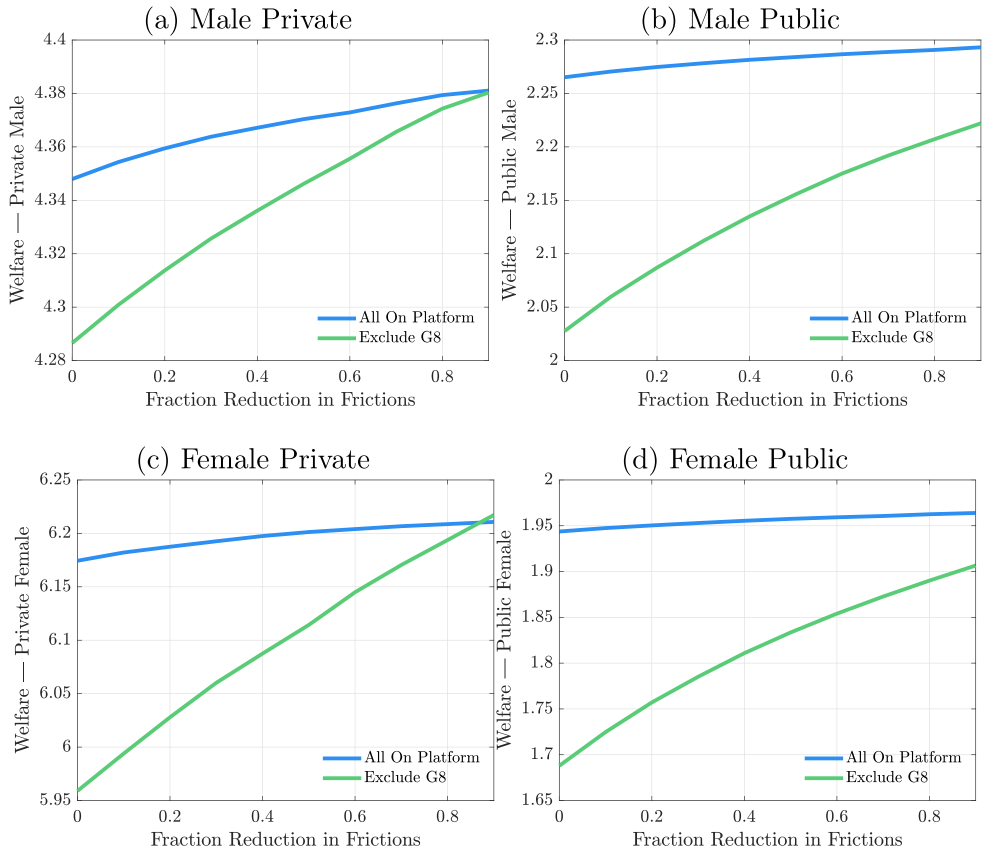
Figure 2.10: Impacts of Reducing Frictions (α)



(a) Welfare (1m CLP) (b) Enrollment (Pct.) (c) Graduation (Pct.)

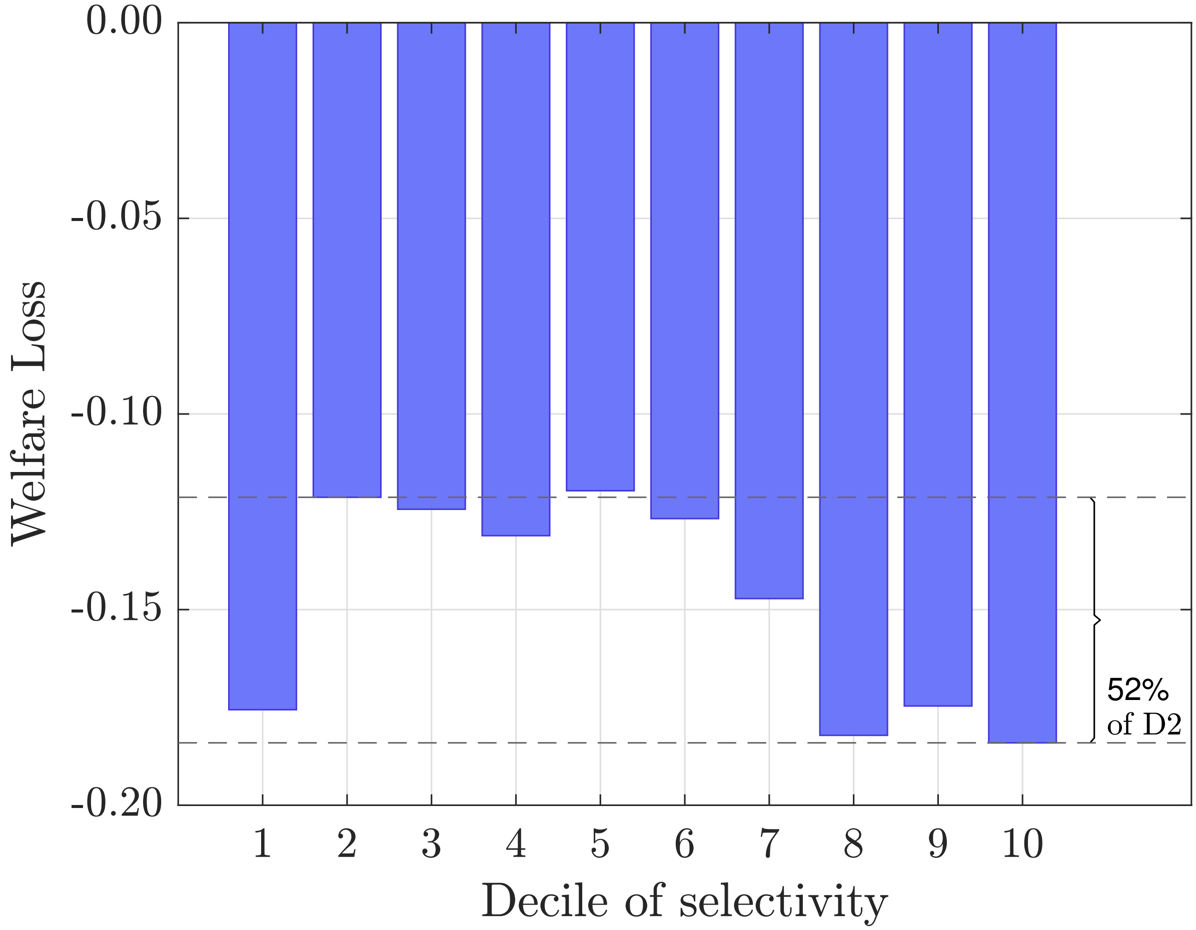
Note: Blue lines (■) indicate when all programs are in the platform, while green lines (■) indicate when G8 programs are excluded. All model-predicted failed-contact probabilities $Pr(a_{ij} = 0)$ multiplied by $(1 - p)$, where p is "fraction reduction in frictions" on X-axis.

Figure 2.11: Welfare Impacts of Reducing Frictions (α): Heterogeneity by Type



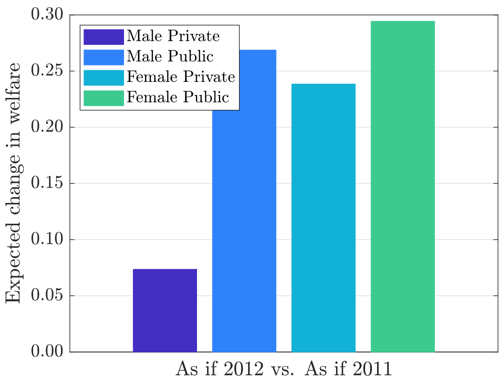
Note: Model-predicted failed-contact probability $Pr(a_{ij} = 0)$ multiplied by $(1-p)$, where p is "fraction reduction in frictions" on X-axis.

Figure 2.12: Utility loss of removing options ordered by selectivity

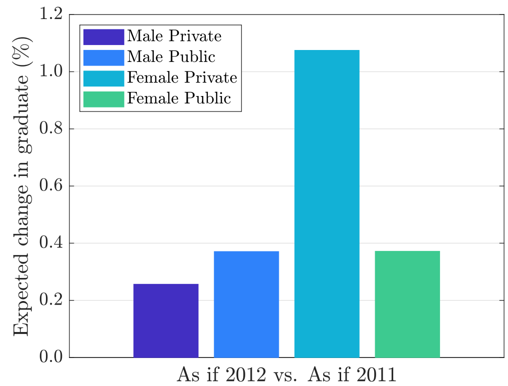


Note: Loss is calculated as the difference in mean utility, in units of 1m Chilean Pesos, between the model-simulated 2012 baseline and the counterfactual in which all program seats in the d'th decile of selectivity—as measured by programs' 2012 mean math+verbal scores—are withheld from the platform. Negative (positive) values indicate losses (gains) relative to baseline.

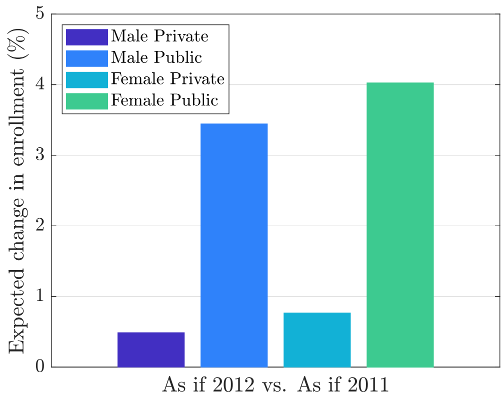
Figure 2.13: Change in Welfare, Graduation, and Enrollment Rates by Type



(a) Δ welfare for students by type



(b) Δ graduation for students by type



(c) Δ enrollment for students by type

Chapter 3

Screening and Recruiting Talent At Teacher Colleges Using Pre-College Academic Achievement

[1] **Abstract**

This paper studies screening and recruiting policies that restrict or incentivize entry to teacher colleges. Using historical records of college entrance exam scores since 1967 and linking them to administrative data on the population of teachers in Chile, we first document a robust positive and concave relationship between pre-college academic achievement and several short and long run teacher outcomes. We use an RD design to evaluate two recent policies that increased the share of high-scoring students studying to become teachers. We then show how data-driven algorithms and administrative data can enhance similar teacher screening and recruiting policies.

3.1 Introduction

Effective teachers matter for students' short and long run outcomes (Chetty et al., 2014; Araujo et al., 2016) and accordingly, governments aim to increase their teachers' productivity (OECD, 2005). A commonly used set of policies looks to increase the effectiveness of teachers

[1]Coauthored with Franco Calle, Sebastian Gallegos and Christopher Neilson.

once they are in the classrooms through incentives, training, accountability measures or rewards (e.g. Biasi, 2021). An alternative set of policies that are less studied are aimed at recruiting or screening candidates before they enter teacher colleges or the teaching profession Jackson et al. (2014).

Recruiting policies can be convenient compared to on-the-job policies for several reasons. The first is that they can prevent students from exposure to ineffective teachers, who are usually difficult to remove once employed. Second, it is logistically and politically hard to implement pay-for-performance schemes that look to encourage effort (Hoxby, 1996; Hanushek, 2011; Biasi, 2021).[2] Third, the evidence suggests that later investments in training have little influence on teacher productivity (Jackson, 2012; Lombardi, 2019).

The design of effective recruiting policies is hard. It requires a good prediction of teachers' effectiveness ex-ante, which has been elusive in the past (Rivkin et al., 2005; Harris and Sass, 2011; Jackson et al., 2014). Recruitment policies would ideally be informed by causal evidence, but prior research has been largely correlational in design (See et al., 2020). Administrative sources and historical records are being digitized and governments are developing the capacity to store and use the data (Figlio et al., 2017). This increasing data availability is likely to help overcome the lack of informative determinants of future teacher productivity and produce research designs helping to identify the causal effects of policies and initiatives on teacher recruitment. In addition, the development of improved predictive algorithms is lowering the cost of making more accurate predictions and influencing decisions, such as hiring, in many markets (Agrawal et al., 2018; Chalfin et al., 2016).

This paper studies policies that use pre-college achievement to recruit or screen out students entering teacher-colleges. We start with a descriptive analysis suggesting that teacher outcomes might be predictable thanks to better data availability. We use recently digitized historical records from 1967 onward and link them to the population of teachers in Chile, to document the relationship between their own academic achievement at age 18 and results as teachers up to 30-40 years later.

We then estimate the causal effects of two recent policies that both restricted and incentivized entry to teacher-colleges. Using administrative records and several regression

[2] Flexible pay schemes might even increase the gender wage gap Biasi and Sarsons (2021).

discontinuities based on the policies' eligibility cutoffs, we assess whether they attracted higher-scoring test-takers to teacher colleges. We also examine later outcomes (like graduation, performance at exit exams and employment in schools) and how the policies interact over time.

Finally, we ask whether the combination of better administrative data and flexible prediction methods can enhance the recruitment procedures implemented by policymakers. We use machine learning methods to find that data-driven algorithms might outperform traditional cutoff-based mechanisms. This data-driven approach seems promising for better targeting of investments in future teachers.

Our main findings are as follows. First, our descriptive analysis shows that there is a robust positive and concave relationship between teachers' pre-college academic achievement and a variety of short and long run teacher outcome measures. These include early measures such as graduation from teacher colleges and college exit exams; and later results such as wages, employment, external classroom teaching evaluations, students' achievement gains and students' perceptions about teaching effectiveness. Broadly, we find that below-average pre-college achievement is systematically associated with lower performance as teachers measured up to thirty and forty years later.

We readily acknowledge that the observed correlation between teachers' entrance exams and later outcomes could be caused by the sorting of teachers into schools and students. There is important work studying teacher sorting in the context of Chile by Tincani and coauthors (e.g. Tincani, 2014; Tincani et al., 2016; Tincani, 2021) which explicitly models the sorting process and simulates related teacher policies. We see our paper as complementing this structural work with empirical descriptive and causal evidence of the relationship between pre-college academic achievement and later outcomes.

Another potential cause (less explored in the literature) is access to higher value-added teacher colleges. We address this question directly by estimating teacher colleges' value-added using a regression discontinuity design building on institutional features of the Chilean centralized admissions system. Using data on the population of applicants to teaching colleges from 1977 to 2011, we find no evidence that any particular teaching college adds more value or contributes to closing or increasing the predicted gap in teacher effectiveness. This result suggests that college training is not enough to undo initial differences and that pre-college academic readiness has a persistent relationship with later teacher productivity.

Finally, the mere fact that lower-scoring students enter the teaching profession is a concern, because the academic abilities of incoming teachers are an indicator of the status of the teaching profession (OECD, 2005; Lankford et al., 2014). Discouraging low-ability individuals (and motivating more high-ability students) to enter teacher colleges might at least weakly increase future teacher quality and the status of the profession.

Overall, the evidence suggests that college entrance exams could be useful to screen out or recruit students entering teacher colleges. We study two related policies implemented in this spirit within the last ten years. The first policy started in 2011 and offered full tuition subsidies for test-takers with scores in the top 20% of the exam distribution. It also required participating teacher colleges to reject applicants with scores below the national mean. The second policy enacted in 2017 extended this requirement to all teacher colleges in the country.

We evaluate these policies using regression discontinuities based on the eligibility score cutoffs for high and low-scoring applicants. We implement this empirical strategy using individual-level data from the population of test-takers in the country, for ten cohorts of students.

Our findings show that the policies increased the number of higher-scoring students enrolled in teacher colleges, with the largest effects at the lower cutoffs of the college entrance distribution (about 37% of an effect size). Effects at the higher cutoffs are large computed as effect sizes (about 100%) but small in levels. This finding serves as a reminder that recruitment incentives are only as good as the next best option and that high-achieving students have many good alternatives, so it is harder to move them toward teaching.

Eight years after the policy was first implemented, we find that the higher-scoring students went on to work in schools later on (effect size of 34% on employment at schools). This finding indicates that the policy was successful at raising the predicted quality of students who entered the teaching profession. We also measured other early indicators such as graduation rates and the exit exams, finding precise zero effects. These results suggest that the higher achieving students graduated and took the teacher exams as we would have predicted using the college entrance scores, indicating that the predicted relationship between pre-college academic achievement and teacher medium run outcomes is policy-invariant in this context.

We finally assess whether the use of data-driven algorithms may enhance the screening procedures planned for the future by the government. We train classification trees that outperform all the different government recruiting policies, by going beyond single-dimensional cutoff rules. Importantly, our classifiers are simple enough to not sacrifice interpretability nor rely on complicated sets of input features. Taken together, the findings support the use of data-driven, machine learning methods as a promising way of aiding screening and recruitment policies.

Our results are important because they have direct policy implications. If teacher effectiveness (or lack thereof) is possible to predict early on, then policies could focus resources on recruiting and retaining the most promising candidates and filtering out applicants who are more likely to become ineffective teachers. This is particularly relevant because teacher labor markets are known to be inefficient (Neal, 2011; Gilligan et al., 2018), misallocation of talent can be widespread in many cases (Bau and Das, 2018), and there is limited scope to sideline or retrain ineffective teachers once they are in the system, especially in the public sector (see, e.g., Estrada (2019) for the Mexican case and Bold et al. (2017, 2019) for seven African countries). Taken together, our findings suggest that, at least in the context of middle-income countries like Chile in the period of our study, resources that look to subsidize teacher training should be targeted towards prospective teachers that have a minimal level of baseline academic achievement.

We contribute to the literature on teacher quality and prediction. We see our results as consistent with the existing evidence on the topic from the US and developed countries Rockoff (2004); Rothstein (2006); Clotfelter et al. (2007). In the case of Chile, most of our ability to predict teacher effectiveness comes from low-achieving students who become teachers and this margin may not be relevant in more developed countries. This evidence is also consistent with recent cross-country descriptive work by Hanushek et al. (2019), who find that in developed economies differences in teacher cognitive skills can explain significant portions of the international differences in student performance (measured by PISA scores). In addition, this analysis uses rich pre-college academic achievement for the population of teachers which may have not be available to researchers in the past.

Our findings highlight avenues for further research in an increasingly data-rich environment where prediction is a key input to policy design (Mullainathan and Spiess, 2017; Kleinberg et al., 2017). In this context, we believe that empirical exercises similar to ours (e.g. Athey, 2019; Sajjadiani et al., 2019) will be increasingly common in the near future.

3.2 Context, Policy and Data

3.2.1 Context

Chile is a country that has reached low levels of teacher absenteeism and a student-teacher ratio close to the levels displayed by OECD countries (World Bank, 2013). Teacher absenteeism is estimated at 5% (Paredes et al., 2015) which is much lower than other countries in Latin America; Chaudhury et al. (2006) estimate absenteeism rates of 15% in Brazil, 14% in Ecuador, and 11% in Peru.

The student-teacher ratio is about 20, which is the result of an increasing number of teachers and a stable population of students over time. The number of classroom teachers[3] has increased from 125,000 in 2008 to 164,000 in 2018 (MINEDUC, 2019), while student enrollment has plateaued and even showed a slight decrease over the last ten years (from 3.1 million in 2008 to 2.9 million in 2018).[4]

With enough teachers in the classrooms and high rates of student enrollment (OECD, 2009), the policy focus in the last ten years has been devoted to bringing more qualified individuals to the teaching profession.

Attracting more skilled individuals to be teachers is challenging because, among other factors, teachers are typically paid less than comparable professionals (Mizala and Nopo, 2016; Hanushek et al., 2019).[5] Consistently, we know from the related literature that college

[3]Teachers in Chile work in public schools, which are funded and administered by the government; voucher schools, which are funded mainly with public funds but administered by privates; and private schools are both funded and administered privately.

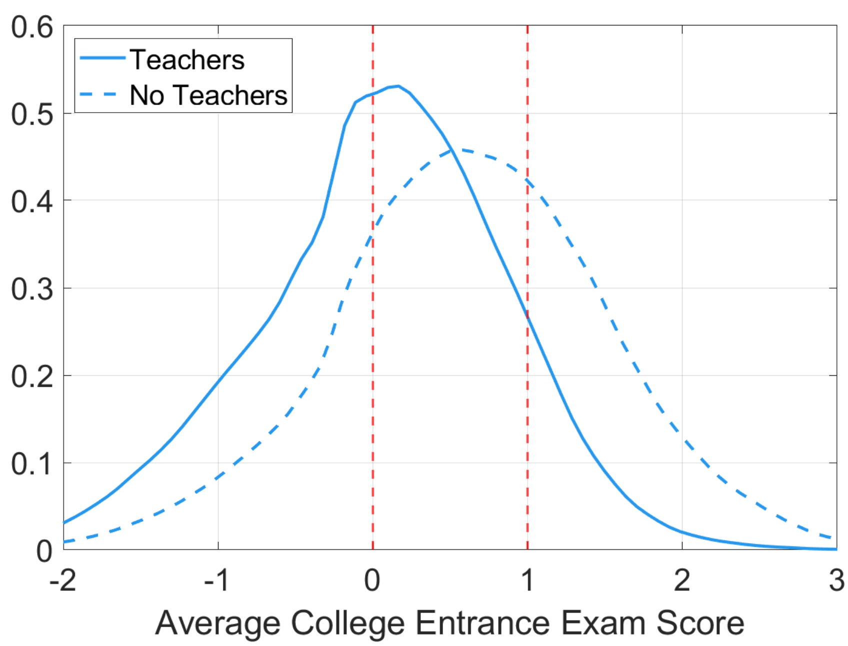
[4]These numbers are consistent with the demographic transition being experienced by Chile, exhibiting low fertility and mortality rates, and relatively high life expectancy (World Bank, 2011).

[5]Mizala and Nopo (2016) estimate a teacher underpayment of about 20% in Latin American countries in 2007 (with a 18% for Chile) after controlling for a set of characteristics linked to productivity. Hanushek

graduates with higher college entrance scores are less likely to enter teaching (Vegas et al., 2001; Hanushek et al., 2019; Estrada and Lombardi, 2020), and Chile is no exception.

Figure 3.1 shows that in year 2010 (before the implementation of teacher recruiting policies described below), teacher colleges' students scored only 0.1 standard deviations (σ) above the national mean in the college entrance exam,[6] while students enrolled in other fields like engineering, law, and medicine scored about 0.6σ above. Also, test scores for education students had been declining over time, since in 1995 students from teacher colleges scored 0.3σ over the national mean (Alvarado et al., 2011). This pattern is similar to the evidence for the U.S. (Bacolod, 2006; Corcoran et al., 2004; Podgursky et al., 2004; Hoxby and Leigh, 2004).

Figure 3.1: Distribution of College Exam Scores: Teachers Colleges vs Other Fields



Note: Figure 3.1 plots the distribution of college entrance exam scores before the implementation of teacher recruiting policies (years 2008-2010), for two groups: freshmen in teacher colleges (continuous line) and freshmen in the health, law and STEM fields (dotted line). The entrance exam score (in standard deviation units) is the average of the math and language exams. We provide further details of the college entrance exam in section 3.2.3.

et al. (2019) estimate that teachers in the U.S are paid some 20 % less than comparable college graduates. Evans et al. (2020) find that, in 7 out of 15 African countries, teachers suffer a deficit in earnings relative to comparable wage workers that averages 26%.

[6]These scores correspond to the average of the math and language exams. We describe the college entrance exam in section 3.2.3.

3.2.2 Teacher Recruitment Policies

In this context, the Chilean government implemented two policies to recruit talent at teacher colleges. The first policy, called *Beca Vocacion Profesor (BVP)* was implemented in 2011. It consisted in full tuition subsidies for prospective students who scored about 1σ above the mean in the college entrance exam. The BVP policy also required participating teacher colleges to reject applicants with scores below the national mean.

The second policy was the *Nueva Ley de Carrera Docente* (NLCD) and started in 2017. The NLCD basically imposed the BVP restrictive requirement for admissions at all teacher colleges across the board. Under this policy all applicants to teacher colleges had to have college entrance exam scores at least as high as the national mean, or have a high-school GPA in the top 30% of their graduating cohort. We provide more specific details and assess both policies in section 3.4.

3.2.3 Data

Data on Pre-College Academic Achievement. The main measure of teachers' pre-college academic achievement that we use in this paper is their scores on college entrance exams taken since 1967. The historical records come from digital copies of old books and newspapers collected as a part of the work done in Hastings et al. (2014). In continued partnership with the national agency in charge (DEMRE) we complemented these data by digitizing additional test scores back to the first test in 1967. For the more recent cohorts of test takers (2004-), the DEMRE has made electronic records available.

The Chilean national college entrance exam is similar to the SAT in the United States. Currently, the exam is called the *Prueba de Selection Universitaria* (PSU) and has been administered once a year since 2004. Prior to that a similar test called *Prueba de Aptitud Academica* had been implemented from 2003 back to 1967, which makes Chile have one of the longest-running centralized college assignment systems in the world.[7] Test-takers complete exams in mathematics and language as well as other specialized subjects. The

[7] A detailed explanation of the application and enrollment process for the period 1980-2009 is presented in Hastings et al. (2014) and a review comparing centralized systems in the world in Neilson (2021).

scores are scaled to a distribution with a mean and median of approximately 500 and a standard deviation of 110. The exam scores are required to apply to all public universities and most private universities and institutes.

Data on Teacher Productivity. We gathered a host of teacher productivity proxies from different sources of administrative records. Our measures of teacher productivity include short run outcomes such as graduation from teacher colleges and college exit exams; and longer run outcomes such as wages, employment, external classroom teaching evaluations, students' achievement gains and students' perceptions about teaching effectiveness.

In the next section, we correlate all these measures of teacher productivity with the digitized pre-college achievement described above. We provide specific details in section C.1, and we describe each measure and data below.

Graduation from Teacher Colleges. We use microdata on the population of teacher college enrollment and graduation, which the Chilean Ministry of Education (MINEDUC) started to collect in 2004 and 2009, respectively. We constructed graduation rates for 105K individuals combining enrollment records from years 2004 to 2010 with graduation data for years 2009 to 2018. This procedure allows us to study graduation rates that were 'on time' (i.e., within 5 years after initial enrollment, at approximately 23 years old) and also late graduation (i.e., up to 8 years after enrollment, at about 26 years old).

Exit Exams. The exit exams were first implemented in 2009. Our data consists of microdata for all the exit exam test-takers between 2009 and 2017. The sample consists of about 35K just-graduated teachers with scores on different exams, like a disciplinary knowledge test (e.g., math knowledge for math teachers) and a pedagogical knowledge test (e.g., capacity to explain concepts in a coherent way). At the time of the exam, test-takers were on average about 25 years old.

Government Evaluations. The government started to implement teacher evaluations in public schools in 2004. We gathered information for 63K classroom teachers, evaluated between 2004 to 2017. Each evaluated teacher receives an overall score at the end of the evaluation process. The MINEDUC uses that score to classify teachers into four categories of performance, from best to worst: outstanding, competent, basic, and unsatisfactory. The overall score is composed of four components: (i) a self-evaluation questionnaire (10%); (ii) a third-party reference report, filled by the school principal or supervisor (10%); (iii) one peer review (20%), and a teacher performance portfolio (60%) that collects direct evidence on teaching skills, pedagogical decisions and classroom practice.[8] Previous research (OECD, 2013; Bruns and Luque, 2015) suggests that the portfolio component has the strongest association with students' progress measured using standardized test scores. Therefore, we use both the overall score and the portfolio score in our analysis later on. On average, evaluated teachers were 40 years old at the time of the assessment.

Employment in Schools. We gathered information for about 240K graduates from teacher colleges in years 1995 to 2017 which we merged with the population of teachers working in schools between 2003 to 2018. We compute whether graduates worked during that period of time and correlate that with entrance exam scores. The age at employment after ten years and twenty of graduation average 37 and 46 years old respectively.

Wages in Schools. The MINEDUC collected information on teacher wages by asking principals about teachers' wages and working hours in the year 2011. Teachers with information on wages are about 117K. Teachers working in public schools are approximately 40 percent of the sample (49K). They benefit from a special labor code, which makes wages grow with tenure and they are not expected to change with productivity. Teachers in voucher schools represent 60 percent of the sample (68K). The voucher sector operates under the regular and more flexible labor code, and thus teacher wages can be given a market-clearing interpretation, associated with productivity. On average, teachers with wage information are 37 years old.

Students' Achievement Gain. We use students' achievement gains during the academic year as another proxy of teacher productivity. The MINEDUC does not implement value-added exams, but De Gregorio and Neilson (2020) implemented math tests specially designed to measure the gain in achievement for students with the same teacher during the academic year. The sample consists of about four thousand students in grades 9th to 11th tested at the beginning and the end of the year in 2016.

[8] As we expand in section C.1, the portfolio component includes two modules. In the first module, teachers plan a class defining its contents and related assessments. They are also asked questions about teaching practices. The second module consists of a videotaped class followed by a questionnaire on the students' behavior and understanding, and the teacher's own performance.

Students' Perceptions. We use students' perceptions regarding effective teaching as an additional measure of teacher quality. The survey implemented by De Gregorio and Neilson (2020) to the same four thousand students follows the recommendations of the Measures of Effective Teaching study carried out in the U.S. (Kane and Cantrell, 2010). Questions are categorized into eight dimensions of teaching practices and classroom environment: positive culture and learning environment, student understanding checked for and ensured, engaging learning environment, expectations held by the teacher, student input and ideas valued, learning fully internalized by students, encouraging and supporting relationships fostered, and classroom participation.

