

# Causal Effects in Matching Mechanisms with Strategically Reported Preferences\*

Marinho Bertanha<sup>†</sup>

Margaux Lufade<sup>‡</sup>

Ismael Mourifié<sup>§</sup>

July 26, 2023

## Abstract

A growing number of central authorities use assignment mechanisms to allocate students to schools in a way that reflects student preferences and school priorities. However, most real-world mechanisms give students an incentive to be strategic and misreport their preferences. In this paper, we provide an identification approach for causal effects of school assignment on future outcomes that accounts for strategic misreporting. Misreporting may invalidate existing point-identification approaches, and we derive sharp bounds for causal effects that are robust to strategic behavior. Our approach applies to any mechanism as long as there exist placement scores and cutoffs that characterize that mechanism's allocation rule. We use data from a deferred acceptance mechanism that assigns students to more than 1,000 university-major combinations in Chile. Students behave strategically because the mechanism in Chile constrains the number of majors that students submit in their preferences to eight options. Our methodology takes that into account and partially identifies the effect of changes in school assignment on various graduation outcomes.

**Keywords:** Matching, Truth Telling, Random Sets, Partial Identification, Constrained School Choice, Regression Discontinuity.

**JEL Classification:** C12, C21, C26.

---

\*This research was conducted in part when Margaux Lufade and Ismael Mourifié were visiting the Becker Friedman Institute (BFI), and Marinho Bertanha was visiting the Kenneth C. Griffin Department of Economics, both at the University of Chicago. The authors thank their respective hosts for their hospitality and support. We are grateful to Nikhil Agarwal, Guillaume Haeringer, YingHua He, Sam Hwang, Yuichi Kitamura, Kory Kroft, Chris Neilson, Marcin Peski, Kevin Song, and Maciej Kotowski for valuable comments and discussions. The paper has also benefited from feedback received during presentations at the 2021 SITE Conference at Stanford, 37th CESG Conference in Vancouver, Chamberlain Online Seminar, 5th Econometrics Workshop at Notre Dame, 2022 CIREQ Conference in Montreal, Northwestern Univ, Univ of Toronto, 2022 IAAE Conference in London, Fund. Getulio Vargas in Rio, MEG Conference at MSU, 2022 SEA Conference, Univ of Michigan, UC Santa Barbara, Boston College, Harvard/MIT, Univ of Texas A&M, Stanford Univ, 2023 Causal Inference Workshop at UIUC, UC Santa Cruz, and the 2023 Econometric Society in Kenya.

<sup>†</sup>Department of Economics, University of Notre Dame. Address: 3060 Jenkins Nanovic Halls, Notre Dame, IN 46556. Email: mbertanha@nd.edu.

<sup>‡</sup>Department of Economics, University of Pennsylvania. Address: 133 South 36th Street, Philadelphia, PA 19104. Email: mlufade@sas.upenn.edu.

<sup>§</sup>Department of Economics, University of Toronto & Washington University in St. Louis & NBER. Address: 150 St. George Street, Toronto ON M5S 3G7, Canada. Email: ismael.mourife@utoronto.ca.

# 1 Introduction

One of the most important decisions made by students is the choice of their fields and institutions of education. Identification of the impact of such choices on future outcomes is a critical step in the study of the decision process as well as public education policies. Causal effects of schooling on outcomes may vary widely across economic agents because of heterogeneous skills and preferences. Moreover, individuals' expectations about their potential returns trigger strategic choices of schools. Heterogeneity and selection make identification of causal effects very challenging, especially when students face a large number of unordered options. On the positive side, a growing number of schools use centralized assignment mechanisms that create credible instruments from discontinuities that randomly assign comparable students into different schools (Kirkeboen et al. (2016) and Abdulkadiroglu et al. (2022)). Discontinuities arise from unpredictable admission cutoffs that characterize the matching of students to schools, which is generated by the assignment mechanism.

Many centralized school assignment mechanisms produce a quasi-experimental design where two groups of individuals share similar scores but are assigned into different schools, based on how their scores relate to admission cutoffs. The assignment in a matching with cutoff characterization depends on the student's preferences over feasible schools. Unlike a typical regression discontinuity (RD) design, this setting does not ensure that students on the same side of the cutoff have the same assignment. For example, individuals with similar scores just above a certain cutoff may all prefer to go to the same school  $a$ ; however, they could have very different second-best options if they were on the other side of that cutoff. In such a context, Kirkeboen et al. (2016) construct comparable groups of individuals near a cutoff by controlling for local preferences; that is, by selecting students whose preferences yield identical first- and second-best options, respectively, if above and if below that cutoff. Controlling for local preferences that equal a pair of schools, e.g.,  $(a, b)$ , allows the RD to identify causal effects of changing school assignment from  $b$  to  $a$ , averaged over individuals that prefer  $a$  over  $b$ . Although powerful, this relies on students reporting their true preferences to the centralized mechanism. A natural question to ask is how often students submit their true preferences. In fact, most real-world school assignment mechanisms create incentives for misreporting preferences. Agarwal and Somaini (2018) and Fack et al. (2019) provide thorough discussions with several real-world examples.

This paper derives sharp bounds on causal effects of school assignment on future outcomes in mechanisms with a cutoff representation and students behaving strategically. We propose a two-step identification approach that is robust to strategic reporting of preferences. In the first step, the researcher partially identifies local preferences and constructs local-preference sets for each student. We provide several tools for constructing these sets in the context of the student-proposing deferred-acceptance (DA) mechanism, where constraints on submissions lead to strategic behavior. Outside of that context, researchers may employ alternative tools to partially identify local preferences as our second-step procedures do not require a particular method to be used in the first-step. For example, the identification method of Agarwal and Somaini (2018) applies to a

general class of mechanisms with cutoff representation that includes variants of the DA, Boston mechanism, First Preferences First, Chinese Parallel Mechanism, etc. Finally, in the second step, the researcher employs an RD identification strategy that controls for local-preference sets and partially identifies causal effects of school assignment.

The strategic behavior of students depends on the characteristics of the assignment mechanism. A mechanism is said to be strategy proof for students if submitting true preferences is a weakly dominant strategy for all students. For example, Dubins and Freedman (1981) demonstrate that the DA mechanism is strategy proof. However, this result breaks down when the mechanism imposes constraints on students' submissions. In many real-world school assignment mechanisms, the number of schools is too large for students to feasibly rank all schools. The central authorities running these systems may either limit the number of schools that students may rank or impose costs on the number of schools submitted. See Table 1, Panel B by Fack et al. (2019) for examples.

Our first-step tools for partial identification of local preferences naturally require assumptions on the strategic behavior of students. We motivate our assumptions after important contributions of Haeringer and Klijn (2009). They study the game where students submit constrained preference rankings and a central mechanism allocates students to schools. One of their important findings is that it is rational for students to submit partial orders of their true preferences in some mechanisms. Specifically, suppose a mechanism is strategy proof when students are free to rank any number of schools, e.g., the unconstrained DA or Top Trading Cycles (TTC) mechanisms. Then, if preference rankings are constrained to have at most  $K$  schools, a student can do no better than selecting  $K$  schools among acceptable schools and ranking them according to the student's true preferences.

The key assumption for our first-step methods is that students submit partial orders of their preferences. That comes in addition to the cutoff characterization of the matching, which we assume throughout the paper. Cutoff characterization means that students are matched to their best feasible schools, where "best" is defined according to true preferences. A school is feasible for a student if that student's placement score clears the admission cutoff for that school. Our requirement of cutoff characterization is satisfied when the matching outcome is stable (Azevedo and Leshno, 2016). Stability means that each student is matched to an acceptable school and better schools are full with people that have better placement scores. In constrained DA and TTC mechanisms, stability occurs in Nash equilibrium of the preference revelation game under appropriate conditions on the placement scores (Theorems 6.3 and 6.4, Haeringer and Klijn (2009)). Stability occurs in Nash equilibrium without restrictions on scores in the constrained serial dictatorship (SD) mechanism, which is a particular case of DA. We characterize sharp local-preference sets for every individual that are compatible with the observed data and these model assumptions. Our random local-preference sets contain the true local-preference random variable with probability one. We also show a way to shrink these sets by imposing assumptions on students' expectations regarding the outcome of the match and by assuming students maximize expected utility in line with Agarwal and Somaini (2018).

The form of our local-preference sets reveals two important implications for applied work. First,

if we ignore strategic behavior and only control for reported local preferences, the RD strategy may be inconsistent. The reason is that our local-preference sets change discontinuously at a cutoff once we select individuals with the same reported local preferences. The problem is akin to the manipulation problem in RD: we control for a variable that is manipulable, that is, reported preferences. Second, suppose we select individuals with reported local preferences that equal a given pair of schools  $(a, b)$  and see that local-preference sets are all singletons  $\{(a, b)\}$ ; in other words, our selection only contains individuals that do not manipulate. Even in this case, the RD strategy that controls for reported local preferences may still be inconsistent. The reason is that local-preference sets may well contain the pair  $(a, b)$  for someone whose reported local preferences differ from  $(a, b)$ . Thus, selecting individuals that report  $(a, b)$  that does select all individuals whose true local preferences equal  $(a, b)$ .

Our second-step approach relies on local-preference sets constructed in the first step, either with our methods or alternative methods. Given interest in a pair of schools  $(a, b)$ , we select all individuals whose local-preference sets contain  $(a, b)$  and placement scores are close to the cutoff for admission at school  $a$ . This sub-population of individuals contains all individuals whose true local preferences equal  $(a, b)$ , but also other individuals. The average outcome in the sub-population equals a weighted average of two averages: first, the average outcome for individuals with true local preferences  $(a, b)$ , which is interesting for the identification of causal effects; and second, the average outcome for individuals with true local preferences that differ from  $(a, b)$ . We do not know which individuals have preferences  $(a, b)$ , but we do characterize sharp bounds on the proportion of such individuals in the sub-population using the random sets constructed in the first step. Thus, our setting fits the identification problem with corrupted data studied by Horowitz and Manski (1995). This allows us to derive closed-form bounds on the first out of the two average outcomes above, which then leads to bounds on average causal effects. The closed form provides intuition for cases when the bounds become informative or are equal (i.e., point identification). Although practical and intuitive, these closed-form bounds may not be sharp. We also characterize sharp bounds using random set theory that are numerically computable when outcomes take finitely many values.

The combination of RD identification with school matching data has been popular among applied and theoretical researchers in economics (Jackson, 2010; Bertanha, 2020). To the best of our knowledge, our paper is the first to prove RD identification of returns of school assignment in matching mechanisms with strategically reported preferences. We note that our two-step approach differs from the usual control function approach because our first step partially identifies the control variable instead of point identifying it as in the usual approach. This paper unifies and complements two branches of the literature. One branch has methods proposed by Agarwal and Somaini (2018) and Fack et al. (2019) that take into account strategic reporting and identify students' true preferences; their focus is not on causal effects of matched school on future outcomes. The other branch has methods by Kirkeboen et al. (2016) and Abdulkadiroglu et al. (2022) that identify causal effects of different assignments but control for reported instead of true preferences.

We apply our two-step identification strategy to matching data from Chile. Chile has a cen-

tralized DA mechanism that assigns students to university-major pairs. In 2010, 88,000 students ranked at most eight university-major pairs out of a total of 1,092 options available. Thus, the mechanism constrains the preference rankings, and students have incentives to behave strategically. We focus on popular university-major pairs and compute bounds on the average effect of matching assignment on various outcome variables such as graduating from program of assignment, ever graduating, reapplying, or graduating from a STEM program.

The methods proposed by Kirkeboen et al. (2016) and Abdulkadiroglu et al. (2022) are not directly applicable to the Chilean case even if students reported their preferences truthfully. In fact, Kirkeboen et al. (2016)’s methods apply to the SD mechanism, which is the particular case of the DA mechanism when all schools utilize the same placement score. In the SD case, the counterfactual set of schools for students just above or just below a cutoff does not vary across students. For example, for students just above the cutoff, their counterfactual set has all schools whose cutoffs are smaller than the cutoff in question. The same does not apply to DA. Students have multiple placement scores and the set of feasible schools may vary widely across students. The definition of a counterfactual set is an important step in the RD identification strategy, because local preferences are defined over these sets, which then become the control variable in the RD strategy. Abdulkadiroglu et al. (2022) propose a clever solution to the problem by constructing a propensity score variable in the DA case. They study the New York City public high schools, where placement scores are functions of integer priority scores plus a continuously distributed variable with full support. That particular structure of school priorities does not cover the Chilean case. In Chile, there are program-specific placement scores that are computed as functions of five primitive scores: math, language, history, science, and an average score from high school. These functions are different across programs and sometimes non-linear. This requires a careful definition of counterfactual set of schools such that controlling for local preferences does not violate the continuity assumptions required by RD. We therefore propose a general method that applies to such empirical contexts and leads to point identification if students are truthful but partial identification otherwise.

The rest of this paper proceeds as follows. Section 2 lays down the matching model for a continuum population of students and a finite number of schools. Section 3 studies point identification of average treatment effects when students are truth tellers. Section 4 studies partial identification when students strategically report their preferences. That section has two sub-sections. Section 4.1 provides tools for constructing local-preference sets that apply to constrained DA mechanisms. Section 4.2 uses local-preference sets to derive bounds on average treatment effects. We illustrate our identification approach with the Chilean data in Section 5. The appendix collects all proofs in the paper.

## 2 Model

We consider a continuum population of students and a set of  $J$  schools,  $\mathcal{J} := \{1, \dots, J\}$ , that

have capacities  $\{q_1, \dots, q_J\}$  defined in terms of shares of the student population. Denote by  $\Omega$  the set of all students in the universe of interest and use  $\omega$  to index an individual student type. The student type consists of three objects. First,  $Q(\omega)$  denotes the true (strict) preference relation of student  $\omega$  over the set of options  $\mathcal{J}^0 := \mathcal{J} \cup \{0\}$ , which includes schools  $\mathcal{J}$  and an outside option 0. For example, if  $J = 2$  and  $Q(\omega) = \{1, 2, 0\}$ , then 1 is preferred to 2 (i.e.,  $1Q(\omega)2$ ), 1 is preferred to 0 (i.e.,  $1Q(\omega)0$ ), and 2 is preferred to 0 (i.e.,  $2Q(\omega)0$ ). Let  $\mathcal{Q}$  be the set of all strict preference relations over  $\mathcal{J}^0$  that admit at least one school that is acceptable. A school  $j \in \mathcal{J}$  is “acceptable” for student  $\omega$  if it is preferred to that student’s outside option, i.e.,  $jQ(\omega)0$ . We define  $\bar{Q}$  as the weak preference relation induced by  $Q$ , i.e.,  $j\bar{Q}k \Leftrightarrow jQk$  or  $j = k$ . The second object of the student type is a vector of scores  $\mathbf{R}(\omega) := (R_1(\omega), \dots, R_J(\omega)) \in \mathcal{R} \subseteq \mathbb{R}^J$ , where each school  $j$  utilizes  $R_j$  to rank students for admission. Third and last object,  $Y(\omega, d)$  is the potential outcome of student  $\omega$  if the student is assigned to option  $d \in \mathcal{J}^0$ . Each student has a potential outcome function  $Y(\omega, \cdot)$  that maps from  $\mathcal{J}^0$  to  $\mathcal{Y} \subseteq \mathbb{R}$ . We call  $\mathbf{\Gamma}$  the set of all possible potential outcome functions. The set of all student types is  $\Omega := \mathcal{Q} \times \mathcal{R} \times \mathbf{\Gamma}$ . In a continuum economy, there is a probability measure  $\mathbb{P}$  over  $\Omega$  and the Borel  $\sigma$ -algebra of the product space  $\Omega$ . We suppress the argument  $\omega$  whenever it is unnecessary for ease of notation, e.g.,  $Y(d)$  vs.  $Y(\omega, d)$  and  $Q$  vs.  $Q(\omega)$ .

A “matching” is described by a measurable function  $\mu : \Omega \rightarrow \mathcal{J}^0$  that satisfies two conditions: for every  $j \in \mathcal{J}$ , (i) the mass of students matched to  $j$  is less than or equal to the capacity of school  $j$ , i.e.,  $\mathbb{P}\{\omega : \mu(\omega) = j\} \leq q_j$ ; and, (ii) the set of students that weakly prefer option  $j \in \mathcal{J}^0$  over their matching, i.e.,  $\{\omega : j\bar{Q}(\omega)\mu(\omega)\}$ , is an open set.<sup>1</sup> For every student type  $\omega$ ,  $\mu(\omega)$  is either the school  $j$  that the student is matched or zero. When  $\mu(\omega) = 0$ , the student is unmatched and takes an outside option. An important definition for this paper is stability.

**Definition 1** (Stability). *We say the matching  $\mu : \Omega \rightarrow \mathcal{J}^0$  is a stable matching if three conditions are satisfied for every  $\omega \in \Omega$ : (i)  $\mu(\omega)\bar{Q}(\omega)0$  (individual rationality); (ii) for any  $j \in \mathcal{J}$ , if  $jQ(\omega)\mu(\omega)$ , then  $j$  is full (no waste); and (iii) for any  $j \in \mathcal{J}$  that is full, if  $\mu(\omega') = j$  and  $jQ(\omega)\mu(\omega)$ , then  $R_j(\omega') > R_j(\omega)$  (no justified envy).*

A mechanism  $\varphi$  matches students to schools by mapping scores and submitted preference lists to schools. Student  $\omega$  submits a preference list  $P(\omega) \subseteq \mathcal{J}$  which is an ordered list of acceptable schools for student  $\omega$ . For example, for  $J = 3$ , if  $Q(\omega) = \{1, 2, 0, 3\}$ , then  $P(\omega) = \{1, 2\}$ , as long as the student submits the true list of acceptable schools. The number of schools in  $P$ , denoted  $|P|$ , is at least one because everyone participating in the match has at least one acceptable school. As with  $\bar{Q}$ , we also define  $\bar{P}$  as the weak preference relation induced by  $P$ . A mechanism takes as inputs the submitted preferences of everyone (i.e., a correspondence  $P : \Omega \rightrightarrows \mathcal{J}$ ) and the scores of everyone (i.e., a vector-valued function  $\mathbf{R} : \Omega \rightarrow \mathcal{R}$ ) and gives rise to a matching function. Formally,  $\varphi(P, \mathbf{R}) : \Omega \rightarrow \mathcal{J}^0$ . We say a mechanism  $\varphi$  is “strategy proof” if submitting the true ranking of acceptable schools is weakly dominant for every student. In other words, misreporting  $P$  for any student  $\omega$  never leads to a better option and sometimes leads to a worse option, depending on what

---

<sup>1</sup>Azevedo and Leshno (2016) impose the same condition to rule out multiplicity of stable matchings that differ in a set of types with measure zero.

other students submit. We say a student is a truth-teller if the student's  $P$  equals the student's true ranking of acceptable schools. Otherwise, we say the student is strategic or not a truth-teller.

The ability to characterize a matching allocation using cutoffs is fundamental for this paper.

**Definition 2** (Cutoff Characterization). *For placement scores  $\mathbf{S} : \Omega \rightarrow \mathcal{S} \subseteq \mathbb{R}^J$ ,  $\mathbf{S}(\omega) := (S_1(\omega), \dots, S_J(\omega))$ , and admission cutoffs  $\mathbf{c} \in \mathcal{S}$ ,  $\mathbf{c} := (c_1, \dots, c_J)$ , the set of feasible options of a student  $\omega$  equals all schools for which the student's placement scores clear the admission cutoffs plus the outside option:  $\{0\} \cup \{j \in \mathcal{J} : S_j(\omega) \geq c_j\}$ ; student  $\omega$ 's best feasible option is the option that ranks first according  $Q(\omega)$  among the student's best feasible options. We say the matching  $\mu : \Omega \rightarrow \mathcal{J}^0$  has cutoff characterization if there exists placement scores  $\mathbf{S} : \Omega \rightarrow \mathcal{S}$ , and admission cutoffs  $\mathbf{c} \in \mathcal{S}$  such that for every  $\omega \in \Omega$  the matching  $\mu(\omega)$  equals student  $\omega$ 's best feasible option according to  $Q(\omega)$ .*

This paper considers mechanisms that produce matching functions that have a cutoff characterization according to Definition 2. Placement scores  $\mathbf{S}$  may or may not equal school priority scores  $\mathbf{R}$ . The definition gives researchers freedom to construct special placement scores  $\mathbf{S}$  in case their mechanism does not admit cutoff characterization using priority scores  $\mathbf{R}$ . The idea comes from the general class of mechanisms of Agarwal and Somaini (2018). Moreover, we assume throughout that for any  $j \in \mathcal{J}$ ,  $S_j$  has distribution that is absolutely continuous with respect to (wrt) the Lebesgue measure and support  $\mathcal{S}_j$  that contains a closed interval around  $c_j$ . The support of  $\mathbf{S}$  is  $\mathcal{S}$ . For any two scores,  $S_j$  and  $S_l$ , we assume that either  $\mathbb{P}[S_j = S_l] = 1$  or  $\mathbb{P}[f(S_j) = S_l] < 1$  for any measurable function  $f$ . This says that the only deterministic function relating any two scores may be the identity function.

