Essays on the Empirics of School Choice

by

Margaux Luflade

Department of Economics
Duke University

Date: ____________________

Approved:

Arnaud Maurel, Supervisor

Peter Arcidiacono

Vincent Joseph Hotz

Matthew Masten

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2018

ABSTRACT

Essays on the Empirics of School Choice

by

Margaux Luflade

Department of Economics
Duke University

Date: ______________________
Approved:

____
Arnaud Maurel, Supervisor

____
Peter Arcidiacono

____
Vincent Joseph Hotz

____
Matthew Masten

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics in the Graduate School of Duke University
2018

Abstract

This dissertation is an empirical study of school choice under various perspectives. Chapter 2 looks at school choice from the application perspective. It focuses on assignment mechanisms used by school districts to allocate available seats to students. Chapters 3 and 4 turn to the admission perspective, investigating the consequences of school choice on students' outcomes. Chapter 3 studies the effects on students' academic performance of admission to one type of selective schools, namely elite schools. Chapter 4 explores the mechanisms underlying these effects. The empirical analysis in each chapter of this dissertation is supported by administrative data from Tunisia.

With school choice comes the necessity to devise rules to decide who gets to enroll in a school or academic program when more students express the willingness to attend than the school's capacity allows. Centralized assignment mechanisms based on the deferred acceptance algorithm (DA) are used by school districts around the world to assign students to schools. Theoretical analyses of the DA consider that students are allowed to list all the alternatives of the choice set in their application rankings. However, in virtually all places where these mechanisms are implemented, students are restricted to list only a small number of choices. As a consequence, students need to take their admission chances into account, and be strategic in their choice.

In Tunisia, high school graduates are assigned to university programs using a sequential variant of the DA. Chapter 2 use data on this Tunisian mechanism to empirically

examine the effect of enabling students to update their expectations about their admissions probabilities. The sequential implementation induces quasi-experimental variation in the information available to students about remaining vacancies, and allows for the identification of students' preferences and expected admission probabilities.

When students cannot revise their expectations, and relative to a benchmark situation in which students are given perfect information about which programs would admit them, their average indirect utility is decreased by the equivalent of a 41km-increase in the distance home-university —40% of the median distance traveled by students in the data. While easy to implement, the sequential implementation of the DA procedure reduces this expected utility loss by 67% in Tunisia. The increase in expected welfare is driven by a decrease in the share of students rejected by all their listed choices. Gains disproportionately accrue to low-ability and low-SES students. Counterfactuals suggest that a better targeting of low-priority students by the information provision would increase welfare gains.

Underlying school choice is the idea that giving students and their family more freedom in their schooling decisions can improve academic outcomes. Although documented in many papers, the impact of attending a better school on future achievement is unclear and varies greatly depending on the context. Chapter 3 examines the impact of being admitted to a high school with high achieving peers in Tunisia, particularly on post-secondary choices. The admission mechanism creates admission cutoffs that we exploit in a sharp regression discontinuity (RD) design. However, despite the validity of the RD design, sample selection and measurement error induce the standard RDD identification argument to fail, and the naive RD techniques to produce biased estimates. Chapter 3 proposes bounds for the average and quantile treatment effects. Results suggest that admission to an elite high school increases students' end-of-high-school performance, and the selectivity level of post-secondary programs students in the higher end of the distribution get assigned to.

Chapter 4 investigates one type of mechanisms possibly driving the effect of selective high schools on students' outcomes: the change in educational inputs induced by admission to an elite school. To explore this mechanism, we link the students database to data on schools infrastructures and teachers. Allowing effects of admission to an elite high school to vary across the twelve Tunisian elite high school institutions, we evaluate the link between the magnitude of the treatment effects on students' outcomes, and the intensity with which treatment modifies various dimensions of the school environment. Results suggest that, although average teachers' quality and student monitoring are increased by admission to an elite high school, the higher peer achievement seems to be the main mediator of treatment effects on students' outcomes.

Contents

List of Tables

List of Figures

Acknowledgements

I am extremely grateful to Arnaud Maurel, Peter Arcidiacono, Joe Hotz, and Matt Masten for the help and advice they have shared over the past four years, as well as their continuous support and encouragement. I would like to thank Arnaud more particularly for having agreed to supervise my doctoral studies; having been generous with his time over the past few years; having been patient and reassuring when it took me time to pin down research ideas; having expressed enthusiasm and confidence in my work in the moments I had lost them; having put energy in setting up deadlines for me when I needed something to work towards; and having shared advice and his experience as I prepared for the job market.

I am deeply indebted to Meryam Zaiem, whose support and collaboration have been instrumental at many stages of my dissertation work. Meryam shared access to the data used in this dissertation. Chapters 3 and 4 are directly based on joint work with Meryam; and her insights about the context, and encouragements have also been critical to the realization of Chapter 2. Meryam's contribution to this dissertation goes beyond our scientific collaboration, though. Sharing doubts and frustrations, exchanging tips and encouragements, Meryam was the first to make me realize and face that doctoral research is more than a scientific adventure. She helped me understand how to rebound and grow from failed attempts. I know —and I hope Meryam feels so as much as I do— that a strong friendship was born out these long hours of work.

For providing the data, I thank the Ministry of Education, as well as the Ministry of Higher Education of Tunisia for providing the data. I also acknowledge funding from the Duke Graduate School, and the James A. and Deborah Allen Hewitt Fellowship.

In addition, Chapter 2 particularly benefited from discussions with Atila Abdulkadiroğlu, Eli Liebman, Seth Sanders, Modibo Sidibé, and participants at the Duke Labor and Microeconometrics workshops, and at the 2017 Young Economists Symposium at Yale University.

Finally, I would like to thank those who, more generally, helped me make it reasonably sanely and healthily through the Ph.D. experience. As already mentioned, the experience would have probably been much more lonely without Meryam. Nicolas-Aldebrando Benelli and Maria Marta Laurito have helped me keep in mind that life goes on when research is not as fruitful as one hopes. Eli Liebman's encouragements and confidence made me find focus and energy in the most stressful and hopeless moments. Papa and Maman's unconditional support gave me the strength to go on each time I was tempted to give up.

Chapter 1

Introduction

This dissertation is an empirical study of school choice under various perspectives. Chapter 2 looks at school choice from the application perspective. It focuses on assignment mechanisms used by school districts to allocate available seats to students. Chapters 3 and 4 turn to the admission perspective. They investigate the consequences of school choice on students' outcomes. Specifically, Chapter 3 studies the effects on students' academic performance of admission to one type of selective schools, namely elite schools. Chapter 4 explores the mechanisms underlying these effects. The empirical analysis in each chapter of this dissertation is supported by administrative data from Tunisia.

The next two sections of this introduction present the questions studied in Chapters 2 to 4. Section 1.3 introduces the secondary and post-secondary education systems in Tunisia.

1.1 School choice and centralized assignment mechanisms

With school choice comes the necessity to devise rules to decide who gets to enroll in a school or academic program when more students express the willingness to attend than the school's capacity allows.

The deferred acceptance (DA) algorithm (Gale and Shapley (1962)) is one of the most extensively used algorithms to assign students to schools around the world. In the US, it has been used in New York City since 2003 to assign students to high schools, by Boston Public Schools since 2006, and in Chicago since 2009. Abroad, Chile, Norway, Spain, Taiwan, Tunisia, Turkey, etc. use a DA-based mechanism to assign high school graduates to university programs. The use of DA-based mechanisms has been advocated over alternative mechanisms (Abdulkadiroğlu and Sönmez (2003); Balinski and Sönmez (1999)) on the grounds that, when students are allowed to list all the alternatives of the choice set in their application rankings, it is strategy-proof and eliminates justified envy (Dubins and Freedman (1981); Roth (1982)). However, in virtually all places where DA-based mechanisms are implemented, students are restricted to list only a small number of choices. When the length of her application list is restricted, a student may actually be rejected from all the academic programs she applies to. As a consequence, students need to take into account their admission chances to the programs, and be strategic in their choice (Haeringer and Klijn (2009); Calsamiglia et al. (2010)).

Chapter 2 shows that a simple modification of assignment mechanisms based on the DA can improve the quality of student-school matches. In this chapter, I empirically examine the effect of enabling students to update their expectations about their admissions probabilities. Taking restrictions on the number of applications as fixed, I investigate how providing students with information can improve the quality of school-student matches. Doing so, I consider a simple information provision design that can be easily embedded in commonly implemented DA-based assignment mechanisms. The standard (one-phase) implementation of the DA involves the whole cohort of N students simultaneously submitting their application lists, and then being assigned via the DA. In contrast, I consider a sequential implementation, which involves first partitioning the cohort in $K \leqslant N$ assignment groups that successively submit lists and are assigned. After a group is assigned, and before next group students submit their application lists, information about which vacancies remain and which programs filled up is publicly updated. To quantify the gains in student welfare from the updating allowed by such informational updates, I estimate a model of school application portfolio choice, and use my estimates in a counterfactual exercise. In the model, students can be strategic in their choices, and may not know their true admission probabilities. To deal with the empirical challenge of identifying students' preferences for programs when observed choices are the result of expected utility maximization, I use administrative data from Tunisia, where high school graduates are assigned to university programs using a three-phase sequential variant of the DA. I show that the revelation of information gives incentives to a subset of applicants to simply list their most-preferred programs in their application portfolio. From these students, I recover population utility parameters using standard discrete choice techniques. In a second step, I use other students' application portfolios to identify applicants' expectations about their admission chances; expectations rationalize submitted lists given utility parameters. I find, that, when students cannot revise their expectations, and relative to a benchmark situation in which students are given perfect information about which programs would admit them, their average indirect utility is decreased. The utility loss is economically significant; it is equivalent to a 41km-increase in the distance home-university –40% of the median distance traveled by students in the data. While easy to implement, the sequential implementation of the DA procedure reduces this expected utility loss by 67% in Tunisia, and drastically decreases the share of students getting rejected by all their listed choices. Gains disproportionately accrue to low-ability and low-SES students, and counterfactuals suggest that a better targeting of low-priority students by the information provision would lead to larger welfare gains.

1.2 School choice and students' outcomes

Underlying school choice is the idea that giving students and their family more freedom in their schooling decisions can improve academic outcomes. Chapters 3 and 4 (based on joint work with Meryam Zaiem) explore this hypothesis.

Chapter 3 exploits in a sharp regression discontinuity (RD) design the admission cutoffs generated by the mechanism used to admit students into elite school in Tunisia. From the policy perspective, this paper expands the already existing literature on reduced-form effects of selective schools by considering post-secondary outcomes on a nationwide scope. While most of the literature has focused on outcomes measured no later than the end of high school, by linking high-school application data with the above-mentioned data on the Tunisian post-secondary assignment procedure, we are able to document the effect of admission to an elite high school on the field, location, and selectivity level of the programs students apply and get assigned to. From the methodological perspective, this paper shows that, despite the validity of the RD design, average and quantile treatment effects are not immune to biases resulting from sample selection and missing outcome data. We propose and estimate bounds for the true effects. We find that admission to an elite high school increases student performance at the end of high school. In addition, it increases the selectivity level of post-secondary programs students in the higher end of the distribution get assigned to. These conclusions differ significantly from those that would be drawn from naive RD estimates, hence highlighting the composition effects that need to be accounted for. Naive RD estimates show no significant difference between observed marginally treated and untreated girls' test scores and post-secondary application choices. However, they show that treated girls have a larger exam-taking rate than untreated girls. Our bounds account for the change in the composition of control and treatment groups following the treatment-induced increase in girls' probability to take the end-of-high-school exam. Correcting for the induced selection shows that treatment does have a significant and positive effect on exam scores for girls who would take the exam regardless of their treatment status ('always-exam-takers'). Naive estimates being downward-biased is consistent with the intuition that students taking the exam only under treatment might be academically weaker than always-exam-takers, and would therefore perform worse than them when taking the end-of-high-school exam.

The heterogeneity of the effects we find in Chapter 3, as well as the variety of results in the literature, call for a better understanding of the mechanisms at play behind the effects of admission to an elite high school on students' outcomes. In Chapter 4, we investigate one type of mechanisms: the change in educational inputs induced by admission to an elite school. To explore this mechanism, we link the students database to data on schools infrastructures and teachers. Allowing effects of admission to an elite high school to vary across the twelve Tunisian elite institutions, we evaluate the link between the magnitude of the treatment effects on students' outcomes, and the intensity with which treatment modifies various dimensions of the school environment. Results suggest that, although average teachers' quality and student monitoring are increased by admission to an elite high school, the higher peer achievement seems to be the main mediator of treatment effects on students' outcomes.

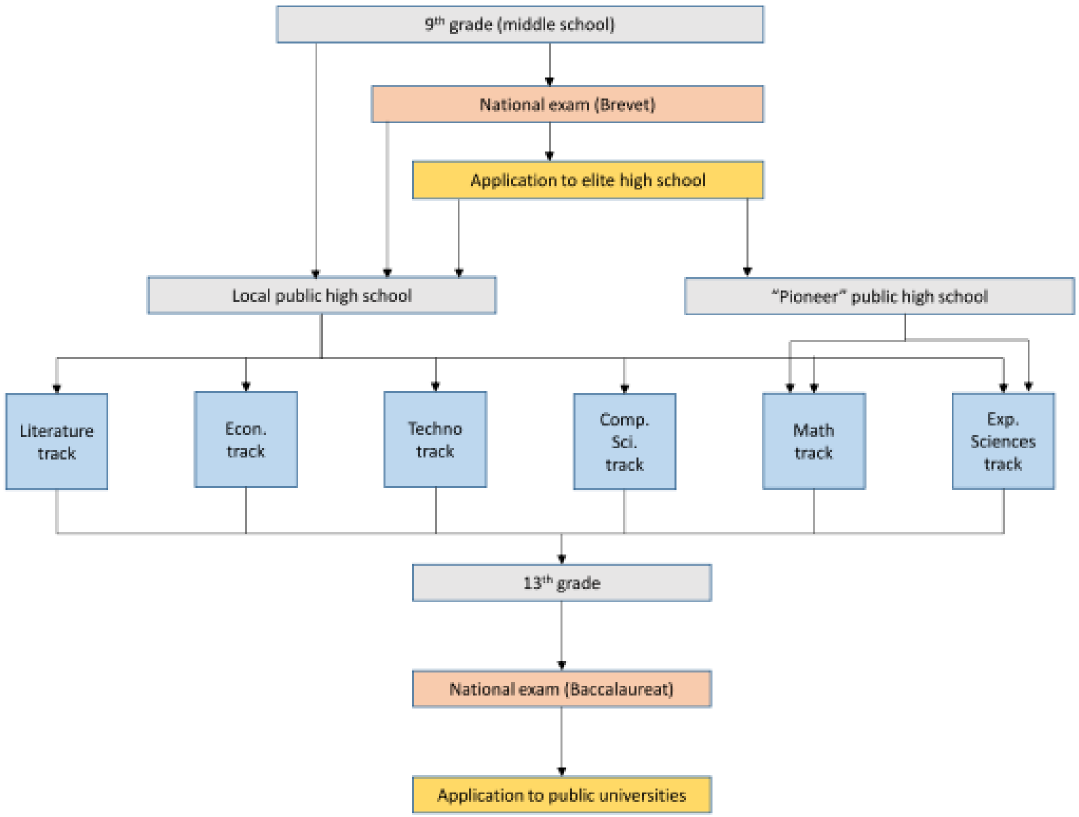


FIGURE 1.1: Public education in Tunisia

1.3 Empirical context: high school and college application in Tunisia

Compulsory schooling in Tunisia starts at age 6 with six years of primary education, followed by three years of preparatory education (in middle schools) and then four years of secondary education (in high schools). After that, high-school graduates may decide to pursue post-secondary studies at one of the country's universities. School is compulsory until age 16. An overview of Tunisia's public secondary education system is pictured in Figure 1.1.

1.3.1 Tunisia's *lycées pilotes*

Chapters 3 and 4 of this dissertation focus on the high school education system in Tunisia. Central to the analysis in these chapters is data gathered at the "Application to elite high school" stage depicted on Figure 1.1. These chapters also rely on outcome data gathered at the "National exam (Baccalaureat)" and "Application to public universities" stages of Figure 1.1. This section provides general institutional details about Tunisia's high school system, and, in particular, its selective high schools. More precise details about the application and admission procedure to these selective high schools, as relevant to the identification strategy in Chapters 3 and 4, are given in Section 3.2.1.

At the end of grade 9, students may take the national end-of-middle-school exam, which consists of written tests in Arabic, French, English, mathematics and natural sciences. Tests are anonymously graded by teachers from schools other than the test-taker's. Taking this exam is not compulsory. In 2011, 24%[1] of the students registered in the last year of middle school took it. The decision of students to take the exam seems to be related to three main reasons. (i) Passing the exam (i.e. scoring above 10 out of 20) grants a place in a high school, regardless of performance during the academic year –so for some students the exam is an opportunity not to repeat the last grade of middle school. (ii) Taking the exam is required to apply to selective high schools, and admission to these schools is exclusively based on exam scores. (iii) It is a terminal credential for those who do not plan to pursue additional schooling beyond middle school.

---

[1] Source: Authors' calculations based on Direction Générale des Études, de la Planification et des Systèmes d'Information (République tunisienne, Ministère de l'Éducation, Secrétariat Général) (2013) and end-of-middle-school exam data provided by the Ministry of Education.

In both middle school and high school, students are not free to choose the courses they take. Two years before the end of high school, they simply choose to major in one of six possible fields[2]. In each field, the Ministry of Education establishes the curriculum and required courses, which are common to all public high schools, selective or not. At the end of grade 13, students may take the national end-of-high-school exam (*Baccalauréat*), which consists of six to eight tests depending on the student's high school major. The major also determines the weights associated with each test in the computation of the global end-of-high-school exam score. Again, taking the exam is not compulsory; but completion is required to enroll in universities and pursue higher education. It is also seen as a terminal credential for those who do not plan to pursue additional schooling beyond high school.

There are three kinds of high schools in Tunisia: private high schools, which mainly enroll students who struggle to stay in the public system[3]; (non-selective) public high schools; and selective public high schools. The first selective public high schools – called *lycées pilotes*, i.e. *pioneer schools*[4]– opened in 1983 with the goal of providing an elite education to the country's brightest students. Selective schools teach the same curriculum and courses as non-selective high schools, but officially differ from regular high schools in the following dimensions. Students are subject to permanent selection; passing to the next grade requires minimum scores in some subjects and grade repetition is not allowed. Additional computer science classes are mandatory and taught at all levels. Selective high schools only offer the math and experimental

[2] The six possible fields of specialization are: literature, experimental sciences, math, economics, technology, and computer science. In 2011, physical education was added.

[3] Private high school accounted for 10% of the students in high schools in 2008. Source: Direction Générale des Études, de la Planification et des Systèmes d'Information (République tunisienne, Ministère de l'Éducation, Secrétariat Général) (2013).

[4] In the remainder of this dissertation, I use "selective high schools", "elite high schools", "pioneer high schools" and "*lycées pilotes*" interchangeably.

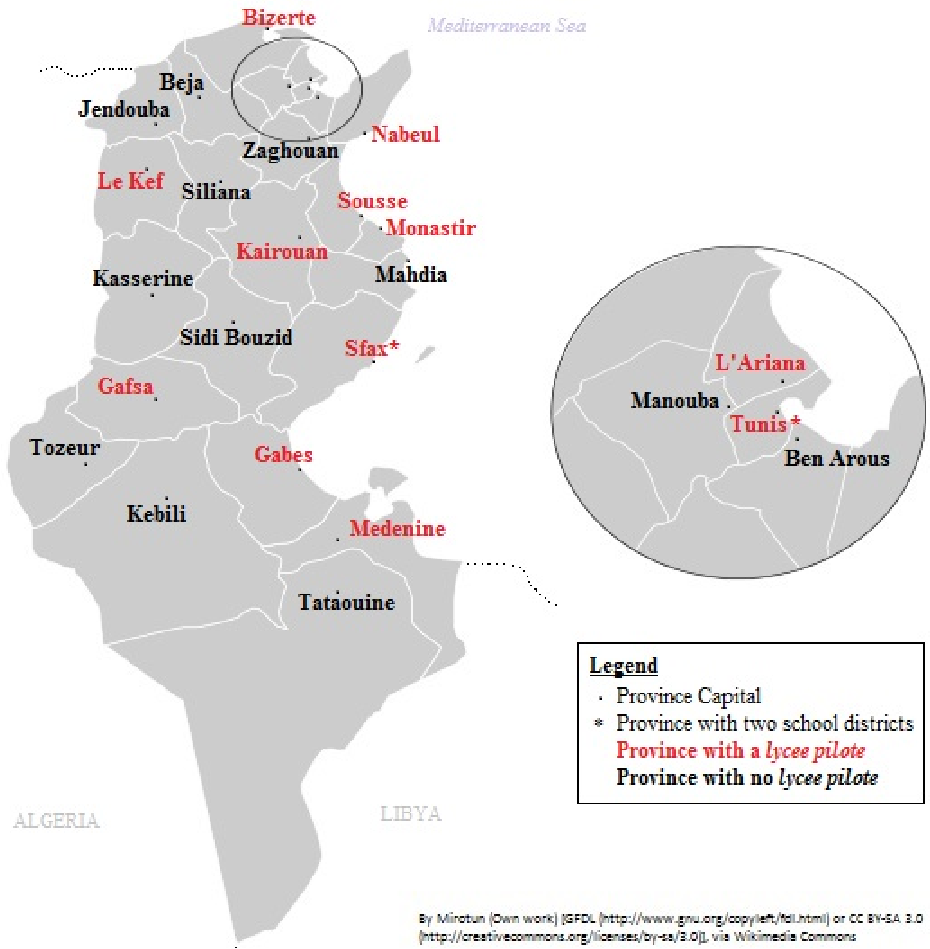


FIGURE 1.2: A map of Tunisia

sciences majors; and students have to wear a uniform.

Figure 1.2 illustrates the location of the *lycées pilotes* in Tunisia over the period considered (2006–08).

1.3.2 Post-secondary studies in Tunisia

Chapter 2 of this dissertation focuses on post-secondary education system in Tunisia. Central to the analysis in this chapter is data gathered at the "Application to public universities" stage depicted on Figure 1.1. This section provides general institutional details about Tunisia's university system. More precise details about the centralized application and admission procedure to university programs, as relevant to the identification strategy in Chapter 2, are given in Section 2.3.1.

Every June, high-school seniors in Tunisia take the national end-of-high-school exam. Passing this exam —that is, scoring at least 10 out of 20 on average over the eight to ten tests of the exam— is a sufficient and necessary condition to graduate from high school and gain access to public post-secondary education in Tunisia. Tunisia counts fourteen public universities, each delivering a wide range of degrees. Degrees are field-specific; each of them requires the completion of a standard curriculum approved by the Ministry of Higher Education, which generally involves undergraduate students specializing in one field of study as soon as their first semester. Hence, when deciding on her post-secondary education, a student needs to choose a university *and* a field of study. Here, I refer to such pair (university, field) as a *program* or *track*. While graduating from high school guarantees Tunisian students access to public higher education (graduating seniors are automatically registered in the centralized post-secondary application system), the particular program they will be allowed to enroll in is determined by a centralized assignment mechanism, which is detailed in Section 2.3.1.

Chapter 2

The Value of Information in Centralized School Choice Systems

2.1 Introduction

New York City, Paris, Spain, Finland, Turkey, Chile, Norway, Ghana, Tunisia all use a similar centralized procedure to assign students to public schools or university programs. This mechanism is based on the deferred acceptance (henceforth, DA) algorithm (Gale and Shapley (1962)), and has been recommended to policy-makers by the school choice market design literature (Abdulkadiroğlu and Sönmez (2003); Balinski and Sönmez (1999)) on the grounds of its desirable theoretical properties. The mechanism involves students submitting to a clearing house an ordered list of schools they would like to attend, schools giving priorities to students over admission offers, and the algorithm processing application lists and priorities to assign students to programs. Not only do all these places use a similar assignment mechanism, but they also implement it with the same departure from the theoretical design. While theoretical analyses of the algorithm are based on students being able to apply to all schools in their choice set, in practice, applicants in all these places are only allowed to list a restricted, and often small, number of programs in their preference report —for instance, twelve in New York, six in Ghana, ten in Tunisia, out of more than six or seven hundreds of alternatives.

This paper empirically examines the students' application portfolio choice problem when they are not able to apply to all academic programs in their choice set, and investigates the effects of enabling applicants to update their expectations about their admission chances. When students are not restricted in the number of applications they can make, mechanisms based on the DA ensure that it is dominant for applicants to simply report schools by order of preference in their application list (Dubins and Freedman (1981); Roth (1982)). List-size restrictions break this property (Haeringer and Klijn (2009)). When they can only apply to a subset of programs, students face the possibility to be rejected from all their listed choices. To avoid rejection, students need to choose their application portfolio taking into account not only their preferences for academic programs, but also their probability to be admitted to these programs. Students' expectations about their probabilities of admission are then a crucial determinant of where they apply and are ultimately accepted.

Taking restrictions on the number of applications as fixed[1], this paper also investigates how providing students with information can improve the quality of school-student matches. Focusing on guiding the formation of expectations about admission chances, I consider the provision of updated information about programs filling up and vacancies remaining, at various points in time in the assignment process. I examine an information provision design that can be easily embedded in commonly implemented DA-based assignment mechanisms.

[1] There is evidence that many policy-makers are reluctant to let students submit lists as large as their choice set when the choice set is large, as this quote from Roth (2015) about the New York City match illustrates: "*[I]n my description [...] students can list as many schools as they like. We economists recommended that students be allowed to do just that, but on this important detail we did not prevail. So New York City students today can list only up to twelve programs among the hundreds that the city offers. Students who want to list more than that face a strategic choice of which twelve to list.*" See also Pathak and Sönmez (2013).

This paper uses administrative data from Tunisia, where college applications and assignments are made using a nationwide centralized assignment mechanism based on the DA. A unique institutional feature, the Tunisian mechanism is implemented in a sequential way, and involves different pools of applicants having different information about the available vacancies. This special implementation yields a quasi-experimental setting that enables me to empirically document three facts.

First, I use the quasi-experiment generated by the implementation of the mechanism to give evidence that students behave strategically when forming application lists. Despite the increasing number of school systems implementing the restricted-list DA, there is little empirical evidence on the performance of DA-based assignment procedures when the number of applications students can submit is restricted. This paper gives evidence that, under such restrictions, students may not find it optimal to truthfully apply to their most preferred schools.

Second, I show that application list size restrictions can decrease student welfare and increase inequality, relative to a setting in which students are freed from the need to form expectations about their admission chances and to engage in strategic behavior. I consider a context in which students' expectations about their admission chances may not coincide with their true probabilities of admission. I estimate a model of application portfolio choice, and use a counterfactual analysis to compare assignment outcomes resulting from the implementation of the restricted-list DA, to those obtained under a strategy-proof setting. Taking advantage of the quasi-experimental variation in information available to students, and in line with findings in the literature, the model allows expectation formation and the use of public information to differ across socioeconomic status (SES) and related variables (e.g. Hoxby and

Turner (2015)).

Third, I find that a simple modification in the implementation of restricted-list DA can improve student welfare. A sequential implementation of the DA, as done in Tunisia, permits the provision of updated information to students about programs filling up and remaining vacancies. The effect of additional information on expected welfare is *a priori* ambiguous. On the one hand, more information about the vacancies that have been taken and remain may prevent students from applying to programs that turn out to be all full, and it may increase the quality of the matches made. On the other hand, it may decrease applicants' ability to signal the magnitude of their preference for the different alternatives (Abdulkadiroğlu et al. (2015)).

Comparing students' outcomes when they under the most common implementation of the restricted-list DA, relative to the strategy-proof benchmark, I find that students' average expected indirect utility is decreased. In magnitude, the average decrease in indirect utility is equivalent to the counterfactual decrease induced by, keeping all other things equal, having students attend a university 41km (25 miles) further away from home —about 38% of the median distance traveled by students in the data. While easy to implement, the 2010 Tunisian three-phase implementation of the restricted-list DA reduces this welfare loss by 67% . The increase in expected welfare is essentially driven by enabling a larger share of students to be assigned to an element of their application list —rather than to assigned students improving their match. Gains disproportionately accrue to low-ability, unsophisticated, and low-SES students. In fact, providing information about vacancies, even through a small number sequential of sequential phases, reduces the expected indirect utility gap existing between high- and low-SES students. Finally, while the 2010 Tunisian implementation of the three-phase procedure does increase welfare and the average match rate, I show that a better targeting of low-priority students by the information provision —through a different sequential partition of the cohort of applicants— could increase gains to students.

I face two main empirical challenges. The first is an identification challenge generally faced by the empirical literature on matching mechanisms. The mapping from students' preferences to their application choices depends on their expectations about which schools may be available for them. With most school applications datasets, the econometrician cannot separately identify students' preferences and expectations about admission chances (Agarwal and Somaini (2014)). The quasi-experiment induced by the Tunisian sequential procedure helps me circumvent this identification problem. I argue that the sequential design induces a subset of students to truthfully report their most-preferred programs. For this subset of students, perceived admission chances can be ignored, and I can identify and estimate students' preferences for post-secondary programs. In a second step, I characterize students' expectations about their admission chances as those rationalizing other students' observed application lists, given identified preferences. The second challenge is computational. Given a student's preferences and expectations about her admission chances, finding the optimal application portfolio —that is, the expected-utility maximizing ordered list of up to ten programs among more than 600 alternatives— is intractable. The two-step approach, that identifies and estimates separately preferences parameters and expectations about admission chances, partially alleviates this issue.

Related literature

This paper contributes to three branches of the literature. It adds evidence to the small empirical literature on the DA. The theoretical literature on mechanism design is large and influential. In the context of school choice, it is part of an active dialogue between economists and policy-makers that has highlighted strategy-proofness as a way to pursue "transparency, fairness, and equal access to public facilities" (Abdulkadiroğlu et al. (2006a)). The use of the strategy-proof DA has been recommended over other mechanisms (e.g. the Boston mechanism) that reward strategic behavior. It avoids penalizing students and families who do not strategize or do not do it well —which has been showed to be correlated with socioeconomic background (Kapor et al. (2016)). Despite the widespread use of the DA, there is little empirical evidence of the consequences of a central feature of its implementation —the restriction imposed on the number of schools students can apply to.[2] Ajayi and Sidibé (2016) is, to my knowledge, the only empirical paper that addresses this question. Using data from Ghana, where the DA is used to assign students to high schools, they quantify the effect of changing the number of programs students are allowed to apply to. Fack et al. (2015) also document strategic behavior in assignment systems based on the (restricted-list) DA. In their analysis of the Paris high-school match, they test and reject the null hypothesis that students are truth-telling. These two recent analyses deliver an empirical counterpart to the experimental findings in Calsamiglia et al. (2010). In contrast, though, Abdulkadiroğlu et al. (2017) provide empirical evidence from the high-school match in New York City (NYC) that students may find it optimal to truthfully report their preferences, even when constrained to submit a application list strictly smaller than their choice set.

My paper differs from these empirical papers in two respects. First, it is the only one to document the effects of a practical and simple policy that provides decision-

[2] A number of studies have compared the unrestricted-list DA to alternative mechanisms (e.g. Agarwal and Somaini (2014); Calsamiglia et al. (2014); Dur et al. (2016); He (2016)). A number of studies have also analyzed other less common assignment mechanism (e.g. Carvalho et al. (2014).

makers with updated information about vacancies, and enables them to update their expectations about their admission chances. Second, the analysis of students' preferences for academic programs and expectations about their admission chances does not *a priori* constrain students to be all strategic, nor to all truthfully apply to their most-preferred programs. Rather, the two-step identification strategy used in this paper allows me to recover the share of students engaging in each type of behavior.

This paper also relates to studies on the role of information and students' imperfect sophistication in the context of centralized school choice systems. The questions tackled in my paper are similar to those Kapor et al. (2016) explore using a survey of about 200 parents of kindergartners and ninth-graders participating in the New Haven school choice mechanism. They show that subjective beliefs about their child's admission chances differ from true admission probabilities, and that the magnitude of the deviation depends on parental effort and demographics. Mitigated empirical evidence about the effect of information on applications was earlier provided by Hastings et al. (2007) and Hastings and Weinstein (2009) using a field experiment conducted in the Charlotte-Mecklenburg Public School District in 2006.
My work complements these studies, as it considers application in a DA framework, while they focus on variants of an alternative assignment mechanism, the Boston mechanism.

More generally, beliefs about admission chances are part of a larger set of expectations students form about variables entering their application decisions and educational choices in general.[3] Recent studies have acknowledged that the expectations students and families form about the outcomes of their investment and application choices may be inaccurate (for instance on future wages, see, among others,

[3] See Altonji et al. (2012), and Altonji et al. (2015) for a review.

Wiswall and Zafar (2013); Jensen (2010); Stinebrickner and Stinebrickner (2014b)). More broadly, studies have shown as well that agents need to form beliefs about the features of their educational decisions they do not have perfect knowledge of, and that providing them with additional information may actually affect their choices — whether it is information about schools and/or curricula characteristics over which the decision-makers may have preferences (e.g., on school quality, see Hastings et al. (2007), Hastings and Weinstein (2009)); or information about one's own ability in the curriculum or taste for these characteristics (see Pistolesi (2016); Arcidiacono et al. (2012); Stinebrickner and Stinebrickner (2014a)).

Outline

The rest of this paper is organized as follows. The next section reviews the theoretical properties of the deferred-acceptance algorithm, illustrates the inefficiencies likely to arise when it is implemented with application list size restrictions, and presents the alternative sequential procedure. Section 3 introduces the empirical setting of this paper. It presents the post-secondary assignment procedure in Tunisia, describes the data, and highlights reduced-form effects of the sequential information revelation on application behaviors. Section 4 describes my strategy to recover students' preferences for university programs, and shows my estimates. Section 5 describes my strategy to characterize students' expectations about their admission chances, and shows that not all students truthfully list their most-preferred programs. Finally, Section 6 compares students' outcomes under the sequential DA procedure and the standard implementation of the restricted-list DA, and discusses the value of information in a centralized school choice system. Section 7 concludes.

2.2 Theoretical background

In this section, I review the theoretical properties of the DA; I describe the consequences of its implementation in a school choice context in which applications are constrained or costly; and I present the alternative sequential procedure at the center of this paper. This section serves two main purposes. First, it establishes key properties of the mechanism that will ground the identification strategy in later parts of the paper. Second, it highlights the questions and trade-offs of interest for the policy-maker that will guide the counterfactual analysis presented at the end of this paper.

2.2.1 The deferred-acceptance algorithm: theoretical properties and tradeoffs.

School choice problems. Formally, a *school choice problem* (Abdulkadiroğlu and Sönmez (2003)) consists of two finite sets: a set of N students, and a set of J schools (or programs). Each school has a finite capacity that determines how many students it can enroll. Students have preferences over schools, while schools rank students by order of priority for admission.[4] Priority orders can be common or differ across schools, may or may not be known to students, and are taken as given. In the empirical setting of this paper, priority is merit-based and determined as a function of past academic performance[5]; it is known to students. A solution to a school choice problem —that is, an allocation in which each student is assigned to at most one school and no school is assigned more students than its capacity— is called a

[4] *School choice* refers to one-sided many-to-one matching problems; while *college admissions* refer to two-sided many-to-one matching problems. In the context of college admissions, students and schools both have preferences over the other side of the market.

[5] In particular, the priority ranking is fine, rather than coarse. When ties occur, they rarely involve more than a handful of students.

matching. A *mechanism* is a systematic rule or procedure that, given any school choice problem, selects a matching. In general, centralized school choice mechanisms involve (1) students simultaneously submitting an ordered list of academic programs to attend; and (2) a central authority assigning students to programs according to a pre-specified rule or algorithm. Because it determines the school or academic program students attend, a centralized mechanism can have significant consequences on students' outcomes such as their academic achievement (e.g. Kapor et al. (2016)). A substantial theoretical literature has been guiding policy-makers in their choices of what mechanism to use by studying their properties, and highlighting some of them as desirable.

Desirable properties for matching mechanisms. Three properties have acquired a central place in the theoretical literature on matching —stability, strategy-proofness and efficiency. Here, I define them and discuss their desirability. In the context of school choice, a matching is *stable* if no student is matched to a school over which she prefers not being matched (it is *individually rational*); and if no student prefers to her assignment a school which has a vacancy in the final match (it is *non-wasteful*), or which admitted a student with lower priority than her (it is *justified-envy free*). A mechanism is stable if it always selects a stable matching. A mechanism being stable means that the outcome will be fair, in the sense that no student will lose a seat at a desired school to a student with lower priority than her at this school. It also means that the implementation of the outcome will be successful, in the sense no student-school pair will be willing to block the final assignment —empirical evidence indeed seems to suggest that failure of stability is a key reason why some mechanisms have been abandoned in practice (Roth (2008)).

A mechanism is *strategy-proof* if for all agents, truthfully reporting one's preferences over schools is always a weakly dominant strategy. A mechanism being strategy-proof means that the application game is easy to play for families. Mechanisms in which manipulating one's preferences can be profitable may put at a disadvantage students who are not able to strategize, or do not strategize well (Abdulkadiroğlu et al. (2006b)). Moreover, there is empirical evidence that students' ability to play the game induced by the assignment mechanism depends on demographics, such as their socioeconomic background (Kapor et al. (2016)). A manipulable mechanism can then possibly foster the persistence of inequalities from one generation to the next. By contrast, strategy-proofness enables the policy-maker to purse "transparency, fairness, and equal access to public facilities" (Abdulkadiroğlu et al. (2006b)). In addition, application lists submitted by students under a strategy-proof mechanism constitute reliable data on families' preferences, which can inform broader policy-making (Abdulkadiroğlu et al. (2006b)).

A matching μ Pareto-dominates another matching ν if every student weakly prefers her assignment under μ over her assignment under ν, and at least one student strictly prefers her assignment under μ over her assignment under ν. A matching is *Pareto-efficient* if it Pareto-dominates all other matchings. A mechanism is Pareto-efficient if the matching it selects always Pareto-dominates the matching selected by other mechanisms. A mechanism being Pareto-efficient means that no welfare is wasted, in the sense that no student could be made better off without hurting someone else.

The deferred-acceptance algorithm. The deferred-acceptance algorithm (DA) is strategy-proof, stable, and Pareto-dominates[6] all other strategy-proof and stable mechanisms (Gale and Shapley (1962); Dubins and Freedman (1981); Roth (1982)).

[6] The matching produced by the DA Pareto-dominates all other stable matches if priorities are strict (i.e. there are no ties). If ties must be broken, the resulting match may not be Pareto-optimal among stable matches (Erdil and Ergin (2008)).

Based on these theoretical properties, its use has been recommended over other mechanisms (e.g. the Boston mechanism) in the school choice context (Balinski and Sönmez (1999); Abdulkadiroğlu and Sönmez (2003); Abdulkadiroğlu et al. (2006b)).

The DA algorithm introduced by Gale and Shapley (1962) takes in two sets of inputs. For each school, a priority ranking of all students over admission offers; and for each student, a preference ranking (the *application list*) of *all* schools of the choice set, from most to least preferred. In the simple case when a unique priority ordering is used by all schools, the DA proceeds as follows, once all students have submitted application lists:[7]

> **DA**
>
> *Step 1/* The first-ranked student is assigned to her first-listed program.
>
> *Step (k+1)/* For any $k \geqslant 1$, once the k^{th} student in the priority ranking has been assigned, the student ranked $(k+1)^{th}$ is assigned to the highest-ranked element of her list that still has a vacancy. If all of her listed choices are full at that point, she is left unassigned and the algorithm proceeds to the next student.
>
> *Stop/* The algorithm stops after all students have been processed.

Trade-offs. While Pareto-efficient among stable and strategy-proof mechanisms, DA is not efficient (Abdulkadiroğlu and Sönmez (2003)). Elimination of justified envy requires, when two students have the same ordinal preferences over two seats, the higher-priority student to be assigned his more-preferred school, regardless of the cardinal intensities of students' preferences. If, for instance, the lower-priority student likes more the preferred seat (or dislikes more the less-preferred seat) than

[7] This simple case is the one relevant for the empirical analysis in this paper. When a unique priority ordering is used by all schools, the DA boils down to the so-called *serial dictatorship* algorithm. A more general version of the DA allows for school-specific priorities. It is not directly relevant for the empirical analysis in this paper; it is described in Appendix ??.

the higher-priority student does, elimination of justified envy can create a welfare loss. Hence, when choosing to implement the DA, the policy-maker demonstrates her willingness to pursue elimination of justified envy and strategy-proofness, and to pay the cost of foregoing efficiency.

2.2.2 List restrictions, uncertainty: implementation constraints and consequences

Versions of the DA are used in many places to assign students to schools (e.g. in NYC, Chicago, Paris but also nationwide in Turkey, Ghana) or colleges (e.g. in Turkey, Taiwan, Tunisia). Most implementations feature one common departure from the theoretical set-up: the size of application list students may submit is restricted to be strictly smaller than the size of the choice set. For instance, in NYC students can list 12 of the 500+ public high schools programs offered in the city; in Ghana, students can apply to 6 of the 1,900+ high school programs in the country.[8] Under such list-size restrictions, the DA *a priori* not strategy-proof. In a restricted-list application setting, students face the possibility of not being assigned to any school, if they get rejected from all the schools they apply to. A student who expects her most-preferred schools to be popular among higher-priority students, may then decide not to submit an application list that truthfully reflects her ordinal preferences over programs, and instead include less-preferred, safer schools. As a consequence of strategic reporting, the final matching may not be stable (with respect to the students' true preferences), and some welfare may be lost. For instance, a student may decide not to apply to a preferred program if she thinks her chances of receiving an offer are low, and then end up being assigned to a less-preferred school, while, *ex post*, the preferred program would have had a seat available for her.

[8] Implementations in Boston and Romania are exceptions. For more examples and details on restrictions, see Appendix E in Fack et al. (2015), or `matching-in-practice.eu`.

In the paragraphs below, I describe the student's problem in a restricted-list setting, and illustrate consequences of the uncertainty faced by students on their incentives to be truthful and on welfare by an example. As they will be useful in the rest of this paper, I also review a couple of simple theoretical results on truth-telling and dominant strategies in the restricted-list setting.

Uncertainty and strategic incentives

The students' problem. At the time of application, i is assumed to know the flow utility she would derive from any element of her list. The only uncertainty she faces is due to her not knowing which element of her list (if any) she will gain admission to. While she does not know which program she will be offered admission to, she has (subjective) beliefs about her probability to be admitted to the different programs. She maximizes her subjective expected utility —the weighted sum of the flow utilities of elements of her ordered application list, with weights equal to her perceived admission chances to these programs —within the set of all ordered lists of up to M alternatives:

$$EU_i(\mathcal{L}_i) = \sum_{k=1}^{M} \Big[\pi_i(\mathcal{L}_i(k)) \times u_i(\mathcal{L}_i(k))\Big] + \overline{\pi_i} \times V_i(0) \tag{2.1}$$

where $\pi_i(\mathcal{L}_i(k))$ denotes i's expectations about her *admission* chances to the k^{th}-ranked element of her application list, $u_i(\mathcal{L}_i(k))$ denotes the flow utility derived from admission to this k^{th}-ranked element, and $V_i(0)$ denotes the option value of being left unassigned.[9] $\overline{\pi_i}$ denotes student i's probability to be rejected from all her listed

[9] If a student who fails to be assigned to any element of her list is left unassigned, the value of unassignment $V_i(0)$ simply corresponds the value of the outside option. Alternatively, if students who fail to be assigned to any element of their list then participate in a secondary application procedure, the option value $V_i(0)$ is equal to the i's expected utility to be derived when participating to this secondary procedure.

choices.[10]

Example 1 below shows that when they face uncertainty about their admission chances and can only apply to a subset of their choice set, it may be optimal for students to submit an ordered list that does not coincide with their most-preferred programs.

Example 1. *Suppose there are two programs A and B, each with two seats. Suppose there are three students, ranked from 1 to 3 by strict priorities. Students know their priority ranking; and that there are twice two seats to be apportioned. Preferences for programs are private information, but their distribution is common knowledge:*

$$u_{iA} = 6.35 + \varepsilon_{iA}$$

$$u_{iB} = 5 + \varepsilon_{iB}$$

where ε_{iA}, $\varepsilon_{iB} \sim$ *i.i.d.* $N(0,1)$.[11] *Suppose students can only apply to one program, and that students who do not get assigned to any program obtain the outside option, that yields a value of 0.*

Student 1 knows she has highest priority, and that, hence, she will be assigned by the DA to the program she ranks first in her list. It is strictly dominant for her to (truthfully) list her most-preferred program in her application list. Student 2 knows she is ranked second. She knows she will be assigned to her first-ranked element since no matter which school Student 1 gets assigned to, both schools still have at

[10] $\overline{\pi_i}$ corresponds to the joint probability of $i$not clearing, *ex-post*, the admission cutoff of all her listed choices. I call Student i (*ex-post*) *eligible* to program ℓ if program ℓ has at least one open seat when it is i's turn to be considered for assignment by the DA algorithm —that is, after all students with higher priority score than i have been assigned (or kept aside for a leftover spot), and none of the students with lower priority score than i has been considered for assignment. When assignments are made via the DA algorithm, the *admission* of Student i to program ℓ requires that (i) i has listed ℓ in her application ranking; (ii) i is *eligible* to program ℓ; (iii) i is *non*-eligible to all programs ranked above ℓ in i's ordered application list. Hence, the need to distinguish between between *eligibility* and *admission* probabilities.