3.3 Pre-college Achievement and Teacher Outcomes

In this section, we document the correlation between pre-college academic ability and the teacher outcomes described in the previous section. We estimate regressions using outcomes at different moments of their careers on their own entrance exam scores taken at age 18. We also describe the empirical relationship showing non-parametric plots leveraging on our large sample sizes.

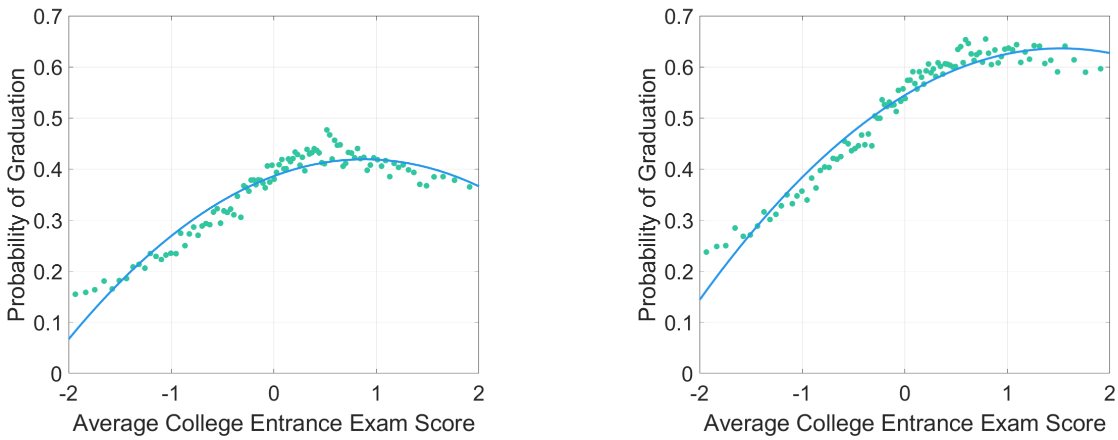
The general takeaway of this descriptive exercise is that the empirical relationship between pre-college skills and own teacher outcomes later on is positive and concave. We report regression coefficients in Table 3.1, organized in five panels of teacher outcomes: graduation, exit exams, productivity measures, labor market results and their students' outcomes. The coefficients come from separate regressions of different measures of teacher outcomes on the college entrance exam score (in standard deviation units and labeled 'PSU Score') and its square. The estimates on scores are all positive and significant, and most coefficients on the squared term are negative.

We discuss our estimation results complemented with simple visual evidence below. All figures in this section plot the y-axis variable within 100 equal-sized bins of the average college entrance exam score and fit estimated lines using all the underlying data.

In Figure 3.2 and Figure 3.3 we examine early outcomes of students from teacher colleges, like graduation rates and exit exams. Figure 3.2 shows that graduation rates correlate positively with test scores at entry, and that the relationship is concave. The results in Figure 3.2a, Figure 3.2b and the first panel of Table 3.1 show that an increase in one standard deviation on the college entrance exam scores relates to an increase in graduation rates of 7.0 and 11.2 percentage points after 5 and 8 years of initial enrollment. These are increases of approximately 20% relative to the baseline graduation rates of 35% and 50%, respectively. The positive correlation is much flatter for scores above the mean.

Figure 3.2: College Entrance Exam and Graduation from Teacher Colleges

(a) Graduation within 5 years (b) Graduation within 8 years

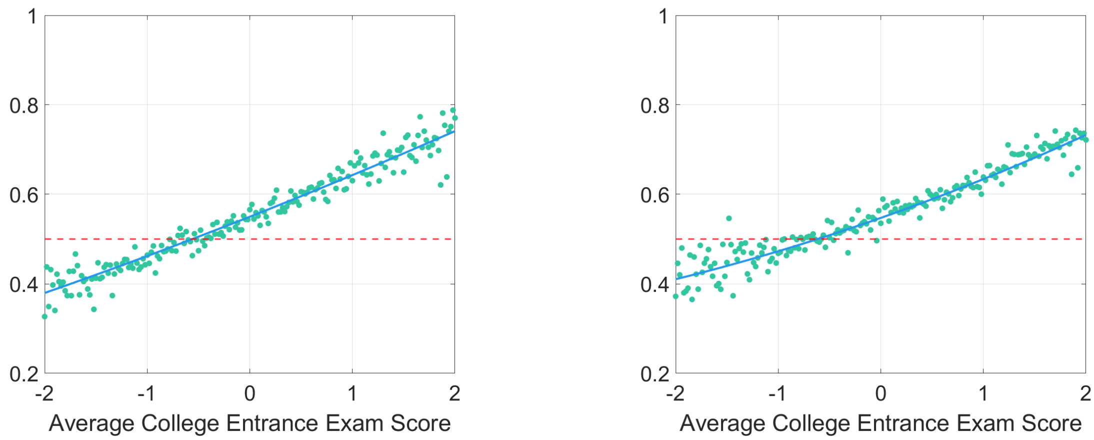


Note: The figures plot the probability of graduation after 5 years (Figure 3.2a) and 8 years (Figure 3.2b) of first enrollment, within 100 equal-sized bins of the average college entrance exam score and fits estimated lines using all the underlying data. The data consists in students enrolled in years 2004 to 2010 who graduated between 2009 and 2019. In both figures, the sample size is of $N = 105,422$.

Figure 3.3 shows the correlation between test scores at college entry and exit exams taken just before graduation. The graphs and the corresponding coefficients in Table 3.1 show that one standard deviation on the college exam test scores is associated with an increase of 0.50σ on the disciplinary and pedagogical skills measured in the exit exam. Table 3.1 also reports that one standard deviation in test scores is related to an increase of 0.28σ and 0.54σ in the writing skills and ICT skills exit exams, respectively.

Figure 3.3: College Entrance Exam and Teacher College Exit Exams

(a) Disciplinary Subject (b) Pedagogy Subject

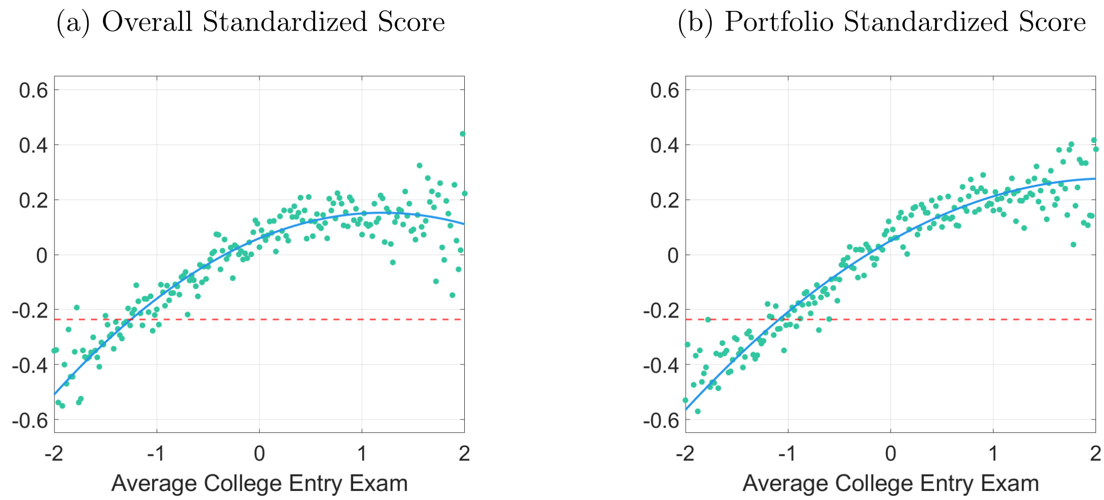


Note: The figures plot the standardized scores of two exit exams within 100 equal-sized bins of the average college entrance exam score (in std. dev. units) and fit estimated lines using all the underlying data. The two exams are disciplinary skills (Figure 3.3a) and pedagogical skills (Figure 3.3b). The data consists of exit exam test takers between the years 2009 and 2017. The sample sizes are $N = 35,355$ in Figure 3.3a, and $N = 33,409$ in Figure 3.3b.

We now present results for later outcomes, when individuals are teaching and working in schools. Figure 3.4 describes the bivariate relation between college entry exam scores and teacher evaluations taken up to 30 years later. As in Figure 3.2, the relationship is concave, suggesting that early scores may have a higher potential for identifying low-performance teachers thirty years later. Coefficients in Table 3.1 show that an increase of one standard deviation in entry exam scores is linked to increases of 0.14σ and 0.19σ on the overall and portfolio evaluations scores, respectively.

Consistent with the concave relationship for the scores, Table 3.1 shows that one standard deviation in the entry exam scores is associated with a drop of 18% (5 percentage points over a mean of 28 percent) in the the likelihood of being classified as basic or unsatisfactory. Similarly, one standard deviation is related to a relatively smaller increase of 7% (5 percentage points over a mean of 72 percent) in the probability of being outstanding or competent.

Figure 3.4: College Entrance Exam and In-Class Teacher Evaluation



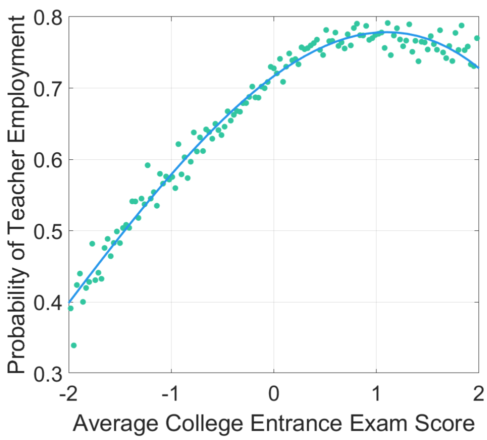
Note: The figures plot the teacher evaluation scores (overall in Figure 3.4a and the portfolio component in Figure 3.4b), within 100 equal-sized bins of the average college entrance exam score and fit estimated lines using all the underlying data. The data consists of teachers evaluated between the years 2004 and 2017. In both figures the sample size is of $N = 63,539$.

Figure 3.5 exhibits correlations between the entrance exam scores and labor market outcomes, which are consistent with the concave patterns described so far. Figure 3.5a plots the probability of working in schools for graduates from teacher colleges versus their college entrance scores. An increase of one standard deviation in scores increases the likelihood of working as a teacher by 10 percentage points (pp.) relative to a baseline of 68%, accompanied by a negative coefficient of 4 pp in the square of the scores. This result suggests that a fraction of teachers in the right tail of the distribution of college preparedness quit the profession by that time.

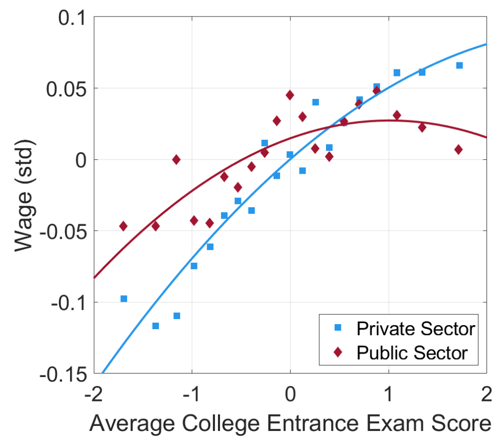
Figure 3.5b shows how hourly wages vary with scores, for teachers working in public and private schools. The slope is much steeper for teachers working in the private sector, and rather flat for teachers working in the public sector. The change in wages in the private sector seems to be driven by both experience and scores, meanwhile for the public sector experience is the most relevant factor since salary increases occur on the base of seniority. Consistently, the coefficients in Table 3.1 show that a standard deviation increase in scores is associated with 0.064σ and 0.024σ of hourly wages for teachers working in the private and public sector respectively (p-value of the difference=0.0001). The magnitude of the coefficient over wages is nearly 3 times higher for the sample of teachers in the private sector, where schools can adjust salaries almost unrestrictedly as teacher productivity changes. The same dynamic does not occur in the public sector where wages are much less flexible and determined primarily by years of service (seniority).

Figure 3.5: College Entrance Exam and Labor Market Outcomes

(a) Employment in Schools



(b) Wages in Schools



Note: Figure 3.5a and Figure 3.5b plot the fraction of teachers employed and their wages (in standard deviation units), respectively, within equal-sized bins of the average college entrance exam score, and fit estimated lines using all the underlying data. The data in Figure 3.5a consists in $N = 240,549$ graduates from teacher colleges in the years 1995 to 2017, who are employed (or not) between 2003 to 2018. The data in Figure 3.5b consists of $N = 117,105$ teachers working in private and public schools in 2011.

The bottom panel of Table 3.1 shows that students' outcomes (like math gains and perceptions about teaching effectiveness) are also positively related to their teachers' entrance exams. The first three columns in the panel show that an increase of one standard deviation in the teacher PSU score is associated with an increase of 0.43σ to 0.29σ in gains in algebra, numbers and geometry. Results are also suggestive of a concave relationship though the negative coefficients are not precisely estimated.

The results for students' perceptions, in the last column of the bottom panel, follow the same pattern. We use factor analysis to produce an index for student perception using the eight categories of teacher effectiveness reported by students. One standard deviation in PSU scores is associated with 0.08σ in the students' perceptions index, with a negative and significant coefficient on the square of the PSU score. These results suggest that the relation between students' outcomes and teacher entrance exams is positive and concave, consistent with the previous set of teacher outcomes.

Table 3.1: College Entrance Exam and Teacher Outcomes

	(1)	(2)	(3)	(4)
Graduation	Years After Enrollment 5 Years	8 Years		
PSU Score	0.070***	0.112***		
	(0.001)	(0.001)		
$(\text{PSU Score})^2$	-0.029***	-0.029***		
	(0.001)	(0.001)		
Observations	105,422	105,422		
Dep. Var. Mean	0.350	0.498		
Exit Exams	Disciplinary Test	Pedagogical Test	Writing Test	Technology Test
PSU Score	0.493***	0.505***	0.282***	0.539***
	(0.005)	(0.005)	(0.009)	(0.014)
$(\text{PSU Score})^2$	0.043***	0.033***	-0.019***	-0.043***
	(0.003)	(0.003)	(0.006)	(0.011)
Observations	35,355	33,409	11,300	5,517
Dep. Var. Mean	0.000	0.000	0.000	0.000
Teacher Evaluation	Overall Score	Portfolio Score	Basic or Unsatisfactory	Outstanding or Competent
PSU Score	0.143***	0.189***	-0.051***	0.051***
	(0.004)	(0.004)	(0.002)	(0.002)
$(\text{PSU Score})^2$	-0.050***	-0.037***	0.020***	-0.020***
	(0.003)	(0.003)	(0.001)	(0.001)
Observations	63,539	63,539	63,539	63,539
Dep. Var. Mean	0.000	0.000	0.283	0.717
Labor Market	Employment	Wages	Private Wages	Public Wages
PSU Score	0.096***	0.047***	0.064***	0.024***
	(0.001)	(0.003)	(0.004)	(0.005)
$(\text{PSU Score})^2$	-0.036***	-0.003	-0.002	-0.010**
	(0.001)	(0.002)	(0.003)	(0.004)
Observations	240,549	117,105	67,909	49,196
Dep. Var. Mean	0.679	0.000	-0.000	0.000
Student Outcomes	Δ Algebra Tests	Δ Numbers Tests	Δ Geometry Tests	Perceptions Index
PSU Score	0.377***	0.292***	0.433***	0.077***
	(0.060)	(0.075)	(0.072)	(0.016)
$(\text{PSU Score})^2$	-0.053	-0.079	-0.078	-0.031**
	(0.048)	(0.061)	(0.053)	(0.012)
Observations	3,756	3,756	3,756	3,612
Dep. Var. Mean	0.000	0.000	0.000	0.000

Note: Table 3.1 reports results from separate regressions of teacher outcomes on college entrance exam scores (labeled 'PSU Score') and its square. The PSU score is expressed in terms of standard deviations in all cases. The table is organized into five panels: graduation, exit exams, productivity measures, labor market outcomes and student outcomes. All results in panels 1-4 come from estimations at the teacher level and include year and teacher specialization fixed effects. Results in panel 5 come from estimations at the student level, with standard errors clustered at the classroom level and controls for teacher experience, class size, school socioeconomic status, and type (public or private). Robust standard errors are in parentheses. ***, ** and * indicate statistical significance at the 1, 5 and 10 percent level respectively.

3.3.1 Teacher Colleges' Value Added

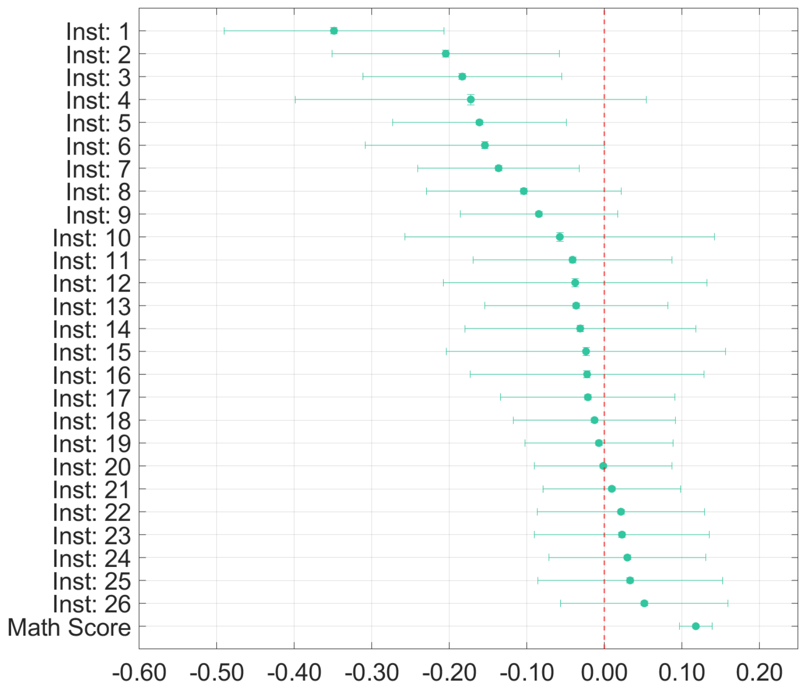
We readily acknowledge that the observed correlation between teachers' entrance exams and later outcomes could be caused by different factors, like sorting or access to different training at teacher colleges.

The main takeaway of section 3.3 is that (independent of those factors) the data shows that if we draw one teacher from the left tail of the population of teachers in Chile, he is likely to display relatively low measures of performance on a host of indicators later on.

In this subsection, we include a complementary exercise examining whether access to higher value-added teacher colleges causes these observed correlations. We combine a regression discontinuity design with data on the population of applicants to teaching colleges from 1977 to 2011, to estimate the value added of teaching colleges in which they enroll versus the next best teacher college. We show our main findings in Figure 3.6.

Figure 3.6 plots the regression discontinuity estimates for each institution on (overall) teaching evaluation scores as a proxy for value added. The results indicate that the institutions' value added cannot be distinguished from zero for most teaching colleges. This result suggests that teaching colleges are not adding differential value to the predicted gap in teacher effectiveness.

Figure 3.6: Education Institutions Value Added to Teacher Evaluation



Note: Figure 3.6 plots the regression discontinuity estimates on teacher evaluation scores by institution, using data on the population of applicants to teaching colleges from 1977 to 2011.

Taken together, the findings of this section indicate that below-average pre-college achievement is systematically associated with lower performance as teachers measured up to thirty and forty years later. Colleges do not seem to be generating the correlation and therefore there might be space to use it for policy.

In the next section, we study the causal effect of recruiting policies that used pre-college scores to attract higher-achieving students to teachers colleges.

3.4 Assessing Teacher Screening Policies

3.4.1 A *Carrot & Sticks* Approach to Recruiting and Screening

The *Beca Vocacion Profesor* (BVP) used college entrance exams to recruit and screen out students entering teacher-colleges. We assess the policy on teacher college's enrollment and medium run outcomes like graduation, exit exams and employment in schools, all measured up to eight years after first implemented.

The results indicate that the BVP policy increased the number of higher-scoring students in teacher colleges, who went on to work in schools eight years later. This finding suggests that the policy was successful at raising the predicted quality of students who entered the teaching profession.

BVP Policy Specifics

The BVP policy started in 2011 and offered full scholarships, stipends and paid semesters abroad for high-scoring test-takers who enroll as freshmen at teacher colleges.[9]

Test-takers with scores in the top 20% (i.e., 600 points or more) were eligible for a full tuition scholarship.[10] Those with scores at approximately the top 5% (700 points or above) were eligible for the full tuition scholarship plus a monthly stipend of about $US150, which was close to 50% of the minimum wage. The top 2% scorers (720 points or higher) would benefit from the tuition, stipend and a paid semester abroad at a prestigious teaching college. For instance, advertisements mentioned a semester abroad at Stanford or in Finland.

The policy also imposed participating teacher colleges to screen out low-scoring applicants. In particular, colleges were required to implement a minimum cutoff score at the national mean of 500 points if they wanted their students to benefit from the BVP.[11] In addition, participating teacher colleges needed to be accredited for at least 2 years at all campuses as determined by the National Commission of Accreditation (CNA).

Empirical Strategy and Data

We use a regression discontinuity (RD) exploiting the BVP score cutoffs to evaluate whether the policy attracted higher-scoring test-takers to teacher colleges.

[9] In practice, the only requirement to be eligible was taking the entrance exam on December 2010 aiming to start as a new first-year student at a teacher college in March 2011. Students already enrolled in teaching careers were not eligible for the scholarship.

[10] If the student had obtained another scholarship called *Beca Excelencia Academica* the cutoff will be 580. These are a handful of students (N=61) and do not change our results if included in the analysis.

[11] The cutoff was lax, allowing colleges to enroll a maximum of 15% of their entering class starting in 2011 with scores below the cutoff.

Our identifying assumptions are standard for RD designs. We assume that there are no other changes occurring at the thresholds that could confound our estimates. In section C.2 we run a series of robustness tests showing that there are no differences in a host of covariates around the thresholds and no evidence of score manipulation.

Our main estimating equation is

$$Y_i = \alpha_0 + \alpha_1 Z_i + f(S_i) + \alpha_2 X_i + \mu_i. \tag{1}$$

where Y_i represents a particular outcome such as enrollment at teacher colleges for the test-taker i. Our parameter of interest is α_1, which is an *intention-to-treat* effect of the BVP policy on the outcome Y_i. The indicator variable Z_i is equal to 1 if the test-taker i scored above a particular threshold and zero otherwise. For simplicity, we estimate separate regressions for the 500, 600, 700 and 720 policy cutoffs.[12] $f(S_i)$ is a smooth function of scores that includes interactions with Z_i to allow for different slopes on each side of the cutoff, and μ_i represents the error term that we cluster within the college entrance exam scores. We also include a set of predetermined variables as controls in X_i, such as test-takers' gender, household income, parents' education, region of residence, and whether they attended a public or private high school. In practice, these control variables have very little effect on our RD estimates and serve mainly to improve precision.

We implement our empirical strategy using individual-level data from the population of test-takers in the country. We first present results for the 2011 cohort, for whom we can estimate the immediate take-up and enrollment effects, but also later outcomes like graduation, exit exams and employment in schools up to 2019. We also compute short-run estimates for later cohorts in the following section.

In Table 3.2 we show descriptive statistics for all test-takers in 2011, organized by information on scores, demographics, and higher education enrollment.[13] A total of 250,758

[12] We also ran a more complex version of Equation 1 to estimate all threshold effects jointly with no differences in our results.

[13] Test-takers complete a survey providing information on their gender, date of birth, household income bracket and parental schooling among other characteristics. We combine this data with the scores information at the individual level, which we merge with administrative records of higher education enrollment coming from the MINEDUC. The enrollment records have information for the population of students enrolled in higher education institutions in the country.

high school graduates[14] took the college entrance exam in December of 2010, aiming to start classes at the beginning of the academic in March of 2011. All of these test-takers were potentially eligible for the BVP had they achieved scores above the policy cutoffs. The scores on each subject (mathematics, language, history and science) have a mean of about 500 points. The college entrance exam score is the math-language average score. The math and language tests are mandatory for all test-takers, while the history and science tests are optional exams.

Test takers are on average 19 years old at the moment of the test, and about half of them are girls. Their parents have on average slightly more than 11 years of completed schooling, and about 40% lives in the capital city. All these statistics are consistent with data coming from national censuses and surveys (CASEN 2016). About 55%, 35% and 10% of the test takers graduated from voucher, public and private high schools, which again are consistent with population figures on enrollment in the country (MINEDUC 2018).

The last panel in Table 3.2 shows the fraction of test takers who enroll in higher education after the exam (in March 2011). A 63% enrolls at any institution, 44% enrolls at colleges and half of that enrolls at the more selective universities.[15] About 20K test takers (8% of the total) enroll at any teacher college and approximately 8K (a 3% of the total) enroll at teacher colleges that were BVP-eligible.

[14]The college entrance exam take-up among high school graduates is high. Each year, about 260K students graduate from high-school in Chile. Test-takers are typically a mix of just-graduated high-schoolers (75%) and graduates from previous years (25%).

[15]These selective universities are non-profit institutions, grouped in the Council of Rectors of the Universities of Chile (CRUCH), which receive students with highest scores in the country.

Table 3.2: Descriptive Statistics for all Test-Takers

Variable	(1) Observations	(2) Mean	(3) Std. Deviation	(4) Min	(5) Max
Scores					
College Exam Score	250,758	501.06	102.34	178	850
Math Score	250,758	501.07	111.27	150	850
Language Score	250,758	501.04	108.34	150	850
Takes History Test	250,758	0.62	0.49	0	1
History Score	154,790	500.41	109.55	150	850
Takes Science Test	250,758	0.56	0.50	0	1
Science Score	139,783	500.52	109.47	150	850
High School GPA Score	248,807	535.81	99.88	208	826
Demographics					
Female	250,758	0.52	0.50	0	1
Age at Test (years)	250,758	19.38	3.17	15	78
Income (1-12 bracket)	250,758	3.40	2.88	1	12
Private Health Insurance	250,758	0.21	0.40	0	1
Father Schooling (years)	215,105	11.45	3.77	0	17
Mother Schooling (years)	233,044	11.30	3.57	0	17
Capital City	248,462	0.40	0.49	0	1
Public High School	248,462	0.35	0.48	0	1
Private High School	248,462	0.10	0.30	0	1
Voucher High School	248,462	0.55	0.50	0	1
Enrollment					
Enroll Higher Education	250,758	0.63	0.48	0	1
Enroll College	250,758	0.44	0.50	0	1
Enroll Selective College	250,758	0.21	0.41	0	1
Enroll Any Teacher College	250,758	0.08	0.28	0	1
Enroll Eligible Teacher College	250,758	0.03	0.18	0	1

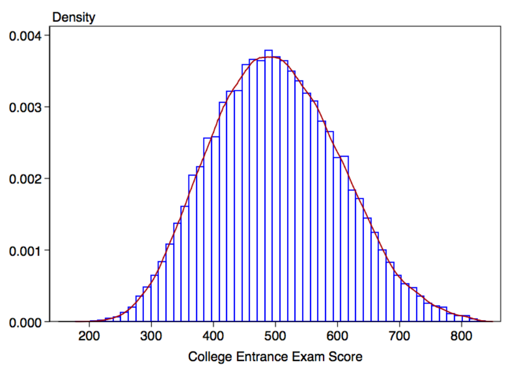
Notes: Table 3.2 shows descriptive statistics for the 250,758 students who took the college entrance exam in December 2010. The college entrance exam score is the math-language average score. The math and language tests are mandatory for all test-takers, while the history and science tests are optional exams. The high school GPA score has valid data for 99.2% of the test-takers (248,807 of 250,758). The age variable corresponds to the age at the moment of the test. The variables of parental schooling have missing information due to both non-response and test-takers not knowing the answer. The Capital City variable indicates whether the test-taker lives in the capital of the country at the moment of the test, while the variables Public, Private and Voucher High School indicate the type of high school from which the test-takers graduated. These last four variables have a response rate of 99.1%. The enrollment variables come from population records collected by the Ministry of Education, indicating whether individuals were enrolled during the academic year 2011. Enroll in Higher Education takes value one if the test-taker enrolled at any institute or university. Enroll College is equal to one if the test-taker enrolled at any college; enroll selective does the same if the test taker enrolled at universities belonging to the *Consejo de Rectores*, a group of non-profit institutions that enroll the students with highest scores in the country. Enroll at any teacher college (TC) takes value one if the test taker enrolled in any education major in the country, and Enroll Eligible TC does the same for enrollment at eligible teacher colleges.

Results

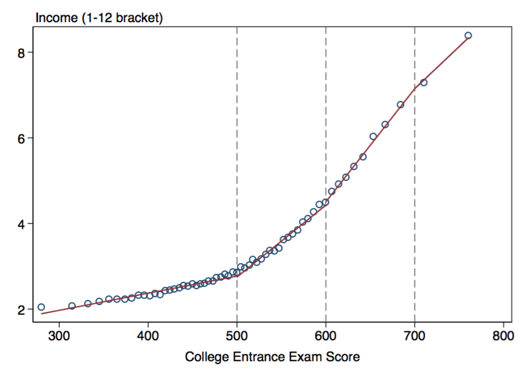
Our main results show that the policy attracted higher-scoring test-takers to teacher colleges. Figure 3.7 summarizes the first set of findings. Figure 3.7a and Figure 3.7b are robustness tests, showing no manipulation of the running variable (the college entrance exam score) and that other covariates, such as household income behave smoothly near the policy thresholds. Figure 3.7c and Figure 3.7d illustrate effects on enrollment at any teacher colleges (TC) and at eligible TC, respectively. Both Figures reveal a sharp discontinuity at the 500 and 600 points and a smaller increase at 700 points, indicating that test-takers with very similar scores around those cutoffs experienced a different likelihood of enrolling at teacher colleges.