If the matching function is stable, Azevedo and Leshno (2016) demonstrate that the matching has cutoff characterization with  $\mathbf{S} = \mathbf{R}$  and admission cutoffs constructed as follows. For each  $j \in \mathcal{J}$ ,  $c_j := \inf\{S_j(\omega) : \text{for } \omega \text{ with } \mu(\omega) = j\}$  if some individuals are matched to  $j$  or  $c_j := \inf \mathcal{S}_j$  if nobody is matched to  $j$ . Many mechanisms produce stable matchings. For example, SD and DA are strategy proof and lead to stable matchings if agents are truth-tellers. If agents are not truth-tellers, e.g., because they face constraints in the submission of  $P$ , Haeringer and Klijn (2009) study settings where the SD, DA, Boston, and TTC mechanisms lead to stable matchings.

This paper is about identification of moments of treatment effects  $Y(d') - Y(d)$ . The econometrician has access to an infinite amount of data and observes the joint distribution of the following random objects:  $P(\omega)$ ,  $\mathbf{S}(\omega)$ ,  $\mu(\omega)$ , and  $Y(\omega) := Y(\omega, \mu(\omega))$ , where we abuse the notation and employ the letter  $Y$  for both observed outcome,  $Y(\omega)$ , and potential outcome of being assigned to school  $d$ ,  $Y(\omega, d)$ .

### 3 Identification with Truthfull Reports

In this section, we consider identification of causal effects when all students are truth tellers, that is, they submit their true ranking of acceptable schools. That is the rational behavior in mechanisms that are strategy proof. For instance, Dubins and Freedman (1981) and Roth (1982)

show that the DA mechanism without constraints on preference submission is strategy proof in the finite economy; Abdulkadiroglu et al. (2015) show the same for the continuum economy.<sup>2</sup> The SD is also strategy proof as it is a special case of DA.

**Assumption 1** (Truth-telling). *Students submit their true list of acceptable schools.*

Although submitting the true list of acceptable schools is a weakly dominant strategy under strategy-proof mechanisms, Assumption 1 explicitly says that students do so in light of recent evidence against this behavior (Chen and Sönmez (2006), Pais and Pintér (2008)).

The identification strategy of this paper resembles a regression discontinuity design that is sharp. Our goal is to identify effects of school of assignment on future outcomes. Another interesting problem is the effect of school of graduation on future outcomes, which resembles a fuzzy RD because some students do not graduate from the same school they are assigned to. We defer this identification problem to future work as several issues beyond the scope of this paper arise in that case, e.g., multiple compliance types with unordered treatments.

Unlike a standard sharp RD, the fact that  $S_j(\omega)$  clears the cutoff  $c_j$  does not automatically determine that student  $\omega$  is allocated to school  $j$ . It is only when  $j$  is the most preferred school among the schools that the student affords; that is, when  $j$  is the favorite school in the set of schools for which the student clears the cutoff.

The first step in RD is to correctly identify the marginal individuals for a given cutoff and a given change in school, e.g., cutoff  $c_j$  and a change from school  $k$  to school  $j$ . For example, for any individual with score  $S_j$  just to the right of  $c_j$ , we need to determine two things: that the individual is matched to school  $j$ ; and that the individual would have been matched to school  $k$  if the score were just to the left of  $c_j$ . The cutoff representation implies that these two things depend on the counterfactual sets of available schools in either side of the cutoff and on the individual's preferences over these sets.

It is straightforward to obtain counterfactual sets of available schools in the case where all schools rely on the same placement score, that is,  $S_j = S_1$  for every  $j$ . For example, this is the case of the SD mechanism. In this case, the set of feasible schools is all schools with cutoff less than or equal to score  $S_1$ . Note that everybody just above (or just below) cutoff  $c_j$  has exactly the same set of feasible schools. For someone with  $S_1 \geq c_j$ , the counterfactual scenario has the score crossing to the left of cutoff  $c_j$ , and school  $j$  gets dropped from the set of feasible schools. Vice versa, for someone with  $S_1 < c_j$ , the counterfactual scenario adds school  $j$  to the set of feasible schools.<sup>3</sup>

---

<sup>2</sup>See also work by Azevedo and Budish (2019), who advocate for a robust notion of strategy-proofness in a large economy.

<sup>3</sup>A common practice in applied work consists of “cleaning” the submitted preference lists from irrelevant schools in cases where  $S_j = S_1$  for every  $j$ . For instance, say an individual submits  $P = \{1, 2, 3\}$  and  $c_2 > S_1 > c_1 > c_3$ . Given the cutoff characterization and truth telling, the matching assignment of this individual is school 1; the counterfactual assignment is school 3 even though  $2P3$ ; this is the case because school 2 has cutoff higher than the cutoff of school 1. The irrelevant school to be cleaned from  $P$  is school 2 in this case. The general idea is to remove all schools ranked below 1 that have cutoffs higher than  $c_1$ . See a description of this practice by Estrada and Gignoux (2017). The practice cannot be used to identify counterfactual assignments in cases where different schools use different placement scores, e.g., the DA mechanism.



Unlike SD, agents close and on the same side of a cutoff in DA differ in their sets of feasible schools. It is not immediate to think about their counterfactual sets. In DA, schools use different scores, and these scores may be functions (e.g., weighted averages) of a small set of primitive scores. That is the case in our application with the Chilean data. This makes the joint support of the distribution of scores highly dependent and complicates the counterfactual analysis. Dealing with this complexity is empirically relevant since many higher-education assignment mechanisms in the real world use DA. See Table 1, Panel B by Fack et al. (2019) for a list of examples.

Our framework allows for a variety of joint distributions on the vector of scores  $\mathbf{S}$  and works with the definition of counterfactual budget sets below.

**Definition 3** (Counterfactual Budget Sets). *Consider a student with a vector of scores  $\mathbf{S}$ . The budget set for such student is the student's set of feasible options,*

$$B(\mathbf{S}) := \{0\} \cup \{m \in \mathcal{J} : S_m \geq c_m\}.$$

*Fix a school  $j \in \mathcal{J}$  with cutoff  $c_j$ . The right-counterfactual budget set for this student at cutoff  $c_j$  is  $B_j^+(\mathbf{S}) := B(\mathbf{S}) \cup \{m : S_m = S_j \text{ and } c_m = c_j\}$ ; the left-counterfactual is  $B_j^-(\mathbf{S}) := B(\mathbf{S}) \setminus \{m : S_m = S_j \text{ and } c_m = c_j\}$ , where  $C \setminus D$  equals the set  $C$  minus the elements of set  $D$ .*

To help fix ideas, we run a simple example throughout the paper in the context of the SD mechanism.

**Example** (SD Example). *In SD,  $S_j = S_1$  for every  $j$ , and the definition of budget set above equals to:  $B(\mathbf{S}) = \{m : S_1 \geq c_m\}$ . Suppose we have four schools with cutoffs  $c_1 < c_2 < c_3 < c_4$ . Individuals in this economy have five different budget sets possible:  $\{0\}$ ,  $\{0, 1\}$ ,  $\{0, 1, 2\}$ ,  $\{0, 1, 2, 3\}$ , and  $\{0, 1, 2, 3, 4\}$ . For individuals near cutoff  $c_4$ , the counterfactual budget sets are:  $B_4^-(\mathbf{S}) = \{0, 1, 2, 3\}$  and  $B_4^+(\mathbf{S}) = \{0, 1, 2, 3, 4\}$ .<sup>4</sup>*

Next, we define the concept of local preferences which involves the first- and second- best choices for a marginal individual at any given cutoff.

**Definition 4** (Local Preferences). *Fix a school  $j \in \mathcal{J}$  with cutoff  $c_j$ . Consider a student  $\omega$  with preference  $Q(\omega)$  and scores  $\mathbf{S}(\omega)$ . For any pair of options  $(k, l) \in \mathcal{J}^0 \times \mathcal{J}^0$ , we say that  $(k, l)$  is the local preference of student  $\omega$  at cutoff  $c_j$  if the favorite feasible option of student  $\omega$  shifts from  $l$  to  $k$  as we exogenously increase  $S_j(\omega)$  from being smaller than  $c_j$  to being larger than  $c_j$ . We define the true local preference of this student as the pair  $Q_j(\omega) := (k, l)$ . Formally, for a set of options  $B \subseteq \mathcal{J}^0$ , define the best option in  $B$  according to  $Q$  as  $Q(B)$ . We have that  $Q(B) = m \Leftrightarrow m \in B$  and  $m \bar{Q}(\omega) n \forall n \in B$ . Finally,  $Q_j(\omega) = (k, l)$  if, and only if,  $Q(B_j^+(\mathbf{S})) = k$  and  $Q(B_j^-(\mathbf{S})) = l$ . Moreover, the reported local preference  $P_j(\omega)$  is defined in a similar fashion. If  $B \cap P \neq \emptyset$ ,  $P(B) = m \Leftrightarrow m \in B$  and  $m \bar{P}(\omega) n \forall n \in B$ ; otherwise, if  $B \cap P = \emptyset$ ,  $P(B) = 0$ . We have that  $P_j(\omega) = (k, l)$  if, and only if,  $P(B_j^+(\mathbf{S})) = k$  and  $P(B_j^-(\mathbf{S})) = l$ .*

<sup>4</sup>In SD or in DA with independent placement scores, the definitions of counterfactual equal to:  $B_j^+(\mathbf{S}) = B(\mathbf{S}) \cup \{m : c_m = c_j\}$  and  $B_j^-(\mathbf{S}) = B(\mathbf{S}) \setminus \{m : c_m = c_j\}$ ; moreover, if cutoffs are unique:  $B_j^+(\mathbf{S}) = B(\mathbf{S}) \cup \{j\}$  and  $B_j^-(\mathbf{S}) = B(\mathbf{S}) \setminus \{j\}$ .

**Example** (SD Example, Continued). Consider four individuals whose submitted preferences are:  $P^{(1)} = \{3, 4, 2, 1\}$ ,  $P^{(2)} = \{4, 1, 2, 3\}$ ,  $P^{(3)} = \{4, 2, 3, 1\}$ , and  $P^{(4)} = \{4, 3, 1, 2\}$ . Their corresponding local preferences are:  $P_4^{(1)} = (3, 3)$ ,  $P_4^{(2)} = (4, 1)$ ,  $P_4^{(3)} = (4, 2)$ , and  $P_4^{(4)} = (4, 3)$ .

There is no distinction between  $P_j$  and  $Q_j$  at this stage because of Assumption 1. Students submit their true list of acceptable schools so that  $P$  equals the schools that are listed higher than 0 in  $Q$  and  $P_j = Q_j$ . Section 4 considers the case of misreporting due to strategic behaviors. In that case,  $P_j$  is observed but  $Q_j$  is not. Cutoff characterization implies that an individual with  $Q_j = (k, l)$  is matched to school  $k$  if  $S_j$  is just above  $c_j$  or to school  $l$  if  $S_j$  is just below  $c_j$ . Same thing applies to individuals with  $P_j = (k, l)$  under Assumption 1.

The local preference pair  $(j, k)$  at a certain cutoff  $c_j$  is only useful for identification if there exists a positive fraction of individuals in the data near cutoff  $c_j$  with those local preferences. We collect such useful pairs in the set  $\mathcal{P}$ .

**Definition 5** (Comparable Pairs). We say  $(j, k) \in \mathcal{J} \times \mathcal{J}$ ,  $j \neq k$ , is a comparable pair of alternatives if (i)  $c_j$  is an interior point of the support  $S_j$ ; and (ii)  $\mathbb{P}[Q_j = (j, k) | S_j = s]$  is bounded away from zero for  $s$  in an open neighborhood of  $c_j$ . Finally, we define  $\mathcal{P} \subseteq \mathcal{J} \times \mathcal{J}$  as the set of all comparable pairs.

We adopt the convention that comparable pairs do not involve the outside option 0. We do this because it may be hard to interpret treatment effects involving changes from an outside option to a school when the outside option varies across individuals. We do not consider pairs with  $j = k$  because the initial school assignment does not change for those individuals. We also exclude pairs  $Q_j = (j', k)$  with  $j' \neq j$  from  $\mathcal{P}$  to avoid redundancy. We may find individuals with  $Q_j = (j', k)$  whenever schools  $j$  and  $j'$  use the same score and have the same cutoff. As the score  $S_j = S_{j'}$  crosses the cutoff  $c_j = c_{j'}$ , access is granted to both schools  $j$  and  $j'$  and individuals may differ in their preferences for these schools. Individuals that prefer  $j'$  won't appear in  $\mathcal{P}$  as having  $Q_j = (j', k)$ , but they will appear in  $\mathcal{P}$  with  $Q_{j'} = (j', k)$ .

The purpose of defining counterfactual sets and local preferences is to construct a local preference variable for every student and use that as a control variable in the RD. In this section, this variable is  $P_j$ , which equals to  $Q_j$  because of Assumption 1. By focusing on students with  $P_j = (j, k)$ ,  $k \neq j$ , the marginal switch in allocation around cutoff  $c_j$  becomes a function of  $S_j$ . Controlling for  $P_j$  ensures that we apply the RD strategy to all individuals that switch assignment from  $k$  to  $j$  at the cutoff; this identifies the effect of changing assignment, as long as the typical RD continuity assumptions are satisfied.

RD identification requires continuity assumptions on the distribution of individual types conditional on the relevant placement score. In our case, we also need to verify continuity after we condition on  $P_j = Q_j$ . After all, we don't want to condition on something that hurts the central argument for identification in RD: individuals on the right and on the left of the cutoff are "similar on average". Below, we state an assumption on the continuity of types and prove that it implies the kind of smoothness required by RD. Before we do so, we define the following set of events. For every

school  $j \in \mathcal{J}$ , partition the placement scores  $\mathbf{S}$  as the score of school  $j$  and all other scores:  $\mathbf{S} \equiv (S_j, \mathbf{S}_{-j})$ . Define  $\mathbf{A}_{-j}$  to be the collection of events on  $\mathbf{S}_{-j}$  that determine the availability of all non- $j$  schools. There are  $2^{J-1}$  such events in  $\mathbf{A}_{-j}$ . For example, if  $J = 2$ ,  $\mathbf{A}_{-1} = \{\{S_2 \geq c_2\}, \{S_2 < c_2\}\}$ ; if  $J = 3$ ,  $\mathbf{A}_{-1} = \{\{S_2 \geq c_2, S_3 \geq c_3\}, \{S_2 \geq c_2, S_3 < c_3\}, \{S_2 < c_2, S_3 \geq c_3\}, \{S_2 < c_2, S_3 < c_3\}\}$ ; and so forth.

**Assumption 2.** (*Continuity of Types*) Consider a school  $j$  with cutoff  $c_j$  in the interior of the support  $\mathcal{S}_j$ . Assume the following functions of  $s$  are all continuous at  $s = c_j$ : (i)  $\mathbb{P}[\mathbf{S}_{-j} \in A_0, Q = Q_0 | S_j = s]$  for any  $A_0 \in \mathbf{A}_{-j}$  and  $Q_0 \in \mathcal{Q}$ ; and (ii)  $\mathbb{E}[g(Y(d)) \mathbb{I}\{\mathbf{S}_{-j} \in A_0, Q = Q_0\} | S_j = s]$  for any  $A_0 \in \mathbf{A}_{-j}$ ,  $Q_0 \in \mathcal{Q}$ , and  $g \in \mathcal{G}$ , where  $\mathcal{G}$  is a set of measurable functions  $g : \mathbb{R} \rightarrow \mathbb{R}$  that includes the constant function  $g(y) = 1$  and the identity function  $g(y) = y$ .

**Lemma 1.** Suppose Assumption 2 holds. Consider a school  $j$  with cutoff  $c_j$  in the interior of the support  $\mathcal{S}_j$ , and choose two schools  $k, l \in \mathcal{J}^0$  such that  $\mathbb{P}[Q_j = (k, l) | S_j = c_j] > 0$ . Then, for any function  $g \in \mathcal{G}$  and any  $d \in \mathcal{J}^0$  we have that  $\mathbb{E}[g(Y(d)) | Q_j = (k, l), S_j = s]$  and  $\mathbb{P}[Q_j = (k, l) | S_j = s]$  are continuous functions of  $s$  at  $s = c_j$ .

The proof of this lemma and all other proofs are found in the appendix. Finally, Assumptions 1 and 2 give sufficient conditions for identification for comparable pairs of school changes.

**Proposition 1.** Suppose Assumptions 1–2 hold. For any pair  $(j, k) \in \mathcal{P}$ ,

$$\mathbb{E}[g(Y(j)) | Q_j = (j, k), S_j = c_j] = \mathbb{E}[g(Y) | P_j = (j, k), S_j = c_j^+]$$

$$\mathbb{E}[g(Y(k)) | Q_j = (j, k), S_j = c_j] = \mathbb{E}[g(Y) | P_j = (j, k), S_j = c_j^-]$$

$$\begin{aligned} \mathbb{E}[g(Y(j)) - g(Y(k)) | Q_j = (j, k), S_j = c_j] \\ = \mathbb{E}[g(Y) | P_j = (j, k), S_j = c_j^+] - \mathbb{E}[g(Y) | P_j = (j, k), S_j = c_j^-], \end{aligned}$$

where the condition  $S_j = c_j^+$  denotes the limit as  $S_j \downarrow c_j$  and the condition  $S_j = c_j^-$  denotes the limit as  $S_j \uparrow c_j$ .

Proposition 1 shows that a standard RD is valid in the truth-telling case, as long as we control for  $P_j$ . The parameter of interest is the average treatment effect on  $g(Y)$  from changing the school of assignment from  $j$  to  $k$ , averaged over individuals at the cutoff  $c_j$  and with true local preferences  $(j, k)$ . Indeed, these are the parameters of interest for the rest of this paper,

$$\mathbb{E}[g(Y(j)) - g(Y(k)) | Q_j = (j, k), S_j = c_j], \quad (j, k) \in \mathcal{P}, \quad g \in \mathcal{G}. \quad (1)$$

Parameters like these are of economic interest to evaluate the current mechanism in place but also to study the effects of counterfactual assignments, for example, when students near  $c_j$  currently at school  $k$  are offered admission into school  $j$ . The fact that the parameter is conditional on  $Q_j$  instead of  $P_j$  is important for the external validity of (1) when  $P$  is manipulable by agents.

## 4 Identification with Strategic Reports

This section studies identification of causal effects when students strategically report their rankings of acceptable schools. Strategic reports make  $P_j$  generally different from  $Q_j$  and  $Q_j$  is not observed. Identification strategies that condition on the observed  $P_j$  are potentially invalid for two reasons. First, unlike Proposition 1, controlling for  $P_j$  does not identify the parameters of interest in (1), which condition on  $Q_j$ . Second, controlling for  $P_j$  breaks the internal validity of the RD in some cases because  $P_j$  is a variable subject to manipulation by agents.

We propose a two-step identification approach that is robust to strategic reporting. In the first step, the researcher characterizes the set of true local preferences  $Q_j$  that is compatible with the data and appropriate behavioral assumptions. In the second step, the researcher controls for local-preference sets and partially identifies the parameters in (1). We discuss first and second steps in Sections 4.1 and 4.2, respectively. Note that our two-step approach differs from the usual two-step control function approach in econometrics. The usual approach point identifies the control variable in the first step, while our approach partially identifies the control variable.

Section 4.1 presents several identification tools for local-preference sets. These tools rely on assumptions that are known to be appropriate in SD and DA contexts, although we do not exclude other mechanisms; e.g., the TTC mechanism satisfies one of those assumptions and some of the identification tools from Section 4.1 are still useful. More generally, researchers may utilize preference identification tools that work under alternative assumptions, for example, the methods of Agarwal and Somaini (2018) and Fack et al. (2019). Either way, the researcher must construct a set of local preferences for each individual in the first step.

Section 4.2 describes our second step procedure. The procedure places high-level assumptions on local-preference sets that do not restrict the researcher to the methods of Section 4.1.

### 4.1 Partial Identification of Local Preferences

This section provides tools for set identification of local preferences using assumptions on the agents' behavior and the mechanism. These assumptions are specific to this subsection and we motivate them in the context of the constrained DA mechanism studied by Haeringer and Klijn (2009). Haeringer and Klijn (2009) study the game where students submit constrained preference rankings and a mechanism matches students to schools as a function of  $P$ ,  $\mathbf{R}$ , and schools capacities. Although the unconstrained DA mechanism is strategy proof, many real-world implementations of DA restrict the number of schools that students submit in their rankings. In this case, there is a quota  $K < J$  such that  $1 \leq |P| \leq K$ , and the submitted ranking  $P$  is generally different from the list of acceptable schools in  $Q$ . The DA mechanism implemented in this way is not strategy proof, and there are no dominant strategies. Strategy-proofness also breaks down if, instead of a quota constraint, students incur a cost as a function of the number of schools submitted (Fack et al., 2019).