[11] Note that $6.35 = 5 + 2 \times \alpha_N(.75)$, with $\alpha_N(.75)$ such that: $Pr(X < \alpha_N(.75)) = .75$ if $X \sim N(0,1)$.

least one remaining vacancy. It is strictly dominant for her to (truthfully) list her most-preferred program in her application list. Student 3 knows that two seats are taken, and one program may be full by the time the algorithm processes her list. If she happens to list a program that is full, she will be left unassigned and get utility 0. She solves the maximization problem:

$$u_3 = \max_{s \in \{A,B\}} \{p_{3A} \cdot u_{3A};\ p_{3B} \cdot u_{3B}\}$$

where her eligibility chances to A and B respectively, are given by the distribution of preferences: $p_{3A} = 1 - .75 \times .75 = .4375$*; and the probability that there is a seat available in Program B is* $p_{3B} = 1 - .25 \times .25 = .9375$. *Denote* s_3^* *the school Student 3 applies to:*

$$s_3^* = \begin{cases} A & \text{if } \varepsilon_{3B} \leqslant \frac{p_{3A}}{p_{3B}} \times (6.35 + \varepsilon_{3A}) - 5 \\ B & \text{otherwise} \end{cases}$$

There is no general dominant strategy in the application game when one can only apply to $M < J$ alternatives, and little can *a priori* be said about students' behavior in such setting. The next two propositions give partial characterizations of students' behavior that will be useful in the rest of the paper. While reporting on one's list a vector of programs that differs from one's most-preferred vector may be dominant, Proposition 1 (Haeringer and Klijn (2009)) establishes that one never benefits from ranking the reported alternatives differently than by decreasing order of preference.

Proposition 1. *[Haeringer and Klijn (2009)] (a) If a student finds at most M schools acceptable, then she can do no better than submitting her true preferences.*
(b) If a student finds more than M schools acceptable, then she can do no better than employing a strategy that selects M schools among the acceptable schools and ranking them according to her true preferences.

The next proposition establishes a sufficient condition for truth-telling to be a dominant strategy. A proof is given in Appendix **??**.

Proposition 2. *(a) Condition 1 (below) is a sufficient condition for students not to have a strict incentive to misreport their preferences over their choice set.*
(b) Under Assumption 1 (below), Condition (1) is a sufficient condition for students not to misreport their preferences over their choice set.

Condition 1. *Student i has a perceived* eligibility *probability 1 for (at least) one of her M most-preferred programs.*

Assumption 1. *When indifferent between doing so or not, a student does not misrepresent her unconstrained preference ranking. In other words, a student does not report her most-preferred programs in her application list only when it is* strictly *profitable to do so.*

Uncertainty and welfare

Example 1 illustrates the way in which uncertainty can generate inefficiencies *ex post*. Student 3 may not be assigned to her most-preferred program *ex post* available if she does not apply to it —even though not applying to it may be optimal *ex ante*. For instance, consider a case in which $-2.037 < \varepsilon_{3B} - \varepsilon_{3A} < 1.35$,[12] and $\varepsilon_{2B} \geqslant 1.35 + \varepsilon_{2A}$. Student 3 prefers A to B, but finds it *ex ante* optimal to apply to B. Student 2 prefers B to A, therefore applies to B and gets in. In this case, Student 3 gets either assigned to B (if Student 1 got in A) or is left unassigned (if Student 1 got in B) while she prefers Program A to both these alternatives.

2.2.3 Sequential implementation and information revelation

In this subsection (and in the rest of this paper), I take as given and fixed any restriction on the size M of the list to be submitted. I describe a simple alternative implementation of the DA in which information about available seats is regularly

[12] that is, $\varepsilon_{3B} < 1.35 + \varepsilon_{3A}$, and $\varepsilon_{3B} > -5 + \frac{p_{3A}}{p_{3B}} \times (6.35 + \varepsilon_{3A})$

publicly updated. I explain how this version of the DA, while easy to implement, may partially restore incentives for truthful reporting and increase welfare, relative to the standard single-phase, restricted-list DA.

Sequential implementation of the DA. The standard (one-phase) implementation of the DA involves the whole cohort of N students simultaneously submitting their application lists, and then being assigned via the DA. In contrast, a sequential implementation involves first dividing the cohort in $K \leqslant N$ assignment groups that successively submit lists and are assigned. In the case in which the same priority order is used by all schools, as is relevant for the empirical analysis in this paper, the division of the cohort can be straightforwardly made along this priority order. Suppose the N students are ranked by a strict priority order from 1 to N. Let $K_1, K_2, \ldots, K_K$ be the sizes of each of the K groups to be created. Assign students with priority ranks 1 to K_1 to Group 1, students with priority ranks K_1+1 to K_1+K_2 to Group 2, etc. Priority order is preserved within groups. Given these groups, the assignment procedure goes as follows:

K-phase DA

Phase 1/ The number of seats open in each program is publicly revealed. Group 1 students submit application lists, and are then assigned using the DA algorithm.

Phase (n+1)/ For any $1 \leqslant n \leqslant (K-1)$, vacancies remaining after the assignment of Group n students are publicly revealed. Unassigned Group n students are added at the top of Group $n+1$ according to their initial priority order. Group $n+1$ students submit application lists, and are then assigned using the DA algorithm.

In the limit, if K is equal to the number of students, sequentially implementing the DA puts students in a perfect information setting and is equivalent to allowing for the submission of an unrestricted list.

Information can restore incentives for truthful reporting

Example 2 illustrates the sequential implementation of the restricted-list DA and how it can restore, for some students, incentives for truthful reporting.

Example 2. *Consider again the setting of Example 1, keeping unchanged the programs, vacancies, priority order and preferences. Suppose however that the pool of applicants is divided into two groups* $\{1, 2\}$ *and* $\{3\}$*. The application procedure is ran sequentially, in two phases, and the information about vacancies is updated between the two phases. In Group 1, Student 1 and Student 2 face the exact same application problem as in Example 1. They both submit their list as described in Example 1. Before Student 3 submits her application list, the information about vacancies is updated, and Student 3 knows she will be assigned first in Group 2. It is therefore dominant for her to truthfully report in her application list her most-preferred program among those that have not been publicly declared full.*

In each group, the student ranked first faces perfect information when applying. She knows exactly which programs have available seats, and therefore optimally apply by truthfully listing her most-preferred programs among those that have not been publicly declared full. Proposition 2 shows that other students at the top of each group, beyond the very first student, may also face incentives to be truthful after information is revealed.

Information and welfare

Example 2 illustrates one way in which revelation of information, via a sequential implementation of the assignment procedure, can improve welfare. It restores, for some students (e.g. Student 3), a choice situation similar to perfect information. Thereby, it ensures that these students always apply and get assigned to their most-preferred programs among those with remaining seats, rather than possibly being assigned to *ex post* suboptimal programs, or failing to be assigned.

More generally, the revelation of information can (weakly) increase welfare even if it does not fully restore incentives for truth-telling. From a revelation of information about remaining seats, rational students update their expectations about their eligibility chances to the true conditional (on the information received) distribution governing unobservables. Expected utility maximization based on this conditional distribution allows them to choose an application list that is better *ex ante* (given the pre-information revelation realizations of unobservables) than the list they would choose without the information. This is illustrated by Example 3.

Example 3. *Consider again the setting of Example 1, keeping unchanged the programs, vacancies, priority order and preferences. Suppose however that the pool of applicants is divided into two groups* $\{1\}$ *and* $\{2,3\}$*. The application procedure is ran sequentially, in two phases, and the information about vacancies is updated between the two phases. In Group 1, Student 1 faces the exact same application problem as in Example 1, and applies truthfully. Student 2 faces a similar application problem as in Example 1, and applies truthfully. The information revelation allows Student 3 to update her beliefs, but does not fully restore incentives for her to be truthful. Student 3 chooses her application list by solving:*

$$\tilde{V}_3 = \max_{s \in \{A,B\}} \{\tilde{p}_{3A} \cdot u_{3A};\ \tilde{p}_{3B} \cdot u_{3B}\}$$

where $\tilde{p}_{3A}$ *and* $\tilde{p}_{3B}$ *are Student 3's expected eligibility chances to A and B, conditional on the information she received about Student 1's assignment:*

$$(\tilde{p}_{3A}, \tilde{p}_{3B}) = \begin{cases} (.25, 1) & \textit{if Student 1 chose A, i.e. } u_{1A} > u_{1B} \\ (1, .25) & \textit{otherwise.} \end{cases}$$

Denote $\tilde{s}_3^*$ *the school Student 3 applies to, conditional on the information she received* *about Student 1's assignment:*

$$\tilde{s}_3^* = \begin{cases} \tilde{s}_3^{*(a)} & \textit{if Student 1 chose A, i.e. } u_{1A} > u_{1B} \\ \tilde{s}_3^{*(b)} & \textit{otherwise.} \end{cases}$$

Ex ante welfare under the information scenario of Example 1 writes:

$$W = \int_{\mathbf{e}} \Big[\max\{u_{1A};\ u_{1B}\} + \max\{u_{2A};\ u_{2B}\} + u_3(s_3^*) \Big] d\mathbf{e}$$

$$= \int_{\mathbf{e}_1} \max\{u_{1A};\ u_{1B}\} d\mathbf{e}_1 + \int_{\mathbf{e}_2} \max\{u_{2A};\ u_{2B}\} d\mathbf{e}_2 + \int_{\mathbf{e}} u_3(s_3^*) d\mathbf{e}$$

Ex ante welfare under the present information scenario writes:

$$\tilde{W} = \int_{\mathbf{e}} \Big[\max\{u_{1A};\ u_{1B}\} + \max\{u_{2A};\ u_{2B}\} + u_3(\tilde{s}_3^*) \Big] d\mathbf{e}$$

$$= \int_{\mathbf{e}_1} \max\{u_{1A};\ u_{1B}\} d\mathbf{e}_1 + \int_{\mathbf{e}_2} \max\{u_{2A};\ u_{2B}\} d\mathbf{e}_2 + \int_{\mathbf{e}} u_3(\tilde{s}_3^*) d\mathbf{e}$$

To see why $W \leqslant \tilde{W}$, decompose the expected indirect utility of Student 3 in each case:

$$\int_{\mathbf{e}} u_3(s_3^*) d\mathbf{e} = Pr(u_{1A} > u_{1B}) \int_{\mathbf{e}|(u_{1A} > u_{1B})} u_3(s_3^*) d[\mathbf{e}|(u_{1A} > u_{1B})]$$

$$+ Pr(u_{1A} \leqslant u_{1B}) \int_{\mathbf{e}|(u_{1A} \leqslant u_{1B})} u_3(s_3^*) d[\mathbf{e}|(u_{1A} \leqslant u_{1B})]$$

and

$$\int_{\mathbf{e}} u_3(\tilde{s}_3^*) d\mathbf{e} = Pr(u_{1A} > u_{1B}) \int_{\mathbf{e}|(u_{1A} > u_{1B})} u_3(\tilde{s}_3^{*(a)}) d[\mathbf{e}|(u_{1A} > u_{1B})]$$

$$+ Pr(u_{1A} \leqslant u_{1B}) \int_{\mathbf{e}|(u_{1A} \leqslant u_{1B})} u_3(\tilde{s}_3^{*(b)}) d[\mathbf{e}|(u_{1A} \leqslant u_{1B})]$$

By $\left(\tilde{s}_3^{*(a)}, \tilde{s}_3^{*(b)}\right)$ being solution to the conditional optimization problem faced by Student 3:

$$\int_{\mathbf{e}|(u_{1A} > u_{1B})} u_3(\tilde{s}_3^{*(a)}) d[\mathbf{e}|(u_{1A} > u_{1B})] \geqslant \int_{\mathbf{e}|(u_{1A} > u_{1B})} u_3(s_3^*) d[\mathbf{e}|(u_{1A} > u_{1B})]$$

and

$$\int_{\mathbf{e}|(u_{1A} \leqslant u_{1B})} u_3(\tilde{s}_3^{*(b)}) d[\mathbf{e}|(u_{1A} \leqslant u_{1B})] \geqslant \int_{\mathbf{e}|(u_{1A} \leqslant u_{1B})} u_3(s_3^{*(b)}) d[\mathbf{e}|(u_{1A} \leqslant u_{1B})].$$

On the other hand, the revelation of information, and the possibly induced incentives for truthfullness may have an negative effect on welfare. By restoring incentives for truthful-reporting, the revelation of information increases the probability that any student gets assigned to her most-preferred program among those that are available *ex post*. As explained in Section 2.2.1, elimination of justified envy can conflict with welfare maximization.

2.2.4 The value of information: an empirical question

Beyond perfect knowledge of true admission probabilities. The existence of gains from information revelation, both in terms of incentives for truthfulness and welfare, crucially depends on students' ability to understand the information they are given, to update their expectations about their admission chances given this information. Example 3 shows that the revelation of information in the sequential DA (weakly) increases welfare when students are *perfectly rational*, that is, able to perfectly update their beliefs to their true conditional eligibility chances, from the information revealed about remaining seats. While straightforward in this three-student, two-school example, in which mean utilities and the distribution of unobserved preferences are common knowledge, the expectations-formation problem can get hard as the choice set gets large. Previous studies in the empirical school choice literature have recognized that, even given the distribution of preferences, deducing one's probability of admission to all programs is a hard problem, which high-school students and their families may not be able to solve (e.g. Agarwal and Somaini (2014); Calsamiglia et al. (2014); Ajayi and Sidibé (2016); Kapor et al. (2016)). In practice, assessing the effect on students' behaviors of providing information about vacancies requires testing whether students understand the information they are given, and characterizing the way they use it to update their expectations about their admission chances.

Magnitude of gains. Students' preferences over the programs in their choice set are a crucial determinant of the magnitude of gains, both in terms of incentives for truthfulness and welfare. Proposition 2 shows that whether a student may find optimal or not to be truthful in a restricted-list application scenario depends on her expectations about her admission chances to her most-preferred programs. The change in welfare induced by the change in students' application behaviors and assignments when more information is provided naturally depends on students' utility for the alternatives. Quantifying the change in welfare induced by the revelation of information therefore requires recovering students' preferences for programs.

Heterogeneity in gains. From a policy perspective, and given the *a priori* ambiguous effect of information revelation on expected welfare, it is important to characterize who may win or lose from the sequential implementation of the restricted-list DA. In addition, an empirical analysis allows to investigate differential effects of information across ability and demographic groups, which have been documented in other settings (e.g. Hoxby and Turner (2015)).

Frequency of informational updates. Finally, when it comes to implementation of the sequential DA, the policy-maker needs to decide on how many phases to implement —that is how frequently to reveal information. If the revelation of information may have benefits, the sequential implementation certainly has its costs too. A fully sequential implementation, with a number of phases equal to the number of students, would give applicants perfect information, and restore the desirable properties of the unrestricted-list DA. When the number of students is large, though, updating the number of vacancies after every single assignment can take a prohibitive amount of time. In addition, as shown by Example 2, a fully sequential implementation may not be needed to induce perfect information. The characterization of the optimal information revelation structure is beyond the scope of this paper, and the data does not allow to identify implementation costs of the assignment mechanism. However, the empirical analysis can provide evidence on the marginal effects of an extra revelation of information, as a function of the information already revealed. It can also inform the cost-effectiveness of information updates as a function of the position, in the priority ranking, at which they are provided.

2.3 The university match in Tunisia

This section introduces the empirical setting of this paper. It describes the data, and presents the practical features of the sequential implementation of the DA in Tunisia, to assign high-school graduates to universities at the nationwide level. Importantly, taking advantage of the cutoffs generated by the division of the applicant pool in group, it provides reduced-form evidence of the consequences of the sequential implementation on students' application behaviors and assignments.

2.3.1 Institutional background

Graduating from high school guarantees Tunisian students access to public higher education (graduating seniors are automatically registered in the centralized post-secondary application system). The particular program each student is allowed to enroll in is determined by a central assignment mechanism. Assignment is made according to a sequential variant of the DA algorithm, similar to that described in Section 2.2.3. Context-specific implementation features are presented now.

Priority score. In year 2010, a common priority ranking of students was used by all programs. A student's priority score was determined as a function of the student's grades at the various tests of the national end-of-high-school exam, which can be viewed as a standardized test. A student with a higher score is given priority over a student with a lower score for admission offers to the post-secondary programs. Students know their priority ranking.

Application groups. The application-assignment process is split into three successive phases. Namely, the cohort of applicants is divided into three groups based on their priority score —in this particular case: the top 30% of students ("Group 1" students), the middle 40% ("Group 2"), and finally the bottom 30% ("Group 3").

Public information. All high school graduates are given a handout containing information about the available post-secondary programs over the country. The handout indicates, for each existing program, the number of vacancies open for the next academic year and the past-year admission cutoff, that is, the priority score of the marginal student admitted in the previous year. After each group has gone through the assignment algorithm, the number of vacancies in each program is publicly updated, so next-group students are told which vacancies remain before submitting their application list.

Application lists. Students may submit an ordered list of up to 10 post-secondary programs.

Unmatched students. Application lists are processed using the DA. Students who fail to be admitted to any of their listed choices are pooled on top of the next application group —if there is one— and proceed to submitting a new application list after the information about vacancies is publicly updated.[13] If there is no next application group, unmatched students are administratively assigned to left-over seats.

[13] The new list is formed based on the programs available at the time it is submitted, and not based on the programs available when the student submitted her initial list. In the data, only the very last list submitted by each student is recorded.

2.3.2 Sample description

I use administrative data from the Tunisian Ministry of Higher Education and Scientific Research. The database contains the ordered application lists and assignment of all students applying to post-secondary programs in public institutions in Tunisia in 2010, as well as an identifier of the high school they graduated from and the grades obtained at the various tests of the national end-of-high-school exam. It also contains a limited number of demographic characteristics, such as gender, date and region of birth, and a category indicator of father's occupation. In this subsection, I describe the student sample and students' choice set of post-secondary programs, as well as patterns in their applications and assignments.

Students. In 2010, 82,748 students graduated from high school and were registered in the centralized post-secondary application system. The assignment procedure is run in parallel for each of the six majors students may graduate from high school with. Here, I focus on students graduating with the Math major.[14] They were 11,029 in 2010. Among them, I drop students for whom high school information and/or all end-of-high-school test scores are missing,[15] as well as the 65 students recorded in the data as not having submitting any application list[16]. Table 2.1 describes the 10,935 students in the final sample, as well as the division of the sample into the

[14] More detail about high school and high school majors in Tunisia is provided in Luflade and Zaiem (2016). The Math high school major is among those allowing students to pursue the widest range of fields of study in their post-secondary career. A similar and separate analysis could be made for the other high-school majors. The comparative analysis of application and updating behaviors across students who graduated from high school with different majors is part of a future project.

[15] High school information and/or all end-of-high-school test scores are missing for 25 students. In addition, I drop 32 students, whose application lists comprise only programs out of their choice set (that is, publicly declared full before these students submit their list) and/or some of the six programs I drop because they did not exist, nor have an equivalent existing program, in the previous year.

[16] Among the 65 students recorded not to submit an application list, 3 are in Group 1, 18 in Group 2, and 44 in Group 3.

three application groups. The ratio of sexes is roughly constant across groups; and slightly more than half of sample is female. High-SES students represent 60% of the student sample; and their share decreases along the priority ranking. The represent 78% of Group 1 students, 58% of Group 2, and 47% of Group 3. A similar pattern is observed for geographical origin. Students from Tunis, and from the dynamic coastal regions account for 30% and 48% of the sample, respectively; their respective shares are highest in Group 1, and decrease along the priority ranking.

Programs. In 2010, 616 post-secondary programs had seats available for students who graduated high-school with a Math major. 54 of them filled up by the end of the first assignment phase; 326 by the end of the second phase; and 100 (16%) did not get assigned as many students as allowed by their capacity. Table 2.2 shows programs characteristics, and illustrates the changes in the choice set faced by students as the sequential assignment procedure moves from one phase to the other. Programs are offered in 10 fields of study,[17] four of them in STEM. Seats in STEM represent 67% of initially offered seats; 66% of the seats still available at the beginning of the second phase; and 60% of the seats remaining at the beginning of the third phase. About two thirds of initially offered seats are in programs preparing to the equivalent of a Bachelor degree (*Licence*), to be earned after the successful completion of three years of classes. There are two types of Bachelor degree: '*Licence appliquée*' (LA), which prepares students who plan to enter the labor market after graduation; and '*Licence fondamentale*' (LF), which prepares students who plan to pursue their education (in a Master's program) after graduation. The remaining third of initially offered seats are in programs preparing students to more advanced degrees, essentially in

[17] Each post-secondary program is associated a six-digit code that reflects the fields classification defined by the Ministry of Higher Education. I use this classification. The ten fields are: Humanities; Arts; Education (incl. Physical Education); Economics/Business/Management; Social Sciences; Law; Health and Life Sciences; Earth Sciences; Physics/Chemistry/Engineering; and Math/Computer Science.

engineering and medical fields. By the end of the first and second phases, seats in advanced-degree programs represent only 19 and 3% of the seats still available, respectively. About a third of the seats initially available are offered in Tunis; they represent only 7% of the seats still available at the end of Phase 2. In contrast, 18% of initial seats are in western and southern regions of the country; while they constitute 36% of the seats that remain vacant at the beginning of Phase 3.

Application behaviors and assignment patterns. Students included an average of 9.1 programs in their list out of an allowed maximum of 10. 70% of students rank 10 programs.[18] On average, applicants got admitted to their second or third choice (2.6^{th} choice), with 39% of them admitted to their first-listed choice, and 76% of them to one of their five first-listed choices. Although shares vary, the pattern is the same across groups —45% of Group 1 students are admitted to their first-listed choice, against 35% of Group 3; 86% of Group 1 students are admitted to one of their five first-listed choices, against 65% of Group 3. 4% of the students fail to be assigned to any element of their initial list. Half of them (and most of Group 1 and Group 2 unassigned students) were assigned in a later round, to a program they listed when pooled in the next group. The other half (essentially Group 3 unassigned students) ended up being administratively assigned.

2.3.3 Local effects of the sequential implementation of the DA on application behaviors and assignments

Figure 2.1 illustrates the changes in behaviors observed at the information revelation cutoffs. I plots, as a function of students' priority, the rank in their application list of

[18] The small and selected share of students listing strictly fewer choices than they are allowed to prevents from using an identification argument similar to the one used by Abdulkadiroğlu et al. (2017) when trying to recover students' preferences for programs.

the choice they are assigned to. Top ranked students, at the left, are assigned to their first-listed choice. As priority goes down in Group 1, and as popular programs fill up, students get assigned to increasingly lower choices in their application portfolio. When updated information about vacancies is provided, as the limit between Groups 1 and 2, application behaviors change in such a way that students get assigned to their top-listed choice again. The cycle starts again until the next revelation of information, at the limit between Groups 2 and 3.

In this subsection, I further documents local application and assignment changes at the information revelation cutoffs. The division of the applicant pool into application subgroups creates cutoffs (henceforth *information revelation* cutoffs or *group* cutoffs) that I use in a sharp regression discontinuity (RD) design in order to document the local effects of providing updated information to applicants.[19]

The effect of a change in groups on any outcome Y of interest is estimated by local linear regression

$$\min_{\alpha,\beta,\tau,\gamma} \sum_{i=1}^{N} \mathbf{1}_{[c-h\leqslant T_i\leqslant c+h]} \cdot \Big(Y_i - \big[\alpha + \beta(T_i - c) + \Delta\mathbf{1}_{[T_i<c]} + \gamma(T_i - c)\mathbf{1}_{[T_i<c]}\big]\Big) \quad (2.2)$$

where T_i is student i's priority score (running variable), c denotes the group cutoff, and h is the estimation bandwidth.[20] $\mathbf{1}_{[T_i<c]}$ is a indicator of i being assigned to the 'informed group' (that is, Group 2 at the Group 1/ Group 2 cutoff; and Group 3 at the Group 2/ Group 3 cutoff) or not, $(T_i - c)$ is the distance of i's score to the group cutoff, and $(T_i - c)\mathbf{1}_{[T_i<c]}$ is an interaction term that allows the slope (of the

[19] Standard graphical evidence supporting the sharpness and validity of the RD design can be found in Appendix A.2.

[20] For each outcome and subsample, the 'optimal' bandwidth is chosen using the Imbens and Kalyanaraman (2011) method and may vary from one outcome to another.

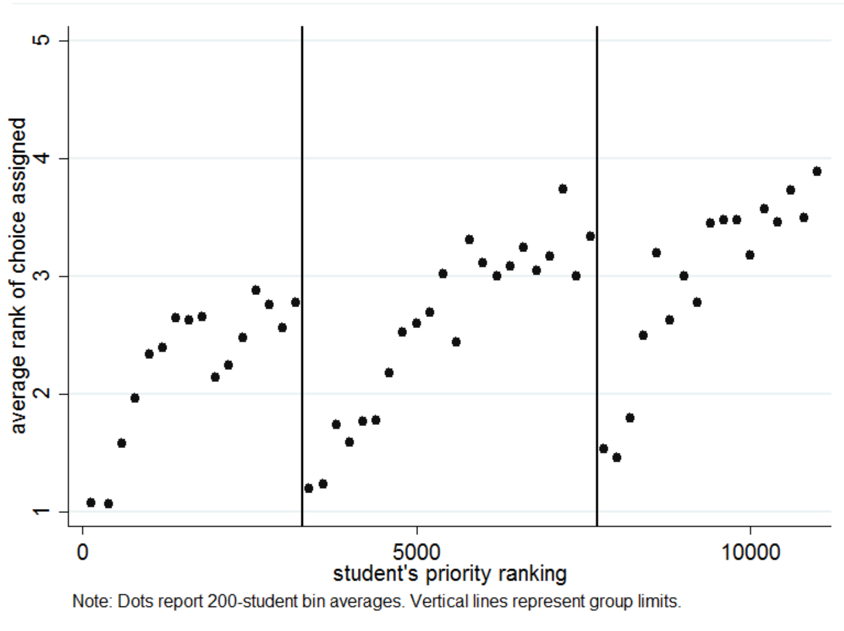


FIGURE 2.1: Choice assigned as a function of priority

outcome as a function of distance to the cutoff) to differ on either side of the group cutoff. Δ is the coefficient of interest, it measures the change outcome Y induced by the revelation of information.

I estimate (2.2) using as Y various application list characteristics (e.g. length, selectivity measures), and assignment outcomes (e.g. probability of assignment, rank of the choice assigned). In addition, to understand further how changes in application rates at the cutoffs correlate with the information revealed about programs, I estimate, for each program j, the change in application rate at the group cutoffs, using the binary indicator of whether or not student i ranked the track in her list as the dependent variable $Y_i^{(j)}$ in equation (2.2). I then regress the estimated change in application rate on various program characteristics. Results are shown in Tables 2.4 and 2.5, and summarized below.

Application behaviors. (1) *Marginally informed students submit shorter and less selective application lists than marginally uninformed students.* The top panel of Table 2.4 shows that, as compared to students that are marginally non-informed, marginally informed students list slightly fewer choices —.26 fewer at the Group 1/ Group 2 cutoff, and .63 fewer at the Group 2/ Group 3 cutoff, with only the latter difference being statistically different from 0 at conventional levels. Marginally informed students apply to programs that are less selective than their marginally uninformed counterparts. For instance, the most selective of their choices, has a past-year cutoff that is about .4 standard deviation lower, which corresponds to 9 percentiles of the priority distribution at the Group 1/ Group 2 cutoff, and 11 percentiles at the Group 2/ Group 3 cutoff. Interestingly, the same is true for safe choices as well. The least selective choice listed by marginally informed students has a past-year cutoff that is about .7 percentiles lower in the priority distribution than that of their marginally informed counterparts.

(2) *Marginally informed students increase their application rate to safer and more popular programs among those with remaining vacancies.* The top panel in Table 2.5 shows that a program being declared full (for the first time) at the group cutoff is correlated with a drop in application rate —by 14 and 5 percentage points at the Group 1/ Group 2 and Group 2/ Group 3 cutoffs, respectively. The middle and bottom panels show that, for programs that are declared full, the magnitude of the drop in application rates increases with the program's initial number of vacancies, and its selectivity level. Symmetrically, for programs that are not full, a larger number of remaining vacancies and a higher past-year cutoff are correlated with a larger surge in application rates.

Assignment patterns. (3) *Marginally informed students are more likely to be assigned to an element of their list than their marginally uninformed counterparts.* The bottom panel of Table 2.4 shows that students' probability to be actually assigned to one of their listed choices, rather than being rejected from all of them, is increased by 9 percentage points at the Group 2/ Group 3 cutoff, and by 2 percentage points at the Group 1/ Group 2 cutoff (but this latter effect is not statistically different from 0).

(4) *Marginally informed students are assigned to higher-ranked elements of their lists than marginally uninformed students.* The bottom panel of Table 2.4 also shows that marginally informed students end up clearing the *ex-post* admission cutoff of a larger share of their listed choices (+32% at the Group 1/ Group 2 cutoff, and +58% at the Group 2/ Group 3 cutoff) that marginally uninformed students do. This induces them to be assigned to a higher-listed element of their list —1.9 and 2.4 ranks higher at the Group 1/ Group 2 and Group 2/ Group 3 cutoffs, respectively.

Discontinuities in application behaviors at the information-revelation cutoffs are evidence of students' lack of perfect foresight and use of information. Validity of the RD design means that the assignment to students in one group or the next is, locally, as good as random: students on either side of a group cutoff have, on average, the same observable and unobservable characteristics. In particular, they also have, on average, the same preferences for post-secondary programs. As a consequence, if students were not using the information they are given, or if students had perfect foresight (in which case they would be able to predict the information they are given, which would then be redundant), application behaviors would not to change discontinuously at the cutoffs. In addition, Table 2.5 shows that the changes in application rates are consistent with students understanding and using the information they are given at the group cutoffs.

However, this reduced-form analysis does not inform whether students actually *benefit* from the informational updates. It does not inform either about the behavior and gains of students located further away from the information-revelation cutoffs in the priority ranking. Conducting a welfare evaluation of the effects of information provision requires comparing how students fare (here, in terms of indirect utility, which students derive from the program they are assigned to) under alternative counterfactual scenarios of information revelation. Performing the comparison requires simulating students' applications and assignment under the alternative scenarios. Generating students' application lists in turn requires to know the flow utility they associate with each program, and to understand how they derive beliefs about their admission chances from the available information. In the next section, I recover students' preferences for post-secondary programs. In Section 2.5, I turn to characterizing students' perceived admission chances. Finally, in Section 2.6, I present the results of the counterfactual analysis.

Table 2.1: Descriptive statistics: students

	All		Group 1		Group 2		Group 3	
	Mean	Sdev	Mean	Sdev	Mean	Sdev	Mean	Sdev
Demographics								
Female	.53	.50	.52	.50	.54	.50	.52	.50
Birth year	1991	.91	1991	.39	1991	.68	1990	1.23
High SES	.60	.49	.78	.41	.58	.49	.47	.50
From Tunis	.30	.46	.33	.47	.30	.46	.27	.44
From Coast (not Tunis)	.48	.50	.53	.50	.49	.50	.43	.49
From West/Interior	.19	.39	.13	.36	.18	.39	.26	.47
From South	.03	.17	.01	.11	.03	.17	.05	.22
Priority and academic performance								
Raw priority score	123.2	29.0	160.2	13.2	119.3	11.0	91.1	6.9
Stdized priority score	0	1	1.28	.46	-.13	.38	-1.10	.34
STEM high-sch. perf.	0	.85	1.04	.39	-.09	.40	-.92	.33
non-STEM h.-s. perf.	0	.79	.75	.54	-.07	.58	-.66	.59
Applications								
List 10 choices	.70	.46	.67	.47	.76	.43	.65	.48
# of choices listed	9.02	1.8	8.86	1.9	9.3	1.5	8.86	1.9
All choices in 1 field	.06	.24	.03	.16	.09	.28	.07	.26
$\geqslant$ 75% ch. in 1 field	.20	.40	.11	.32	.25	.43	.22	.41
All choices in STEM	.36	.48	.19	.40	.46	.50	.40	.49
All choices in 1 univ.	.03	.18	.01	.09	.04	.20	.05	.21
$\geqslant$ 75% ch. in 1 univ.	.08	.27	.02	.13	.09	.29	.12	.33
All choices in 1 region	.26	.44	.10	.31	.35	.48	.31	.46
Assignments								
Admitted to 1st choice	.39	.49	.45	.50	.39	.49	.35	.48
Admitted to 2nd choice	.15	.36	.16	.36	.16	.37	.12	.33
Admitted to 3rd choice	.10	.30	.11	.31	.11	.31	.07	.26
Admitted to 4th choice	.07	.26	.08	.28	.07	.26	.06	.24
Admitted to 5th choice	.05	.22	.06	.23	.05	.22	.05	.21
Admitted to 6th choice	.04	.19	.05	.20	.04	.20	.03	.18
Admitted to ch. # $\geqslant$ 7	.06	.12	.02	.09	.07	.12	.11	.16
Admin. assigned	.02	.14	0	.06	0	.07	.06	.23
Assigned in later round	.02	.14	.02	.13	.04	.19	0	0
Sample size	10,935		3,299		4,384		3,252	

Note: In the second panel, STEM (resp. non-STEM) high-school performance is the unweighted average of the student's standardized scores at the Math, Physics, Natural Sciences, and Comp. Sci. (resp. English, French, Arabic, and Philosophy) tests of the end-of-high-school national exam.

Table 2.2: Descriptive statistics: programs

	Phase 1		Phase 2		Phase 3	
	Mean	Sdev	Mean	Sdev	Mean	Sdev
Filling up in 2010	.84	.36	.83	.38	.67	.47
Filling up in 2009	.90	.31	.89	.32	.80	.40
2009 cardinal cutoff	-.47	.79	-.62	.63	-1.07	.44
2009 ordinal cutoff	.61	.25	.66	.21	.81	.15
# of seats	22	45	17	27	16	19
# at least 1 applicant	.73	.44	.98	.50	.98	.15
# of applicants	77	251	80	123	96	78
# of applicants/seat	2.9	4.9	8.7	16.5	10.0	10.6
Sample size						
Total # of programs	616		562		290	
Total # of seats	13,580		9,574		4,516	

Note: In the second panel, 2009 marginal admission score are shown conditional on programs filling up in 2009 –hence the change in sample size from 616 to 552. In the rest of the paper, for programs which did not fill up in 2009, the marginal admission score is set to the score of the very last student in the priority ranking (1 in percentiles terms).

Table 2.3: Descriptive statistics: programs

	Phase 1		Phase 2		Phase 3	
	Programs	Seats	Programs	Seats	Programs	Seats
Field						
Humanities	.14	.04	.14	.05	.08	.04
Arts	.08	.11	.08	.13	.08	.17
PE & Educ.	.02	.05	.01	.01	0	0
Social sci.	.05	.01	.05	.01	.07	.01
Econ. & Mgmt.	.15	.11	.02	.01	.20	.02
Law	.02	.01	.02	.01	.01	.01
Health & Life sci.	.10	.09	.08	.05	.01	.02
Earth sci.	.05	.03	.05	.04	.06	.04
Math & Comp. sci.	.12	.11	.12	.16	.15	.20
Phys., Chem., Engineer.	.28	.44	.28	.41	.32	.34
Degree						
Bachelor equiv. (LA)	.67	.36	.70	.48	.77	.59
Bachelor equiv. (LF)	.27	.29	.26	.33	.22	.39
Beyond Bachelor	.07	.34	.04	.19	.01	.03
Location						
In Tunis	.30	.30	.27	.21	.07	.07
On Coast (not Tunis)	.51	.52	.53	.58	.60	.56
In West/Interior	.16	.16	.17	.21	.29	.34
In South	.02	.01	.02	.01	.04	.02
Abroad	.01	0	0	0	0	0

Table 2.4: Reduced-form effects of informational updates on application behaviors and assignment patterns

	Groups 1/2 cutoff		Groups 2/3 cutoff	
	Change	Base level	Change	Base level
Application behaviors				
# listed choices	-0.264	9.081	-0.627***	9.289
	(0.256)	(1.754)	(0.239)	(1.506)
Obs.	727		917	
PYc. of most selective ch.	-0.094***	0.845	-0.110***	0.582
	(0.019)	(0.101)	(0.019)	(0.131)
Obs.	535		848	
PYc. of least selective ch.	-0.073***	0.437	-0.078***	0.132
	(0.024)	(0.180)	(0.027)	(0.180)
Obs.	819		944	
Avg. PYc over choices	-0.092***	0.653	-0.091***	0.363
	(0.017)	(0.099)	(0.016)	(0.116)
Obs.	478		820	
Assignment patterns				
Proba. to be assigned	0.025	0.961	0.092***	0.909
	(0.022)	(0.195)	(0.031)	(0.288)
Obs.	676		683	
# of choices eligible to	2.784***	5.406	4.206***	4.079
	(0.318)	(2.397)	(0.500)	(2.243)
Obs.	767		356	
% of choices eligible to	0.322***	0.589	0.583***	0.444
	(0.036)	(0.236)	(0.047)	(0.236)
Obs.	487		323	
Rank of choice assigned	-1.888***	2.200	-2.412***	3.057
	(0.365)	(2.287)	(0.455)	(2.630)
Obs.	397		360	

Note: The '*Change*' column gives the estimated average change in outcome at the group cutoff. Std. errors are reported in parentheses below estimates. The '*Base level*' column gives control-group statistics about the outcome. Std. deviations are reported in parentheses below mean values. Past-year cutoffs (PYc) are expressed in percentiles ($\times 100$) of the previous-year priority score distribution, with 0 indicating the lowest scores, and 1 the highest scores. *Read*: The most selective program listed by marginally uninformed students at the Group 1/2 cutoff has, on average, a past-year cutoff (PYc) at the 85th percentile of the priority score distribution. The most selective program listed by their marginally informed counterpart has a standardized past-year cutoff that is 9 percentiles lower in the priority score distribution. Estimation bandwidth is 1/8 of Imbens and Kalyanaraman (2011) optimal bandwidth. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.5: Correlations between local change in application rates and information

	Groups 1/2 cutoff	Groups 2/3 cutoff
Regression 1		
Just full	-.1400***	-.0492***
	(.0079)	(.0022)
Constant	.0122***	.0204***
	(.0011)	(.0013)
R-sqr	0.208	0.257
Obs.	616	616
Regression 2		
Remaining vacancies (10s) × Not just full	.0006	.0005***
	(.0001)	(.0003)
Earlier vacancies (10s) × Just full	-.0018***	-.0013***
	(.0001)	(.0001)
Constant	-.0007	.0037
	(.0180)	(.0012)
R-sqr	0.440	0.186
Obs.	616	616
Regression 3		
Ordinal past-year cutoff × Just full	-.1617***	-.0705***
	(.0094)	(.0047)
Ordinal past-year cutoff × Not just full	.0303***	.0867***
	(.0066)	(.0138)
Constant	0.026	.0071**
	(.0022)	(.0017)
R-sqr	–	–
Obs.	616	616

Note: For all regressions, the outcome variable is '*Estimate change in application rate at the group cutoff*'. *Just full* is an indicator of the program being declared full for the first time at the Group cutoff considered. In the 'Regression 2' panel, *remaining* vacancies correspond to the number of vacancies declared to be remaining at the Group cutoff considered. *Earlier* vacancies correspond to the number of vacancies declared to be remaining at the most recent revelation of information *previous* to that of the Group cutoff considered. In the 'Regression 3' panel, *ordinal* past-year cutoffs are expressed in percentiles (×100) of the previous-year priority score distribution, with 0 indicating the lowest scores, and 1 the highest scores. Bootstrap std. errors in parentheses, account for two-step estimation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

2.4 Recovering students' preferences for post-secondary programs

In this section, I present my approach to recover students' preferences for post-secondary programs. The identification strategy takes full advantage of local incentives for truth-telling induced by a sequential implementation of the DA. Using standard discrete choice methods, I am then able to estimate utility parameters without taking a stand on the way students form their expectations. At the end of this section, I present estimation results.

2.4.1 Identification strategy

When observed choices are the result of expected utility maximization, the econometrician who does not know the agents' expectations generally faces an identification problem (in the context of school choice, see Agarwal and Somaini (2014)). Without additional variation or assumption, it is not possible to disentangle the extent to which students' decisions are driven by what they like and the extent to which they are driven by what they think they can get. The quasi-experimental design induced by the sequential implementation of the DA enables me to circumvent this identification problem. My strategy to recover students' preferences for post-secondary programs directly builds on the three-group structure of the Tunisian mechanism. First, I show that the particular structure of information revelation embedded in the Tunisian assignment mechanism gives incentives to a subset of students to truthfully report their most-preferred programs in their application list (i.e. to 'be truthful'). The choices made by these students can be used to recover their preferences without characterizing their expectations about their admission chances. Then, I argue that this subset of truthful students is informative about, and identifies, the utility parameters governing the preferences of the population of students.

To fix ideas, I first explain how, given a set of truthful students, the preferences of this set of students can be recovered. In the next paragraph, I characterize such set of students. A student being truth-telling (or truthful) means that the alternatives listed in her application ranking coincide with her most-preferred programs among those that have not been declared full. In what follows, I call student i's *choice set* (denoted $\mathcal{J}_i$) the subset of all post-secondary programs that have not been publicly declared full at the time student i chooses and submits her application list. Then, precisely, student i being truthful means that (i) her first-listed choice has higher flow utility than any alternative in her choice set, (ii) her second-listed choice has higher flow utility than any alternative in her choice set *but* her first-listed choice, and so on, until her last-listed choice.

For a given subset of truthful students, maximizing Problem (2.1) is equivalent (in the sense that the solution sets of the two problems coincide) to the following discrete choice problem:

$$\begin{cases} \ell_i(1) & = arg\max_\ell \left[u_i(\ell) \mid \ell \in \mathcal{J}_i\right] \\ \ell_i(2) & = arg\max_\ell \left[u_i(\ell) \mid \ell \in \mathcal{J}_i \backslash \{\ell_i(1)\}\right] \\ & \vdots \\ \ell_i(M_i) & = arg\max_\ell \left[u_i(\ell) \mid \ell \in \mathcal{J}_i \backslash \{\ell_i(1), \ell_i(2), \dots, \ell_i(M_i-1)\}\right] \end{cases} \tag{2.3}$$

where $M_i \leqslant 10$ is the length of the application list submitted by student i,[21] and $\ell_i(k),\ k = 1, \dots, M$ denote the *ordered* elements of i's application list.

[21] I do not model the choice of M_i in $\{1, 2, \dots, 10\}$, nor do I control for the change of M_i across students.

'Top' students. It is natural to assume that the very first-ranked student in each group, who knows she is ranked first, truthfully reports her most-preferred programs in her application list (among those that have not been publicly declared full). Indeed, given the information revelations publicly made before each group submits applications, the very first student in each group is faced with making a choice under perfect information. She knows she has probability one to be assigned to the first-ranked element of the list she submits (as long as she lists a program that has not been publicly declared full). It is therefore strictly dominant for her to list her most-preferred program first in her application ranking. And it is weakly dominant for her to also list her second to tenth most-preferred programs.[22]

Because students may apply to up to ten programs and because most programs have more than one vacancy, not only the first-ranked student, but a subset of applicants at the top of each group have incentives to truthfully report their most-preferred programs. Proposition 2 and Condition 1 stated in Section 2.2.2 give a sufficient condition for students to truthfully report their most-preferred programs. They imply that, when going down along the priority ranking within a group, students will truthfully report their preferences as long as they think they have probability one to have a seat in one of their ten most-preferred programs.

'Short-list' students. Regardless of their priority rank, students who submit application list with strictly less than the allowed ten programs (henceforth, 'short-list students') can be interpreted as truthfully reporting their most-preferred programs

[22] The possibility to be tied in the priority order may encourage students to list choices beyond the very first rank.

(Abdulkadiroğlu et al. (2017)[23]). On the one hand, submitting a list of size smaller than ten is a (weakly) dominant strategy for students who like less than ten schools. Furthermore, this list will coincide with their preferences. Indeed, Proposition 1(a) (Haeringer and Klijn (2009)) in Section 2.2 establishes that it is dominant for students interested in strictly less than ten programs to truthfully report their preferences. On the other hand, it is (weakly) dominant for students who like ten schools or more to submit a full list of 10 programs. Indeed, it is always (weakly) profitable for such a student to add a program to an application list of less than ten programs.

Extrapolation

I argue that each of the two subsets of students described in Section 2.4.1 is sufficient to identify preferences parameters representative of those of the whole population of applicants. This result partially relies on assuming that conditional on observables, students all have the same mean flow utility for a given program. The available data however allows for enough flexibility in the specification of the flow utility function to make this assumption reasonable.

Utility specification. I assume flow utilities have the additively separable form:

$$u_i(\ell) = \delta_\ell + v_{i\ell} + \varepsilon_{i\ell}$$

δ_ℓ is a program fixed effect; it corresponds to the mean flow utility students derive from program ℓ. $v_{i\ell}$ is the part of i's demeaned (across i) flow utility for program ℓ that depends on individual characteristics observable to the analyst. $\varepsilon_{i\ell}$ is an individual- and program- specific utility shock which is privately known to the student at the time of decision-making, but remains unobserved to the analyst. I assume the program fixed effect δ_ℓ can be written as a linear (in the parameters) combination

[23] In the setting of Abdulkadiroğlu et al. (2017), 80% of students submit an application list with strictly fewer schools than the twelve allowed in the NYC high-school match. They use this subset of students to identify the preferences of NYC eighth-graders for high schools.

of program characteristics. I further assume that $v_{i\ell}$ can be written as a linear (in the parameters) combination of individual and individual-program characteristics.[24] Specifically:

$$\delta_\ell = Z_\ell' \gamma$$

$$v_{i\ell} = W_{i,\ell}' \beta$$

where Z_ℓ are program-specific attributes; $W_{i,\ell}$ are characteristics specific to the (individual, program) pair; and γ, β are utility parameters of interest, assumed to be invariant across programs and individuals. All program and individual characteristics Z, W are thought of as observed by the student at the time of decision-making, and by the analyst. In the empirical part, program attributes consist in the field of study, the degree to be received upon completion of the program, and, as a proxy for selectivity/quality/popularity, the admission score of the marginally admitted student in the previous year. Individual and program-student characteristics include distance between the student's home (as proxied by her high school) and the university hosting the program;[25] the student's high-school performance in the field of the program and outside this field; interactions between distance traveled and SES as well as program quality; and interactions between gender and field of study, as well SES and terminal degree.