Figure 3.7: Main Results

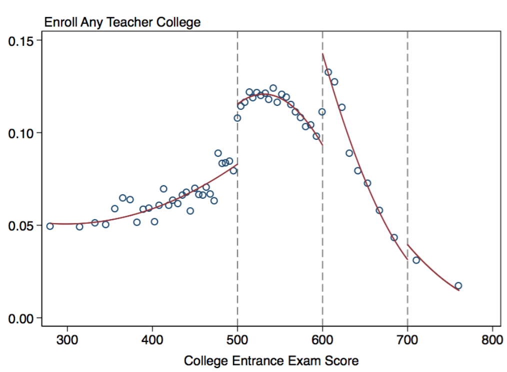
(a) Score Distribution



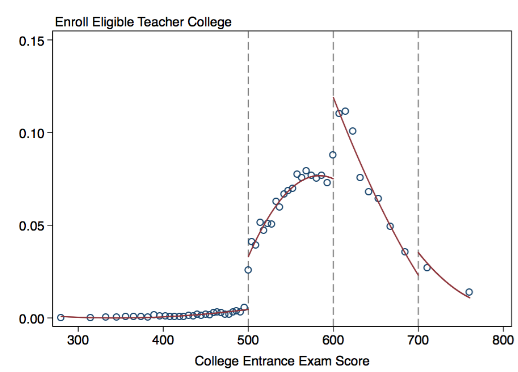
(b) Household Income



(c) Enrollment at Teacher Colleges



(d) Enrollment at Eligible Teacher Colleges



Note: Figure 3.7a plots the distribution of scores for all test takers. Figure 3.7b, Figure 3.7c and Figure 3.7d plot the mean of the y-axis variable within bins of scores, and fit estimated lines using all the underlying data. The sample size in each graph in Figure 3.7 is of N=250,758 observations.

Table 3.3 provides the regression analog of graphs 3.7c and 3.7d in panels 1 and 2. The columns report the RD estimates from Equation 1 at the 500, 600, 700 and 720 cutoffs, with MSE-optimal bandwidths (Cattaneo et al., 2018) for each threshold. These are our preferred estimates, which are robust to different bandwidths and specifications.

The estimates from Panel 1 show sizable effects near cutoffs. The magnitude of the estimates represents relative increases of 37% at 500 points (3.2pp over 8.6pp just below the cutoff), 37% at 600 (3.5pp over 9.5pp) and 100% at 700 (2.5pp over 2.5pp). We find a precise null effect at the highest cutoff of 720 points.

Panel 2 in Table 3.3 shows similar point estimates for the respective cutoffs on enrollment at eligible teacher colleges. The main difference is that the enrollment rate at eligible teacher colleges just before the cutoff of 500 points is zero, consistent with the policy design.

Approximating Aggregated Effects

We now add additional microdata data for the population of test-takers in 2010 to approximate impacts beyond the local effects estimated above. We use the policy time variation to compare outcomes along the distribution of test scores before and after the BVP was implemented.

Figure 3.8 motivates the analysis with visual evidence showing how the policy shifted the distribution of scores in teaching colleges. Figure 3.8a shows distributions of entrance exam scores for the 2010 cohort (before the BVP) and the 2011 cohort (after), by enrollment at eligible and non-eligible teacher colleges. The distribution of scores at eligible teaching colleges shifts markedly to the right after the policy, while it remains the same for students at non-eligible institutions.

Table 3.3: BVP Effects on Enrollment

Panel 1. Dep. Variable: Enrollment at Teacher Colleges

	(1)	(2)	(3)	(4)
RD_Estimate	0.032***	0.035***	0.025**	-0.010
	(0.004)	(0.007)	(0.009)	(0.008)
Mean Just Below Cutoff	.086	.095	.025	.032
Optimal Bandwidth	48.3	34.3	26.3	34.5
Cutoff Value	500	600	700	720
Effective Observations	86,457	40,559	8,423	8,210
All Observations	250,758	250,758	250,758	250,758

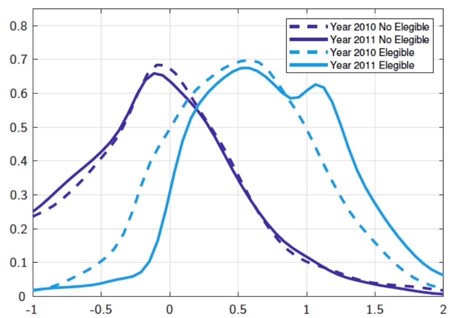
Panel 2. Dep. Variable: Enrollment at Eligible Teacher Colleges

	(1)	(2)	(3)	(4)
RD_Estimate	0.033***	0.029***	0.022**	-0.008
	(0.002)	(0.006)	(0.008)	(0.007)
Mean Just Below Cutoff	.005	.073	.022	.027
Optimal Bandwidth	41.8	30.7	28.4	33.3
Cutoff Value	500	600	700	720
Effective Observations	75,825	36,437	9,178	7,719
All Observations	250,758	250,758	250,758	250,758

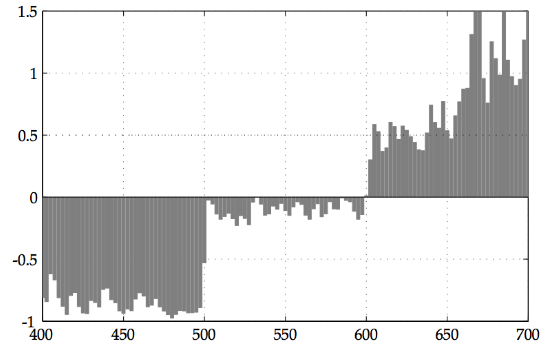
Notes: Table 3.3 shows regression discontinuity estimates from Equation 1 using local polynomial regressions at the 500, 600, 700 and 720 cutoffs. The dependent variables are Enrollment at Teacher Colleges and Enrollment at Eligible Teacher Colleges in Panels 1 and 2, respectively. All estimates are computed using a triangular kernel and robust variance estimators, with bandwidths that are data-driven MSE-optimal. All regressions control for the demographics described in Table 3.2. These are our preferred estimates, which are robust to different bandwidths and specifications.

Figure 3.8: Aggregate Effects on the Distribution of Scores

(a) Scores at eligible and non-eligible teacher colleges, before & after the BVP



(b) Change in Probability of Choosing a Teacher College at Eligible Institutions



Note: In Figure 3.8a the continuous (dotted) line shows the distribution before (after) the BVP policy. The ■ color depicts the distribution for non-eligible colleges while the ■ does the same for eligible colleges. Figure 3.8b shows the before-after change in the probability of enrolling in an eligible teacher college conditional on enrollment, along the test score distribution.

Because all the action takes place at eligible institutions, we also examine the before-after change in the probability of choosing teaching as a major, conditional on enrollment. We plot the results along the distribution in Figure 3.8b. The figure illustrates an increase close to 40% in the probability of enrollment at an eligible teaching college around 600 points, which increases above 100% at 700 points. Under 500 points the probability decreases by almost 100%, because students with scores below that threshold could not enroll using the BVP policy.

Our findings show that the BVP policy was successful in screening out lower-scoring test-takers and attracting higher-scoring applicants. However, it is useful to put the results in perspective. While the results show that the BVP raised choice probabilities significantly, the number of students at those margins is still relatively modest compared to the population of students in teaching colleges. We estimate that about 1,000 additional students entered teaching colleges from the top 30% of the distribution and that the screening restrictions reduced the bottom tail of the distribution by about 4,000. While meaningful, these figures must be considered taking into account that the total number of freshmen students at teachers' colleges is close to 20,000.

Medium Run Effects. Our previous results show that the policy attracted higher-scoring test takers to enroll at teachers colleges. In this subsection we examine results on a host of medium run outcomes described before, like graduation, exit exams and employment in schools, all measured up to eight years after initial enrollment.

Table 3.4 reports our estimates. The estimates in Panel 1 show that the policy increased employment at schools of the higher-scoring test-takers near the cutoffs of 500 and 600 points. The effect sizes are of 12% at 500 (1.2pp over 6.4pp) and 34% at 600 (2.3pp over 6.7pp).

Panels 2 to 4 present estimates on college graduation, and the likelihood of taking the exit exam and the teacher evaluation. We find zero effects on these outcomes, with small standard errors. These precise null effects suggest that higher achieving students graduated and took the teacher exams as we would have predicted using the college entrance scores.[16]

Taken together, these results indicate that the BVP policy increased the number of higher-scoring students in teacher colleges, who went on to work in schools eight years later. This finding indicates that the policy was successful at raising the predicted quality of students who entered into the teaching profession.

The rest of the early productivity indicators in Table 3.4 suggest that the higher achieving students graduated and took the teacher exams as we would have predicted using the college entrance scores. This finding is most useful from a policy perspective, because it suggests that the predicted relationship between pre-college academic achievement and teacher medium run outcomes is invariant, and can be used in policy design.

3.4.2 Policy Effects Over Time

In the previous section we focused our analysis on the first cohort that benefited from the BVP policy. In this section we study the BVP policy effects for different cohorts over time, and examine how the results change when new policies are introduced.

We report results for ten different cohorts, from 2008 to 2018 in Figure 3.9 and Table 3.5. For each cohort, we estimate equation Equation 1 and report regression discontinuity estimates near the 500, 600 and 700 cutoffs.

Our main findings indicate that the effects of setting the minimum scores at the 500

[16]The take-up rates of the exit exam and the teacher evaluation are very low because these exams were not mandatory for the cohort of test takers under analysis.

Table 3.4: BVP Effects on Medium Run Outcomes (8 years)

Panel 1. Dep. Variable: Employment at Schools

	(1)	(2)	(3)	(4)
RD_Estimate	0.012***	0.023***	0.006	-0.010
	(0.003)	(0.005)	(0.008)	(0.007)
Mean Just Below Cutoff	.064	.067	.033	.029
Optimal Bandwidth	60.7	52.5	32.1	38.3
Cutoff Value	500	600	700	720
Effective Observations	107,517	62,410	10,612	9,042
All Observations	250,758	250,758	250,758	250,758

Panel 2. Dep. Variable: Graduation

	(1)	(2)	(3)	(4)
RD_Estimate	-0.000	0.003	-0.013	0.019
	(0.006)	(0.008)	(0.021)	(0.020)
Mean Just Below Cutoff	.522	.575	.600	.626
Optimal Bandwidth	63.9	54	31.8	43.6
Cutoff Value	500	600	700	720
Effective Observations	112,474	63,569	10,328	10,523
All Observations	250,758	250,758	250,758	250,758

Panel 3. Dep. Variable: Takes Exit Exam

	(1)	(2)	(3)	(4)
RD_Estimate	0.006**	-0.003	-0.006	0.006
	(0.002)	(0.003)	(0.004)	(0.005)
Mean Just Below Cutoff	.019	.024	.013	.004
Optimal Bandwidth	66.9	47.2	47.0	34.3
Cutoff Value	500	600	700	720
Effective Observations	117,261	55,975	16,288	8,034
All Observations	250,758	250,758	250,758	250,758

Panel 4. Dep. Variable: Takes Teacher Evaluation

	(1)	(2)	(3)	(4)
RD_Estimate	0.004***	0.003*	-0.001	-0.005
	(0.001)	(0.001)	(0.001)	(0.003)
Mean Just Below Cutoff	.004	.006	.002	.005
Optimal Bandwidth	68.2	62.2	44.2	23
Cutoff Value	500	600	700	720
Effective Observations	119,378	73,348	15,287	5,279
All Observations	250,758	250,758	250,758	250,758

Notes: Table 3.4 shows regression discontinuity estimates from Equation 1 using local polynomial regressions at the 500, 600, 700 and 720 cutoffs. The dependent variables are Employment at Schools, Graduation, and Taking the Exit Exam and Teacher Evaluation in Panels 1 through 4, respectively. All estimates are computed using a triangular kernel and robust variance estimators, with bandwidths that are data-driven MSE-optimal. The regressions control for all the demographics described in Table 3.2.

cutoff remain high and persistent for cohorts over time. Second, the effects at the 600 cutoff tend to vanish when another 'free college' policy kicks in; and third, there is essentially no action at the top (at the 700 cutoff) no matter what policy was in place.

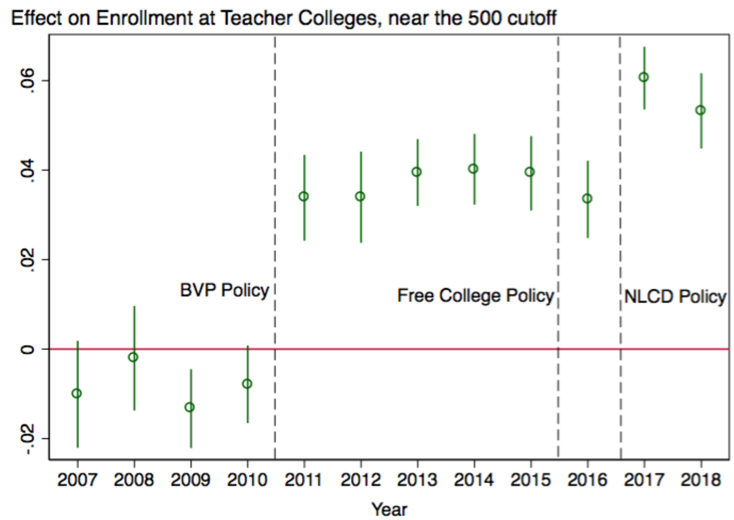
We plot the effects near the 500 cutoff over time in Figure 3.9a. As expected, the figure shows no effects for cohorts 2008 to 2010, before the BVP policy was implemented. Once the BVP was implemented, in 2011, enrollment at teacher colleges jumped 3.2pp, as we described in the previous section. The magnitude of this effect is similar for the next five cohorts until 2017, when the NLCD policy (described in section 3.2) was first implemented. We discuss NLCD policy and its results in detail in the next section.

Figure 3.9b plots the RD estimates near the 600 points threshold. As in Figure 3.9a, the figure shows zero effects before the BVP policy was implemented and positive effects after. In this case, the effects diminish in more recent years and disappear in 2016. The country started with a nationwide policy to make tuition free, which was fully implemented in 2016 (Bucarey, 2018). This *free college* policy appears to naturally have reduced the financial incentives generated by the BVP. Consistently the regression discontinuity estimates show that the effectiveness of the policy was significantly diminished for the newer cohorts. These results are aligned with contemporaneous work by Castro-Zarzur et al. (2019) and Castro-Zarzur and Mendez (2019).

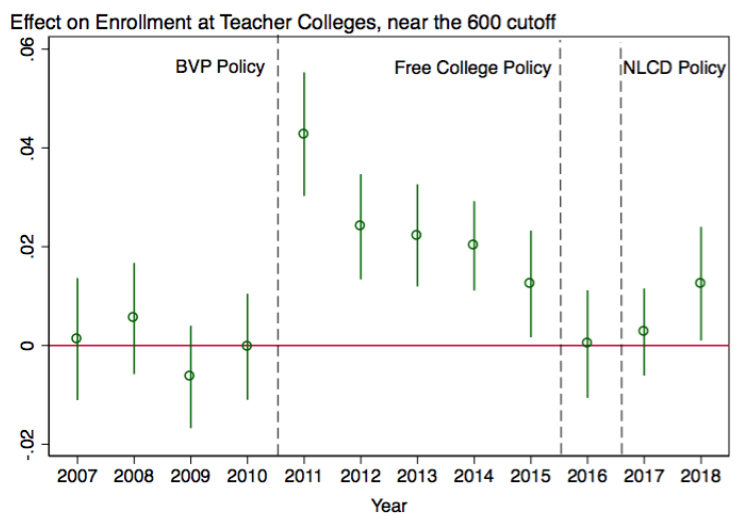
Finally, we find precisely estimated zero effects for higher-scoring test takers near the 700 and 720 cutoffs. Figure 3.9c graphs the estimates for the 700 threshold, while Table 3.5 presents the estimates for both 700 and 720 cutoffs. This finding serves as a reminder that recruitment incentives are only as good as the next best option and that high-achieving students have many good alternatives, and it is harder to move them toward teaching careers.

Figure 3.9: Effects on Enrollment over Time

(a) Figure 3.9a



(b) Figure 3.9b



(c) Figure 3.9c

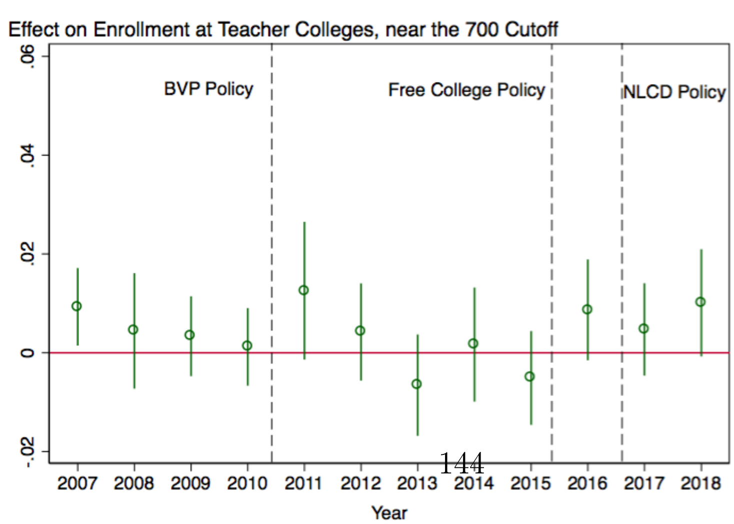


Table 3.5: BVP Effects on Enrollment over Time 2008-2018

Panel 1. Dep. Variable: Enrollment at Teacher Colleges near the 500 Cutoff

	(1) 2008	(2) 2009	(3) 2010	(4) 2011	(5) 2012	(6) 2013	(7) 2014	(8) 2015	(9) 2016	(10) 2017	(11) 2018
$\widehat{\alpha_1}$	-0.002	-0.013***	-0.008*	0.034***	0.034***	0.039***	0.040***	0.039***	0.033***	0.061***	0.053***
	(0.006)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
$\widehat{\alpha_0}$	0.125***	0.130***	0.134***	0.085***	0.087***	0.068***	0.058***	0.052***	0.063***	0.042***	0.051***
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)
Eff Size	-.016	-.102	-.059	.396	.39	.58	.695	.752	.532	1.451	1.044
Band	50	50	50	50	50	50	50	50	50	50	50
Cutoff	500	500	500	500	500	500	500	500	500	500	500
N	77,865	87,108	90,169	90,450	84,773	86,341	86,955	90,065	90,725	93,455	97,357

Panel 2. Dep. Variable: Enrollment at Teacher Colleges near the 600 Cutoff

	(1) 2008	(2) 2009	(3) 2010	(4) 2011	(5) 2012	(6) 2013	(7) 2014	(8) 2015	(9) 2016	(10) 2017	(11) 2018
$\widehat{\alpha_1}$	0.005	-0.006	-0.000	0.043***	0.024***	0.022***	0.020***	0.012**	0.000	0.003	0.013**
	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.004)	(0.006)
$\widehat{\alpha_0}$	0.094***	0.098***	0.093***	0.096***	0.087***	0.074***	0.069***	0.074***	0.070***	0.068***	0.075***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.004)
Eff Size	.058	-.065	-.003	.448	.277	.3	.291	.169	.004	.04	.168
Band	50	50	50	50	50	50	50	50	50	50	50
Cutoff	600	600	600	600	600	600	600	600	600	600	600
N	52,485	58,302	60,345	59,437	59,044	60,076	59,428	64,005	60,442	62,200	64,579

Panel 3. Dep. Variable: Enrollment at Teacher Colleges near the 700 Cutoff

	(1) 2008	(2) 2009	(3) 2010	(4) 2011	(5) 2012	(6) 2013	(7) 2014	(8) 2015	(9) 2016	(10) 2017	(11) 2018
$\widehat{\alpha_1}$	0.004	0.003	0.001	0.013*	0.004	-0.007	0.002	-0.005	0.009*	0.005	0.010*
	(0.006)	(0.004)	(0.004)	(0.007)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
$\widehat{\alpha_0}$	0.013***	0.012***	0.013***	0.030***	0.024***	0.031***	0.034***	0.027***	0.024***	0.023***	0.022***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.003)
Eff Size	.339	.272	.094	.423	.175	-.209	.049	-.19	.36	.208	.451
Band	50	50	50	50	50	50	50	50	50	50	50
Cutoff	700	700	700	700	700	700	700	700	700	700	700
N	15,426	17,509	17,775	17,586	18,692	18,403	18,097	18,405	17,556	17,677	18,864

Panel 4. Dep. Variable: Enrollment at Teacher Colleges near the 720 Cutoff

	(1) 2008	(2) 2009	(3) 2010	(4) 2011	(5) 2012	(6) 2013	(7) 2014	(8) 2015	(9) 2016	(10) 2017	(11) 2018
$\widehat{\alpha_1}$	0.002	-0.002	0.007*	-0.007	0.006	0.004	0.002	0.013**	-0.003	0.001	0.006
	(0.005)	(0.004)	(0.004)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)
$\widehat{\alpha_0}$	0.008**	0.010***	0.006*	0.027***	0.022***	0.022***	0.026***	0.018***	0.023***	0.022***	0.020***
	(0.004)	(0.003)	(0.003)	(0.005)	(0.003)	(0.005)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Eff Size	.235	-.232	1.327	-.256	.253	.183	.071	.759	-.114	.032	.289
Band	50	50	50	50	50	50	50	50	50	50	50
Cutoff	720	720	720	720	720	720	720	720	720	720	720
N	10,720	12,166	12,630	12,488	13,123	12,864	12,496	12,755	12,270	12,275	13,275

Notes: Table 3.5 shows regression discontinuity estimates from Equation 1 using local polynomial regressions at the 500, 600, 700 and 720 cutoffs. The dependent variable is Enrollment at Teacher Colleges for every regression. All estimates are computed using a triangular kernel and robust variance estimators, with bandwidths that are data-driven MSE-optimal. The regressions control for high school GPA and all the demographics described in Table 3.2.

3.4.3 A Mandatory Screening Policy

The NLCD (*Nueva Ley de Carrera Docente*)[17] was enacted in 2017. While the screening component of the BVP policy prevented *participating* teacher colleges from admitting applicants with scores below the national mean, the NLCD policy extended the requirement to *all* teacher colleges in the country.[18]

The data suggest that the NLCD screening policy was successful at reducing the fraction of low-scoring students enrolled in teacher colleges. As illustrated in Figure 3.9a the threshold crossing effect at 500 points jumps markedly in 2017 when the NLCD takes place, even though the BVP policy had already been in place for six years.

We argue that the difference between the coefficients in 2016 and 2017 gives us the NLCD effect on top of the BVP policy. Both Figure 3.9a and Table 3.5 show that the coefficient was 3.3 pp in 2016 and jumped to 6.1 pp in 2017 (and 5.3pp in 2018). These parameters suggest that the NLCD policy reduced significantly the fraction of low scorers who enrolled at teacher colleges. As a robustness (placebo) test, the results in Figures 3.9b and 3.9c and their regression analogs in Table 3.5 show no effects at the higher scoring cutoffs of 600 and 700 points.

3.5 Towards Data-Driven Screening Policies

We now turn to assessing whether the use of data-driven algorithms may enhance the screening procedures planned for the future by the MINEDUC policymakers.

The future screening procedures. The NLCD policy states two screening procedures for the future, one for the admissions in years 2023 through 2025 (which we call P23) and another from 2026 onward (which we label P26).

According to P23, all applicants to teaching colleges must either achieve an average

[17]The NLCD is a broad policy aimed at enhancing the system of professional development for teachers in the country. The law is available in the Congress' website here.

[18]The requirements for the screening policy affect admissions to all teacher colleges and are designed to be implemented gradually. During the first six years (2017-2022), the screening policy (P17) requires students to either achieve an entrance exam score above the 50th percentile of the distribution when averaging math and language or, alternatively, students can also avoid the screening rule if their high school GPA is above the 70th percentile within their high school graduating cohort.

entrance exam score above 525 points, or have a high school GPA above the 80th percentile of their cohort. Alternatively, applicants with less than 525 points but more than 500 points might also enroll if they also have a high school GPA above the 60th percentile.

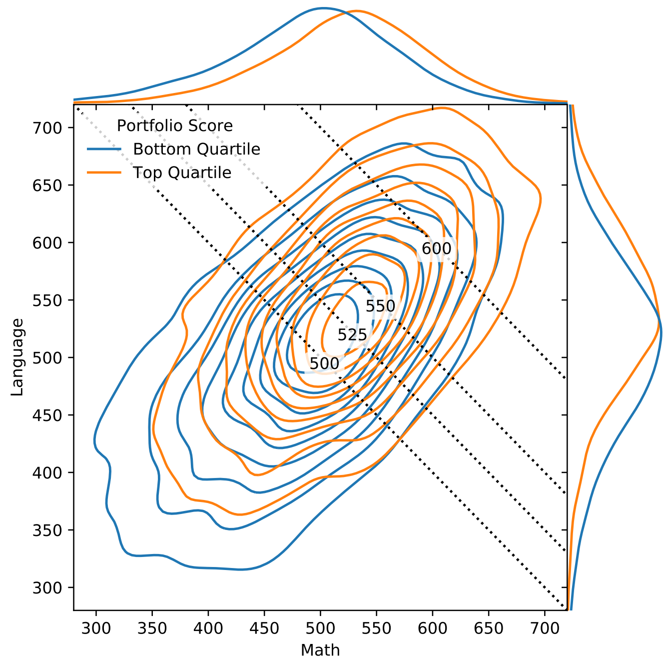
From 2026 onward (P26), the NLCD requires applicants to have either entrance exam scores above 550 points, or belong to the top 10% of their cohort high school GPA. If a student's GPA is in the top 30% *and* her average score is at least 500, then that student may also enroll at teaching colleges.[19]

Using machine learning tools. The screening rules aim to recruit promising students into the teaching profession, and therefore their effectiveness critically hinges on their predictive capacity. The availability of rich individual-level data and the low cost of prediction deems the use of machine learning (ML) algorithms as a natural way of complementing and augmenting the teacher-selection procedure, just as in other recruitment use cases (Agrawal et al., 2018). In our setting, these methods could serve as tools to augment the current cutoff-based selection rules.

Figure 3.10 motivates the promise of a data-driven approach. The figure plots the bivariate density of math and language scores for teachers in the top (in orange) and bottom (in blue) quartiles of the portfolio evaluation, which we use here as a proxy for teacher performance. The graph shows a substantial overlap in the level curves of the score densities for both teachers categorized as top or low performers. This *prima facie* evidence suggests that there is information in the data that would be lost if classifiers rely on single-dimensional cutoff rules (Elizondo et al., 2012).

[19] All of these conditions are designed as minimal requirements for admission to teacher colleges. Each institution is allowed to consider stricter conditions, and define the number of vacancies or slots and application mechanisms. However, all the requirements must be informed before the beginning of the admission process each year.

Figure 3.10: Level-curves for math and language scores, by portfolio evaluations



Note: This figure shows the level curves of the bivariate density of math and language test scores, conditional on groups defined by portfolio scores. Blue lines correspond to teachers who performed in the bottom quartile of the portfolio-score distribution, while orange lines depict those who performed in the top quartile of this distribution. Nine level curves divide each distribution into homogeneous segments that contain 10% of the data. Relevant score cutoffs are plotted as referential dotted lines.

Procedures and Data. We show how simple classifiers can outperform MINEDUC's screening procedures. Our first exercise avoids fitting complex classifiers and is limited to families of algorithms that are close to the cutoff rules used by MINEDUC: shallow classification trees that only use math and language scores as features.

Our decision-tree classifiers use the applicants' individual math and language test scores, along with their arithmetic and geometric means, as training features.[20]

Our full sample consists of roughly 50K observations (N=49,274), corresponding to all teachers who took their college entrance exam and exit exams in 2004-2018. We randomly split this sample into a training set (70% of the sample) and a testing set (the remaining

[20]The geometric mean $\sqrt{Math \cdot Language}$ is a measure of the complementarity between these two scores. For example, a high-math-score applicant might not be a good teacher if she performs poorly in the language test and is not able to communicate well when lecturing. This applicant could have a high arithmetic, but a low geometric mean and, in principle, would be more likely to be screened out in our procedure vis-à-vis MINEDUC's screening rules.

30%). Our target variable is an indicator that equals 1 when a teacher scores at least -0.5 in both standardized exit exams. We picked this threshold to create fairly well-balanced groups, with 57% of the data belonging to the positive class.

Results. Figure 3.11 shows that our classification tree outperforms all the different government recruiting policies.

The figure in the left panel displays in blue the testing accuracy of our tree (in the y-axis), trained with 1 to 14 depth hyperparameters (in the x-axis). The horizontal red lines correspond to the full-sample accuracy achieved by the different screening rules enacted by MINEDUC.[21] We include a shaded area that denotes the depth interval going from four (which maximizes accuracy) to seven (which achieves maximum sensitivity).