Lemma 4.2 by Haeringer and Klijn (2008) shows that, if a mechanism is strategy proof when

$K = J$ , then, in the game with  $K < J$ , any constrained ranking of schools is weakly dominated by the same set of schools ranked according to true preferences. This result implies that a student cannot lose and may possibly gain by taking any arbitrary list with less than or equal to  $K$  schools, drop unacceptable schools, and rank acceptable schools according to the student's true preferences. A further implication is that if a student's number of acceptable schools is less than or equal to  $K$ , then the student's dominant strategy is to submit the true list of acceptable schools (Proposition 4.2 by Haeringer and Klijn (2009)). These implications lead to a class of undominated strategies according to the following definitions of partial order.

**Definition 6.** (*Weak and Strong Partial Order*) We say  $P$  is a weak partial order of  $Q$  if  $P$  is any selection of up to  $K$  schools among the acceptable schools in  $Q$ , and that selection of schools is ranked according to  $Q$ . Formally, (i)  $1 \leq |P| \leq K$ ,  $P \subseteq \{d \in Q : dQ0\}$ ; and (ii) for every  $d, d' \in P$ ,  $d'Pd \Leftrightarrow d'Qd$ . We say  $P$  is a strong partial order of  $Q$  when a third condition holds in addition to (i) and (ii). Namely, (iii)  $|P| = \min\{K, |\{d \in Q : dQ0\}|\}$ . In other terms, if the number of acceptable schools in  $Q$  is less than or equal to  $K$  and  $P$  is a strong partial order of  $Q$ , then  $P$  equals the list of acceptable schools in  $Q$ ; otherwise, if the number of acceptable schools in  $Q$  is greater than  $K$ ,  $P$  is a subset of  $K$  schools among the acceptable schools in  $Q$ .

Lemma 2 in Section A.3 of the appendix summarizes the implications of the result on partial orders by Haeringer and Klijn (2008, 2009) in terms of Definition 6. In short, for a student with true preferences  $Q$ , any  $P$  is weakly dominated by a weak partial order  $P^*$  of  $Q$  that has the same acceptable schools as  $P$ ; in turn,  $P^*$  is weakly dominated by a strong partial order  $P^{**}$  of  $Q$  that contains the same set of acceptable schools as  $P^*$ . Every strong partial order is a weak partial order, but not the other way around.

Assuming agents always submit a strong partial order implies they reveal their true ordered list of acceptable schools whenever they submit  $P$  with less schools than the quota constraint  $K$ . That could be a strong behavioral assumption in some contexts where agents have more than  $K$  acceptable schools but have a strong prior for admission into a smaller than  $K$  set of schools. In this case, they may submit  $|P| < K$  not because that's their full list of acceptable schools but simply because they may not want to incur the mental costs of ranking all schools up to  $K$ . In the rest of this subsection, we consider mechanisms that impose a quota  $K$  on  $P$  and assume that students submit a weak partial order of their true preferences.

**Assumption 3** (Submission of Weak Partial Order). *Students submit a weak partial order of their true preferences.*

Assumption 3 replaces Assumption 1 to accommodate mechanisms that are not strategy proof. Submitting a weak partial order is the rational thing to do in DA mechanisms with quota constraints. That is true in any mechanism that becomes strategy proof once we remove the quota constraint, for example, the TTC mechanism.

A maintained assumption throughout this paper is the cutoff characterization of Definition 2. That assumption says that  $\mu(\omega) = Q(B(\mathbf{S}(\omega)))$  for every  $\omega \in \Omega$ . Azevedo and Leshno (2016)

show that stability is equivalent to cutoff characterization with  $\mathbf{S} = \mathbf{R}$  and cutoffs that equal the minimum score of admitted students in each school. Thus, it is worth discussing stability of the constrained DA mechanism. Theorem 6.3 by Haeringer and Klijn (2009) demonstrates that any Nash equilibrium in constrained DA where  $\mathbf{R}$  satisfies Ergin-acyclicity leads to a stable matching in the finite economy.<sup>5</sup> Even without Ergin-acyclicity, some Nash equilibria still produce stability. SD always satisfies Ergin-acyclicity, thus every Nash equilibrium in SD produces a stable matching. Fack et al. (2019) also study the constrained DA mechanism. They extend Theorem 6.3 of Haeringer and Klijn (2009) to (pure-strategy) Bayesian Nash equilibria in the continuum economy (Proposition A3, Online Appendix A.2.5, Fack et al. (2019)). Fack et al. (2019) also provide primitive conditions for finite economies where students play partial orders to converge to a continuum economy with a stable equilibrium (Proposition 5, Fack et al. (2019)). They further provide a test for implications of stability and do not find empirical or simulation evidence against it. In the context of the constrained TTC mechanism, any Nash equilibrium leads to a stable matching as long as  $\mathbf{R}$  satisfies Kesten-acyclicity (Theorem 6.4 by Haeringer and Klijn (2009)). Therefore, constrained SD, DA, and TTC all satisfy the cutoff characterization of Definition 2 with  $\mathbf{S} = \mathbf{R}$  under the appropriate conditions.

There is another interesting feature of the cutoff characterization of DA mechanisms. We know that DA produces a stable matching if agents are truth tellers. In case agents are not truth tellers, the matching outcome continues to be “stable” if we replace  $Q$  with  $P$  in the definition of stability.

**Definition 7** (Stability wrt  $P$ ). *We say the matching  $\mu : \Omega \rightarrow \mathcal{J}^0$  is a stable matching wrt  $P$  if three conditions are satisfied for every  $\omega \in \Omega$ : (i)  $\mu(\omega)\bar{P}(\omega)0$  (individual rationality); (ii) for any  $j \in \mathcal{J}$ , if  $jP(\omega)\mu(\omega)$ , then  $j$  is full (no waste); and (iii) for any  $j \in \mathcal{J}$  that is full, if  $\mu(\omega') = j$  and  $jP(\omega)\mu(\omega)$ , then  $R_j(\omega') > R_j(\omega)$  (no justified envy), where we adopt the convention that  $mP0$  for every  $m \in P$ . This is the same as Definition 1 except for  $P$  in the place of  $Q$ .*

The DA mechanism, constrained or unconstrained, produces a matching that is stable wrt reported preferences  $P$ . Stability wrt  $P$  leads to a cutoff characterization wrt  $P$  by the work of Azevedo and Leshno (2016). That cutoff characterization has scores  $\mathbf{S} = \mathbf{R}$  and admission cutoffs that equal the smallest scores of admitted students in each school. In other words, this is the same cutoff characterization from Definition 2 except that  $Q$  is replaced with  $P$ . Cutoff characterization wrt  $P$  is natural in DA but not necessarily in other mechanisms, so we state it in the following assumption.

**Assumption 4** (Cutoff Characterization wrt  $P$ ). *In addition to the maintained assumption of cutoff characterization as in Definition 2, the matching function  $\mu$  also satisfies  $\mu(\omega) = P(B(\mathbf{S}(\omega)))$  for every  $\omega \in \Omega$ .*

Assumption 4 essentially says that agents are matched to their best feasible options, where best is now defined according to  $P$ . Assumption 4 is convenient because it allows us to write

---

<sup>5</sup>Ergin-acyclicity ensures that no student can block a potential improvement for any other two students without affecting the first student’s own assignment. See Ergin (2002) for the formal definition.

a simple expression for the identified set of local preferences in Proposition 2 below; however it is not a necessary assumption to identify those sets. The convenience comes from the fact that Assumption 4 implies  $\mu = P(B(\mathbf{S})) = Q(B(\mathbf{S}))$ , where both  $\mu$  and  $P$  are observed and  $P$  and  $Q$  are related via the weak partial order assumption. If we drop Assumption 4, we only have one equality  $\mu = Q(B(\mathbf{S}))$ , which leads to bigger sets of local preferences in Proposition 2 below. This is useful to know for cases like the TTC mechanism, which is not stable wrt  $P$  and thus not a natural candidate for Assumption 4.

Next, we characterize all possible pairs of local preferences at a cutoff that are compatible with the data and Assumptions 3 and 4.

**Proposition 2** (Identification of Local-Preference Sets). *Suppose Assumptions 3 and 4 hold. Select a school  $j$  with cutoff  $c_j$ . Consider a student with scores  $\mathbf{S} \equiv (S_j, \mathbf{S}_{-j})$  and submitted preferences  $P$ . Call  $(a, b) = P_j$ . For such student, define  $N_j^+ = B_j^+(\mathbf{S}) \setminus \{P \cup \{0\}\}$  and  $N_j^- = B_j^-(\mathbf{S}) \setminus \{P \cup \{0\}\}$ , respectively, the sets of unlisted feasible schools in the counterfactual budget sets on the right and on the left of the cutoff. Then, the  $Q_j$  of this student belongs to  $\mathbf{Q}_j$ , where the set  $\mathbf{Q}_j$  is defined as follows:*

$$\mathbf{Q}_j = \begin{cases} \{(a, b)\}, & \text{if } S_j \geq c_j \text{ and } a = b, \\ \{(a, b)\} \cup \left(\{a\} \times N_j^-\right), & \text{if } S_j \geq c_j \text{ and } a \neq b, \\ \{(a, b)\} \cup \left((N_j^+ \setminus N_j^-) \times \{b\}\right) & \text{if } S_j < c_j, \end{cases} \quad (2)$$

where  $\left(\{a\} \times N_j^-\right)$  denotes the set formed by the Cartesian product of  $a$  and elements in  $N_j^-$  and  $\left(\{a\} \times N_j^-\right) = \emptyset$  if  $N_j^- = \emptyset$ .

Moreover, assume  $P$  is a strong partial order of  $Q$ . Then,  $\mathbf{Q}_j$  becomes:

$$\mathbf{Q}_j = \begin{cases} \{(a, b)\}, & \text{if } |P| < K, \text{ or if } |P| = K, S_j \geq c_j, \text{ and } a = b, \\ \{(a, b)\} \cup \left(\{a\} \times N_j^-\right), & \text{if } |P| = K, S_j \geq c_j, \text{ and } a \neq b, \\ \{(a, b)\} \cup \left((N_j^+ \setminus N_j^-) \times \{b\}\right) & \text{if } |P| = K \text{ and } S_j < c_j. \end{cases} \quad (3)$$

Finally, the characterization in (2) is sharp if the distribution of  $Q$  conditional on  $P$  and  $\mathbf{S}$  has full support, that is, every  $Q$  that satisfies Assumptions 3 and 4 is in that support. Likewise, (3) is sharp if the distribution of  $Q$  conditional on  $P$  and  $\mathbf{S}$  has full support under Assumptions 3, 4, and  $P$  being a strong partial order.

**Example** (SD Example, Continued). *Suppose the quota constraint is  $K = 3$  and the four schools are acceptable for everyone. We consider all agents whose  $P_4 = (4, 2)$ . For example, if agents submit strong partial orders, they submit either  $P = \{4, 2, 1\}$  or  $P = \{4, 2, 3\}$ . The assumption of cutoff characterization wrt  $P$  (Assumption 4) says that these agents are matched to school 4 if  $S_1 \geq c_4$ , otherwise to school 2. To keep things simple, consider five different types of true preferences:  $Q^{(1)} = \{4, 2, 3, 1, 0\}$ ,  $Q^{(2)} = \{4, 3, 2, 1, 0\}$ ,  $Q^{(3)} = \{4, 1, 3, 2, 0\}$ ,  $Q^{(4)} = \{3, 4, 2, 1, 0\}$ , and  $Q^{(5)} = \{1, 3, 2, 4, 0\}$ . The weak partial order assumption rules out  $Q^{(5)}$  because  $2Q^{(5)}4$*

contradicts 4 being reported preferred to 2. The maintained assumption of cutoff characterization (Definition 2) further rules out more types of  $Q$ , depending on whether  $S_1 \geq c_4$  or  $S_1 < c_4$ :

1. if  $S_1 \geq c_4$ ,  $Q^{(4)}$  is not possible because the matching assignment is 4 but the best feasible option according to  $Q^{(4)}$  is 3; in this case, the possible true local preferences are:  $Q_4^{(1)} = (4, 2)$ ,  $Q_4^{(2)} = (4, 3)$ , and  $Q_4^{(3)} = (4, 1)$ ; for a student that submits  $P = \{4, 2, 1\}$ ,  $\mathbf{Q}_4 = \{(4, 2), (4, 3)\}$ ; otherwise, for someone that submits  $P = \{4, 2, 3\}$ ,  $\mathbf{Q}_4 = \{(4, 2), (4, 1)\}$ ;
2. if  $S_1 < c_4$ , none of  $Q^{(2)}$ ,  $Q^{(3)}$ , or  $Q^{(4)}$  is possible because the matching assignment is 2 but the best feasible options according to these  $Q$ s differ from 2; in this case, the only possible true local preference is  $Q_4^{(1)} = (4, 2)$ , so that  $\mathbf{Q}_4 = \{(4, 2)\}$ .

This example illustrates why an RD that only controls for  $P_4 = (4, 2)$  may fail to identify the treatment effect of changing assignment from school 4 to school 2. As the score  $S_1$  crosses the cutoff  $c_4$ , the support of the distribution of true local preference types generally changes discontinuously from having only  $(4, 2)$  to having  $(4, 1)$ ,  $(4, 2)$ , and  $(4, 3)$ .

Proposition 2 identifies all possible values of  $Q_j$  for students near a cutoff  $c_j$  as function of their scores and submitted preferences. In some settings, there may be a large number of unlisted programs that are feasible to students, which translates into sets  $\mathbf{Q}_j$  that have many values. For example, the Chilean data has  $K = 8$  but more than 1,000 options; a student may have numerous feasible options but not consider many of them. It is possible to obtain smaller sets  $\mathbf{Q}_j$  by imposing further assumptions on the expectations that students have when they submit  $P$ .

Agarwal and Somaini (2018) propose a general framework to rationalize strategic reports as being the optimal solution to an expected utility maximization problem. Agents have private information about their preferences and scores. They form beliefs about the distribution of other people's preferences and scores. These beliefs plus knowledge of the mechanism lead the rational agent to derive probabilities of admission into the various schools as a function of the agent's private information and expectations on other agents. The agent then chooses the submission  $P$  that maximizes expected utility.

We assume that agents are expected-utility maximizers and note that it is sufficient for each individual to consider beliefs on admission cutoffs. This is the case in our continuum economy with cutoff characterization because cutoffs and scores fully characterize the agent's budget set. Given the student's scores, a distribution of possible cutoffs translates into a distribution of possible budget sets. Under Assumption 4, the student is admitted to the best school according to the submission  $P$  among the available schools in the budget set. Therefore, beliefs on cutoffs translate into probabilities of admission into various schools for any given  $P$ . We make an assumption on the distribution of cutoffs expected by agents that has to do with the concept of uniformly more accessible schools.

**Definition 8** (Uniformly More Accessible Schools). *For a pair of distinct schools  $(d, e)$ , we say  $e$  is uniformly more accessible than  $d$  if two conditions are satisfied. First, access to school  $d$  implies*



access to school  $e$ ,

$$\{\omega : S_d(\omega) \geq c_d\} \subseteq \{\omega : S_e(\omega) \geq c_e\}.$$

Second, replacing option  $d$  with option  $e$  in any submission  $P$  alters the likelihood of admission of at least one school listed in  $P$ ; formally, for any two fixed (i.e., non-random) submissions  $P$  and  $\tilde{P}$  such that  $P$  has  $d$  but does not have  $e$  and  $\tilde{P}$  equals  $P$  except for  $e$  in the place of  $d$ , there exists  $u \in \{0, 1, \dots, |P|\}$  for which

$$\mathbb{P}[P(B(\mathbf{S})) = P^u] \neq \mathbb{P}[\tilde{P}(B(\mathbf{S})) = \tilde{P}^u],$$

where  $P(B)$  denotes the best choice in set  $B$  according to  $P$  (Definition 4) and  $P^u$  denotes the school ranked in the  $u$ -th position in  $P$ . In short, we say  $(d, e) \in \text{UMAS}$ , where  $\text{UMAS} \subseteq \mathcal{J} \times \mathcal{J}$  is the set of all such pairs.

Definition 8 says that  $e$  is uniformly more accessible than  $d$  if everyone that qualifies for school  $d$  also qualifies for school  $e$ . Schools  $d$  and  $e$  must also be relevant in the sense of the second condition: there is always a strictly positive fraction of individuals for which listing  $e$  in the place of  $d$  changes their best feasible options. In the SD case, a sufficient condition for Assumption 5 is that  $c_d > c_e$  and cutoffs are distinct interior points in the support of the placement score. Uniformly more accessible schools do not always exist. Existence depends on the mechanism in place and the joint distribution of the placement scores. We use this definition to impose a mild restriction on the expectations of agents regarding cutoffs.

**Assumption 5.** Consider a student with scores  $\mathbf{s} \in \mathcal{S}$  who views uncertain cutoffs as random variables  $C_1, \dots, C_J$  before the matching assignment. Let  $\tilde{B} = \{0\} \cup \{j \in \mathcal{J} : s_j \geq C_j\}$  be the corresponding random budget set of the student. For every pair  $(d, e) \in \text{UMAS}$ , the distribution of cutoffs for this student is such that two conditions are satisfied. First,

$$\{s_d \geq C_d\} \subseteq \{s_e \geq C_e\}.$$

Second, for any two fixed (i.e., non-random) submissions  $P$  and  $\tilde{P}$  such that  $P$  has  $d$  but does not have  $e$  and  $\tilde{P}$  equals  $P$  except for  $e$  in the place of  $d$ , there exists  $u \in \{0, 1, \dots, |P|\}$  for which

$$\mathbb{P}[P(\tilde{B}) = P^u] \neq \mathbb{P}[\tilde{P}(\tilde{B}) = \tilde{P}^u].$$

This is true for every student in the economy.

Assumption 5 says that students correctly anticipate which schools will be uniformly more accessible after the matching assignment. For example, agents may learn that by observing past realizations of the matching in the economy. If a school  $e$  is well known to be accessible to everyone who has access to school  $d$ , then it is natural that a student expects to have access to  $e$  if he ever has access to school  $d$ . Note that the assumption does not pin down the expected probability of admission or the set of schools the student will have access to in the ex-post economy. It only

restricts the hierarchy of school access according to  $UMAS$ . This assumption has implications for the joint distribution of  $(P, Q)$ .

**Proposition 3.** *Suppose Assumptions 3–5 hold. Consider a student with reported preference ranking  $P$ . Let  $(P \times P^c)$  be the Cartesian product of listed and unlisted schools, respectively,  $P$  and  $P^c$ . If  $(d, e) \in UMAS \cap (P \times P^c)$ , then  $dQe$ .*

Proposition 3 says that if an agent lists school  $d$  but does not list the uniformly more accessible school  $e$  it must be that this agent prefers  $d$  over  $e$ . The result offers a refinement to Proposition 2 above.

**Corollary 1.** *Consider the setup of Proposition 2 and suppose Assumption 5 holds. Define  $A_j^- = N_j^- \setminus \{e : \exists d \in P \text{ with which } (d, e) \in UMAS \text{ and } b\bar{P}d\}$  and  $A_j^+ = (N_j^+ \setminus N_j^-) \setminus \{e : \exists d \in P \text{ with which } (d, e) \in UMAS \text{ and } a\bar{P}d\}$ . Then, under weak partial order,*

$$\mathbf{Q}_j = \begin{cases} \{(a, b)\}, & \text{if } S_j \geq c_j \text{ and } a = b, \\ \{(a, b)\} \cup \left(\{a\} \times A_j^-\right), & \text{if } S_j \geq c_j \text{ and } a \neq b, \\ \{(a, b)\} \cup \left(A_j^+ \times \{b\}\right) & \text{if } S_j < c_j. \end{cases} \quad (4)$$

*This characterization is sharp if (2) is sharp and  $\mathbb{P}[eQd|P, \mathbf{S}] = 0$  for every  $(d, e) \in UMAS \cap (P \times P^c)$ .*

*Under strong partial order,*

$$\mathbf{Q}_j = \begin{cases} \{(a, b)\}, & \text{if } |P| < K, \text{ or if } |P| = K, S_j \geq c_j, \text{ and } a = b, \\ \{(a, b)\} \cup \left(\{a\} \times A_j^-\right), & \text{if } |P| = K, S_j \geq c_j, \text{ and } a \neq b, \\ \{(a, b)\} \cup \left(A_j^+ \times \{b\}\right) & \text{if } |P| = K \text{ and } S_j < c_j. \end{cases} \quad (5)$$

*This characterization is sharp if (3) is sharp and  $\mathbb{P}[eQd|P, \mathbf{S}] = 0$  for every  $(d, e) \in UMAS \cap (P \times P^c)$ .*

Corollary 1 describes how to use Proposition 3 to potentially reduce the number of elements in the  $\mathbf{Q}_j$  constructed in Proposition 2. The intuition runs as follows. Suppose a student submits  $P$  and  $P_j = (a, b)$ . If school  $e$  is uniformly more accessible than school  $d$ ,  $d$  is listed in  $P$ , but  $e$  is not listed in  $P$ , then we know the student truly prefers  $d$  over  $e$ . This excludes some possibilities of  $Q_j$  in the  $\mathbf{Q}_j$  defined by Proposition 2. For instance, this person cannot have  $Q_j = (a, e)$  if  $b\bar{P}d$  because the latter implies that  $b\bar{Q}dQe$ , a contradiction with  $dQe$ . Likewise, this person cannot have  $Q_j = (e, b)$  if  $a\bar{P}d$ .