The distribution of unobservables $\varepsilon_{i\cdot} := (\varepsilon_{i\ell})_\ell$ is assumed to be known, and independent of programs' and students' observable characteristics.[26] To facilitate estimation,

[24] This specification rules out preferences that depends on identified peers' assignments.

[25] As a proxy for student's i distance to university j, I use the distance between the capital city of their respective regions. Hence, student i is at distance 0 of any university in her home region. The distance between regions capitals is provided to students in the application handout made available by the Ministry of Higher Education.

[26] This rules out students sorting based on unobservable preferences —for instance, students systematically choosing their geographical residence at the time of high school to be next to the university programs they like. This guarantees that coefficients on school attributes identify the

I later assume $\varepsilon_{i\cdot}$ are i.i.d. type-1 extreme value. Normalizing to zero the coefficient on a reference field and on a reference terminal degree then fully identies the model. This means that, for each student, the value of every post-secondary program is interpreted as relative to the mean value of a local (in that distance traveled is 0) program that is not selective (past-year cutoff is 0 for programs that did not fill to capacity in 2009[27]), and upon completion of which the student would earn an 'LA' Bachelor degree (the reference degree) in Humanities (the reference field of study)[28].

The specified function controls flexibly for the two main determinants of college choices: distance between the university and the student's home, and the student's academic performance (Altonji et al. (2015)). Distance from home enters in a quadratic way. It is interacted with the student's socioeconomic status, to account for the fact that traveling may be more costly to economically disadvantaged students. It is also interacted with the program's selectivity level, to account for the fact that students may be willing to travel more to have better peers. The data, which contain students' scores at the national exam in eight different subjects, allows me to control separately for the student's high-school performance in STEM fields and non-STEM fields. The student's high-school performance in each field is also interacted with the program's field of study to account for individual comparative advantages in studying one subject vs. another, and for the fact that studying a given field may require more effort from students with lower high-school performance in the field.

students' valuation for that attribute, and does not capture correlated variation with unobservable tastes. However, note that programs' and students' observable characteristics will be taken as given and fixed in the counterfactual analysis and welfare evaluations in this paper.

[27] Past-year cutoffs are expressed in percentiles of the distribution of priority scores; non-selective programs have cutoff at the 0^{th} percentile.

[28] In other words, student i derives flow utility $u_{i,\ell} = \beta_{\text{SES}_i} \times \text{distance}_{i,\ell} + \gamma_{\text{SES}_i} \times \text{past-year cutoff}_{\ell} + \varepsilon_{i,\ell}$ from a program ℓ preparing her to receive an 'LA' Bachelor degree in Humanities. If program ℓ is local ($\text{distance}_{i,\ell} = 0$) and non-selective ($\text{past-year cutoff}_{\ell} = 0$), then i's utility for ℓ is $u_{i,\ell} = \varepsilon_{i,\ell}$.

Representativeness of truthful students and their choices. The particular structure of the subset of truthful students is crucial to the identification of such a flexible utility function, which renders the extrapolation credible. Tables A.1 and A.2 in Appendix A.3 show descriptive statistics for key student and choice characteristics, comparatively for three samples of interest: the whole population (for which we would like to recover utility parameters); and each of the two truthful subsamples. Table A.1 shows that students' characteristics have, in each of the truthful subsamples, similar variation and support as they have in the population. In that sense, each of the truthful samples is representative of the population. There is one main exception: high-school performance variables have smaller support in the 'top' sample than they have in the population. This is a consequence of 'top' students being sampled from three points in the priority (a deterministic function of high-school performance) distribution. When using 'top' students to recover population utility parameters, the relationship between preferences and high-school performance in the population is therefore extrapolated from the relationship between preferences and high-school performance among 'top' students, via the continuity of the utility function. The validity of such extrapolation would be a strong assumption if 'top' students were sampled from *one* point of the priority distribution. However, Table A.1 shows that the three-point sampling allowed by the Tunisian design ensures sufficient range and variance in 'top' students performance to allow for a reasonable extrapolation by continuity.[29]

[29] The argument is made clear by comparing 'top' students characteristics in the Tunisian design with the characteristics of those who would be 'top' students in a single-phase implementation of the assignment mechanism —that is, students at the very top of Group 1. Descriptive statistics for these students, also provided in Table A.1, show that there is very little variation in 'top Group 1' students' high-school performance, and the variable has a very small support relative to its support in the population. This is unsurprising given the very selected nature of the top of Group 1 data. As a consequence, extrapolating to the population the utility function recovered from the top of Group 1 data would require unreasonable assumptions about the homogeneity of students' preferences across the range of high-school performance. As shown by Table A.1, The Tunisian design, which gives incentives to be truthful to students at three point of the priority distribution rather than

Table A.2 shows that the characteristics of the choices made by students in each of the truthful subsamples described in Section 2.4.1 span the full support of programs' characteristics in the initial choice set. In that sense, students in each of the truthful samples express preferences over all relevant tradeoffs existing in the choice set. This is a consequence of the choice set restrictions imposed by the informational updates. Students in low-priority groups are induced to express preferences over and solve tradeoffs involving programs others than the programs most popular among high-priority students and publicly declared to be full.[30]

2.4.2 Estimation

A local discrete choice estimation procedure.

I assume that the unobservable components $\varepsilon_{i,j}$ are i.i.d. type-1 extreme value. Estimation proceeds by maximum likelihood, using the sample of truthful students.[31] Independence of unobservables across individuals allows to write the sample likelihood L as the product of individual likelihoods p_i; independence of unobservables across alternatives further allows to write individual likelihoods as the product of

one, crucially alleviates this issue.

[30] As an illustration, Table A.2 shows that there is also little variation in the characteristics of the programs chosen by 'top Group 1' students. Only 10% of the existing programs are listed by top of Group 1 students in their application lists. There is very little variation in the selectivity level of the listed programs by students at the top of Group 1, relative to what is observed in the population. Moreover, some program characteristics do not have full support in 'top Group 1' students' choices. This is the case not only for some fields of study (no program in Social Sciences and Law, while in the population students do express preferences regarding these fields), but also for key interaction variables such as *distance from home* × *program selectivity*. As a consequence, important aspects of students' preferences, such as the way they solve trade-offs between traveling further from home, attending more selective institutions, and studying a field they like, could not be identified by the sample of 'top Group 1' students.

[31] In Appendix A.3, I discuss an alternative estimation strategy that could yield more precise utility parameter estimates by using all students' application lists, rather than only those from truthful students.

logit probabilities, yielding the well-known ranked-ordered (or exploded) logit form:

$$L = \prod_{a=1}^{N_E} p_{i_a}$$

$$\text{with} \quad p_i = \frac{\exp(u_{i\ell_i(1)})}{\sum_{k \in J_i} \exp(u_{ik})} \times \frac{\exp(u_{i\ell_i(2)})}{\sum_{k \in J_i \setminus \{\ell_i(1)\}} \exp(u_{ik})} \times \dots \times \frac{\exp(u_{i\ell_i(M_i)})}{\sum_{k \in J_i \setminus \{\ell_i(1),\dots,\ell_i(M_i-1)\}} \exp(u_{ik})} \tag{2.4}$$

where $M_i \leqslant 10$ is number of programs included by i in her application list; and $(i_a)_{a=1,\dots,N_E}$ denote the estimation sample.

While the subset of 'short-list' students is readily observable from the data and can straightforwardly be used for estimation, the subset of 'top' students is *a priori* unobserved. Indeed, Condition 1, which ensures that students who *think* they have probability 1 to clear the *ex-post* admission cutoff of (at least) one of ten most-*preferred* programs (among those not declared to be full) truthfully report their preferences, cannot be used in practice as students' preferences and expectations are unknown at this stage. In the rest of this subsection, I explain how I select the 'top' estimation sample.

Estimation from 'top' students: choosing the 'top' sample in practice

A standard bandwidth choice problem. The choice of the 'top' estimation sample is akin to the choice of the estimation bandwidth in any local estimation procedure (e.g. local linear regression, as in Section 3.3.2). Selecting the 'top' estimation sample involves solving a trade-off between bias and variance of the estimator. The sample should be sufficiently large to have identification power and for the estimates to be precise. However, the sample should be small enough not to include any non-truthful students, whose inclusion would bias the estimates.

In practice, I include in the 'top' estimation sample all students with priority in the top 200 ranks within each group.[32] In the next paragraph, I present empirical evidence that the selected 'top' students are aware of the incentives they face and behave truthfully. In Section 2.4.3, I discuss the robustness of my results to changes in the estimation bandwidth.

Empirical validation of the bandwidth choice. I present three pieces of empirical evidence suggesting that, within the chosen 'top' bandwidth, students truthfully report their most-preferred programs —that is, that my estimates are unlikely to be biased by the presence of students misreporting their preferences. First, given Condition 1, it is crucial that students use and understand the public information about vacancies. In particular, it must be reasonable to assume that they *understand* that, given their priority ordering and the number of vacancies said to be remaining, they do have eligibility probability 1 to a range of programs. The RDD analysis in Section 3.3.2 strongly suggests that this is indeed true. In particular, Table 2.5 suggests that students understand the incentives they face. In addition, marginally informed students submitting shorter lists than their marginally uninformed counterparts (Table 2.4) suggests that students at the top of each group recognize that their eligibility to some programs is certain. It is then natural to infer that they understand that a truthful report of their most preferred alternatives (in their choice set) is dominant.

Second, I show that suggestive evidence of students censoring themselves (in the sense

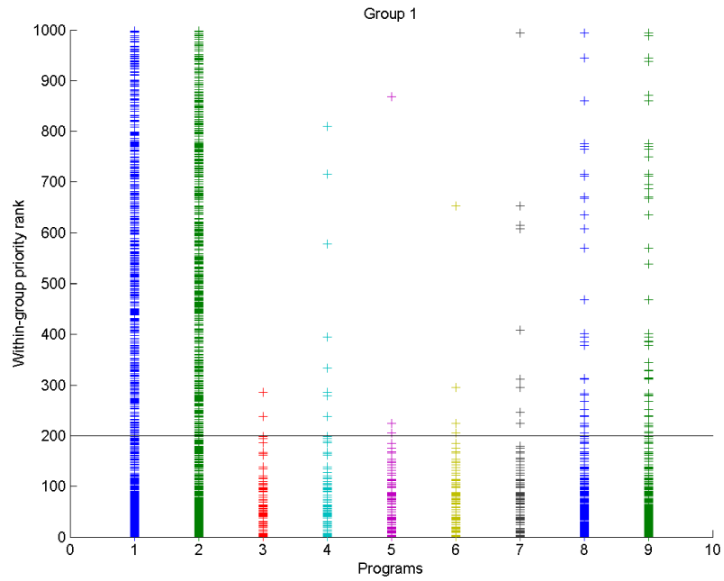
[32] The size of the subset of truthful students is increasing in the size of the application list students are able to submit, and in programs' capacities. Indeed, keeping preferences and everything else fixed, the number of students who perceive to be eligible with probability 1 to one of their $M' > 10$ most-preferred programs is weakly larger than the number of students who perceive to be eligible with probability 1 to one of their $M = 10$ most-preferred programs. Similarly, all other things equal, if the number of available seats to all programs (weakly) increases, the number of students who perceive to be eligible with probability 1 to one of their 10 most-preferred programs weakly increases as well.

that they do not apply to their most-preferred but very popular alternatives) can be found only among students outside the chosen bandwidth. It is uncontroversial that the very first student at the top of each group is certain about her eligibility chances and is truth-telling. Ambiguity about whether other applicants think that reporting truthfully is dominant for them increases as within-group priority decreases. When students stop being truthful, we expect to observe a decreased frequency of application to programs listed by the uncontroversially truthful students. Figure 2.2 shows that such frequency decrease does not happen within the chosen 'top' estimation bandwidth. Figure 2.2 considers the ten programs listed the first three students at the top of Group 1,[33] and shows the frequency at which these programs are listed by Group 1 students as a function of students' priority. Programs are represented on the x-axis, priority on the y-axis. A dot in position (a, b) in the graph means that student ranked b in Group 1 included program a in her list. The vertical line at rank 200 represents the limit of the 'top' estimation bandwidth —so students with priority rank lower than 200 are included in the estimation sample. The frequency of application to the programs listed by the top three students does not decrease within the 'top' estimation sample. Passed the bandwidth limit, it then decreases more or less abruptly for some programs—suggesting that omission or censoring start occurring.

Third, I show *ex-post* evidence that students in the 'top' estimation sample submit a truthful report of their preferences. For this, I rely on another unique feature of the implementation of the DA algorithm in the Tunisian context: a *reassignment round*. After students of all three groups have been assigned by the DA algorithm[34] but before the new academic year starts, students are invited to express any dis-

[33] Figure A.2 in Appendix A.3 shows similar evidence for Groups 2 and 3.

[34] or administratively for those not eligible to any element of any of their application lists



Legend: This graph considers the ten programs listed the first three students at the top of Group 1, and shows the frequency at which these programs are listed by Group 1 students as a function of students' priority. Programs are represented on the x-axis, priority on the y-axis. A dot in position (a, b) in the graph means that student ranked b in Group 1 included program a in her list. The vertical line at rank 200 represents the limit of the estimation bandwidth —students with priority rank lower than 200 are included in the estimation sample.

FIGURE 2.2: Persistence of the top-ranked students' listed choices over the priority ranking

satisfaction about their assignment. Precisely, students may submit a new ordered list of four programs they would prefer to attend over their assigned match. Any program can be included in this new list, irrespective of whether it was part of the student's initial ranking or not, of whether it has vacancies left or not, and of the program's realized admission cutoff relative to the student's priority score. Importantly, students do not have to forgo their initial assignment to participate in the reassignment round: unless their reassignment request is approved, they keep their initial match. No precise procedure is explicitly defined regarding the processing and approval of requests. It is generally understood that priority within students is preserved in the reassignment round and that approval depends on the ability of the requested track to welcome an additional student. The top panel in Table 2.6 shows the shares of top and non-top students applying for reassignment. The bottom panel describes further the behavior of students applying for reassignment. Students in the 'top' bandwidth apply for reassignment at a significantly lower rate than other students (17.6 vs 24.6%) and when they do, they submit fewer requests (2.7 vs 3 programs included in the reassignment list). Most importantly, about 84% of the programs students in the 'top' bandwidth request are outside their choice set, that is, had already been declared full when the students applied. On the contrary, the majority (54%) of requests submitted by other students are within their choice set. The large difference between the shares of students in and out of the 'top' bandwidth who request reassignment within their choice set (2 vs 11%[35]) suggests that students in the estimation sample did not censor themselves —that is, indeed reported their most-preferred programs. Indeed, students out of the 'top' bandwidth reveal by their reassignment requests that they prefer some of the alternatives in their choice set that they did not initially list over the ones they initially applied to.[36]

Table 2.6: Students at the top reveal to be satisfied with their assignments

	In the 'top' estimation sample	Out of the 'top' estimation sample
# students	636	10,368
% requesting reassignment	17.61	24.60
Conditional on request		
Average # requests per student	2.71 (1.28)	3.02 (1.22)
% requesting w/in choice set	13.39	43.39
% of all requests w/in choice set	16.17	54.32

Standard deviations in parentheses, next to sample means.

2.4.3 Results

Main estimates

Table 2.7 shows maximum likelihood (ML) estimates obtained from pooling both subsets of truthful students ('top' and 'short-list'). Results for each of the two estimation samples are similar to the ones shown here, and displayed in Table A.3 in Appendix A.3. Results being similar across the two subsets of truthful students supports the validity of the extrapolation argument made above. As expected, estimates obtained from the larger 'top + short-list' sample are more precise —most standard-errors are reduced by half relative to the sample of 'top' students. Results of a sensitivity analysis regarding the bandwidth choice for the 'top' sample are also

[35] Shares: $.02 = .176 \times .13$ and $.11 = .246 \times .43$.

[36] Calsamiglia et al. (2014) and Kapor et al. (2016) rationalize students "changing their mind" by them receiving a post-assignment, pre-enrollment utility shock. The contexts in these two papers are different from the one here: they consider families applying for seats in public schools and kindergarten in Barcelona and Cambridge, MA respectively (both cities use a variant of the Boston mechanisms). In their context, a student "changing her mind" is a student who is matched to her first choice by the centralized mechanism but ends up not enrolling in the school —and supposedly enrolling in a private school instead. Despite this difference of settings, a post-assignment, pre-enrollment utility shock could be used here as well to justify students requesting reassignment. However, there is no reason *a priori* why students 'at the top' would be induced to change their mind at such a much lower rate relative to other students. At the very least, the share of non-top students asking for reassignment in excess of the share of students at the top can reasonably be attributed to forecasting errors on their part.

Table 2.7: Utility parameter estimates (1/3)

	(1) Main	(2) Lin. in distance
Distance (100km)	-2.010***	-1.009***
	(0.07)	(0.05)
× high SES	0.026	0.137*
	(0.04)	(0.06)
Distance (100km) sq.	0.221***	
	(0.01)	
Past-year marginal admit	2.282***	3.015***
	(0.30)	(0.30)
× high SES	0.534	0.713
	(0.39)	(0.39)
Past-year marginal admit sq.	-1.029***	-0.712*
	(0.31)	(0.33)
× high SES	0.958*	0.580
	(0.38)	(0.38)
Distance (100km) × Past-year marginal adm.	0.884***	
	(0.07)	
Degree: Bachelor (LF)	0.547***	0.548***
	(0.06)	(0.06)
× h-s perf.	0.379***	0.384***
	(0.05)	(0.05)
× high SES	0.033	0.009
	(0.07)	(0.07)
Degree: Adv. degree	2.544***	2.529***
	(0.08)	(0.08)
× h-s perf.	1.838***	1.849***
	(0.08)	(0.08)
× high SES	-0.160	-0.193*
	(0.09)	(0.09)
Sample	Top + Short	Bdw
PseudoObs.	24,961	24,961
Obs.	3,629	3,629

Distance (100km) gives the distance (in 100km) between the program's region and the region of the student high school (as a proxy for home); *Distance (100km) sq.* is the square of this distance. *Past-year (PY) marginal admit* gives the priority score (in percentiles of the priority scores distribution) of the program's marginally admitted student in the past year. A higher score number means a higher priority —for instance, a value of *Past-year marginal admit* of .01 means that the program's 2009 marginally admitted student was at the bottom 1% of the 2009 priority distribution. *Degree:'. . . '* are indicators of whether the program prepares to the degree considered; these coefficients are allowed to differ continuously × *h-s perf.*, where *h-s perf.* is the student's (standardized) unweighted average score at the end-of-high-school exam. LA (*Licence appliquée*) is used as the reference group for degree dummies.

Utility parameter estimates (2/3)

	(1)	(2)
Field: Arts	2.788***	2.787***
	(0.28)	(0.28)
× STEM h-s perf.	1.672***	1.672***
	(0.18)	(0.18)
× non-STEM h-s perf.	-1.261***	-1.245***
	(0.23)	(0.22)
× female	-0.949**	-0.944**
	(0.31)	(0.31)
Field: Educ.	2.042***	2.073***
	(0.39)	(0.39)
× STEM h-s perf.	1.287***	1.242**
	(0.37)	(0.39)
× non-STEM h-s perf.	-0.725	-0.743
	(0.51)	(0.53)
× female	-1.568**	-1.560**
	(0.51)	(0.52)
Field: Soc. Sc.	0.911**	0.935**
	(0.35)	(0.35)
× STEM h-s perf.	0.820**	0.841**
	(0.32)	(0.32)
× non-STEM h-s perf.	-0.912**	-0.901**
	(0.34)	(0.33)
× female	-1.337**	-1.336**
	(0.41)	(0.41)
Field: Eco/Mgmt	3.647***	3.632***
	(0.29)	(0.29)
× STEM h-s perf.	1.168***	1.200***
	(0.19)	(0.18)
× non-STEM h-s perf.	-1.194***	-1.186***
	(0.23)	(0.23)
× female	-1.056***	-1.050***
	(0.30)	(0.30)
Field: Law	2.189***	2.177***
	(0.40)	(0.39)
× STEM h-s perf.	0.434	0.481
	(0.35)	(0.34)
× non-STEM h-s perf.	-0.408	-0.415
	(0.34)	(0.33)
× female	-1.187**	-1.190**
	(0.42)	(0.42)
Field: Math/Comp.Sci.	4.296***	4.284***
	(0.28)	(0.28)
× STEM h-s perf.	1.317***	1.341***
	(0.17)	(0.17)
× non-STEM h-s perf.	-1.551***	-1.542***
	(0.23)	(0.22)
× female	-1.230***	-1.220***
	(0.30)	(0.30)

Utility parameter estimates (3/3)

	(1)	(2)
Field: Phys./Chem./Engin.	4.074***	4.054***
	(0.27)	(0.27)
× STEM h-s perf.	1.291***	1.311***
	(0.17)	(0.16)
× non-STEM h-s perf.	-1.598***	-1.586***
	(0.22)	(0.22)
× female	-1.391***	-1.383***
	(0.29)	(0.29)
Field: Health/Life Sc.	3.491***	3.445***
	(0.28)	(0.28)
× STEM h-s perf.	1.143***	1.177***
	(0.17)	(0.17)
× non-STEM h-s perf.	-1.310***	-1.286***
	(0.23)	(0.23)
× female	-0.346	-0.324
	(0.30)	(0.30)
Field: Earth Sc.	2.169***	2.173***
	(0.28)	(0.28)
× STEM h-s perf.	0.293	0.301
	(0.18)	(0.18)
× non-STEM h-s perf.	-1.633***	-1.626***
	(0.24)	(0.23)
× female	-0.863**	-0.870**
	(0.30)	(0.30)
Program location: Tunis	0.520***	0.748***
	(0.08)	(0.10)
Program location: Coast	0.352***	0.395***
	(0.07)	(0.08)
Program location: Abroad	-9.626***	-10.676***
	(1.43)	(1.51)
× STEM h-s perf.	3.448***	3.754***
	(0.72)	(0.77)
× non-STEM h-s perf.	2.254***	2.371***
	(0.35)	(0.38)
× high SES	-0.116	0.177
	(0.36)	(0.37)

Field:'. . . ' are indicators of whether the program is in the field considered; these coefficients are allowed to differ across sexes, and continuously with high-school performance in STEM and non-STEM fields via the interactions × *female*, × *STEM h-s perf.* and × *non-STEM h-s perf.*. *STEM h-s perf.* and *nonSTEM h-s perf.* are the student's (standardized) unweighted average score at the STEM and non-STEM tests taken in the end-of-high-school exam. 'Humanities' is used as the reference group for field dummies. *Program location: '. . . '* are indicators of whether the program is located in the region considered. '*Southern and western regions*' is used as the reference group for program location dummies.

Std. errors in parentheses, clustered at the high school level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

presented in Appendix A.3. Table A.4 in Appendix A.3 shows results derived with alternative bandwidth sizes (twice, five and ten times the original bandwidth size, namely). Utility coefficients derived under the original bandwidth and a bandwidth of twice its size are not identical, but, for many of them, similar. This suggests that results are robust to small changes in the bandwidth size. As the bandwidth size increases, estimated coefficients are increasingly different from the original estimates from Table 2.7, illustrating the increasing bias affecting estimates as more non-truthful students get included in the estimation sample.

Interpretation is easier when distance is used as a numeraire. Column (2) in the table shows ML estimates obtained for a utility function with no quadratic term on distance. While Column (1) shows preferred estimates that will be used in later parts of the analysis (Sections 2.5 to 2.6), I use the linear-in-distance utility estimates to comment on parameters in this paragraph.

A positive coefficient on *Past-year marginal admit* means that students value program quality, as measured by the priority ranking of the past-year median admit of the program —an increase in *Past-year marginal admit* means that the 2009 median admit had higher priority, hence corresponds to a increase in the program quality. A positive coefficient on squared *Past-year marginal admit* means that the marginal value of an increase in program selectivity increases with the program level of selectivity. In other words, students' willingness to travel to attend a program with marginally higher quality increases as quality gets higher. The magnitudes estimated with specification (2) suggest that low-SES students are willing to travel 3.0km ($\approx$ 1.875 miles[37]) for a 1-percentile increase of in program quality, all other things equal, against 4.3km for high-SES students..

[37] $3.015 \times .01/(1.009 \times .01)=3.0$; $(3.015+.713)/(1.009-.137)=4.3$

A positive coefficient on a field dummy means that, on average, students prefer the field to the reference field (Humanities). All field-dummy coefficients being positive means that Humanities is the least-preferred field for average-performing[38] male students. Comparison of field-dummy coefficients show that STEM fields are preferred over non-STEM fields, and that Math/Comp.Sci. is the most-preferred field of study —which is not surprising given that students in the sample graduated from high-school with a Math major. All other things equal, low-SES (resp. high-SES) male students are willing to travel 23km (resp. 26km[39]) to study Math/Comp.Sci. rather than Physics/ Chemistry/ Engineering (the second most-popular STEM field among male students), and 65km (resp 75km[40]) to study Math/Comp.Sci. rather than Economics/ Business/ Management (the most-popular non-STEM field among male students). Female students prefer Math/Comp.Sci. over Physics/ Chemistry/ Engineering by the same magnitude as males, and their most-preferred field is also a STEM field —again, as to be expected from high-school Math-majors. In contrast to males, though, female students strongly prefer Health and Life Sciences over Math/Comp.Sci.. Economics/ Business/ Management is also female students' most preferred non-STEM field (at the mean performance levels in STEM and non-STEM), and they generally dislike less non-STEM fields than male students. For students of both sexes, Earth Sciences is the least preferred STEM field. Preferences for field of study are correlated with high-school performance in STEM and non-STEM fields, although most coefficients on field-performance interactions are not statistically significant. Notably though, in my sample of Math-high-school graduates, preference for Earth Sciences strongly decreases as high-school performance in STEM increases. As for non-STEM fields, preference for Humanities strongly increase as high-school

[38] High-school performance in STEM and non-STEM are standardized to have mean 0 and standard-deviation 1 in the population of students graduating from high-school with a Math major.

[39] (4.284-4.054)/1.009=.23; and (4.284-4.054)/(1.009-.137)=.26

[40] (4.284-3.632)/1.009=.65; and (4.284-3.632)/(1.009-.137)=1.75

performance decreases in STEM subjects and/or increases in non-STEM subjects. A positive coefficient on a degree dummy means that, on average, students prefer the degree in question over the reference degree ('*licence appliquée*', a type of Bachelor degree designed for students likely to enter the labor market upon graduation). On average, students prefer both the '*licence fondamentale*' (a type of Bachelor degree designed for students likely to pursue a graduate studies upon graduation) and the more advanced degrees to the reference Bachelor-equivalent. Preference for these two types of degree increases with high-school performance —especially for the latter, as expected given the higher academic level certified by this type of degree. On average, students with high-school performance equal to the mean are willing to travel about 54 to 64km, depending on SES, to work towards a *licence fondamentale* rather than a *licence appliquée*; and between 251 and 268km to work towards an advanced degree (e.g., Masters' and M.D.) instead of a *licence appliquée*.[41]

2.5 Expectations about admission chances

Students' application lists depend not only on their preferences, but also on their expectations about their admission chances. When studying the effect of provision of information about program vacancies on students' chosen application lists, it is important to have a sense of how their expectations about their admission chances relate to the information available to them. Taking preferences as known from the previous section, this section characterizes students' expectations about their admission chances.

A student's admission to a given university program is the result of all students' application decisions. Indeed, whether or not a seat in that program will be offered

[41] LF: .548/1.009=.54; and (.548+.009)/(1.009-.137)=.64. Adv. degrees: 2.529/1.009=2.51; and (2.529-.193)/(1.009-.137)=2.68.

to the student by the algorithm depends not only on whether or not the student listed the program in her portfolio, but also on whether or not the program was already filled up as a consequence of other students' assignments and applications. It has been acknowledged in previous studies (e.g., Agarwal and Somaini (2014); Calsamiglia et al. (2014); Ajayi and Sidibé (2016); Kapor et al. (2016)) that the complexity of forming expectations about admission chances in such a game setting and solving the expected-utility maximization problem (2.1) likely exceeds the computational capacity of high-school students and their families, especially as the number of programs students can choose from and the size of the application list they may submit get large. Hence, students have been allowed to behave with limited sophistication, viewing the application decision as a single-agent problem rather than a game. Previous analyses have also established the possible existence of differences in expectations formation and use of public information across socioeconomic status (SES) and related variables (e.g., Hoxby and Turner (2015)).

The approach taken in this section is in line with these concerns. I specify *types* of expectations-formation behavior for students who take their application choice as a single-agent problem. My analysis then proceeds to recovering the *share* of each type in the student population (conditional on students' observables). The identification strategy takes advantage of the fact that utility parameters were recovered from a strict subsample and without taking a stand on students' expectations about their admission chances. Perceived admission probabilities are sought to rationalize, given utility parameters, the application lists submitted by the students, who can *a priori* be truthful or strategic. Given utility parameters, I use maximum simulated likelihood (MSL) to estimate the share of various types of expectations formation behaviors in the population.

2.5.1 Forming expectations about one's admission chances: a model

I allow for two main types of application behavior among students —sophisticated and unsophisticated. I describe these types now, along with the effect of informational updates about programs filling up and vacancies remaining on each of them.

Unsophisticated students. Unsophisticated students simply report in their application list their most-preferred alternatives in their choice set (Agarwal and Somaini (2014); Calsamiglia et al. (2014)). No matter their position in the priority ranking, they are truthful. The provision of information about vacancies, made through the sequential implementation of the application procedure, enables them to update their choice set: they only consider programs that have not been publicly declared full at the time they submit their portfolio.

Sophisticated students. Sophisticated students maximize expected-utility (2.1) to choose their application portfolio. I assume that sophisticated students form their expectations about their eligibility chances on the grounds of the programs' past-year admission cutoffs and their own priority score, rather than based on a model for other students' behavior. This is a natural approach given the availability of past-year marginal admission scores to students. Specifically, to report the expected-utility-maximizing list, such students derive their expectations assuming marginal admission scores follow, from one year to the next, some AR(1) process of the form:

$$\text{cutoff}_{j,2010} = a + b \times \text{cutoff}_{j,2009} + \eta_j \quad \text{with } \eta_j \sim N(0, \sigma^2). \tag{2.5}$$

Taking the parameters a, b, σ of the relationship (2.5) as given, student i's expectation about her probability to clear the admission cutoff for program j is:

$$\Pr\left(\text{priority}_i \geqslant \text{cutoff}_{j,2010}\right) = \Phi\left(\frac{1}{\sigma}(\text{priority}_i - a - b \times \text{cutoff}_{j,2009})\right) \tag{2.6}$$

In this framework, the effect on eligibility chances of an informational update, such as those provided in the Tunisian mechanism, is to reset to 0 the perceived probability of admission to programs which are declared to be full. In addition, when a student's priority ranking within her group is such that there are fewer students in the group to be assigned before her than vacancies publicly declared to be remaining in the program, the student's perceived probability of admission to this program is reset to 1 by the information revelation.

In the Tunisian sequential design, Group 1 and Group 2 students who fail to be assigned to any of their listed choices are pooled at the top of the next group and allowed to participate again time in the application process (see Section 2.3.1). At that time, they can only pretend to alternatives still available in the choice set of this next group. This 'second chance' affects students' option value of being rejected from all their listed choices —in Equation (2.1), $V_i(0)$. In the present model of sophisticated behavior, I assume that a student i in Group 1 or 2 computes $V_i(0)$ as follows. Let $groupCutoff_i$ denote the position in the priority ranking of the division between i's application group and the next. Student i can form expectations about the choice set she would face at the time of her 'second chance' if she were to use it by using $groupCutoff_i$ instead of $priority_i$ in Equation (2.6). Then, $V_i(0)$ is the value of the program with highest expected utility, based on these 'second-chance' expectations. The probability that i will use that second chance ($\overline{\pi_i}$ in Equation (2.1)) is determined by her admission probabilities to her listed choices.

Estimation of the model of expectations formation proceeds in two parts, both described in the next subsection. First, I fix the AR(1) parameters characterizing the sophisticated type; then I recover the respective shares of sophisticated and unsophisticated students in the population. The second part uses data on students' application behaviors; the first part uses panel data on admission cutoffs. While I fix AR(1) parameters and do not recover them from students application choices *per se*, I maintain flexibility in the model by allowing for different sophisticated (sub)types. Hence, different sophisticated students may use different sets of AR(1) parameters to form their expectations about their admission chances.

2.5.2 Using observed choices to recover types shares

Paragraphs 2.5.2 and 2.5.2 present the second part of the estimation strategy. I explain how, when types are taken as fixed and known, their population shares can be recovered. Paragraph 2.5.2 turns to the first part of the estimation strategy, and describes the way I fix AR(1) parameters for the sophisticated (sub)types.

In the next two paragraphs, it is useful to view each sophisticated expectations-formation type t as a *known* function that takes as inputs the student's priority score and the information about programs that is publicly available at the time she submits her application list, and returns a vector of J probabilities, which is interpreted as the student's perceived eligibility probabilities to all university programs. A type-t student then uses the J probabilities outputed by the type-t function, along with her preferences, as inputs in the expected-utility maximization problem (2.1). (Consistently, the unsophisticated type can be viewed as the function that, given the student's preferences and the information about programs that is publicly available at the time she submits her application list, returns the programs with highest flow utility among those that have not been declared full).

A maximum simulated likelihood approach

Suppose each student has one of T discrete expectations-formation types. The probability P_i to observe the actual application list $\mathcal{L}_i$ submitted by student i given her mean flow utilities for all programs $(\bar{u}_{i,j})_j$, and her observable characteristics X_i, writes

$$P_i = \sum_{t=1}^{T} \underbrace{\Pr\left(\mathcal{L}_i \mid (\bar{u}_{i,j})_j, X_i, \theta_i = t\right)}_{:=p_i(t)} \times \underbrace{\Pr\left(\theta_i = t \mid (\bar{u}_{i,j})_j, X_i\right)}_{:=\rho(t,X)} \tag{2.7}$$

where, for types $t = 1, \ldots, T$, $p_i(t)$ is the probability to observe the actual application list $\mathcal{L}_i$ submitted by student i conditional on her being of type t, and $\rho(t, X)$ is her probability to be of type t given her observable characteristics. Since, conditional on observable characteristics, a student's expectations-formation type is independent of her vector of mean flow utilities,[42] $\rho(t, X) = \Pr(\theta_i = t \mid X_i)$.

The type functions being known means that, assuming a student has type t, and given her priority score, I can straightforwardly derive her perceived probabilities of admission to all university programs using Equation (2.6). Hence, for any fixed type t, and for each student i, since utility parameters are known, $p_i(t)$ can be estimated by simulation (of choices, over preferences unobservables). Given the set of possible types, and estimated conditional choice probabilities $\hat{p}_i(t)$ for all $(i, t) \in \{1, \ldots, N\} \times \{1, \ldots, T\}$, the type shares $\rho := (\rho(t, X))_t$ can then be recovered by maximizing the (simulated) sample likelihood:

$$\mathbf{L}(\rho) = \prod_{i=1}^{n} P_i = \prod_{i=1}^{n} \left(\sum_{t=1}^{T} \hat{p}_i(t) \times \rho(t, X_i) \right).$$

Identifying variation

If two expectations-formation types t_1 and t_2 are such that $p_i(t_1) = p_i(t_2)$ for all $i = 1, \ldots, N$ such that $X_i = X$, their respective shares $\rho(t_2, X)$ and $\rho(t_1, X)$ cannot

[42] This follows from the fact that, conditional on observable characteristics, utility parameters are independent of student's expectations and level of sophistication. This holds under the assumptions made in Section 2.4.

be separately identified. The estimation strategy exploits the fact that, given her (known) mean utility, a student does not have the same probability to submit the application list she did submit (which is observed in the data) conditional on being of two different (sub)types. Precisely, the identification of conditional type shares $\rho(t, X)$ in (2.7) for the specified types $t = 1, \ldots, T$ relies on $p_i(t)$, the (simulated) likelihood to observe the data conditional on X_i, varying across t. Figure A.4 in Appendix A.4 illustrates the identifying variation in the data, and the way it is enters in the estimation strategy.

In practice, given the size of the choice set and programs' observable characteristics, there is more variation across expectations-formation types in the likelihood (given preference parameters and individual characteristics) of observing the *characteristics* of students' chosen programs (shown in Figure A.4) than in the likelihood of observing the *identity* of their chosen programs. Hence in the implementation of the MSL, rather than the definition given in Equation (2.7), I use

$$p_i(t) = \Pr\left(\mathbf{Y}(\mathcal{L}_i) \mid (\bar{u}_{i,j})_j, X_i, \theta_i = t\right)$$

where $\mathbf{Y}(\mathcal{L}_i)$ is a vector of program characteristics of the programs listed by i. For instance, $\mathbf{Y}(\mathcal{L}_i)$ may include the selectivity level (in terms of past-year admission cutoff), the distance home-university for student i, the number of vacancies publicly announced to be remaining a the beginning of i's application group.[43]

Fixing types

Fixing the types can be done in a very flexible way. The set of possible types *a priori* allowed can be large, and the estimation procedure allows the data to dictate which types have positive probability in the population. The only limitations on

[43] Because the estimator does not use all the available data, it is not efficient.

the set of possible types are imposed by identification requirements. If two sets of expectations-formation parameters induce the same application behavior for all students, their shares in the population cannot be separately identified —regardless of whether they induce the same expectations for all students or not.

The results shown in Section 2.5.3, are based on allowing for six sophisticated (sub)types, in addition to the unsophisticated type. Each sophisticated (sub)type corresponds to a different specification of the AR(1) equation (2.5). Specifications differ from one another in the level of observable heterogeneity across programs accounted for in (2.5). I characterize these specifications more in detail in Section 2.5.3, as I describe and interpret results. The choice of AR(1) coefficients for these specifications is based on data on programs' 2009 and 2010 marginal admission scores. Namely, using these cutoffs data, I estimate the AR(1) processes characterizing each sophisticated type. Estimated coefficients, which I use in simulations to recover conditional choice probabilities (2.7), are showed in Appendix A.4.[44]

[44] One may think that the ideal approach would allow to recover from students' choices not only the shares of each type, but the parameters characterizing the types as well. Such approach would however face two pitfalls. The first is the identification issue already mentioned at the beginning of paragraph 2.5.2. The second is a computational issue. Consider the following (already restrictive) framework. Suppose each student is of one of $T < +\infty$ discrete types of expectations formation processes. Suppose a student of type t forms her expectations about her admission chances by assuming that the changes in admission cutoffs from one year to the next are normally distributed so that: $\text{cutoff}_{j,2010} = \text{cutoff}_{j,2009} + \nu_j$ where $\nu_j \sim \mathcal{N}(0, \sigma_t^2)$. The variances $(\sigma_t^2)_t$, along with the shares $(\rho)_t$, are to be recovered. *Each* evaluation of the likelihood $\mathbf{L}(\sigma^2, \rho)$ requires simulating choices for *all* students under type t to estimate the conditional choice probabilities $p_i(t)$. Optimization over a set of values for σ^2 quickly gets very demanding. Fixing the types, simulations and estimation of the conditional choice probabilities $p_i(t)$ can be done once and for all outside the optimization routine, and hence greatly simplify estimation. In this case, an evaluation of the likelihood $\mathbf{L}(\rho)$ simply involves reweighing this estimated conditional choice probabilities $p_i(t)$.

2.5.3 Results

Estimated types.

Table 2.8 shows types shares estimated conditional on socioeconomic status (SES) and for six sophisticated subtypes. Estimated shares of unsophisticated students are robust to changes in the set of AR(1) specifications characterizing sophisticated subtypes, and conditioning on other demographics, such as sex, region of origin, or interactions of these demographics with SES. Shares for the main two types — sophisticated and unsophisticated– are shown in bold font, along with a breakdown of the sophisticated type into its subtypes. Slightly less than two thirds of the low-SES students are estimated to behave naively, against half of high-SES students. These large shares of unsophisticated behaviors are not necessarily surprising. In the context of the NYC high school match, which is based on a single-phase DA with a restricted list of 12 choices, Abdulkadiroğlu et al. (2017) show evidence that at least 80% of applicants (and very possibly all of them) truthfully report their most preferred programs in their application list. Focusing on an alternative assignment mechanism that highly rewards sophisticated behavior (a variant of the Boston mechanism), Agarwal and Somaini (2014) estimate that a third of families participating in the elementary school match in Cambridge, Massachusetts behave in a unsophisticated way, the other two thirds behaving as if they knew their true admission probabilities. Consistent differences in expectations-formation, sophistication, and application behaviors across SES and related variables have also been documented in other studies (e.g., Hoxby and Turner (2015)).

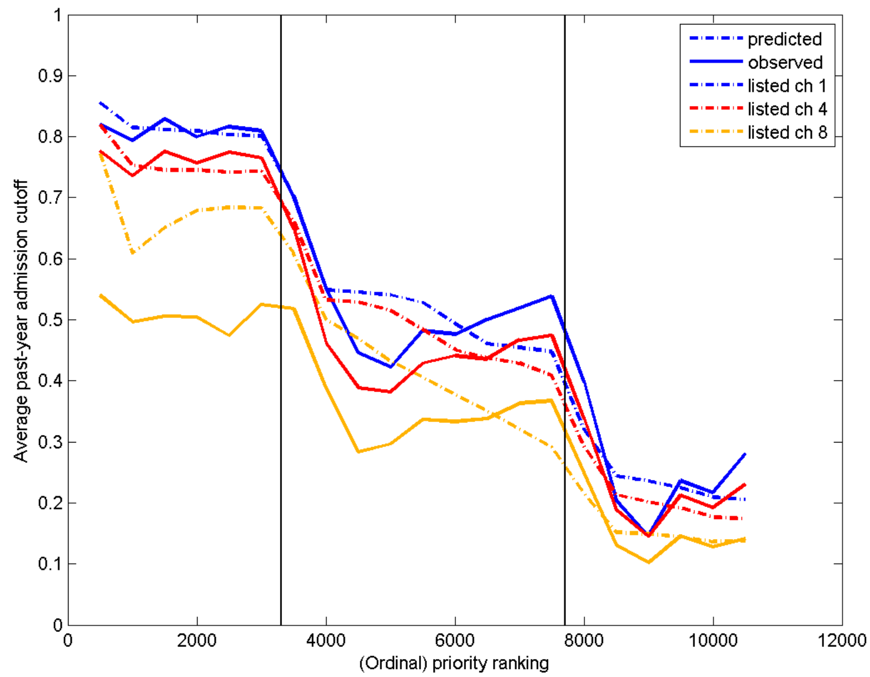
Among sophisticated students, most students, regardless of SES, form expectations about their admission chances using an AR(1) whose parameters differ (at least) for programs in different fields of study. About 12% of both low- and high-SES students use an AR(1) whose parameters differ only across programs' fields of study. In addition, 26% of high-SES students and 16% of low-SES use a finer AR(1) process allowing different parameters not only across field of study, but also across programs' capacity filling status in the previous year or (quantiles) of selectivity level. 6% of high-SES students and 3% of low-SES students use an AR(1) whose parameters differ only along either or these two dimensions —programs' capacity filling status in the previous year or (quantiles) of selectivity level. Finally, 4% of students, regardless of SES, are estimated to form expectations about their admission chances using an AR(1) specification that does not account for any kind of heterogeneity across programs.

Table 2.8: Estimated shares of expectations formation behaviors

	low-SES students	high-SES students
Unsophisticated	**0.64**	**0.50**
Sophisticated	**0.36**	**0.50**
Homogeneous AR(1)	0.04	0.04
AR(1) w/ heterog. across fields	0.12	0.13
AR(1) w/ heterog. across capacity filling status in 2009	0.02	0.03
AR(1) w/ heterog. across selectivity levels	0.01	0.03
AR(1) w/ heterog. across field and cap. filling in 2009	0.08	0.12
AR(1) w/ heterog. across field and selectivity level	0.08	0.14

Model fit

Figure 2.3 plots the selectivity level of students' choices as a function of their priority ranking. For clarity, it focuses on students' first, fourth, and eighth-listed choices. Solid lines represent choices observed in the data; dotted lines represent choices predicted given utility parameter estimates from Section 2.4 and estimated expectations-formation types shown in Table 2.8. Figure 2.3 suggests that the estimated model is able to reproduce the variation in the selectivity level of students' listed choices as a function of their priority rank; and predicts reasonably well the diversification of students' application portfolios in terms of selectivity level.



This graph plots the selectivity level (in terms of past-year admission cutoff) of students' choices as a function of their priority ranking —focusing on students' first, fourth, and eighth-listed choices. Solid lines represent choices observed in the data; dotted lines represent choices predicted given utility parameter estimates from Section 2.4 and estimated expectations-formation types shown in Table 2.8.