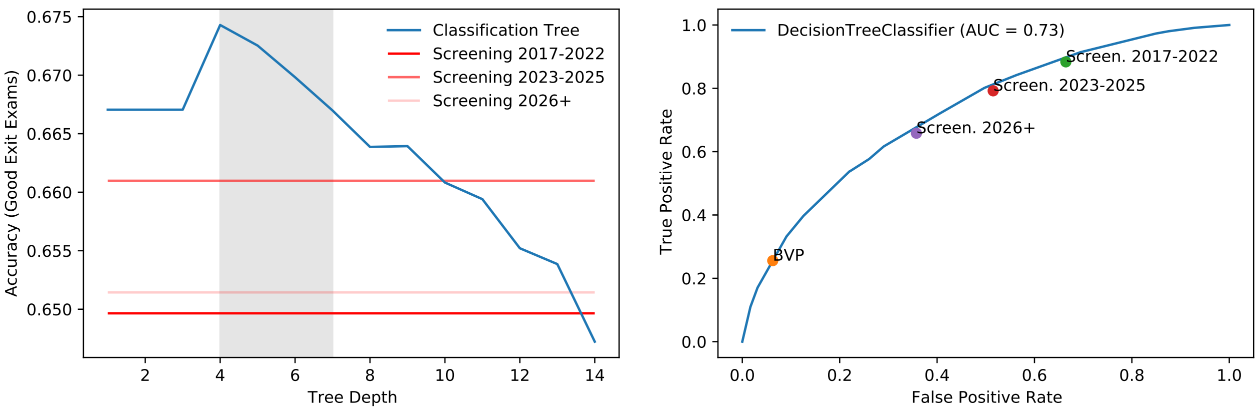
Our shallow classification tree achieves a higher accuracy compared to all NLCD screening policies by between 1.7-2.8 percentage points (a 2.5-4.2% increase). It also outperforms the BVP policy (not depicted in the left graph) by over 12 percentage points (a 22% increase).

The graph in the right panel depicts the ROC curve of our trained classification tree with four levels of depth, plotting the true positive and the false positive rates in the testing sample. The annotated dots correspond to the location of MINEDUC's policies in this space.

The results show that our classifier achieves a higher performance than the government recruitment policies, because its ROC curve lies above all of the MINEDUC's alternatives. In addition, the area under the ROC curve (AUC) is 73% of the unit square, which is higher than the standard for predicting behavioral outcomes (Chalfin et al., 2016).

[21] The accuracy of the BVP policy is considerably lower (55%) and is excluded from the figure for aesthetic purposes.

Figure 3.11: Classification Tree performance, compared to MINEDUC policies



Note: [Left Panel] This figure shows the testing-accuracy of our classification tree, trained with different depths. Red lines correspond to the full-sample accuracy achieved by the different screening rules enacted by MINEDUC (considering GPA rankings as well). The shaded area denotes the interval going from the depth that achieves maximum accuracy to the depth that achieves maximum sensitivity (not shown in this graph). The accuracy of the BVP policy is considerably lower (55%) and is excluded from the figure for aesthetic purposes. [Right Panel] This figure shows the True Positive and False Positive rates along the ROC curve of our trained classification tree, with 4 levels of depth, in the testing sample. The annotated dots correspond to the location of MINEDUC's policies in this space. All policies are below the ROC curve, which achieves an area of 73% of the unit square.

These indicators of good performance are conservative because we kept the procedure simple. First, we are restricting our optimization space to shallow classification trees, which are low-complexity and easily interpretable algorithms. Second, we do not rely on additional features, such as the high school GPA that is used in the policy screening rules. We relax these procedures below.

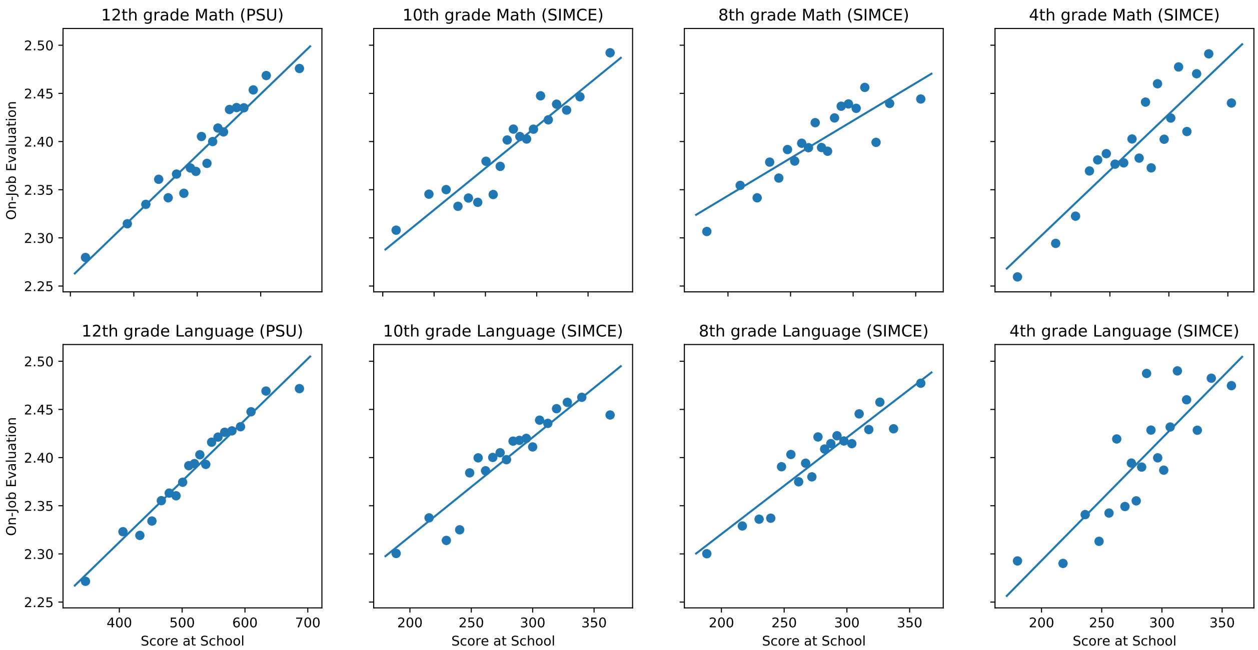
Including additional pre-college characteristics. Using the two test scores with a 4-step, if-else resulting rule[22] makes this previous classifier a useful policy tool, especially when it comes to having applicants understand the screening mechanism.

The algorithm performance might be enhanced by including additional features that also determine future performance as a teacher. We explore if other pre-college characteristics, unaffected by the screening policy, are highly correlated with on-the-job performance. We include teachers' own test scores on standardized exams taken back in 4th, 8th, 10th and 12th grades and correlate them with their teacher evaluations as adults in Figure 3.12. The

[22] Our best classification tree in terms of out-of-sample accuracy requires 4 levels of depth, which translates into 4 "if-else" branching points when classifying a data point.

graphs in Figure 3.12 show a high correlation between teachers' own earlier scores and their performance later on, suggesting that additional features could indeed increase the predictive power of our classification algorithms.

Figure 3.12: School Test Scores Correlate with On-Job Performance



Note: This figure shows binned scatter plots and fitted regression lines. The dependent variable is the raw on-the-job evaluation score (portfolio). The independent variables are their own math and language test scores on standardized exams (called SIMCE exams) taken in 4th, 8th, 10th grades, and the PSU exam taken at the end of 12th grade.

A More Complex Classifier. We now train a more complex classifier using a richer feature space to enhance the accuracy of our data-driven screening rule. Besides features, our previous algorithm might also be too constrained in terms of flexibility,[23] so we train a 16-input Multi-Layer Perceptron (MLP) with 3 hidden layers. We also use random dropout to prevent overfitting (Srivastava et al., 2014), as the number of parameters in our network (1666) is large relative to our data (19,271 observations without missing features, 70% of which are used in training).

[23]The family of 4-layer binary classification trees is limited in its capacity to classify datasets that are not "purely" separable in 4 steps.

Table 3.6 shows that our MLP outperforms our previous simple classification tree and all the government recruiting policies.[24] The table displays key performance metrics (in columns) for classifiers (in rows), starting with the MLP and decision tree classifiers, and following with the MINEDUC's recruiting and screening policies: BVP, NLCD and the future NLCD policies P23 and P26.

The MLP exhibits two percentage points of additional accuracy than our previous classification tree, and beats the F1-score of the latter by two percentage points as well. Sensitivity and Precision are individually outperformed by other "extreme" rules, such as the 2017-2022 screening rule, which most applicants comply with (and therefore high sensitivity comes at the cost of a high false-positive rate), or BVP, which screens out the vast majority of applicants (and therefore high precision comes at the cost of a low sensitivity). Overall, the estimated MLP would be preferred in our setting, and it would be followed by our estimated classification tree.

Table 3.6: Performance comparison among different classifiers

Classifier	**Accuracy**	**Sensitivity**	**Precision**	**F1-Score**
Multi-Layer Perceptron	0.69	0.83	0.70	0.76
Classification Tree	0.67	0.80	0.68	0.74
Beca Vocacion de Profesor	0.55	0.26	0.85	0.39
Screening 2017-2022 (NLCD)	0.65	0.88	0.64	0.74
Screening 2023-2025 (P23)	0.66	0.79	0.67	0.73
Screening 2026+ (P26)	0.65	0.66	0.71	0.68

Note: This table shows various performance metrics for different classifiers. All metrics are computed in the testing data, and use as target variable an indicator that equals 1 when both standardized exit exams are at least -0.5. Accuracy: fraction of correctly classified observations. Sensitivity: true-positive rate (fraction of "good" teachers that were predicted to be "good"). Precision: positive predictive value (fraction of predicted "good" teachers that are effectively "good"). F1-Score: harmonic mean between sensitivity and precision.

The results should be interpreted with caution; this is a predictive exercise, with a specific sample where data on both outcomes (exit exams) and features (the predictors) are available. With that said, the findings of this section support the idea that data-driven, machine learning methods might be a promising way to aid screening and recruitment policies. Importantly, we find that these methods can outperform traditional cutoff-based mechanisms

[24]Other classifiers with richer features and more complicated architectures might outperform our MLP as well.

without sacrificing interpretability nor relying on complicated sets of input features.

3.6 Conclusions

In this paper, we put together historical datasets with administrative records on the population of teachers in a middle-income country to show that (i) pre-college academic achievement is systematically related to a series of measures of long-run teacher outcomes; (ii) teacher recruitment policies can bring higher-scoring individuals to teacher colleges, and they will work later in schools; (iii) it remains challenging to attract very top students to the profession, and (iv) there might be an opportunity to combine improved algorithms with better data to improve hiring in the teacher labor markets.

In our analysis, we find a concave relationship between pre-college academic achievement and later teacher outcomes, which we interpret as evidence that in a developing country context such as Chile, basic academic competency might be a necessary condition to be an effective teacher.

We readily acknowledge that the observed correlation between teachers' entrance exams and later outcomes could be caused by sorting, selection into teaching or other factors. [25]

The main takeaway of this descriptive exercise is that (independent of those factors) when drawing one teacher from the left tail of the population of teachers in Chile, he is likely to display relatively low measures of performance on a host of indicators later on. Discouraging low-ability individuals (and motivating more high-ability students) to enter teacher colleges might at least weakly increase future teacher quality and the status of the profession (OECD, 2005; Lankford et al., 2014).

We then evaluate two policies implemented in Chile that look to shape the pool of students entering teaching colleges by screening out low-performing students or setting incentives for high-performing students based on their pre-college academic achievement.

The first policy, implemented in 2011, offered full tuition subsidies for high-scoring appli-

[25] Tincani and coauthors (see,.e.g., Tincani, 2014; Tincani et al., 2016; Tincani, 2021) explicitly model the sorting process and simulate related teacher policies in Chile. We see our paper as complementing this structural work with empirical descriptive and causal evidence of the relationship between pre-college academic achievement and later outcomes.

cants and also required participating institutions to reject low-scoring students. We evaluate this *'carrots and sticks'* policy using a regression discontinuity based on the eligibility score cutoffs for high and low-scoring applicants. Our findings show that the policy increased the number of higher-scoring students in teacher colleges, with the highest effects at the lower cutoffs of the college entrance distribution (about 37% of an effect size). This finding serves as a reminder that recruitment incentives are only as good as the next best option and that high-achieving students have many good alternatives, so it is harder to move them toward teaching.

In addition, many higher education options became tuition-free as part of another government policy years later (2016). This new policy changed relative prices and generated suggestive evidence helping to disentangle effects attributed to the components described above. In practice, we find that the effectiveness of the financial incentives at the 600 cutoff was significantly reduced. The results suggest that inducing colleges to voluntarily exclude the lowest-performing students was the most effective aspect of the policy. The results also highlight that the effectiveness of targeting highly talented students with recruiting efforts is highly context-dependent and expensive because they have many other valuable options.

A second screening policy implemented in 2017 barred all teaching colleges from admitting students with below-average scores unless they had a very high GPA. Our regression discontinuity estimates near this cutoff suggest that the policy screening out about half of the least academically prepared applicants near the threshold. To evaluate the policy relevance of a minimum standard for entering teaching colleges, we develop a model that classifies potential teacher productivity based on the rich set of pre-college information including GPA course transcripts and entrance exam scores. This model provides feasible cutoff rules that exclude students with a higher chance of being a low-performing teacher.

In both policies studied, the most effective aspect of the policy comes with screening policies aimed at excluding prospective students with scores below the median rather than with recruiting the highest-ability students. This is both a function of the higher ability to identify low-productivity teachers from the bottom of the academic achievement distribution and that it is difficult to recruit high-ability students. Taken together, this suggests that increasing the predicted productivity of a cohort of future teachers can be increased first by excluding the lower tail of the distribution of academic achievement and potentially using any resources saved to incentivize a large group of simply above-average students to enroll in teaching colleges, with the former being the more effective of the two.

The policy relevance of screening policies is important for countries that, like Chile, have seen tremendous growth in the supply of higher education options. Teaching is a relatively cheap degree to offer and supply expanded faster than any other option in Chile after government-backed loans were provided by the government for the first time. Many students with low scores then find themselves with limited options, but teaching is virtually always feasible for them. Minimal standards for entry or access to subsidies can also help regulate the supply of degrees that are being oversupplied by reducing demand from groups that are less likely to benefit from those studies. In this context, it might do a country well to consider growing more slowly, sticking with minimal standards for entry into the teaching profession and (flexible) higher wages (Biasi, 2021). This setup should ease a smoother transition from a system that provides quantity to one that provides quality.

In this paper, we have outlined that screening and recruiting policies implemented before candidates enter college could be feasible and useful in some contexts. A data-driven approach to determining the specific details of the policies seems promising. Future related work should study the equilibrium effects of these policies (in the vein of Tincani, 2021), as they will likely affect the dynamic incentives for universities. Research is needed to understand how to improve models and data to better screen candidates, or to realize they should not screen, in new contexts and consider the objectives and priorities of the policy-maker.

Appendix A

Chapter 1

A.1 Additional Tables

Table A.1: Effects on applications - Full sample

Outcome:	(1) List Length	(2) Share of HTS	(3) Ranked HTS	(4) First is HTS
Treatment	0.001	0.007	0.012	0.025
	(0.011)	(0.003)	(0.006)	(0.006)
Observations	27164	27164	27164	27164
Control Mean	4.908	0.431	0.865	0.402
p -value	0.897	0.012	0.034	<0.001

Robust standard errors, clustered at the district level, in parentheses.
Treatment p -values reported in the last row.

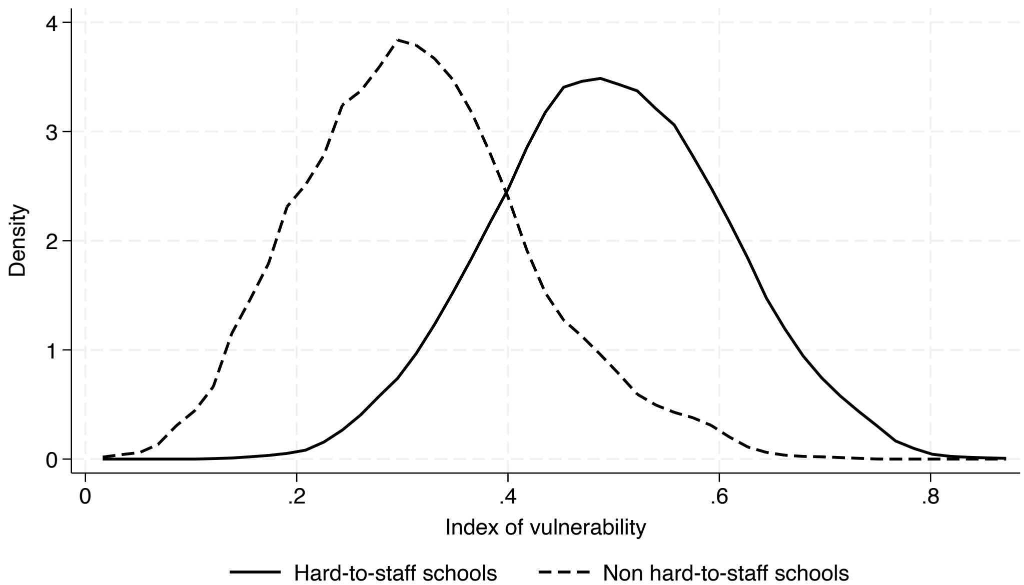
Table A.2: Effects on applications - Without failed treatment arm

Outcome:	(1) List Length	(2) Share of HTS	(3) Ranked HTS	(4) First is HTS
Treatment	0.003	0.006	0.011	0.026
	(0.012)	(0.003)	(0.007)	(0.007)
Observations	18108	18108	18108	18108
Control Mean	4.908	0.431	0.865	0.402
p -value	0.811	0.06	0.102	<0.001

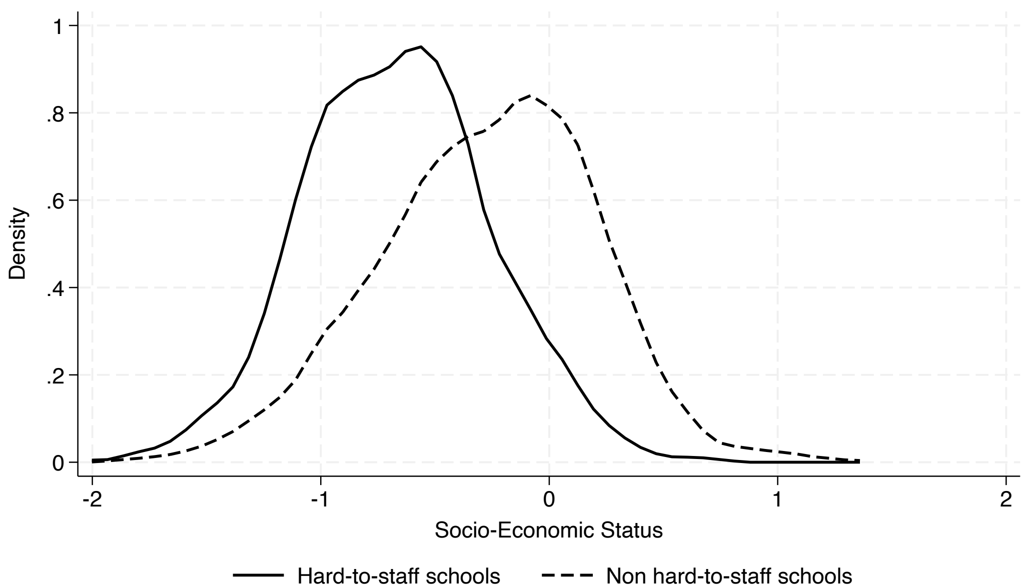
Robust standard errors, clustered at the district level, in parentheses.
Treatment p -values reported in the last row.

A.2 Additional Figures

Figure A.1: School Characteristics - Socio-Economic Variables



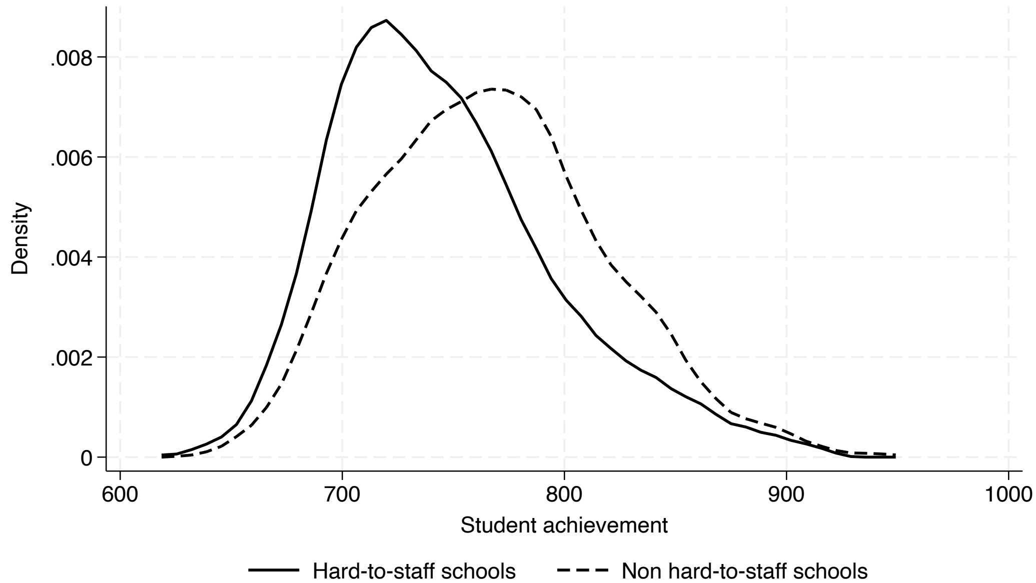
(a) School Vulnerability Index (Poverty Rate)



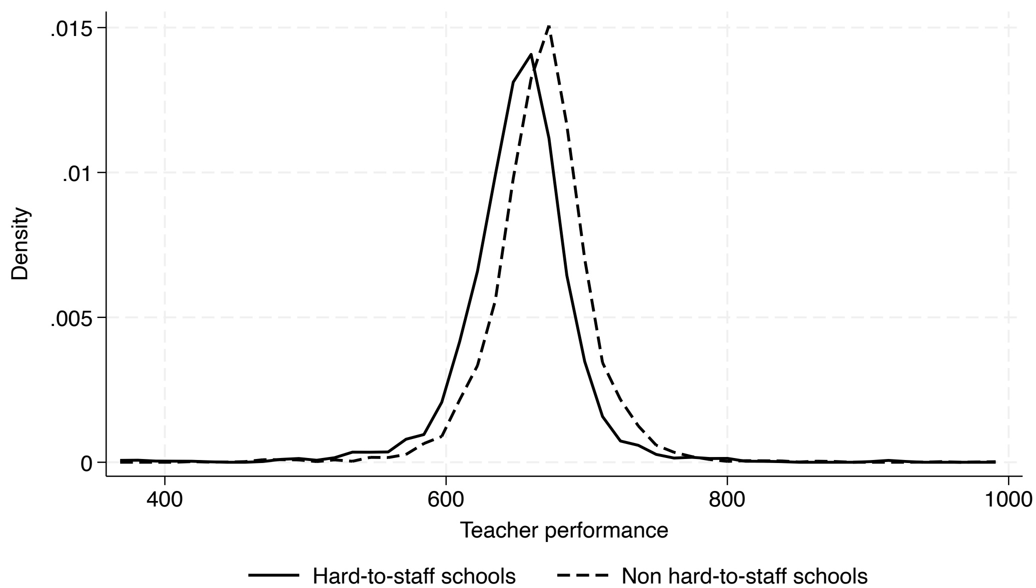
(b) Student SES Index

Note: This figure is based on Ajzenman et al. (2021) and describes socioeconomic characteristics of hard-to-staff schools and other schools. Panel (a) shows the estimated densities of the school-level poverty rates, while panel (b) plots the estimated densities of an index for the socioeconomic status of students.

Figure A.2: School Characteristics - Achievement Variables



(a) Student Test Scores



(b) Teacher Test Scores

Note: This figure is based on Ajzenman et al. (2021) and describes academic achievement variables of hard-to-staff schools and other schools. Panel (a) shows the estimated densities of the student test scores at schools, while panel (b) plots the estimated densities of teachers' base composite scores.

A.3 Additional Proof

Let $r \in (0, 1]$ *and* $\theta_i^d \perp r$ *for* $d \in \{0, 1\}$*. If* Assumption 1 *and* Assumption 2 *hold, then* $\hat{\tau}(r) := \bar{Y}^1(r) - \bar{Y}^0$ *is a consistent estimator of* $\tau(r)$*, where* $\bar{Y}^1(r)$ *is defined as the output of* algorithm 2 *and* $\bar{Y}^0$ *is the output of* algorithm 1*.*

Proof. As d_i is randomly assigned, Lemma 3 implies that

$$\bar{Y}_0 \to \mathbb{E}[Y_i|\theta_I^0] = \mathbb{E}[Y_i|i \in T, \theta_I^0],$$

almost surely.

Similar to the proof of Lemma 3, the empirical distribution $F_{\theta_T}^n$ of reported types in the treatment group θ_T is an independent random sample of size $|T|$ from the population distribution F_{θ^1}. Analogously, $F_{\theta_C}^n$ is an independent random sample of size $|C|$ from the population distribution F_{θ^0}. Thus, $\widehat{\theta}_I^r$ comes from independently drawing nr types from the empirical measure $F_{\theta_T}^n$ and $n(1-r)$ types from the empirical measure $F_{\theta_C}^n$. This induces the empirical measure $\widehat{F}_{(\theta^1,\theta^0),r}^n$.

By the Glivenko-Cantelli theorem,

$$\lim_{n\to\infty} \sup |\widehat{F}_{(\theta^1,\theta^0),r}^n - F_{(\theta^1,\theta^0),r}| = 0,$$

almost surely.

As φ complies with CD-ETE (Assumption 2), then the plug-in empirical measure $\widehat{F}_{\varphi,r}^n$, induced by $\varphi(\widehat{\theta}_I^r)$, also converges almost surely to the population distribution $F_{\varphi,r}$:

$$\lim_{n\to\infty} \sup |\widehat{F}_{\varphi,r}^n - F_{\varphi,r}| = 0.$$

As φ is a random mechanism, we can draw empirical bootstrap samples from this consistent estimator of the counterfactual distribution $F_{\varphi,r}$ by randomly drawing match-outcomes of size nr for treatment individuals in the simulation $\{Y^1[:, b]\}_{b=1}^B$ with a large B, as is standard in bootstrap methods. Convergence in the simulated outcomes resulting from the random matches $\widehat{\mu}|T$ is guaranteed by the Continuous Mapping Theorem because Y_i is a linear function of $\mu(i)|T$.

Finally, as the empirical distribution of simulated bootstrap outcomes for the treatment group comes from a consistent estimator of the counterfactual distribution $F_{\varphi,r}|T$, the continuous (linear) functional of this bootstrap distribution $\bar{Y}^1(r)$ is a sample average that converges almost surely to the population expectation $\mathbb{E}[Y_i|i \in T, \theta_T^1, \theta_C^0, r]$:

$$\lim_{B\to\infty}\lim_{n\to\infty}\sum_{b=1}^{B}\sum_{i=1}^{nr} Y^1[i,b]/(nrB) = \mathbb{E}[Y_i|i \in T, \theta_T^1, \theta_C^0, r],$$

almost surely.