**Example** (SD Example, Continued). *The set of uniformly more accessible schools is  $UMAS = \{(2, 1), (3, 2), (3, 1), (4, 3), (4, 2), (4, 1)\}$ . Assumption 5 shrinks the set  $\mathbf{Q}_4$  of those agents with  $S_1 \geq c_4$  and  $P = \{4, 2, 3\}$ . Applying the assumption changes  $\mathbf{Q}_4 = \{(4, 2), (4, 1)\}$  to  $\mathbf{Q}_4 = \{(4, 2)\}$  because 4 is uniformly more accessible than 1, 4 is listed, 1 is not listed, and  $2P1$ .*

## 4.2 Partial Identification of Causal Effects

In this section, we lay down conditions and derive bounds on average treatment effects. We assume that the researcher has already identified the set of local preferences at a cutoff of interest. That means the researcher has a set-valued variable  $\mathbf{Q}_j$  for all students in the vicinity of a cutoff  $j$  corresponding to a comparable pair  $(j, k)$  in  $\mathcal{P}$ . Researchers may construct  $\mathbf{Q}_j$  using the methods of Section 4.1 if they find reasonable to rely on at least some of the specific assumptions of that subsection; otherwise, they may use any other method to construct  $\mathbf{Q}_j$ . There is no restriction on the choice of the method to construct  $\mathbf{Q}_j$  except for a couple of high-level conditions that we assume in this section. We start by defining the conditional support of partially-identified true local preferences.

**Definition 9** (Support of Local-Preference Sets). *Consider a pair  $(j, k) \in \mathcal{P}$  and corresponding cutoff  $c_j$ . The support of partially-identified true local preferences conditional on  $S_j = s$  is defined as,*

$$\mathbf{\Lambda}_j(s) = \{B \subseteq \mathcal{J}^0 \times \mathcal{J}^0 : \mathbb{P}[\mathbf{Q}_j = B | S_j = s] > 0\}.$$

The union set of that support is defined as the collection of all unions of sets in  $\mathbf{\Lambda}_j(s)$ , namely,

$$\mathbf{\Lambda}_j^\cup(s) = \{B^\cup \subseteq \mathcal{J}^0 \times \mathcal{J}^0 : \exists B_1, B_2, \dots \in \mathbf{\Lambda}_j(s) \text{ with } B^\cup = \cup_i B_i\}.$$

The set  $\mathbf{\Lambda}_j(s)$  collects all values of  $\mathbf{Q}_j$  that occur with positive probability conditional on  $S_j = s$ . In the specific context of Section 4.1,  $\mathbf{Q}_j$  is constructed from mapping observables  $(P, \mathbf{S})$  to a subset of  $\mathcal{J}^0 \times \mathcal{J}^0$ , i.e.,  $\mathbf{Q}_j = \psi_j(P, \mathbf{S})$ . For example, Proposition 2 and Corollary 1 give examples of such mapping  $\psi_j$ . A set  $B$  of pairs  $(a, b) \in \mathcal{J}^0 \times \mathcal{J}^0$  belongs to the support set  $\mathbf{\Lambda}_j(s)$  if there is a set of values in the support of the conditional distribution of  $(P, \mathbf{S})$  given  $S_j = s$  such that  $\psi_j$  maps those values to the set  $B$ . The union set  $\mathbf{\Lambda}_j^\cup(s)$  collects all possible unions of support points of  $\mathbf{Q}_j$  conditional on  $S_j = s$ . These definitions are instrumental in the computation of the partially identified distribution of  $Q_j$ , as explained in Proposition 4 below.

Sharpness of identification of the distribution of  $Q_j$  requires sharpness in the construction of the sets  $\mathbf{Q}_j$ . Proposition 2 and Corollary 1 gave conditions for sharpness of  $\mathbf{Q}_j$  in the context of Section 4.1. Outside of that context, researchers may construct  $\mathbf{Q}_j$  in a different way, so we impose sharpness of  $\mathbf{Q}_j$  in the general form of the assumption below.

**Assumption 6** (Sharp Local Preference Sets). *Consider a pair  $(j, k) \in \mathcal{P}$  and corresponding cutoff  $c_j$ . Assume that:*

- (i) *the random variable  $Q_j$  and the random set  $\mathbf{Q}_j$  are both measurable maps on the same probability space and  $\mathbb{P}[Q_j \in \mathbf{Q}_j | S_j] = 1$  with probability 1;*
- (ii)  *$\text{supp}[Q_j | \mathbf{Q}_j, S_j] = \mathbf{Q}_j$  with probability 1, where  $\text{supp}[Y | X]$  denotes the support set of the distribution of  $Y$  conditional on  $X$ .*

Assumption 6(i) says that  $\mathbf{Q}_j(\omega)$  of individual  $\omega$  contains the true pair of local preferences  $Q_j(\omega)$  of that individual (for almost all individuals), which is a minimum requirement for the construction

of  $\mathbf{Q}_j(\omega)$ . This does not say anything on the sharpness of  $\mathbf{Q}_j$ . For example,  $\mathbf{Q}_j = \mathcal{J}^0 \times \mathcal{J}^0$  is completely uninformative and trivially satisfies Assumption 6(i). The sharpness requirement is stated in Assumption 6(ii). It says that all possibilities of local preferences listed in  $\mathbf{Q}_j$  actually occur in the data with positive probability. That rules out unnecessarily big sets  $\mathbf{Q}_j$ . Assumption 6(ii) may be dropped at the cost of not having sharp identified sets in the rest of this section.

Partial identification of true local preferences and treatment effects occur at the limit, as  $S_j$  approaches  $c_j$ , and is conditional on  $\mathbf{Q}_j$ . For that to work, we impose regularity conditions on the distribution of potential outcomes and  $\mathbf{Q}_j$  conditional on  $S_j$  at the limit  $c_j$ .

**Assumption 7** (Distribution of Local Preference Sets). *Consider a pair  $(j, k) \in \mathcal{P}$  and corresponding cutoff  $c_j$ . Assume that:*

(i) *there exist a small  $\varepsilon > 0$  and collections of subsets of  $\mathcal{J}^0 \times \mathcal{J}^0$  denoted  $\Lambda_j^+$  and  $\Lambda_j^-$  such that  $\Lambda_j^+ = \Lambda_j(c_j + e) \forall e \in [0, \varepsilon)$  and  $\Lambda_j^- = \Lambda_j(c_j - e) \forall e \in (0, \varepsilon)$ ; consistent with Definition 9, we define  $\Lambda_j^{\cup+}$  and  $\Lambda_j^{\cup-}$  as union sets of  $\Lambda_j^+$  and  $\Lambda_j^-$ , respectively;*

(ii) *for any  $g \in \mathcal{G}$  of Assumption 2 and any  $\underline{\tau}, \bar{\tau} \in \mathbb{R} \cup \{-\infty, +\infty\}$ ,  $\underline{\tau} < \bar{\tau}$ , the side-limits of the following expectations are well defined:  $\mathbb{E} \left[ g(Y) \mathbb{I}\{\mathbf{Q}_j = A, \underline{\tau} < g(Y) < \bar{\tau}\} \mid S_j = c_j^+ \right] \forall A \in \Lambda_j^+$  and  $\mathbb{E} \left[ g(Y) \mathbb{I}\{\mathbf{Q}_j = A, \underline{\tau} < g(Y) < \bar{\tau}\} \mid S_j = c_j^- \right] \forall A \in \Lambda_j^-$ .*

Assumption 7(i) concerns the distribution of  $\mathbf{Q}_j$  conditional on  $S_j$ : the support set of  $\mathbf{Q}_j$  is constant as  $S_j = s$  approaches the cutoff  $c_j$  from either side of it. Part (ii) of the assumption concerns the joint distribution of potential outcomes and  $\mathbf{Q}_j$  conditional on  $S_j$ . For example, Assumption 7(ii) implies that  $\mathbb{P} \left[ \mathbf{Q}_j = A \mid S_j = c_j^+ \right]$  and  $\mathbb{E}[Y \mid \mathbf{Q}_j = A, Y < \tau, S_j = c_j^+]$  are well defined limits for any  $A \in \Lambda_j^+$  and  $\tau \in \mathbb{R} \cup +\infty$  provided that  $\mathbb{P}[\mathbf{Q}_j = A, Y < \tau \mid S_j = c_j^+] > 0$ . The next result gives inequalities to construct bounds on  $\mathbb{P}[Q_j = (a, b) \mid S_j = c_j]$  for any pair  $(a, b)$ .

**Proposition 4** (Sharp Set of Distributions of Local Preferences). *Consider a pair  $(j, k) \in \mathcal{P}$ . Suppose Assumptions 2, 6, and 7 hold. Then, the sharp set of all possible discrete probability distributions of  $Q_j$  conditional on  $S_j = c_j$  is characterized as follows. For every  $A \in \Lambda_j^{\cup+} \cup \Lambda_j^{\cup-}$ , each probability distribution in that set implies a value for  $\mathbb{P}[Q_j \in A \mid S_j = c_j]$  that satisfies one of the three inequalities below:*

(i) *if  $A \in \Lambda_j^{\cup+} \cap \Lambda_j^{\cup-}$ ,*

$$\mathbb{P}[Q_j \in A \mid S_j = c_j] \geq \max \left\{ \mathbb{P} \left[ \mathbf{Q}_j \subseteq A \mid S_j = c_j^+ \right] ; \mathbb{P} \left[ \mathbf{Q}_j \subseteq A \mid S_j = c_j^- \right] \right\};$$

(ii) *if  $A \in \Lambda_j^{\cup+} \setminus \Lambda_j^{\cup-}$ ,*

$$\mathbb{P}[Q_j \in A \mid S_j = c_j] \geq \mathbb{P} \left[ \mathbf{Q}_j \subseteq A \mid S_j = c_j^+ \right]; \text{ or}$$

(iii) *if  $A \in \Lambda_j^{\cup-} \setminus \Lambda_j^{\cup+}$ ,*

$$\mathbb{P}[Q_j \in A \mid S_j = c_j] \geq \mathbb{P} \left[ \mathbf{Q}_j \subseteq A \mid S_j = c_j^- \right].$$

Proposition 4 provides a way to construct the sharp partially-identified set of all possible distributions of  $Q_j$  conditional on  $S_j = c_j$ . A distribution of  $Q_j$  conditional on  $S_j = c_j$  consists of values  $p_{a,b} \in [0, 1]$  for every  $(a, b) \in \mathcal{J}^0 \times \mathcal{J}^0$  such that  $\sum_{(a,b) \in \mathcal{J}^0 \times \mathcal{J}^0} p_{a,b} = 1$ , where  $p_{a,b} = \mathbb{P}[Q_j = (a, b) | S_j = c_j]$ . The sharp set is constructed by finding all values of  $p_{a,b}$  where  $\sum_{(a,b) \in A} p_{a,b}$  satisfies the inequalities of Proposition 4 for every  $A \in \mathbf{\Lambda}_j^{\cup+} \cup \mathbf{\Lambda}_j^{\cup-}$ .

**Example** (SD Example, Continued). *Continue to assume that the four schools are acceptable for everyone. Suppose for a moment that all combinations of  $(P, Q)$  that satisfy Assumptions 3–5 exist in the economy, both for  $S_1 \geq c_4$  and  $S_1 < c_4$ . Then, this is the list of all possible  $\mathbf{Q}_4$ :*

1. if  $S_1 \geq c_4$ ,  $\{(1, 1)\}, \{(2, 2)\}, \{(3, 3)\}, \{(4, 1)\}, \{(4, 2)\}, \{(4, 3)\}, \{(4, 1), (4, 2), (4, 3)\}, \{(4, 1), (4, 3)\}$ , and  $\{(4, 2), (4, 3)\}$ ;
2. if  $S_1 < c_4$ ,  $\{(1, 1)\}, \{(2, 2)\}, \{(3, 3)\}, \{(4, 1)\}, \{(4, 2)\}, \{(4, 3)\}, \{(1, 1), (4, 1)\}, \{(2, 2), (4, 2)\}$ , and  $\{(3, 3), (4, 3)\}$ .

Let's focus on the case that  $S_1 \geq c_4$ . To keep things simple, suppose three types of  $\mathbf{Q}_4$  occur with positive probability conditional on  $S_1 = s$  for any  $s \geq c_4$ :  $\{(4, 2)\}$  with probability 0.1,  $\{(4, 3)\}$  with probability 0.3, and  $\{(4, 2), (4, 3)\}$  with probability 0.6. It follows that  $\mathbf{\Lambda}_4(s) = \mathbf{\Lambda}_4^+ = \{\{(4, 2)\}, \{(4, 3)\}, \{(4, 2), (4, 3)\}\}$  and  $\mathbf{\Lambda}_4^{\cup}(s) = \mathbf{\Lambda}_4^{\cup+} = \{\{(4, 2)\}, \{(4, 3)\}, \{(4, 2), (4, 3)\}\}$ . The lower bounds  $\mathbb{P}[\mathbf{Q}_4 \subseteq A | S_1 = c_4^+]$  of Proposition 4 are as follows: 0.1 for  $A = \{(4, 2)\}$ ; 0.3 for  $A = \{(4, 3)\}$ ; and 1 for  $A = \{(4, 2), (4, 3)\}$ . Thus,  $\mathbb{P}[Q_4 = (4, 2) | S_1 = c_4^+]$  has lower bound 0.1,  $\mathbb{P}[Q_4 = (4, 3) | S_1 = c_4^+]$  has lower bound 0.3, and the sum of the two equals 1. Looking at each individual probability, the bounds are  $[0.1, 0.7]$  on  $\mathbb{P}[Q_4 = (4, 2) | S_1 = c_4^+]$  and  $[0.3, 0.9]$  on  $\mathbb{P}[Q_4 = (4, 3) | S_1 = c_4^+]$ .

Recall that the construction of the random set  $\mathbf{Q}_j$  depends on assumptions regarding the behavior of agents when they submit  $P$ . For example, Section 4.1 characterizes  $\mathbf{Q}_j$  by assuming weak partial order and cutoff characterization wrt  $P$ . Alternatively, the identification approach of Agarwal and Somaini (2018) makes different types of assumptions on agents' expectations and requires data variation on the choice environment. The theoretical credibility of these types of assumptions depends on the mechanism faced by agents; in practice, these assumptions imply testable implications for the observed data. It is therefore useful to characterize a falsification test based on these implications to aid researchers to screen out assumptions rejected by the data. The test relies on the fact that the right-hand sides of the inequalities in Proposition 4 must provide a lower bound for a probability mass function if the model assumptions are correct. For any partition  $\mathcal{A}$  of  $\mathcal{J}^0 \times \mathcal{J}^0$ , we must have  $\sum_{A \in \mathcal{A}} \mathbb{P}[Q_j \in A | S_j = c_j] = 1$  for any given distribution in the sharp set of Proposition 4. Thus the same sum applied to the right-hand sides of the inequalities above must be less than or equal to one.

**Corollary 2** (Model's Falsification Test). *Assume the setup of Proposition 4, which presupposes that the model assumptions utilized to construct  $\mathbf{Q}_j$  are true. Then, for any partition  $\mathcal{A}$  of  $\mathcal{J}^0 \times \mathcal{J}^0$*

we have that

$$\sum_{A \in \mathcal{A}} \left\{ \begin{aligned} & \mathbb{I} \left\{ A \in \Lambda_j^{U+} \cap \Lambda_j^{U-} \right\} \max \left\{ \mathbb{P} \left[ \mathbf{Q}_j \subseteq A | S_j = c_j^+ \right] ; \mathbb{P} \left[ \mathbf{Q}_j \subseteq A | S_j = c_j^- \right] \right\} \\ & + \mathbb{I} \left\{ A \in \Lambda_j^{U+} \setminus \Lambda_j^{U-} \right\} \mathbb{P} \left[ \mathbf{Q}_j \subseteq A | S_j = c_j^+ \right] \\ & + \mathbb{I} \left\{ A \in \Lambda_j^{U-} \setminus \Lambda_j^{U+} \right\} \mathbb{P} \left[ \mathbf{Q}_j \subseteq A | S_j = c_j^- \right] \end{aligned} \right\} \leq 1.$$

Partial identification of the distribution of local preferences allows us to bound the fraction of individuals near cutoff  $c_j$  that have  $Q_j = (j, k)$ . The average outcome near the cutoff is a weighted average of average outcomes from two different groups: first, individuals with  $Q_j = (j, k)$ , who interest us for the identification of treatment effects; and second, individuals with  $Q_j \neq (j, k)$ . The overall average is identified but not the average in each of the groups. A strictly positive lower bound on the fraction of individuals in the  $Q_j = (j, k)$  group allows us to construct lower and upper bounds on the average outcome for that group.

Start with all individuals above and near cutoff  $c_j$  whose  $\mathbf{Q}_j$  contain the comparable pair of interest,  $(j, k) \in \mathcal{P}$ . The fraction of those individuals that have  $Q_j = (j, k)$  equals to

$$\delta_{j,k}^+ = \frac{\mathbb{P} [Q_j = (j, k) | S_j = c_j]}{\mathbb{P} \left[ \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset | S_j = c_j^+ \right]},$$

where both numerator and denominator are strictly positive by virtue of  $(j, k)$  being a comparable pair (Definition 5) and sharpness of  $\mathbf{Q}_j$  (Assumption 6). The denominator of  $\delta_{j,k}^+$  is identified from the data, and Proposition 4 bounds the numerator. All we need for identification of treatment effects is a lower bound on  $\delta_{j,k}^+$ , which comes from a lower bound on its numerator. Let  $\underline{p}_{j,k}$  denote the infimum over all probability values for  $\mathbb{P} [Q_j = (j, k) | S_j = c_j]$  that belong to the partially identified set of Proposition 4. The sharp lower bound on  $\delta_{j,k}^+$  equals to

$$\underline{\delta}_{j,k}^+ = \frac{\underline{p}_{j,k}}{\mathbb{P} \left[ \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset | S_j = c_j^+ \right]}.$$

The denominator of  $\underline{\delta}_{j,k}^+$  is strictly positive, but  $\underline{p}_{j,k}$  may or may not be strictly positive.

Same idea applies for individuals just below the cutoff. Select all individuals whose  $\mathbf{Q}_j$  contain the comparable pair of interest,  $(j, k) \in \mathcal{P}$ . The fraction of those that have  $Q_j = (j, k)$  equals to

$$\delta_{j,k}^- = \frac{\mathbb{P} [Q_j = (j, k) | S_j = c_j]}{\mathbb{P} \left[ \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset | S_j = c_j^- \right]},$$

and the sharp lower bound on  $\delta_{j,k}^-$  is

$$\delta_{j,k}^- = \frac{\underline{p}_{j,k}}{\mathbb{P}[\mathbf{Q}_j \cap \{(j,k)\} \neq \emptyset | S_j = c_j^-]}.$$

**Example** (SD Example, Continued). *We have that  $\mathbb{P}[\mathbf{Q}_4 \cap \{(4,2)\} \neq \emptyset | S_1 = c_4^+] = 0.7$  and the bounds on  $\mathbb{P}[Q_4 = (4,2) | S_1 = c_4]$  are  $[0.1, 0.7]$ . These imply  $\underline{p}_{4,2} = 0.1$  and  $\underline{\delta}_{4,2}^+ = 1/7$ .*

The following result utilizes the proportions  $\underline{\delta}_{j,k}^+$  and  $\underline{\delta}_{j,k}^-$  to partially identify average outcomes for individuals with  $Q_j = (j,k)$  on either side of the cutoff. Taking differences of these bounds yield bounds for averages of treatment effects  $Y(j) - Y(k)$ .