FIGURE 2.3: Selectivity level of predicted vs. observed choices

2.6 Understanding the value of information

In this section, I evaluate the effects of informational updates in a restricted-list DA mechanism. To evaluate these effects, I use simulations, and compare students' outcomes under the standard single-phase implementation of the DA and alternative multiple-phase implementations of the mechanism, in which information about available vacancies is publicly updated between phases.

Welfare. I measure average student welfare as the uniformly weighted sum of utilities that students derive from their assignment (indirect utilities). Given the distribution of preferences, a simulation exercise allows me to compute the expected average student welfare induced by a given mechanism. Welfare comparisons made in this section are based on expected average student welfare:

$$W = \frac{1}{N}\sum_{i=1}^{N}\left[u_{i,\mu(i)} + \varepsilon_{i,\mu(i)}\right]$$

where $\mu(i)$ denotes student i's assignment, and the expected value is estimated by simulations over ε.

Counterfactual scenarios. I comparatively evaluate the effects of informational updates by considering four multiple-phase scenarios: dividing the cohort in two, three, four, or five groups by order of priority. When simulating applications in the three-phase mechanism, I divide the cohort into groups as was done in the 2010 Tunisian mechanism (i.e. top 30, middle 40, and bottom 30%). When simulating applications in the two-, four-, and five-phase mechanisms, I divide the cohort in equally-sized groups. As a benchmark, I also simulate applications in a perfect information setting, publicly updating vacancies after every single assignment. This corresponds to a limit N-phase scenario, where N is the total number of students in the population. The number of phases is the only difference between the scenarios I simulate.

I first show that, when applying under the single-phase restricted list DA, and relative to the perfect-information benchmark, their average expected indirect utility is significantly decreased. While easy to implement, the 2010 Tunisian three-phase implementation of the restricted-list DA reduces by 67% the welfare loss induced by the implementation of a standard (single-phase) restricted-list DA, relative to the perfect information benchmark. Investigating the mechanisms underlying these changes in indirect utility, I show that expected indirect utility gains essentially accrue to students who fail to be admitted to any of their listed elements under a single-phase mechanism and who gain assignment because of the informational updates —rather than to assigned students improving their match. Exploring heterogeneous effects across students with different ability, sophistication, and socioeconomic backgrounds, I find that gains disproportionately accrue to low-ability, unsophisticated, and low-SES students. In fact, providing information about vacancies, even through a small number sequential of sequential phases, reduces the expected indirect utility gap existing between high- and low-SES students. Finally, while the 2010 Tunisian implementation of the three-phase procedure does increase welfare and the average match rate, I show that a better targeting of low-priority students by the information provision —through a different division of the cohort into three groups— could increase gains to students.

2.6.1 Effects of information-revelation on expected welfare and assignment rates

Welfare

Average welfare gains. Figure 2.4 shows, as a function of the number of phases implemented, the difference in student welfare relative to the single-phase scenario. The horizontal dotted red line shows the difference in expected average student welfare between perfect information and to the one-phase implementation. A positive difference in welfare means that, on average, students derive more utility from their assignment under a multiple-phase mechanism than under the standard single-phase DA. Under the perfect-information benchmark, the average indirect utility is higher than in the single-phase DA by the equivalent of a 41km-reduction in distance traveled. As a reference, distance to actual assignments has mean 145km ($\approx$ 90 miles), median 107km, and standard deviation 200km in the data. Comparing alternative multiple-phase scenarios suggests that welfare gains increase as the number of information revelations made increases, although the marginal value of an extra phase seems to be decreasing. Under the two-, three-, four-, and five-phase scenarios, indirect utility gains average an equivalent of about 17km, 28km, 31km, and 34km travel-distance reductions, respectively. In particular, this means that a three-phase implementation of the DA, as done in Tunisia, reduces by about 68% the loss in average utility induced by using a single-phase restricted-list DA in a environment where students face uncertainty about their admission chances, relative to the perfect-information benchmark.

Distribution of indirect utility gains. Table 2.9 reports selected quantiles of the distributions of indirect utility gains under each scenario, relative to the single-phase implementation. Under each multiple-phase scenario, the range of gains is pretty large. Under the 2010 three-phase mechanism for instance, the first and last percentiles of indirect utility gains are equivalent to a 41km *in*crease in distance traveled, and a 192km *de*crease in distance traveled, respectively. Although individual losses can be significant, under each multiple-phase scenario, the share of students hurt by the multiple-phase mechanism (relative to the standard single-phase DA) is relatively small as compared to the share of students who (weakly) benefit from it (null gains correspond to the 15th, 11th, 10th, and 9th percentile under the two-, three-, four-, and five-phase scenario, respectively). Interestingly, the perfect infor-

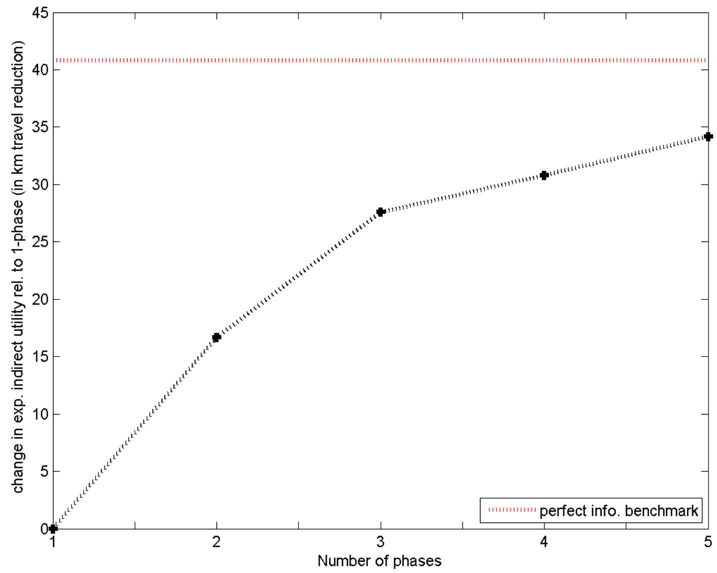


FIGURE 2.4: Change in expected average student welfare relative to single-phase implementation of the restricted-list DA

mation setting does not constitute a Pareto improvement relative to the single-phase implementation of the DA. Indeed, 8% of students derive less utility from the assignment they obtain under perfect information than under the single-phase DA. The next section sheds light on the mechanisms underlying these facts.

Table 2.9: Distribution of utility gains in *ex-post* flow utility (in km): selected quantiles

Percentiles	1st	5th	25th	50th	75th	95th	99th
Two-phase	-67	-22	0	5	25	93	147
Three-phase	-41	-5	0	7	38	134	192
Four-phase	-28	-4	1	9	43	143	200
Five-phase	-16	-3	1	9	49	152	207
Perfect Info.	-8	-2	1	11	62	174	230

Underlying mechanisms

There are two margins through which a student's indirect utility may change from the single-phase implementation of the DA to a multiple-phase implementation. The first is a change in assignment status; the student fails to be assigned to any element of her list under the single-phase implementation (and is therefore administratively assigned to a left-over seat) while she managed to be assigned to one of the programs she listed under the multiple-phase implementation —or vice-versa. The second is a change in assignment, holding the assignment status fixed; under both implementations, the student is assigned to one of her listed elements, but the program she is assigned to changes —or symmetrically, she is administratively assigned in both cases but her assigned left-over program changes. The former can be seen as a change in indirect utility at the extensive margin, and the latter as a change at the intensive margin. Table 2.10 decomposes the welfare gains shown in Figure 2.4 into these two mechanisms. It shows the average share of students who, under the multiple-phase scenarios and relative to the single-phase implementation, switch assignment status, and the share of those who do not. Table 2.10 also shows the average change in indirect utility experienced by students with each of the four assignment-status pairs. Figure A.5 in Appendix A.5 provides additional information by showing the distribution of indirect utility changes (relative to the single-phase DA) within each assignment-status-pair group.

Table 2.10 shows that an increase in the match rate is the main mechanism underlying the increase in average indirect utility induced by the revelation of information. Under the single-phase implementation, 9.1% of students end up administratively assigned. These students all gain assignment (i.e. match) under the perfect information benchmark. As a consequence, they experience a large average expected indirect utility gain —equal to more than 10 times the population-average expected indirect utility gain. By contrast, the 90.9% of students who are assigned under both the single implementation and the perfect information setting experience on average little expected indirect utility changes. As Figure A.5 shows, students who are assigned under both scenarios may experience an increase or decrease in expected indirect utility when information in revealed. Increases in indirect utility are due to some students failing to apply to a desired program under the single-phase implementation, because they expect their admission chances to be low, while the program would actually have had a seat for them (which they are able to claim under perfect information). The slight worsening of some assignments among 'always-assigned' students is the result of equilibrium effects, given programs' finite capacities. Some students with higher-priority improving their assignment or gaining assignment takes away from lower-priority applicants spots that are available under the one-phase implementation. That is, some students are better off under the single-phase mechanism than under a perfect-information setting (last row of Table 2.9) because they benefit from other students' misplacement in the absence of informational updates.

While under perfect information no one is administratively assigned, a few students fail to be admitted to any element of their list under the other multiple-phase scenarios. Among the 9.1% of students administratively assigned under the single-phase scenario, 82% (hence 7.5% of the population) switched to being assigned to an element of their list when applying under the 2010 three-phase mechanism. These students experience large welfare gains. The other 18% of the students administratively assigned under the single-phase scenario keep this assignment status under the three-phase scenario. On average, these students experience a decrease in expected indirect utility —in magnitude lower than the average gain. This is again the consequence of equilibrium effects. As more students get assigned to desired programs, leftover programs that remain available for administrative assignment get worse[45] (as suggested by the larger loss experienced by 'never-assigned' students when moving from one to five phases, than to three phases). Finally, a few students fail to be assigned to any element of their application portfolio under the multiple-phase mechanism while they were under the single-phase mechanism —less than 1% (resp. .5%) of students when switching from one to three (resp. five) sequential phases.

Table 2.10: Changes in assignment status and associated changes in utility

	Changes on the extensive margin				Changes on the intensive margin			
In 1-phase, students are …	*… unmatched*		*… matched*		*… matched*		*… unmatched*	
In multi.-phase, students are …	*… matched*		*… unmatched*		*… matched*		*… unmatched*	
	%	ΔW	%	ΔW	%	ΔW	%	ΔW
Perfect info.	9.1	510	0	–	90.9	-6	0	–
Three-phase	7.5	544	0.9	-648	90	-8	1.6	-31
Five-phase	8.5	523	0.4	-635	90.4	-8	0.6	-38

2.6.2 Heterogeneous effects by ability, sophistication and SES

Gains and priority ranking

The left panel of Figure 2.5 plots indirect utility gains as a function of students' priority ranking under the three- and five-phase DA (relative to the single-phase DA), and under the perfect-information benchmark. Other multiple-phase scenarios yield similar graphs. Close to all welfare gains accrue to students in the second half of the

[45] No explicit rule is provided relative to how administrative assignments are made. In simulations, for each student who fail to be assigned to any element of her list, I randomly set the administrative assignment using a uniform distribution over the 50% closest seats that are leftover at the end of the DA. Random administrative assignments using a uniform distribution over all leftover seats yield similar results.

priority distribution. The right panel of Figure 2.5 plots students' probability to be assigned to one of their listed choices as a function of their priority ranking under the various multiple-phase implementations. It echoes the findings of the previous section that the larger gains in indirect utility accrue to students with larger drops in their probability to be administratively assigned. These graphs highlight the fact that, when students do not know their true admission probabilities, the accuracy of their expectations decreases as the state of the world to be forecast gets further from the one they have information on. Under the single-phase restricted-list DA, the lower a student's priority ranking, the larger the number of students to be assigned before her. In other words, the larger the number of random events (assignments) to alter the initial state of the world before she gets to be assigned. The revelation of information, as done in the Tunisian mechanism, increases low-priority students' average expected indirect utility by bringing them 'closer' to up-to-date information.

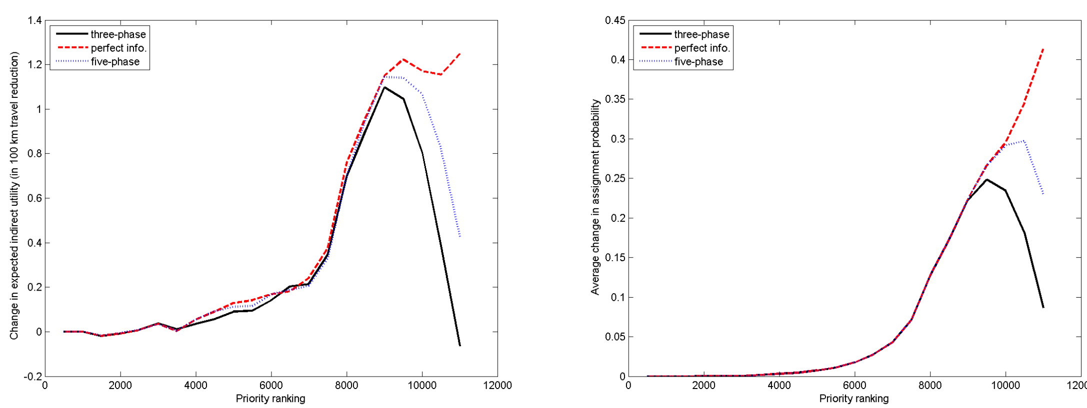


FIGURE 2.5: Changes in indirect utility and assignment probability as a function of priority ranking

As mentioned earlier, the sequential implementation of the DA may affect applicants' behaviors through two channels. The provision of information about programs filling up enable (later groups) students to update their expectations about their admission chances. In addition, the introduction of a 'second chance' to (early group) students if they fail to be assigned to any of their listed choices increase students' option value of being rejected. Gains being small in the first part of the priority ranking (and in particular in what becomes Group 1 under the three- and five-phase scenario) suggests that the latter channel has a small effect on welfare relative to the former.

Under the three-phase scenario, average welfare gains are negative at the very end of the priority ranking; the very last students are the most likely to have their assignment worsened as a result of equilibrium effects, while at the same time remaining relatively far from the information and so relatively likely to fail to be assigned to any of their choices.

Gains and sophistication

Figure 2.6 is analogous to Figure 2.5 but shows average welfare gains as a function of priority separately for sophisticated and unsophisticated students. The left panel shows that increases in the match rate are mostly experienced by unsophisticated students. Most sophisticated students maintain their assignment status relative to the single-phase scenario. At the very end of the priority ranking though, sophisticated students, on average, experience a decrease in their match rate. This is the result of equilibrium effects described earlier; as higher-priority students who end administratively assigned under the single-phase implementation manage to gain a match with the sequential implementation, fewer seats become available at the end of the priority ranking, increasing students probability to be rejected from all their listed choices. This also explains why the increase in match rate dips down at the end of the priority ranking for unsophisticated students.

The right panel shows that largest increase in average indirect utility accrue to unsophisticated students. Given that unsophisticated students are those who benefit from an increase in their match rate, this is consistent with the mechanisms established in Section 2.6.1. Among sophisticated students, significant increases in average expected indirect utility are essentially experienced by mid-priority students (ranks 3,000–9,000 under three phases; ranks 3,000–10,000 under five phases). The decrease in average expected indirect utility experienced by sophisticated students at the bottom the priority ranking experience follows the same causes as the decrease in their match rate.

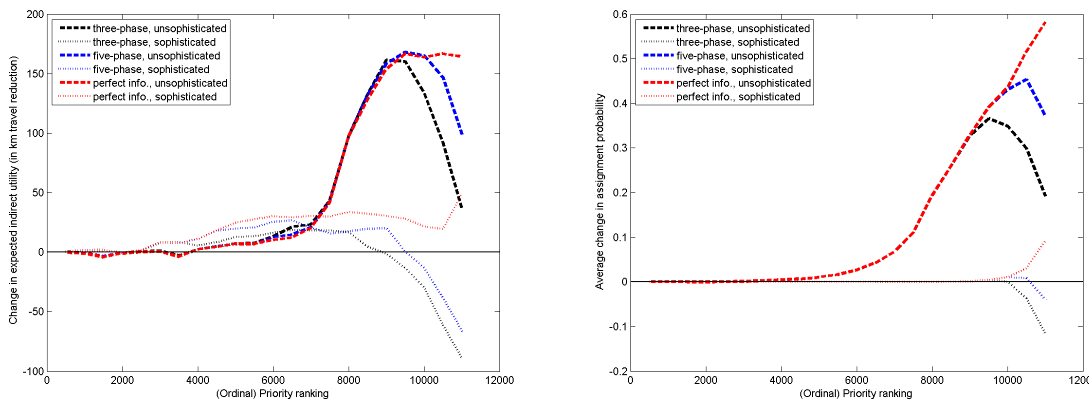


FIGURE 2.6: Changes in indirect utility and assignment probability as a function of priority ranking by sophistication type

Gains by demographics

The previous two paragraphs have established that the extent to which a student gains or not from the revelation of information depends on her position in the priority ranking, and the level of sophistication with which she forms expectations about her admission chances. Statistics from Table 2.1 and estimates from Table 2.8 show that a student's position in the priority ranking and the level of sophistication of her beliefs are correlated with her socioeconomic background. As a consequence, welfare gains from information revelation differ across SES.

The left panel of Figure 2.7 shows, as function of the number of sequential phases implemented, the average expected indirect utility gains experienced relative to the single-phase scenario, separately for low-SES (thicker dashed black line) and high-SES (thinner dotted black line) students. The horizontal red lines show, separately for low- (thicker dashed red line) and high- SES (thinner dotted red line), the difference in expected average student welfare between the perfect-information and one-phase settings. On average, and in terms of expected indirect utility, low-SES students benefit more from the information revelations than high-SES students do — in other words, low-SES students are more hurt by the single-phase implementation of the DA than their high-SES counterparts are. Switching from the single-phase implementation to the perfect information setting increases average expected indirect utility for low-SES students by an equivalent of a 59-km reduction in travel distance, against 29km for high-SES students. Switching from the single-phase implementation to the 2010 Tunisian three-phase setting increases average expected indirect utility for low-SES students by an equivalent of a 43-km reduction in travel distance, against 18km for high-SES students.

As a consequence of low-SES students being more hurt by the single-phase implementation of the DA than their high-SES counterparts are, the provision of information, through sequential implementation of the mechanism, reduces the welfare gap existing across low- and high-SES students. The right panel of Figure 2.7 shows, as function of the number of sequential phases implemented, the difference in average expected indirect utility between high- and low-SES students. The horizontal dotted red line show the level of this welfare gap in the perfect-information setting[46].

[46] The welfare gap across SES that persists under the perfect information benchmark is due to low-SES students being matched with lower quality programs (because they are more heavily distributed

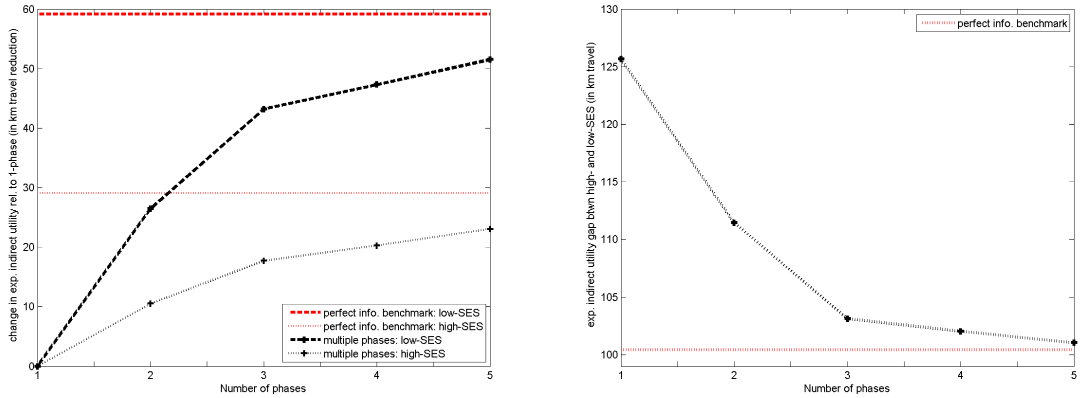


FIGURE 2.7: Changes in average expected indirect utility by SES, and changes in the average expected indirect utility gap across SES as a function of the number of sequential phases

Switching from the perfect information setting to the single-phase implementation increases the welfare gap across SES by 25%. Switching from the single-phase implementation to the 2010 Tunisian three-phase procedure reduces this increase by 88%.

2.6.3 How much information to give? vs. whom to give information to?

The previous subsection established that, even under perfect information or with information being provided early on as in the three- and five- phase scenarios, most of the welfare gains are generated in the second half of the priority ranking. This suggests that, beyond the amount of information provided (i.e. the number of sequential phases implemented), the points in the priority rankings at which revelations are crucial determinants of welfare gains. In this section, I test this hypothesis by comparing student welfare and match rate under different three-phase scenarios. While I hold

at the bottom of the priority ranking), and having different preferences for program characteristics. In particular, low-SES students have a larger disutility from traveling (see Table 2.7); and, due to lower average high-school performance in both STEM and non-STEM fields, derive on average less utility from the different fields of study (see Table **??**).

the number of sequential phases constant, the division of the cohort in application groups differ across scenarios. I compare the 2010 Tunisian design, in which groups correspond to the top 30%, middle 40% and bottom 30% on the priority distribution (denoted ‘30/40/30’), to divisions that allow to focus information provision on the lower end of the priority ranking —50/25/25 and 50/37.5/12.5.

The right panel of Figure 2.8 reproduces Figure 2.4. The horizontal dotted red line shows the difference in expected average student welfare under perfect information, relative to the one-phase implementation. The black dotted curve shows welfare gains for the multiple-phase scenarios documented in Section 2.6.1. On this curve, the black dot at the three-phase mark of the horizontal axis represents welfare gains, relative to the single-phase scenario, achieved under the 30/40/30 three-phase implementation. The blue and green dots at the three-phase mark represent welfare gains, relative to the single-phase scenario, achieved under the 50/25/25 and 50/37.5/12.5 three-phase implementations, respectively. The colored dots being above the mark-three black dot means that, relative to the single-phase scenario, average expected indirect utility is increased more under the latter two implementations than by the 2010 Tunisian three-phase procedure. Switching from the single-phase restricted-list DA to the 50/25/25 (resp. 50/37.5/12.5) three-phase scenario achieves 76% (resp. 90%) of the average expected indirect utility increase generated by switching from the single-phase restricted-list DA to a perfect information setting. That is, the 50/25/25 scenario achieves as much as the four-phase implementation documented in Section 2.6.1; and the 50/37.5/12.5 scenario achieves more than the five-phase implementation. (To ease comparison, the two horizontal thinner black lines show the levels welfare gains achieved by the equally-spaced four- and five-phase scenarios.)

The left panel of Figure 2.8 shows that, again, the main mechanism under the increase in average expected indirect utility is an decrease in the share of students administratively assigned. It shows, as a function of the number of phases implemented, the share of students administratively assigned. The blue and green dots at the three-phase mark show the administrative assignment rate under the 50/25/25 and 50/37.5/12.5 three-phase implementations, respectively —which is in both case smaller than under the 30/40/30 implementation (black dot at the three-phase mark).

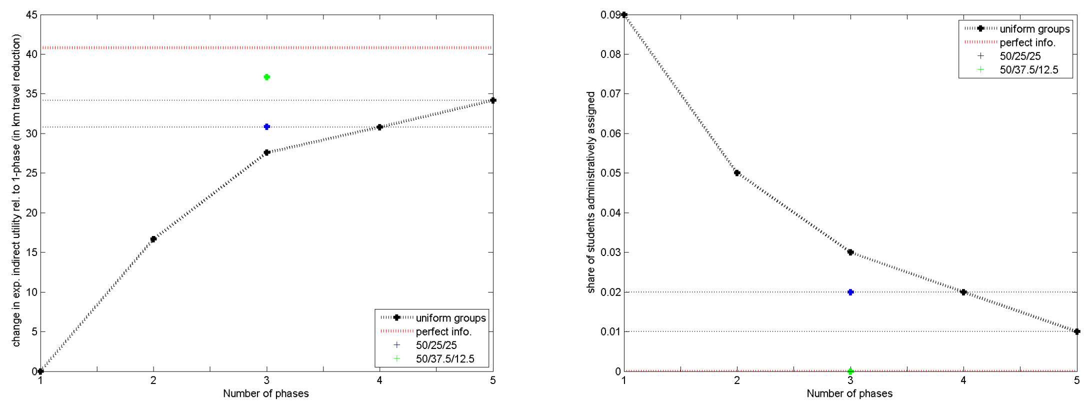


FIGURE 2.8: Change in average expected indirect utility relative to single-phase scenario and assignment rate, as a function of sequential phases

Figure 2.9 is analogous to Figure 2.6. It shows average welfare gains (relative to the single-phase implementation of the DA) and match rate as a function of priority separately for sophisticated (thinner plots) and unsophisticated (thicker plots) students. Plots are showed for the perfect-information benchmark (red dotted lines), as well as for the three alternative three-phase scenarios. The black plots are the same as those in Figure 2.6; they show outcomes for the 30/40/30 implementation. The blue dotted plots and green line plots show outcomes for the 50/25/25 and 50/37.5/12.5 three-phase implementations, respectively. The left panel shows that the 50/37.5/12.5 implementation allows to maintain the upward-sloping increase in matching rate through the entire priority ranking, where the increase dips under the

30/40/30 and 50/25/25 scenarios (and under the five-phase implementation, see the left panel of Figure 2.6). The increase in matching rate in particularly high for unsophisticated students, but it also benefits very-low-priority sophisticated students —by contrast, these students experience a decrease in match probability (relative to the single-phase implementation) under the other multiple-phase scenarios documented here.

As shown in the right panel of Figure 2.9, the upward-sloping increase in match rate translates, for unsophisticated students, into a larger increase in average expected indirect utility than under the other multiple-phase scenarios. Just as they do in other multiple-phase implementations, very-low-priority sophisticated students experience, under the 50/37.5/12.5 implementation, a decrease in average expected indirect utility relative to the single-phase scenario. However, this average decrease starts much later (about 1,500 ranks later) in the priority ranking than it does in the other three-phase scenarios, meaning that a larger share of students experience increases in expected indirect utility than under other multiple-phase scenarios. Furthermore, the existing decreases in expected indirect utility (relative to the single-phase DA) are, on average, smaller in magnitude under the 50/37.5/12.5 implementation than under other implementations. In fact, the late provision of information (at the 87.5 percentile of the priority distribution) mitigates the negative equilibrium effects affecting low-priority sophisticated students by enabling them to have more accurate expectations about available seats at the very end of the assignment process.

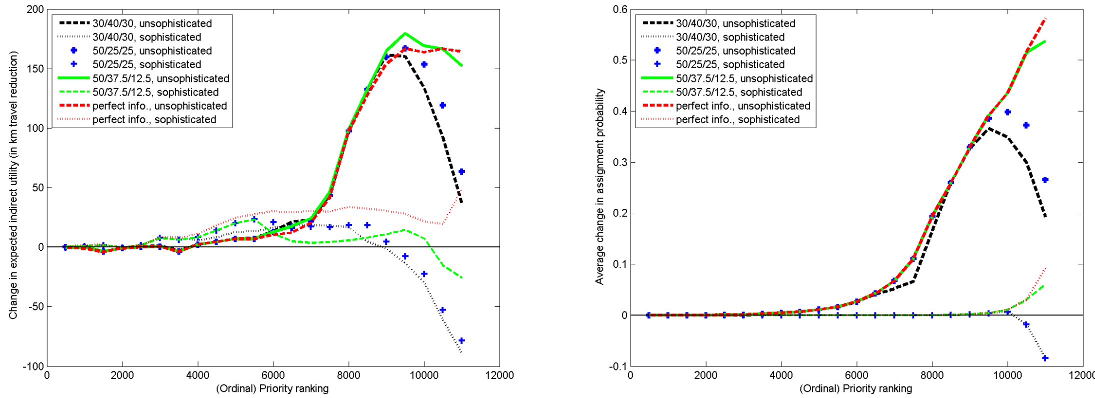


FIGURE 2.9: Changes in indirect utility and matching probability as a function of priority ranking by sophistication type

2.7 Conclusion

This paper quantifies the welfare effects of enabling students to update their expectations about their admission chances to academic programs in a setting where they cannot apply to all the alternatives in their choice set. It documents a simple way to enable this updating in the context of DA-based assignment mechanisms, which are extensively used around the world to assign students to schools. I estimate a model of application portfolio choice and perform a counterfactual analysis to compare students' application and assignments under scenarios with different levels of updating. I take advantage of a rich administrative data set from Tunisia, where a variant of the DA is used nationwide to assign high-school graduates to university programs. Building on the quasi-experimental design induced by the Tunisian procedure, I am able to recover students' preferences for university programs without taking a stand on their expectations, hence circumventing a common identification challenge in the empirical literature on school choice. I then take preferences as given and characterize students' expectations. This two-step approach partially alleviates the computational intractability of the application portfolio choice problem —the other main challenge faced by the empirical literature on school choice.

I combine preferences estimates and findings about expectations to show that, while easy to implement, a sequential version of the DA can reduce the welfare loss and inequality induced by the standard restricted-list implementation. The 2010 Tunisian three-phase implementation of the restricted-list DA reduces the welfare loss induced by the implementation of a standard (single-phase) restricted-list DA by 67%, relative to the perfect information benchmark. Gains disproportionately accrue to low-ability, unsophisticated, and low-SES students; so providing information about vacancies, even through a small number sequential of sequential phases, reduces the expected indirect utility gap existing between high- and low-SES students. My results suggest that, while the 2010 Tunisian implementation of the three-phase procedure does increase welfare and the average match rate, a better targeting of low-priority students by the information provision —through a different division of the cohort into three groups— could increase gains to students.

The findings of this paper show that a simple twist in the implementation of the DA can effectively mitigate the consequences of imperfect knowledge of admission chances on welfare and inequality. No data is available on the implementation costs of the sequential procedure, and it is therefore not possible to rigorously compare the costs and benefits of such a policy. However, the small number of sequential phases needed to restore a large share of the loss suggests that the benefits of a sequential implementation are likely to exceed its costs. The Tunisian example also demonstrates that the information revelation intervention can be implemented at a large scale.

The application problem faced by Tunisian high school graduates is no different from the one faced by students in other places. The inefficiencies that can arise with the single-phase implementation of the restricted-list DA, such as a large rate of administrative assignment, have been documented in other instances (e.g., Ajayi and Sidibé, 2016). It is likely that a sequential implementation of the DA would have the same benefits there as in Tunisia —at a limited cost since fixed costs are already being paid in the non-sequential implementation of the DA. Beyond school and college choice, the DA is used as an assignment mechanism in other contexts where information is imperfect and application is costly. A sequential implementation may improve outcomes there as well.

This paper takes as given and fixed the restrictions placed, in virtually every school choice implementation of the DA, on the number of alternatives students can list in their application portfolio. There is evidence that while they value the strategy-proofness implied by the DA when no constraint in imposed on the size on the application portfolio, policy-makers have been unwilling to lift list size restrictions (Pathak and Sönmez, 2013; Roth, 2015). An interesting question for future research is to understand the reasons behind this reluctance. In settings where the choice set faced by students is large (such as in Tunisia, but also for instance in NYC where students can choose from 700+ high schools), a natural hypothesis is that it is costly for applicants to process information, learn about, and precisely assess their preferences for all existing alternatives. In this case, allowing students to downsize their choice set, by revealing which programs are full by the time they get to apply, may be an additional benefit of the sequential design.

Chapter 3

Do Elite Schools Improve Students Performance?

This chapter is based on a joint work with Meryam ZAIEM.

3.1 Introduction

This paper is concerned with identifying whether elite schools may have a lasting effect on students' trajectories. Over the past few decades, there has been a renewed interest in selective schools and other schooling options such as charter schools, as part of a larger move toward giving parents more flexibility in their children schooling decisions. This resurgence of popularity calls for the evaluation of existing systems, and possibly, the development of new, improved ones. In this perspective, it is imperative to have a sense of the strengths and weaknesses of the intervention design, and a comprehensive appreciation of the mechanisms at play. Identifying the value-added of schools, to be disentangled from the selected students' high ability, is critical.

The present analysis uses nationwide administrative data from Tunisia. It takes advantage of a special feature of the Tunisian secondary education system that offers a quasi-experimental setting to study the impact of being admitted to one of the twelve[1] selective public high schools in Tunisia: students are selected into these elite high schools based on their results on a national exam, which creates admission cut-offs that we exploit in a sharp regression discontinuity (RD) design.

This paper expands the already large existing literature on reduced-form effects of selective schools by considering longer-term outcomes. The literature has mostly focused on outcomes measured no later than the end of high school. In Tunisia, high-school graduates apply and are assigned to post-secondary programs via a centralized State-run mechanism. The mechanism is a modified version of the college-proposing Deferred Acceptance algorithm (Gale and Shapley Gale and Shapley (1962)), and involves students submitting rank-ordered preference lists of post-secondary programs they would like to attend. By linking our high-school data with rich data on this post-secondary assignment procedure, we are able to document the effect of admission to an elite high school on the field, location, and selectivity level of the programs they apply and get assigned to.

Investigating the effects of admission to an elite high school on college application and admission outcomes, which realize conditional on students graduating from high school, requires paying special attention to selection and composition biases that may arise if admission to an elite high school affects high school graduation rates in the first place. We highlight that this mechanism can cause the standard RD analysis to produce biased estimates; and account for this fact to produce bounds for the treatment effects.

We find that admission to an elite high school in Tunisia does have significant posi-

[1] Twelve during the period considered (cohorts entering high school between 2006 and 2008); their number increased since then –there were 14 selective high schools in 2012–13.

tive effects on high-school performance outcomes, and that, for some students, these effects do translate into changes in post-secondary assignments. End-of-high-school exam scores are increased for students who would take the exam irrespective of their treatment status[2]. Girls' retention rate is decreased: marginally admitted girls are (at least) two percentage points more likely than their non-admitted counterparts to take the end-of-high-school exam within four years of starting high school. We find that, on average, marginally admitted and non-admitted students do not have significantly different post-secondary application portfolios in particular in terms of the selectivity of the programs applied to. Allowing heterogeneity in treatment effects along the outcome distribution, we find that positive treatment effects on high school performance occur mostly at the top quartile of the distribution.

These conclusions are built on estimates accounting for sample selection and missing outcome data which plague our data. When considering treatment effects (TEs) on end-of-high-school exam score and post-secondary outcomes, we face the same sample selection issue as Lee (2009) in his influential study of the wage effects of the Job Corps program. Test scores are observed only for students who did take the exam; post-secondary outcomes are observed only for students who did pass the exam. If exam-taking or exam-passing are affected by treatment, then comparing mean exam scores and tertiary outcomes across treatment groups may fail to capture the causal treatment effect. In addition, our exam-taking variable are affected by non-classical measurement error due to failures in the linkage of the databases at the heart of our analysis. We highlight in this paper that sample selection and missing outcome data, two pervasive issues in applied work, induce the standard RDD identification argument to fail, and the naive RD techniques to produce biased estimates. The importance of correcting for these biases is made clear by the stark

[2] We will call them always-exam-takers.

differences between the conclusions we derive from our bias-corrected estimation and those one would come to based on naive RD estimates. Naive estimates suggest that girls benefit from admission to an elite high school only through a decreased retention rate. In particular, due to unaccounted for composition effects, they fail to detect the significant increase in their end-of-high-school exam scores induced by treatment. From naive estimates, one would also infer that admission to an elite high school has large and significant effects on boys' outcomes beyond high school; in particular, marginally admitted male students seem to apply and get assigned to more selective post-secondary programs.

Related literature

This paper naturally pertains to the literature on elite schools, though its methodological standpoint and its consideration of an extended set of outcomes also make it relevant to broader fields of research. The effects of admission to and/or attendance at a selective school on short-term effects have been widely documented, however, a review of this literature shows little consensus. On the one hand, using data from Trinidad and Tobago, Jackson (2010) finds that attending a middle school with high achieving peers has a positive effect on students' grades in English and their probability to pass the end-of-middle-school certification exam (although no effect on the probability to take it). Pop-Eleches and Urquiola (2013), using Romanian data, also find a positive impact of attending a higher quality school on Baccalaureate grades, along with no effect on the probability of taking the exam. Estrada and Gignoux (2014) find that admission to the Instituto Politécnico Nacional system in Mexico results in significantly improved performance in mathematics as well as more positive attitudes toward school. On the other hand, in a recent paper looking at Boston and

New York exam schools, Abdulkadiroğlu et al. (2014) find that attending a school with higher quality peers does not significantly affect test scores. Ajayi (2014) similarly finds that assignment to a high school with higher achieving peers in Ghana results in small improvements in exam performance at the end of secondary school. Clark (2010), studying the effects of attending a selective grammar school in the UK, finds no effect on Reading, Math and Science test scores. Using evidence from randomized lotteries in Chicago Public Schools, Cullen et al. (2006) find that attending a high school with higher attainment rates and peer quality does not lead to improved cognitive outcomes.

Very few of the papers interested in selective schools discuss post-secondary outcomes. Dobbie and Fryer Jr (2011) find that admission to an exam school in New York has little effect on the rate at which students later enroll to college, and on the quality of the colleges they attend. The same is found for Boston exam schools by Abdulkadiroğlu et al. (2014), who also show an absence of effect on the number of Advanced Placement tests taken and on the score obtained at these exams. In the UK context, Clark (2010) finds that pupils attending a grammar school are more likely to take courses from the academic curriculum in high school, which is known as a good preparation to university courses; furthermore, boys show a higher, although imprecisely estimated, university enrollment rate[3]. Taking advantage of the

[3] In a different context, looking at the effects of gaining admission to a charter school in Boston, Angrist et al. (2016) are able study a broader set of outcomes and do find some positive results. Admitted students are found to be more likely to take Advanced Placement exams, particularly in science; to get higher SAT scores; and to attend 4-year (rather than 2-year) colleges at higher rates, although they do not seem more likely to enroll in college, overall. In a more descriptive analysis, Booker et al. (2009) also document that students attending charter middle and high schools in Chicago and in Florida have an increased likelihood of enrolling in college within five years of starting high school.
This focus on outcomes measured no later than graduation from the selective school is driven by data limitations. Students are often hard to follow beyond graduation as comprehensive datasets on post-secondary stages of life that can be linked with high-school data may not be available to the researcher.

centralized market for university applications and admissions in Tunisia, this paper is able to investigate a broader set of university programs characteristics when documenting the effects of admission to an elite high school on post-secondary outcomes.

Even beyond elite schools, few studies focus on the effects of high school curriculum on college choices (see Altonji et al. (2012) for a survey of the literature on the post-secondary effects of high school curriculum; and Altonji et al. (2015) for a recent review of the literature on the determinants of and returns to college major choice). As reviewed by Altonji et al. (2012), complex selection issues, intertwined and diverse administrative policies and requirements, as well as data limitations make identifying a causal relation between high school curriculum and post-secondary choices challenging. In this paper, the reduced-form analysis of students' post-secondary applications and assignments provide new nationwide empirical evidence that high school programs and curricula do causally affect post-secondary trajectories in some dimensions.

Finally, this paper relates to the literature on RD designs. Over the past decades, the popularity of these designs has grown substantially in economics (see surveys by Imbens and Lemieux (2007), van der Klaauw (2008), and Lee and Lemieux (2010)). The identification argument resulting from the RD design is intuitive; and methods have been developed making empirical use and estimation simple (see for instance Imbens and Lemieux (2007) and Imbens and Kalyanaraman (2011), as well as Frandsen et al. (2012) about quantiles treatment effects). Sample selection and attrition have been noted to jeopardize the consistency of RD estimates. To clear concerns, past studies have mainly relied on parametric approaches (e.g. McCrary and Royer (2011), Martorell and McFarlin (2011)) or the argument that a non-significant average effect of treatment on the selection variables is evidence of negligible sample selection bias (e.g. Pop-Eleches and Urquiola (2013), Estrada and Gignoux (2014), Jackson (2010)). In a study of the effects of developmental mathematics courses, Kim (2012) takes a partial identification approach, extending Lee (2009) to sharp RD designs (Jackson (2010) also mentions using Lee (2009) bounds as a robustness check). Our approach differs in the assumptions we make. We argue that the central assumption underlying Lee (2009) bounds (monotonicity) is very strong in our context and, in the spirit of Zhang and Rubin (2003), suggest an alternative way to refine the identified set. Closest to our methodological approach to sample selection is the contemporaneous and independent work by Dong (2016). Just as we do, Dong (2016) builds on insights from Zhang and Rubin (2003) to show the partial identification of average and quantile treatment effects in the presence of sample selection in an RD design. While our setting requires the use of a sharp RD, Dong (2016) also provides extensions for the fuzzy RD design and evaluates the effects of placement on academic probation on students' probability to complete college, as well as on their GPA at graduation. While our post-secondary outcomes are affected by sample selection, we are also faced with missing data in our end-of-high-school outcome (exam-taking). Missing data on this binary outcome amounts to non-classical measurement error, and results in biased standard RD estimates. We also address this issue by taking a partial identification approach and providing bounds for the treatment effects.

Outline

The paper is organized as follows. The next section describes the Tunisian education system and presents the data. Sections 3.3 and 3.4 expose our treatment effects identification and estimation strategies. Section 3.5 shows corrected estimates accounting for selection and missing outcome data, and highlights the bias plaguing standard results by comparing them with naive RD estimates. Section 3.6 concludes.

3.2 Institutional background & Data

3.2.1 Application and admission procedure to Tunisia's *Lycées Pilotes*

In the public system, students are assigned to a (non-selective) high school based on the middle school they attended: each middle school is associated with a high school, and students graduating from this middle school pursuing to secondary education are assigned to the associated high school[4]. Application to selective high schools also depends on the school district of the student's middle school. Figure 1.2 in Section 1.3.1 illustrates the location of the *lycées pilotes* in Tunisia over the period considered (2006–08). Appendix B.1 provides details about application regions. At that time, there were twenty-six school districts in Tunisia; and twelve selective high schools, each in a distinct district. In 18 (19 in 2008) districts, students could *not* choose the selective school they would apply to, as their middle school was associated with only one selective high school, located either in their own school district or in a neighboring district. In our data, this corresponds to 58% of the applicants[5]. The remaining 42% of applicants, attending middle school in the other 8 (7 in 2008) districts, had the choice between two to four selective schools, and could formulate an ordered preference list.

Students graduating from middle school who apply to an elite high school receive admissions offers based on their scores on the national end-of-middle-school exam,

[4] There exist exceptions, for instance if the student's family moves.

[5] More precisely, 57% in 2006, 55.3% in 2007 and 61% in 2008.

according to following process.

1. 9th-graders decide whether or not to take the *Brevet* and whether or not to signal their interest in attending an elite high school[6]. If they have the choice between several selective high schools, applicants provide a ranking. At the end of June, interested students take the exam in the middle school they attend.
2. Prior to the exam, the Ministry of Education decides on the number of spots available in each selective school for the next academic year.
3. Once middle-school exam scores are known, students who signaled their interest in attending an elite high school are nationally ranked based on their score.
4. Students are then assigned to elite high schools following a *serial dictatorship algorithm*: The first-ranked student is assigned to her most preferred (or unique choice of) school. Once the $(k-1)^{th}$ student in the list has been assigned, the student ranked k^{th} is assigned to her most preferred (or unique choice of) school if there is a spot left. If not, and if she listed a second preferred school, she is allocated to this second school provided a spot is available. If not, the process goes down her preference ranking until finding an available spot. If no spot is available in any of her listed schools, the student is declined admission to any selective school and is assigned to the non-selective public high school associated with her middle school. The assignment process continues until all spots in selective schools are allocated or all applicants are assigned.

[6] Application starts early in the last year of middle school. At the end of the fall semester, students are asked about their interest in attending a *lycée pilote*. Those who indicate interest are then given information about the application process: they are given the list of selective schools they are eligible to apply to, are told they need to register for the end-of-middle-school exam, and that they will be given an application form later in the spring. Application forms are out filled by the families, and returned to the middle schools which send them back to the Ministry. By May, at the latest, all students willing to take the end-of-middle-school exam register. Finally, mid-June, all exam-takers, irrespective of their application to a selective school, sit in the same conditions.

3.2.2 Data sources & Sample construction

The data at the heart of our analysis consists of two databases provided by the Tunisian Ministry of Education: (i) the official data on the end-of-middle-school exam for 2006 to 2008; (ii) the official data on the end-of-high-school exam for 2010 to 2014. The former database contains information on all the students who took the end-of-middle-school exam between June 2006 and June 2008; it documents their first and last names, date and administrative region of birth, gender, an identifier of their middle school, as well as their end-of-middle-school exam grade, an indicator of whether or not they applied to a selective high school (and their ordered list if they had to submit one), and an indicator of whether or not they got admitted to a selective high school (and an identifier of this school if they applied to more than one). The latter database contains information on all the students who took the end-of-high-school exam between June 2010 and June 2014; it documents their first and last names, date and administrative region of birth, gender, an identifier of their high school (at the date of the exam), as well as their grades at the various tests of the end-of-high-school exam. The core student sample used for the analysis in this paper is the result of linking these two datasets.

In addition, our exploration of college outcomes relies on the official data on post-secondary applications and assignments, which was made available to us by the Tunisian Ministry of Higher Education. This database contains the ordered application lists and assignment of all students applying to post-secondary programs in public institutions in Tunisia in 2010. When looking at post-secondary outcomes, we are constrained by the data to restrict our analysis to the cohort graduating from middle school in Spring 2006 and high school in Spring 2010, and entering college in Fall 2010.