Then,

$$\begin{aligned}\hat{\tau}(r) =& \bar{Y}^1(r) - \bar{Y}^0 \\ \to& \mathbb{E}[Y_i|i \in T, \theta_T^1, \theta_C^0, r] - \mathbb{E}[Y_i|i \in T, \theta_I^0] \\ =& \tau(r)\end{aligned}$$

almost surely. □

Appendix B

Chapter 2

B.1 Additional Tables

Table B.1: Selectivity by Institution

Univ.	Mean Score	Std. Dev.	P10	P25	P50	P75	P90	Total Adm.	Tier
11	691.5	46.0	630.9	658.2	688.9	723.5	751.9	5424	1
12	694.7	52.5	630.6	652.8	690.4	734.2	769.9	4754	1
16	633.9	36.64	591.8	610.7	631.75	654.9	676.6	4725	2
42	656.2	37.1	610.3	633.7	655.5	678.3	702.3	1868	2
43	657.2	53.5	586.0	613.6	657.6	694.9	723.9	1221	2
13	602.8	62.8	521.3	554.3	601.8	645.0	687.4	6377	3
15	615.2	78.9	506.9	547.7	629.2	680.5	708.2	4346	3
19	596.5	54.4	532.7	558.7	589.8	625.3	671.5	3984	3
14	610.9	43.0	557.4	581.4	608.3	637.7	667.3	3448	3
17	582.6	58.1	514.5	541.6	577.1	617.0	660.4	3076	3
44	605.9	55.0	537.8	566.2	603.0	636.1	673.9	2903	3
38	617.2	40.0	571.2	590.0	613.3	638.85	668.6	2726	3
30	592.7	56.1	525.1	549.9	586.3	627.1	669.2	2354	3
18	589.3	55.3	520.4	546.3	583.6	625.1	664.8	2322	3
34	616.1	52.1	551.4	582.9	615.0	646.9	685.9	1904	3
35	581.5	48.0	527.8	546.1	573.2	606.7	644.3	1639	3
45	595.6	38.8	547.4	567.5	593.9	619.6	646.8	1539	3
40	580.9	59.6	509.7	540.8	575.7	613.7	660.0	1145	3
20	605.7	33.6	570.0	580.7	603.3	624.95	649.0	1092	3
41	550.3	53.7	484.5	511.9	546.8	583.7	618.6	12615	4
39	571.8	56.5	501.4	529.0	568.0	604.4	645.8	4895	4
36	563.2	51.9	502.7	525.6	557.6	593.4	625.4	2451	4
29	577.5	43.7	523.4	544.5	572.4	604.35	639.7	2449	4
21	548.7	36.0	503.2	524.5	548.3	574.25	595.0	2392	4
37	543.9	43.0	492.0	509.6	540.0	571.35	598.8	2206	4
25	575.4	45.0	521.4	542.5	569.4	605.7	635.4	1791	4
26	556.0	37.4	512.0	530.0	553.0	578.55	604.0	1748	4
22	557.0	51.9	493.7	517.1	549.0	594.4	627.8	1720	4
24	566.1	65.2	487.2	515.7	558.3	606.7	662.1	1433	4
23	550.8	47.1	493.8	514.0	541.3	583.8	616.4	928	4
27	550.6	50.0	486.8	511.9	549.5	585.7	616.6	815	4
32	541.0	44.7	484.4	508.2	535.8	571.0	603.7	785	4
33	554.0	50.7	492.1	516.1	547.4	586.0	626.0	499	4

Note: This table summarizes the admission scores for each on-platform institution in 2012. Universities

Table B.2: Institutional Aid Summary - 2012

	Scholarships				Discounts		Supplement
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Institution	Any	Deterministic	PSU-based	SES-based	Any	PSU-based	Any
P. Universidad Católica de Chile	Yes	Yes	Yes	No	Yes	Yes	Yes
P. Universidad Católica de Valparaiso	Yes	Yes	Yes	No	Yes	No	Yes
Universidad Adolfo Ibanez	No	No	No	No	Yes	Yes	No
Universidad Alberto Hurtado	No	No	No	No	Yes	Yes	No
Universidad Andres Bello	No	No	No	No	No	No	No
Universidad Austral de Chile	Yes	Yes	Yes	No	Yes	Yes	No
Universidad Católica de Temuco	No	No	No	No	No	No	No
Universidad Católica de la S. Concepcion	No	No	No	No	No	No	Yes
Universidad Católica del Maule	No	No	No	No	Yes	No	Yes
Universidad Católica del Norte	No	No	No	No	No	No	Yes
Universidad Diego Portales	Yes	Yes	No	No	Yes	Yes	Yes
Universidad Finis Terrae	No	No	No	No	Yes	No	No
Universidad Mayor	No	No	No	No	Yes	Yes	No
Universidad Metrop. de Cs. de la Educación	No	No	No	No	No	No	No
Universidad Técnica Federico Santa Maria	Yes	Yes	Yes	Yes	Yes	Yes	No
Universidad Tecnológica Metrop.	Yes	Yes	Yes	No	Yes	Yes	No
Universidad de Antofagasta	Yes	No	Yes	No	No	No	Yes
Universidad de Arturo Prat	Yes	No	No	No	Yes	No	No
Universidad de Atacama	No	No	No	No	No	No	No
Universidad de Chile	Yes	No	Yes	Yes	No	No	No
Universidad de Concepción	Yes	No	Yes	No	Yes	Yes	No
Universidad de La Frontera	Yes	Yes	No	No	Yes	Yes	Yes
Universidad de La Serena	No	No	No	No	No	No	No
Universidad de Los Andes	No	No	No	No	No	No	No
Universidad de Los Lagos	No	No	No	No	No	No	No
Universidad de Magallanes	No	No	No	No	No	No	No
Universidad de Playa Ancha	Yes	Yes	Yes	No	Yes	No	No
Universidad de Santiago de Chile	Yes	Yes	Yes	No	No	No	Yes
Universidad de Talca	Yes	Yes	Yes	No	No	No	No
Universidad de Tarapacá	Yes	Yes	Yes	No	Yes	Yes	No
Universidad de Valparaiso	Yes	Yes	Yes	No	Yes	Yes	No
Universidad del Bio-Bio	Yes	No	Yes	No	No	No	No
Universidad del Desarrollo	No	No	No	No	Yes	No	No

Note: This table presents each institution with indicators of whether they offer different types of scholarships, discounts, stipends, or supplementary help to complement MINEDUC scholarships and credits obtained by the student. We only consider scholarships for entering (not continuing) students in this summary. Ex ante deterministic means that there is certainty about the amount of the scholarship/discount at the time of deciding where to apply. PSU-based indicates a requirement to obtain a PSU score above a high threshold, usually 1 or 2 standard deviations above the mean, depending on the institution. SES-based indicates a requirement to belong to the income quintiles 1, 2, or (sometimes) 3. PSU-based and SES-based are not mutually exclusive in most cases. Source is CRUCh's official document on university services and benefits, circulated in newspapers one month before the PSU.

Table B.3: Event study: Admission, Enrollment, Dropout, Graduation (without student covariates)

	Admission	Enrollment	Dropout	Graduation
Year 2010×Male×Private	-0.002	0.006	-0.003	0.006
	(0.006)	(0.008)	(0.005)	(0.010)
Year 2012×Male×Private	0.095***	0.107***	-0.005	0.030***
	(0.005)	(0.007)	(0.005)	(0.010)
Year 2013×Male×Private	0.104***	0.127***	0.002	
	(0.005)	(0.007)	(0.005)	
Year 2014×Male×Private	0.114***	0.125***	-0.003	
	(0.005)	(0.007)	(0.005)	
Year 2015×Male×Private	0.099***	0.134***	0.003	
	(0.005)	(0.007)	(0.005)	
Year 2010×Male×Public	-0.045***	0.014***	0.001	-0.002
	(0.003)	(0.004)	(0.003)	(0.005)
Year 2012×Male×Public	0.082***	0.051***	-0.011***	0.009*
	(0.003)	(0.003)	(0.003)	(0.005)
Year 2013×Male×Public	0.084***	0.069***	-0.001	
	(0.003)	(0.003)	(0.003)	
Year 2014×Male×Public	0.089***	0.081***	-0.004	
	(0.002)	(0.003)	(0.003)	
Year 2015×Male×Public	0.058***	0.084***	-0.002	
	(0.003)	(0.003)	(0.003)	
Year 2010×Female×Private	-0.018**	0.004	0.008	-0.024**
	(0.007)	(0.009)	(0.005)	(0.010)
Year 2012×Female×Private	0.090***	0.139***	0.000	0.022**
	(0.006)	(0.008)	(0.004)	(0.010)
Year 2013×Female×Private	0.107***	0.156***	0.006	
	(0.006)	(0.008)	(0.004)	
Year 2014×Female×Private	0.114***	0.166***	0.002	
	(0.006)	(0.008)	(0.004)	
Year 2015×Female×Private	0.096***	0.155***	0.013***	
	(0.006)	(0.008)	(0.005)	
Year 2010×Female×Public	-0.045***	0.020***	-0.008**	-0.001
	(0.003)	(0.004)	(0.003)	(0.005)
Year 2012×Female×Public	0.060***	0.057***	-0.023***	0.016***
	(0.003)	(0.004)	(0.003)	(0.005)
Year 2013×Female×Public	0.072***	0.093***	-0.013***	
	(0.003)	(0.004)	(0.003)	
Year 2014×Female×Public	0.084***	0.110***	-0.017***	
	(0.003)	(0.004)	(0.003)	
Year 2015×Female×Public	0.051***	0.112***	-0.013***	
	(0.003)	(0.004)	(0.003)	
Constant	0.778***	0.691***	0.107***	0.569***
	(0.002)	(0.003)	(0.002)	(0.003)
Observations	606280	393193	318809	163531

Note: This table shows estimates of the average difference in each outcome, for each type of student, and for each year after 2009. The base year is 2011 and the base type is Female-Public. Admission refers to the probability of being assigned a seat in the platform; Enrollment refers to the probability of enrolling in a platform program conditional on being admitted in a G25 option; Dropout refers to the probability of not being enrolled in any option the year after enrolling in a G25 program; and Graduation refers to the probability of graduating within 7 years of enrolling in a G25 program. The estimating equation does not include student covariates, except for student-type fixed effects. These estimated coefficients are not reported in the table. The results on graduation rates are constrained to years before 2013 because we do not have data after 2019. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.4: Event study: Admission, Enrollment, Dropout, Graduation (score-decile covariates)

	Admission	Enrollment	Dropout	Graduation
Year 2010×Male×Private	-0.002	0.004	-0.003	0.006
	(0.006)	(0.008)	(0.005)	(0.010)
Year 2012×Male×Private	0.117***	0.109***	-0.008*	0.035***
	(0.005)	(0.007)	(0.005)	(0.010)
Year 2013×Male×Private	0.125***	0.127***	-0.001	
	(0.005)	(0.007)	(0.005)	
Year 2014×Male×Private	0.132***	0.128***	-0.006	
	(0.005)	(0.007)	(0.005)	
Year 2015×Male×Private	0.119***	0.135***	0.002	
	(0.005)	(0.006)	(0.005)	
Year 2010×Male×Public	-0.042***	0.013***	0.001	-0.002
	(0.003)	(0.003)	(0.003)	(0.004)
Year 2012×Male×Public	0.091***	0.058***	-0.017***	0.015***
	(0.002)	(0.003)	(0.003)	(0.004)
Year 2013×Male×Public	0.093***	0.076***	-0.008***	
	(0.002)	(0.003)	(0.003)	
Year 2014×Male×Public	0.093***	0.087***	-0.010***	
	(0.002)	(0.003)	(0.003)	
Year 2015×Male×Public	0.062***	0.086***	-0.005*	
	(0.002)	(0.003)	(0.003)	
Year 2010×Female×Private	-0.018***	0.003	0.009*	-0.022**
	(0.007)	(0.009)	(0.005)	(0.010)
Year 2012×Female×Private	0.115***	0.141***	-0.001	0.026***
	(0.006)	(0.008)	(0.004)	(0.010)
Year 2013×Female×Private	0.135***	0.163***	0.002	
	(0.005)	(0.008)	(0.004)	
Year 2014×Female×Private	0.137***	0.171***	-0.001	
	(0.005)	(0.008)	(0.004)	
Year 2015×Female×Private	0.126***	0.164***	0.007	
	(0.006)	(0.008)	(0.005)	
Year 2010×Female×Public	-0.042***	0.022***	-0.009***	-0.002
	(0.003)	(0.004)	(0.003)	(0.005)
Year 2012×Female×Public	0.074***	0.064***	-0.026***	0.024***
	(0.003)	(0.004)	(0.003)	(0.005)
Year 2013×Female×Public	0.087***	0.102***	-0.019***	
	(0.003)	(0.004)	(0.003)	
Year 2014×Female×Public	0.097***	0.120***	-0.025***	
	(0.003)	(0.004)	(0.003)	
Year 2015×Female×Public	0.062***	0.118***	-0.018***	
	(0.003)	(0.004)	(0.003)	
Constant	0.557***	0.464***	0.260***	0.321***
	(0.013)	(0.005)	(0.004)	(0.007)
Observations	606280	393193	318809	163531

Note: This table shows estimates of the average difference in each outcome, for each type of student, and for each year after 2009. The base year is 2011 and the base type is Female-Public. Admission refers to the probability of being assigned a seat in the platform; Enrollment refers to the probability of enrolling in a platform program conditional on being admitted in a G25 option; Dropout refers to the probability of not being enrolled in any option the year after enrolling in a G25 program; and Graduation refers to the probability of graduating within 7 years of enrolling in a G25 program. The estimating equation includes student covariates (GPA and test score deciles) and student-type fixed effects. These estimated coefficients are not reported in the table. The results on graduation rates are constrained to years before 2013 because we do not have data after 2019. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B.5: Preference estimates: inside-option parameters

Parameters	Male Private	Male Public	Female Private	Female Public
Preferences (ψ^o)				
Price	-0.17	-0.146	-0.109	-0.142
	(0.012)	(0.004)	(0.008)	(0.003)
Same City	1.296	1.388	1.289	1.418
	(0.012)	(0.002)	(0.012)	(0.003)
STEM x Math	0.331	0.277	0.254	0.336
	(0.012)	(0.004)	(0.008)	(0.002)
Humanities x Math	-0.067	-0.105	-0.062	-0.034
	(0.004)	(0.002)	(0.008)	(0.003)
STEM x Verbal	-0.071	-0.065	-0.035	-0.075
	(0.005)	(0.002)	(0.008)	(0.003)
Humanities x Verbal	0.179	0.182	0.145	0.133
	(0.007)	(0.002)	(0.011)	(0.003)
Score*Teaching	-0.258	-0.113	-0.15	-0.17
	(0.02)	(0.007)	(0.02)	(0.005)
Score2*Teaching	-0.086	-0.145	-0.045	-0.222
	(0.019)	(0.005)	(0.022)	(0.007)
Score3*Teaching	0.072	0.077	0.04	0.121
	(0.008)	(0.003)	(0.007)	(0.003)
Aftermarket frictions (α)				
G25	-1.551	-0.88	-1.349	-0.84
	(0.03)	(0.014)	(0.035)	(0.014)
G8 On	-0.827	-0.616	-0.739	-0.595
	(0.042)	(0.037)	(0.045)	(0.032)
G8 Off	0.255	-0.435	0.183	-0.739
	(0.024)	(0.016)	(0.021)	(0.01)
Local	0.232	-0.046	0.187	0.041
	(0.027)	(0.014)	(0.029)	(0.01)
SD of program FE				
σ_{FE}	1.2	0.803	1.429	0.853
	(0.041)	(0.005)	(0.054)	(0.01)
RC covariance matrix (ψ^u)				
STEM	0.318	0.248	0.483	0.403
	(0.019)	(0.003)	(0.011)	(0.004)
Humanities	0.337	0.226	0.39	0.264
	(0.022)	(0.002)	(0.017)	(0.004)
Humanities vs STEM (ρ)	0.166	0.09	0.28	0.178
	(0.013)	(0.001)	(0.01)	(0.003)

Note: Preference parameters were estimated via Gibbs sampling and include program fixed effects. The number of observations used for the estimation are 484549 and the number of options are 1334 over three years.

Table B.6: Preference estimates: outside-option and individual-level parameters

Parameters	Male Private	Male Public	Female Private	Female Public
First Outside Option (γ^0)				
Constant	2.602	2.405	3.153	2.553
	(0.08)	(0.01)	(0.062)	(0.01)
Math	0.099	-0.087	-0.012	-0.136
	(0.018)	(0.006)	(0.017)	(0.006)
Verbal	-0.008	-0.041	0.028	0.032
	(0.015)	(0.006)	(0.015)	(0.005)
Big City	0.891	0.688	0.886	0.679
	(0.033)	(0.008)	(0.028)	(0.008)
Current Cohort	-0.117	0.127	-0.107	0.113
	(0.027)	(0.008)	(0.024)	(0.008)
1(2011)	0.063	0.078	0.046	0.082
	(0.031)	(0.009)	(0.024)	(0.008)
1(2012)	-0.095	0.14	-0.074	0.154
	(0.032)	(0.009)	(0.026)	(0.009)
Scholarship Amount	-0.29	0.032	-0.185	0.011
	(0.118)	(0.014)	(0.104)	(0.013)
Score2	0.303	0.371	0.216	0.354
	(0.027)	(0.008)	(0.031)	(0.007)
Score3	-0.078	-0.086	-0.046	-0.095
	(0.009)	(0.004)	(0.012)	(0.005)
$\sigma_{0,0}$	0.569	0.584	0.546	0.578
	(0.012)	(0.003)	(0.009)	(0.003)
Second Outside Option (γ^1)				
Constant	1.845	2.135	2.378	2.339
	(0.089)	(0.027)	(0.075)	(0.019)
Math	-0.33	-0.41	-0.402	-0.386
	(0.039)	(0.017)	(0.033)	(0.012)
Verbal	-0.211	-0.223	-0.092	-0.138
	(0.03)	(0.01)	(0.02)	(0.009)
Big City	0.807	0.884	0.697	0.671
	(0.036)	(0.014)	(0.035)	(0.009)
Current Cohort	-0.607	-0.001	-0.392	0.036
	(0.044)	(0.01)	(0.032)	(0.008)
1(2011)	-0.064	0.063	0.027	0.096
	(0.043)	(0.012)	(0.038)	(0.013)
1(2012)	0.608	0.4	0.572	0.525
	(0.041)	(0.012)	(0.038)	(0.01)
Scholarship Amount	0.259	0.344	-0.103	0.183
	(0.109)	(0.017)	(0.1)	(0.014)
Score2	0.233	0.225	0.219	0.175
	(0.048)	(0.01)	(0.038)	(0.012)
Score3	-0.018	0.007	-0.018	0.008
	(0.015)	(0.005)	(0.014)	(0.007)
$\sigma_{0,1}$	1.815	1.277	1.54	1.205
	(0.093)	(0.036)	(0.055)	(0.024)

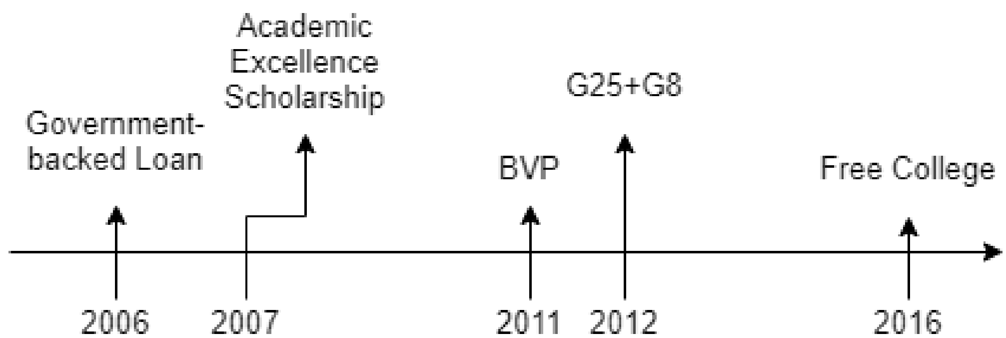
Table B.7: Parameter estimates: outcomes

Parameters	Male Private	Male Public	Female Private	Female Public
Production Function				
utility shock	0.571	0.402	1.15	0.321
	(0.372)	(0.026)	(1.117)	(0.031)
Constant	-2.608	-1.851	-4.57	-1.319
	(1.299)	(0.118)	(4.191)	(0.129)
Math	0.279	0.298	0.327	0.361
	(0.049)	(0.012)	(0.043)	(0.012)
Verbal	0.095	0.031	0.117	0.043
	(0.03)	(0.01)	(0.053)	(0.01)
Big City	-0.196	-0.183	-0.753	-0.203
	(0.217)	(0.02)	(0.711)	(0.025)
Current Cohort	0.208	0.144	0.288	0.124
	(0.036)	(0.011)	(0.114)	(0.01)
1(2011)	-0.0	-0.012	0.055	0.022
	(0.025)	(0.012)	(0.03)	(0.013)
1(2012)	0.036	0.022	0.147	0.059
	(0.041)	(0.014)	(0.065)	(0.014)
Scholarship Amount	0.073	0.138	0.111	0.121
	(0.107)	(0.019)	(0.125)	(0.015)
Score2	-0.161	-0.133	-0.313	-0.102
	(0.064)	(0.018)	(0.204)	(0.019)
Score3	0.019	0.014	0.044	-0.005
	(0.017)	(0.006)	(0.048)	(0.008)
Price	-0.117	-0.065	-0.066	-0.015
	(0.08)	(0.025)	(0.09)	(0.019)
Same City	0.614	0.492	1.268	0.424
	(0.366)	(0.032)	(1.136)	(0.038)
STEM x Math	0.076	0.059	0.239	0.085
	(0.093)	(0.015)	(0.214)	(0.019)
Humanities x Math	-0.07	-0.093	-0.035	-0.068
	(0.03)	(0.015)	(0.049)	(0.017)
STEM x Verbal	0.024	0.035	0.006	0.005
	(0.028)	(0.014)	(0.037)	(0.015)
Humanities x Verbal	0.102	0.081	0.124	0.056
	(0.056)	(0.014)	(0.137)	(0.013)
Score*Teaching	-0.149	-0.107	-0.101	-0.15
	(0.164)	(0.049)	(0.264)	(0.043)
Score2*Teaching	-0.04	-0.004	0.057	0.008
	(0.189)	(0.059)	(0.15)	(0.043)
Score3*Teaching	0.019	0.012	-0.024	0.015
	(0.079)	(0.028)	(0.062)	(0.03)

Note: Preference parameters were estimated via Gibbs sampling and include program fixed effects. The number of observations used for the estimation are 484549 and the number of options are 1334 over three years.

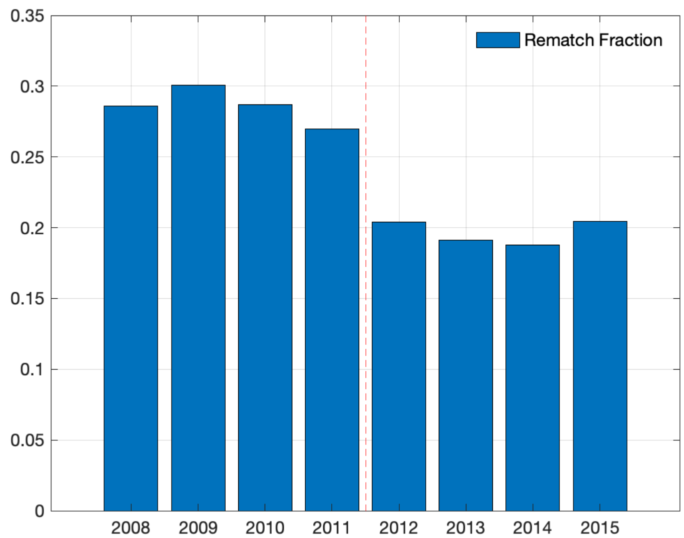
B.2 Additional Figures

Figure B.1: Policies Expanding Access to Higher Education in Chile - Timeline



Note: This figure shows a timeline with the major policies in higher education implemented in Chile before and after the time period under study in this paper. The first one is the *Government-backed Loan (CAE)*, aid open to students applying to CRUCh or accredited non-CRUCH higher education institutions. Importantly, eligibility was not tied to participation on the centralized platform. The second one is the *Academic Excellence Scholarship*, aimed to cover part of the annual fee of students belonging to the 10% of higher achievement. They have to apply to CRUCh or accredited non-CRUCH higher education institutions, come from public or private voucher schools, belong to the 80% most vulnerable population, and enter the year right after they graduated high school. This policy is also unrelated to participation on the centralized platform. The *Teacher Scholarship* (BVP) which began in 2011 provided full scholarships for high-scoring students at eligible teacher training programs. This was unrelated to participation on the platform. Finally, the Free College policy established that 50% of most vulnerable students do not have to pay tuition or annual fee in CRUCh or accredited non-CRUCH higher education institutions attached to the agreement.

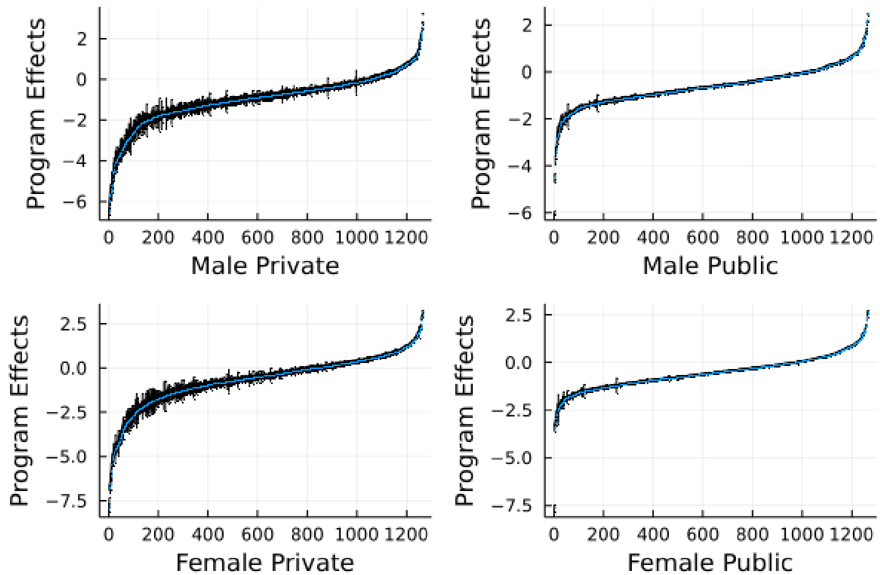
Figure B.2: Rematch fraction when (ex-ante) dropping students who decline placements



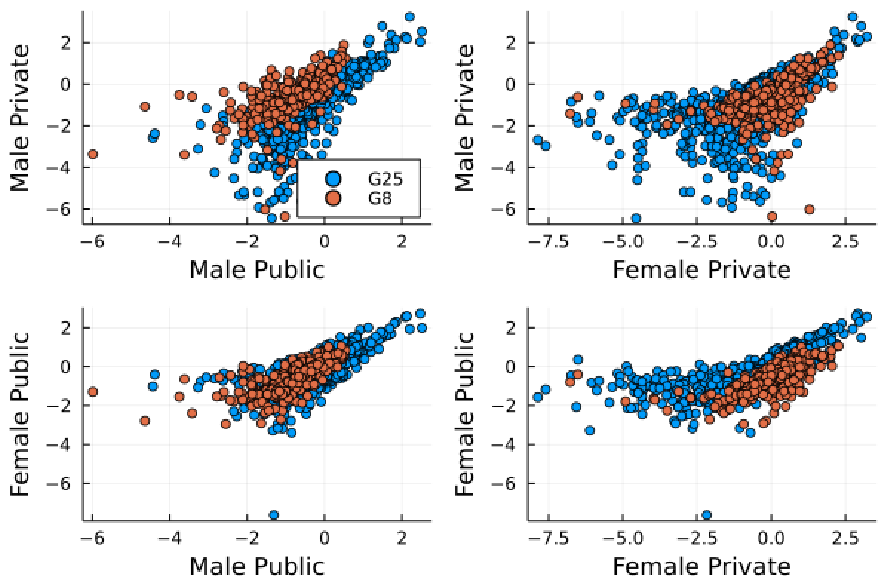
Note: This figure shows the fraction of match participants by year, other than those who renege on offers, whose initial assignment would change if students who decline their offers are removed from the match ex ante.

Figure B.3: Distribution of Program Fixed Effects (δ)

(a) Distribution of Program Fixed Effects δ



(b) Comparisons across types



Note: Figures display estimates of program fixed effects δ. Parameters are estimated separately by type. Left panel: sorted within each type, black lines represent 95% posterior probability intervals. Right panel: scatter plots comparing means of each program across types. Blue indicates G25, red indicates G8.

B.3 Additional Proof

Our model and setting involve an initial matching phase in which "on-platform" programs participate, followed by a second phase in which all programs participate. In this section we formally define multi-phase and single-phase CPDA procedures and show that, given any profile of reports, priorities and capacities, the multi-phase and single-phase DA procedures produce identical allocations in the absence of frictions. Thus, viewed as mechanisms—i.e. mappings from preference profiles to allocations—the two are identical. It follows that the two mechanisms have identical properties with respect to stability and incentives. In particular, with no frictions, the two-phase procedure inherits "large market" results that have been established for college-proposing DA procedures. Moreover, in the absence of frictions, enrollment probabilities at program j are continuous at the initial cutoffs for programs which admit students off of waitlists whenever students' preferences are continuous in $index_{ij}$ and no other program $k \neq j$ has a final cutoff that coincides with j's initial cutoff.

Let $i = 1, \ldots, N$ be a set of students, and J be a set of programs. Let $J^{\text{on}} \subset J$ be any subset of J, and $J^{\text{off}} = J \setminus J^{\text{on}}$ the remaining programs. For $j \in J$, let program j's capacity be given by k_j. Suppose that each program j has responsive preferences, ranking all applicants according to some score or ranking $index_{ij} \in [0, \overline{s}] \subset \mathbb{R}$. Let student i have true ordinal preferences R_i, and let $R = \times_i R_i$ the profile of student preferences. Student i is acceptable to j if i satisfies j's exogenously-given eligibility requirements.

Consider the following two matching procedures:

Claim 1. *The two-phase college-proposing deferred acceptance procedure and the college-proposing deferred acceptance procedure produce identical assignments.*

Proof. Fix a set of programs $J^{\text{on}} \subseteq J$. We define a mapping $F^{J^{\text{on}}} : \mathbb{R}_J \to \mathbb{R}_J$ as follows:

Given a vector $\pi \in \mathbb{R}^J$, define student i's choice set as $Ch_i(\pi) = \{j \in J : index_{ij} \geq \pi_j\}$. Let $c_i(Ch)$ be i's preferred element of a choice set $Ch \subseteq J$ (or 0, if i prefers the outside

Two-phase college-proposing deferred acceptance procedure:

1. Step 1: each college $j \in J^{\text{on}}$ proposes to its preferred k_j acceptable students. Students receiving multiple offers hold their preferred offer, if any, and reject the rest.

2. Step $k \geq 1$: each college $j \in J^{\text{on}}$ proposes to its preferred k_j acceptable students among those that have not yet rejected it. (If there are fewer than k_j such students, the college proposes to all students who have not yet rejected it). Students receiving multiple offers hold their preferred offer, if any, and reject the rest.