**Proposition 5.** *Suppose Assumptions 2, 6, and 7 hold. Consider a pair  $(j,k) \in \mathcal{P}$  such that  $\underline{p}_{j,k} > 0$ .*

(i) *If  $g(Y)$  is a continuous random variable for  $g \in \mathcal{G}$  of Assumption 2, then we have the following bounds on  $\mathbb{E}[g(Y(j)) | Q_j = (j,k), S_j = c_j]$  and  $\mathbb{E}[g(Y(k)) | Q_j = (j,k), S_j = c_j]$ :*

$$\begin{aligned} & \mathbb{E} \left[ g(Y) \mid \mathbf{Q}_j \cap \{(j,k)\} \neq \emptyset, g(Y) < F_{j,k+}^{-1}(\underline{\delta}_{j,k}^+), S_j = c_j^+ \right] \\ & \leq \mathbb{E} [g(Y(j)) | Q_j = (j,k), S_j = c_j] \leq \\ & \mathbb{E} \left[ g(Y) \mid \mathbf{Q}_j \cap \{(j,k)\} \neq \emptyset, g(Y) > F_{j,k+}^{-1}(1 - \underline{\delta}_{j,k}^+), S_j = c_j^+ \right], \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E} \left[ g(Y) \mid \mathbf{Q}_j \cap \{(j,k)\} \neq \emptyset, g(Y) < F_{j,k-}^{-1}(\underline{\delta}_{j,k}^-), S_j = c_j^- \right] \\ & \leq \mathbb{E} [g(Y(k)) | Q_j = (j,k), S_j = c_j] \leq \\ & \mathbb{E} \left[ g(Y) \mid \mathbf{Q}_j \cap \{(j,k)\} \neq \emptyset, g(Y) > F_{j,k-}^{-1}(1 - \underline{\delta}_{j,k}^-), S_j = c_j^- \right], \end{aligned}$$

where  $F_{j,k+}^{-1}(u) := \inf \left\{ y : \mathbb{P} \left[ g(Y) \leq y \mid \mathbf{Q}_j \cap \{(j,k)\} \neq \emptyset, S_j = c_j^+ \right] \geq u \right\}$  and  $F_{j,k-}^{-1}(u) := \inf \left\{ y : \mathbb{P} \left[ g(Y) \leq y \mid \mathbf{Q}_j \cap \{(j,k)\} \neq \emptyset, S_j = c_j^- \right] \geq u \right\}$ .

(ii) *If  $Y$  is a binary random variable, then we have the following bounds on  $\mathbb{E}[Y(j) | Q_j = (j,k), S_j = c_j]$  and  $\mathbb{E}[Y(k) | Q_j = (j,k), S_j = c_j]$ :*

$$\begin{aligned} & \max \left\{ 1 - \frac{1}{\underline{\delta}_{j,k}^+} \mathbb{P} \left[ Y = 0 \mid \mathbf{Q}_j \cap \{(j,k)\} \neq \emptyset, S_j = c_j^+ \right], 0 \right\} \\ & \leq \mathbb{E} [Y(j) | Q_j = (j,k), S_j = c_j] \leq \\ & \min \left\{ \frac{1}{\underline{\delta}_{j,k}^+} \mathbb{P} \left[ Y = 1 \mid \mathbf{Q}_j \cap \{(j,k)\} \neq \emptyset, S_j = c_j^+ \right], 1 \right\}, \end{aligned}$$

and

$$\begin{aligned} & \max \left\{ 1 - \frac{1}{\underline{\delta}_{j,k}^-} \mathbb{P} \left[ Y = 0 \mid \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, S_j = c_j^- \right], 0 \right\} \\ & \leq \mathbb{E} [Y(k) \mid Q_j = (j, k), S_j = c_j] \leq \\ & \min \left\{ \frac{1}{\underline{\delta}_{j,k}^-} \mathbb{P} \left[ Y = 1 \mid \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, S_j = c_j^- \right], 1 \right\}. \end{aligned}$$

The bounds in Proposition 4 build on work by Horowitz and Manski (1995). To see the intuition, take part (i) of the proposition, make  $g(Y) = Y$ , and focus on individuals just above the cutoff. Among all individuals in the sub-population with  $\mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset$  and  $S_j = c_j^+$  there is  $\delta_{j,k}^+$  of them that have  $Q_j = (j, k)$  and  $Y = Y(j)$  by the cutoff characterization. We don't know who these individuals are among those in the sub-population. However, the lowest possible value for  $\mathbb{E}[Y(j) \mid Q_j = (j, k), S_j = c_j]$  occurs if all such individuals are located at the lower tail of the distribution of outcomes in the sub-population. Likewise, the highest possible value for  $\mathbb{E}[Y(j) \mid Q_j = (j, k), S_j = c_j]$  occurs if that same fraction of individuals are all located at the upper tail of the distribution of outcomes. We don't know  $\delta_{j,k}^+$ , but we do know that it is no smaller than  $\underline{\delta}_{j,k}^+ > 0$ . The bounds only get wider as the fraction  $\delta_{j,k}^+$  decreases, so the bounds evaluated at  $\delta_{j,k}^+ = \underline{\delta}_{j,k}^+$  take into account all possible values for  $\delta_{j,k}^+$ . The expressions for the bounds in the case that  $g(Y) = Y$  is binary change relatively to  $g(Y)$  continuous, but the derivation of the bounds follows the same intuition. We refer the reader to the proof in the appendix for the details (Section A.8).

Although intuitive, these bounds are not necessarily sharp because  $\underline{\delta}_{j,k}^+$  is not exogenously given as in Horowitz and Manski (1995);  $\underline{\delta}_{j,k}^+$  is constructed using only the distribution of  $Q_j$ . The complete characterization of sharp bounds is complex because it involves bounds on the joint distribution of potential outcomes and  $Q_j$  and may not be practical to implement when potential outcomes are continuous. For the sake of simplicity, we relegate the sharp characterization to Section B in the appendix. The bounds of Proposition 4 contain the sharp bounds of Section B, as long as our model assumptions are true. The lack of sharpness may not matter in practice when Proposition 4 yields tight bounds for a given dataset. However, if our assumptions are not true, the bounds of Proposition 4 may not contain the sharp bounds of Section B. That being said, we may obtain tight bounds from the data using Proposition 4 but that does not mean they contain the true parameter (Kédagni et al., 2020). Therefore, it is advisable to assess the testable implications of Corollary 2 as a matter of routine.

## 5 Assignment to College Majors in Chile

In this section, we illustrate our method using data from college applications in Chile. Before estimating bounds for the effect of assignment to a program on graduation outcomes, we exemplify the presence of strategic behavior in this setting.



## 5.1 Data, Institutional Setting, and Evidence of Strategic Behavior.

**Centralized college application in Chile.** We use publicly available data on centralized college applications and assignments in Chile from 2004 to 2010, as well as graduation from 2007 to 2020. The institutional setting is described in detail by Hastings et al. (2013); Larroucau and Rios (2020, 2021), among others. College choice in Chile is organized as a semi-centralized system —a subset of universities participate in a centralized market in which a clearinghouse collects ranked order lists from applicants and determines assignments using a variant of the DA algorithm. Students can submit ranked order lists of up to eight major-university pairs (“programs”) out of more than 1,000.<sup>6</sup> Priorities are program-specific and determined by a weighted average of scores obtained at a national standardized test (“PSU”, for “Prueba de Selección Universitaria”) and of high-school GPA.<sup>7</sup> Descriptive statistics about the sample of students and programs are shown in the left panel of Tables 2 and 3 in Appendix C.

**Strategic behavior.** The fact that Chilean college applicants behave strategically has been thoroughly documented by Larroucau and Rios (2020, 2021). Using a 2014 survey linked to administrative data on applications, they show that listed programs often do not coincide with the truly preferred programs explicitly elicited by the survey. Focusing on application to Medicine programs, they find for instance that “among the 40,000 students who answered the survey, close to 10% (3,797) reported Medicine as their top preference, and 2,987 of these students ended-up applying to the system. Among these, only 1,360 listed Medicine as their top preference.” Moreover, students’ probability of not including Medicine at the top of their submitted list (while declaring it as most preferred in the survey) increases as application scores decrease; and the probability to apply to Medicine drops after the 600–750 application score range, where most Medicine cutoffs lie. This suggests that students tend to omit Medicine as their admission chances get lower, despite preferring it over other programs. These findings are confirmed in a 2019 survey, where they find that “most students who did not include their first true preference in their application list expected a higher cutoff than the cutoff for their first listed preference.”

We construct Figure 1 to provide additional evidence of strategic behavior. Consider a program  $j$  with cutoff  $c_j$ . Suppose, (consistently with the general finding of the empirical literature on college choice), that students tend to prefer programs of higher quality. If applicants behave strategically, one would expect applications to program  $j$  to peak among students with application score close to  $c_j$ . If cutoffs tend to remain in the same neighborhood across years, students with application scores much higher than  $c_j$  can expect to be admissible to more selective, higher-quality programs than  $j$ , which they prefer over  $j$ . Hence, we expect very few of these students to include  $j$  in their list. As application scores get lower and closer to  $c_j$ , students’ admission chances to the most selective programs decrease, and program  $j$  becomes one of the most selective (desirable) program among

---

<sup>6</sup>The quota was increased to ten in 2012.

<sup>7</sup>In 2014, students’ relative rank within their high school has been added as one of the “primary” scores to be averaged to construct priorities.

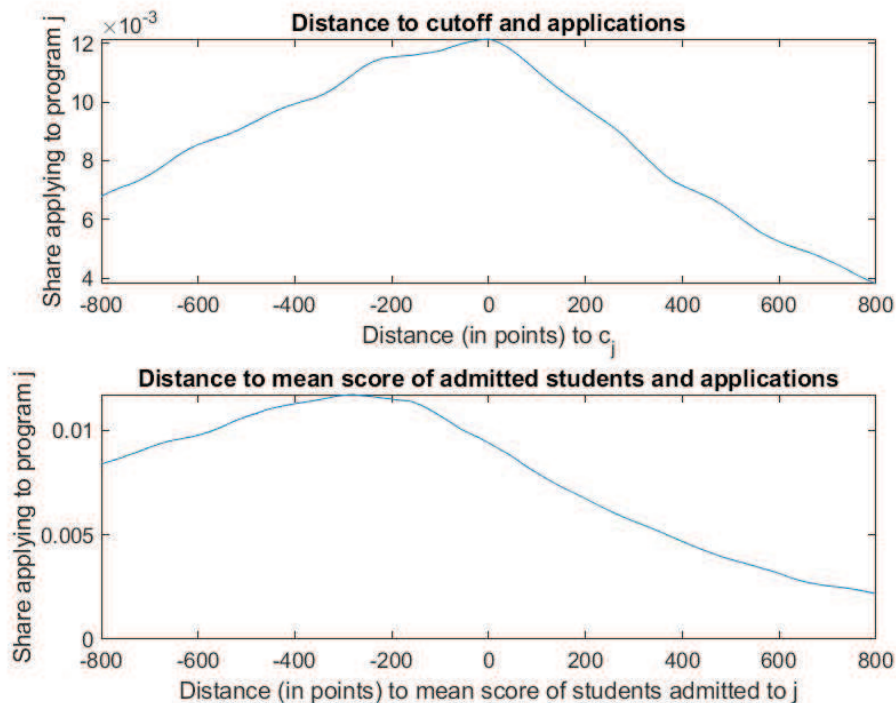
those they still have a high admission probability. Hence, we expect applications to  $c_j$  to increase as application scores decrease and get closer to  $c_j$ . As application scores decrease below  $c_j$ , students realize that their admission probability to program  $j$  gets lower, and while program  $j$  remains a relatively desirable (selective) alternative, we expect these expectations to drive applications down. This application pattern, expected if students behave strategically, is exactly what we observe on the top panel of Figure 1. Pooling all programs  $j$  together, the top panel of Figure 1 shows the fraction of students listing program  $j$  in their rank-order lists (ROL or  $P$  in terms of our notation), as a function of the distance between their application for program  $j$  and the cutoff  $c_j$ .

It may be difficult to disentangle the role of preferences from the role of expectations about admission probabilities when both may enter students' choice of which programs to include in their ROLs (Manski (2004); Agarwal and Somaini (2018)). The pattern observed in the top panel of Figure 1 could, alternatively, be consistent with students not behaving strategically but preferring programs that “are a good fit in terms of quality”, that is, programs in which their skill level would be close to the average skill level. If that is the case, applications should peak among students whose skills (proxied by application score) are close to the mean skill level in the program. This is not what we observe in the bottom panel of Figure 1. Pooling all programs  $j$  together, the bottom panel of Figure 1 shows the fraction of students including program  $j$  in their list, as a function of the distance between their application for program  $j$  and the mean application score among students admitted to  $j$ . Conditional on application score, the share of students applying to a program  $j$  is not the highest for application scores close to the mean score among students admitted to  $j$ . It reaches its highest point well below this level, showing that students do not systematically prefer programs in which they would be the “average” student. This further confirms the hypothesis that students behave strategically.

## 5.2 Results

Regardless of strategic behavior, the aggregation of treatment effects across different programs brings additional challenges not tackled in this paper; therefore, to more clearly illustrate the importance of accounting for strategic behavior, we focus on a single program. We are interested in identifying the effects of assignment to a given post-secondary program on college graduation. College returns are typically thought of as tied to college graduation, motivating our focus on graduation-related outcomes (see Kirkeboen et al. (2016) and Altonji et al. (2016) for a detailed argument). For the average college program, initial enrollment and eventual graduation are far from being perfectly correlated (OECD 2019, Larroucau and Rios (2021)). We first look at whether assignment to a given program increases one's probability to graduate from that same program. Students who don't graduate from their initially assigned program either drop out entirely or participate again in the application procedure in a subsequent year, and possibly go on to graduate from another program. Therefore, we also consider outcomes such reapplication and graduation from any program to investigate alternative trajectories to graduation from the assigned program.

Figure 1: Applications to a program peak among applicants with score close to the cutoff



Finally, we consider graduation from a STEM program and graduation from a top university as other outcomes. This is in line with the literature on returns to education and college choice, which focuses on the role of field of study and institution quality in driving returns (see Kirkeboen et al. (2016) and Altonji et al. (2016) for more references). We follow the classification of the OECD and define as STEM programs those pertaining to the physical sciences, life sciences, engineering, and mathematics. This for instance includes chemistry and biology, but excludes economics, medicine, or logistics. We define as “top” universities Pontificia Universidad Católica de Chile (PUC) and Universidad de Chile (UCHile).

As noted earlier, the identification proposed in this paper resembles an RD design that is sharp, and is therefore well suited to study the effects of assignment to a program on graduation-related outcomes. In contrast, estimating the effects of graduation from a program on earnings typically involves a fuzzy design, which we study in separate ongoing work.

In practice, we use Bachelor of Arts (BA) in Business Administration (“Licencia, Ingeniería Comercial”) at PUC in Santiago as our illustrative program  $j$ . The choice of this program is driven by two considerations. First, to get reasonably precise estimates, we need a sufficient number of observations, and this program is popular. Second, with regard to the outcomes above, assignment to this program is interesting in itself; the chosen program is a very popular program from a selective university, with a field typically of interest for both STEM-oriented and non STEM-oriented students, making it particularly interesting to look at the outcomes above. Descriptive

statistics about the program and its applicants are shown in the right panel of Tables 2 and 3 in Appendix C.

The first step of the exercise is to recover  $Q_j$  for each student. Absent strategic behavior, next-best alternatives are straightforwardly given by application lists. In other terms, direct observation of  $P$  and placement scores lead us to compute cutoffs, budget sets, and finally  $P_j$ , which equals  $Q_j$  under truth telling. When accounting for strategic behavior, we need to construct  $\mathbf{Q}_j$ , a set that contains the true  $Q_j$  of each individual. Section 4.1 shows how to construct  $Q_j$  under the Strong Partial Order (SPO) or the Weak Partial Order (WPO) assumptions. The WPO assumption implies that any submitted list is a subset of acceptable programs ordered just as in the student’s true preferences. The SPO assumption implies, in addition, that any student submitting a list strictly shorter than permitted reveals the student’s complete true list of acceptable schools. Under the SPO assumption, true local preferences  $Q_j$  are observed for students submitting fewer than eight choices (in the case of our empirical application). For students ranking eight choices,  $Q_j$  is only observed if  $\mathbf{Q}_j$  is a singleton. The SPO assumption is a strong assumption, especially in light of the argument in Larroucau and Rios (2020). Deriving bounds under the SPO assumption is, however, sufficient to illustrate the importance of our method relative to the naive approach of comparing mean outcomes of marginal applicants by simply conditioning on  $P_j$ , that is, the reported local preferences.

Results are shown in Table 1.<sup>8</sup> For simplicity and clarity, we focus on just the four most frequent second-best choices of  $Q_j$  among applicants to our program  $j$ ; they are shown in the first column of Table 1.

First, it is worth noting that estimated bounds are relatively informative. The width of the estimated set is smaller than .1 in eleven out of twenty cases and exceeds .3 only once. The bounds identify the sign of the effect in sixteen out of twenty cases. Naive estimates fall within the bounds except in one case, but the sign and magnitude of naive estimates may disagree with those of the true effect in cases where bounds are wide and uninformative about the sign. In our context, the narrowness of bounds is little surprising, given the SPO assumption and the large share of students ranking fewer than eight choices. Interestingly, the narrowness does not depend on whether we impose the UMAS assumption. Reassuringly, estimates show that admission to the case-study program, Business Administration at PUC Santiago, increases students’ probability to graduate from that program. The effect is particularly strong for those whose next-preferred alternative is Economics at UChile or Social Sciences and Humanities at PUC Santiago.

Admission to our case-study program also strongly decreases students’ probability to reapply through the centralized system in subsequent years for students whose local next-preferred option is Science and Engineering at UChile, Economics at UChile, or Social Sciences and Humanities at PUC Santiago. Interestingly, for students with the latter two next-best alternatives, it also decreases the probability to graduate from any program. For students whose next-best alternative is Science &

---

<sup>8</sup>Inference is beyond the scope of this paper. Besides the fact that admission cutoffs are estimated, the fact that the  $\delta$ ’s used in the construction of bounds are partially identified needs to be taken into account when deriving confidence bands —a task that we leave for future research.

Engineering at UChile, the naive estimate indicates a small negative effect on graduation from any program while our bounds show that this effect cannot be signed and may be 50 times larger in magnitude. Bounds do suggest that admission to Business Administration at PUC Santiago has a negative effect on reapplication for these students. While we generally see magnitudes of 0.1 or 0.2 in the effects on graduation from Business Administration or any program, we see much smaller magnitudes on the students' probability to graduate from a STEM program—which is reassuring given that our definition of STEM does not include Business Administration. Finally, we find a strong positive effect on the probability to graduate from a top institution for those whose local next-best alternative is Law or Social Sciences and Humanities at PUC Santiago.

## 6 Conclusion

Centralized mechanisms for the assignment of students to educational programs are growing in popularity across the world. These systems provide a valuable source of exogenous variation through discontinuities around admission cutoffs. This along with individual-level data on admissions, preferences, and future outcomes allow researchers to identify a wide range of causal effects of education on outcomes. The variation is useful for identification as long as we control for students' true preferences; however, true preferences generally differ from what students report to the mechanism given that most implementations of mechanisms generate incentives for students to behave strategically.

This paper provides a novel approach to partially identify effects from mechanism assignment on future outcomes that is robust to strategic behavior. We illustrate our approach using data from a deferred-acceptance mechanism that assigns roughly 70,000 students to more than 1,000 university-major programs every year in Chile. First, we find substantial evidence of strategic behavior which confirms earlier findings from the literature on Chile. Second, we compute bounds on the average effect of program assignment on graduation outcomes. We do so for students whose scores are marginal to the cutoff of admission into one of the most popular and selective programs in the system: business administration at the Catholic University of Santiago. Although admission into this program increases the likelihood of graduating from that same program, its effect on the likelihood of ever graduating from university varies depending on what students prefer as their second-best option. We compare our bounds to naive estimates that simply control for reported preferences as if they were true preferences. Naive estimates generally fall within our bounds but may lead to erroneous conclusions in some cases where bounds are wide and do not pin down the sign of the true effect.

Table 1: Results under SPO

	Estimate bounds	Naive estimate
<i>Graduation from Bus. Adm., PUC Santiago</i>		
Economics, UChile Santiago	[.319 ; .418]	.369
Science & Engineering, UChile Santiago	[.000 ; .188]	.159
Social Sciences & Humanities, PUC Santiago	[.230 ; .239]	.272
Law, PUC Santiago	[.087 ; .407]	.321
<i>Reapplication</i>		
Economics, UChile Santiago	[-.178 ; -.079]	-.114
Science & Engineering, UChile Santiago	[-.369 ; -.107]	-.199
Social Sciences & Humanities, PUC Santiago	[-.164 ; -.155]	-.156
Law, PUC Santiago	[.065 ; .174]	.116
<i>Graduation from any program</i>		
Economics, UChile Santiago	[-.218 ; -.029]	-.084
Science & Engineering, UChile Santiago	[-.126 ; .164]	-.003
Social Sciences & Humanities, PUC Santiago	[-.055 ; -.064]	-.063
Law, PUC Santiago	[.032 ; .139]	.058
<i>Graduation from STEM program</i>		
Economics, UChile Santiago	[-.016 ; .021]	.004
Science & Engineering, UChile Santiago	[.000 ; .076]	.064
Social Sciences & Humanities, PUC Santiago	[.000 ; .015]	.000
Law, PUC Santiago	[.000 ; .000]	.000
<i>Graduation from a top university (PUC Santiago or UChile)</i>		
Economics, UChile Santiago	[-.086 ; .013]	-.040
Science & Engineering, UChile Santiago	[-.058 ; .232]	.060
Social Sciences & Humanities, PUC Santiago	[.209 ; .218]	.210
Law, PUC Santiago	[.123 ; .411]	.193

Outcomes and second-best options  $k$  in  $Q_j = (j, k)$  are shown in the leftmost column of the table. For our parameter of interest, the first-best option in  $Q_j$  is always  $j = \text{“BA in Business administration at PUC Santiago”}$ . The second column shows bounds for our parameter of interest (Equation 1). Bounds are estimated using local linear polynomials, and without imposing UMAS (Assumption 5). Bandwidths are set to 30 points, except if another program  $\ell$  has its cutoff within 30 points of  $c_j$ . Then, the bandwidth is reduced to the distance between  $c_j$  and  $c_\ell$ . Naive estimates in the third column are computed as the difference in mean outcomes between applicants above vs. below the admission cutoff, within the same bandwidths, and conditional on reported local preferences  $P_j$  being equal to  $(j, k)$ , where  $k$  is given in the leftmost column of the table.