Linking databases. The database on post-secondary applications and assignments can be linked to the high school sample via a unique end-of-high-school-exam identifier. However, no such identifier is used at the time of taking the end-of-middle-school exams. Hence, the matching of the former two databases needs to be made on other variables –students' first and last names, gender, as well as date and region of birth. Beyond the possible non-uniqueness of matches based on these variables, the matching task is made all the more challenging as names are in Arabic, a language in which some characters and accents can casually be omitted, inducing students' names to be possibly spelled slightly differently from one database to the other. Every year, about 17% of middle-school exam-takers are not linked to high-school exam-takers. Among those, one cannot distinguish those who actually did not take the end-of-high-school exam from those who did so but failed to be matched for reasons suggested earlier. This latter subgroup can be thought of as affected by 'missing outcome data'.

Figure 3.1 shows the distribution of the running variable for middle-school-exam-takers that are, respectively, found and not found in the end-of-high-school-exam data, as well as separately for (found) students taking the end-of-high-school exam in four, five and six years. Distributions for students taking the exam in four, five and six years are skewed each time more to the left; which suggests that drop-outs would have a distribution even more skewed to the left. The distribution of the running variable for non-matched students being closest to the rightmost distribution for matched students taking the end-of-high-school exam in four years suggests that failure to be found in the end-of-high-school-exam data is for the most part not correlated with performance at the end-of-middle-school exam, and could be due to data linkage failures. Consequences of this fact will be discussed in Section 3.3.

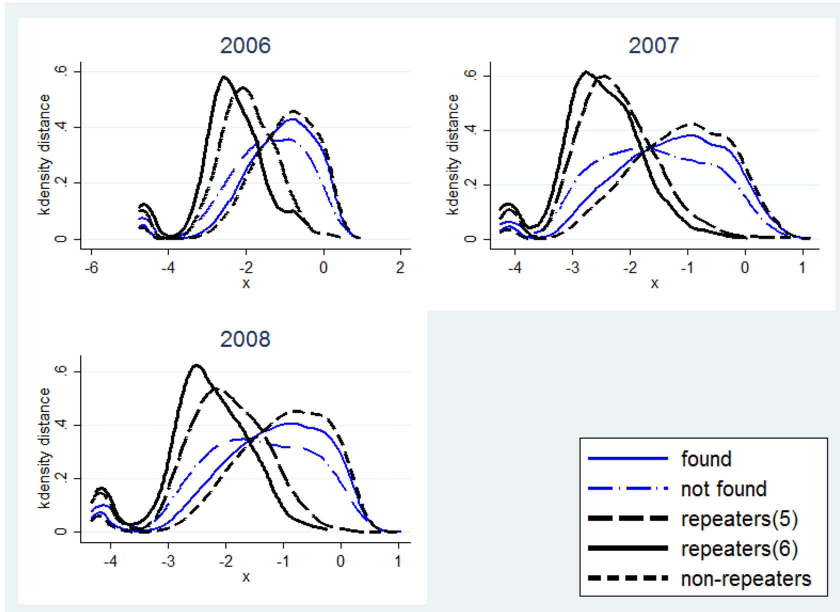


FIGURE 3.1: Distribution of end-of-middle-school exam score per end-of-high-school exam-taking record (and year).

3.2.3 Sample & Outcomes of interest

In this section, we describe the estimation sample and present the outcomes and subsamples of interest. The population consists of the 80,624 students who took the end-of-middle-school exam *and* applied to a selective high school between June 2006 and June 2008 (three cohorts). Among them, 6,645 students were admitted to one of the twelve Tunisian selective high schools. These 80,624 applicants represent more than half of the 151,401 middle-schoolers who took the exam over the period considered. In Appendix B.2, Tables B.1 and B.2 provide a description of the applicants and admitted populations, as well as for the whole population of end-of-middle-school exam takers. Importantly, appendix Table B.2 shows that compliance to the admission decision is far from being perfect. 71% of the students admitted to a selective school and 0.3% of non-admitted applicants are observed to be enrolled in a selective school at the time they take the end-of-high-school exam. From another perspective, 93.5% of the students enrolled in a selective school at the time they take the end-of-high-school exam were indeed admitted to a selective school upon middle school graduation while 6.5% were not. Imperfect compliance will be accounted for in the estimation strategy and discussed in Section 3.3.

Table 3.1: Outcomes: Descriptive statistics –Control Group

	Boys	Girls	P-value	All
End-of-high-school outcomes				
Take exam in 4 y.	0.86	0.85	0.171	0.85
Score at exam	0.62	0.58	0.013	0.60
Final score at exam	0.62	0.59	0.195	0.60
College application outcomes				
2009 cutoff for 1st-ranked	1.66	1.62	0.350	1.64
Log-dist. from 1st-ranked to h-sch.	2.59	2.64	0.589	2.63
College admission outcomes				
2009 cutoff for assigned track	1.23	1.09	0.000***	1.14
Log-dist. from assigned univ to h-sch.	2.53	2.41	0.146	2.45

Figures reported are sample means except in the 'P-value' column which presents the p-value in a test of the significance of the difference between boys and girls' means. Stars give the outcome of this significance test: * significant at the 10% level; ** significant at the 5% level; *** significant at the 1% level.

Table 3.1 provides mean levels of outcomes of interest in the control group. In this table, statistics describe *marginal* applicants, that is, whose end-of-middle-school exam score is *close* to the threshold for admission to an elite high school[7]. We consider three sets of outcomes. Outcomes measured at the time of taking the end-of-high-school exam (probability to take the exam after 4 years of entering high school, score at the exam, and the priority score induced by this exam score); outcomes related to college application (selectivity level and distance from home of the programs applied to); and outcomes related to college admission (selectivity level and distance from

[7] Admission thresholds are characterized in more detail in Section 3.3. "Close" is understood is as *within the RD estimation bandwidth.*For each outcome and (sub)sample, the 'optimal' bandwidth is chosen using the Imbens and Kalyanaraman (2011) method and may vary from one outcome to another. Keeping in mind that this method selects the bandwidth solving a bias/efficiency tradeoff, to assess the sensitivity of our results to the bandwidth choice, we present descriptive statistics based both on a bandwidth equal to the optimal bandwidth and on a bandwidth equal to half of it.

home of the program admitted to).

Exam-taking and selection. It is important to notice than exam score, as well as college application and admission outcomes, realize conditional on students taking the end-of-high-school exam. If admission to an elite high school affects high school graduation rates in the first place, it may complicates the analysis of treatment effect of admission to an elite high school on subsequent outcomes. Table 3.2 shows the results of a test of this hypothesis. It shows that female students marginally admitted to selective high school are about 3 percentage point more likely to take the end-of-high-school exam (after four years of entering high school) than their marginally non-admitted counterparts. The consequences of this fact will be discussed in Section 3.3.

Table 3.2: Local mean differences in end-of-high-school exam-taking rate across treatment groups

	Opt.Bdw.	Half Opt.Bdw.
Full sample	0.021***	0.024**
	(0.008)	(0.011)
Obs.	27899	15390
Boys	0.011	0.013
	(0.013)	(0.017)
Obs.	10443	5497
Girls	0.027***	0.029**
	(0.01)	(0.013)
Obs.	17456	9893

Differences in means presented here are estimated using standard RD techniques (local linear regression). Robust standard errors in parentheses. Standard errors adjust for clustering at the school district time year level. The 'Half Opt.Bdw.' column shows estimates derived using a bandwidth equal to (half) the IK Imbens and Kalyanaraman (2011) optimal bandwidth. *significant at the 10% level; **significant at the 5% level;***significant at the 1% level.

3.3 Identification

To study the effects of (admission to) selective schools, many studies rely on a quasi-experimental setting often resulting from the assignment process used to determine students' eligibility. The present analysis is no exception as it exploits admission cutoffs within an RD design. However, we highlight that the nature of the outcomes of interest causes the RD analysis to produce biased estimates; and account for this fact to produce bounds for the true effects. In this section, we first describe the TE parameters of interest, their interpretation and detail the derivation of the cutoffs involved in the RDD analysis. We then review standard identification arguments within a sharp RD design. (Standard graphical evidence supporting the sharpness and validity of the RD design can be found in Appendix B.3.) Finally, we discuss the failure of the standard strategy in case of sample selection and/or measurement error. An attenuation bias may arise due to measurement. Sample selection affects outcomes (e.g. exam scores) existing as a byproduct of other outcomes (e.g. exam-taking) altered by treatment.

3.3.1 Treatment & Relevant admission cutoffs

Students are observed twice in the data. First, at the time of taking the end-of-middle-school exam; next, at the time of taking the end-of-high-school exam[8]. In each occurrence, we observe which school they are enrolled in; but we have no direct information about their path during the years in-between. In particular, if a student was admitted to a selective school but appears to take the end-of-high-school exam in another school, we do not know whether she never enrolled in the selective school

[8] For most students, these the two observations are four years apart; thus the general statements made later for simplicity about the "four years of high school"; however we do consider in our sample repeaters whose observations are 5 or 6 years apart. Also, as noted in the data section, we observe students each time they enroll to take an exam so we may observe students more than twice if they take the end-of-high-school exam several times (4% of the students in our sample take it twice, 0.2% of the students take it three times).

or did so for some time but changed schools at some point within the four years of high school. In the data, about 37% of the students admitted to a selective school in a given year and taking the end-of-high-school exam four years later end up not taking the exam in a selective school. Conversely, about .4% of the students not admitted to a selective school in a given year and taking the end-of-high-school exam four years later appear to be attending one as they take the exam. We interpret the second kind of students as those being offered a spot in a second round, after initially admitted students declined the offer. Given the large difference between these two statistics, we infer that a good number of students that are offered admission accept the offer and indeed attend the school for some time but change high schools before graduation. Lack of information about who these students are and how long each of them attends a selective school prevents us from identifying the effect of *attending* a selective high school. We therefore focus on estimating the effect of *being admitted* to one of the twelve Tunisian *lycées pilotes* on various outcomes; our estimates can then be interpreted as an intent to treat (ITT).

In this paper, we consider the system of selective high schools as a whole, and focus on the effect of being admitted to any selective high school rather than distinguishing between them. To determine the admission cutoffs relevant to each student, as needed for the RDD, we proceed as Estrada and Gignoux (2014). Let s_k, $k = 1, \ldots, 12$ denote the twelve *lycées pilotes*. The *school cutoff* $c_{k,t}$, associated to school k in year t is the score of the last student offered admission (ignoring a possible wait-list) to school k in year t. Consider a student i who, given the location of her middle school, cannot choose and may only apply to one selective school, say k_i. If her score at the end-of-middle-school exam falls above c_{k_i} she will receive admission, otherwise she will not. $c_i^* := c_{k_i}$ is then the cutoff relevant for student i as it fully determines her admission status. Alternatively, consider a student j who, given the location of her middle school, can choose between several selective schools, and rank them in order of preference. Let $s_{k_1}, \dots, s_{k_n}$ be the schools ranked by j. The relevant cutoff for student j then is $\min\{c_{k_1}, \dots c_{k_n}\} := c_j^*$. Indeed, if j's score is strictly lower than this c_j^*, then it is lower than the admission cutoffs of all the schools ranked by j and therefore j is not admitted to any selective school. If j's score is above this c_j^*, then j is admitted to a selective school –either the school with cutoff c_j^* or a school preferred to that one if j's score is also above the cutoff of a school she ranked before. c_j^* therefore determines j's admission status.

3.3.2 Identification in a standard RD setting

Potential outcome framework and the fundamental problem of causal inference

Let D denote the treatment indicator. Let Y_i denotes any outcome of interest, for instance, whether or not student i takes the end-of-high-school exam within four years of the treatment assignment. Adopting the potential outcome framework, suppose the existence of the latent random variables Y_i^0 and Y_i^1 indicating the value of outcome Y for student i under, respectively, absence of treatment and treatment. The individual effect of admission to a selective school (i.e. treatment) on outcome Y, for student i, is defined as $\widetilde{\Delta}_{Y,i} := Y_i^1 - Y_i^0$. For any student i, only one[9] version of Y^d is observed by the analyst: Y^1 is observed for treated individuals, Y^0 is observed for untreated individuals. This is the source of the fundamental problem of causal inference (Holland (1986)): $\widetilde{\Delta}_{Y,i}$ is not identified. Individual treatment effect being fundamentally out of reach, the analyst's goal may be two-fold: (i) determine other treatment effect parameters of interest, and derive conditions under which they are (at least *partially*) identified; and conversely (ii) scrutinize identified

[9] Actually, *at most* one. See our discussion of missing outcome data issues later in this section

parameters, derive conditions under which they can be given a causal interpretation, precisely interpret the treatments they identify and assess the extent to which these are of interest. We focus on point (i). Average treatment effects on various performance outcomes have drawn the attention of most of the literature on elite schools. Recent analyses (e.g. Abdulkadiroğlu et al. (2014), Behaghel et al. (2015)) have highlighted the possibility that significant changes in the school environment, such as switching schools or attending a boarding school, may have negative effects on some students. To allow heterogeneity in treatment effects along the distribution of student performance, we also consider quantile treatment effects.

Standard identification of LATE and LQTE in an RD setting

In this paragraph, we recall the definition of the TE parameters of interest, and review conditions for their identification in an RD setting.

In a sharp RD design, treatment is assigned deterministically based on a *running variable*. Let T denote the running variable: here, T_i is student i's end-of-middle-school exam score. The treatment indicator D_i is a *deterministic* function of T_i: $D_i := \mathbf{1}_{[T_i \geqslant c_i]}$, where c_i is the value of the admission cutoff relevant for student i. Let Y be an outcome of interest (e.g. students' exam-taking rate at the end of high school). Table 3.3 gives the definition of the TE parameters of interest using this notation. In what follows, F_Z denotes the cdf of random variable Z.

Table 3.3: Treatment effect parameters

LATE	$\lim_{t\to 0}\left\{[Y \mid D=1, \lvert T_i - c\rvert \leqslant t] - [Y \mid D=0, \lvert T_i - c\rvert \leqslant t]\right\}$
LQTE	$\lim_{t\to 0}\left\{Q_{Y\mid D=1,\lvert T-c\rvert\leqslant t}(q) - Q_{Y\mid D=0,\lvert T-c\rvert\leqslant t}(q)\right\}$ with $Q_{Z\mid D,\lvert T-c\rvert\leqslant t}(q) := \inf\left\{u \mid F_{Z\mid D,\lvert T-c\rvert\leqslant t}(u) \geqslant q\right\}$ $F_{Y\mid D=d,\lvert T-c\rvert\leqslant t}(u) = \left[\mathbb{1}_{[Y\leqslant u]} \mid D=d, \lvert T-c\rvert \leqslant t\right]$

The intuition underlying the RD design is that if this assignment method is as good as random, students whose score fell just below the admission threshold are similar to, and therefore a good control group for, students whose score barely made it above the cutoff value. More formally, consider the following conditions:

Condition C1. Conditional independence of treatment and potential outcomes: $Y_i^0, Y_i^1 \perp D_i \mid T_i$.

Condition C2. Continuity:

C2a. Continuity in the forcing variable of conditional expectations of potential outcomes (at the cutoff value): $t \mapsto \mathbb{E}[Y^d \mid T = t]$ is continuous at $t = c$ for $d \in \{0, 1\}$.

C2b. Continuity in the forcing variable of conditional distributions of potential outcomes (at the cutoff value): $t \mapsto F_{Y^d|T}(y \mid t)$ is continuous in t at $t = c$ for $d \in \{0, 1\}$.

Condition C3. Density at the threshold: $F_T(t)$ is differentiable at c and $\lim_{t \to c} f_T(t) > 0$.

Conditions C1 and C2 formalize the idea that students on each side of the cutoff should be, on average, similar in terms of observables and unobservables. C3 is a technical condition ensuring the mathematical well-definition of the terms in the expression for LQTE given in Table 3.3.

If the RD design is valid (that is, if Condition C1 holds), comparison of the mean outcome across these two groups identifies the average treatment effect on Y under Condition C2a (see, for instance, Hahn et al. (2001), Imbens and Lemieux (2007) or Lee and Lemieux (2010)). Comparison of the q^{th} quantile of the outcome distribution across the two groups identifies the quantile-q treatment effect on Y under Conditions C2b and C3 (see Frandsen et al. (2012)).

3.3.3 Beyond the RD setting: identification challenges

In this subsection, we expose the identification issues arising due to sample selection and data linkage failures. We also suggest alternative identification strategies addressing each of these concerns. In our setting, failures in the linkage of the original databases –end-of-middle-school exam data and end-of-middle-school exam data– result in non-classical measurement error in the outcome variable, and prevent standard RD estimates from being given a causal interpretation. In addition, when considering TEs on end-of-high-school exam score, we face the same sample selection issue as Lee (2009) in his influential study of the wage effects of the Job Corps program. Test scores are observed only for students who did take the exam. If exam-taking is affected by treatment, then comparing mean score at the end-of-high-school exam across treatment groups may fail to capture a causal treatment effect. Sample selection is not the only threat to the identification strategy outlined in the previous subsection.

Measurement error

Let B denote the indicator variable for whether or not the student took the end-of-high-school-exam[10] (B for *Baccalauréat*). Let R denote the end-of-high-school exam record indicator: $R_i = 1$ if student i is recorded as taking the end-of-high-school exam (four years after treatment assignment), and $R_i = 0$ otherwise. $R_i = 1$ only if $B_i = 1$; however $R_i = 0$, meaning that the student was not found in the end-of-high-school exam data 4 years after treatment assignment, may correspond to two distinct situations. *Either* student i indeed did not take the end-of-high-school

[10] within four years of the treatment assignment

exam (i.e. $B_i = 0$); *or* student i did take the exam but her information failed to be matched between the two databases (i.e. $B_i = 1$). R is perfectly observed; it is a noisy signal of B. Let M denote the matching (linkage) indicator: $M_i = 1$ if student i's information can be linked between the end-of-high-school exam and end-of-middle-school exam databases, and $M_i = 0$ if it cannot. $M_i = 0$ can be due, for instance, to typos occurring during data inputting of the student's identifiers. M is not perfectly observed. $M_i = 1$ if $R_i = 1$: student i was indeed found in both databases. If $R_i = 0$ though, M_i is unknown: one does not know whether the student did not take the end-of-high-school exam at all ($B_i = 0$), or did take it but failed to be matched ($B_i = 1, M_i = 0$). This missing outcome data issue can be framed in terms of *non-classical* measurement error. It is clear that the noise resulting from them is correlated with the true value of the end-of-high-school-exam-taking outcome: only students who did take the exam ($B = 1$) can be affected by the error. Students who did take the exam may fail to be matched, in which case their exam-taking record ($R_i = 0$) differ from their true exam-taking value ($B_i = 1$). Students who did not take the exam, however, cannot fail to be unmatched (i.e. be matched by mistake), so their exam-taking record ($R_i = 0$) and true exam-taking value ($B_i = 0$) always coincide. Formally, one can show that that TE estimates derived from R as a proxy for B suffer from an *attenuation bias*: $\widehat{\tau} = \tau \cdot p_M < \tau$ where $\widehat{\tau}$ and τ denote, respectively, TE on R and B, and $p_M \in (0,1)$ is the probability of an exam-taker to be matched across the two databases.

If p_M were known, bias correction would be straightforward, as $\widetilde{\tau} \cdot p_M^{-1}$ would be a consistent estimate of τ. In our case, p_M is unknown but the interval:

$$\Big[\min\Big\{ \lim_{t\to 0} \Pr(R = 1 \mid B = 1, D = 1, |T_i - c| \leqslant t), \\ \lim_{t\to 0} \Pr(R = 1 \mid B = 1, D = 0, |T_i - c| \leqslant t) \Big\}; \quad 1 \Big]$$

can be shown to be the sharp identified set for p_M. The upper bound amounts to assuming perfect matching, which the data cannot reject. The lower bound corresponds to the lowest value of p_M compatible with the matching rate being the same for all students (conditional on exam-taking), and in particular, across treated and control groups. Our lower bound for ATE on outcomes with missing values is then based on assuming perfect matching; while our upper bound is based on assuming the lowest data-compatible value for P_M[11].

Note that data linkage failures as they happen in this context do not affect identification of TE on exam performance and other outcomes realizing after or conditional on exam-taking. It is indeed reasonable to assume that whether the student is successfully linked or not is not correlated with how he performs at the end-of-high-school exam: conditional on taking the exam, failure to be matched across databases occurs at random. In other words, the only difference between the sample of all exam-takers and the sample of matched exam-takers is their size –they are drawn from the same distribution. As a consequence, linkage failures only affect the precision of estimated TE outcomes realizing after or conditional on exam-taking.[12]

Sample selection

Consider a test-taker's performance at the end-of-high-school exam, which we denote Y in this paragraph. For practicality, we follow Lee (2009) or Angrist et al. (2006) in their analysis of the long-run effects of Colombia's PACES program on high-schoolers' test scores, and assume that potential outcomes Y^1, Y^0 exist for *all* students, regardless of their actual exam-taking and treatment statuses[13] . Let B

[11] Aside from the RD setting, these bounds are similar to those introduced in Horowitz and Manski (1995).

[12] This means that, in our context, no TE of interest is cumulatively affected by both sample selection and linkage failures.

[13] Alternatively, one could follow Zhang and Rubin (2003) and assume that for those who do not take the end-of-high-school exam four years after entering high school, exam score is not defined.

be the indicator of whether or not a student took the end-of-high-school exam (B for *Baccalauréat*). Comparison of mean test scores across treatment groups, as prescribed by the definition of ATE in Table 3.3, can only be made using students for which test scores are observed. Making exam-taking explicit in the conditioning event:

$$
\begin{aligned}
&\lim_{t\to 0}\Big\{ [Y_i \mid B_i = 1, D_i = 1, |T_i - c| \leqslant t] - [Y_i \mid B_i = 1, D = 0, |T_i - c| \leqslant t] \Big\} \\
&= \lim_{t\to 0}\Big\{ \big[Y^1 \mid B_i^1 = 1, D_i = 1, |T_i - c| \leqslant t\big] - \big[Y_i^0 \mid B_i^0 = 1, D_i = 0, |T_i - c| \leqslant t\big] \Big\} \\
&= \lim_{t\to 0}\Big\{ \big[Y^1 \mid B_i^1 = 1, |T_i - c| \leqslant t\big] - \big[Y^0 \mid B_i^0 = 1, |T_i - c| \leqslant t\big] \Big\} \\
&= \lim_{t\to 0}\Big\{ \big[Y^1 - Y^0 \mid B_i^1 = 1, B^0 = 1, |T_i - c| \leqslant t\big] \times \Pr[B_i^1 = 1 \mid B_i^0 = 1] \\
&\qquad + \big[Y_i^1 \mid B_i^1 = 1, B_i^0 = 1, |T_i - c| \leqslant t\big] \times \Big(\Pr[B_i^0 = 1 \mid B_i^1 = 1] - \Pr[B_i^1 = 1 \mid B_i^0 = 1] \Big) \\
&\qquad + \big[Y_i^1 \mid B_i^1 = 1, B_i^0 = 0, |T_i - c| \leqslant t\big] \times \Pr[B_i^0 = 0 \mid B_i^1 = 1] \\
&\qquad - \big[Y_i^0 \mid B_i^0 = 1, B_i^1 = 0, |T_i - c| \leqslant t\big] \times \Pr[B_i^1 = 0 \mid B_i^0 = 1] \Big\},
\end{aligned}
$$

where the second equality follows from condition C1. This derivation suggests the absence of (interesting) causal interpretation to be given to the *feasible* difference in means. Indeed, the above equals the LATE on the $(B^1 = 1, B^0 = 1)$-subpopulation[14] if and only if treatment does not affect exam-taking. Otherwise, despite randomization of treatment assignment, the groups of treated exam-takers and untreated exam-takers, whose average scores are *de facto* being compared, may not be valid counterfactuals for each other. Since only one of these potential outcomes is observed, the identity of the students taking the exam regardless of their treatment status (call them 'always *exam*-takers') is unknown to the analyst, who therefore cannot directly restrict her analysis to this subpopulation.

This assumption means that the potential outcome framework is fundamentally non-appropriate, and requires the definition of an alternative, extended structure. The TE bounds one can derive following either approaches are the same.

[14] That is, $\lim_{t\to 0}\Big\{ \big[Y^1 - Y^0 \mid B_i^1 = 1, B^0 = 1, |T_i - c| \leqslant t\big] \Big\}$.

Similar conclusions can be drawn for LQTEs.

To address sample selection, we derive bounds for TE parameters on the population of always-exam-takers. In spirit, this strategy is similar to that used by Zhang and Rubin (2003), Lee (2009) or Angrist et al. (2006) in random trial settings. However, our particular procedure differ from those as it needs to account for the quasi-experimental setting we are facing. In contemporaneous work, Dong (2016) also applies this strategy to provide a partial identification framework to address sample selection in a fuzzy RD design. In addition, as we explain in more detail later, we depart from Lee (2009) and Angrist et al. (2006) in the assumptions we make to refine our bounds. If the (limit) shares of always exam-takers within the control and treatment groups ($s_0 := \lim_{t \to 0} \Pr[B^1 = 1 \mid B^0 = 1, |T_i - c| \leqslant t]$ and $s_1 := \lim_{t \to 0} \Pr[B^0 = 1 \mid B^1 = 1, |T_i - c| \leqslant t]$, respectively) were known, agnostic sharp bounds on the treatment effect on the end-of-high-school exam score would be straightforwardly derived. To derive the upper bound, consider the scenario in which always exam-takers are (i) at the *top* of the distribution of end-of-high-school exam scores for the treated group; and (ii) at the *bottom* of the distribution of end-of-high-school exam scores for the control group. Under this scenario, one can isolate the subgroups of always *exam*-takers within the control and treatment groups. Comparison (of the limit as the running variable goes to the cutoff value) of the mean end-of-high-school exam score within each subgroup gives the upper bound for the TE of interest. To derive the lower bound, proceed symmetrically, considering the scenario in which always exam-takers are (i) at the *bottom* of the distribution of end-of-high-school exam scores for the treated group; and (ii) at the *top* of the distribution of end-of-high-school exam scores for the control group. Table 3.4 gives the formal expression of these bounds[15], as well as their extension to LQTEs.

[15] Derivations are shown in Appendix B.4.

Table 3.4: Agnostic bounds with known (s_1, s_0)

LATE	**Lower:** $\Delta_L(s) := \lim_{t\to 0}\Big\{ \big[Y \mid D=1, \lvert R_i - c\rvert \leqslant t, Y_i \leqslant y_1^-(s)\big] - \big[Y \mid D=0, \lvert R_i - c\rvert \leqslant t, Y_i \geqslant y_0^+(s)\big] \Big\}$ **Upper:** $\Delta_U(s) := \lim_{t\to 0}\Big\{ \big[Y \mid D=1, \lvert R_i - c\rvert \leqslant t, Y_i \geqslant y_1^+(s)\big] - \big[Y \mid D=0, \lvert R_i - c\rvert \leqslant t, Y_i \leqslant y_0^-(s)\big] \Big\}$
LQTE	**Lower:** $\delta_L(q;s) = \lim_{t\to 0}\Big\{ Q^-_{Y\mid D=1,\lvert R-c\rvert\leqslant t}(q) - Q^+_{Y\mid D=0,\lvert R-c\rvert\leqslant t}(q) \Big\}$ **Upper:** $\delta_U(q;s) = \lim_{t\to 0}\Big\{ Q^+_{Y\mid D=1,\lvert R-c\rvert\leqslant t}(q) - Q^-_{Y\mid D=0,\lvert R-c\rvert\leqslant t}(q) \Big\}$ with $Q^\pm_{Z\mid D,\lvert R-c\rvert\leqslant t}(q) := \inf\Big\{u \mid F^\pm_{Z\mid D,\lvert R-c\rvert\leqslant t}(u) \geqslant q\Big\}$ $F^+_{Y\mid D=d,\lvert R-c\rvert\leqslant t}(u) = \big[\mathbb{1}_{[Y\leqslant u]} \mid D=d, \lvert R-c\rvert \leqslant t, Y_i \geqslant y_d^+(s)\big]$ $F^-_{Y\mid D=d,\lvert R-c\rvert\leqslant t}(u) = \big[\mathbb{1}_{[Y\leqslant u]} \mid D=d, \lvert R-c\rvert \leqslant t, Y_i \leqslant y_d^-(s)\big]$

We denote $y_d^-(s) := y_{D=d,|R_i-c|\leqslant t}\left(s_{(d)}\right)$ and $y_d^+(s) := y_{D=d,|R_i-c|\leqslant t}\left(1-s_{(d)}\right)$.

Because the identity of always exam-takers is unknown, shares of always-exam-takers within the control and treatment groups are not known from the data. They can however be bounded. Using basic algebra, one can rewrite:

$$s_0 := \lim_{t\to 0}\left\{1 - \frac{\Pr[B_0=1, B_1=0 \mid |T_i-c|\leqslant t]}{\Pr[B=1 \mid D=0, |T_i-c|\leqslant t]}\right\}$$

$$s_1 := \lim_{t\to 0}\Bigg\{\Big(\Pr[B=1 \mid D=0, |T_i-c|\leqslant t] - \Pr[B_0=1, B_1=0 \mid |T_i-c|\leqslant t]\Big)\times$$

$$\Big(\Pr[B=1 \mid D=0, |T_i-c|\leqslant t] + \Pr[B=0 \mid D=0, |T_i-c|\leqslant t] -$$

$$\Pr[B=0 \mid D=1, |T_i-c|\leqslant t]\Big)^{-1}\Bigg\}$$

All probability terms in the right-hand side are identified, except for the term $\pi_{01} := \lim_{t\to 0}\Pr[B_0=1, B_1=0 \mid |T_i-c|\leqslant t]$ as the full pair (B_0, B_1) is never observed. It is only partially identified. Indeed, the interval

$$\Big[\lim_{t\to 0}\max\big\{0, \Pr[B=1 \mid D=0, |T_i-c|\leqslant t] - \Pr[B=1 \mid D=1, |T_i-c|\leqslant t]\big\},$$
$$\lim_{t\to 0}\min\big\{\Pr[B=1 \mid D=0, |T_i-c|\leqslant t], 1-\Pr[B=1 \mid D=1, |T_i-c|\leqslant t]\big\}\Big]$$

can be shown to be the identified set for the (limit) share of students whose exam-taking status is negatively affected by treatment, π_{01}. In this paper, we derive TE bound estimates under two distinct scenarios. Our *most conservative* scenario allows for the larger share of negatively affected students; π_{01} is set to its upper bound, π_{01}^{+}. Our *least conservative* scenario imposes the maximal restriction on the share of negatively affected students, π_{01} is set to its lower bound, π_{01}^{-}. The monotonicity assumption underlying identified bounds in Lee (2009) and Angrist et al. (2006) corresponds to setting π_{01} to 0. It cannot be rejected by the data and has become a standard assumption to be made. However. there are reasons to feel hesitant about it. In our setting, a number of admitted students are observed to be registered in a non-elite school at the time of taking the end-of-high-school-exam. As discussed in Section 3.3.1, this suggests that some treated students declined treatment from the beginning, possibly expecting it not to be beneficial for them; while others did enroll in the elite system for a while before switching for a non-selective school, possibly realizing the treatment was harming them. More generally in the selective schools context, studies (e.g. see Abdulkadiroğlu et al. (2014), Behaghel et al. (2015)) have found that the change of environment associated with attending a selective school may be disruptive to some students and have adverse effects.

Following Zhang and Rubin (2003), we refine the identified set for TE parameters on outcomes Y whose values can be naturally ordered in terms of quality using a stochastic dominance (SD) assumption, which states that, in each treatment group, the distribution of exam scores of always-exam-takers first-order stochastically dominates the distribution of exam scores of non-always exam-takers. Formally,

Stochastic dominance. For $d \in \{0,1\}$ and $d' \neq d$, and for any $y \in Supp(Y)$,

$$F_{Y|D=d,B^d=1,B^{d'}=1}(y) \leqslant F_{Y|D=d,B^d=1,B^{d'}=0}(y).$$

The intuition underlying this assumption is that those students able to take the exam regardless of their treatment status may have higher ability or motivation than those who would fail to take the exam under one treatment arm only; and that, on average, this higher ability may drive their exam scores up. Note that the SD assumption implies that, on average and in each treatment group, always-exam-takers perform better at the exam than non-always exam-takers. Formally,

Ranked conditional average outcomes. For $d \in \{0,1\}$ and $d' \neq d$,

$$\left[Y_i \mid D_i = d, B_i^d = 1, B_i^{d'} = 1\right] \geqslant \left[Y_i \mid D_i = d, B_i^d = 1, B_i^{d'} = 0\right].$$

Table 3.5 presents refined bounds under the most and least conservative scenarios regarding π_{01}.

3.4 Estimation

Our estimate procedures are a natural hybrids of estimation methods for TE bounds in random trial settings (e.g. Horowitz and Manski (1995) for measurement error; Lee (2009), Zhang and Rubin (2003) for sample selection) on the one hand, and estimation methods for point-identified TE parameters in RD designs (e.g. Hahn et al. (2001), Imbens and Lemieux (2007) for ATE; and Frandsen et al. (2012) for QTE) on the the other hand.

Table 3.5: Bounds under the SD assumption

Bounds under most conservative scenario ($\pi_{01} = \pi_{01}^{+}$)	
LATE	**Lower:** $\Delta_L(s(\pi_{01}^{+})) := \lim_{t\to 0}\Big\{ \big[Y \mid D=1, \lvert R_i - c\rvert \leqslant t\big] - \big[Y \mid D=0, \lvert R_i - c\rvert \leqslant t, Y_i \geqslant y_0^{+}(s_0(\pi_{01}^{+}))\big] \Big\}$ **Upper:** $\Delta_U(s(\pi_{01}^{+})) := \lim_{t\to 0}\Big\{ \big[Y \mid D=1, \lvert R_i - c\rvert \leqslant t, Y_i \geqslant y_1^{+}(s_1(\pi_{01}^{+}))\big] - \big[Y \mid D=0, \lvert R_i - c\rvert \leqslant t\big] \Big\}$
LQTE	**Lower:** $\delta_L(q; s(\pi_{01}^{+})) = \lim_{t\to 0}\Big\{ Q_{Y\mid D=1,\lvert R-c\rvert\leqslant t}(q) - Q^{+}_{Y\mid D=0,\lvert R-c\rvert\leqslant t}(q) \Big\}$ **Upper:** $\delta_U(q; s(\pi_{01}^{+})) = \lim_{t\to 0}\Big\{ Q^{+}_{Y\mid D=1,\lvert R-c\rvert\leqslant t}(q) - Q_{Y\mid D=0,\lvert R-c\rvert\leqslant t}(q) \Big\}$
Bounds under least conservative scenario ($\pi_{01} = \pi_{01}^{-}$)	
LATE	**Lower:** $\Delta_L(s(\pi_{01}^{-})) := \lim_{t\to 0}\Big\{ \big[Y \mid D=1, \lvert R_i - c\rvert \leqslant t\big] - \big[Y \mid D=0, \lvert R_i - c\rvert \leqslant t\big] \Big\}$ **Upper:** $\Delta_U(s(\pi_{01}^{-})) := \lim_{t\to 0}\Big\{ \big[Y \mid D=1, \lvert R_i - c\rvert \leqslant t, Y_i \geqslant y_1^{+}(s_1(\pi_{01}^{-}))\big] - \big[Y \mid D=0, \lvert R_i - c\rvert \leqslant t\big] \Big\}$
LQTE	**Lower:** $\delta_L(q; s(\pi_{01}^{-})) = \lim_{t\to 0}\Big\{ Q_{Y\mid D=1,\lvert R-c\rvert\leqslant t}(q) - Q_{Y\mid D=0,\lvert R-c\rvert\leqslant t}(q) \Big\}$ **Upper:** $\delta_U(q; s(\pi_{01}^{-})) = \lim_{t\to 0}\Big\{ Q^{+}_{Y\mid D=1,\lvert R-c\rvert\leqslant t}(q) - Q_{Y\mid D=0,\lvert R-c\rvert\leqslant t}(q) \Big\}$

$Q^{\pm}_{Z\mid D,\lvert R-c\rvert\leqslant t}$ and $F^{\pm}_{Y\mid D=d,\lvert R-c\rvert\leqslant t}$ are defined in Table 3.4. The least conservative scenario (monotonicity if the lower bound for π_{01} is 0, as it will be in our case) implies all non-treated exam-takers are always-exam-takers. Hence, no trimming is made on the outcome distribution for the untreated group ($s_0 = 1$). In addition, s_1 being a decreasing function of π_{01}, the portion of the outcome distribution that is trimmed for the treated group is minimal ($s_1(\pi_{01}^{-}) > s_1(\pi_{01}^{+})$).

3.4.1 Bounds accounting for measurement error

Start by choosing an estimation bandwidth h_n.

Step 1: Estimate bounds for the matching probability.

1. For each $b \in \{0,1\}$, estimate $p_{Md} := \lim_{t\to 0}\Pr(R = 1 \mid B = 1, D = d, \lvert T_i - c\rvert \leqslant t)$ by kernel regression. We use a rectangular kernel:

$$\hat{p}_{Md}(h_n) = \frac{\sum_{i=1}^{n} \mathbb{1}_{[c-h_n\leqslant T_i\leqslant c+h_n]} \times \mathbb{1}_{[D_i=d]} \times \mathbb{1}_{[R_i=1]}}{\sum_{i=1}^{n} \mathbb{1}_{[c-h_n\leqslant T_i\leqslant c+h_n]} \times \mathbb{1}_{[D_i=d]}}.$$

2. The matching probability is assumed to be the constant in the population: $P_M := \Pr(R = 1 \mid B = 1) = \lim_{t \to 0} \Pr(R = 1 \mid B = 1, |T_i - c| \leqslant t)$.

$$\text{Lower bound:} \quad \hat{p}_M^{-}(h_n) := \min\left\{\hat{p}_{M0}(h_n), \hat{p}_{M1}(h_n)\right\};$$

$$\text{Upper bound:} \quad \hat{p}_M^{+}(h_n) := 1.$$

Our lower (resp. upper) bound ATE estimate will be derived assuming the largest (resp. smallest) value of P_M compatible with the data, that is, $\hat{P}_{M(+)}(h_n)$ (resp. $\hat{P}_{M(-)}(h_n)$).

Step 2: Derive naive RDD estimates. Our naive estimate is given Δ solving:

$$\min_{\alpha,\beta,\Delta,\gamma} \sum_{i=1}^{n} \mathbb{1}_{[c-h_n \leqslant T_i \leqslant c+h_n]} \cdot \Big(Y_i - \big[\alpha + \beta(T_i - c) + \Delta \mathbb{1}_{[T_i > c]} + \gamma(T_i - c)\mathbb{1}_{[T_i > c]}\big]\Big).$$

Step 3[ATE]: Obtain bounds. Finally, our derivation of estimates for the bounds for ATE mimics bias correction:

$$\hat{\Delta}_L(h_n) = \begin{cases} \hat{\Delta}(h_n) & \text{if } \hat{\Delta}(h_n) > 0 \\ \hat{\Delta}(h_n) \times \left(\hat{p}_M^{-}(h_n)\right)^{-1} & \text{otherwise} \end{cases}$$

$$\hat{\Delta}_U(h_n) = \begin{cases} \hat{\Delta}(h_n) \times \left(\hat{p}_M^{-}(h_n)\right)^{-1} & \text{if } \hat{\Delta}(h_n) > 0 \\ \hat{\Delta}(h_n) & \text{otherwise} \end{cases}$$

By construction, $\hat{\Delta}_U(h_n) \geqslant \hat{\Delta}_L(h_n)$ for all $n \in \mathbb{N}\backslash\{0\}$ and $h_n > 0$.

3.4.2 Bounds accounting for sample selection

Start by choosing an estimation bandwidth h_n[16].

Step 1: Estimate trimming shares.

[16] For better comparability with naive RD estimates presented in Section **??**, we choose the Imbens and Kalyanaraman (2011)-optimal bandwidth.

1. For each pair $(b, d) \in \{0, 1\}$, estimate $p_{bd} := \lim_{t \to 0} \Pr(B = b \mid D = d, |T_i - c| \leqslant t)$ by kernel regression. We use a rectangular kernel:

$$\hat{p}_{bd} = \frac{\sum_{i=1}^{n} \mathbb{1}_{[c-h_n \leqslant T_i \leqslant c+h_n]} \times \mathbb{1}_{[D_i=d]} \times \mathbb{1}_{[B_i=b]}}{\sum_{i=1}^{n} \mathbb{1}_{[c-h_n \leqslant T_i \leqslant c+h_n]} \times \mathbb{1}_{[D_i=d]}}.$$

2. Derive bounds for data-compatible values of $\pi_{01} := \lim_{t \to 0} \Pr(B_0 = 1, B_1 = 0 \mid |T_i - c| \leqslant t)$:

$$\text{Lower bound:} \quad \hat{\pi}_{01}^{-}(h_n) := \max\{0, \hat{p}_{10}(h_n) - \hat{p}_{11}(h_n)\};$$

$$\text{Upper bound:} \quad \hat{\pi}_{01}^{+}(h_n) := \min\{\hat{p}_{10}(h_n), 1 - \hat{p}_{11}(h_n)\}.$$

3. Assume a value $\tilde{\pi}_{01} \in \left[\hat{\pi}_{01}^{-}(h_n), \hat{\pi}_{01}^{+}(h_n)\right]$ and derive estimates of the selection probabilities s_0 and s_1:

$$\hat{s}_1(h_n, \tilde{\pi}_{01}) = \frac{\hat{p}_{10}(h_n) - \tilde{\pi}_{01}}{\hat{p}_{10}(h_n) + \hat{p}_{00}(h_n) - \hat{p}_{01}(h_n)};$$

$$\hat{s}_0(h_n, \tilde{\pi}_{01}) = 1 - \frac{\tilde{\pi}_{01}}{\hat{p}_{10}(h_n)}.$$

Step 2: Trim outcome distributions. Given $(\hat{s}_0(h_n, \tilde{\pi}_{01}), \hat{s}_1(h_n, \tilde{\pi}_{01}))$, create trimming indicators $\tilde{\Gamma}^{-}$ and $\tilde{\Gamma}^{+}$:

- $\tilde{\Gamma}^{-}(h_n, \tilde{\pi}_{01}))$ takes the value 0 for observations trimmed away in the estimation of the lower bound of the ATE, that is, treated observations *above* the $\hat{s}_1^{th}$ quantile of $\{Y \mid D = 1\}$, as well as control observations *below* the $(1 - \hat{s}_0)^{th}$ quantile of $\{Y \mid D = 0\}$:

  $$\tilde{\Gamma}_{[h_n, \tilde{\pi}_{01}], i}^{-} = \left\{D_i = 1, Y_i \leqslant y_1\left[\hat{s}_1(h_n, \tilde{\pi}_{01})\right]\right\} + \left\{D_i = 0, Y_i \geqslant y_0\left[1 - \hat{s}_0(h_n, \tilde{\pi}_{01})\right]\right\}$$

  where $y_d[s]$ denotes the s^{th} quantile of $\{Y \mid d = 0\}$.

- $\tilde{\Gamma}_{[h_n, \tilde{\pi}_{01}]}^{+}$ takes the value 0 for observations trimmed away in the estimation of the upper bound of the ATE, that is, treated observations *below* the $(1 - \hat{s}_1)^{th}$

quantile of $\{Y \mid D=1\}$, as well as control observations *above* the $\hat{s}_0^{th}$ quantile of $\{Y \mid D=0\}$:

$\tilde{\Gamma}^{+}_{[h_n,\tilde{\pi}_{01}],i} = \{D_i = 1, Y_i \geqslant y_1[1-\hat{s}_1(h_n,\tilde{\pi}_{01})]\} + \{D_i = 0, Y_i \leqslant y_0[\hat{s}_0(h_n,\tilde{\pi}_{01})]\}$

Step 3[ATE]: Estimate ATE bounds. Our lower bound ATE estimate is given by Δ_L solving:

$$\min_{\alpha,\beta,\Delta,\gamma} \sum_{i=1}^{n} \mathbb{1}_{[c-h_n \leqslant T_i \leqslant c+h_n]} \cdot \tilde{\Gamma}^{-}_{[h_n,\tilde{\pi}_{01}],i} \cdot \Big(Y_i - \big[\alpha + \beta(T_i - c) + \Delta \mathbb{1}_{[T_i>c]} + \gamma(T_i - c)\mathbb{1}_{[T_i>c]}\big]\Big).$$

Our upper bound estimate is given by Δ_U solving the analogous minimization problem with trimming indicator $\tilde{\Gamma}^{+}$ instead of $\tilde{\Gamma}^{-}$. By construction, $\hat{\Delta}_U(h_n,\tilde{\pi}_{01}) \geqslant \hat{\Delta}_L(h_n,\tilde{\pi}_{01})$ for all $n \in \mathbb{N}\backslash\{0\}, h_n > 0$ and $\tilde{\pi}_{01} \in [0,1]$.