3. The initial phase ends when there is a round with no rejected proposals. Each student who is holding a program is provisionally assigned to that program.

4. The second phase begins. Each college $j \in J$ proposes to its preferred k_j acceptable students among those that have not yet rejected it.[1] (If there are fewer than k_j such students, the college proposes to all students who have not yet rejected it). Students receiving multiple offers hold their preferred offer, if any, and reject the rest. This step repeats until there is a round with no rejected proposals.

5. Each student who is holding a program is assigned to the program that he is holding.

College-proposing deferred acceptance procedure:

1. Step 1: each college $j \in J$ proposes to its preferred k_j acceptable students. Students receiving multiple offers hold their preferred offer, if any, and reject the rest.

2. Step $k \geq 1$: each college $j \in J$ proposes to its preferred k_j acceptable students among those that have not yet rejected it. (If there are fewer than k_j such students, the college proposes to all students who have not yet rejected it). Students receiving multiple offers hold their preferred offer, if any, and reject the rest.

3. The algorithm terminates when there is a round with no rejected proposals. Each student who is holding a program is assigned to the program that he is holding.

option to all elements of Ch). Let

$$Rej_{ij}(\pi) = 1(j \in Ch_i(\pi)) * 1(c_i(Ch_i(\pi)) \neq j)$$

be an indicator for the event that student i rejects program j when the cutoffs are π.

Let

$$F_j^{J^{\text{on}}}(\pi) = \begin{cases} \pi_j & \text{if } j \notin J^{\text{on}} \\ \sup\{\tilde{\pi} : \sum_{i:index_{ij} \geq \tilde{\pi}} (1 - Rej_{ij}(\pi)) \geq k_j\} & \text{if } j \in J^{\text{on}} \text{ and } \sum_i (1 - Rej_{ij}(\pi)) \geq k_j \\ 0 & \text{otherwise} \end{cases}$$

Let $F^{J^{\text{on}}}(\pi) = (F_1^{J^{\text{on}}}(\pi), \ldots, F_J^{J^{\text{on}}}(\pi))$.

By construction, the college-proposing deferred acceptance procedure consists of iterating F^J to convergence, starting at $[\overline{s}, \ldots, \overline{s}]$.[2] The two-phase CPDA procedure consists of iterating $F^{J^{\text{on}}}$ to convergence, starting at $[\overline{s}, \ldots, \overline{s}]$, and subsequently, letting π^1 denote the value to which repeated iterations of $F^{J^{\text{on}}}$ converge, iterating F^J to convergence starting at π^1.

Because $Rej_{ij}(\pi)$ is decreasing in π, it's clear that $\pi' \geq \pi$ implies $F^{J^{\text{on}}}(\pi') \geq F^{J^{\text{on}}}(\pi)$, i.e. $F^{J^{\text{on}}}$ is isotone. Let π^* denote the fixed point cutoffs obtained by iterating F^J, beginning at $\pi_0 = [\overline{s}, \ldots, \overline{s}]$. Because F^J is isotone, iterating F^J beginning at any $\pi \geq \pi^*$ will converge to π^*. Thus it suffices to show $\pi^1 \geq \pi^*$.

By construction, $F^J([\overline{s}, \ldots, \overline{s}]) \leq F^{J^{\text{on}}}([\overline{s}, \ldots, \overline{s}])$. Hence, for all $n \geq 1$ we have

$$(F^J)^n([\overline{s}, \ldots, \overline{s}]) \leq (F^{J^{\text{on}}})^n([\overline{s}, \ldots, \overline{s}]).$$

□

[2] The function $F^{J^{\text{on}}}$ represents a single college-proposing DA step in which the programs in J^{on} participate. The cutoffs π represent the state of the algorithm, and a student's current held option is given by $c_i(Ch_i(\pi))$.

B.4 Additional Details

B.4.1 Preliminaries

For computational convenience in estimation, we reparametrize our model as follows. We emphasize that the likelihood is unchanged under this reparameterization, and hence the reparametrized model is equivalent to our original specification.

First, without loss, we modify the location normalization by subtracting $z_i \lambda_g^z$ from all "inside" utilities u_{ij} and all "outside" utilities u_{i0}^k for $k = 0, 1$. We do so because this transformation allows us to reduce the size of the matrices required to store all covariates that enter the "inside" goods' utilities.[3]

For each group g, define $\gamma_g^0 = -\lambda_g^z$, and $\gamma_g^1 = \gamma_g - \lambda_g^z$. We have

$$\begin{aligned}
u_{ij} &= \delta_{j,g(i)} + w_{ij}\lambda_{g(i)}^w + x_j \eta_i^x + p_{ij}\lambda_{g(i)}^p + \epsilon_{ij} \\
u_{i0} &\sim N(z_i \gamma_{g(i)}^0, \sigma_{0,0,g(i)}^2) \\
u_{i1} &\sim N(z_i \gamma_{g(i)}^1, \sigma_{0,0,g(i)}^2).
\end{aligned}$$

Second, again without loss, we substitute u_{ij} into the human-capital index h_{ij} as follows: Define $\tilde{\overline{\beta}}_g = \overline{\beta}_g - \rho_g \delta_g$ and $\tilde{\beta}_g = \beta_g - \rho_g \lambda_g$. Whenever i enrolls in j, let $\tilde{\nu}_{ij} = \nu_{ij} - \rho_{g(i)}(x_{j,t(i)}\eta_i^x + \epsilon_{ij}))$. We have

$$h_i = \tilde{\overline{\beta}}_{g(i)} + z_i \beta_{g(i)}^z + w_{ij}\tilde{\beta}_{g(i)}^w + p_{ij}\tilde{\beta}_{g(i)}^p + u_{ij}\rho_{g(i)} + \tilde{\nu}_{ij},$$

where $\tilde{\nu}_{ij} \sim N(0, 1)$.

As all shocks are normal, this transformed model is equivalent to including the error term $x_j \eta_i^x + \epsilon_{ij}$ in the human capital index, or to allowing correlation between the human-capital shock ν_{ij} and the utility shock, in the sense that the likelihood is invariant to this transformation.

If person i does not enroll in any inside option, we adopt the convention $h_{i,0} = 0$.

[3] For each person i we store a $J \times K$ matrix of match-level covariates, where K is the number of such variables. This transformation lets us avoid repeating the student-level variables J times.

B.4.2 Constructing the relevant set and bound indices

Before estimation, we first restrict each person i's rank-order list to programs in i's relevant set, J_i^{relevant}, dropping programs outside this set. To do so, we iterate through i's list, dropping programs that are ex-ante clearly infeasible. In the event that we reach a program k which is ex-ante clearly feasible, we keep program k but drop the remainder of the list. Let ℓ_i denote the resulting restricted list.

Next, we find the indices of the programs that provide the least upper bound and greatest lower bound for each option in application data. Let $\overline{J}_t$ be a $(J_t \times N_t)$ matrix denoting upper bound indices. We set:

- $\overline{J}_t[j,i] = -1$, denoting no upper bound on u_{ij}, if j was i's first choice, j was ex-ante clearly infeasible for i, or j was off-platform in market t.
- $\overline{J}_t[j,i] = 0$, denoting that the least upper bound on u_{ij} is given by u_{i0}^0, if j was on-platform, not ex-ante clearly infeasible, not ranked by i, and J_i^{relevant} does not contain a program that is clearly feasible for i.
- $\overline{J}_t[j,i] = k > 0$, denoting that the least upper bound on u_{ij} is given by u_{ik}, if j is ranked in rth place for $r \geq 2$ and k is ranked in $r-1$th place within J_i^{relevant}, or if j was on-platform, not ex-ante clearly infeasible, not ranked by i, and J_i^{relevant} contains an ex-ante clearly-feasible program k.

Analogously, we construct a matrix of lower-bound indices, $\underline{J}_t$, as follows. We set:

- $\underline{J}_t[j,i] = -2$, denoting that the greatest lower bound on u_{ij} is given by

$$\max\{u_{i0}^0, \max\{u_{ik} : k \text{ unranked, on-platform, not ex-ante clearly infeasible}\}\},$$

  in the event j is the final program in J_i^{relevant}, and either the list is of full length or j is ex-ante clearly feasible.[4]

[4]We know that the final listed program is preferred to the outside option and to everything that was not listed. The subtlety is that in this case we do not know how the outside option u_{i0}^0 compares to the "inside" options that could have been listed but were not.

- $\underline{J}_t[j, i] = -1$, denoting no lower bound (i.e. the greatest lower bound on u_{ij} is $-\infty$), if $j \notin J_i^{\text{relevant}}$. This case occurs when j is available but not listed, as well as when j is off-platform or ex-ante clearly infeasible.

- $\underline{J}_t[j, i] = 0$, denoting the greatest lower bound is $u_{ij} < u_{i0}^0$, when j is the final element of J_i^{relevant}, the list is not full, and j is not ex-ante clearly feasible.

- $\underline{J}_t[j, i] = k > 0$, denoting the greatest lower bound on u_{ij} is given by u_{ik}, if j is ranked in rth place and k is ranked in $r + 1$th within J_i^{relevant}.

Similarly, we construct vectors $\overline{O}$ and $\underline{O}$, of length N_t, giving the indices of the programs providing bounds on u_{i0}^0. If i's rank-order list is empty we have $\overline{O}[i] = -1$, denoting no upper bound, otherwise $\overline{O}[i] = k$ where k is the final element of i's list. If i's list is not of full length and does not contain an ex-ante clearly feasible program, we have $\underline{O}[i] = -2$, denoting that the greatest lower bound on u_{i0}^0 is given by $\max\{u_{ik} : k \text{ unranked, on-platform, not ex-ante clearly infeasible}\}$, otherwise we have $\underline{O}[i] = -1$, indicating that there is no lower bound.

Next, we construct upper and lower bounds, $\overline{A}$ and $\underline{A}$ respectively, on aftermarket availability. If i is assigned to j in the match, or enrolls in j, we have $\overline{A}[j, i] = \underline{A}[j, i] = 1$, indicating that the program is always available. If i enrolls in on-platform program k, and ranked j ahead of k on her rank-order list, then j must not have offered i a position, and hence we have $\overline{A}[j, i] = \underline{A}[j, i] = 0$. In the remaining cases, we have $\overline{A}[j, i] = 1$ and $\underline{A}[j, i] = 0$.

Finally, let

$$J_i^{\text{uncertain}} = \{j \in J_{t(i)} : j \text{ ex-post aftermarket-feasible for } i\} \setminus (\{0\} \cup \{\text{placement}_i\})$$

denote the set of programs which are available to i if and only if there is a positive realization of a_{ij}^*. We use this set to determine which (i, j) cells are relevant for the steps in which we draw a_{ij} and α.

B.4.3 Starting values

We find feasible starting values for utility and availability as follows. If j is on-platform and is ranked in the n_jth position on i's relevant rank-order list, ℓ_i, then $u_{ij} = 1/n_j$. If j is on-platform but not ranked, then $u_{ij} = -1$. If j is off-platform and i does not enroll in it, then $u_{ij} \sim \text{Uniform}[-1e^{-4}, 1e^{-4}]$. If j is off-platform and i enrolls in it, then $u_{ij} = 2$.

At all programs for which $a_{ij} = 1$ is possible given these utilities, we set $a^*_{ij} = 1.0$. Otherwise, we set $a^*_{ij} = -1.0$. We set $u^0_{i0} = -0.001$ unless there is a clearly-feasible school on i's rank-order list, in which case we have $u^0_{i0} = .001$. We set $u^1_{i0} = 0.002$ for all students.

These starting values ensure that i enrolls in his most-preferred program among those that he could have attended, and that the rank-order list is truthful after restricting to the relevant set.

We start with $h_{i,enroll_i} = 1$ if i graduated, and $h_{i,enroll_i} = -.5$ if i did not. We draw initial random coefficients $\eta^x_i \sim N(0, I)$, i.i.d. across people.

B.4.4 Estimation procedure

Our procedure iterates the following draws from conditional posterior distributions.

1. For each market t, for each type $g \in G$, for each $i \in \{1, \ldots, N_t\}$ of type g:

   (a) Draw $u^0_{i0} | u_i, u^1_{i0}, \ell_i, enroll_i, \gamma^0_{g(i)}, \sigma^2_{0,0,g(i)}$.

   (b) Draw $u^1_{i0} | u_i, u^0_{i0}, \ell_i, enroll_i, \gamma^1_{g(i)}, \sigma^2_{0,1,g(i)}$.

   (c) for each $j \in \{1, \ldots, J_t\}$

   - If $j \neq enroll_i$:

     i. Draw $u_{ij} | a_i, u^0_{i0}, u^1_{i0}, u_{i,-j}, \ell_i, enroll_i, \eta^x_i, \delta_{j,g(i)}, \lambda_{g(i)}$.

     ii. Draw $a^*_{ij} | u_i, u^0_{i0}, u^1_{i0}, \ell_i, enroll_i, \alpha_{g(i)}$.

   - Else:

     i. Draw $u_{ij} | a_i, u^0_{i0}, u^1_{i0}, u_{i,-j}, \ell_i, enroll_i, \eta^x_i, \delta_{j,g(i)}, \lambda_{g(i)}, h_{ij}, \tilde{\bar{\beta}}_{g(i)}, \tilde{\beta}^{w,p}_{g(i)}, \beta^z_{g(i)}$.

     ii. Draw $h_{ij} | graduate_{ij}, u_{ij}, \tilde{\bar{\beta}}_{g(i)}, \tilde{\beta}^{w,p}_{g(i)}, \beta^z_{g(i)}$.

     iii. Draw $a^*_{ij} | u_i, u^0_{i0}, u^1_{i0}, \ell_i, enroll_i, \alpha$.

(d) Draw $\eta_i^x | u_i, \Sigma^{rc}_{g(i)}$

2. for each type $g \in G$:

   (a) Draw $\sigma^2_{0,0,g} | \{u^0_{i0} : g(i) = g\}, \gamma^0_g$.

   (b) Draw $\sigma^2_{0,1,g} | \{u^1_{i0} : g(i) = g\}, \gamma^1_g$.

   (c) Draw $\gamma^0_g | \{u^0_{i0} : g(i) = g\}, \sigma^2_{0,0,g}$.

   (d) Draw $\gamma^1_g | \{u^1_{i0} : g(i) = g\}, \sigma^2_{0,1,g}$.

   (e) Draw $(\delta_g, \lambda_g) | \{(u_i, \eta^x_i) : g(i) = g\}$.

   (f) Draw $\Sigma^{rc} | \{\eta^x_i : g(i) = g\}$

   (g) Draw $\alpha | \{a^*_{ij} | g(i) = g\}$.

   (h) Draw $(\tilde{\bar{\beta}}_g, \tilde{\beta}^{w,p}_g, \beta^z_g, \rho_g) | \{h_{i,\text{enroll}_i}, u_i : g(i) = g\}$.

B.4.5 Priors

We use standard conjugate priors. We choose the prior parameters to be uninformative. Program effects and other linear-index parameters $(\delta, \lambda, \eta, \gamma^1, \gamma^0, \tilde{\bar{\beta}}, \tilde{\beta}^{w,p}, \beta^z, \alpha)$ have independent $Normal(0, 10 * I)$ priors, where I is the identity matrix. Scalar variances have $InverseGamma(10, 10)$ priors. Variance-covariance matrices of size (k, k) have $InverseWishart(k+1, 10 * I)$ priors.

B.4.6 Counterfactuals, Uniqueness, Model Fit, and Convergence

We conduct 7500 iterations. We throw out the first 2500 as burn-in. After dropping the initial burn-in draws, we simulate all counterfactuals and model-fit exercises at every 200th iteration. To simulate the counterfactuals discussed in the paper, we use the latent utility values (u, u_0) associated with the relevant MCMC iteration, but draw aftermarket frictions a^* and human-capital shocks h freshly from their distribution conditional on α and human-capital parameters, respectively.

To obtain starting values for parameters for our main chain of 7500 iterations, we first conduct 20000 iterations on a random 15% subsample. Taking the final parameter draw from this procedure, we construct starting values of latent variable for the full sample, then iterate the Gibbs steps that update the latent variables 2500 times, holding the parameters fixed. Once we've obtained starting values in this way, we begin the 7500 iterations described above.

For the model fit exercises in the paper, we do not condition on observed choices or on estimated latent availability, human capital, or utility draws consistent with those choices. Rather, we draw utilities, human-capital shocks, and availability shocks from their distributions conditional on parameters.[5] We conduct two draws of latent variables at every 200th iteration along the Markov chain and compute model fit under each draw. We then report averages over parameters and latent utility/availability/human-capital draws.

To check whether there is a unique stable matching under plausible estimates of true preferences, we use these model fit draws, as well as the draws used for counterfactual simulations. Uniqueness matters because students cannot gain by deviating if there is a unique stable matching under the true preferences; in contrast, if the student-optimal and college-optimal stable matchings do not coincide, then under the college-proposing DA algorithm it is possible for students to misreport their preferences in such a way as to achieve the student-optimal matching with respect to their true preferences (Demange et al., 1987; Dubins and Freedman, 1981).

We consider all of the utility draws that we used for counterfactuals. For computational reasons, when we revisit the "model fit" draws, we use the first set of draws of utilities at every 200th iteration along the Markov chain in each year after "burn-in". Thus we consider a total of 78 "model fit" draws, consisting of 26 draws per year for the years 2010-2012 which do not condition on observed behavior, in addition to 26 draws from 2012 which condition on the observed rank-order lists and which were used for our counterfactuals. In each simulated market, we simulate the student-proposing and college-proposing deferred acceptance procedures under truthful reporting and check whether they coincide.[6]

[5]Re-simulating all latent variables is important for assessing fit. If we were to use the utility values that were drawn in the Markov chain, which are by construction consistent with optimality of the applicant's observed behavior, we would exactly replicate the initial match, and the characteristics of students matched to each program in simulations would be identical to those in the data.

[6]We consider the "final" allocation produced by running the SPDA or CPDA algorithm once with all programs, including the second outside option, participating. We maintain priorities and eligibility as in the data. Because we are allowing potentially very long lists, we ignore the "top-4" restrictions imposed by the

Averaging over simulated markets, the mean number of students who receive different assignments under the CPDA and SPDA algorithms is less than 0.001% of the total. In particular, we find that in 19 out of 26 draws used for counterfactuals there is a unique final stable matching. In six more cases the student-optimal and college-optimal stable match assignments differ for two students, and in one draw the assignments differ for four students. In the "model fit" draws, results are similar. In the 2012 "model fit" simulations, there are 21 out of 26 draws in which there is a unique stable matching, four draws in which two students' assignments differ between the student-optimal and college-optimal stable matchings, and one draw in which four students' assignments differ. In 2011, there are 18 draws with a unique stable matching, 7 draws in which two students' assignments differ, and one in which four students' assignments differ. In 2010, 19 of 26 draws have a unique stable matching. There are 5 of 26 draws in which two students' assignments differ, one in which three students' assignments differ, and one with five differences.

Finding few cases in which students have possible manipulations, with a similar probability under observed ROLs as under the "true preferences," is consistent with our Assumption 1, which requires truthtelling within the relevant set.

B.4.7 Details on updating latent variables

Conditional on other variables and parameters, the terms a^*_{ij}, u_{i0}, u_{ij}, and h_{i,enroll_i} are drawn from truncated normal distributions. Truncation bounds on u_{ij} and u_{i0} come from two sources: optimality of the submitted application and optimality of the enrollment decision. Bounds on h come from the observed graduation outcome. Constraints on a come from i's choices and utilities, as we describe below. Our approach here builds on McCulloch and Rossi (1994) to allow for partial rank-order data as well as the constraints implied by the enrollment decision when there is a latent "availability set".

When updating u^0_{i0}, the construction of bounds is as follows. The first outside option u^0_{i0} is bounded below by the maximum over the utilities of all programs that are not clearly infeasible and are not listed, and by $\max\{u_{ik} : a_{ik} = 1\}$ whenever $\text{enroll}_i = 0$ and $u^1_{i0} <$

University of Chile and the Pontifical Catholic University of Chile which drop applications that rank these programs lower than fourth place.

$\max\{u_{ik} : a_{ik} = 1\}$. u_{i0}^0 is bounded above by the utility of the final listed program, if any, and by u_{i,enroll_i} if $\text{enroll}_i \neq 0$. We draw u_{i0}^0 from a $\text{Normal}(z_i\gamma_{g(i)}^0, \sigma_{0,0,g(i)}^2)$ distribution truncated at the min of the upper bounds and the max of the lower bounds.

The second outside option, u_{i0}^1, is bounded below by $\max\{u_{ik} : a_{ik} = 1\}$ whenever we have $\text{enroll}_i = 0$ and $u_{i0}^0 < \max\{u_{ik} : a_{ik} = 1\}$, and is bounded above by u_{i,enroll_i} if $\text{enroll}_i \neq 0$. We draw u_{i0}^1 from a $\text{Normal}(z_i\gamma_{g(i)}^1, \sigma_{0,1,g(i)}^2)$ distribution truncated at these bounds.

Applications are optimal if and only if the following constraints on u_{ij} are satisfied: If $j \in J_i^{relevant}$ was ranked mth on ℓ_i, then u_{ij} is bounded above by the utility of the $m-1$th program whenever $m > 1$, and bounded below by the utility of the $m+1$th option if one exists. If j is the final option in ℓ_i, then u_{ij} is bounded below by u_{i0} if ℓ_i is not of full length and j is ex-ante marginal, and otherwise by the max of u_{i0}^0 and highest-utility program in $J_i^{relevant}$ that was not listed. If j was on-platform, not ex-ante clearly infeasible, and not listed, its utility is bounded above by u_{i0}^0 if ℓ_i does not contain an ex ante strictly feasible program, and by the final listed program otherwise.

These constraints are captured by the matrices $\overline{J}$ and $\underline{J}$.

The decision to enroll in j implies that it is preferred to all other programs k for which $a_{ik} = 1$, providing additional constraints. If $j = enroll_i$ then

$$u_{ij} > \max\{u_{ik} : k \geq 0,\ k \neq j,\ k \text{ ex-post feasible for } i,\ a_{ik} = 1\}.$$

If $j \neq enroll_i$ then whenever $a_{ij} = 1$ we must have $u_{ij} < u_{i,enroll_i}$. We adopt the convention $a_{i0} = 1$.

The lower bound on u_{ij} is the maximum of the lower bound from the enrollment decision (if any) and the lower bound from applications (if any). The upper bound is analogous.

let

$$\mu_{ij}^u \equiv \delta_{j,g(i)} + w_{ij}\lambda_{g(i)}^w + x_{j,t(i)}\eta_i^x + p_{ij}\lambda_{g(i)}^p.$$

If $j > 0$ is not the program in which i enrolls, then u_{ij} is drawn from a truncated $\text{Normal}(\mu_{ij}^u, 1)$ distribution with truncation bounds given above.

To update u_{i,enroll_i}, we must condition on h_{i,enroll_i}, which affects the mean and variance of the truncated normal as follows. Fixing $j = \text{enroll}_i$, let

$$\mu_{ij}^h \equiv h_{ij} - u_{ij}\rho_{g(i)} = \tilde{\tilde{\beta}}_{j,g(i)} + z_i\beta_{g(i)}^z + w_{ij}\tilde{\beta}_{g(i)}^w + p_{ij}\tilde{\beta}_{g(i)}^p + \tilde{\nu}_{ij}$$

denote the portion of h_{ij} that is independent of utility shocks, conditional on observables. The likelihood of $u_{ij}|h_{ij}$ is proportional to $\phi(u_{ij} - \mu_{ij}^u)\phi(h_{ij} - \mu_h)$. With some algebra one can show:

$$u_{ij}|h_{ij}, \eta_i^x, \delta, \lambda \sim N\left((\mu_{ij}^u + \rho_{g(i)}(h_{ij} - \mu_{ij}^h))\tilde{\sigma}^2, \tilde{\sigma}^2\right),$$

where $\tilde{\sigma}^2 = \frac{1}{\rho_{g(i)}+1}$.

Conditional on enrolling in $j > 0$ and on u_{ij}, human capital h_{ij} is distributed according to a Normal$(\mu_{ij}^h + u_{ij}\rho_{g(i)}, 1)$ distribution truncated from above at zero if i does not graduate and from below at zero if i graduates. An advantage of our "tilde" reparameterization is that u_{ij} may be treated as an ordinary covariate at this step.

Some elements of a_{ij} are observed. If i enrolls in program j or was placed in j then $a_{ij}^* > 0$. If i enrolls in his original match or in a waitlist offer, then $a_{ik} = 0$ for all waitlisted programs k that i ranks above where he enrolls. When a_{ij} is not observed, and is potentially relevant because j is off-platform or i is waitlisted at j, we have $a_{ij} = 0$ whenever $u_{ij} > u_{i,enroll_i}$, and $a_{ij} \sim Bernoulli(1-\alpha)$ otherwise.

We update a_{ij} only for (i,j) cells for which $j \in J_i^{\text{uncertain}}$, and use only these values of (i,j) when updating the availability parameters α.

B.4.8 "Special" Admissions

There are a handful of cases in the data in which people enroll in on-platform options that were infeasible, that they did not apply to via the platform, or that they ranked below their placed option. These cases receive special treatment as follows.

Let program j denote the program in which person i enrolls, and suppose that j is on platform but is not part of the relevant set (e.g. because i did not apply to it, or because it was ex-ante clearly infeasible), or is relevant but ranked below i's placed program. We assume that the constraints implied by optimal enrollment decisions continue to apply, and

j is assumed to be available (i.e. to have $a_{ij} = 1$) and maximize utility among available options. However, the application decision does not impose any constraints involving j. In the event that j was part of the relevant set, it is removed from this set. In addition, because selection into enrollment may differ for these programs, the student's outcome at this program is not used when updating graduation/human capital parameters, nor do we use this person-program pair to update friction parameters.

A leading example is the case in which person i enrolls in some program j that was ex-ante clearly infeasible. In most cases, this event occurs because the person applied to the program, and was admitted off of the waitlist, but the program's cutoff was outside of the person's bandwidth. In a handful of additional cases, people match to programs via alternative channels such as the BEA program, which reserves some seats for students with high class rank.[7] We emphasize that our assumptions here are weaker than if we had treated every ex-post waitlist-feasible program as ex-ante marginal, in the sense that we impose fewer constraints on utilities.[8]

B.4.9 Details on updating other parameters

The remaining steps are standard. We provide details here for completeness. Similar derivations are available for linear parameters and covariance matrices in the online appendix to Agarwal and Somaini (2018).

Posterior distribution of regression coefficients: We first state a standard result which we use below. Suppose we have a multilinear regression model of the form

$$v_i = X_i b + e_i,$$

[7]In a very small number of cases, people enroll in programs that were part of the relevant set but that they dispreferred to their placed program. It impossible to enroll via the standard process in an option that was ranked below the placed program, but it may be possible for people to do so if they have received a BEA offer, or other offer outside of the match at the dispreferred program, in addition to their "standard" on-platform offer.

[8]We have estimated a version of the model in which all ex-post waitlist-feasible programs that would have been ex-ante clearly infeasible are coded as ex-ante marginal for the relevant applicants. This version makes use of more waitlist data because "long-shot" admisssions offers off of waitlists are not treated as "special admissions". Results are quantitatively very similar.

where $v \in \mathbb{R}^k$ for $i = 1, \ldots, N$ is a vector of "outcomes", x_i is a matrix of "regressors" of dimension (J, K), and the error term is distributed as $e_i \sim \text{MvNormal}(0, \Sigma)$ for some positive-definite matrix Σ, iid across i. Let us place a $\text{MvNormal}(\bar{b}, A^{-1})$ prior on b, for some positive-definite k-by-k matrix A and k-vector $\bar{b}$. Then the posterior distribution of "regression coefficients" b is given by:[9]

$$(b|\{v_i, X_i\}_{i=1,\ldots,N}) \sim \text{MvNormal}(\tilde{b}, V),$$

where

$$\begin{aligned}
V &= (X^* X^{*\prime} + A)^{-1} \\
\tilde{b} &= V(X^* v^* + A\bar{b}) \\
X^* &= \begin{bmatrix} X_1^* \\ \vdots \\ X_N^* \end{bmatrix} \\
X_i^* &= C X_i \\
v_i^* &= C v_i,
\end{aligned}$$

and C is the upper-triangular Cholesky factor from the Cholesky decomposition of Σ.

We now give details about specific steps:

- For each i, we draw η_i^x using the "regression coefficients" formula above in which the coefficients are $b = \eta_i^x$, the "outcomes" are $v_i = u_i - (\delta_{j,g(i)} + z_i \lambda_{g(i)}^z + w_{ij} \lambda_{g(i)}^w + p_{ij} \lambda_{g(i)}^p)$, the regressors are $X_{j,t(i)}$, the covariance of utility shocks is the identity matrix and hence $C' = I$, and the prior is $\eta_i^x \sim N(0, \Sigma^{g(i)})$

- For each type $g \in G$:

  - Draw $\sigma_{0,0,g}^2 | \{u_{i0}^0 : g(i) = g\}$: When the prior is $\sigma_{0,0,g}^2 \sim \text{InverseGamma}(\alpha, \beta)$, the posterior is distributed as $\text{InverseGamma}(\alpha + \sum_t N_{gt}, \beta + \sum_t (u_{i0} - z_i \gamma^0)^2)$ where N_{gt} denotes the number of people in cohort t who are of type g.