## References

- ABDULKADIROGLU, A., J. D. ANGRIST, Y. NARITA, AND P. PATHAK (2022): “Breaking Ties: Regression Discontinuity Design Meets Market Design,” *Econometrica*, 90, 117–151.
- ABDULKADIROGLU, A., Y.-K. CHE, AND Y. YASUDA (2015): “Expanding “Choice” in School Choice,” *American Economic Journal: Microeconomics*, 7, 1–42.

- AGARWAL, N. AND P. SOMAINI (2018): “Demand analysis using strategic reports: An application to a school choice mechanism,” *Econometrica*, 86, 391–444.
- ALTONJI, J., P. ARCIDIACONO, AND A. MAUREL (2016): “The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects,” *R. Hanushek, S. Machin, and L. Worrseman (eds) The Handbook of the Economics of Education*, V.
- AZEVEDO, E. M. AND E. BUDISH (2019): “Strategy-proofness in the Large,” *The Review of Economic Studies*, 86, 81–116.
- AZEVEDO, E. M. AND J. D. LESHNO (2016): “A Supply and Demand Framework for Two-sided Matching Markets,” *Journal of Political Economy*, 124, 1235–1268.
- BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2012): “Partial Identification Using Random Set Theory,” *Journal of Econometrics*, 166, 17–32.
- BERTANHA, M. (2020): “Regression Discontinuity Design with Many Thresholds,” *Journal of Econometrics*, 218, 216–241.
- CHEN, Y. AND T. SÖNMEZ (2006): “School Choice: An Experimental Study,” *Journal of Economic Theory*, 127, 202–231.
- CHESHER, A. AND A. M. ROSEN (2017): “Generalized Instrumental Variable Models,” *Econometrica*, 85, 959–989.
- DUBINS, L. E. AND D. A. FREEDMAN (1981): “Machiavelli and the Gale-Shapley Algorithm,” *American Mathematical Monthly*, 88, 485–494.
- ERGIN, H. I. (2002): “Efficient Resource Allocation on the Basis of Priorities,” *Econometrica*, 70, 2489–2497.
- ESTRADA, R. AND J. GIGNOUX (2017): “Benefits to Elite Schools and the Expected Returns to Education: Evidence from Mexico City,” *European Economic Review*, 95, 168–194.
- FACK, G., J. GRENET, AND Y. HE (2019): “Beyond Truth-telling: Preference Estimation with Centralized School Choice and College Admissions,” *American Economic Review*, 109, 1486–1529.
- HAERINGER, G. AND F. KLIJN (2008): “Constrained School Choice,” *Working paper version*.
- (2009): “Constrained School Choice,” *Journal of Economic Theory*, 144, 1921–1947.
- HASTINGS, J., C. NEILSON, AND S. ZIMMERMAN (2013): “Are Some Degrees Worth More Than Others? Evidence From College Admissions Cutoffs in Chile,” *NBER Working paper No. 19241*.
- HOROWITZ, J. L. AND C. F. MANSKI (1995): “Identification and Robustness with Contaminated and Corrupted Data,” *Econometrica: Journal of the Econometric Society*, 281–302.

- JACKSON, C. K. (2010): “Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago,” *The Economic Journal*, 120, 1399–1429.
- KÉDAGNI, D., L. LI, AND I. MOURIFIÉ (2020): “Discordant Relaxations of Misspecified Models,” *arXiv preprint arXiv:2012.11679*.
- KIRKEBOEN, L. J., E. LEUVEN, AND M. MOGSTAD (2016): “Field of Study, Earnings, and Self-selection,” *Quarterly Journal of Economics*, 131, 1057–1111.
- LARROUCAU, T. AND I. RIOS (2020): “Do “Short-List” Students Report Truthfully? Strategic Behavior in the Chilean College Admissions Problem,” *Working paper*.
- (2021): “Dynamic College Admissions,” *Working paper*.
- MANSKI, C. (2004): “Measuring Expectations,” *Econometrica*, 72, 1329–1376.
- MOLINARI, F. (2020): “Microeconometrics with Partial Identification,” in *Handbook of Econometrics, Volume 7A*, ed. by S. N. Durlauf, L. P. Hansen, J. J. Heckman, and R. L. Matzkin, Elsevier, vol. 7 of *Handbook of Econometrics*, chap. 5, 355–486.
- PAIS, J. AND Á. PINTÉR (2008): “School Choice and Information: An Experimental Study on Matching Mechanisms,” *Games and Economic Behavior*, 64, 303–328.
- ROTH, A. E. (1982): “The Economics of Matching: Stability and Incentives,” *Mathematics of Operations Research*, 7, 617–628.



# A Proofs

## A.1 Proof of Lemma 1

Consider a school  $j$  with cutoff  $c_j$  in the interior of the support  $\mathcal{S}_j$ . Once you fix an event  $A \in \mathbf{A}_{-j}$ , you fix the availability of those schools with non- $j$  scores  $\mathbf{S}_{-j}$ . The right- and left-counterfactual budget sets  $B_j^-(\mathbf{S})$  and  $B_j^+(\mathbf{S})$  become fixed (i.e., non random), regardless of the value of  $S_j$ . Once we fix  $S_j$  on top of that, the actual budget set  $B(\mathbf{S})$  is also fixed.

Fix an event  $A \in \mathbf{A}_{-j}$  and  $S_j$  in a small neighborhood of  $c_j$ . Pick schools  $k, l \in \mathcal{J}^0$  such that  $\mathbb{P}[Q_j = (k, l) | S_j = c_j] > 0$ . Define  $\mathbb{Q}(A)$  to be the set of all preference relations  $Q \in \mathcal{Q}$  such that  $Q(B_j^+(\mathbf{S})) = k$  and  $Q(B_j^-(\mathbf{S})) = l$ . Note that  $\mathbb{Q}(A)$  is a fixed set of preferences (i.e., non random).

Enumerate the mutually exclusive events in  $\mathbf{A}_{-j}$  as  $A_1, \dots, A_M$ , where  $M = 2^{J-1}$ . We have that  $Q_j = (k, l)$  is equivalent to  $Q \in \mathbb{Q}(A_1)$  if  $A_1, \dots, Q \in \mathbb{Q}(A_M)$  if  $A_M$ . Therefore, for  $s$  in a small neighborhood of  $c_j$ ,

$$\begin{aligned} & \mathbb{E}[g(Y(d))\mathbb{I}\{Q_j = (k, l)\} | S_j = s] \\ &= \sum_{l=1}^M \mathbb{E}[g(Y(d))\mathbb{I}\{Q_j = (k, l), \mathbf{S}_{-j} \in A_l\} | S_j = s] \\ &= \sum_{l=1}^M \mathbb{E}[g(Y(d))\mathbb{I}\{Q \in \mathbb{Q}(A_l), \mathbf{S}_{-j} \in A_l\} | S_j = s] \\ &= \sum_{l=1}^M \sum_{Q_0 \in \mathbb{Q}(A_l)} \mathbb{E}[g(Y(d))\mathbb{I}\{Q = Q_0, \mathbf{S}_{-j} \in A_l\} | S_j = s]. \end{aligned}$$

It follows that  $\mathbb{E}[g(Y(d))\mathbb{I}\{Q_j = (k, l)\} | S_j = s]$  is a continuous function of  $s$  at  $s = c_j$  because  $\mathbb{E}[g(Y(d))\mathbb{I}\{Q = Q_0, \mathbf{S}_{-j} \in A_l\} | S_j = s]$  is continuous at  $s = c_j$  for every  $l$  and  $Q_0$  by assumption.

Likewise,

$$\begin{aligned} & \mathbb{P}[Q_j = (k, l) | S_j = s] \\ &= \sum_{l=1}^M \mathbb{P}[Q_j = (k, l), \mathbf{S}_{-j} \in A_l | S_j = s] \\ &= \sum_{l=1}^M \mathbb{P}[Q \in \mathbb{Q}(A_l), \mathbf{S}_{-j} \in A_l | S_j = s] \\ &= \sum_{l=1}^M \sum_{Q_0 \in \mathbb{Q}(A_l)} \mathbb{P}[Q = Q_0, \mathbf{S}_{-j} \in A_l | S_j = s], \end{aligned}$$

which is continuous at  $s = c_j$  for every  $l$  and  $Q_0$  by assumption.

Therefore,

$$\begin{aligned} & \mathbb{E}[g(Y(d)) | Q_j = (k, l), S_j = s] \\ &= \frac{\mathbb{E}[g(Y(d))\mathbb{I}\{Q_j = (k, l)\} | S_j = s]}{\mathbb{P}[Q_j = (k, l) | S_j = s]} \end{aligned}$$

is continuous at  $s = c_j$

□

## A.2 Proof of Proposition 1

Take  $(j, k) \in \mathcal{P}$ . Start with school  $j$ ,  $s \geq c_j$ ,

$$\begin{aligned} \lim_{s \downarrow c_j} \mathbb{E}[g(Y)|P_j = (j, k), S_j = s] &= \lim_{s \downarrow c_j} \mathbb{E}[g(Y)|Q_j = (j, k), S_j = s] \\ &= \lim_{s \downarrow c_j} \mathbb{E}[g(Y(j))|Q_j = (j, k), S_j = s] \\ &= \mathbb{E}[g(Y(j))|Q_j = (j, k), S_j = c_j]. \end{aligned}$$

For school  $k$ ,  $s < c_j$ ,

$$\begin{aligned} \lim_{s \uparrow c_j} \mathbb{E}[g(Y)|P_j = (j, k), S_j = s] &= \lim_{s \uparrow c_j} \mathbb{E}[g(Y)|Q_j = (j, k), S_j = s] \\ &= \lim_{s \uparrow c_j} \mathbb{E}[g(Y(k))|Q_j = (j, k), S_j = s] \\ &= \mathbb{E}[g(Y(k))|Q_j = (j, k), S_j = c_j]. \end{aligned}$$

□

## A.3 Proof that Partial Orders Dominate Non-Partial Orders

First, some definitions. The strategy for each student  $\omega$  is an ordered list of schools  $P(\omega) \subseteq \mathcal{J}$  that has at least one and at most  $K < J$  schools in it. The strategy profile of the economy is a correspondence  $P : \Omega \rightrightarrows \mathcal{J}$  (random set). The score profile of individuals in the economy is denoted by the random vector  $\mathbf{S} : \Omega \rightarrow \mathcal{S}$ . A mechanism  $\varphi$  takes the whole correspondence  $P$  and function  $\mathbf{S}$  as givens and produces school assignments for each individual  $\omega \in \Omega$ , such that  $\varphi(P, \mathbf{S}) : \Omega \rightarrow \mathcal{J}^0$ . The assignment of student  $\omega$  for profiles  $(P, \mathbf{S})$  is  $\varphi(P, \mathbf{S})[\omega]$ .

For this proof, it is convenient to focus on the assignment of an individual  $\omega_0$  as a function of his individual ranking submission  $p \subseteq \mathcal{J}$ , the ranking submissions of others  $P^{-\omega_0} : \Omega \setminus \{\omega_0\} \rightrightarrows \mathcal{J}$ , and the scores of everybody  $\mathbf{S} : \Omega \rightarrow \mathcal{S}$ . We write that assignment as  $\varphi((p, P^{-\omega_0}), \mathbf{S})[\omega_0]$ .

For individual  $\omega_0$  with true preferences  $Q(\omega_0)$ , we say  $p'$  **weakly dominates**  $p$  if  $\varphi((p', P^{-\omega_0}), \mathbf{S})[\omega_0] \bar{Q}(\omega_0) \varphi((p, P^{-\omega_0}), \mathbf{S})[\omega_0]$  for every  $P^{-\omega_0}$ . In addition, we say  $\varphi$  is strategy proof with unrestricted lists if submitting the true list of acceptable schools weakly dominates submitting anything else for every individual  $\omega$ .

**Lemma 2.** *Assume  $\varphi$  is strategy proof with unrestricted lists. Consider student  $\omega$  with true preference  $Q(\omega)$ . Fix an arbitrary ranking of schools  $p \subseteq \mathcal{J}$ . For such student,  $p$  is weakly dominated by any weak partial order  $p'$  of  $Q(\omega)$  that contains all the acceptable schools in  $p$ . In turn,  $p'$  is weakly dominated by any strong partial order  $p''$  of  $Q(\omega)$  that contains all the acceptable schools in  $p'$ .*

*Moreover, suppose the number of acceptable schools in  $Q(\omega)$  is less than or equal to  $K$ . Then, the dominant strategy is to submit the unique strong partial order of  $Q(\omega)$  which equals the true list of acceptable schools.*

### Proof of Lemma 2:

This proof is due to Haeringer and Klijn (2008), Lemma 4.2. We expand it here in terms of our framework and definitions of partial order.

From the main text, everybody has at least one acceptable school. Take student  $\omega$  and consider an arbitrary list of schools  $p$  that has at least one acceptable school for that student. If  $p$  does not have any acceptable schools, then it is clearly dominated by any weak partial order.

First, remove the non-acceptable schools from  $p$  (if any), keep the relative ordering of the acceptable schools, and call the resulting list  $\bar{p}$ . It follows that  $\bar{p}$  weakly dominates  $p$  for student  $\omega$ .

Second, let  $p'$  be a weak partial order of  $Q(\omega)$  that contains the schools listed in  $\bar{p}$ . Construct a new “true” preference ranking  $q \subseteq \mathcal{J}^0$  as follows: (a) take the acceptable schools of  $p'$  and place them first in  $q$ , in the same order as they are in  $p'$ ; (b) add a 0 to  $q$  after the last school in part (a); (c) fill the remaining positions below 0 in  $q$  with the schools not listed in  $p'$ , in any order. Note that  $p'$  equals the true list of acceptable schools from  $q$ ,  $p'$  is a weak partial order of  $q$ , and  $p'$  is the unique strong partial order of  $q$ .

Third, suppose for a moment that the true preference of individual  $\omega$  were  $q$  instead of  $Q(\omega)$ . In that case, strategy-proofness of  $\varphi$  implies that

$$\varphi((p', P^{-\omega}), \mathbf{S})[\omega] \bar{q} \varphi((\bar{p}, P^{-\omega}), \mathbf{S})[\omega] \text{ for every } P^{-\omega}.$$

Given that  $p'$  is a weak partial order of both  $Q(\omega)$  and  $q$ , for any two options  $d, d'$  in  $p'$ , we have  $d' \bar{q} d$  implies  $d' \bar{Q}(\omega) d$ . Therefore,

$$\varphi((p', P^{-\omega}), \mathbf{S})[\omega] \bar{Q}(\omega) \varphi((\bar{p}, P^{-\omega}), \mathbf{S})[\omega] \text{ for every } P^{-\omega}.$$

It then follows that  $p'$  weakly dominates  $\bar{p}$ , which weakly dominates  $p$ , so  $p'$  weakly dominates  $p$ .

It follows that any strong partial order  $p''$  of  $Q(\omega)$  that contains the same acceptable schools as  $p'$  weakly dominates  $p'$ . To see that, repeat the argument above by replacing  $p$  with  $p'$  and replacing  $p'$  with  $p''$ .

The second claim of the lemma follows from strategy-proofness of  $\varphi$  since the quota constraint is not binding for an individual whose number of acceptable schools is less than or equal to  $K$ , and submission of the true list of acceptable schools is feasible.

□

## A.4 Proof of Proposition 2

The true preference list  $Q$  is unobserved. The submitted preference list  $P$  is observed and a weak partial order of  $Q$ . Note that all the schools listed in  $P$  appear in  $Q$  ranked before 0 (i.e., as acceptable schools). There may be other elements in  $Q$  that were not listed in  $P$ . These remaining schools might appear anywhere in  $Q$ , as long as the relative ordering of schools in  $P$  is preserved in  $Q$ . Our focus is on students with  $P_j = (a, b)$ , so we know that  $a \bar{Q} b$ , where  $a \bar{Q} b$  if  $a \neq b$ .

We consider all possibilities of  $Q$  that are consistent with the observed  $P$  and assumptions, and that affect  $Q_j$ . In such cases, the acceptable schools in  $Q$  include all the schools in  $P$  and possibly more; in fact, since  $|P| \leq K < J$ , the  $J - |P| > 0$  unlisted schools in  $P$  may appear as acceptable in  $Q$ . The additional acceptable schools in  $Q$  may be schools that are feasible or not within the budget set of the individual.

The proposition defines two sets of unlisted feasible schools for an individual with scores  $\mathbf{S}$  and submitted preferences  $P$ :  $N_j^+ = B_j^+(\mathbf{S}) \setminus \{P \cup \{0\}\}$  and  $N_j^- = B_j^-(\mathbf{S}) \setminus \{P \cup \{0\}\}$ . The only difference between  $B_j^+$  and  $B_j^-$  is the set of schools whose priority scores equal  $S_j$  and cutoffs equal  $c_j$ .

The rest of the proof builds on the following reasoning. Ignore Assumption 4 for a moment. The first coordinate of  $Q_j$  depends on  $P_j$  and the set  $N_j^+$ . In fact, the first coordinate of  $P_j$  is the best option in  $B_j^+(\mathbf{S})$  according to  $P$ : that's option  $a$ . The best option in  $B_j^+(\mathbf{S})$  according to  $Q$  may also be  $a$  as long as there are no unlisted options in  $P$  that are available in  $B_j^+(\mathbf{S})$  and rank higher than  $a$  in  $Q$ . A similar

argument applies to the second coordinate of  $Q$ : that is a function of  $P_j$  and the set  $N_j^-$ . Now, Assumption 4 further restricts  $Q_j$  because it implies that  $P(B(\mathbf{S})) = Q(B(\mathbf{S}))$  and  $P(B(\mathbf{S}))$  is observed.

Regarding the outside option 0, the only way that  $Q_j$  will have a zero is if  $P_j$  has a zero. In fact, if  $a \neq 0$  and  $b \neq 0$ , we have that  $a\bar{P}bP0$  which implies  $a\bar{Q}bQ0$ , so none of the coordinates of  $Q_j$  will be zero. That's why the sets of unlisted feasible options,  $N_j^+$  and  $N_j^-$ , do not contain zero.

**Case 1:**  $S_j \geq c_j$ .

By Assumption 4 we have that the mechanism assignment  $\mu$  equals the best option according to  $P$  in the set  $B(\mathbf{S}) = B_j^+(\mathbf{S})$ . An individual with  $P_j = (a, b)$  has  $a = P(B_j^+(\mathbf{S})) = P(B(\mathbf{S}))$ , therefore  $\mu = a$ . The cutoff characterization dictates that  $a = Q(B(\mathbf{S})) = Q(B_j^+(\mathbf{S}))$ , so the first coordinate of  $Q_j$  equals  $a$ . It remains to determine the second coordinate of  $Q_j$ , which depends on  $P_j$  and the set  $N_j^-$ .

**Case 1.1:**  $N_j^- = \emptyset$ .

None of the unlisted schools in  $P$  are feasible in the counterfactual below the cutoff. These unlisted schools may rank higher than  $b$  in  $Q$ , but none of them will ever be the best feasible option in the counterfactual below the cutoff. Thus, the 2nd coordinate of  $Q_j$  equals  $b$  and  $\mathbf{Q}_j = \{P_j\}$ .

**Case 1.2:**  $N_j^- \neq \emptyset$ .

For any option  $d \in N_j^-$ , we have that  $d \neq a$ ,  $d \neq b$ ,  $d \in N_j^+$ , and  $aQd$ .

**Case 1.2.1:** If  $a \neq b$ , we have  $aQb$ . We can always find a  $Q$  such that  $dQb$  and the second coordinate of  $Q_j$  equals  $d$ ; and we can always find another  $Q$  such that  $bQd$  and the second coordinate of  $Q_j$  equals  $b$ . Therefore,  $\mathbf{Q}_j = \{P_j\} \cup \{(a \times N_j^-)\}$ .

**Case 1.2.2:** If  $a = b$ , then we have  $bQd$  because  $aQd$ . Thus,  $\mathbf{Q}_j = \{P_j\}$ .

**Case 2:**  $S_j < c_j$ .

By Assumption 4 we have that the mechanism assignment  $\mu$  equals the best option according to  $P$  in the set  $B(\mathbf{S}) = B_j^-(\mathbf{S})$ . An individual with  $P_j = (a, b)$  has  $b = P(B_j^-(\mathbf{S})) = P(B(\mathbf{S}))$ , therefore  $\mu = b$ . The cutoff characterization dictates that  $b = Q(B(\mathbf{S})) = Q(B_j^-(\mathbf{S}))$ , so the second coordinate of  $Q_j$  equals  $b$ . It remains to determine the first coordinate of  $Q_j$ , which depends on  $P_j$  and the set  $N_j^+$ .