Step 3[QTE]: Estimate QTE bounds. Our lower bound QTE estimate at quantile q is derived as follows:

1. we estimate the local distributions of potential outcomes $F_{Y^1|T=c}$ and $F_{Y^0|T=c}$ given in Table 3.4 using local linear regression on the trimmed data:

$$\hat{F}_{Y^1|T=c}(y) := a_1(y)$$

$$\hat{F}_{Y^0|T=c}(y) := a_0(y)$$

where for any $y \in \mathbb{R}$, $a_1(y)$ and $a_0(y)$ are such that

$$\begin{aligned}
\big(a_1(y), b_1(y)\big) &:= arg\min_{a,b} \sum_{i=1}^{n} \mathbb{1}_{[c-h_n \leqslant T_i \leqslant c+h_n]} \cdot \tilde{\Gamma}^{-}_{[h_n,\tilde{\pi}_{01}],i} \\
&\qquad \cdot \Big(\mathbb{1}_{[Y_i \leqslant y]} \cdot D_i - a - b(T_i - c)\Big)^2 \\
\text{and} \quad \big(a_0(y), b_0(y)\big) &:= arg\min_{a,b} \sum_{i=1}^{n} \mathbb{1}_{[c-h_n \leqslant T_i \leqslant c+h_n]} \cdot \tilde{\Gamma}^{-}_{[h_n,\tilde{\pi}_{01}],i} \\
&\qquad \cdot \Big(\mathbb{1}_{[Y_i \leqslant y]} \cdot (1 - D_i) - a - b(T_i - c)\Big)^2
\end{aligned}$$

2 we derive the q^{th} quantile of the estimated local distributions: $\hat{Q}_{Y^1|T=c}(q) := \inf\{u \mid \hat{F}_{Y^1|T=c}(u) \geqslant q\}$ and $\hat{Q}_{Y^0|T=c}(q) := \inf\{u \mid \hat{F}_{Y^0|T=c}(u) \geqslant q\}$;

3 finally, we deduce the lower bound QTE estimate $\hat{\delta}_L(q; h_n, \tilde{\pi}_{01}) := \hat{Q}_{Y^1|T=c}(q) - \hat{Q}_{Y^0|T=c}(q)$.

Our upper bound $\hat{\delta}_U(q; h_n, \tilde{\pi}_{01})$ is derived in an analogous way using trimming indicator $\tilde{\Gamma}^+$ instead of $\tilde{\Gamma}^-$ in the first-stage minimization problems. By construction, $\hat{\delta}_U(q; h_n, \tilde{\pi}_{01}) \geqslant \hat{\delta}_L(q; h_n, \tilde{\pi}_{01})$ for all $q \in [0,1], n \in \mathbb{N}\backslash\{0\}, h_n > 0$ and $\tilde{\pi}_{01} \in [0,1]$.

3.4.3 A word on inference

To account for the multiple steps in estimation, our confidence intervals and standard errors are estimated by the bootstrap. It is important to understand that the bootstrap gives confidence intervals and standard errors for the bounds rather than the TE parameters (of interest).

As noted by Imbens and Manski (2004), if $[L_{(Y)-}, L_{(Y)+}]$ and $[U_{(Y)-}, U_{(Y)+}]$ are intervals of $\mathbb{R}$ covering with $\alpha\%$ (for instance 95%) probability the lower and upper bounds, respectively, of the effect of treatment on outcome Y, then the interval $[L_{(Y)-}, U_{(Y)+}]$ covers the TE parameter with probability larger than $\alpha\%$. Deriving confidence intervals covering the TE parameters with a fixed probability α is a challenge that we leave for future work[17]. Given the remark from Imbens and Manski (2004), significance levels given in Section 3.5 can be seen as conservative.

[17] In the case of Lee (2009), confidence intervals covering the TE parameters are with probability α can be straightforwardly derived applying results in Imbens and Manski (2004). Lee (2009) estimates bounds for TE from a randomized experiment, and use a fully parametric one-step estimation method. In this paper, the estimation procedure is local, requires the use of a bandwidth and proceeds in multiple steps. This makes the derivation confidence intervals covering the TE parameters are with some fixed probability significantly less straightforward.

3.5 Results

This section shows estimated bounds for TEs derived following the methodology described in Section 3.4.

3.5.1 Local average effects

As expected, bound estimates reveal the attenuation bias likely plaguing the LATE on exam-taking rate due to measurement error. Recall that lower bounds are derived under the assumption of perfect data merging ($p_M = 1$). This means that unmatched students are assumed not to be taking the end-of-high-school exam. The difference between upper and lower bounds shown in Table 3.6 therefore reveals the magnitude of the potential attenuation bias affecting estimates derived without accounting for measurement error. Qualitatively, the conclusions drawn with and without accounting for measurement error are similar. Admission to an elite high school is estimated to significantly decrease the retention rate of female students, and not affect the retention rate of male students. Quantitatively, female students are estimated to be 3.2 to 3.6 percentage points more likely to take the end-of-high-school exam in four years after entering high school (using half optimal bandwidth estimates) due to their admission to an elite high school –hence a potential attenuation bias of .4 percentage point. Estimates for male students are lower than a percentage point and therefore little affected by a change in the merging probability (within the data-compatible range).

Absence of a significant effect on the (binary) selection variable is often invoked as evidence that post-selection outcomes are not affected by sample selection. Note that such statement is valid if and only if the treatment does not negatively affect anyone. In light of recent evidence that selective and boarding schools may negatively affect students performance, (e.g. Abdulkadiroğlu et al. (2014), Behaghel et al. (2015)), turning to other outcomes, we do not *a priori* make the assumption that male outcomes are not affected by sample selection.

Table 3.6: Bound estimates for TEs on exam-taking rate

	Low. B.	Upp. B.	Obs.	$\hat{p}_{M(-)}$	$\hat{p}_{M(+)}$
Optimal Bandwidth					
Total sample	0.021***	0.024***	27899	0.89	1
Boys	0.011	0.012	10443	0.89	1
Girls	0.027***	0.03***	17456	0.89	1
Half Optimal Bandwidth					
Total sample	0.024**	0.027**	15390	0.89	1
Boys	0.013	0.015	5497	0.89	1
Girls	0.029**	0.033**	9893	0.89	1

The two rightmost columns give the range of matching probabilities p_M compatible with the data. For each subpopulation, the TE lower (resp. upper) bound is derived assuming the largest (resp. smallest) value of p_M compatible with the data. Bounds in the top panel are estimated using the Imbens and Kalyanaraman (2011) optimal bandwidth; bounds in the bottom panel are estimated using half of this bandwidth. Significance levels are based on bootstrap confidence intervals for bounds (5,000 repetitions). Bootstrap confidence bands for bound estimates are given in Appendix B.5, Table B.4. *significant at the 10% level; **significant at the 5% level;***significant at the 1% level.

Table 3.7 shows estimated bounds for local ATEs on always-exam-takers, in each of the estimation subsamples considered in Section 3.2.2. Appendix B.5, Table B.5 shows associated confidence bands.

As a reference, estimates derived without accounting for sample selection are shown in the rightmost column of Table 3.7 ('naive' RD estimates). Results labeled as *most conservative* are derived under the lowest shares of always-exam-takers compatible with the data in both the treatment and control groups. This corresponds to the scenario in which the share of students that can be negatively affected by the treatment

(π_{01}) is set to its largest value allowed by the data. This is the scenario that possibly differs the most from the assumptions underlying 'naive' RD results, which implicitly assume that all observed exam-takers are in fact always-exam-takers. Results labeled as *least conservative* are derived under the highest shares of always-exam-takers compatible with the data, that is, under the scenario restricting the share of students that can be negatively affected to its minimal value (compatible with the data). It corresponds to the scenario that is the closest to the assumptions underlying the 'naive' RD results.

Under the most conservative scenario, our estimates suggest[18] that admission to an elite high school has a positive effect on male always-exam-takers' exam performance, as well on the selectivity of the university programs they apply and get assigned to. Bounds for the effect on exam-performance range from .015 to about .60 standard deviation, against about .13 for the naive point estimates. Male always-exam-passers seem to apply to more selective post-secondary programs, in particular regarding their top-ranked choice (+.14 to +.26 std. dev., against about .19 for the naive point estimates.). Most importantly, male always-exam-passers seem to be assigned to more selective post-secondary programs (+.10 to +.24 std. dev., against about .12 for the naive point estimates.).

The least conservative scenario gives similar results, with tighter bounds and lower bounds now significantly different from 0. In each case, the large width of the identified under the most conservative scenario is not too surprising as this scenario assumes both a large share of negatively affected students and a large share of positively affected students. The identifying power of the monotonicity assumption is

[18] For each of these outcomes, both lower and upper bounds are positive, although only the upper bound reaches statistical significance at conventional levels.

illustrated by the shrinking of the estimated set under the least conservative scenario.

Interestingly, for girls, results reveal the potential for a different picture of average treatment effects than the one drawn from 'naive' RD estimates –although, again, precision is not high enough to conclude to the statistical significance of the lower bound in each case. Female always-exam-takers' exam and final score are increased by about .05 to .3 std. dev.. While the extent to which the top-ranked track in their application rankings is affected by treatment cannot be signed, female always-exam-passers are potentially assigned to more selective programs (.036 to .08 std. dev.). 'Naive' RD estimates show no significant difference between observed marginally treated and untreated girls' test scores and post-secondary application choices. Bound estimates account for the change in the composition of control and treatment groups following the treatment-induced increase in girls' probability to take the end-of-high-school exam. They show that treatment does have a significant and positive effect on exam and final scores for girls who *would take the exam regardless of their treatment status* (always-exam-takers). Naive estimates being downward-biased (under the least conservative scenario) is consistent with the intuition that students taking the exam only under treatment might be academically weaker than always-exam-takers, and would therefore perform worse than them when taking the end-of-high-school exam. In other words, female always-exam-takers' exam and final scores are both increased by treatment, and failure to detect these effects with 'naive' RD estimates is due to unaccounted for composition effects.

To some extent, similar composition effects may bias LATE estimates for boys, although the change in conclusion is not as stark. Table 3.6 suggests that a failure to detect a significant treatment effect on boys' exam-taking rate due to imperfect data linkage is possible. In such case, just as the female subsample is, the male subsample would be subject to a change in composition.

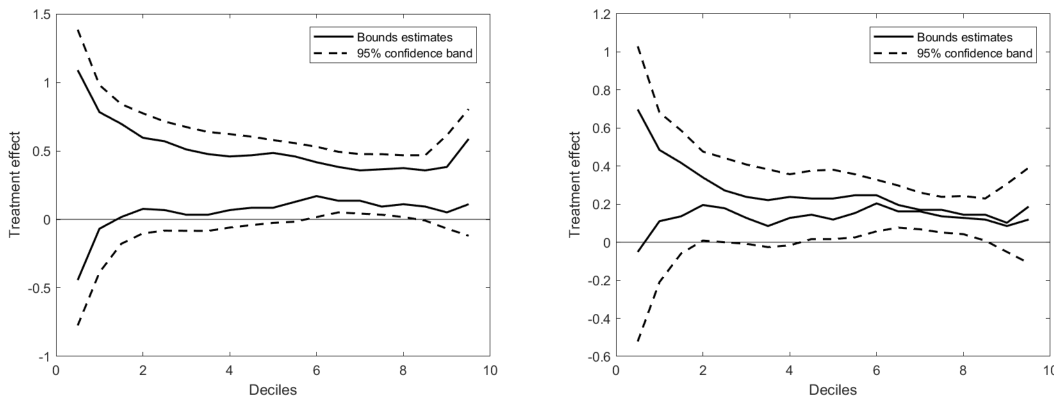
Table 3.7: Bounds estimates for TEs on end-of-high-school and post-secondary outcomes

	Most conservative		Least Conservative		Naive RD
	Low. B.	Upp. B.	Low. B.	Upp. B.	estimate
Boys					
End-of-high-school outcomes					
Exam score	0.015	0.584***	0.125**	0.306***	0.125***
Priority score	0.017	0.603***	0.124**	0.311***	0.124***
College application outcomes (first-ranked choice)					
2009 cutoff	0.141	0.264***	0.186*	0.223***	0.196***
Log-dist. from h-sch.	0.276	0.571	0.406	0.547	0.43
College admission outcomes (assigned program)					
2009 cutoff	0.098	0.24**	0.115	0.193**	0.123
Log-dist. from h-sch.	-0.036	0.247	0.076	0.173	0.103
Girls					
End-of-high-school outcomes					
Exam score	-0.065*	0.686***	0.063	0.343***	0.063
Final exam score	-0.077*	0.655***	0.046	0.327***	0.046
College application outcomes (first-ranked choice)					
2009 cutoff	-0.085	0.057	-0.086	0.006	-0.078
Log-dist. from h-sch.	0.059	0.336	0.131	0.2	0.153
College admission outcomes (assigned program)					
2009 cutoff	0.032	0.142*	0.036	0.082	0.046
Log-dist. from h-sch.	0.119	0.382*	0.178	0.254	0.25

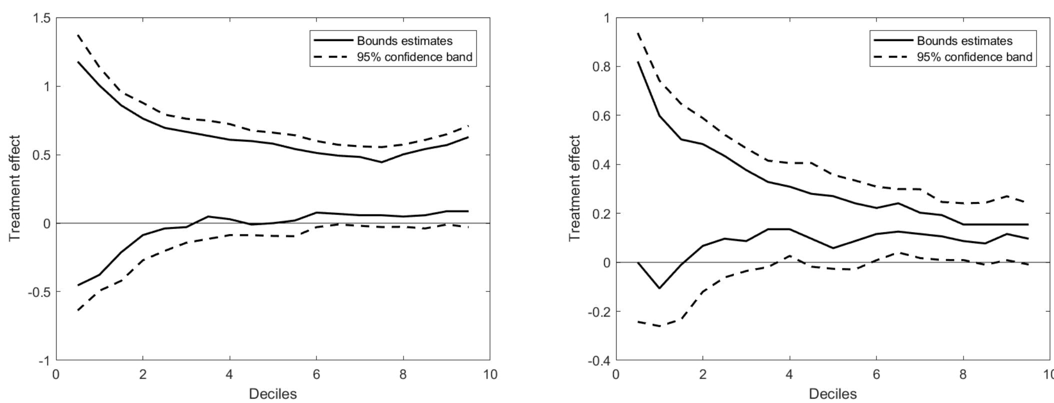
Bounds for TE on outcomes whose values can be naturally ordered in terms of quality (exam-passing rate, exam scores, 2009 cutoff of top-ranked and assigned programs) are derived under the SD assumption. Significance levels are based on bootstrap confidence intervals for bounds (5,000 repetitions). *significant at the 10% level; **significant at the 5% level;***significant at the 1% level.

3.5.2 Heterogeneous effects along outcomes distributions

Finally, Figure 3.2 depicts bounds for LQTEs on always-exam-takers derived under the SD assumption. Under the most conservative scenario, our estimates establish



(a) Boys



(b) Girls

Estimates depicted on the left (resp. right) panels are based on the most (resp. least) conservative scenario regarding sample selection shares (s_0, s_1). Estimated bounds are derived under the SD assumption. For outcome Y and quantile q the confidence band is constructed as $[L_{(Y,q)-}, U_{(Y,q)+}]$ where $[L_{(Y,q)-}, L_{(Y,q)+}]$ and $[U_{(Y,q)-}, U_{(Y,q)+}]$ are intervals of $\mathbb{R}$ covering with 95% probability, respectively, the lower and upper bounds of the TE on the q^{th} quantile of outcome Y. 95% confidence intervals for bounds are estimated by bootstrap (5,000 repetitions).

FIGURE 3.2: Bounds for QTEs on end-of-high-school exam score

that admission to an elite high school has a positive (and statistically significant) effect on the top half of the exam score distributions (plots for the priority score look similar). Again, under the most conservative scenario, making a quantitatively precise statement about the magnitude of the effect is impossible. At their narrowest point, bounds suggest an increase by .15 to .5 std. dev of the 65^{th} percentile of the score distribution for boys; and an increase by .05 to .6 std. dev of the 77^{th} percentile of the score distribution for girls. As one reduces the extent to which admission to an elite high school may negatively affect students' exam-taking rate, bounds get narrower (reaching about a .05-std dev-width for the top quartile of the distribution for boys, and a .12-std dev-width for the top quartile of the distribution for girls). The share of the score distribution positively affected by the treatment remains about the same under both scenarios..

These findings point at the importance of going beyond LATE, and looking at treatment effect heterogeneity. The LQTE analysis reveals that admission to an elite high school has positive effects on top students' exam performance. Notably, these effects are robust to sample selection and to the possible non-monotonicity of treatment effects on exam-taking and -passing rates.

3.6 Conclusion

This work highlights the importance of carefully addressing sample selection and missing outcome data, even in quasi-experimental settings such as the one offered by a RD design. It emphasizes that the nature of the outcomes of interest in the literature on selective schools (and, more generally, on treatment effects in education) makes sample selection and missing outcome data potentially pervasive. This paper extends the idea underlying well-known Lee bounds to the RD design, as well as to LQTE. It also discusses an intuitive assumption –the ranked average outcome assumption– that can be used as alternative or as a complement to the standard monotonicity assumption to refine identified bounds.

Chapter 4

An Exploration of the Mechanisms Through Which Elite Schools Affect Students' Outcomes

This chapter is based on a joint work with Meryam ZAIEM.

4.1 Introduction

Findings from the previous chapter, that reduced form estimates of the effects of admission to a selective high school vary across students and schools, suggest that valuable information may be missed when considering "admission to a selective high school" as a binary treatment. In this chapter, to explore the mechanisms through which schools affect students' academic outcomes, we broaden the analysis by looking at schools, selective or not, as vectors of inputs: being admitted to a given school means having access to a particular student body, set of teachers, and type of infrastructure.

Related literature

The multiplicity of inputs entering the acquisition and development of skills and human capital has been formalized in the skills production function model (see, for instance, Todd and Wolpin (2003), Cunha and Heckman (2008), Cunha and Heckman (2010), and Cunha et al. (2010)).

In the reduced-form literature on selective schools, peers effects are probably the most documented mechanisms, although there again, little consensus can be found in the literature. For instance, on the one hand, Ding and Lehrer (2007) show that in Chinese elite high schools, a higher average as well as a lower variance in peer performance do benefit students. On the other hand, Abdulkadiroğlu, Agarwal, and Pathak (2017) conclude from their study of the Boston exam schools that the change in peer characteristics that occurs as a marginal student gains admission to an elite school has little effect on outcomes[1].

The multiplicity of environmental changes induced by admission to a selective school has however been acknowledged and analyzed by richer studies[2]. In a study of the oversubscribed French "boarding schools of excellence", which admits students via a lottery, Behaghel et al. (2015) record that admitted students gain access to smaller and less disrupted classes; more demanding and more engaged teachers; academically stronger peers; and the generally more academically oriented environment of a boarding school where, for instance, more time is dedicated to homework and less to TV-watching. Very interestingly, the authors suggest that, even though these changes may eventually induce students to perform better by the end of their second

[1] This finding is similar to that of Angrist et al. (2016) in the charter school context.

[2] In the Massachusetts charter schools context, Angrist et al. (2013) explore possible sources of the heterogeneity in the effects of charter schools within the assortment of characteristics and practices each type of school is associated with. They note, for instance, that, as compared to other charter schools, the highly effective 'No Excuses' urban schools have a likelihood of exerting strict discipline, teacher selection (e.g. implementation of methods to provide teaching feedback to instructors, such as recording lectures; or hiring of former students from the Teach for America program), and special classroom practices such as cold-calling.

year at the school, the disruption created in their environment may cause them, in their first year, to experience a lower level of well-being, and therefore perform worse. In a similar vein, Pop-Eleches and Urquiola (2013) probe the behavioral responses –from students, parents and schools– to admission to higher-quality high schools in Romania. In particular, they report that marginally admitted students spend more time on homework; receive less help for it from their parents, who "may view educational quality and their own effort as substitutes"[3]; and are sorted by performance level within each school in such a way that leads them to have lower-achieving classmates and less certified instructors than the average elite school student.

Outline

The chapter is organized as follows. The next section describe the data supporting the analysis in this chapter. Section 4.3 presents the empirical strategy. Section 4.4 discusses the results and concludes.

4.2 Data

4.2.1 Data sources & Educational inputs

The analysis in this chapter relies on the datasets used in Chapter 3 and described in Section 3.2.2, as well as on an additional database of the Tunisian ministry of Education containing information about teachers, schools personnel and infrastructures. and a data set about the end of primary school exam and access to elite middle schools. Information on this latter database can be linked to the student data via school identifiers. Below, we briefly describe the construction of educational inputs of interest.

[3] Pop-Eleches and Urquiola (2013), page 4.

School inputs. The teacher database lists, for each high school and for the academic year 2015, all teachers employed in the school, along with the following characteristics: gender; years of teaching experience; subject(s) taught; tenure status; and degree(s) obtained (level and field). The school database lists, for each school and each academic year, a headcount of enrolled students by gender; a headcount of teachers, administrative staff and supervising staff; the number of classrooms and specialized rooms such as computer labs, libraries or gym infrastructures. We use this database to construct within-high-school measures of the following inputs: share of teachers with higher certification than the standard teaching certification level; share of tenured teachers; share of female teachers; average teaching experience; as well as average class size, and student/teacher ratio.

Peer characteristics. For the purpose of decomposing the causal effect of elite high schools on student outcomes into the respective effect of different educational inputs, peer characteristics should ideally be measured before peers are affected by the elite-school treatment, that is, at the end of middle school. Hence, performance at the national end-of-middle-school exam appears as a good measurement for peer achievement. As explained in Section 3.2.2, however, for the cohorts of interest, this exam was not compulsory and taken by only 25% of students, nationally. As a consequence, we resort to using end-of-high-school exam scores[4] to build measures of within-high-school peer achievement. Similarly, students' high-school major and socio-economic status (SES), which we use to compute the within-school shares of high-SES peers, and peers majoring in a scientific field, are also observed at the time of taking the end-of-high-school exam.

[4] which are, naturally, available only for those students who took this exam. A selection bias is therefore likely to affect our measures of peer characteristics. Given the share of high-school students taking the end-of-high-school exam relative to the share of middle-school students taking the end-of-middle-school exam, we however suppose this bias is not as significant as the one that would result from using end-of-middle-school exam grades to construct our measures.

Table 4.1: Descriptive statistics: High schools

	Non-select. h.sch.	Selective h.sch.	All h.sch.	Diff.
Peers: academic profile				
Mean exam score of peers	-.002	1.701	0.028	1.702***
	(0.506)	(0.160)	(0.551)	(.146)
Sd exam score of peers	0.781	0.522	0.777	-.259***
	(0.180)	(0.073)	(0.182)	(.052)
Share of students majoring	0.523	1	0.533	.477***
in science	(0.131)	(0)	(0.148)	(.038)
Peers: demographic profile				
Share of female students	0.591	0.619	0.592	.023**
	(0.056)	(0.030)	(0.056)	(.016)
Share of high-SES students	0.476	0.869	0.483	.392***
	(0.278)	(0.090)	(0.281)	(.081)
Share of students whose	0.036	0.230	0.039	.194***
father is a teacher	(0.026)	(0.061)	(0.036)	(.008)
Teacher characteristics				
Share of female teachers	0.468	0.478	0.468	.010
	(0.117)	(0.083)	(0.117)	(.034)
Share of tenured teachers	0.934	0.995	0.936	.061**
	(0.086)	(0.011)	(0.086)	(.025)
Mean teaching experience (years)	14.68	20.93	14.82	6.24***
	(3.865)	(1.966)	(3.941)	(1.12)
Share of teachers with	0.027	0.052	0.028	.024**
advanced degrees	(0.026)	(0.031)	(0.028)	(.008)
Teacher-student measures				
Mean class size	22.69	22.29	22.68	-.391
	(3.402)	(3.052)	(3.393)	(.991)
Students/teacher ratio	11.37	9.502	11.32	-1.85***
	(1.874)	(1.605)	(1.888)	(.548)
Total enrollment	802.7	704.3	800.5	-97.1
	(372.1)	(319.4)	(371.1)	(108.5)
Obs.	653	12	665	

The first three columns show sample means, with std. deviations in parentheses. The last column shows estimated mean differences between selective and non-selective schools, as well as their significance: *** significant at the 1% level; ** significant at the 5% level; * significant at the 10% level. Std. errors are reported in parentheses.

4.2.2 Descriptive statistics

Table 4.1 summarizes disparities in school characteristics between *lycées pilotes* and other public high schools. As expected given the description of selective high schools provided in the previous chapter, students enrolled in selective high schools are, on average, higher performers and more likely to major in science in high school, than those enrolled in non-selective high school. In addition, selective high schools typically enroll larger shares of female students and of high-SES students than non-selective high schools.

Regarding educational inputs, and as compared to non-selective high schools, selective high schools employ significantly more experienced teachers, with higher levels of certification, and who are also slightly more likely to have tenure. In addition, a larger share of teachers from selective schools hold a higher degree than a bachelor degree. On average, selective high schools also have a lower students-to-teacher ratio than non-selective high schools.

Table 4.2 shows statistics on the distribution, over selective high schools, of the treatment effects of admission to an elite high school on students' outcomes.

4.3 Empirical strategy

Admission to an elite school can be seen as a multidimensional treatment that changes at once multiple inputs entering the production of students' skills. The skills production function framework is a natural one to study the mechanisms through which admission to an elite school affect students' outcomes. In this framework, skills acquisition is seen as a cumulative process: a child's stock of skill at age a is the product of parental and school inputs provided to the child up to age a. It is also commonly

Table 4.2: Heterogeneity in TE on students' outcomes across selective schools

	Mean	Sd	Min	Max
End-of-high-school outcomes				
Take exam in 4 y.	0.01	0.018	-0.01	0.04
Pass exam in 4 y.	0.01	0.022	-0.03	0.04
Pass exam in 6 y.	0.02	0.023	-0.03	0.04
Score at exam	0.16	0.106	-0.02	0.32
College application outcomes				
Selectivity of 1st choice	0.13	0.134	-0.07	0.35
Distance to 1st choice	0.08	0.231	-0.21	0.53
College admission outcomes				
Selectivity of assignment	0.12	0.136	-0.09	0.36
Distance to assignment	0.25	0.226	-0.07	0.65

For each students' outcome, the table shows the mean, standard deviation, minimum, and maximum, over the twelve *lycées pilotes*, of the treatment effects of admission to each *lycée pilote*.

assumed that there is an innate, genetic determinant of the child ability.

However, Todd and Wolpin (2003) describe the challenges met when trying to estimate such model while some relevant family and school inputs, past or current, are unobserved to the econometrician. In addition, they highlight that the use of proxies for these unobserved variables may not be sufficient to alleviate the bias resulting from missing inputs. While the data available to us does not allow to derive a causal decomposition of the effects of admission to an elite high schools on students outcomes, possible channels through which elite schools affect students' outcomes may be suggested by an exploration of the associations that exist between change in inputs caused by treatment and changed in outcomes cause by treatment.

To assess the extent to which causal effect on student outcomes correlate with treatment-induced changes in the academic environment, we use the following procedure. (i) For each student outcome of interest Y, we obtain school-specific RD-estimated of the treatment effect on Y. We denote by $\tau_{Y,s}$ the estimate for outcome

Y and elite school $s = 1, \ldots, 12$. (ii) For each characteristic Z of the academic environment described in Table 4.1, we obtain a school-specific RD-estimate of the treatment-induced change in Z. We denote by $\tau_{Z,s}$ the estimate for characteristic Z and elite school $s = 1, \ldots, 12$. (iii) We evaluate the correlation of interest by estimating the following equation for each pair (Y, Z):

$$\tau_{Y,s} = b \cdot \tau_{Z,s} + \epsilon_s. \tag{4.1}$$

The associations of interest are captured by the parameter b. Estimates are shown in the next section.

4.4 Results

4.4.1 Treatment-induced changes in educational inputs

Table 4.3 shows results from part (ii) of the empirical strategy described in the previous section. Admission to an elite high school changes dramatically the profile of peers marginally accepted students get exposed to. It increases by half a standard deviation the mean performance of peers at the end-of-high school exam, and decreases the variance in performance at the school level (standard deviation of test scores is lowered by .08). Relative to marginally rejected students, students marginally admitted to selective high schools also have a larger share of peers majoring in science (+13 percentage points), that are female (+4 percentage points), from high-SES backgrounds (+5 percentage points), and whose father is a teacher (+6 percentage points).

Table 4.3: Treatment-induced changes in peers and school characteristics

	Opt. Bdw.	.5× Opt. Bdw.
Peers: academic profile		
Mean exam score of peers	0.561***	0.455***
	(0.054)	(0.069)
Obs.	7738	3964
Sd exam score of peers	-0.109***	-0.086***
	(0.009)	(0.012)
Obs.	8878	4608
Share of students majoring in science	0.158***	0.127***
	(0.016)	(0.020)
Obs.	7722	3938
Peers: demographic profile		
Share of female students	0.043***	0.036***
	(0.005)	(0.006)
Obs.	10879	5833
Share of high-SES students	0.063***	0.045***
	(0.010)	(0.011)
Obs.	12944	6863
Share of students whose father is a teacher	0.070***	0.054***
	(0.008)	(0.009)
Obs.	9043	4608
Teacher characteristics		
Share of female teachers	-0.014**	-0.007
	(0.005)	(0.006)
Obs.	15890	8517
Mean teaching experience (years)	1.196***	0.835***
	(0.201)	(0.212)
Obs.	12079	6405
Share of tenured teachers	0.007***	0.006***
	(0.001)	(0.002)
Obs.	13869	7356
Share of teachers with advanced degrees	0.016***	0.015***
	(0.002)	(0.002)
Obs.	19286	10620
Teacher-student measures		
Class size	-1.152***	-0.911***
	(0.194)	(0.182)
Obs.	17285	9362
Student/professor ratio	-1.024***	-0.671***
	(0.109)	(0.135)
Obs.	10879	5833

Robust std errors in parentheses. Std errors adjust for clustering at the (school district × year)-level. *significant at the 10% level; **significant at the 5% level;***significant at the 1% level.

Regarding teachers, as compared to marginally rejected students, students marginally admitted to selective high schools are exposed to more experienced teachers (by about a year), with slightly, but significantly, higher rates of tenure and advanced certification. Marginally admitted students also benefit from a slightly lower students-to-teacher ratio than their non-admitted counterparts.

Interestingly, the estimated differences in Table 4.3 have same sign but are much smaller in magnitude than the differences shown in Table 4.1. This means that the average high school attended by marginally rejected applicants is much more similar to the average *lycée pilote* than the average high school in the country is.

4.4.2 Correlations between changes in inputs and in outcomes

Tables 4.4 and 4.5 show results from part (iii) of the empirical strategy described in Section 4.3. A striking result is the general lack of statistically significant correlation between treatment-induced changes in inputs and treatment-induced changes in students' outcomes.

If anything, among the inputs studied here, peer-related inputs seem to be the main mediators of elite schools' effects on students' outcomes. Treatment effect on end-of-high-school exam performance (Table 4.4) is positively and significantly correlated with a treatment-induced change in the within-school average peer performance; and strongly negatively correlated with a treatment-induced change in the within-school variance in peer performance. These correlations are intuitive: a (treated) student's performance benefit from higher-performing and more homogeneous peers. Treatment effect on end-of-high-school exam performance is also positively correlated with a treatment-induced change the within-school share of students whose father is a teacher. Peer-related characteristics are also correlated with measures of students' future geographic mobility (Table 4.5). A larger (positive) treatment effect on the distance from home to the universities students applied to is positively associated with a larger treatment-induced change in the within-school average peer performance; and negatively associated with a treatment-induced increase in the within-school variance in peer performance. It is also positively associated with an increase in the share of peers majoring in scientific fields in high-school, and in the share of peers whose father is a teacher. The positive association with an increase in the share of peers whose father is a teacher is also true for the change in the distance from home to the the university students get assigned to.

Changes in teacher characteristics, as well as in the gender and socioeconomic composition of students, on the other hand, do not seem to explain much of the treatment-induced changes in students' outcomes.

Table 4.4: Correlations between treatment-induced changes in inputs and changes in students' end-of-high-school outcomes

	Take exam in 4 y.	Pass exam in 4 y.	Pass exam in 6 y.	Score at the exam
Peers: academic profile				
Mean exam score of peers	-0.035*	0.026	0.005	0.213*
	(0.018)	(0.024)	(0.026)	(0.103)
Sd exam score of peers	0.267**	-0.093	0.070	-1.250*
	(0.093)	(0.148)	(0.156)	(0.628)
Share of students majoring in science	-0.132*	0.107	0.026	0.633
	(0.063)	(0.084)	(0.095)	(0.400)
Peers: demographic profile				
Share of female students	-0.166	0.393	0.352	0.559
	(0.209)	(0.226)	(0.246)	(1.259)
Share of high-SES students	0.022	0.047	-0.008	-0.525
	(0.106)	(0.127)	(0.134)	(0.607)
Share of students whose father is a teacher	-0.341***	-0.059	-0.247	1.793**
	(0.106)	(0.180)	(0.173)	(0.692)
Teacher characteristics				
Share of female teachers	-0.136	-0.060	-0.127	-0.631
	(0.225)	(0.273)	(0.284)	(1.336)
Teaching experience (years)	0.004	0.019	0.017	0.035
	(0.011)	(0.011)	(0.012)	(0.062)
Share of tenured teachers	0.088	0.224	-0.288	-3.516
	(0.719)	(0.859)	(0.899)	(4.102)
Share of teachers with advanced degrees	-0.775*	0.253	-0.028	2.960
	(0.356)	(0.512)	(0.543)	(2.376)
Teacher-student measures				
Class size	-0.015**	-0.014	-0.017*	0.057
	(0.006)	(0.008)	(0.008)	(0.042)
Student/teacher ratio	-0.004	-0.030*	-0.019	-0.032
	(0.016)	(0.016)	(0.019)	(0.091)

Std errors in parentheses. *significant at the 10% level; **significant at the 5% level;***significant at the 1% level.

Table 4.5: Correlations between treatment-induced changes in inputs and changes in students' post-secondary outcomes

	Selectivity of 1st ch.	Distance to 1st ch.	Selectivity of assignment	Distance to assignment
Peers: academic profile				
Mean exam score of peers	0.073	0.569**	0.170	0.119
	(0.151)	(0.194)	(0.145)	(0.256)
Sd exam score of peers	-1.136	-3.547**	-1.239	-0.948
	(0.890)	(1.219)	(0.890)	(1.592)
Share of students majoring in science	0.417	1.560**	0.686	0.578
	(0.416)	(0.569)	(0.385)	(0.714)
Peers: demographic profile				
Share of female students	0.111	3.218	0.756	-1.516
	(1.776)	(2.892)	(1.783)	(2.960)
Share of high-SES students	0.536	0.364	-0.280	2.484
	(0.961)	(1.680)	(0.984)	(1.448)
Share of students whose father is a teacher	0.908	3.475**	0.959	2.702*
	(0.954)	(1.322)	(0.962)	(1.448)
Teacher characteristics				
Share of female teachers	1.349	-1.886	-1.297	-0.069
	(1.647)	(2.875)	(1.673)	(2.871)
Teaching experience (years)	0.023	0.066	0.055	-0.091
	(0.065)	(0.110)	(0.063)	(0.106)
Share of tenured professors	-2.084	-3.942	-0.722	8.625
	(5.744)	(9.900)	(5.852)	(9.370)
Share of teachers with advanced degrees	2.220	10.397*	1.968	2.621
	(3.282)	(4.768)	(3.341)	(5.603)
Teacher-student measures				
Class size	0.036	0.063	0.013	0.249***
	(0.052)	(0.089)	(0.054)	(0.043)
Student/professor ratio	0.008	-0.267	-0.033	0.158
	(0.111)	(0.171)	(0.112)	(0.180)

Std errors in parentheses. *significant at the 10% level; **significant at the 5% level;***significant at the 1% level.

Chapter 5

Conclusion

Using administrative data from Tunisia, the goal of this dissertation was to provide an empirical study of some aspects of school choice. From the point of view of application, Chapter 2 focused on assignment mechanisms used by school districts to allocate available seats to students. From the point of view of admission, Chapters 3 and 4 investigated the consequences of school choice on students' outcomes.

Focusing on the extensively-used deferred acceptance algorithm (DA), Chapter 2 quantified the welfare effects of enabling students to update their expectations about their admission chances to academic programs in a setting where they cannot apply to all the alternatives in their choice set. Using Tunisia's post-secondary education application system as a model, it documented a simple way to enable this updating in the context of DA-based assignment mechanisms, which are extensively used around the world to assign students to schools. The main empirical challenge faced in this chapter was the need, in order to perform a counterfactual analysis to compare students' application and assignments under scenarios with different levels of updating, to disentangle the extent to which students' application choices are driven by their preferences vs. their expectations about their admission chances. To circumvent this common identification challenge in the empirical literature on school choice, Chapter

2 built on the quasi-experimental design induced by the Tunisian procedure. This design allowed to, first, recover students' preferences for university programs without taking a stand on their expectations; and to then take preferences as given and characterize students' expectations.

Focusing on selective high schools, Chapter 3 and 4 investigated the extent to which and channels through which elite high school can affect students' secondary and post-secondary academic trajectories. The main empirical challenge faced in this chapter was the need to account for measurement error and sample selection when identifying the treatment effects of admission to a selective high schools on students' outcomes. To address this issue, Chapter 3 extended the idea underlying well-known Lee bounds to the RD design, as well as to LQTE, and discussed an intuitive assumption – the ranked average outcome assumption– that can be used as alternative or as a complement to the standard monotonicity assumption to refine identified bounds.

Appendix A

Appendices to Chapter 2

A.1 Theory

This appendix provides supplemental information to Section 2.2. Part A.1.1 gives a more general description of the DA. Part A.1.2 gives a proof of Proposition 2.

A.1.1 Deferred acceptance algorithm (Gale and Shapley (1962))

DA

Step 1/ Schools receive applications from students who ranked them first in their list. Schools that received fewer applications than their capacity hold on to these applications. Each school j that received more applications than its capacity q_j sends rejection decision to applicants: it temporarily hold on to the q_j applicants with highest priority, and rejects all others (if any). Students receive the rejection notifications sent.

Step (k+1)/ For any $k \geqslant 1$, students who received a rejection notification at step k send an application to the school ranked next on their list. Schools received these applications. Schools then consider their total pool of applications –those just received, and those they held on at step k (if any). Schools which have fewer applications than their capacity hold on to these applications. Each school

j with excess applications sends rejection decision to applicants: it temporarily hold on to the q_j applicants with highest priority, and rejects all others (if any). Students receive the rejection notifications sent.

Stop/ The algorithm stops after all students who received rejections have exhausted their list of acceptable schools. School formally admit applicants they hold on to at this stage.

A.1.2 Proof of Proposition 2

Proposition 2. (a) Condition 1 (below) is a sufficient condition for students not have a strict incentive to misreport their preferences over their choice set.

(b) Under Assumption 1 (below), Condition (1) is a sufficient condition for students not misreport their preferences over their choice set.

Condition 1. Student i has a perceived *eligibility* probability 1 for (at least) one of her ten most-preferred programs (among those not declared to be full).

Assumption 1. When indifferent between doing so or not, a student does not left-censor nor mis-order her application list relative to her unconstrained preference ranking (over her own choice set). In other words, a student left-censors or mis-orders her application list relative to her unconstrained preference ranking (over her own choice set) only when it is *strictly* profitable to do so.

Proof. Deviation from truth-telling involve at least one of the following:

- Misrepresenting one's preferences by not reporting one's M most-preferred alternatives –i.e. reporting a subset of alternatives that is left-censored or not consecutive relative to one's unrestricted preference ranking.

- Misrepresenting one's preferences by not reporting alternatives in decreasing order of flow utility –i.e. reporting a subset of alternatives that is not well ordered relative to one's unrestricted preference ranking.

First, reporting a subset of alternatives that is not well ordered relative to one's unrestricted preference ranking is never *strictly* profitable since one needs to be rejected from an higher-ranked choice in order to be considered for admission into a lower-ranked choice.

Now, suppose Condition 1 holds for some student i. WLOG, say student i thinks she has probability 1 to clear the *ex-post* cutoff of program j^*, which is ranked second in her unrestricted preference ranking. Denote j_1 the alternative ranked first in her unrestricted preference ranking.

- If i thinks she has probability 0 to clear the cutoff of j_1, she is indifferent between any ordered list starting with: $\{j_1, j^*\}$ or $\{j^*\}$. In that case, it is therefore not strictly profitable to omit j_1 from the list.
- Suppose i thinks she has probability $p_1 > 0$ to clear the cutoff of j_1. Submitting ordered list $\{j_1, j^*\}$ she thinks she will be assigned to j^* unless she is assigned to j_1, which she prefers the j^*, and which occurs with non-0 probability. In expectation, she is then better of submitting $\{j_1, j^*\}$ than $\{j^*\}$, with which she would be admitted to j^* with probability 1.

This shows Part (a) of Proposition 1. Part (b) follows directly from Part (a) and Assumption 1.

A.2 Sharpness & validity of the regression discontinuity design

This appendix provides standard graphical evidence supporting the sharpness and validity of the regression discontinuity (RD) design used in the analysis of Section 2.3.3.

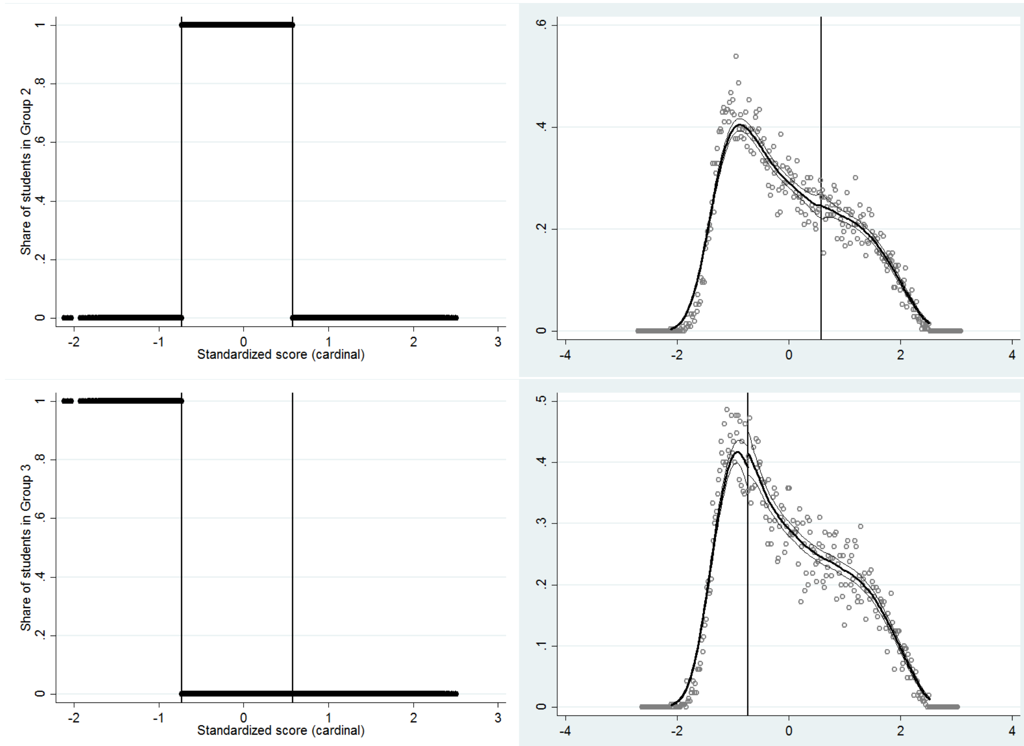
The division of the applicant pool in three groups creates assignment cutoffs. The top left panel of Figure A.1 displays students' probability to be assigned to Group 2 as a function of the running variable, that is, their priority score (or priority ranking, in ordinal terms). This probability jumps from 0 to 1 at standardized score .578 (score, 3,307 the Group 1/Group 2 cutoff), indicating the sharpness of the discontinuity. The probability drops back from 1 to 0 at standardized score -.726 (rank 7,708, the Group 2/Group 3 cutoff). The bottom left panel of Figure A.1 shows students' probability to be assigned to Group 3 as a function of the running variable. It jumps from 0 to 1 at standardized score -.726.

I use McCrary (2008)'s test as an evidence of the validity of the design, and reject the presence of a discontinuity of the running variable at standardized scores -.578 (estimated discontinuity -.004, with standard error .080), and -.726 (estimated discontinuity .069, with standard error .065). The right panel of Figure A.1 illustrates the test, showing the density of the standardized test score.

A.3 Preferences

This appendix gathers supporting materials and details to Section 2.4.

Part A.3.1 shows descriptive statistics supporting the extrapolation argument made in Section 2.4.1. Part A.3.2 shows utility parameters estimates separately for both truthful samples; as well as for alternative bandwidth choice for the sample of 'top'



Note: The top (resp. bottom) left panel shows students' probability to be assigned to Group 2 (resp. 3) as a function of the running variable. It presents a sharp discontinuity at rank 3,307 (resp. 7,709). The top (resp. bottom) right panel shows the density of the running variable, whose continuity at rank 3,307 (resp. 7,709) cannot be rejected.

FIGURE A.1: Sharpness and validity of the RD design: graphical evidence

students. Complementing Section 2.4.2, part A.3.3 shows additional empirical validation for the bandwidth choice. Part A.3.4 discuss an alternative estimation strategy that could yield more precise utility parameters by using all students' application lists, rather than only those for truthful students.

A.3.1 Extrapolation: tables

Tables A.1 and A.2 show descriptive statistics for key student and choice characteristics, comparatively for three subsamples of interest: the whole population, as well as each of the two truthful subsamples. In addition, it shows statistics for what would be the top sample in a single-phase implementation of the DA: the top of Group 1 only.

A.3.2 Estimates: sensitivity analysis

Table A.3 shows utility parameter estimates obtained for each of the truthful subsamples separately. Table A.4 shows estimation results for alternative choices for the bandwidth of top students. Columns (1), (2) and (3) show estimates for a bandwidth size equal to twice, five and ten times the original bandwidth size, respectively.