[9] This result is standard. e.g. see Gelman et al. (2013).

– Draw $\sigma^2_{0,1,g}|\{u^0_{i0} : g(i) = g\}$: When the prior is $\sigma^2_{0,1,g} \sim \text{InverseGamma}(\alpha, \beta)$, the posterior is distributed as $\text{InverseGamma}(\alpha + \sum_t N_{gt}, \beta + \sum_t (u_{i0} - z_i\gamma^1)^2)$.

– Draw $\gamma^0_g|\{u^0_{i0} : g(i) = g\}, \sigma^2_{0,0,g}$: use the "regression coefficients" formula above, with "outcomes" u^0_{i0}, "regressors" z_i, and error terms distributed according to a $N(0, \sigma^2_{0,0,g_i})$.

– Draw $\gamma^1_g|\{u^1_{i0} : g(i) = g\}, \sigma^2_{0,1,g}$: use the "regression coefficients" formula above, with "outcomes" u^1_{i0}, "regressors" z_i, and error terms distributed according to a $N(0, \sigma^2_{0,1,g_i})$.

– Draw $(\delta_g, \lambda_g)|\{(u_i, \eta^x_i) : g(i) = g\}$: use the "regression coefficients" formula above, with outcomes $(u_i - x_{j,t(i)}\eta^x_i)$, regressors (I, z_i, w_{ij}, p_{ij}), and $\Sigma = I$.

– Draw $\Sigma^g|\{\eta^x_i : g(i) = g\}$: when the prior distribution is $\Sigma^g \sim IW(v_0, S)$, the posterior is

$$\Sigma^g|\{\eta^x_i : g(i) = g\} \sim IW\left(v_0 + \sum_t N_{gt}, S + \sum_t \sum_{i=1}^{N_{gt}} \eta^x_i \eta^{x\prime}_i\right).$$

– Draw $\alpha_g|\{a^*_{ij}|g(i) = g\}$: Restrict attention to (i, j) cells for which $j \in J^{\text{uncertain}}_i$. Use the "regression coefficients" formula above, with iid standard normal error terms.

– Draw $(\tilde{\tilde{\beta}}_g, \tilde{\beta}_g, \rho_g)|\{h_{i,\text{enroll}_i}, u_i, \text{graduate}_i : g(i) = g\}$: use the "regression coefficients" formula above, restricting to people who enrolled in some $j > 0$, with outcomes h_{i,enroll_i}, regressors $(I, z_i, w_{i,\text{enroll}_i}, p_{i,\text{enroll}_i}, u_{i,\text{enroll}_i})$, and $\Sigma = I$.

B.4.10 Notation table

Notation Table - Model

Indices	
i	Student
j	Program
t	Application cohort

g	Observable student group; interaction of student gender and type of high school attended
N_t	Set of students in year t
$\text{G8}, \text{G25}$	Denotes whether program belongs to G8 or G25 university
$J_t^{\text{on}}, J_t^{\text{off}}$	On-platform options in year t, Off-platform options in year t
$J_t = J_t^{\text{on}} \cap J_t^{\text{off}}$	Set of all inside options in year t
Observables	
$w_{i,j}$	Observed student-program match-level observables
$x_{j,t}$	Program-year characteristics
z_i	Vector of student-level observables
$p_{i,j}$	Price that student i would pay to attend program j after government and institutional aid
$v_{i,j}$	Observed student-level and student-program match-level variables that affect aftermarket availability
placement_i	Program j that student i was assigned in the match, 0 if none
enroll_i	Program j in which student i enrolled, 0 if student i did not enroll
graduate_i	graduation outcome of student i
$index_{i,j}$	Application score of eligible student i at program j; weighted sum of exam scores; 0 if ineligible
$\pi_{j,t}$	Minimum value of $index_{i,j}$ among students placed in program j during the initial match
$\underline{\pi}_{j,t}, \overline{\pi}_{j,t}$	Cutoff values that determine whether program j is ex-ante infeasible, marginal, or clearly feasible
J_i^{relevant}	On-platform programs that are relevant for i: not ex-ante infeasible or dispreferred to the best clearly-feasible option
ℓ_i	(Restricted) submitted rank-order list; consists of all elements of J_i^{relevant} ranked according to preferences
$J_i^{\text{uncertain}}$	Programs that are ex-post aftermarket-feasible for i but not ex-post match-feasible; admit i if and only if $a_{ij}^* > 0$
Shocks and Latent Variables[10]	
$\epsilon_{i,j}, \aleph_{i,j}, \nu_{i,j}$	iid match-level shocks to preferences, aftermarket availability, human capital
$\boldsymbol{\eta_i^x}$	Unobserved random tastes for program-year characteristics
$\mathbf{u_{ij}}$	Utility that student i derives from attending program j
u_{i0}	Utility that student i derives from their outside option. Max of u_{i0}^0 and u_{i0}^1
$\mathbf{u_{i0}^0}$	Value of student i's first outside option; known to i before applications are due
$\mathbf{u_{i0}^1}$	Value of student i's second outside option; learned during the aftermarket
$\mathbf{a_{ij}^*}$	Latent variable determining whether j is able to successfully contact i in the aftermarket
a_{ij}	Indicator for whether program j is available to student i

[10] Bolded variables are stored, updated, and tracked during estimation. Non-bolded shocks and indicators are constructed as needed from the values of the parameters and tracked latent variables.

$\mathbf{h_{ij}}$	Human capital index determining whether student i graduates from program j
Parameters	
α_g	Aftermarket friction parameters
$\overline{\beta}_{j,g}, \beta_g^z, \beta_g^w, \beta_g^p$	Coefficients of human capital function
$\delta_{j,g}$	Mean utility of program j for students in group g
λ_g^w	Group specific coefficients for match-level term w_{ij}
λ_g^z	Group specific coefficients on student-level terms z_i, which shift the value of all inside options relative to the value of attending a non-G33 program
λ_g^p	Group specific coefficient for price $p_{i,j}$
γ_g^z	Group specific coefficients on student-level terms z_i that enter the second outside option component.
ρ_g	Group specific correlation between utility shock and human-capital shock
Σ_g	Group specific covariance matrix of random coefficients

This section describes the tradeoffs involved in our choice of bandwidth in constructing the set of *relevant* applications for each student. In practice we choose a bandwidth

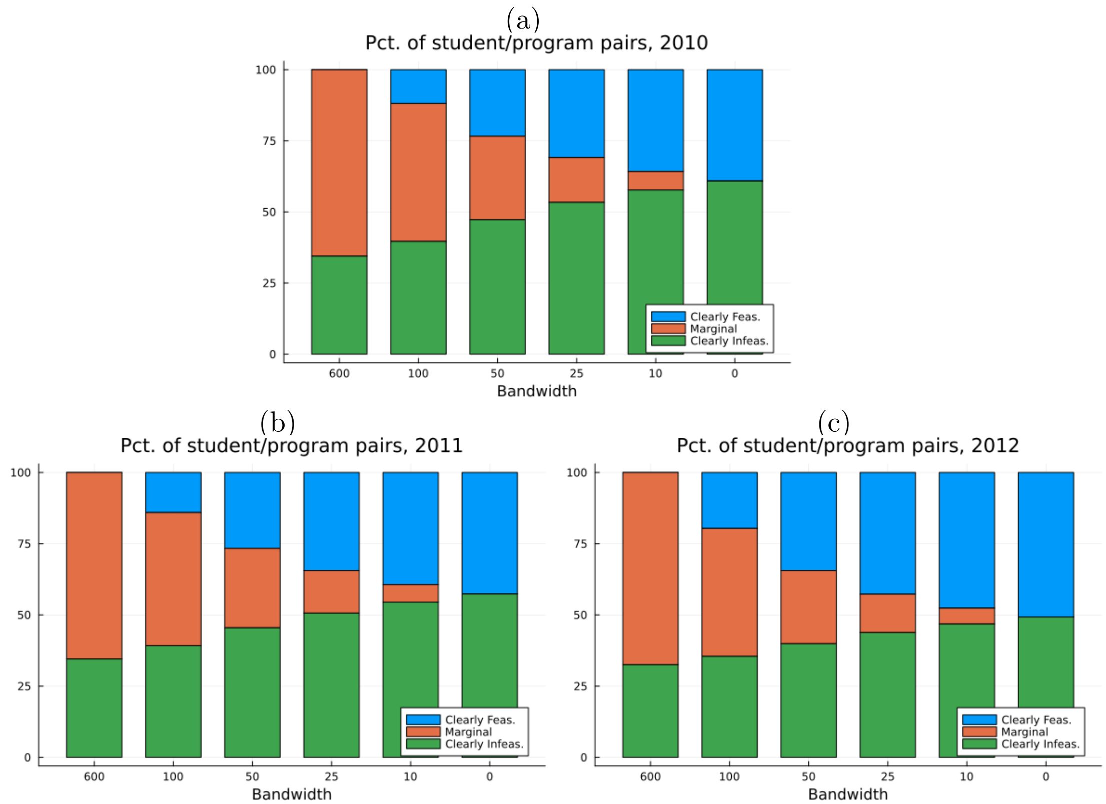
$$\overline{\pi}_{jt} - \pi_{jt} = \pi_{jt} - \underline{\pi}_{jt} = 25$$

for all j, t (recall that the standard deviation of student test scores is 110).

In this section, we consider bandwidth specifications of 600, 100, 50, 25, 10, and 0 points. At a bandwidth of 600 points, Every program at which i is eligible to apply is ex-ante marginal, and hence only programs in which the student violates an eligibility rule are clearly infeasible. As bandwidths shrink, we approach the case in which the relevant set contains at most one program: namely, the most-preferred ex-post feasible program, if there is one that is preferred to the outside option.

We first show how the share of ex-ante strictly-feasible, strictly infeasible, and marginal programs varies with bandwidth. Figure B.4 shows that as the bandwidth shrinks, fewer program-student pairs are classified as ex-ante marginal. Roughly 40% of program-student pairs are impossible because the students do not meet program-specific eligibility rules; these pairs are always ex-ante strictly infeasible, even in the case of bandwidth=600. As the bandwidth approaches zero, every program becomes either strictly infeasible or strictly feasible.

Figure B.4: Ex-ante classification of (program,student) pairs

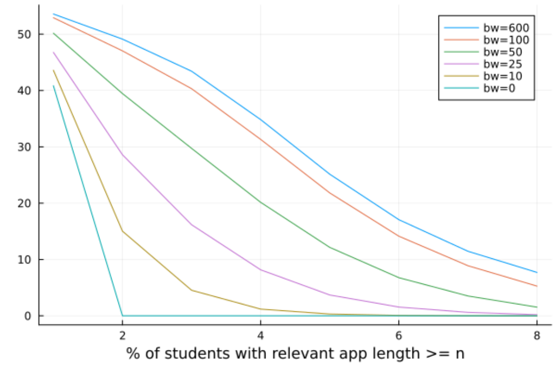


Note: the bars show the share of student-program pairs which are ex-ante clearly feasible, ex-ante marginal, and ex-ante clearly infeasible as a function of year and bandwidth choice.

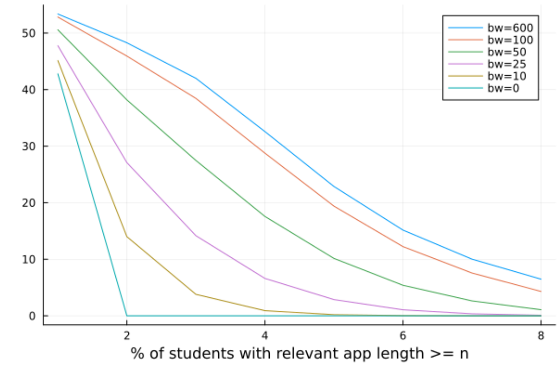
Next, consistent with the shrinking share of programs that are ex-ante marginal, we show that the length of relevant applications declines as the bandwidth shrinks. Figure B.5 shows inverse-CDFs of relevant application length by year and bandwidth. An important observation is that, for bandwidths less than 50, the list length constraint very rarely binds.

Figure B.5: Length of relevant application lists

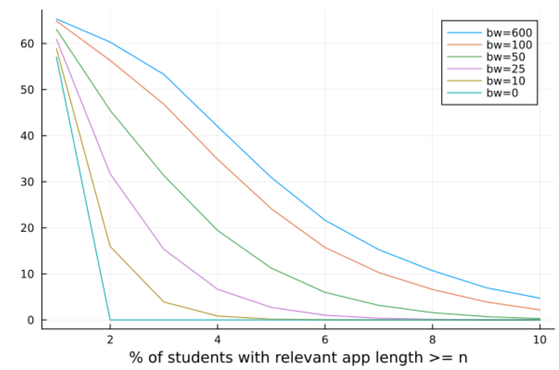
(a) Relevant Application Length, 2010



(b) Relevant Application Length, 2011



(c) Relevant Application Length, 2012



Note: lines show the probability that the application contains at least n programs, by year and bandwidth.

Finally, Table B.9 describes the share of students for whom list length constraints might possibly bind after dropping irrelevant applications. There are two relevant measures here. We say that "list length may bind" for estimation if the restricted application is of the maximum length and does not contain an ex-ante clearly feasible program. In this case, it is possible that a student who reports elements of his relevant set in order of true preferences could run out of space to rank all options. Second, we say an application is "possibly unstable" if the restricted application is of maximum length and does not contain an ex-post match-feasible program. Because every ex-ante clearly feasible program is ex-post match-feasible, this set is a subset of the set of applications for which list length may bind. A person whose application is not "possibly unstable" is necessarily matched to his most-preferred match-feasible program, if any, and hence does not have justified envy. Therefore, the share of "possibly unstable" applications places an upper bound on the amount of justified envy. We find that this share is negligible under a 25-point bandwidth specification. The table shows that there is at most one person, who applied in 2010, who could have justified envy.

Table B.9: When do Constraints on List Length Bind

Year	Bandwidth	% List Length Binds	% Possibly Unstable	N Possibly Unstable
2010	600	7.7	0.85	1387
2010	100	5.03	0.33	543
2010	50	1.28	0.03	45
2010	25	0.13	0.0	1
2010	10	0.0	0.0	0
2010	0	0.0	0.0	0
2011	600	6.48	0.51	804
2011	100	4.11	0.19	296
2011	50	0.85	0.01	16
2011	25	0.05	0.0	0
2011	10	0.0	0.0	0
2011	0	0.0	0.0	0
2012	600	4.71	0.23	366
2012	100	1.99	0.06	103
2012	50	0.21	0.0	1
2012	25	0.01	0.0	0
2012	10	0.0	0.0	0
2012	0	0.0	0.0	0

This table shows the share of students for whom list length constraints may bind after restricting to relevant applications J_i^{relevant}. Bandwidth: symmetric bandwidth around π_{jt}; j is ex-ante marginal for i if i is eligible to apply to i and $|index_{ij} - \pi_{j,t(i)}| <$ Bandwidth. "% List Length Binds": percentage of all students with ℓ_i of full length and containing no ex-ante strictly-feasible program. "% Possibly Unstable", "N Possibly Unstable": Percentage and count of students with ℓ_i of full length and containing no ex-post match-feasible program.

Appendix C

Chapter 3

C.1 Additional Tables

C.1.1 Description and Implementation

The teacher exit exam, called 'INICIA', consists of a set of tests taken by newly graduated teachers, implemented for the first time in year 2009 and up to 2015.[1] The exam intends to measure four dimensions: (1) disciplinary knowledge (e.g. math knowledge for math teachers); (2) pedagogical knowledge (intended to measure if test takers can explain concepts in a coherent way); (3) writing skills, and (4) capacity to use ICT (information and communication technologies) for teaching purposes. In 2016 the ministry of education administered 'Diagnostica' which also evaluates disciplinary and pedagogical skills and is taken the year before graduation in different universities.

INICIA and Diagnostica's main objective is to assess the qualification of recent teacher graduates. The information produced by the exit exams is thought to be useful for the institutions training teachers, policy makers and the test-takers themselves, although there are no associated consequences to its results.[2] Results are published at the institution level, with individual-level information remaining confidential. The exam's application was gradually

[1] Typically, students in their last semester of class, or just graduated students taking the exam before getting a job.

[2] This may change in the near future. During 2015, the Ministry of Education sent a bill to the Congress in order to make the INICIA test mandatory and to establish minimum performance levels to be allowed to teach at least in the public sector.

expanded by year and by the level at which teachers specialize (i.e. pre-school, primary and secondary), as summarized in Table C.1.[3]

Table C.1: Teacher Exit Exam: Tests Implemented by Year and Teacher Specialization Level

		Year						
Level	Test	2009	2010	2011	2012	2014	2015	2016
Pre-school	Disciplinary	✓	✓	✓	✓	✓	✓	✓
	Pedagogical		✓	✓	✓	✓	✓	✓
	Writing	✓	✓	✓	✓			
	ICT							
Primary	Disciplinary	✓	✓	✓	✓	✓	✓	✓
	Pedagogical	✓	✓	✓	✓	✓	✓	✓
	Writing	✓	✓	✓	✓			
	ICT	✓	✓	✓				
Secondary	Disciplinary				✓	✓	✓	✓
	Pedagogical				✓	✓	✓	✓
	Writing				✓			
	ICT							

Notes: 'Disciplinary' stands for the test measuring disciplinary knowledge; 'Writing' stands for the writing skills test; 'Pedagogical' stands for the Pedagogical knowledge test; 'ICT' stands for the test measuring the information and communications technology skills. Source: MINEDUC (2012).

The Inicia exam is voluntary. Formally, the Ministry of Education invites graduate institutions that train teachers (i.e. institutions offering teacher or education degrees) to participate in the INICIA exam every year. In the case of Diagnostica, the exam is mandatory and is administered to all students of pedagogy in certified institutions of education. Table C.2 summarizes the number of institutions invited, and those that participated. From years 2009 to 2012, around 80% of the invited institutions participated, which means that at least some of its graduates took the test. Institutions can encourage their graduates to participate, but can not force them to do so.

Table C.3 summarizes the number of potential test-takers, the ones that sign-up and those that take at least one test, by year. Every INICIA test before 2012 was held in December of

[3] In 2013 the exam was not applied.

each year, which coincides with the end of the academic year in Chile. Due to administrative issues, the 2012 INICIA test was held in April of 2013. This delay seems to be the reason behind the low take-up of that year's test (see Table C.3). By that time of the year, most new teachers would be working, because the academic year starts in March. Also, it is likely that graduates lose the connection with their universities after a while. After this episode, the Ministry of Education decided to postpone the application of the 2013 INICIA, supposed to be held in December 2013, to December of 2014, combining evaluations 2013 and 2014 into a single sitting.

Table C.2: Teacher Exit Exam: Invited and Participating Institutions by Year

	Application	Number of Institutions		Participation
Year	Date	Invited	Participating	Percentage
2009	Dec. 2009	54	43	80%
2010	Dec. 2010	56	43	77%
2011	Dec. 2011	59	49	83%
2012	Apr. 2013	58	50	86%
2014	Apr. 2014		50	-
2015	Dec. 2015		50	-
2016*	Apr. 2016		50	-

Notes: Invited institutions correspond to those that train primary school teachers (every year), pre-school teachers (years 2009-2012 and 2016) and secondary school teachers (year 2012 and 2016). Participating institutions are the ones for which at least one of their graduates takes one or more of the tests described in Table C.1. Participation percentage displays the number of participating institutions as a percentage of the number of invited institutions. * In 2016 the corresponding exit exam was Diagnostica and was mandatory. Source: MINEDUC (2012).

C.1.2 Teacher Exit Exam Results and PSU Scores

Institutional Reports. The Ministry of Education publishes each year a presentation with the INICIA exam results.[4] According to these institutional reports, the results achieved by the education graduates are below what is needed to perform adequately as a teacher.

[4]For years 2008 to 2010, results were mainly published as the percentage of correct answers achieved in each test, without a statement on what was considered a good outcome. For the 2011 and later exams, the MINEDUC implemented three categories to classify test-takers according to their performance, based on the knowledge and skills necessary to begin their career as a classroom teachers: Outstanding, Acceptable and Unsatisfactory.

Table C.3: Teacher Exit Exam: Test-Takers by Year

	Number of Test-Takers			
Year	Potential	Signed-up	Participated	Take-up
2008	5,250	3,006	1,994	38%
2009	7,979	4,527	3,223	40%
2010	8,594	4,681	3,616	42%
2011	8,069	4,874	3,271	41%
2012	10,351	2,443	1,443	14%
2014	15,013	714	682	4%
2015	14,472	1,993	1,916	13%
2016*	20,215	17,971	12,741	63%

Notes: Potential Test-Takers correspond to the number of graduates from previous year. Those that sign-up to take the test are displayed in column 2. The number individuals that took at least one test described in Table C.1 is shown in column 3. Column 4 presents the number of actual test-takers as a percentage of the potential test-takers. * In 2016 the corresponding exit exam was Diagnostica and was mandatory. Source: MINEDUC (2012).

More than 60% of the test-takers that graduated as primary teachers fall in the 'unsatisfactory' category for the disciplinary tests in 2011 and 2012. The percentage is approximately 40% for the pedagogical test. For secondary teachers, the disciplinary tests by subject show the worse results in Mathematics, Biology, Physics and Chemistry, where about 70% of the test-takers fall in the 'unsatisfactory' category.

Microdata. The Ministry of Education provided us information from 2009 to 2015 on the INICIA exam, at the individual level and 2016 data on Diagnostica exam. We have microdata for more than 16K teachers with INICIA scores in at least one test and arround 13k teachers evaluated in Diagnostica. Table C.4 provides summary statistics for the four available tests. The first three rows report the percent of correct answers for the disciplinary, pedagogical and the ICT tests.[5] The last row shows the scores in the standardized writing test.

The Teacher's Public Evaluation System[6] (*Evaluación Docente* in Spanish, or ED onwards) is a mandatory assessment for all classroom teachers working in the public sector in

[5]The difference in the samples is explained by the fact that the Pedagogical test was not held in 2009, and the ICT test was not applied in 2012.

[6]For more details, see the institutional website www.docentemas.cl.

Table C.4: Exit Exam Summary statistics

Variable	Mean	Std. Dev.	Min	Max	N	Corr(PSU)
% of Correct Answers in:						
Disciplinary Test	0.57	0.14	0	1	20224	0.53
Pedagogical Test	0.58	0.13	0.05	1	18025	0.51
ICT Test	0.65	0.14	0.1	1	5517	0.51
Writing Test*	0.02	0.99	-6.65	3.26	11300	0.28

Notes: the last column displays the displays Spearman's rank correlations for each variable and PSU scores. The percentage of correct answers for the Disciplinary, Pedagogical and ICT tests has an associated a cutoff above which the performance is considered 'acceptable'. These cutoffs are .61, .61 and .65 for the Disciplinary, Pedagogical and ICT tests, respectively. These thresholds vary slightly over years, so they should be interpreted as proxies. For the writing test score, the cutoff is about -.09 SD from the mean. * Writing test is a standardized variable of the scores achieved by students by years.

Chile. The ED declared objective is 'to strengthen the teaching profession and the quality of education'. The assessment is composed by four components, with different weights: (i) a self-evaluation questionnaire (10%); (ii) a third-party reference report, filled by the school principal or supervisor (10%); (iii) one peer review (20%), and a teacher performance portfolio (60%). The portfolio component aims to collect direct evidence on teaching skills, pedagogical decisions and classroom practice. It includes two modules. In the first module, teachers plan a class defining its contents and related assessments. They are also asked questions about teaching practices. The second module consists in a videotaped class followed by a questionnaire on the students behavior and understanding, and the teacher's own performance.

The ED assigns a weighted score for each teacher using the components (i) to (iv) above. Then, the score is used to classify each teacher performance in one of four categories: unsatisfactory, basic, competent or outstanding. As opposed to the INICIA exit exam, the ED has consequences associated to performance. Teachers classified in the 'competent' or 'outstanding' categories can opt to receive a monetary bonus. Teachers classified in the unsatisfactory level need to retake the ED. If they remain in the unsatisfactory category after three times, they must leave their schools and can not teach again.

The ED has been implemented gradually since 2004 according to the level at which teachers specialize (pre-school, primary, secondary).[7] Table C.5 shows its year-level coverage for ten years 2004 to 2016.

Table C.5: Teacher Evaluation Implementation by Year and Level Taught

	Level		
Year	Preschool	Primary	Secondary
2004		✓	
2005		✓	✓
2006		✓	✓
2007		✓	✓
2008	✓	✓	✓
2009	✓	✓	✓
2010	✓	✓	✓
2011	✓	✓	✓
2012	✓	✓	✓
2013	✓	✓	✓
2014	✓	✓	✓
2015	✓	✓	✓
2016	✓	✓	✓
2017	✓	✓	✓

Notes: There are also other levels that have been incorporated to the teacher evaluation, like special education and education for adults, but we focus on primary and secondary levels in our analysis.

[7]There are also other levels that have been incorporated to the teacher evaluation, like special education and education for adults, but we focus on preschool, primary and secondary levels in our analysis.

The ED has carried out more than 174207 assessments for preschool, primary and secondary teachers from 2004 to 2013. Table C.6 exhibits the number of evaluations per teacher by year. The system has evaluated 101423 teachers at least once.[8] Approximately half of those teachers have been evaluated twice[9] (∼51K), and a about 35K have been evaluated more than three times.

Table C.6: Number of times teachers were evaluated from 2004-2013

Year	N:1	N:2	N:3	N:4	N:5	N:6
2004	1719	0	0	0	0	0
2005	10631	34	0	0	0	0
2006	13931	255	4	0	0	0
2007	10178	208	27	0	0	0
2008	14890	1104	21	0	0	0
2009	8567	5524	25	0	0	0
2010	3873	6422	121	3	0	0
2011	3498	7274	158	9	0	0
2012	3875	10496	693	17	0	0
2013	4343	6447	3818	57	4	0
2014	4993	3536	5118	167	7	0
2015	4620	2828	3889	339	16	0
2016	5707	3229	6118	899	35	1
2017	6667	3657	4757	2080	47	2
All	101423	50744	20456	1518	65	1

Notes: The table above represent the number of tests administered each year by the number of times a teacher was evaluated until each year.

For purposes of the analysis we will restrict the sample to teachers of primary or secondary education that were evaluated. This sample consist on 78513 teachers from the total of 101K evaluated (%77 of the total sample). Table C.7 reports the first ED results per category for all the 78.5K teachers in its first column. Only a 2 percent of the teachers resulted in an ‘unsatisfactory’ performance; 28% were classified as ‘basic’, 61% as ‘competent’ and 9% as

[8]From the 101K evaluated teachers a fraction has already retired from teaching. To get a sense of the coverage regarding those working currently in the public sector, consider that in year 2016 130K classroom teachers were working in the public sector (in either the preschool, primary or secondary level) and about 101K of them (∼78%) had been evaluated at least once.

[9]All teachers are supposed to be re-evaluated every four years, which the data does not fully support; teachers first classified in the unsatisfactory or basic category should be re-evaluated the year after or two years after the first evaluation respectively.

'outstanding'. It also shows the maximum scored achieved by category for some years. The thresholds to be in each category vary by year.

Table C.7: Teacher Evaluation Results 2004-2016

Classification	N obs	%	Max: 2004	Max: 2008	Max: 2012	Max: 2016
Outstanding	6875	8.8	3.63	3.59	3.21	3.37
Competent	48130	61.3	3.11	3.25	3	3.15
Basic	22091	28.1	2.64	2.67	2.79	2.9
Usatisfactory	1417	1.8	2	2.26	1.95	2.1
Total:	78513	100	2.84	2.94	2.74	2.88

C.1.3 Teacher Evaluation Results and PSU Scores

From the sample of first test taken by primary and secondary teachers we examine the correlation between ED and PSU scores. From the 78,513 teachers of primary and secondary education with ED scores about 63K (or 81%) have an available PSU score, while 14974 (or 19%) have not. As we explained in detail in the PSU Section, we collected data on the national college exam (PSU) that teachers took up to 35 years ago (from 1980 onwards). Therefore, we do not have information for the older teachers, many of whom have retired from teaching anyway. On average, the teachers with ED scores but no PSU scores were 61 years old in 2016, and a 44% of them was not teaching during year.

Table C.8 shows the teacher evaluation results by availability of PSU scores. Panel A shows the results by the four categories of performance. We also have information on the overall ED score and also the portfolio component score, whose results we present in Panel B of Table C.8.

Teachers with PSU scores tend to perform better in the ED. Panel A shows that they fall more in the upper two categories (competent and outstanding) and less in the lower (basic and unsatisfactory). Consistently, teachers with PSU scores also achieve higher ED scores, both overall and in the portfolio component as shown in Panel B. Differences in both Panels are significant at the 1% level.