**Case 2.1:**  $N_j^+ = \emptyset$ .

None of the unlisted options in  $P$  are feasible in the counterfactual above the cutoff. These unlisted options may rank higher than  $a$  in  $Q$ , but none of them will ever be the best feasible option in the counterfactual above the cutoff. Thus, the first coordinate of  $Q_j$  equals  $a$  and  $\mathbf{Q}_j = \{P_j\}$ .

**Case 2.2:**  $N_j^+ \neq \emptyset$

For any option  $d \in N_j^+ \cap N_j^-$ , we have that  $d \neq a$ ,  $d \neq b$ ,  $d$  is in  $B_j^-(\mathbf{S})$  but  $bQd$  since the 2nd coordinate of  $Q_j$  equals  $b$ . We have that  $a\bar{Q}b$  and it follows that  $aQd$ . In this case, we can never find a  $Q$  such that the best choice in  $B_j^+(\mathbf{S})$  is  $d$ .

Consider there is an option  $d \in N_j^+ \setminus N_j^-$ . We have that  $d \neq a$ ,  $d \neq b$ ,  $d$  is in  $B_j^+(\mathbf{S})$  but not in  $B_j^-(\mathbf{S})$ .

It is possible to find  $Q$  such that  $dQa$ , in which case the best choice in  $B_j^+(\mathbf{S})$  is  $d$ . It is also possible to find another  $Q$  such that  $aQd$ , in which case the best choice in  $B_j^+(\mathbf{S})$  is  $a$ . Therefore,  $\mathbf{Q}_j = \{P_j\} \cup ((N_j^+ \setminus N_j^-) \times \{b\})$ .

Moreover, assume  $P$  is a strong partial order of  $Q$ . This implies that  $|P| = \min\{K, |\{d \in Q : dQ0\}|\}$  in addition to  $P$  being a subset of  $\{d \in Q : dQ0\}$ . If  $|P| < K$  then  $|P| = |\{d \in Q : dQ0\}|$  and  $P$  equals the true list of acceptable schools in  $Q$ . Therefore,  $Q_j = P_j$  and  $\mathbf{Q}_j = \{P_j\}$ . On the other hand, if  $|P| = K$ ,  $P$  may or may not be the true list of acceptable schools in  $Q$ , and  $\mathbf{Q}_j$  continues to be as defined in the case of weak partial order.

Finally, the proof above is constructive as it considers all possibilities of  $Q$  given  $P$  and  $\mathbf{S}$  that are consistent with the assumptions. Thus, it leads to the sharp set of possible  $Q_j$ s under the full support assumption.

□

## A.5 Proof of Proposition 3

Before the matching mechanism is run, the agent knows his placement scores  $\mathbf{s} = (s_1, \dots, s_J)$ , his true preferences  $Q$ , but does not know what the admission cutoffs will be after the matching is run. He sees admission cutoffs as random variables  $(C_1, \dots, C_J)$ . The strict preference relation  $Q$  is represented by a vector of distinct utility values  $(U_0, U_1, \dots, U_J)$  so that  $aQb \Leftrightarrow U_a > U_b$  for any  $a, b \in \mathcal{J}^0$ . We normalize  $U_0 = 0$  for simplicity.

The agent has to decide on a ranking of acceptable schools  $P$  to submit. The number of schools ranked in  $P$  is  $|P|$  and a feasible ranking has  $1 \leq |P| \leq K$ . The set of all feasible rankings is defined as  $\Delta P$ . We let  $P^u$  denote the  $u$ -th school listed in  $P$ , for  $u = 1, \dots, |P|$ . We define  $L_u^P$  to be the agent's expected probability of being assigned to school  $u$  when submitting ranking  $P$ ,  $u = 1, \dots, |P|$ .  $L_0^P$  denotes the expected probability of remaining unassigned, that is, of taking the outside option. Naturally,  $L_u^P \geq 0$  for every  $u$  and  $\sum_{u=0}^{|P|} L_u^P = 1$ . Cutoff characterization wrt  $P$  (Assumption 4) implies:

$$L_0^P = \mathbb{P} \left[ \bigcap_{v=1}^{|P|} \{s_{P^v} < C_{P^v}\} \right],$$

$$L_u^P = \mathbb{P} \left[ \bigcap_{v=1}^{u-1} \{s_{P^v} < C_{P^v}\} \cap \{s_{P^u} \geq C_{P^u}\} \right], \quad u = 1, \dots, |P|,$$

where we adopt the convention that  $\bigcap_{v=1}^{u-1} \{s_{P^v} < C_{P^v}\} \cap A = A$  for any measurable set  $A$  if  $u = 1$ ; in other words, if the intersection  $\bigcap_{v \in V}$  is to be computed over an empty set of indices,  $V = \emptyset$ , then  $\mathbb{P}[\bigcap_{v \in V} \{s_{P^v} < C_{P^v}\} \cap A] = \mathbb{P}[A]$ , for any measurable set  $A$ .

The agent's optimal ranking to be submitted is the solution to the following problem,

$$\max_{P \in \Delta P} \sum_{u=1}^{|P|} U_{P^u} L_u^P.$$

From now on, let  $P$  be the optimal choice of this agent. Suppose there is a pair of schools  $(d, e) \in UMAS \cap (P \times P^c)$ . Let  $n$  be the position in  $P$  where  $d$  appears, i.e.,  $P^n = d$ . Construct  $\tilde{P}$  by taking  $P$  and replacing option  $d$  with option  $e$ . This implies that  $\tilde{P}^u = P^u$  for every  $u \neq n$  and  $\tilde{P}^n = e$ .

Suppose  $eQd \Leftrightarrow U_e > U_d$  by contradiction. In what follows, we compare the expected utility of submitting  $P$  to the expected utility of submitting  $\tilde{P}$  for this agent and show a contradiction. In order to do that, we first establish some implications of Assumption 5 on the probabilities of admission:

$$L_u^P = L_u^{\tilde{P}} \text{ for any } u \text{ such that } 1 \leq u < n \text{ if } n > 1, \tag{6}$$

$$L_n^P \leq L_n^{\tilde{P}}, \tag{7}$$

$$L_u^P \geq L_u^{\tilde{P}} \text{ for any } u \text{ such that } n < u \leq |P| \text{ if } n < |P|, \tag{8}$$

$$L_0^P \geq L_0^{\tilde{P}}, \tag{9}$$

where at least one of the inequalities (7)–(9) is strict if  $n < |P|$ ; or, if  $n = |P|$ , at least one of the inequalities (7) and (9) is strict. Below, we prove (6)–(9).

Equation 6 comes from the fact that  $\tilde{P}^u = P^u$  for every  $1 \leq u < n$  if  $n > 1$ . The admission probability depends on  $u$  being the lowest ranked school for which the agent qualifies. Thus,

$$\mathbb{P} \left[ \bigcap_{v=1}^{u-1} \{s_{P^v} < C_{P^v}\} \cap \{s_{P^u} \geq C_{P^u}\} \right] = \mathbb{P} \left[ \bigcap_{v=1}^{u-1} \{s_{\tilde{P}^v} < C_{\tilde{P}^v}\} \cap \{s_{\tilde{P}^u} \geq C_{\tilde{P}^u}\} \right],$$

for every  $1 \leq u < n$  if  $n > 1$ .

For Equation 7, we have that  $\mathbb{P}[s_d \geq C_d] \leq \mathbb{P}[s_e \geq C_e]$  is implied by the first condition of Assumption 5. This further implies that

$$\mathbb{P} \left[ \bigcap_{v=1}^{n-1} \{s_{P^v} < C_{P^v}\} \cap \{s_d \geq C_d\} \right] \leq \mathbb{P} \left[ \bigcap_{v=1}^{n-1} \{s_{P^v} < C_{P^v}\} \cap \{s_e \geq C_e\} \right],$$

which is equivalent to  $L_n^P \leq L_n^{\tilde{P}}$ .

For Equation 8, note that the first condition of Assumption 5 implies  $\mathbb{P}[s_d < C_d] \geq \mathbb{P}[s_e < C_e]$ . This further implies that,

$$\begin{aligned} L_u^P &= \mathbb{P} \left[ \bigcap_{v=1}^{n-1} \{s_{P^v} < C_{P^v}\} \cap \{s_d < C_d\} \cap \bigcap_{v=n+1}^{u-1} \{s_{P^v} < C_{P^v}\} \cap \{s_{P^u} \geq C_{P^u}\} \right] \\ &\geq \\ &\mathbb{P} \left[ \bigcap_{v=1}^{n-1} \{s_{P^v} < C_{P^v}\} \cap \{s_e < C_e\} \cap \bigcap_{v=n+1}^{u-1} \{s_{P^v} < C_{P^v}\} \cap \{s_{P^u} \geq C_{P^u}\} \right] = L_u^{\tilde{P}}, \end{aligned}$$

for  $u$  such that  $n < u \leq |P|$  if  $n < |P|$ .

For Equation 9, again we have that  $\mathbb{P}[s_d < C_d] \geq \mathbb{P}[s_e < C_e]$  implies

$$\begin{aligned} L_0^P &= \mathbb{P} \left[ \bigcap_{v=1}^{n-1} \{s_{P^v} < C_{P^v}\} \cap \{s_d < C_d\} \cap \bigcap_{v=n+1}^{|P|} \{s_{P^v} < C_{P^v}\} \right] \\ &\geq \\ &\mathbb{P} \left[ \bigcap_{v=1}^{n-1} \{s_{P^v} < C_{P^v}\} \cap \{s_e < C_e\} \cap \bigcap_{v=n+1}^{|P|} \{s_{P^v} < C_{P^v}\} \right] = L_0^{\tilde{P}}. \end{aligned}$$

Finally, the second condition in Assumption 5 (relevance condition) implies that at least one of the inequalities (7)–(9) is strict if  $n < |P|$ ; or, if  $n = |P|$ , at least one of the inequalities (7) and (9) is strict. Having established these facts, we now move on to compare the expected utility of submitting  $P$  to the expected utility of submitting  $\tilde{P}$ .

Define  $\epsilon = U_e - U_d = U_{\tilde{P}^n} - U_{P^n} > 0$ . Note that  $L_0^P = 1 - \sum_{u=1}^{|P|} L_u^P$  and  $L_0^P - L_0^{\tilde{P}} = \sum_{u=1}^{|P|} (L_u^{\tilde{P}} - L_u^P)$  =  $(L_n^{\tilde{P}} - L_n^P) + \sum_{u=n+1}^{|P|} (L_u^{\tilde{P}} - L_u^P)$ , where we adopt the convention that a sum over an empty set of indices equals zero, i.e.,  $\sum_{u=n+1}^{|P|} (L_u^{\tilde{P}} - L_u^P) = 0$  if  $n+1 > |P|$ . This leads to  $(L_n^{\tilde{P}} - L_n^P) = (L_0^P - L_0^{\tilde{P}}) - \sum_{u=n+1}^{|P|} (L_u^{\tilde{P}} - L_u^P)$ . Next, we combine these definitions with the inequalities in Equations 7–9 to evaluate the difference between the expected utility of submitting  $P$  and the expected utility of submitting  $\tilde{P}$ :

$$\begin{aligned} &\sum_{u=1}^{|P|} U_{P^u} L_u^P - U_{\tilde{P}^u} L_u^{\tilde{P}} = U_{P^n} L_n^P - U_{\tilde{P}^n} L_n^{\tilde{P}} + \sum_{u=n+1}^{|P|} U_{P^u} (L_u^P - L_u^{\tilde{P}}) \\ &= U_{P^n} (L_n^P - L_n^{\tilde{P}}) - \epsilon L_n^{\tilde{P}} + \sum_{u=n+1}^{|P|} U_{P^u} (L_u^P - L_u^{\tilde{P}}) \\ &= U_{P^n} \left[ - (L_0^P - L_0^{\tilde{P}}) + \sum_{u=n+1}^{|P|} (L_u^{\tilde{P}} - L_u^P) \right] - \epsilon L_n^{\tilde{P}} + \sum_{u=n+1}^{|P|} U_{P^u} (L_u^P - L_u^{\tilde{P}}) \end{aligned}$$

$$\begin{aligned}
&= -U_{P^n} \left( L_0^P - L_0^{\tilde{P}} \right) - U_{P^n} \sum_{u=n+1}^{|P|} \left( L_u^P - L_u^{\tilde{P}} \right) - \epsilon L_n^{\tilde{P}} + \sum_{u=n+1}^{|P|} U_{P^u} \left( L_u^P - L_u^{\tilde{P}} \right) \\
&= -U_{P^n} \left( L_0^P - L_0^{\tilde{P}} \right) - \epsilon L_n^{\tilde{P}} - \sum_{u=n+1}^{|P|} (U_{P^n} - U_{P^u}) \left( L_u^P - L_u^{\tilde{P}} \right) < 0,
\end{aligned}$$

where we use the fact that weak partial order (Assumption 3) implies  $U_{P^1} > \dots > U_{P^{|P|}} > 0$ ,  $\epsilon > 0$ , and at least one of  $\left( L_0^P - L_0^{\tilde{P}} \right)$ ,  $L_n^{\tilde{P}}$ , and  $\left( L_u^P - L_u^{\tilde{P}} \right)$  for  $u > n$  is strictly positive if  $n < |P|$ . If  $n = |P|$ , at least one of  $\left( L_0^P - L_0^{\tilde{P}} \right)$  and  $L_n^{\tilde{P}}$  is strictly positive. The inequality above shows that submitting  $\tilde{P}$  increases the expected utility relatively to submitting  $P$ , which is a contradiction with  $P$  being the optimal choice. Therefore,  $dQe$ .

□

## A.6 Proof of Corollary 1

Let  $(d, e)$  be an arbitrary pair of distinct schools such that  $e$  is uniformly more accessible than school  $d$ ;  $d \in P$ ,  $e \notin P$ . Call  $(a, b) = P_j$ . Proposition 3 implies  $dQe$ . This fact weakly decreases the set of possible  $Q$ s each individual has and thus potentially affects only the non-singleton cases in the definition of  $\mathbf{Q}_j$  in Proposition 2. These correspond to cases 1.2.1 and 2.2 in the proof of Proposition 2. We re-examine these cases below for the arbitrary pair  $(d, e)$ .

**Case 1.2.1:**  $S_j \geq c_j$ ,  $N_j^- \neq \emptyset$ , and  $a \neq b$ .

We have  $aQb$ . For any option  $f \in N_j^-$ , we have that  $f \neq a$ ,  $f \neq b$ ,  $f \in B_j^-(\mathbf{S})$ , and  $f \in B_j^+(\mathbf{S})$ .

**Case 1.2.1(a):** Suppose  $b\bar{P}d$ .

$b\bar{P}d$  implies that  $b\bar{Q}d$ . Given that  $dQe$ , we have  $bQe$ . In case  $e \in N_j^-$ , it is no longer true that we can construct  $Q$  such that  $fQl$  for any  $f \in N_j^-$ . We can only do so for  $f \neq e$ . Therefore,  $Q_j = (a, e)$  does not belong to  $\mathbf{Q}_j$ .

**Case 1.2.1(b):** Suppose  $dPb$ .

This implies that  $dQb$ . The fact that  $dQe$  does not restrict us from having two possibilities:  $dQeQb$  or  $dQbQe$ . It is again possible to construct  $Q$  such that  $fQb$  for any  $f \in N_j^-$ , including  $f = e$  if  $e \in N_j^-$ . Therefore,  $Q_j = (a, e)$  does belong to  $\mathbf{Q}_j$  in this case.

**Case 2.2:**  $S_j < c_j$ ,  $N_j^+ \neq \emptyset$

Consider there is an option  $f \in N_j^+ \setminus N_j^-$ . We have that  $f \neq a$ ,  $f \neq b$ ,  $f$  is in  $B_j^+(\mathbf{S})$  but not in  $B_j^-(\mathbf{S})$ .

**Case 2.2(a):** Suppose  $a\bar{P}d$

$a\bar{P}d$  implies that  $a\bar{Q}d$ . Given that  $dQe$ , we have that  $aQe$ . If  $f = e$ , we cannot construct  $Q$  such that  $fQa$ . Thus, it is no longer true that we can find  $Q$  such that  $fQa$  for every  $f \in N_j^+ \setminus N_j^-$ ; that's true only for  $f \neq e$ . Therefore,  $Q_j = (e, b)$  does not belong to  $\mathbf{Q}_j$  in this case.

**Case 2.2(b):** Suppose  $dPa$ .

$dPa$  implies  $dQa$ . The fact that  $dQe$  does not restrict us from having two possibilities:  $dQeQa$  or  $dQaQe$ . It is again possible to construct  $Q$  such that  $fQa$  for every  $f \in N_j^+ \setminus N_j^-$ , including  $f = e$  if  $e \in N_j^+ \setminus N_j^-$ . Therefore,  $Q_j = (e, b)$  does belong to  $\mathbf{Q}_j$  in this case.

Sharpness follows because we remove all  $Q$ s that violate the implication of Proposition 3.

□

## A.7 Proof of Proposition 4

The random set  $\mathbf{Q}_j$  is a measurable map from  $\Omega$  to  $\mathcal{J}^0 \times \mathcal{J}^0$ , so it is compact valued. Following Definition A.1 by Molinari (2020), we say that  $\mathbf{Q}_j$  is a random closed set because for every compact set  $\mathcal{K} \in \mathbb{R}^2$ , the set  $\{\omega \in \Omega : \mathbf{Q}_j(\omega) \cap \mathcal{K} \neq \emptyset\}$  is a measurable event. By Assumption 6(i), the random vector  $Q_j$  and the random set  $\mathbf{Q}_j$  are measurable maps on the same probability space; and  $\mathbb{P}[Q_j \in \mathbf{Q}_j | S_j = s] = 1$  for any  $s \in \mathcal{S}_j$ .

Artstein's inequality (Theorem A.1 by Molinari (2020)) characterizes all possible probability mass functions  $\mathbb{P}[Q_j = (a, b) | S_j = s]$  over  $(a, b) \in \mathcal{J}^0 \times \mathcal{J}^0$  for any  $Q_j$  such that  $\mathbb{P}[Q_j \in \mathbf{Q}_j | S_j = s] = 1$  for any  $s \in \mathcal{S}_j$ . Since  $\mathbf{Q}_j$  is sharp (Assumption 6(ii)), the Artstein's inequality yields the sharp set of all probability mass functions  $\mathbb{P}[Q_j = (a, b) | S_j = s]$ . For any  $s \in \mathcal{S}_j$ , the inequality says that

$$\mathbb{P}[\mathbf{Q}_j \subseteq A | S_j = s] \leq \mathbb{P}[Q_j \in A | S_j = s] \quad \forall A \in 2^{\mathcal{J}^0 \times \mathcal{J}^0}, \quad (10)$$

where  $2^{\mathcal{J}^0 \times \mathcal{J}^0}$  denotes the power set of  $\mathcal{J}^0 \times \mathcal{J}^0$ .

Lemma 1 of Chesher and Rosen (2017) applied to our case shows that the inequalities (10) are equivalent to:

$$\mathbb{P}[\mathbf{Q}_j \subseteq A | S_j = s] \leq \mathbb{P}[Q_j \in A | S_j = s] \quad \forall A \in \mathbf{\Lambda}_j^{\cup}(s).$$

Next, consider any  $s \in [c_j, c_j + \varepsilon)$  for  $\varepsilon$  of Assumption 7(i). We have that

$$\mathbb{P}[\mathbf{Q}_j \subseteq A | S_j = s] \leq \mathbb{P}[Q_j \in A | S_j = s] \quad \forall A \in \mathbf{\Lambda}_j^{\cup+},$$

and taking limits on both sides as  $s \downarrow c_j$  leads to

$$\mathbb{P}[\mathbf{Q}_j \subseteq A | S_j = c_j^+] \leq \mathbb{P}[Q_j \in A | S_j = c_j] \quad \forall A \in \mathbf{\Lambda}_j^{\cup+}, \quad (11)$$

where we use continuity of  $\mathbb{P}[Q_j \in A | S_j = s]$  wrt  $s$  (Assumption 2) and existence of the side-limit  $\mathbb{P}[\mathbf{Q}_j \subseteq A | S_j = c_j^+]$  (Assumption 7(ii)). Applying an analogous argument on the left of the cutoff  $c_j$  leads to

$$\mathbb{P}[\mathbf{Q}_j \subseteq A | S_j = c_j^-] \leq \mathbb{P}[Q_j \in A | S_j = c_j] \quad \forall A \in \mathbf{\Lambda}_j^{\cup-}. \quad (12)$$

If there is  $A \in \mathbf{\Lambda}_j^{\cup+} \cap \mathbf{\Lambda}_j^{\cup-}$ , then both (11) and (12) are true which leads to

$$\max\{ \mathbb{P}[\mathbf{Q}_j \subseteq A | S_j = c_j^+] ; \mathbb{P}[\mathbf{Q}_j \subseteq A | S_j = c_j^-] \} \leq \mathbb{P}[Q_j \in A | S_j = c_j]. \quad (13)$$

In summary, we have an inequality for every  $A \in \mathbf{\Lambda}_j^{\cup+} \cup \mathbf{\Lambda}_j^{\cup-}$ . There are three possibilities:  $A \in \mathbf{\Lambda}_j^{\cup+} \cap \mathbf{\Lambda}_j^{\cup-}$  (Inequality 13),  $A \in \mathbf{\Lambda}_j^{\cup+} \setminus \mathbf{\Lambda}_j^{\cup-}$  (Inequality 11), and  $A \in \mathbf{\Lambda}_j^{\cup-} \setminus \mathbf{\Lambda}_j^{\cup+}$  (Inequality 12).  $\square$

## A.8 Proof of Proposition 5

Define

$$\delta_{j,k}(s) = \frac{\mathbb{P}[Q_j = (j, k) | S_j = s]}{\mathbb{P}[\mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset | S_j = s]},$$

where we know  $\delta_{j,k}(s)$  is well defined for  $s$  in a neighborhood of  $c_j$  because  $(j, k)$  is a comparable pair (Definition 5) and Assumption 6.