Table A.1: Comparative statistics: students

	Sample	Mean	Sdev.	Median	Min	Max	Obs.
Female	Pop.	.53	.50	1	0	1	10,897
	Top	.55	.50	1	0	1	636
	Top G1	.55	.50	1	0	1	205
	Sh.-list	.53	.50	1	0	1	3,236
High SES	Pop.	.60	.49	1	0	1	10,897
	Top	.73	.45	1	0	1	619
	Top G1	.89	.33	1	0	1	205
	Sh.-list	.63	.48	1	0	1	3,236
From Tunis	Pop.	.30	.46	0	0	1	10,897
	Top	.35	.48	0	0	1	619
	Top G1	.38	.49	0	0	1	205
	Sh.-list	.26	.44	0	0	1	3,236
From Coast (excl. Tunis)	Pop.	.48	.50	0	0	1	10,897
	Top	.48	.50	0	0	1	619
	Top G1	.54	.50	1	0	1	205
	Sh.-list	.53	.50	0	0	1	3,236
From West/Interior	Pop.	.19	.39	0	0	1	10,897
	Top	.15	.36	0	0	1	619
	Top G1	.09	.28	0	0	1	205
	Sh.-list	.18	.38	0	0	1	3,236
From South	Pop.	.03	.17	0	0	1	10,897
	Top	.02	.14	0	0	1	619
	Top G1	0	0	0	0	0	205
	Sh.-list	.03	.17	0	0	1	3,236
STEM h.-sch. perf.	Pop.	0	.85	-.10	-2.26	1.96	10,897
	Top	.52	.96	.49	-1.15	1.96	619
	Top G1	1.70	.10	1.70	1.34	1.96	205
	Sh.-list	.04	.93	-.08	-2.15	1.96	3,236
non-STEM h.-sch. perf.	Pop.	0	.79	0	-2.62	2.40	10,897
	Top	.47	.91	.42	-1.99	2.40	619
	Top G1	1.48	.32	1.49	.62	2.40	205
	Sh.-list	.03	.85	.01	-2.24	2.40	10,897

Note: 'Top' refers to the subset of students at the top of each group that is used for estimation of utility parameters. 'Sh.-list' refers to the subset of all students listing strictly fewer than 10 programs, also used for estimation of utility parameters. By contrast to 'Top', 'Top G1' refers to students in 'Top' who are also in Group 1. 'Pop.' indicates population statistics. In the second panel, STEM (resp. non-STEM) high-school performance is the unweighted average of the student's standardized scores at the Math, Physics, Natural Sciences, and Comp. Sci. (resp. English, French, Arabic, and Philosophy) tests of the end-of-high-school national exam.

Table A.2: Comparative statistics: application behaviors

	Sample	Mean	Sdev.	Median	Min	Max	Obs.
Distance home-sch.:	Pop.	39.4	75.7	0	0	1,800	10,897
min over list	Top	51.9	109.2	0	0	1,800	619
(km)	Top G1	89.1	157.4	65	0	1,800	205
	Sh.-list	37.9	71.6	0	0	1,800	3,236
Distance home-sch.:	Pop.	235.9	291.2	191	0	1,800	10,897
max over list	Top	535.7	670.9	235	0	1,800	619
(km)	Top G1	1,209.7	766.9	1,800	0	1,800	205
	Sh.-list	316.3	491.6	163	0	1,800	3,236
Distance home-sch.:	Pop.	123.1	169.1	82.3	0	1,800	10,897
avg. over list	Top	340.6	484.7	100.4	0	1,800	619
(km)	Top G1	824.8	573	1,285.7	0	1,800	205
	Sh.-list	161	272.1	77.7	0	1,800	3,236
2009 cutoff:	Pop.	.33	.23	.29	.002	.991	10,897
min over list	Top	.26	.24	.25	.002	.84	619
	Top G1	.02	.03	.005	.002	.23	205
	Sh.-list	.34	.25	.32	.002	.991	3,236
2009 cutoff:	Pop.	.71	.24	.72	.012	1	10,897
max over list	Top	.57	.34	.57	.012	1	619
	Top G1	.17	.18	.03	.012	.80	205
	Sh.-list	.67	.28	.70	.012	1	3,236
2009 cutoff:	Pop.	.51	.23	.51	.008	.996	10,897
avg over list	Top	.41	.28	.42	.008	.94	619
	Top G1	.08	.08	.01	.008	.34	205
	Sh.-list	.50	.27	.51	.008	.996	3,236
At least 1 choice in	Pop.	.21	.41	0	0	1	10,897
Earth Sc.	Top	.21	.41	0	0	1	619
	Top G1	.02	.14	0	0	1	205
	Sh.-list	.15	.36	0	0	1	3,236
At least 1 choice in	Pop.	.02	.16	0	0	1	10,897
Soc. Sc.	Top	.01	.13	0	0	1	619
	Top G1	0	0	0	0	0	205
	Sh.-list	.02	.11	0	0	1	3,236
At least 1 choice in	Pop.	.03	.17	0	0	1	10,897
Law	Top	.02	.13	0	0	1	619
	Top G1	0	0	0	0	0	205
	Sh.-list	.02	.14	0	0	1	3,236
Total # programs	Pop.	609	n/a	n/a	n/a	n/a	10,897
applied to	Top	429	n/a	n/a	n/a	n/a	619
	Top G1	62	n/a	n/a	n/a	n/a	205
	Sh.-list	609	n/a	n/a	n/a	n/a	3,236

Note: 'Top' refers to the subset of students at the top of each group that is used for estimation of utility parameters. 'Sh.-list' refers to the subset of all students listing strictly fewer than 10 programs, also used for estimation of utility parameters. By contrast to 'Top', 'Top G1' refers to students in 'Top' who are also in Group 1. 'Pop.' indicates population statistics.

Table A.3: Utility parameter estimates – truthful samples (1/3)

	(1) Main	(2) Lin. in distance	(3) Main	(4) Lin. in distance
Distance (100km)	-2.083***	-0.931***	-2.019***	-1.026***
	(0.12)	(0.08)	(0.08)	(0.05)
× high SES	0.085	0.181	0.021	0.147*
	(0.08)	(0.11)	(0.04)	(0.06)
Distance (100km) sq.	0.223***		0.222***	
	(0.02)		(0.01)	
Past-year marginal admit	2.503***	3.613***	2.097***	2.794***
	(0.60)	(0.58)	(0.31)	(0.31)
× high SES	-0.952	-0.706	0.467	0.679
	(0.87)	(0.88)	(0.40)	(0.41)
Past-year marginal admit sq.	0.761	1.055	-1.015**	-0.668*
	(0.70)	(0.71)	(0.32)	(0.34)
× high SES	2.394*	1.771	1.017*	0.616
	(0.95)	(0.93)	(0.40)	(0.40)
Dist. × Past-year marg. adm.	1.107***		0.885***	
	(0.12)		(0.07)	
Degree: Bachelor (LF)	0.302*	0.303*	0.595***	0.597***
	(0.12)	(0.12)	(0.06)	(0.06)
× h-s perf.	0.431**	0.449**	0.393***	0.399***
	(0.14)	(0.14)	(0.06)	(0.06)
× high SES	0.206	0.178	0.009	-0.016
	(0.16)	(0.16)	(0.07)	(0.07)
Degree: Adv. degree	2.577***	2.551***	2.603***	2.592***
	(0.17)	(0.17)	(0.09)	(0.09)
× h-s perf.	0.884***	0.919***	1.941***	1.955***
	(0.15)	(0.16)	(0.09)	(0.08)
× high SES	-0.024	-0.062	-0.175	-0.210*
	(0.20)	(0.20)	(0.10)	(0.10)
Program location: Tunis	0.902***	1.128***	0.499***	0.728***
	(0.14)	(0.15)	(0.09)	(0.10)
Program location: Coast	0.913***	0.951***	0.304***	0.348***
	(0.13)	(0.13)	(0.07)	(0.08)
Program location: Abroad	-18.693***	-19.996***	-10.233***	-11.335***
	(2.33)	(2.49)	(1.36)	(1.45)
× STEM h-s perf.	7.492***	7.837***	4.319***	4.671***
	(1.25)	(1.40)	(0.77)	(0.82)
× non-STEM h-s perf.	3.671***	3.874***	1.969***	2.069***
	(0.42)	(0.44)	(0.30)	(0.33)
× high SES	0.064	0.368	-0.261	0.040
	(0.50)	(0.46)	(0.35)	(0.37)
Sample	Bdw	Bdw	Short	Short
PseudoObs.	4,927	4,927	24,961	24,961
Obs.	624	624	3,629	3,629

Utility parameter estimates – truthful samples (2/3)

	(1)	(2)	(3)	(4)
Field: Arts	1.751**	1.741**	2.962***	2.963***
	(0.57)	(0.57)	(0.32)	(0.32)
× STEM h-s perf.	1.484***	1.478***	1.676***	1.678***
	(0.41)	(0.40)	(0.19)	(0.19)
× non-STEM h-s perf.	-1.064	-1.067*	-1.273***	-1.256***
	(0.55)	(0.54)	(0.24)	(0.24)
× female	-0.090	-0.060	-1.151**	-1.149**
	(0.55)	(0.55)	(0.36)	(0.36)
Field: Educ.	-0.572	-0.625	2.287***	2.318***
	(1.90)	(1.94)	(0.43)	(0.43)
× STEM h-s perf.	2.727**	2.762**	1.142**	1.084**
	(0.88)	(0.88)	(0.35)	(0.36)
× non-STEM h-s perf.	-0.147	-0.145	-0.883	-0.908
	(0.98)	(1.00)	(0.53)	(0.54)
× female	-2.044***	-2.044***	-1.676**	-1.653**
	(0.57)	(0.58)	(0.58)	(0.58)
Field: Soc. Sc.	-0.120	-0.103	1.033**	1.064**
	(0.87)	(0.88)	(0.37)	(0.37)
× STEM h-s perf.	1.490	1.508	0.775*	0.797*
	(0.97)	(0.98)	(0.32)	(0.32)
× non-STEM h-s perf.	-0.944	-0.946	-0.882*	-0.868*
	(0.94)	(0.93)	(0.36)	(0.35)
× female	-0.640	-0.628	-1.455***	-1.461***
	(0.92)	(0.92)	(0.43)	(0.43)
Field: Eco/Mgmt	2.995***	2.931***	3.775***	3.765***
	(0.56)	(0.55)	(0.33)	(0.33)
× STEM h-s perf.	0.981**	1.050**	1.160***	1.191***
	(0.34)	(0.33)	(0.21)	(0.21)
× non-STEM h-s perf.	-1.045*	-1.050*	-1.189***	-1.177***
	(0.49)	(0.49)	(0.25)	(0.25)
× female	-0.478	-0.466	-1.199***	-1.194***
	(0.51)	(0.51)	(0.36)	(0.36)
Field: Law	0.497	0.454	2.386***	2.378***
	(1.07)	(1.06)	(0.43)	(0.42)
× STEM h-s perf.	0.199	0.296	0.451	0.501
	(0.73)	(0.71)	(0.36)	(0.36)
× non-STEM h-s perf.	-1.689*	-1.755*	-0.161	-0.167
	(0.75)	(0.77)	(0.36)	(0.35)
× female	0.943	0.940	-1.455**	-1.454**
	(1.18)	(1.16)	(0.48)	(0.48)
Field: Math/Comp.Sci.	3.863***	3.801***	4.405***	4.399***
	(0.55)	(0.55)	(0.32)	(0.32)
× STEM h-s perf.	1.113**	1.166**	1.338***	1.366***
	(0.36)	(0.35)	(0.19)	(0.19)
× non-STEM h-s perf.	-1.370**	-1.364**	-1.562***	-1.553***
	(0.48)	(0.48)	(0.25)	(0.24)
× female	-0.837	-0.819	-1.337***	-1.328***
	(0.51)	(0.51)	(0.36)	(0.36)

Utility parameter estimates – truthful samples (3/3)

	(1)	(2)	(3)	(4)
Field: Phys./Chem./Engin.	3.530***	3.464***	4.187***	4.171***
	(0.55)	(0.55)	(0.31)	(0.31)
× STEM h-s perf.	1.145***	1.180***	1.301***	1.322***
	(0.33)	(0.33)	(0.18)	(0.18)
× non-STEM h-s perf.	-1.244**	-1.243**	-1.626***	-1.612***
	(0.47)	(0.46)	(0.24)	(0.23)
× female	-0.835	-0.813	-1.517***	-1.511***
	(0.51)	(0.51)	(0.35)	(0.35)
Field: Health/Life Sc.	3.371***	3.286***	3.603***	3.562***
	(0.56)	(0.56)	(0.32)	(0.32)
× STEM h-s perf.	0.907**	0.978**	1.102***	1.135***
	(0.34)	(0.33)	(0.19)	(0.19)
× non-STEM h-s perf.	-1.467**	-1.436**	-1.322***	-1.300***
	(0.50)	(0.50)	(0.25)	(0.24)
× female	0.226	0.259	-0.536	-0.517
	(0.54)	(0.54)	(0.36)	(0.36)
Field: Earth Sc.	2.234***	2.204***	2.105***	2.109***
	(0.56)	(0.56)	(0.32)	(0.32)
× STEM h-s perf.	-0.130	-0.134	0.264	0.276
	(0.35)	(0.35)	(0.21)	(0.21)
× non-STEM h-s perf.	-1.278**	-1.256**	-1.732***	-1.730***
	(0.48)	(0.48)	(0.26)	(0.25)
× female	-0.654	-0.632	-0.929*	-0.941**
	(0.53)	(0.53)	(0.36)	(0.36)

Table A.4: Utility parameter estimates – top bandwidth choice (1/3)

	(1)	(2)	(3)
Distance (100km)	-2.011***	-1.851***	-1.807***
	(0.10)	(0.08)	(0.06)
× high SES	0.035	0.005	0.031
	(0.06)	(0.04)	(0.03)
Distance (100km) sq.	0.215***	0.194***	0.189***
	(0.02)	(0.01)	(0.01)
Past-year marginal admit	2.735***	2.725***	2.983***
	(0.45)	(0.32)	(0.26)
× high SES	-0.090	0.892*	1.141***
	(0.62)	(0.41)	(0.33)
Past-year marginal admit sq.	0.151	0.090	-0.937***
	(0.45)	(0.29)	(0.24)
× high SES	1.146	-0.026	-0.156
	(0.62)	(0.38)	(0.30)
Dist. × Past-year marg. adm.	1.097***	0.959***	0.841***
	(0.09)	(0.07)	(0.06)
Degree: Bachelor (LF)	0.411***	0.407***	0.463***
	(0.09)	(0.06)	(0.05)
× h-s perf.	0.411***	0.442***	0.354***
	(0.09)	(0.06)	(0.04)
× high SES	0.142	0.181**	0.163**
	(0.11)	(0.07)	(0.05)
Degree: Adv. degree	2.783***	2.768***	2.699***
	(0.11)	(0.08)	(0.06)
× h-s perf.	1.047***	1.107***	1.335***
	(0.10)	(0.07)	(0.06)
× high SES	-0.109	0.018	-0.069
	(0.13)	(0.09)	(0.07)
Program location: Tunis	0.800***	0.779***	0.701***
	(0.11)	(0.08)	(0.06)
Program location: Coast	0.753***	0.695***	0.574***
	(0.10)	(0.07)	(0.06)
Program location: Abroad	-25.303***	-23.147***	-18.036***
	(2.10)	(1.98)	(2.05)
× STEM h-s perf.	10.733***	9.989***	7.545***
	(1.08)	(0.97)	(1.02)
× non-STEM h-s perf.	4.245***	3.745***	3.075***
	(0.43)	(0.40)	(0.43)
× high SES	0.290	0.274	0.347
	(0.50)	(0.43)	(0.38)
Sample	Bdw × 2	Bdw × 5	Bdw × 10
PseudoObs.	10,169	26,384	53,485
Obs.	1,252	3,134	6,250

Utility parameter estimates – truthful samples (2/3)

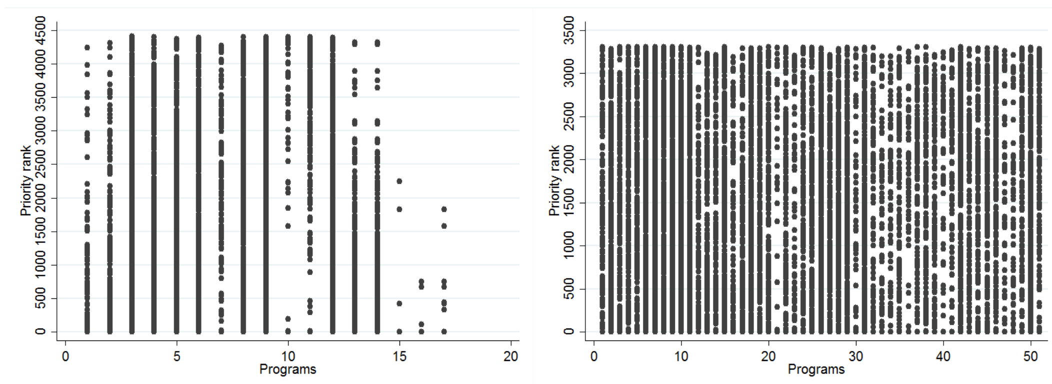
	(1)	(2)	(3)
Field: Arts	1.705***	2.211***	2.473***
	(0.35)	(0.22)	(0.17)
× STEM h-s perf.	1.523***	2.145***	2.036***
	(0.26)	(0.22)	(0.16)
× non-STEM h-s perf.	-0.961*	-1.209***	-1.144***
	(0.39)	(0.25)	(0.16)
× female	0.087	-0.264	-0.440*
	(0.41)	(0.26)	(0.19)
Field: Educ.	-0.181	0.328	1.449***
	(1.38)	(0.71)	(0.28)
× STEM h-s perf.	1.344	2.907***	1.798***
	(1.18)	(0.70)	(0.37)
× non-STEM h-s perf.	0.303	-1.531**	-0.868**
	(0.97)	(0.54)	(0.34)
× female	-1.254	-1.177*	-1.122***
	(0.71)	(0.54)	(0.34)
Field: Soc. Sc.	0.136	0.660*	0.637**
	(0.46)	(0.29)	(0.23)
× STEM h-s perf.	1.216*	1.724***	1.476***
	(0.54)	(0.35)	(0.24)
× non-STEM h-s perf.	-0.824	-1.139*	-0.796**
	(0.63)	(0.48)	(0.29)
× female	-1.079	-0.744*	-0.476
	(0.57)	(0.37)	(0.26)
Field: Eco/Mgmt	3.081***	3.355***	3.513***
	(0.34)	(0.21)	(0.17)
× STEM h-s perf.	1.001***	1.618***	1.413***
	(0.25)	(0.21)	(0.15)
× non-STEM h-s perf.	-0.909*	-1.276***	-1.110***
	(0.36)	(0.24)	(0.16)
× female	-0.450	-0.613**	-0.706***
	(0.36)	(0.23)	(0.18)
Field: Law	1.486**	1.922***	2.286***
	(0.47)	(0.34)	(0.22)
× STEM h-s perf.	0.825	0.907	0.815**
	(0.61)	(0.47)	(0.28)
× non-STEM h-s perf.	-0.708	-0.855**	-0.785***
	(0.48)	(0.29)	(0.22)
× female	-0.198	-0.430	-0.525*
	(0.53)	(0.39)	(0.26)

Utility parameter estimates – truthful samples (3/3)

	(1)	(2)	(3)
Field: Phys./Chem./Engin.	3.559***	3.777***	3.933***
	(0.35)	(0.21)	(0.16)
× STEM h-s perf.	1.149***	1.710***	1.489***
	(0.24)	(0.21)	(0.14)
× non-STEM h-s perf.	-1.289***	-1.662***	-1.537***
	(0.35)	(0.23)	(0.15)
× female	-0.816*	-0.875***	-1.000***
	(0.38)	(0.22)	(0.17)
Field: Health/Life Sc.	3.493***	3.721***	3.633***
	(0.37)	(0.22)	(0.17)
× STEM h-s perf.	1.010***	1.558***	1.411***
	(0.24)	(0.21)	(0.15)
× non-STEM h-s perf.	-1.389***	-1.590***	-1.370***
	(0.36)	(0.24)	(0.16)
× female	0.074	-0.026	-0.091
	(0.40)	(0.24)	(0.19)
Field: Earth Sc.	2.049***	2.094***	2.079***
	(0.35)	(0.21)	(0.17)
× STEM h-s perf.	-0.207	0.577**	0.404**
	(0.26)	(0.22)	(0.15)
× non-STEM h-s perf.	-1.239***	-1.640***	-1.547***
	(0.36)	(0.23)	(0.15)
× female	-0.391	-0.425	-0.436*
	(0.40)	(0.23)	(0.18)
Field: Math/Comp.Sci.	3.871***	4.029***	4.181***
	(0.34)	(0.21)	(0.17)
× STEM h-s perf.	1.205***	1.901***	1.559***
	(0.25)	(0.21)	(0.14)
× non-STEM h-s perf.	-1.307***	-1.726***	-1.523***
	(0.36)	(0.23)	(0.15)
× female	-0.711	-0.621**	-0.810***
	(0.37)	(0.23)	(0.18)

A.3.3 Empirical validation of the bandwidth choice: supplementary figures

Figure A.2 is analogous to Figure 2.2 in Section A.3.3. The right (resp. left) panel considers the programs listed by the ten students of Group 2 (resp. Group 3), and shows the frequency at which these programs are listed by other Group 2 (resp. Group 3) students as a function of their priority ranking.



Legend: This graph on the right (resp. left) panel considers the programs listed the first ten students at the top of Group 2 (resp. Group 3), and shows the frequency at which these programs are listed by Group 2 (resp. Group 3) students as a function of students' priority. Programs are represented on the x-axis, priority on the y-axis. A dot in position (a, b) in the graph means that student ranked b in Group 2 (resp. Group 3) included program a in her list. The vertical line at rank 200 represents the limit of the estimation bandwidth —students with priority rank lower than 200 are included in the estimation sample.

FIGURE A.2: Persistence of the top-ranked students' listed choices over the priority ranking –Groups 2 and 3

A.3.4 An alternative estimation strategy

Proposition 1(b) (Haeringer and Klijn (2009)) in Section 2.2 shows that while students may not report their *most-preferred* programs, they always report programs in decreasing order of preference. That is, while I identify students' preferences from the choices made by a strict subset of students, all application lists reveal a *partial* preference ranking.

Non-truthful students' application lists generate moment *inequalities*, while truthful students' lists imply moment *equalities*. Indeed, if student i is truthful, then the likelihood of observing her list $\mathcal{L}_i$ in the data corresponds to the likelihood of the listed programs being her most-preferred programs:

$$\Pr\Big(\mathcal{L}_i = \{\mathcal{L}_i(1), \mathcal{L}_i(2), \ldots, \mathcal{L}_i(M_i)\}\Big)$$
$$= \Pr\Big(u_i(\mathcal{L}_i(1)) > u_i(\mathcal{L}_i(2)) > \cdots > \mathcal{L}_i(M_i) > u_i(j),\ \forall\, j \in \mathcal{C}_i \backslash \mathcal{L}_i\Big),$$

where $\mathcal{L}_i$, M_i and $\mathcal{C}_i$ denote i's application list, the length of i's application list, and i's choice set, respectively. By contrast, if student i is (a priori) not truthful, then observing her list $\mathcal{L}_i$ in the data only implies an order on the utilities of listed programs:

$$\Pr\Big(\mathcal{L}_i = \{\mathcal{L}_i(1), \mathcal{L}_i(2), \ldots, \mathcal{L}_i(M_i)\}\Big) \leqslant \Pr\Big(u_i(\mathcal{L}_i(1)) > u_i(\mathcal{L}_i(2)) > \cdots > \mathcal{L}_i(M_i)\Big).$$

Extracting the information included in lists of non-truthful students requires an estimation method that allows for the use of both moment equalities and inequalities (e.g., Andrews and Shi (2013); Fack et al. (2015)[1]), which I do not pursue in this paper.

A.4 Students' expectations about their admission chances

This appendix provides supporting materials and information to Section 2.5.

Parts A.4.1 and A.4.2 document the estimation of expectations about admission

[1] Fack et al. (2015) implement such a method to recover the preferences for high schools of middle-schoolers in Paris. Their identifying moment equalities differ from those I describe, though. Identification of students' preferences in their analysis relies on the assumption that the match realized in their data is *stable*. They also implement a partial identification approach, using only moments inequalities, without assuming stability.

chances in a benchmark setting where students know their true probability of admission to the different programs. Turning to the framework of the main text, in which students may not know their true admission probabilities, Part A.4.3 shows the parameters characterizing each sophisticated subtype used in the estimated model. As a complement to Section 2.5.2, Part A.4.4 illustrates the identifying variation grounding the estimation of types shares.

A.4.1 True-admission-probability benchmark

As a standard benchmark, I assume that, given the distribution of preferences, students are all able to form expectations that coincide with their true admission chances (are '*perfectly rational*'). In this setting, perfect rationality and the distribution of preferences are common knowledge.

Recovering true admission probabilities

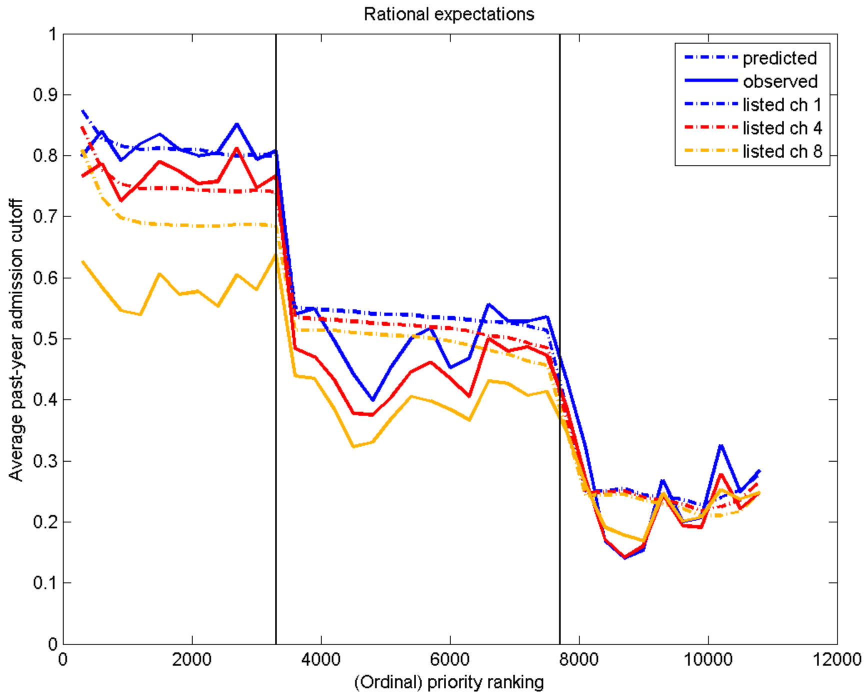
Students having the same priority at all programs allows me compute numerically the joint distribution of each students' admission chances to all programs, in a simple way, given utility parameters and the distribution of preference unobservables.

- *Step 1:* I estimate the distribution of seats left after Student 1's assignment. Given a draw of her unobservable preference term, Student 1 truthfully lists her most-preferred program on top of her application list and get assigned to it. Simulating over her unobservable preference term, I can recover the distribution of her assignment –hence the joint distribution of seats left after her assignment. This distribution gives, for every program, the probability of an available vacancy for Student 2, that is, Student 2's expected admission chances in the rational-expectations benchmark.

- *Step k, $(1 < k \leqslant N)$:* I estimate the distribution of seats left after Student k's assignment. Student k solves Problem (2.1) given her preferences, and her expectations about her admission chances –recovered in Step (k-1). For any draw of Student k's unobservables, I solve Problem (2.1) and deduce Student k's assignment via the DA and the assignment of previous students. Simulating over her unobservable preference term, I can recover the distribution of her assignment –hence the joint distribution of seats left after her assignment. This distribution yields Student $k + 1$'s expected admission chances in the rational-expectations benchmark.

The simulation of optimal application lists is a computational challenge. When the choice set is large, the simulation of application lists of even moderate size is demanding. For instance, in the setting of this paper, finding *one* individual's expected-utility-maximizing ordered list of up to 10 elements among 600 requires evaluating the expected utility function at more than 10^{20} points, and finding the maximum. In practice, to ease computation, I do not solve the optimization Problem (2.1). Rather, I approximate the solution using the Marginal Improvement Algorithm (MIA) proposed by Chade and Smith (2006).[2] The MIA starts by first selecting the application list of size 1 (that is, the alternative) with highest expected utility. It then proceeds to finding the best marginal improvement to that list. That is, it selects the alternative that forms, together with the first pick, the application list of size 2 with highest expected utility (among all the lists containing the first pick). This iterative process continues until the desired list size is reached. A detailed description of the MIA is provided in Appendix Section A.4.2.

[2] When one's eligibility chances are independent across programs, the MIA yields the actual solution of Problem (2.1) (Chade and Smith (2006)). When these eligibility chances are not independent, however, this result is not guaranteed (Ajayi and Sidibé (2016)).



This graph shows the selectivity level (in terms of past-year admission cutoff) of students' choices as a function of their priority ranking. For clarity, it focused on students' first, fourth, and eighth-listed choices. Solid lines represent choices observed in the data; dotted lines represent choices predicted under the rational-expectations benchmark, given utility parameter estimates from Section 2.4.

Figure A.3: Rational-expectations benchmark –Selectivity level of predicted vs. observed choices

Figure A.3 plots the selectivity level of students' choices as a function of their priority ranking. For clarity, it focused on students' first, fourth, and eighth-listed choices. Solid lines represent choices observed in the data; dotted lines represent choices predicted under the rational-expectations benchmark, given utility parameter estimates from Section 2.4. The graph suggests that assuming perfect rationality of students imperfectly captures the variation in the data in two ways. It overshoots the selectivity level of some students' listed choices, and does not reproduce the diversification of students' application portfolios in terms of selectivity levels. In a rational equilibrium, each student understands that when the number of seats in each program is fixed, given the number of students assigned before her, there is a negative correlation between the number of seats remaining available when her turn in the algorithm comes in programs with similar characteristics.[3] In other words, conditional on being rejected from an higher-listed program, they know they have an increased chance to be admitted to a program with similar characteristics if they rank such program lower in their list. Hence the similarity in programs' characteristics across a rational student's listed choices. The diversification of students' portfolios (in terms of selectivity level) in the observed data suggests that students do not fully account for this negative correlation.

Note, from the procedure described in A.4.1, that given utility parameter estimates, predicting choices under the rational-expectations framework does not require the identification of any new parameter. The admission chances expected by perfectly rational students are fully determined once the utility parameters and the distribution of preference unobservables are known. In Section 2.4, preferences were recovered without taking any stand on expectations, and using a strict subset of the students. In the previous paragraph, other students' observed lists were used to assess the ability of the rational-expectations framework to reproduce patterns in the data (Figure A.3). In the next subsection, I use still-unexploited identifying variation in these lists to estimate an alternative framework of expectations formation.

[3] This is mechanical. Suppose there are only two programs A and B, and that they have identical characteristics. Fix students' tastes and the number of students to be assigned before student i in the algorithm. Statistically, and conditional on the characteristics of A and B, when none of the programs are full, the students assigned to A are the ones whose unobservable utility draw for A is larger than for B. The number of students to be assigned before student i in the algorithm being fixed, the larger the number of students assigned to A before i's turn, the smaller the number of students assigned to B before i's turn.

A.4.2 Marginal Improvement Algorithm (Chade and Smith (2006))

Marginal Improvement Algorithm

Step 0: Start with the empty list: $\mathcal{L}_i^{(0)} = \varnothing$.

Discard from choice set all alternative with lower flow utility than the outside option.

Step 1: Select the program with highest expected utility: $\mathcal{L}_i^{(1)} = \{s_1\}$.
Step k ($2 \leqslant k \leqslant 10$)**:** Select the best complement to the current list $\mathcal{L}_i^{(k-1)}$, i.e. solve:

$$\max_{s \in \mathcal{J} \setminus \mathcal{L}_i^{(k-1)}} \quad EU(\mathcal{L}_i') \qquad s.t. \quad \mathcal{L}_i' = \mathcal{O}_i\left(\mathcal{L}_i^{(k-1)} \cup \{s\}\right)$$

where

- $\mathcal{O}_i$ arranges the elements of $\mathcal{L}_i$ in decreasing order of flow utility for i
- $EU_i(\mathcal{L}_i^{(k)}) = \pi_{i,\ell_1} \cdot u_{i,\ell_1} + \pi_{i,\ell_2|\ell_1} \cdot u_{i,\ell_2} + \cdots + \pi_{i,\ell_k|\ell_1,\ell_2,\ldots,\ell_{k-1}} \cdot u_{i,\ell_k}$

A.4.3 Types specification

In Section 2.5, expectations-formation types are specified as AR(1) processes, the coefficients of which are estimated by MLE from 2009-2010 data on marginal admission scores. Types differ from one another in the level of observable heterogeneity allowed in the AR(1) specification:

$$\text{cutoff}_{j,2010} = a_j + b_j \times \text{cutoff}_{j,2009} + \eta_j \quad \text{with } \eta_j \sim N(0, \sigma_j^2).$$

Table A.5 shows estimated coefficients for the different specifications considered.

Table A.5: Estimated AR(1) parameters for marginal admission scores (1/3)

NO HETEROGENEITY					LOG-LIK.: 0.54
	All				
cst.	-0.19				
	0.02				
slope	0.82				
	0.02				
σ	0.42				
	0.02				
Log-lik.	0.54				
Obs.	616				
HETEROGENEITY BY FIELD OF STUDY					Log-lik.: 0.46
	STEM	non-STEM			
cst.	-0.09	-0.32			
	0.02	0.04			
slope	0.89	0.73			
	0.02	0.04			
σ	0.31	0.5			
	0.02	0.03			
HETEROGENEITY BY 2009 FILLING STATUS					Log-lik.: 0.53
	full	not full			
cst.	-0.21	0.02			
	0.02	0.04			
slope	0.82	0.91			
	0.03	0.02			
σ	0.4	0.45			
	0.02	0.05			
HETEROGENEITY BY FIELD $\times$ 2009 FILLING STATUS					Log-lik.: 0.44
	STEM		non-STEM		
	full	not full	full	not full	
cst.	-0.11	0.09	-0.32	-0.63	
	0.02	0.03	0.04	0.42	
slope	0.91	0.92	0.72	0.56	
	0.03	0.01	0.04	0.24	
σ	0.3	0.28	0.48	0.57	
	0.02	0.03	0.03	0.07	
Obs.	616				

Estimated AR(1) parameters for marginal admission scores (2/3)

Heterogeneity by selectivity level						Log-lik.: 0.51
						≤ 5th
5–25th	25–50th	50–75th	75–95th	≥ 95th		
cst.	1.73	-0.5	0	-0.31	-0.25	-0.43
	1.57	0.43	0.19	0.07	0.06	0.15
slope	1.66	0.58	1.09	0.65	0.89	1.17
	0.82	0.27	0.19	0.18	0.15	0.07
σ	0.44	0.41	0.39	0.41	0.39	0.38
	0.09	0.05	0.04	0.03	0.04	0.09
Obs.	616					

Estimated AR(1) parameters for marginal admission scores (3/3)

Heterogeneity by field × selectivity level												Log-lik.: 0.36
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
cst.	-0.76	-0.16	-0.46	-0.27	-0.25	0.08	1.64	-0.95	0.68	-0.37	-0.26	1.86
	2.03	0.26	0.14	0.08	0.04	0.09	1.8	0.83	0.29	0.12	0.1	0.97
slope	0.3	0.8	0.59	0.46	1.21	0.95	1.64	0.3	1.88	0.77	0.67	-1.48
	1.08	0.17	0.15	0.2	0.13	0.05	0.93	0.53	0.33	0.31	0.23	1.02
σ	0.24	0.26	0.31	0.3	0.26	0.11	0.51	0.57	0.45	0.47	0.46	0.25
	0.05	0.03	0.03	0.04	0.03	0.02	0.15	0.08	0.07	0.05	0.07	0.05
Obs.	616											

A.4.4 Identifying variation

Figure A.4 illustrates how, given preferences, the likelihood of observing one's actual choices differs under alternative assumptions about expectations formation process and level of sophistication. This figure illustrates the variation in the data allowing me to characterize students' expectations. It plots, as a function of students' priority ranking, the likelihood (given preferences) of observing the characteristics of one's actual choices under alternative assumptions about the expectations-formation process. Specifically, it focuses on the selectivity level (in terms of past-year admission cutoff) of students' listed programs under the eight distinct scenarios of expectations-formation considered in Section 2.5.2. For the sake of space, I only show plots for students' first, third, sixth, and ninth listed choices –plots for the second, fourth, fifth, seventh, eighth, and tenth listed choices are similar. Lists are simulated assuming that unsophisticated students report their ten most preferred programs among those that have not been publicly declared to be full. Students with any given AR(1) type report the expected-utility-maximizing list, and derive their expectations assuming marginal admission scores follow, from one year to the next, the given AR(1) process.

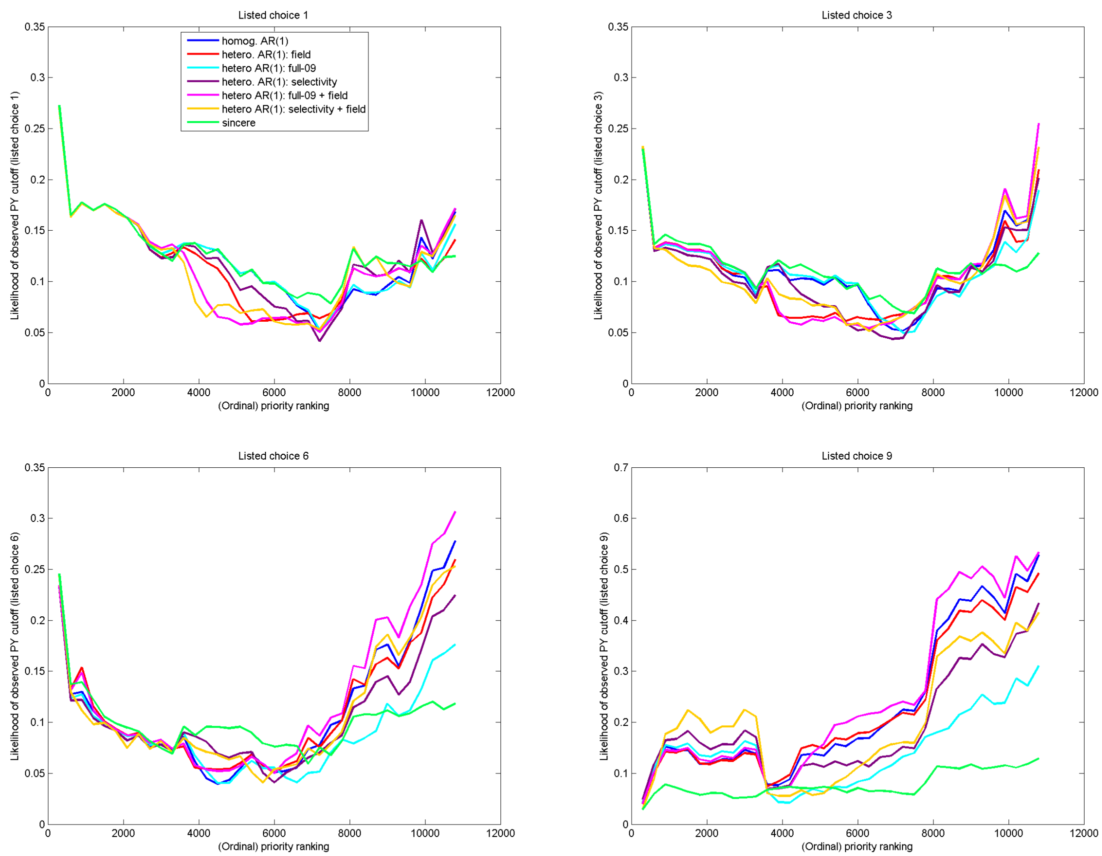


FIGURE A.4: Density of listed-choice characteristics (past-year cutoff) under alternative expectations formation assumptions

A.5 Counterfactual analysis

This appendix provides supplemental figures and tables to Section 2.6.

Part A.5.1 provides an illustration to results in Section 2.6.1. Part A.5.2 discuss gains from information in a setting where students know their true probabilities of admission to each programs.

A.5.1 Supplemental figures

As a complement to Figure 2.10 in Section 2.6.1, Figure A.5 plots the distribution of indirect utility changes (relative to the single-phase restricted-list DA) within each assignment-status-pair group.

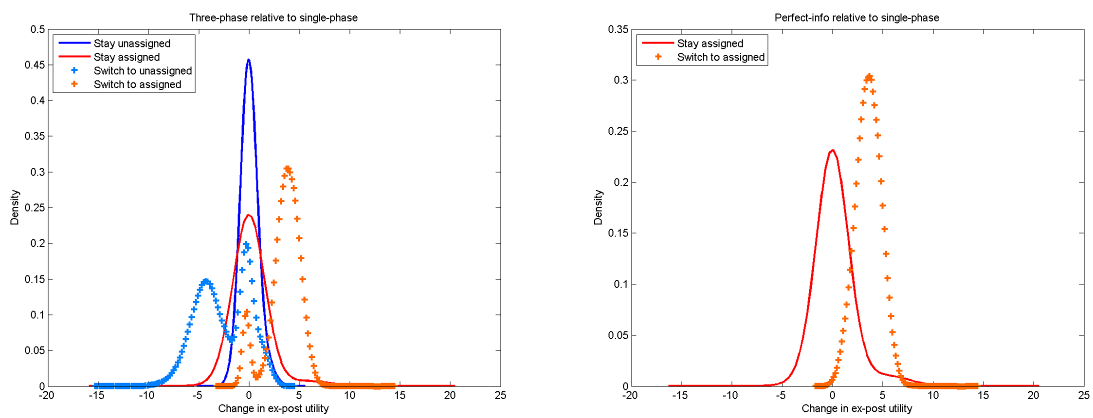


FIGURE A.5: Distribution of changes in expected indirect utility by assignment status pair

A.5.2 True-admission-probabilities benchmark

Table A.6 is the true-admission-probabilities analogue of Figure 2.4: it shows the difference in expected average student welfare between scenarios with and without information revelation –in a setting in which students form expectations about their admission chances that coincide with their admission probabilities (see Appendix Section A.4.1). In this setting, the loss generated by the implementation of the standard (single-phase) restricted-list DA, relative to perfect-information benchmark, is very small. Under the perfect-information benchmark, the average indirect utility is higher than in the single-phase DA by an equivalent of a 0.40km-reduction in distance traveled –a difference 100 times smaller than the one shown in Figure 2.4. This suggests that, rather than the sole incompleteness of information about which seat are available for them, it is students' inability to form accurate expectations about their admission chances that is responsible for most of the welfare loss induced by the implementation of a single-phase restricted-list DA in an incomplete information setting.

Table A.6: Rational expectations benchmark –Change in expected average student welfare relative to single-phase implementation of the restricted-list DA

	Two-phase	Three-phase	Four-phase	Five-phase	Perfect info
equiv. km	0.04	0.22	0.1	0.11	0.39

Appendix B

Appendices to Chapter 3

B.1 Institutional background: Application regions

Application to selective high schools depends on the school district the student's middle school is in. Over the period considered (2006–08), there were twenty-six school districts in Tunisia. These school districts approximately correspond to the Tunisian administrative provinces[1] (see Figure ?? in Section 1.3.1); the main difference being that Tunis and Sfax are both divided into two school districts[2]. In 2006–08, there were twelve selective high schools, each in a distinct district; and two cases. In 18 (19 in 2008) districts, students could not choose the selective school they would apply to, as their middle school was associated to only one selective high school. This case corresponds to three situations. First, schools districts in which there is one selective high school and where students have no choice but apply to that one. This is the case of Nabeul, Sousse, Kairouan, Sfax(1), Gabes, Medenine, Gafsa, Le Kef and Bizerte; as well as Monastir in 2008. Second, school districts in which there is no selective high school and that are paired with one neighbor district in which there is one elite high school, so that all students in the former district have no choice but apply to the selective high school in that latter district. This is the

[1] Administrative provinces are called *gouvernorats*; and school districts *délégations*

[2] Districts named Tunis(1) and Tunis(2); Sfax(1) and Sfax(2)

case of Zaghouan (to Nabeul), Jandouba (to Le Kef), Sidi Bouzid (to Gafsa), Tozeur (to Gafsa), Kebili (to Gabes), Tataouine (to Medenine), and Sfax(2) (to Sfax(1)). Third, school districts in which there is no selective high school but where not all students can apply to the same neighbor-district selective school. This is the case of Siliana (where an area is paired with Le Kef, and another with Kairouan), and Kasserine (where an area is paired with Le Kef, and another with Gafsa).

In the other 8 (7 in 2008) districts, students had a choice of several selective schools, and could formulate an ordered preference list. This is typically the case in relatively small and densely populated areas, and corresponds to two situations. First, in districts with no selective school and that are geographically close to more than one neighboring selective school, students are allowed to apply to all these schools. This is the case of Tunis(2), Ben Arous, Manouba (where students can apply to Tunis(1) and L'Ariana), Beja (where students can apply to Tunis(1), L'Ariana and Le Kef, as well as Bizerte in 2008), and Mahdia (where students can apply to Sfax(1) and Sousse, as well as Monastir in 2008). Second, in dense districts with a selective school but that are also geographically close to the selective schools of neighboring districts, students can apply to their own-district as well as to neighor-district selective high schools. This is the case of Tunis(1), L'Ariana (where students can apply to both Tunis(1) and L'Ariana), and, in 2006–07, Monastir (where students can apply to Monastir and Sousse).

B.2 Data: Population-level descriptive statistics

Tables B.1 and B.2 describe the applicants and admitted populations; as a benchmark, descriptive statistics are also provided for the whole population of end-of-middle-school exam takers.