Table C.8: Teacher Evaluation Results 2004-2013, by PSU Score availability

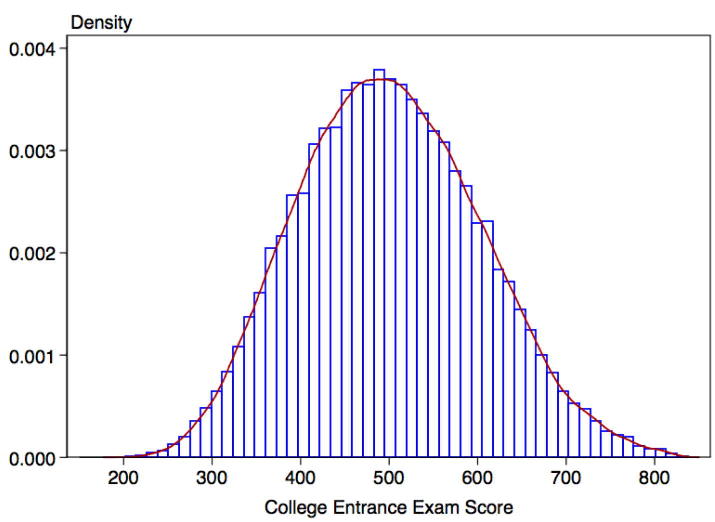
	With PSU Scores		Without PSU Scores		T-Test
Classification	N obs	%	N obs	%	Difference
Outstanding	5931	9.3	944	6.3	3 ***
Competent	39605	62.3	8525	56.9	5.4 ***
Basic	16987	26.7	5104	34.1	-7.4 ***
Unsatisfactory	1016	1.6	401	2.7	-1.1 ***
Total	63539	100	14974	100	
Score	N obs	%	N obs	%	Difference
Overall	2.63	0.28	2.57	0.29	0.06 ***
Portfolio	2.29	0.32	2.2	0.32	0.09 ***

C.2 Additional Figures

In this section we examine whether there is manipulation of the college entrance exams near the cutoffs in our data. In the centralized admission system in Chile scores are administered by a specialized agency (DEMRE), and test-takers do not know how to convert their performance into a score when they are taking the exam. Their raw score is a function of good and bad answers, and their final score is computed after standardizing raw scores taking into account all test-takers in the country.

In Figure C.1 we plot the distribution of the college entrance exam score for all test takers. By construction, it has a smooth bell-shaped distribution, showing no bunching at particular points of the support of the scores.

Figure C.1: College Entrance Exam Density



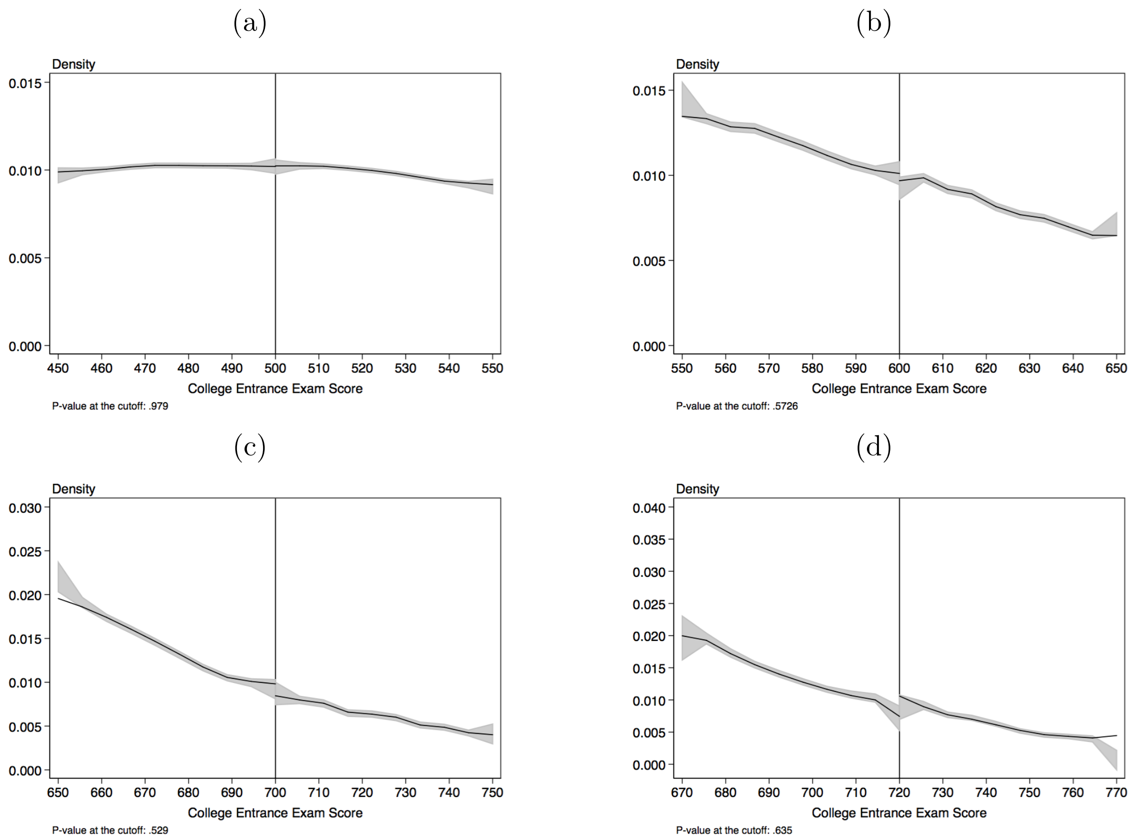
Note: This figure plots a density estimate of the college entrance exams in our data.

However, for the sake of the argument, let's suppose, for example, that some institutions could try to game the system by administratively manipulating their applicants' test scores when reporting to the Ministry of Education that those applicants are eligible for the scholarship. If that was the case, we would observe bunching of scores just above say, the 600 points threshold.

We test for manipulation using a nonparametric test (Cattaneo et al., 2018) of discontinuity in the density of students with scores in the vicinity of the BVP cutoffs of 500, 600, 700 and 720 points. Figure C.2 provides a graphical representation of the continuity in the density test approach, plotting the density of observations by scores in our data.

At the bottom of each graph, we provide the p-value associated with the manipulation test. In all cases, we find a high p-value, which indicates that there is no statistical evidence of systematic manipulation of the running variable. This plot is consistent with the results from the formal test from Cattaneo et al. (2018), as the density estimates above and below the cutoff (the two intercepts in the figure) are very near each other.

Figure C.2: Density Tests



Note: The graphs in Figure C.2 plot the estimated density functions using all the underlying data, allowing for different density estimates on each side of the cutoff.

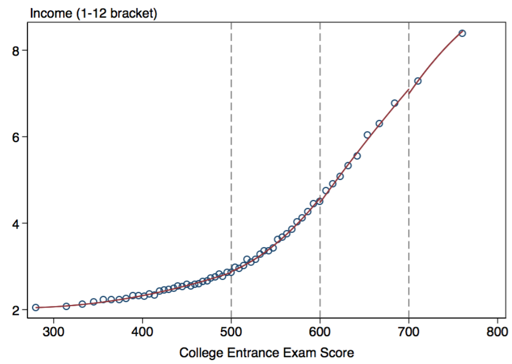
Covariates Smoothness

Our research design mimics a local experiment where test takers are exogenously allocated to receive a scholarship to study at teacher colleges. In this section we show that there are no other changes in our observable covariates occurring at the score threshold that could confound our analysis.

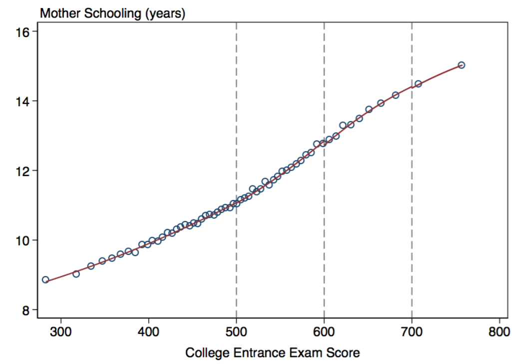
We show a graphical illustration for every covariate in Figure C.3, which provides further evidence of a smooth behavior at the test score cutoff.

Figure C.3: Covariates Smoothness

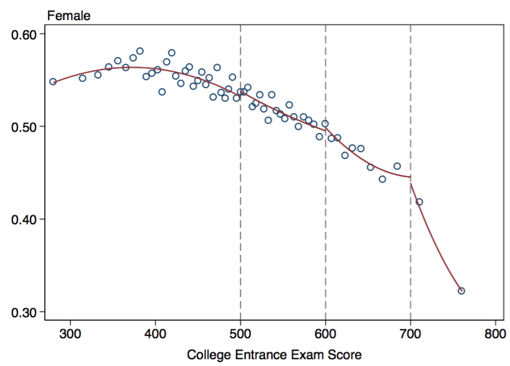
(a) Household Income



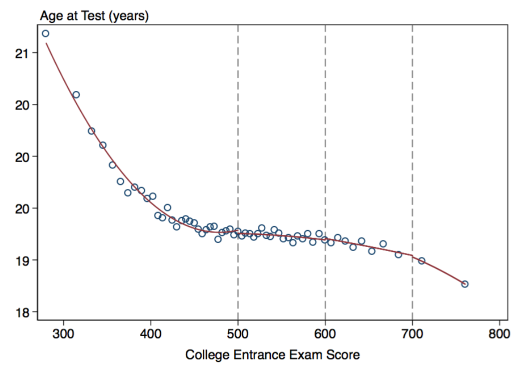
(b) Mother's Schooling



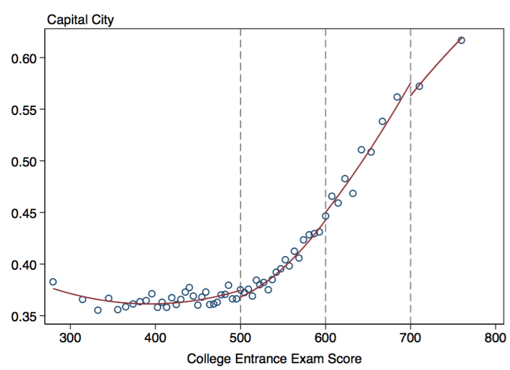
(c) Girl



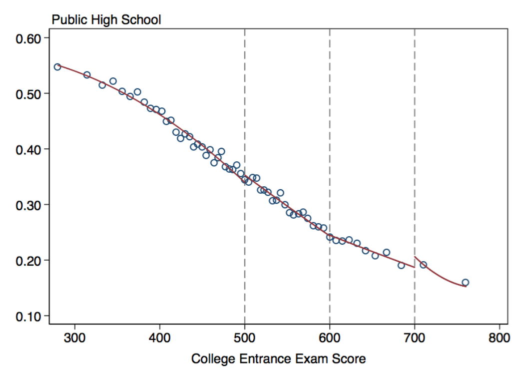
(d) Age



(e) Capital City



(f) Public School



Note: The graphs in Figure C.3 plot the mean of the y-axis variable within bins of scores, and fit estimated lines using all the underlying data, allowing for different slopes on each side of the cutoff.

References

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

Abdulkadiroğlu, A., Angrist, J. D., Narita, Y., and Pathak, P. A. (2017). Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation. Econometrica, 85(5):1373–1432.

Abdulkadiroğlu, A. and Sönmez, T. (2003). School Choice: A Mechanism Design Approach. American Economic Review, 93(3):729–747.

Agarwal, N., Ashlagi, I., Azevedo, E., Featherstone, C. R., and Karaduman, Ö. (2019). Market failure in kidney exchange. American Economic Review, 109(11):4026–70.

Agarwal, N., Hodgson, C., and Somaini, P. (2020). Choices and outcomes in assignment mechanisms: The allocation of deceased donor kidneys.

Agarwal, N. and Somaini, P. (2018). Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism. Econometrica, 86(2):391–444.

Agrawal, A., Gans, J., and Goldfarb, A. (2018). Prediction machines: the simple economics of artificial intelligence. Harvard Business Press.

Ainsworth, R., Dehejia, R. H., Pop-Eleches, C., and Urquiola, M. S. (2020). Information, Preferences, and Household Demand for School Value Added. NBER Working Paper, (13980).

Ajzenman, N., Elacqua, G., Jaimovich, A., and Pérez-Núñez, G. (2023). Humans Versus Chatbots: Scaling-Up Behavioral Interventions to Reduce Teacher Shortages. SSRN Electronic Journal, (16404).

Ajzenman, N., Elacqua, G., Marotta, L., and Olsen, A. S. (2021). Order Effects and Employment Decisions: Experimental Evidence from a Nationwide Program. Technical report, Inter-American Development Bank, Washington, D.C.

Allende, C., Gallego, F., and Neilson, C. (2019). Approximating the Equilibrium Effects of Informed School Choice. Working Paper, (1100623).

Altmejd, A., Barrios-Fernández, A., Drlje, M., Goodman, J., Hurwitz, M., Kovac, D., Mulhern, C., Neilson, C., and Smith, J. (2021). O Brother, Where Start Thou? Sibling Spillovers on College and Major Choice in Four Countries. The Quarterly Journal of Economics, 136(3):1831–1886.

Alvarado, M., Duarte, F., and Neilson, C. (2011). Getting Better Students to Become Teachers: the Effects of Targeting Tuition Subsidies. Working Paper, Centro de Estudios MINEDUC, Chile.

Andersson, T., Dur, U., Ertemel, S., Kesten, O., et al. (2018). Sequential school choice with public and private schools. Unpublished paper, Working Papers, 39.

Angrist, J., Gray-Lobe, G., Idoux, C., and Pathak, P. A. (2022). Still Worth the Trip? School Busing Effects in Boston and New York. SSRN Electronic Journal, (July).

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., and Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. The Quarterly Journal of Economics, 131(3):1415–1453.

Arteaga, F., Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2022). Smart Matching Platforms and Heterogeneous Beliefs in Centralized School Choice. Quarterly Journal of Economics, 137(3):1791–1848.

Artemov, G., Che, Y.-K., and He, Y. (2020). Strategic 'mistakes': Implications for market design research. Work. Pap., Melbourne Univ., Melbourne, Aust.

Athey, S. (2019). The impact of machine learning on economics. The Economics of Artificial Intelligence: An Agenda, Chapter 21:507–547.

Aue, R., Klein, T., and Ortega, J. (2020). What happens when separate and unequal school districts merge? ZEW-Centre for European Economic Research Discussion Paper, (20-032).

Azevedo, E. M. and Budish, E. (2019). Strategy-proofness in the large. The Review of Economic Studies, 86(1):81–116.

Azevedo, E. M. and Leshno, J. D. (2016). A Supply and Demand Framework for Two-Sided Matching Markets. Journal of Political Economy, 124(5):1235–1268.

Bacolod, M. (2006). Do alternative opportunities matter? the role of female labor markets in the decline of teacher quality. Working papers, U.S. Census Bureau, Center for Economic Studies.

Basse, G. W., Feller, A., and Toulis, P. (2019). Randomization tests of causal effects under interference. Biometrika, 106(2):487–494.

Bau, N. and Das, J. (2018). Teacher value-added in a low-income country. American Economic Journal: Economic Policy.

Biasi, B. (2021). The labor market for teachers under different pay schemes. American Economic Journal: Economic Policy, 13(3):63–102.

Biasi, B. and Sarsons, H. (2021). Flexible Wages, Bargaining, and the Gender Gap*. The Quarterly Journal of Economics, 137(1):215–266.

Bold, T., Filmer, D., Martin, G., Molina, E., Stacy, B., Rockmore, C., Svensson, J., and Wane, W. (2017). Enrollment without learning: Teacher effort, knowledge, and skill in primary schools in africa. Journal of Economic Perspectives, 31(4):185–204.

Bold, T., Filmer, D., Molina, E., and Svensson, J. (2019). The Lost Human Capital: Teacher Knowledge and Student Achievement in Africa. The World Bank.

Bruns, B. and Luque, J. (2015). Great Teachers: How to Raise Student Learning in Latin America and the Caribbean. World Bank, Washington, DC.

Bucarey, A. (2018). Who Pays for Free College? Crowding Out on Campus. Job Market Paper, pages 1–71.

Calsamiglia, C., Fu, C., and Güell, M. (2018). Structural Estimation of a Model of School Choices: The Boston Mechanism vs. Its Alternatives.

Castro-Zarzur, R. and Mendez, C. (2019). Effectiveness of the bvp in time. Working paper, Mimeo.

Castro-Zarzur, R., Sarzosa, M., and Espinoza, R. (2019). Unintended consequences of free college: Self-selection into the teaching profession. Working paper, Mimeo.

Cattaneo, M. D., Jansson, M., and Ma, X. (2018). Manipulation testing based on density discontinuity. The Stata Journal, 18(1):234–261.

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., and Mullainathan, S. (2016). Productivity and selection of human capital with machine learning. American Economic Review P&P, 106(5):124–27.

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., and Rogers, F. H. (2006). Missing in Action: Teacher and Health Worker Absence in Developing Countries. Journal of Economic Perspectives, 20(1):91–116.

Che, Y.-K., Hahm, D. W., and He, Y. (2020). Leveraging uncertainties to infer preferences: Robust analysis of school choice. Technical report.

Che, Y.-K. and Kojima, F. (2010). Asymptotic Equivalence of Probabilistic Serial and Random Priority Mechanisms. Econometrica, 78(5):1625–1672.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. American Economic Review, 104(9):2633–79.

Clotfelter, C. T., Ladd, H. F., and Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. Economics of Education Review, 26(6):673 – 682. Economics of Education: Major Contributions and Future Directions - The Dijon Papers.

Corcoran, S., Evans, W., and Schwab, R. M. (2004). Changing labor-market opportunities for women and the quality of teachers, 1957-2000. American Economic Review, 94(2):230–235.

Correa, J., Epstein, R., Escobar, J., Rios, I., Bahamondes, B., Bonet, C., Epstein, N., Aramayo, N., Castillo, M., Cristi, A., and Epstein, B. (2019). School choice in Chile. ACM EC 2019 - Proceedings of the 2019 ACM Conference on Economics and Computation, pages 325–343.

Cox, D. (1958). Planning of Experiments. John Wiley & Sons, Inc., New York.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R., and Zamora, P. (2013). Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment. The Quarterly Journal of Economics, 128(2):531–580.

De Gregorio, S. and Neilson, C. (2020). Teacher preparation and student learning: The ensena chile case. Working paper, Mimeo.

Demange, G., Gale, D., and Sotomayor, M. (1987). A further note on the stable matching problem. Discrete Applied Mathematics, 16(3):217–222.

Dubins, L. E. and Freedman, D. A. (1981). Machiavelli and the gale-shapley algorithm. The American Mathematical Monthly, 88(7):485–494.

Duflo, E., Glennerster, R., and Kremer, M. (2007). Chapter 61 Using Randomization in Development Economics Research: A Toolkit. In Handbook of Development Economics, volume 4, pages 3895–3962.

Efron, B. and Tibshirani, R. (1994). An Introduction to the Bootstrap. Number 3. Chapman and Hall/CRC.

Ekmekci, M. and Yenmez, M. B. (2019). Common enrollment in school choice. Theoretical Economics, 14(4):1237–1270.

Elacqua, G., Westh Olsen, A. S., and Velez-Ferro, S. (2021). The Market Design Approach to Teacher Assignment: Evidence from Ecuador. Technical report, Inter-American Development Bank, Washington, D. C.

Elizondo, D. A., Birkenhead, R., Gamez, M., Garcia, N., and Alfaro, E. (2012). Linear separability and classification complexity. Expert Systems with Applications, 39(9):7796 – 7807.

Estrada, R. (2019). Rules versus Discretion in Public Service: Teacher Hiring in Mexico. Journal of Labor Economics, 37(2):545–579.

Estrada, R. and Lombardi, M. (2020). Skills and selection into teaching: Evidence from latin america. Working paper, Mimeo.

Evans, D., Yuan, F., and Filmer, D. (2020). Are teachers in africa poorly paid? evidence from 15 countries. CGD Working Paper Series 538, Center for Global Development.

Fack, G., Grenet, J., and He, Y. (2019). Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions. American Economic Review, 109(4):1486–1529.

Ferreyra, M. M., Avitabile, C., Botero Álvarez, J., Haimovich Paz, F., and Urzúa, S. (2017). At a crossroads: higher education in Latin America and the Caribbean. The World Bank.

Figlio, D., Karbownik, K., and Salvanes, K. (2017). The promise of administrative data in education research. Education Finance and Policy, 12(2):129–136.

Gale, D. and Shapley, L. S. (1962). College Admissions and the Stability of Marriage. The American Mathematical Monthly, 69(1):9–15.

Gallegos, S., Neilson, C., Calle, F., and Karnani, M. (2022). Screening and recruiting talent at teacher colleges using pre-college academic achievement. Hceo working paper, The University of Chicago.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). Bayesian data analysis. CRC press.

Geweke, J., Gowrisankaran, G., and Town, R. J. (2003). Bayesian inference for hospital quality in a selection model. Econometrica, 71(4):1215–1238.

Gilligan, D. O., Karachiwalla, N., Kasirye, I., Lucas, A. M., and Neal, D. (2018). Educator incentives and educational triage in rural primary schools. Working Paper 24911, National Bureau of Economic Research.

Haeringer, G. and Klijn, F. (2009). Constrained school choice. Journal of Economic theory, 144(5):1921–1947.

Hanushek, E. (2011). The Economic Value of Higher Teacher Quality. Economics of Education Review, 30(3):466–479.

Hanushek, E. A., Piopiunik, M., and Wiederhold, S. (2019). The value of smarter teachers: International evidence on teacher cognitive skills and student performance. Journal of Human Resources, 54(4):857–899.

Harris, D. and Sass, T. (2011). Teacher training, teacher quality, and student achievement. Journal of Public Economics, 95:798–812.

Hassidim, A., Romm, A., and Shorrer, R. I. (2016). "strategic" behavior in a strategy-proof environment. In Proceedings of the 2016 ACM Conference on Economics and Computation, pages 763–764.

Hastings, J., Neilson, C., and Zimmerman, S. (2014). Are some degrees worth more than others? evidence from college admission cutoffs in chile. Working Paper 19241, NBER.

Horowitz, J. L. (2001). The Bootstrap. In Handbook of Econometrics, volume 5, pages 3159–3228.

Hoxby, C. and Leigh, A. (2004). Pulled away or pushed out? explaining the decline of teacher aptitude in the united states. American Economic Review, 94(2):236–240.

Hoxby, C. M. (1996). How Teachers' Unions Affect Education Production*. The Quarterly Journal of Economics, 111(3):671–718.

Hull, P. (2018). Estimating hospital quality with quasi-experimental data. Available at SSRN 3118358.

Jackson, C. K. (2012). Recruiting, retaining, and creating quality teachers. Nordic Economic Policy Review, 3(1):61–104.

Jackson, C. K., Rockoff, J. E., and Staiger, D. O. (2014). Teacher Effects and Teacher-Related Policies. Annual Review of Economics, 6(1):801–825.

Kane, T. and Cantrell, S. (2010). Learning about teaching research report: Initial findings from the measures of effective teaching project. Technical report.

Kapor, A., Karnani, M., and Neilson, C. (Forthcoming, 2024). Aftermarket Frictions and the Cost of Off-Platform Options in Centralized Assignment Mechanisms. Journal of Political Economy.

Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2020). Heterogeneous Beliefs and School Choice Mechanisms. American Economic Review, 110(5):1274–1315.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. (2017). Human Decisions and Machine Predictions*. The Quarterly Journal of Economics, 133(1):237–293.

Kojima, F. and Pathak, P. A. (2009). Incentives and stability in large two-sided matching markets. American Economic Review, 99(3):608–27.

Lafortune, J., Figueroa, N., and Saenz, A. (2018). Do you like me enough? The impact of restricting preferences ranking in a university matching process. Working Paper.

Lankford, H., Loeb, S., McEachin, A., Miller, L. C., and Wyckoff, J. (2014). Who enters teaching? encouraging evidence that the status of teaching is improving. Educational Researcher, 43(9):444–453.

Larroucau, T. and Rios, I. (2018). Do "short-list" students report truthfully? strategic behavior in the chilean college admissions problem. Technical report.

Larroucau, T. and Rios, I. (2020). Dynamic college admissions and the determinants of students' college retention. Technical report.

Leung, M. P. (2020). Treatment and Spillover Effects Under Network Interference. The Review of Economics and Statistics, 102(2):368–380.

Lewis, H. G. (1983). Union Relative Wage Effects: A Survey of Macro Estimates. Journal of Labor Economics, 1(1):1–27.

Lombardi, M. (2019). Is the remedy worse than the disease? the impact of teacher remediation on teacher and student performance in chile. Economics of Education Review, 73:101928.

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

MINEDUC (2019). Sistema de informacion general de estudiantes: Idoneidad docente. Data file, retrieved from http://datos.mineduc.cl/dashboards/19732/bases-de-datos-de-cargos-docentes/, Santiago, Chile: Ministerio de Educacion de Chile.

Mizala, A. and Nopo, H. (2016). Measuring the relative pay of school teachers in Latin America 1997–2007. International Journal of Educational Development, 47(C):20–32.

Mullainathan, S. and Spiess, J. (2017). Machine learning: An applied econometric approach. Journal of Economic Perspectives, 31(2):87–106.

Neal, D. (2011). Chapter 6 - the design of performance pay in education. In Hanushek, E. A., Machin, S., and Woessmann, L., editors, Handbook of The Economics of Education, volume 4 of Handbook of the Economics of Education, pages 495 – 550. Elsevier.

Neilson, C. A. (2021). The Rise of Centralized Mechanisms in Education Markets Around the World.

Neyman, J. (1990). On the Application of Probability Theory to Agricultural Experiments. Statistical Science, 5(4):465–472.

Niederle, M. and Roth, A. E. (2009). Market culture: How rules governing exploding offers affect market performance. American Economic Journal: Microeconomics, 1(2):199–219.

OECD (2005). Teachers Matter: Attracting, Developing and Retaining Effective Teachers. Organisation for Economic Co-operation and Development, Paris.

OECD (2009). Education at a Glance 2009: OECD Indicators. Organisation for Economic Co-operation and Development, Paris.

OECD (2013). OECD Reviews of Evaluation and Assessment in Education: Teacher Evaluation in Chile. Organisation for Economic Co-operation and Development, Paris.

Paredes, V., Aron, A., and Carril, A. (2015). Where is the teacher? short-run effect of teacher absenteeism on student achievement. Technical report, Department of Economics, Universidad de Chile. Mimeo.

Pathak, P. A. (2017). What really matters in designing school choice mechanisms. Advances in Economics and Econometrics, 1:176–214.

Pathak, P. A. and Sönmez, T. (2008). Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism. American Economic Review, 98(4):1636–1652.

Podgursky, M., Monroe, R., and Watson, D. (2004). The academic quality of public school teachers: an analysis of entry and exit behavior. Economics of Education Review, 23(5):507–518.

Rios, I., Larroucau, T., Parra, G., and Cominetti, R. (2021). Improving the chilean college admissions system. Operations Research.

Rivkin, S. G., Hanushek, E. A., and Kain, J. F. (2005). Teachers, Schools, and Academic Achievement. Econometrica, 73(2):417–458.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. American Economic Review, 94(2):247–252.

Rossi, P. E., McCulloch, R. E., and Allenby, G. M. (1996). The value of purchase history data in target marketing. Marketing Science, 15(4):321–340.

Roth, A. E. and Sotomayor, M. (1992). Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis. Econometric Society Monographs.

Roth, A. E. and Xing, X. (1994). Jumping the gun: Imperfections and institutions related to the timing of market transactions. The American Economic Review, pages 992–1044.

Rothstein, J. M. (2006). Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions. American Economic Review, 96(4):1333–1350.

Rubin, D. B. (1974). Estimating causal effects of treatment in randomized and nonrandomized studies. Journal of Educational Psychology, 66(5):688–701.

Rubin, D. B. (1980). Randomization Analysis of Experimental Data: The Fisher Randomization Test Comment. Journal of the American Statistical Association, 75(371):591.

Sajjadiani, S., Sojourner, A. J., Kammeyer-Mueller, J. D., and Mykerezi, E. (2019). Using machine learning to translate applicant work history into predictors of performance and turnover. Journal of Applied Psychology, 104(10):1207–1225.

See, B. H., Morris, R., Gorard, S., Kokotsaki, D., and Abdi, S. (2020). Teacher recruitment and retention: A critical review of international evidence of most promising interventions. Education Sciences, 10(10).

Shorrer, R. I. and Sóvágó, S. (2018). Obvious mistakes in a strategically simple college admissions environment: Causes and consequences. Available at SSRN 2993538.

Solis, A. (2017). Credit access and college enrollment. Journal of Political Economy, 125(2):562–622.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(56):1929–1958.

Tincani, M. (2014). School Vouchers and the Joint Sorting of Students and Teachers. Human Capital and Economic Opportunity Global Working Group (HCEO) Working Paper No. 2014-012.

Tincani, M. (2021). Teacher labor markets, school vouchers, and student cognitive achievement: Evidence from Chile. Quantitative Economics, 12(1):173–216.

Tincani, M., Todd, P., Behrman, J., and Wolpin, K. (2016). Teacher Quality in Public and Private Schools Under a Voucher System: The Case of Chile. Journal of Labor Economics, 34(2).

van Dijk, W. (2019). The socio-economic consequences of housing assistance. Job Market Paper.

Vegas, E., Murnane, R., and Willet, J. (2001). From high school to teaching: Many steps, who makes it? Teachers College Record, 103(3):427–449.

Walters, C. R. (2018). The demand for effective charter schools. Journal of Political Economy, 126(6):2179–2223.

World Bank (2011). Population aging: Is Latin America Ready? David Cotlear (editor), Washington, DC.

World Bank (2013). System Approach for Better Education Results (SABER): What Matters Most in Teacher Policies? A Framework for Building a More Effective Teaching Profession. Washington, DC. Technical report.