By Assumptions 2 and 7, side limits of  $\delta_{j,k}(s)$  as  $s \downarrow c_j$  and  $s \uparrow c_j$  are well defined and equal to  $\delta_{j,k}^+$  and  $\delta_{j,k}^-$ , respectively.

Take  $g \in \mathcal{G}$  of Assumption 2. For  $s \geq c_j$ ,

$$\begin{aligned} & \mathbb{E} [g(Y) \mid \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, S_j = s] \\ &= \delta_{j,k}(s) \mathbb{E} [g(Y) \mid Q_j = (j, k), \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, S_j = s] \\ & \quad + (1 - \delta_{j,k}(s)) \mathbb{E} [g(Y) \mid Q_j \neq (j, k), \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, S_j = s] \\ &= \delta_{j,k}(s) \mathbb{E} [g(Y(j)) \mid Q_j = (j, k), S_j = s] \\ & \quad + (1 - \delta_{j,k}(s)) \mathbb{E} [g(Y) \mid Q_j \neq (j, k), \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, S_j = s], \end{aligned}$$

where we use the cutoff characterization and the fact that  $\{Q_j = (j, k)\} \subseteq \{\mathbf{Q}_j \cap \{(j, k)\}\}$ . Taking the limit as  $s \downarrow c_j$ ,

$$\begin{aligned} & \mathbb{E} [g(Y) \mid \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, S_j = c_j^+] \\ &= \delta_{j,k}^+ \mathbb{E} [g(Y(j)) \mid Q_j = (j, k), S_j = c_j] \\ & \quad + (1 - \delta_{j,k}^+) \mathbb{E} [g(Y) \mid Q_j \neq (j, k), \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, S_j = c_j^+], \end{aligned}$$

where again all limits are well defined by Assumptions 2 and 7. Repeating the derivation for  $s < c_j$  and making  $s \uparrow c_j$ ,

$$\begin{aligned} & \mathbb{E} [g(Y) \mid \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, S_j = c_j^-] \\ &= \delta_{j,k}^- \mathbb{E} [g(Y(k)) \mid Q_j = (j, k), S_j = c_j] \\ & \quad + (1 - \delta_{j,k}^-) \mathbb{E} [g(Y) \mid Q_j \neq (j, k), \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, S_j = c_j^-]. \end{aligned}$$

We know the expectations on the left-hand sides of the last two equations, but we do not know the  $\delta$ s or the expectations on the right-hand sides. Above the cutoff, the goal is to partially identify  $\mathbb{E} [g(Y(j)) \mid Q_j = (j, k), S_j = c_j]$  using the distribution of  $g(Y)$  conditional on  $\mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset$  and  $S_j = c_j^+$  plus knowledge of a strictly positive lower bound on  $\delta_{j,k}^+$ , i.e.,  $\underline{\delta}_{j,k}^+$ . Likewise, below the cutoff, the goal is to partially identify  $\mathbb{E} [g(Y(k)) \mid Q_j = (j, k), S_j = c_j]$  using the distribution of  $g(Y)$  conditional on  $\mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset$  and  $S_j = c_j^-$  plus  $\underline{\delta}_{j,k}^-$ . This problem fits the setting of Horowitz and Manski (1995), specifically, Propositions 1–4 and Corollary 4.1 of their paper.

### Part (i)

Assume  $g(Y)$  is a continuous random variable. Assume for the time being that we know the fraction of individuals with  $Q_j = (j, k)$ , that is,  $\delta_{j,k}^+ > 0$ . The lowest possible value for the mean  $\mathbb{E} [g(Y(j)) \mid Q_j = (j, k), S_j = c_j]$  occurs when all individuals with  $Q_j = (j, k)$  are at the lower tail of the distribution of  $g(Y)$  conditional on  $\mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset$  and  $S_j = c_j^+$ . This gives the lower bound

$$\mathbb{E} [g(Y) \mid \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, g(Y) < F_{j,k+}^{-1}(\delta_{j,k}^+), S_j = c_j^+].$$

On the other hand, the highest possible value for the mean  $\mathbb{E} [g(Y(j)) \mid Q_j = (j, k), S_j = c_j]$  occurs when all individuals with  $Q_j = (j, k)$  are at the upper tail of the distribution of  $g(Y)$  conditional on  $\mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset$

and  $S_j = c_j^+$ . This gives the upper bound

$$\mathbb{E} \left[ g(Y) \mid \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, g(Y) > F_{j,k+}^{-1}(1 - \delta_{j,k}^+), S_j = c_j^+ \right].$$

We don't know  $\delta_{j,k}^+$ , but we know that  $0 < \underline{\delta}_{j,k}^+ \leq \delta_{j,k}^+$ , so all values  $\delta_{j,k}^+ \in [\underline{\delta}_{j,k}^+, 1]$  are possible. As the fraction  $\delta_{j,k}^+$  decreases, the lower bound derived above decreases and the upper bound increases. Thus, we widest bounds occur when  $\delta_{j,k}^+ = \underline{\delta}_{j,k}^+$ , that is,

$$\mathbb{E} \left[ g(Y) \mid \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, g(Y) < F_{j,k+}^{-1}(\underline{\delta}_{j,k}^+), S_j = c_j^+ \right]$$

and

$$\mathbb{E} \left[ g(Y) \mid \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, g(Y) > F_{j,k+}^{-1}(1 - \underline{\delta}_{j,k}^+), S_j = c_j^+ \right].$$

Bounds for  $\mathbb{E} [g(Y(k)) \mid Q_j = (j, k), S_j = c_j]$  are derived in an analogous fashion.

### Part (ii)

Assume  $g(Y) = Y$  and  $Y$  is binary. The expressions for the bounds in Part (i) are not valid in this case. To see that, take the lower bound for the mean of  $Y(j)$  as an example. The quantile function  $F_{j,k+}^{-1}(\underline{\delta}_{j,k}^+)$  equals either 0 or 1, so that the conditioning event  $Y < F_{j,k+}^{-1}(\underline{\delta}_{j,k}^+)$  is either empty or has probability that is generally different from  $\underline{\delta}_{j,k}^+$ .

To start, note that

$$\mathbb{E} [g(Y(j)) \mid Q_j = (j, k), S_j = c_j] = \mathbb{P} [Y(j) = 1 \mid Q_j = (j, k), S_j = c_j].$$

Define

$$q_{j,k}^+ := \mathbb{E} [Y \mid \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, S_j = c_j^+] = \mathbb{P} [Y = 1 \mid \mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset, S_j = c_j^+].$$

Assume for the time being that we know  $\delta_{j,k}^+$ .

**Case (a):** Suppose  $1 - q_{j,k}^+ \geq \delta_{j,k}^+$ , that is, the fraction of individuals with  $Y = 0$  in the population of individuals with  $\mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset$  and  $S_j = c_j^+$  is greater than or equal to the fraction of individuals with  $Q_j = (j, k)$  in that same population. In this case, the lowest possible value for  $\mathbb{P} [Y(j) = 1 \mid Q_j = (j, k), S_j = c_j]$  occurs when all  $\delta_{j,k}^+$  individuals have  $Y$  equal 0. That lowest value is zero.

**Case (b):** Now, suppose  $1 - q_{j,k}^+ < \delta_{j,k}^+$ . It is no longer possible to have all  $\delta_{j,k}^+$  individuals with  $Y$  equal 0. The lowest possible value for  $\mathbb{P} [Y(j) = 1 \mid Q_j = (j, k), S_j = c_j]$  has as many of  $\delta_{j,k}^+$  individuals as we can with  $Y = 0$ , that is,  $1 - q_{j,k}^+$  of them. The remaining  $\delta_{j,k}^+ - (1 - q_{j,k}^+)$  individuals must have  $Y = 1$ . That lowest possible value is  $(1 - q_{j,k}^+)/\delta_{j,k}^+ \times 0 + [\delta_{j,k}^+ - (1 - q_{j,k}^+)]/\delta_{j,k}^+ \times 1 = 1 - (1 - q_{j,k}^+)/\delta_{j,k}^+$ .

The lower bound expression for  $\mathbb{P} [Y(j) = 1 \mid Q_j = (j, k), S_j = c_j]$  that covers both cases (a) and (b) is as follows,

$$\max \left\{ 1 - \frac{1 - q_{j,k}^+}{\delta_{j,k}^+}, 0 \right\}.$$

Next, we consider the upper bound for  $\mathbb{P} [Y(j) = 1 \mid Q_j = (j, k), S_j = c_j]$ .

**Case (c):** Suppose  $q_{j,k}^+ \geq \delta_{j,k}^+$ , that is, the fraction of individuals with  $Y = 1$  in the population of individuals with  $\mathbf{Q}_j \cap \{(j, k)\} \neq \emptyset$  and  $S_j = c_j^+$  is greater than or equal to the fraction of individuals with

$Q_j = (j, k)$  in that same population. The highest possible value for  $\mathbb{P}[Y(j) = 1 | Q_j = (j, k), S_j = c_j]$  occurs when all  $\delta_{j,k}^+$  individuals have  $Y = 1$ . That highest value is 1.

**Case (d):** Now, suppose  $q_{j,k}^+ < \delta_{j,k}^+$ . It is no longer possible to have all  $\delta_{j,k}^+$  individuals with  $Y$  equal 1. The highest possible value for  $\mathbb{P}[Y(j) = 1 | Q_j = (j, k), S_j = c_j]$  has as many of the  $\delta_{j,k}^+$  individuals as we can with  $Y = 1$ , that is,  $q_{j,k}^+$  of them. The remaining  $\delta_{j,k}^+ - q_{j,k}^+$  individuals must have  $Y = 0$ . That highest possible value is  $q_{j,k}^+/\delta_{j,k}^+ \times 1 + (\delta_{j,k}^+ - q_{j,k}^+)/\delta_{j,k}^+ \times 0 = q_{j,k}^+/\delta_{j,k}^+$ .

The upper bound expression for  $\mathbb{P}[Y(j) = 1 | Q_j = (j, k), S_j = c_j]$  that covers both cases (c) and (d) is as follows,

$$\min \left\{ \frac{q_{j,k}^+}{\delta_{j,k}^+}, 1 \right\}.$$

We don't know  $\delta_{j,k}^+$ , but we only know the lower bound  $\underline{\delta}_{j,k}^+ > 0$ . As in Part (i) above we compute the bounds at  $\underline{\delta}_{j,k}^+$  because they get weakly wider as  $\delta_{j,k}^+$  decreases.

Bounds for  $\mathbb{P}[Y(k) = 1 | Q_j = (j, k), S_j = c_j]$  are derived in an analogous fashion.

□

## B Sharp Bounds on Treatment Effects

In this section we utilize Artstein's inequality (Theorem A.1 by Molinari (2020)) to characterize sharp bounds on the joint distribution of potential outcomes and true local preferences. That set of distributions produces sharp bounds on averages of treatment effects  $Y(j) - Y(k)$  conditional on  $Q_j = (j, k)$  and  $S_j = c_j$  for any comparable pair  $(j, k) \in \mathcal{P}$ . Recall that we denote the space of possible outcomes  $Y$  as  $\mathcal{Y}$ . Let  $2^{\mathcal{J}^0 \times \mathcal{J}^0}$  denote the power set of  $\mathcal{J}^0 \times \mathcal{J}^0$  and  $2^{\mathcal{Y}}$  denote the power set of  $\mathcal{Y}$ .

**Theorem 1.** *Suppose Assumptions 2, 6, and 7 hold. Assume  $\mathcal{Y}$  is compact. Consider a pair  $(j, k) \in \mathcal{P}$ . Then, the inequalities below characterize the sharp set of all probability values of  $\mathbb{P}[Y(d) \in A, Q_j = (b, b') | S_j = c_j]$  for  $A \subseteq \mathcal{Y}$ ,  $(b, b') \in \mathcal{J}^0 \times \mathcal{J}^0$ , and  $d \in \{b, b'\}$ :*

$$\mathbb{P}[Y \in A, \mathbf{Q}_j \subseteq B | S_j = c_j^+] \leq \sum_{(b, b') \in B} \mathbb{P}[Y(b) \in A, Q_j = (b, b') | S_j = c_j] \quad \forall A \in 2^{\mathcal{Y}}, \quad B \in \mathbf{\Lambda}_j^{\cup+},$$

$$\mathbb{P}[Y \in A, \mathbf{Q}_j \subseteq B | S_j = c_j^-] \leq \sum_{(b, b') \in B} \mathbb{P}[Y(b') \in A, Q_j = (b, b') | S_j = c_j] \quad \forall A \in 2^{\mathcal{Y}}, \quad B \in \mathbf{\Lambda}_j^{\cup-}.$$

**Proof of Theorem 1:** The random set  $(\{Y\} \times \mathbf{Q}_j)$  is a measurable map from  $\Omega$  to  $\mathcal{Y} \times \mathcal{J}^0 \times \mathcal{J}^0$ , so it is compact valued. Following Definition A.1 by Molinari (2020), we say that  $(\{Y\} \times \mathbf{Q}_j)$  is a random closed set because for every compact set  $\mathcal{K} \in \mathbb{R}^3$ , the set  $\{\omega \in \Omega : (\{Y(\omega)\} \times \mathbf{Q}_j(\omega)) \cap \mathcal{K} \neq \emptyset\}$  is a measurable event. By Assumption 6(i), the random vector  $(Y, Q_j)$  and the random set  $(\{Y\} \times \mathbf{Q}_j)$  are measurable maps on the same probability space, and  $\mathbb{P}[(Y, Q_j) \in (\{Y\} \times \mathbf{Q}_j) | S_j = s] = 1$  for any  $s \in \mathcal{S}_j$ .

Artstein's inequality (Theorem A.1 by Molinari (2020)) characterizes the sharp set of all possible probability distributions for  $(Y, Q_j)$  that are consistent with our observation of  $(\{Y\} \times \mathbf{Q}_j)$  and the fact that  $\mathbb{P}[(Y, Q_j) \in (\{Y\} \times \mathbf{Q}_j) | S_j = s] = 1$ . For any  $s \in \mathcal{S}_j$ , the inequality says that

$$\mathbb{P}[Y \in A, \mathbf{Q}_j \subseteq B | S_j = s] \leq \mathbb{P}[Y \in A, Q_j \in B | S_j = s] \quad \forall A \in 2^{\mathcal{Y}}, \quad B \in 2^{\mathcal{J}^0 \times \mathcal{J}^0}. \quad (14)$$

Next, we use the cutoff characterization and rewrite the right-hand side of (14) as a sum. For  $s \geq c_j$ ,

$$\begin{aligned} \mathbb{P}[Y \in A, Q_j \in B | S_j = s] &= \sum_{(b,b') \in B} \mathbb{P}[Y \in A, Q_j = (b,b') | S_j = s] \\ &= \sum_{(b,b') \in B} \mathbb{P}[Y(b) \in A, Q_j = (b,b') | S_j = s]. \end{aligned} \quad (15)$$

Substitute (15) into (14) and take the limit as  $s \downarrow c_j$  on both sides,

$$\mathbb{P}[Y \in A, \mathbf{Q}_j \subseteq B | S_j = c_j^+] \leq \sum_{(b,b') \in B} \mathbb{P}[Y(b) \in A, Q_j = (b,b') | S_j = c_j] \quad \forall A \in 2^{\mathcal{Y}}, B \in 2^{\mathcal{J}^0 \times \mathcal{J}^0}, \quad (16)$$

where the limits of the left- and right-hand sides of the inequality are well defined by Assumptions 7 and 2, respectively.

Similarly, for  $s > c_j$ ,

$$\begin{aligned} \mathbb{P}[Y \in A, Q_j \in B | S_j = s] &= \sum_{(b,b') \in B} \mathbb{P}[Y \in A, Q_j = (b,b') | S_j = s] \\ &= \sum_{(b,b') \in B} \mathbb{P}[Y(b') \in A, Q_j = (b,b') | S_j = s]. \end{aligned} \quad (17)$$

Use (17) into (14) and take the limit as  $s \uparrow c_j$  on both sides,

$$\mathbb{P}[Y \in A, \mathbf{Q}_j \subseteq B | S_j = c_j^-] \leq \sum_{(b,b') \in B} \mathbb{P}[Y(b') \in A, Q_j = (b,b') | S_j = c_j] \quad \forall A \in 2^{\mathcal{Y}}, B \in 2^{\mathcal{J}^0 \times \mathcal{J}^0}. \quad (18)$$

Lemma 1 of Chesher and Rosen (2017) applied to our case shows that the inequalities (16) and (18) are equivalent to:

$$\mathbb{P}[Y \in A, \mathbf{Q}_j \subseteq B | S_j = c_j^+] \leq \sum_{(b,b') \in B} \mathbb{P}[Y(b) \in A, Q_j = (b,b') | S_j = c_j] \quad \forall A \in 2^{\mathcal{Y}}, B \in \mathbf{\Lambda}_j^{\cup+},$$

$$\mathbb{P}[Y \in A, \mathbf{Q}_j \subseteq B | S_j = c_j^-] \leq \sum_{(b,b') \in B} \mathbb{P}[Y(b') \in A, Q_j = (b,b') | S_j = c_j] \quad \forall A \in 2^{\mathcal{Y}}, B \in \mathbf{\Lambda}_j^{\cup-}.$$

□

Theorem 1 characterizes the sharp set of all possible probability values of  $\mathbb{P}[Y(d) \in A, Q_j = (b,b') | S_j = c_j]$  for  $A \subseteq \mathcal{Y}$ ,  $(b,b') \in \mathcal{J}^0 \times \mathcal{J}^0$ , and  $d \in \{b,b'\}$ . For a fixed  $g \in \mathcal{G}$  of Assumption 2, that set of distributions allows us to define the sharp set of all possible means of potential outcomes,

$$\mathbb{E}[g(Y(d)) | Q_j = (j,k), S_j = c_j],$$

for  $(j,k) \in \mathcal{P}$  such that  $\underline{p}_{j,k} > 0$  and  $d \in \{j,k\}$ . The set of possible means in turn allows us to define the sharp set of all average treatment effects of the form

$$\mathbb{E}[g(Y(j)) - g(Y(k)) | Q_j = (j,k), S_j = c_j].$$

In case of continuous  $Y$ , it is impossible to directly evaluate all inequalities of Theorem 1 because there are uncountably many sets  $A \in 2^{\mathcal{Y}}$ . This is one of the drawbacks of the Artstein’s inequality approach which has been extensively discussed by Beresteanu et al. (2012). In case  $Y$  takes finitely many values, e.g., when  $Y$  is binary, the number of such inequalities is feasible to evaluate because  $2^{\mathcal{Y}} \times \Lambda_j^{\cup+}$  (or  $2^{\mathcal{Y}} \times \Lambda_j^{\cup-}$ ) has finitely many elements. In fact, the number of inequalities is slightly higher than the number of inequalities in Proposition 4 of the main text, which we utilize to compute lower bounds on  $\delta_{j,k}^+$  and  $\delta_{j,k}^-$ .

## C Empirical Appendix

Table 2: Descriptive statistics: students

	All students		Applicants to BA Bus. Adm., PUC Santiago	
	Mean	Std. deviation	Mean	Std. deviation
Number of programs in ROL	4.86	2.20	4.48	1.94
ROL strictly shorter than permitted	.80	.40	.87	.34
Assigned (to any prog.)	.65	.48	.73	.44
Rank of assigned program, cond. on assigned	2.24	1.59	2.10	1.45
Reapplies	.19	.39	.12	.32
Graduates from assigned prog., cond. on assigned	.33	.47	.40	.49
Graduates from any program	.77	.42	.86	.34
Number of students	519,409		9,398	

Table 3: Descriptive statistics: programs

	All programs		BA Business Adm., PUC Santiago (across years)	
	Mean	Std. deviation	Mean	Std. deviation
Number of applicants (per year)	473	369	1,541	261
Number of admitted students (per year)	66	49	245	3
Year-to-year absolute change in cutoff (points)	7.4	93	4.1	4.2
Cutoff			720	10
Number of programs	1,191			