Table B.1: Descriptive statistics: full sample (Part1)

	2006	2007	2008	Total
Take end-of-middle-school exam	51,026	51,116	49,259	151,401
% female	57.45	56.34	56.63	56.81
Apply to *lycée pilote*	23,292	26,277	31,055	80,624
% female	57.44	56.06	56.48	56.62
Admitted to *lycée pilote*	2,261	2,017	2,367	6,645
% female	67.54	66.48	66.88	66.98
Pass end-of-middle-school exam	32,658	23,282	30,153	85,993
% female	60.76	60.81	60.90	60.82
Middle schools				
All	995	1,003	1,016	1,136
Applicants	817	841	842	911
Admitted	481	485	487	656
High schools (in 4)				
All	668	670	681	771
Applicants	538	562	606	664
Admitted	235	225	231	356
Middle schools feeding in a high school				
Non *pilote*	13.0	12.8	13.5	13.0
	(10.0)	(9.0)	(10.2)	(9.8)
Pilote	49.4	48.9	51.7	50.0
	(31.2)	(30.0)	(33.8)	(31.8)

Figures in parentheses are standard deviations, they are reported below sample means.

Gender is the only demographic characteristic available for all applicants[3]. 57% of the students in the sample are girls. With respect to this dimension, the sample is balanced as compared to the whole population of end-of-middle-school exam takers[4]. In contrast, girls make up 60% of the students who pass the end-of-middle-school exam, and 67% of those admitted into a selective high school.

According to several measures, admitted students perform on average better in high school than the average applicant, who, in turn, performs better than the average end-of-middle-school exam-taker. 89% of admitted students take the end-of-high-school exam four years after taking the end-of-middle-school exam, that is, without

[3] Father's occupation is available in the data set on end-of-high-school exam.

[4] The share of girls in the whole population of 9th graders is 53% in 2006.

repeating any grade in high school. This number drops to 73% for applicants and 63% for end-of-middle-school exam-takers. Looking at passing rates within five years enables applicants (81%) and end-of-middle-school exam-takers (74%) to catch up, as the rate for admitted students stagnates at 89.5%. Among end-of-middle-school exam-takers and applicants, girls are significantly more likely to take the end-of-high-school exam than boys. For admitted students no significant difference is observed between girls' and boys' exam-taking rates. The passing rate, conditional on taking the end-of-high-school exam in four years, is close to 1 (99.1%) for admitted students; against 85% for applicants and 81.7% for end-of-middle-school exam-takers. Again, girls in the total sample and the group of applicants are more likely to pass the exam in four years than boys. Exam scores, conditional on taking the exam in four years, confirm the trend: on average, admitted students score at 15.9 (out of 20), while applicants and middle school exam takers score at 11.8 and 11.3 respectively. As a benchmark, 10 is the passing grade at the exam; 16 is the *cum laude* threshold.

Table B.2: Descriptive statistics: full sample (Part2)

	Boys	Girls	Diff	Total
Compliance rates				
% Admitted taking Bac in *lycée pilote* (in 4)	74.00	69.47	0.05***	70.97
% Non-admitted taking Bac in *lycée pilote* (in 4)	0.33	0.33	0.00	0.33
Taking Bac in lycée pilote *(in 4)*				
% Were admitted	92.65	93.95	-0.01*	93.50
% Were not admitted	7.35	6.05	0.01*	6.50
Take high-school exam in 4				
% All middle-school exam	56.69	67.30	-0.106***	62.72
% Applicants	68.78	75.95	-0.072***	72.84
% Admitted	89.24	88.90	0.003	89.03
Take high-school exam w/in 5				
% All middle-school exam	69.86	76.89	-0.070***	73.85
% Applicants	79.57	82.31	-0.27***	81.12
% Admitted	89.88	89.32	0.006	89.51
Take high-school exam w/in 6				
% All middle-school exam	73.37	78.93	-0.056***	76.53
% Applicants	82.20	83.53	-0.013***	82.95
% Admitted	89.88	89.35	0.005	89.52
Passing rate upon taking exam in 4				
% All middle-school exam	79.70	83.02	-0.033***	81.73
% Applicants	83.33	86.17	-0.028***	85.01
% Admitted	99.03	99.19	-0.002	99.14
Score upon taking exam in 4				
All middle-school exam	11.24	11.30	-0.07***	11.28
	(3.36)	(3.03)		(3.16)
Applicants	11.76	11.84	-0.08***	11.81
	(3.29)	(3.12)		(3.19)
Admitted	16.21	15.92	0.29***	16.02
	(2.32)	(2.32)		(2.23)
Math or Science high-school major				
% All middle-school exam	46.09	62.16	-0.16***	55.51
% Applicants	56.26	73.69	-0.17***	66.19
% Admitted	89.71	96.61	-0.07***	96.61

Figures in parentheses are standard deviations, they are reported below sample means. The column 'Diff' presents the difference between boys and girls and its significance: *significant at the 10% level; **significant at the 5% level;***significant at the 1% level.

Compliance to the admission decision is far from being perfect. 71% of the students admitted to a selective school and 0.3% of non-admitted applicants are observed to be enrolled in a selective school at the time they take the end-of-high-school exam. From another perspective, 93.5% of the students enrolled in a selective school at the time they take the end-of-high-school exam were indeed admitted to a selective school upon middle school graduation while 6.5% were not. Imperfect compliance will be accounted for in the estimation strategy.

B.3 Evidence of the sharpness and validity of the RD design

B.3.1 Admission cutoffs & sharpness of the design

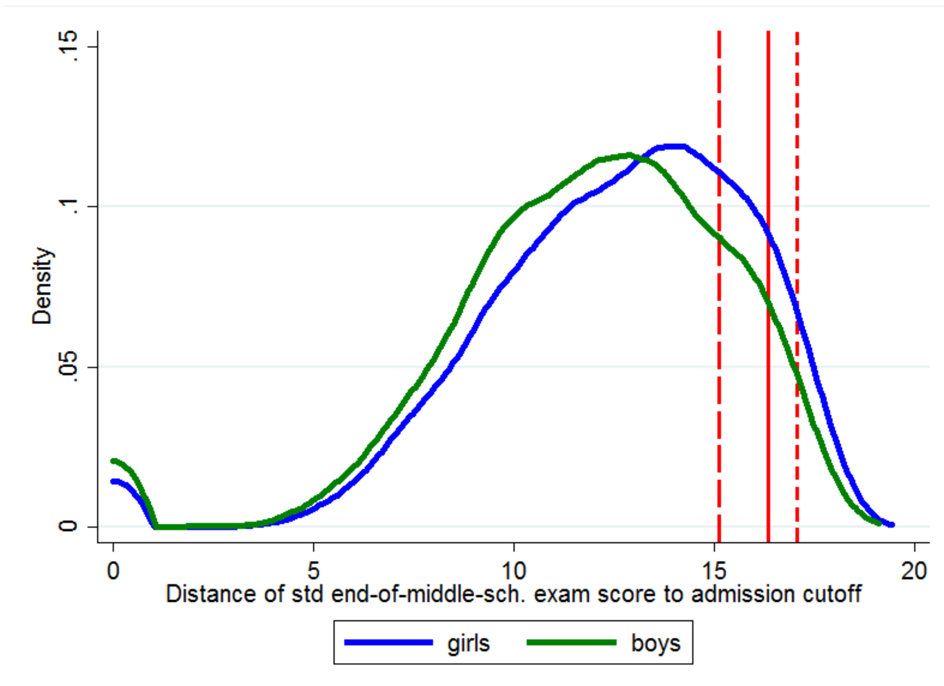
Each year, the distribution of end-of-middle-school exam scores is standardized such that it has mean 0 and standard deviation 1 in our sample. For each school, the admission cutoff is the (standardized) exam score of the student admitted last, that is, the lowest exam score among students admitted to the school. Table B.3 gives the cutoff for each school each year, thereby describing the relative selectivity of the Tunisian *lycées pilotes*. *Lycées pilotes* in big coastal cities (Tunis, L'Ariana, Sfax and Sousse) are the more selective ones, probably due to them drawing from a relatively larger pool of applicants.

Table B.3: Admission cutoffs

Lycée Pilote	2006	2007	2008
Tunis	1.2941	1.4069	1.4011
L'Ariana	1.3828	1.5076	1.4799
Bizerte	1.0361	1.3230	1.2676
Le Kef	1.1340	1.3230	1.2754
Gafsa	1.1340	1.2057	1.1496
Medenine	0.8320	1.2057	0.9139
Gabes	1.0006	1.2140	0.9687
Sfax	1.2761	1.5076	1.3855
Kairouan	0.9386	1.2057	1.0397
Monastir	1.1962	1.3063	1.1888
Sousse	1.2761	1.4488	1.3068
Nabeul	1.1783	1.3734	1.3775

Admission thresholds by year and *lycées pilote*, as measured by standardized exam scores.

Figure B.1 illustrates the location of various admission cutoffs relative to the (standardized) distribution of exam scores. The bell curves are the density plots for girls and boys while the vertical lines mark respectively the lowest (.83), median (1.18) and highest (1.38) admission thresholds for 2006 as measured on the total sample. The figure also makes pictures the fact that girls in our sample are better-achievers than boys, as the pdf of their exam scores is shifted to the right as compared to boys'.



Kernel density estimation of the probability density functions of girls' and boys' end-of-middle-school exam standardized scores. The red vertical lines mark, respectively, the lowest, median and highest admission thresholds for 2006.

FIGURE B.1: Admission cutoffs & middle school exam score density (2006)

Figure B.2 illustrates the sharpness of the present discontinuity design. The probability of admission to a selective school is plotted against the distance of the student's score to her relevant cutoff; the probability jumps, from 0 to 1, at distance 0, that is, as the score equals the cutoff.

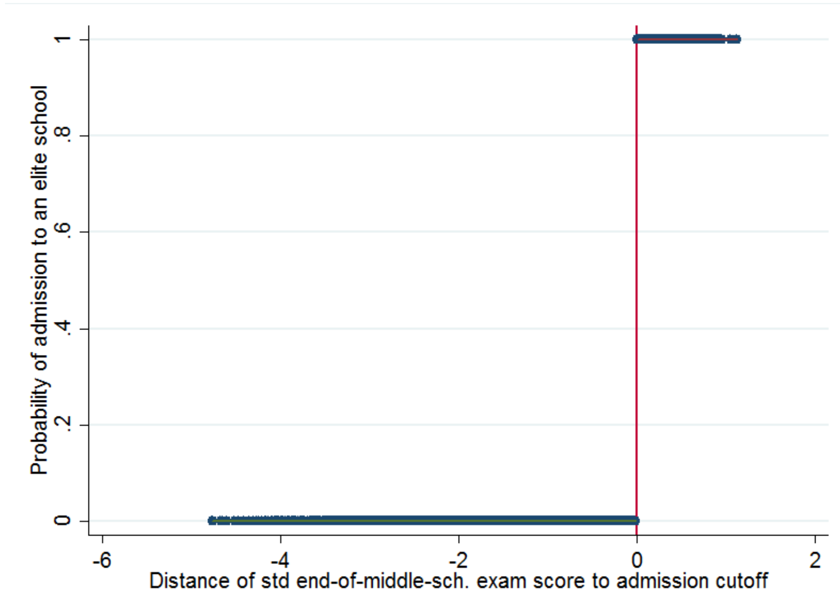


FIGURE B.2: Probability of admission vs. distance of student's score to the cutoff

B.3.2 Validity of the design

We provide some visual evidence of the validity of the design. In particular, we follow McCrary McCrary (2008) and test the existence of a discontinuity of the density of the forcing variable at the cutoff. The presence of bunching at the right of the cutoff would suggest that students have enough control over their grade at the entry exam to be able to score just above the cutoff, ensuring themselves admission. This would invalidate the RDD. Figure B.3 illustrates the test. Formal estimation of the discontinuity yields a jump of .0549 (.0278), which allows us to reject the existence of a significant discontinuity at the cutoff. This is not surprising given the institutional setting. The number of spots in selective high schools is decided by the Administration prior to the students taking the end-of-middle-school exam. The cutoffs being endogenously determined, as a result of students' performances, they are not known to applicants at the time of taking the entry exam. It therefore seems really unlikely that test-takers are able to precisely manipulate their running variable to guarantee themselves admission[5].

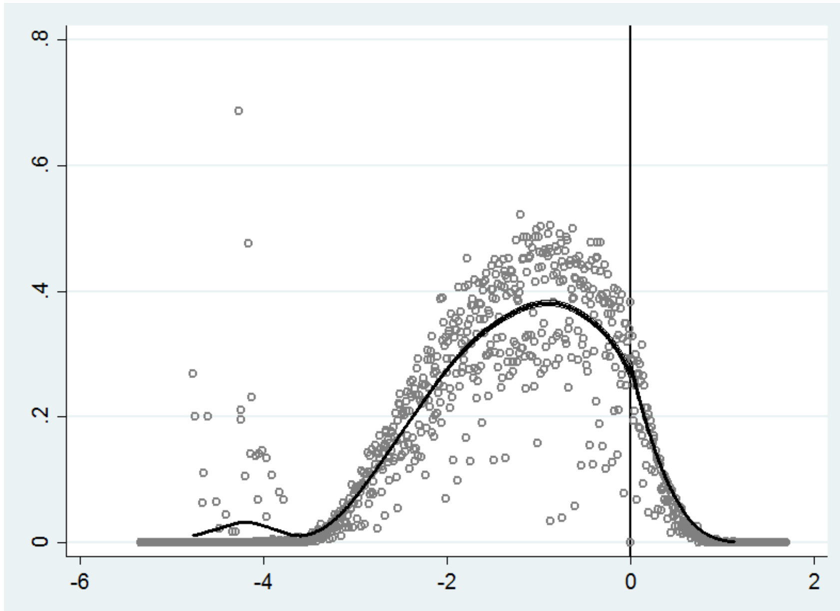


FIGURE B.3: Density of end-of-middle-school exam scores

Another common visual test of the validity of the design is to plot students (pre-treatment) individual characteristics as a function of the running variable. A jump at the cutoff could suggest the presence of confounding factors. Students average grade upon middle-school graduation (equivalent of middle school GPA) is the only pre-treatment individual characteristic contained in our data set. Grades are not standardized across schools (or even within schools, across classes), so individual values are hardly comparable. Due to this concern, we do not use this variable in the formal analysis. However, in Figure B.4, we present a plot of middle-school GPA vs. distance (of the exam grade) to the cutoff as a simple visual hint.

[5] Lee and Lemieux **?** distinguish between precise and imprecise control over the running variable. Imprecise control does not mean total absence of control. They define (p.295): "individuals have imprecise control over T when conditional on [observable characteristics] and [non-observable characteristics affecting the outcome], the density [non-observable characteristics affecting the running variable](and thus [the running variable]) is continuous". For instance a student who, "through her effort, [...] can choose to shift the distribution [of her running variable] to the right [... but is] unable to precisely control [her] score just around the threshold" has imprecise control.

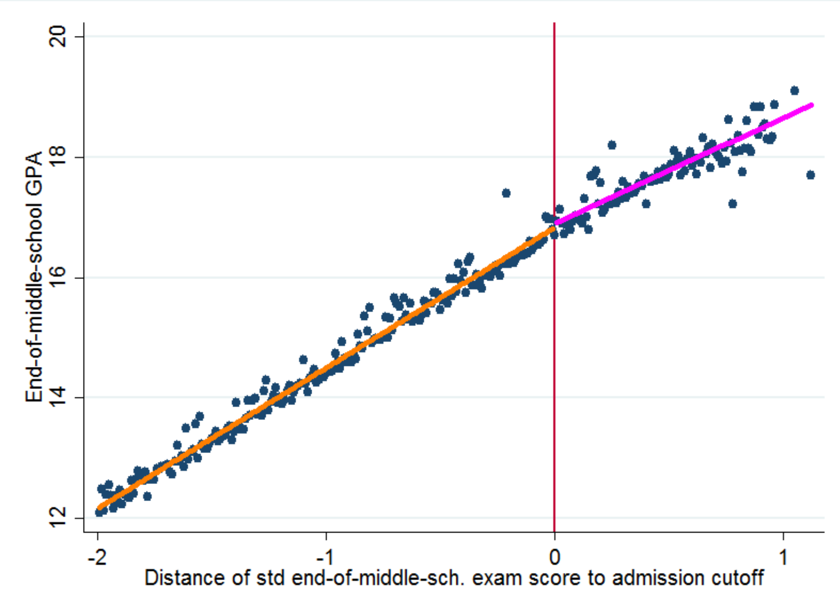


FIGURE B.4: Mean of student's score at end of middle school exam vs. distance of student's score to the cutoff

B.4 Identification

B.4.1 Missing data as non-classical measurement error

From the definition of latent and observed variables, it follows that

$$\Pr(R = 1 \mid T) = \Pr(B = 1,\ M = 1 \mid T) = \Pr(M = 1 \mid T,\ B = 1) \times \Pr(B = 1 \mid T) \tag{B.1}$$

Assuming $M \perp T \mid B$,

$$\underbrace{\Pr(R = 1 \mid T = t)}_{=:p_{Rt}} = \underbrace{\Pr(B = 1 \mid T = t)}_{=:p_{Bt}} \times \underbrace{\Pr(M = 1 \mid B = 1)}_{=:p_M}$$

and (B.1) then rewrites:

$$p_{Rt} = p_{Bt} \cdot p_M = p_{Bt} - (1 - p_M) \cdot p_{Bt}$$

which shows that seeing R as a noisy signal of B frames the problem into one of *non-classical* measurement error: the noise is correlated with the true value.

For $d = 0, 1$ or $t \in supp(T)$, $B \mid (T = t)$ and $R \mid (T = t)$ are Bernoulli random variables with respective parameters p_{Bt} and p_{Rt}. Therefore:

$$(R \mid T = t) = \Pr(R = 1 \mid T = t) = p_{Rt}$$

$$(B \mid T = t) = \Pr(B = 1 \mid T = t) = p_{Bt}$$

Using this notation, the effect τ of treatment on the probability of students to take the end-of-high-school exam (four years after treatment assignment) writes

$$\begin{aligned} \tau &:= \lim_{t \downarrow c}(B \mid T = t) - \lim_{t \uparrow c}(B \mid T = t) \\ &= \lim_{t \downarrow c} p_{Bt} - \lim_{t \uparrow c} p_{Bt} \end{aligned}$$

If B were perfectly observable then τ would identified and a straightforward estimator would be the usual RD estimator:

$$\tilde{\tau} := \widehat{\ }(B \mid c \leqslant T \leqslant c + h_n) - \widehat{\ }(B \mid c - h_n \leqslant T \leqslant c) \tag{B.2}$$

where h_n is the estimation bandwidth (given sample size n), and $\widehat{\ }(B \mid c \leqslant T \leqslant c+h_n)$, $\widehat{\ }(B \mid c-h_n \leqslant T \leqslant c)$ denote (consistent) estimators of the expectations of interest (for instance obtained by local linear regression, see for instance Imbens and Lemieux). B, however, is not observed; only a noisy version of B is available to the econometrician: R. For $t \in supp(T)$

$$\begin{aligned} & (R = 1 \mid T = t) = p_{Rt} = p_{Bt} \cdot p_M = (B = 1 \mid T = t) \cdot p_M \\ \Leftrightarrow \quad & (B = 1 \mid T = t) = (R = 1 \mid T = t) \cdot p_M^{-1} \\ \Rightarrow \quad & \tau = \left(\lim_{t \downarrow c}(R \mid T = t) - \lim_{t \uparrow c}(R \mid T = t) \right) \cdot p_M^{-1} \end{aligned}$$

Treatment effect τ is identified if and only if $p_M = 1$ (no linkage failure), or p_M is known (extent of the linkage failure is perfectly known). In other words, in the presence of linkage failure (of unknown extent), τ is not identified. Estimating τ using R as a proxy for B yields non-consistent estimates. Since $p_M \in (0, 1)$, the presence of an *attenuation bias* is made clear in the following. Consider the following feasible analogue of (B.2) using R instead of B:

$$\widehat{\tau} := \widehat{\ }(R \mid c \leqslant T \leqslant c + h_n) - \widehat{\ }(R \mid c - h_n \leqslant T \leqslant c).$$

Then:

$$\widehat{\tau} \xrightarrow{p} \Big(\lim_{t \downarrow c}(R \mid T = t) - \lim_{t \uparrow c}(R \mid T = t) \Big) = \Big(\lim_{t \downarrow c}(B \mid T = t) - \lim_{t \uparrow c}(B \mid T = t) \Big) \cdot p_M \quad \text{(B.3)}$$

$$= \tau \cdot p_M < \tau \quad \text{(B.4)}$$

Intuitively, and to sum up, linkage failures cause a non classical measurement error because the noise resulting from them is correlated with the true value of the end-of-high-school-exam-taking outcome. Indeed, only students who did take the exam ($B = 1$) can be affected by the error. Students who did take the exam may fail to be matched, in which case their exam-taking record ($R_i = 0$) differ from their true exam-taking value ($B_i = 1$). Students who did not take the exam, however, cannot fail to be unmatched (i.e. matched by mistake), so their exam-taking record ($R_i = 0$) and true exam-taking value ($B_i = 1$) always coincide.

B.4.2 Partial identification under sample selection

Getting (s_0, s_1) as a function of π_{01}

The shares of always-exam-takers within the treatment and control groups $s_{(1)}$ and $s_{(0)}$ are not directly observed since the identity of always-takers among test-takers is not observed. $s_{(1)}$ and $s_{(0)}$ are related to the shares π of potential outcome profiles as follows:

$$s_{(1)} := \Pr(i \in \mathcal{B}11 \mid i \in \mathcal{O}11) = \frac{\pi_{11}}{\pi_{11} + \pi_{10}};$$

$$s_{(0)} := \Pr(i \in \mathcal{B}11 \mid i \in \mathcal{O}01) = \frac{\pi_{11}}{\pi_{11} + \pi_{01}}.$$

Proof

The derivation for $s_{(0)}$ is analogous:

$$
\begin{aligned}
s_{(1)} &:= \Pr(i \in \mathcal{B}11 \mid i \in \mathcal{O}11) \\
&= \frac{\Pr(i \in \mathcal{B}11,\ i \in \mathcal{O}11)}{\Pr(i \in \mathcal{O}11)} \quad \text{by definition of conditional probability} \\
&= \frac{\Pr(i \in \mathcal{B}11,\ D_i = 1)}{\Pr(i \in \mathcal{O}11)} \\
&= \frac{\Pr(i \in \mathcal{B}11)\Pr(D_i = 1)}{\Pr([i \in \mathcal{B}11 \cup i \in \mathcal{B}10], D_i = 1)} \quad \text{by } (B^1, B^0) \perp D \\
&= \frac{\Pr(i \in \mathcal{B}11)\Pr(D_i = 1)}{\Big(\Pr(i \in \mathcal{B}11) + \Pr(i \in \mathcal{B}10)\Big)\Pr(D_i = 1)} \quad \text{by } \mathcal{B}11 \cap \mathcal{B}10 = \varnothing \\
&= \frac{\Pr(i \in \mathcal{B}11)}{\Pr(i \in \mathcal{B}11) + \Pr(i \in \mathcal{B}10)} \\
&= \frac{\pi_{11}}{\pi_{11} + \pi_{10}}
\end{aligned}
$$

Bounding π_{01}

If π shares were known, shares of interest $s_{(1)}$ and $s_{(0)}$ could be straightforwardly deduced. Potential outcome profiles being unobserved, π shares are not directly known though. However, we do know that potential outcome profiles shares are related to observation groups shares as follows:

$$p_{11} = p_D \times \big(\pi_{11} + \pi_{10}\big)$$

$$p_{10} = p_D \times \big(\pi_{00} + \pi_{01}\big)$$

$$p_{00} = (1 - p_D) \times \big(\pi_{00} + \pi_{10}\big)$$

$$p_{01} = (1 - p_D) \times \big(\pi_{11} + \pi_{01}\big)$$

Proof

Derivations are similar for p_{10}, p_{00} and p_{01}:

$$\begin{aligned}
p_{11} &= \Pr(D=1, B=1) = \Pr(D=1, B^1=1) \\
&= \Pr(D=1, B^1=1, B^0=1) + \Pr(D=1, B^1=1, B^0=0) \\
&= \Pr(D=1) \times \Big(\Pr(B^1=1, B^0=1) + \Pr(B^1=1, B^0=0) \Big) \text{ as } (B^1, B^0) \perp D \\
&= p_D \times \big(\pi_{11} + \pi_{10}\big).
\end{aligned}$$

As long as $p_D \in (0,1)$, the four-equation system rewrites:

$$\begin{pmatrix} \frac{p_{11}}{p_D} \\ \frac{p_{10}}{p_D} \\ \frac{p_{00}}{1-p_D} \\ \frac{p_{01}}{1-p_D} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \pi_{11} \\ \pi_{10} \\ \pi_{00} \\ \pi_{01} \end{pmatrix}$$

The (4×4) matrix in this expression is singular, hence π shares are not identified from this system. In addition, we know that $\pi_{11} + \pi_{10} + \pi_{00} + \pi_{01} = 1$, which does not solve the point identification problem of π shares. We can however bound these shares as follows. First, getting all π shares in terms of π_{01}:

$$\pi_{00} = \frac{p_{10}}{p_D} - \pi_{01} \tag{B.5}$$

$$\pi_{11} = \frac{p_{01}}{1-p_D} - \pi_{01} \tag{B.6}$$

$$\pi_{10} = \frac{p_{00}}{1-p_D} - \pi_{00} = \frac{p_{00}}{1-p_D} - \left(\frac{p_{10}}{p_D} - \pi_{01}\right) = \frac{p_{00}}{1-p_D} - \frac{p_{10}}{p_D} + \pi_{01}$$

Now, we can derive the following bounds for π_{01}:

$$\pi_{01} \in \Pi_{01} := \left[\max\left\{0, \frac{p_{01}}{1-p_D} - \frac{p_{11}}{p_D}\right\}; \min\left\{\frac{p_{01}}{1-p_D}, 1 - \frac{p_{11}}{p_D}\right\}\right]$$

Proof

Bounds are derived as follows. First, note that

$$\begin{aligned}
\frac{p_{11}}{p_D} + \frac{p_{10}}{p_D} &= \frac{\Pr(B=1, D=1) + \Pr(B=0, D=1)}{\Pr(D=1)} = 1; \\
\frac{p_{01}}{1-p_D} + \frac{p_{00}}{1-p_D} &= \frac{\Pr(B=1, D=0) + \Pr(B=0, D=0)}{\Pr(D=0)} = 1,
\end{aligned}$$

so that

$$\pi_{10} = 1 - \frac{p_{01}}{1-p_D} - \left(1 - \frac{p_{11}}{p_D}\right) + \pi_{01} = \frac{p_{11}}{p_D} - \frac{p_{01}}{1-p_D} + \pi_{01}. \qquad \text{(B.7)}$$

Now:

$$\begin{aligned} \text{(B.6)} &\Rightarrow 0 \leqslant \pi_{01} \leqslant \frac{p_{01}}{1-p_D} \\ \text{(B.5)} &\Rightarrow 0 \leqslant \pi_{01} \leqslant \frac{p_{10}}{p_D} = 1 - \frac{p_{11}}{p_D} \\ \Rightarrow \quad & 0 \leqslant \pi_{01} \leqslant \min\left\{\frac{p_{01}}{1-p_D}, 1 - \frac{p_{11}}{p_D}\right\}; \end{aligned}$$

and

$$\text{(B.7)} \Rightarrow -\left(\frac{p_{11}}{p_D} - \frac{p_{01}}{1-p_D}\right) \leqslant \pi_{01} \leqslant 1 - \left(\frac{p_{11}}{p_D} - \frac{p_{01}}{1-p_D}\right).$$

Noting that

$$\min\left\{\frac{p_{01}}{1-p_D}, 1 - \frac{p_{11}}{p_D}\right\} \leqslant 1 - \left(\frac{p_{11}}{p_D} - \frac{p_{01}}{1-p_D}\right),$$

we get the stated bounds.

Getting TE bounds as a function of (s_0, s_1)

For each value of π_{01} in the identified set Π_{01}, values for the shares of interest $s_{(1)}(\pi_{01})$ and $s_{(0)}(\pi_{01})$ can be computed. In turn, for each such pair $\left(s_{(1)}(\pi_{01}), s_{(0)}(\pi_{01})\right)$, upper and lower bounds

$$\tau_{11+}(\pi_{01}) := \tau_{11+}(s_{(1)}(\pi_{01}), s_{(0)}(\pi_{01}))$$
$$\tau_{11-}(\pi_{01}) := \tau_{11-}(s_{(1)}(\pi_{01}), s_{(0)}(\pi_{01}))$$

can be computed.

Agnostic vs. refined bounds

The bounds constructed using the steps detailed above are agnostic, in the sense that they do not require additional assumptions about the data generating process. These bounds are sharp, as argued by Zhang and Rubin **?**.

Ranked conditional average outcome assumption

The *ranked conditional average outcome* (RCA) assumption states that on average and in each treatment group, always-exam-takers perform better at the exam than non-always exam-takers. Formally:

Assumption RCA. For $d \in \{0,1\}$ and $d' \neq d$,

$$\left[Y_i \mid D_i = d, B_i^d = 1, B_i^{d'} = 1\right] \geqslant \left[Y_i \mid D_i = d, B_i^d = 1, B_i^{d'} = 0\right].$$

Consequence:

$$\begin{aligned}\left[Y_i \mid D_i = d, B_i^d = 1\right] = &\left[Y_i \mid D_i = d, B_i^d = 1, B_i^{d'} = 1\right] \times \Pr\left(B_i^{d'} = 1 \mid B_i^d = 1, D = d\right) \\ &+ \left[Y_i \mid D_i = d, B_i^d = 1, B_i^{d'} = 0\right] \times \Pr\left(B_i^{d'} = 0 \mid B_i^d = 1, D = d\right)\end{aligned}$$

Then, under (RAS)

$$\left[Y_i \mid D_i = d, B_i^d = 1, B_i^{d'} = 1\right] \geqslant \left[Y_i \mid D_i = d, B_i^d = 1\right] \geqslant \left[Y_i \mid D_i = d, B_i^d = 1, B_i^{d'} = 0\right]$$

so $\left[Y_i \mid D_i = d, B_i^d = 1, B_i^{d'} = 1\right]$ is bounded below by $\left[Y_i \mid D_i = d, B_i^d = 1\right]$.

Why does this tighten our bounds? Note that

$$\left[Y_i \mid D_i = d, B_i^d = 1\right] \geqslant \left[Y_i \mid D_i = d, B_i^d = 1, Y_i \leqslant y_{q,d}\right]$$

where $y_{q,d}$ is some quantile of the distribution of $Y_i \mid D_i = d, B_i^d = 1$. In words: the mean of the LHS is taken over the whole distribution, while the mean in the RHS is taken over the distribution resulting from truncating the top values (above the $(y_{q,d})^{th}$ quantile).

When $d = 1$, substituting the LHS for the RHS in the expression of the lower bound for ATE yields lower bound that is weakly larger. When $d = 0$, substituting the LHS for the RHS in the expression of the upper bound for ATE yields lower bound that is weakly smaller.

B.5 Confidence bands for bound estimates

Table B.4: 95%-confidence bands for bound estimates for TEs on exam-taking rate

	Low. B. CI low	Low. B. CI upp	Upp. B. CI low	Upp. B. CI upp
Optimal bandwidth				
Total sample	0.005	0.036	0.006	0.041
Boys	-0.018	0.037	-0.016	0.041
Girls	0.007	0.046	0.008	0.051
Half optimal bandwidth				
Total sample	0.003	0.044	0.003	0.049
Boys	-0.022	0.045	-0.019	0.05
Girls	0.003	0.054	0.004	0.06

The confidence interval is constructed as $[L_{(Y)-}, U_{(Y)+}]$ where $[L_{(Y)-}, L_{(Y)+}]$ and $[U_{(Y)-}, U_{(Y)+}]$ are intervals of $\mathbb{R}$ covering with 95 % probability, respectively, the lower and upper bounds of the TE on outcome Y. 95 % confidence intervals for bounds are estimated by bootstrap (5,000 repetitions). The two rightmost columns give the range of matching probabilities P_M compatible with the data. For each subpopulation, the TE lower (resp. upper) bound is derived assuming the largest (resp. smallest) value of P_M compatible with the data.

Table B.5: 95%-onfidence bands for bounds estimates for post-exam outcomes

	Most conservative				Least Conservative			
	Low. B.	Upp. B.	Low. B.	Upp. B.	Low. B.	Upp. B.	Low. B.	Upp. B.
Boys								
End-of-high-school outcomes								
Exam score	-0.087	0.12	0.47	0.691	0.008	0.235	0.209	0.408
Priority score	-0.096	0.123	0.477	0.703	0.003	0.243	0.214	0.419
College application outcomes (first-ranked choice)								
2009 cutoff	-0.038	0.326	0.09	0.444	-0.007	0.371	0.062	0.419
Log-dist. from h-sch.	-0.407	1.003	-0.119	1.295	-0.352	1.06	-0.171	1.23
College admission outcomes (assigned program)								
2009 cutoff	-0.102	0.282	0.05	0.425	-0.084	0.31	0.004	0.384
Log-dist. from h-sch.	-0.567	0.634	-0.375	0.856	-0.51	0.687	-0.431	0.779
Girls								
End-of-high-school outcomes								
Exam score	-0.147	0.016	0.6	0.759	-0.022	0.147	0.267	0.413
Priority score	-0.16	0.003	0.578	0.734	-0.04	0.129	0.253	0.397
College application outcomes (first-ranked choice)								
2009 cutoff	-0.244	0.061	-0.087	0.211	-0.232	0.072	-0.141	0.16
Log-dist. from h-sch.	-0.441	0.557	-0.174	0.838	-0.373	0.621	-0.276	0.725
College admission outcomes (assigned program)								
2009 cutoff	-0.132	0.187	-0.021	0.303	-0.124	0.198	-0.062	0.246
Log-dist. from h-sch.	-0.334	0.529	-0.028	0.831	-0.256	0.592	-0.131	0.724

For outcome Y, the confidence interval is constructed as $[L_{(Y)-}, U_{(Y)+}]$ where $[L_{(Y)-}, L_{(Y)+}]$ and $[U_{(Y)-}, U_{(Y)+}]$ are intervals of $\mathbb{R}$ covering with 95 % probability, respectively, the lower and upper bounds of the TE on outcome Y. 95 % confidence intervals for bounds are estimated by bootstrap (5,000 repetitions). Bounds for TE on naturally ordered outcomes (exam-passing rate, exam scores, 2009 cutoffs, distances) are derived under the Ranked Average Outcome assumption.

Bibliography

Abdulkadiroğlu, A., Angrist, J., and Pathak, P. (2014), "The elite illusion: Achievement effects at Boston and New York exam schools," *Econometrica*, 82, 137–196.

Abdulkadiroğlu, A. and Sönmez, T. (2003), "School choice: a mechanism design approach," *The American Economic Review*, 93, 729–46.

Abdulkadiroğlu, A., Pathak, P., Roth, A., and Sönmez, T. (2006a), "Changing the Boston school choice mechanism," *NBER Working Paper No. 11965.*

Abdulkadiroğlu, A., Pathak, P., Roth, A., and Sönmez, T. (2006b), "Changing the Boston school choice mechanism: Strategy-proofness as equal access," *Working paper.*

Abdulkadiroğlu, A., Che, Y.-K., and Yasuda, Y. (2015), "Expanding "choice" in school choice," *American Economic Journal: Microeconomics*, 7, 1–42.

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. (2017), "The welfare effects of coordinated assignment: Evidence from the New York City high school match," *American Economic Review*, 107, 3635–89.

Agarwal, N. and Somaini, P. (2014), "Demand analysis using strategic reports: application to a school choice mechanism," *NBER Working Paper No. 20775.*

Ajayi, K. and Sidibé, M. (2016), "An empirical analysis of school choice under uncertainty," *Working paper.*

Ajayi, K. F. (2014), "Does School Quality Improve Student Performance? New Evidence from Ghana," *Unpublished working paper.*

Altonji, J., Blom, E., and Meghir, C. (2012), "Heterogeneity in human capital investments: high school curriculum, college major, and careers," *NBER Working Paper No. 17985.*

Altonji, J., Arcidiacono, P., and Maurel, A. (2015), "The analysis of field choice in college and graduate school: determinants and wage effects," *NBER Working Paper No. 21655.*

Andrews, D. and Shi, X. (2013), "Inference based on conditional moment inequalities," *Econometrica*, 81, 609–22.

Angrist, J., Bettinger, E., and Kremer, M. (2006), "Long-term educational consequences of secondary school vouchers: Evidence from administrative records in Colombia," *The American Economic Review*, 96, 847–862.

Angrist, J. D., Pathak, P. A., and Walters, C. R. (2013), "Explaining charter school effectiveness," *American Economic Journal: Applied Economics*, 5, 1–27.

Angrist, J. D., Cohodes, S. R., Dynarski, S. M., Pathak, P. A., and Walters, C. R. (2016), "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice," *Journal of Labor Economics*, 34, 275–318.

Arcidiacono, P., Hotz, V. J., and Kang, S. (2012), "Modeling the college major choices using elicited measures of expectations and counterfactuals," *Journal of Econometrics*, 166, 3–16.

Balinski, M. and Sönmez, T. (1999), "A tale of two mechanisms: Student placement," *Journal of Economic Theory*, 84, 73–94.

Behaghel, L., de Chaisemartin, C., and Gurgand, M. (2015), "Ready for boarding?: the effects of a boarding school for disadvantaged students," .

Booker, K., Sass, T., Gill, B., and Zimmer, R. (2009), "Going beyond test scores, evaluating charter school impact on educational attainment in Chicago and Florida," *RAND Education, Working Paper WR-610-BMG.*

Calsamiglia, C., Haeringer, G., and Klijn, F. (2010), "Constrained school choice: an experimental study," *American Economic Review*, 100, 1860–74.

Calsamiglia, C., Fu, C., and Güell, M. (2014), "Structural estimation of a model of school choices: the Boston mechanism vs. its alternatives," *FEDEA Working Paper No. 2014-21.*

Carvalho, J.-R., Magnac, T., and Xiong, Q. (2014), "College choice allocation mechanisms: structural estimates and counterfactuals," *Toulouse School of Economics Working Paper No. TSE-506.*

Chade, H. and Smith, L. (2006), "Simultaneous search," *Econometrica*, 74, 1293–307.

Clark, D. (2010), "Selective schools and academic achievement," *The BE Journal of Economic Analysis & Policy*, 10.

Cullen, J. B., Jacob, B. A., and Levitt, S. (2006), "The effect of school choice on participants: Evidence from randomized lotteries," *Econometrica*, 74, 1191–1230.

Cunha, F. and Heckman, J. J. (2008), "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation," *Journal of human resources*, 43, 738–782.

Cunha, F. and Heckman, J. J. (2010), "Investing in our young people," Tech. rep., National Bureau of Economic Research.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010), "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica*, 78, 883–931.

Ding, W. and Lehrer, S. F. (2007), "Do peers affect student achievement in China's secondary schools?" *The Review of Economics and Statistics*, 89, 300–312.

Direction Générale des Études, de la Planification et des Systèmes d'Information (République tunisienne, Ministère de l'Éducation, Secrétariat Général) (2013), "Statistiques Scolaires –Année scolaire 2012–13," .

Dobbie, W. and Fryer Jr, R. G. (2011), "Exam high schools and academic achievement: Evidence from New York City," Tech. rep., National Bureau of Economic Research.

Dong, Y. (2016), "Regression discontinuity designs with sample selection," *Working Paper*, [Accepted, Journal of Business and Economic Statistics.].

Dubins, L. and Freedman, D. (1981), "Machiavelli and the Gale-Shapley algorithm," *American Mathematical Monthly*, 88, 485–94.

Dur, U., Hammond, R., and Morrill, T. (2016), "Identifying the harm of manipulable school-choice mechanisms," *Working paper.*

Erdil, A. and Ergin, H. (2008), "What's the matter with tie-breaking? Improving efficiency in school choice," *American Economic Review*, 98, 669–89.

Estrada, R. and Gignoux, J. (2014), "Benefits to elite schools and the formation of expected returns to education: Evidence from Mexico City," .

Fack, G., Grenet, J., and He, Y. (2015), "Beyond truth-telling: preference estimation with centralized school choice," *Working paper.*

Frandsen, B. R., Frölich, M., and Melly, B. (2012), "Quantile treatment effects in the regression discontinuity design," *Journal of Econometrics*, 168, 382–395.

Gale, D. and Shapley, L. (1962), "College admissions and stability of marriage," *American Mathematical Monthly*, 69, 9–15.

Haeringer, G. and Klijn, F. (2009), "Constrained school choice," *Journal of Economic Theory*, 144, 1921–47.

Hahn, J., Todd, P., and van Der Klaauw, W. (2001), "Identification and estimation of treatment effects in a regression-discontinuity design," *Econometrica*, 69, 201–9.

Hastings, J. and Weinstein, J. (2009), "Information, school choice and academic achievement: Evidence from two experiments," *The Quarterly Journal of Economics*, 123, 1373–414.

Hastings, J., Weelden, R. V., and Weinstein, J. (2007), "Preferences, information, and parental choice behavior in public school choice," *NBER Working Paper No. 12995*.

He, Y. (2016), "Gaming the Boston school choice mechanism in Beijing," *Working paper*.

Holland, P. W. (1986), "Statistics and causal inference," *Journal of the American statistical Association*, 81, 945–960.

Horowitz, J. L. and Manski, C. F. (1995), "Identification and robustness with contaminated and corrupted data," *Econometrica: Journal of the Econometric Society*, pp. 281–302.

Hoxby, C. and Turner, S. (2015), "What high-achieving low-income students know about college," *American Economic Review: Papers and Proceedings*, 105, 514–17.

Imbens, G. and Kalyanaraman, K. (2011), "Optimal bandwidth choice for the regression discontinuity estimator," *Review of Economic Studies*, 79, 933–59.

Imbens, G. and Lemieux, T. (2007), "Regression discontinuity designs: a guide to practice," *Journal of Econometrics*, 142, 615–35.

Imbens, G. W. and Manski, C. F. (2004), "Confidence intervals for partially identified parameters," *Econometrica*, 72, 1845–1857.

Jackson, K. C. (2010), "Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago," *The Economic Journal*, 120, 1399–1429.

Jensen, R. (2010), "The (perceived) returns to education and the demand for schooling," *The Quarterly Journal of Economics*, 125, 515–48.

Kapor, A., Neilson, C., and Zimmerman, S. (2016), "Heterogeneous beliefs and school choice," *Working paper*.

Kim, B. (2012), "Do developmental mathematics courses develop the mathematics? Addressing missing outcome problem in regression discontinuity design," *Unpublished Manuscript*.

Lee, D. S. (2009), "Training, wages, and sample selection: Estimating sharp bounds on treatment effects," *The Review of Economic Studies*, 76, 1071–1102.

Lee, D. S. and Lemieux, T. (2010), "Regression discontinuity designs in economics," *Journal of economic literature*, 48, 281–355.

Martorell, P. and McFarlin, I. (2011), "Help or hindrance? The effects of college remediation on academic and labor market outcomes," *Review of Economics and Statistics*, 93, 436–454.

McCrary, J. (2008), "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, 142, 698–714.

McCrary, J. and Royer, H. (2011), "The effect of female education on fertility and infant health: Evidence from school entry policies using exact date of birth," *The American Economic Review*, 101, 158–95.

Pathak, P. and Sönmez, T. (2013), "School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation," *American Economic Review*, 103, 80–106.

Pistolesi, N. (2016), "Advising students on their field of study: Evidence from a French university reform," *Working paper*.

Pop-Eleches, C. and Urquiola, M. (2013), "Going to a better school: Effects and behavioral responses," *The American Economic Review*, 103, 1289–1324.

Roth, A. (1982), "The economics of matching: stability and incentives," *Mathematics of Operations Research*, 7, 617–28.

Roth, A. (2008), "Deferred acceptance algorithms: history, theory, practice, and open questions," *International Journal of Game Theory*, 36, 537–69, A Collection of Papers Dedicated to David Gale on the Occasion of His 85th Birthday, Special Issue.

Roth, A. (2015), "Why New York City's high school admission process only works most of the time," in [*Blog*] *Chalkbeat*, Available at [Accessed 9/14/2017]: `https://ny.chalkbeat.org/posts/ny/2015/07/02/why-new-york-citys-high-school-admissions-process-only-works-most-of-the-time/`.

Stinebrickner, R. and Stinebrickner, T. (2014a), "Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model," *Journal of Labor Economics*, 32, 601–44.

Stinebrickner, R. and Stinebrickner, T. (2014b), "A major in science? Initial beliefs and final outcomes for college major and dropout," *Review of Economic Studies*, 81, 426–72.

Todd, P. E. and Wolpin, K. I. (2003), "On the specification and estimation of the production function for cognitive achievement," *The Economic Journal*, 113, F3–F33.

van der Klaauw, W. (2008), "Regression–discontinuity analysis: a survey of recent developments in economics," *Labour*, 22, 219–245.

Wiswall, M. and Zafar, B. (2013), "How do college students respond to public information about earnings?" *Federal Reserve Bank of New York Staff Report No. 516*.

Zhang, J. L. and Rubin, D. B. (2003), "Estimation of causal effects via principal stratification when some outcomes are truncated by "death"," *Journal of Educational and Behavioral Statistics*, 28, 353–368.

Biography

Margaux Luflade was born on May 1st, 1989 in Fontenay-aux-Roses (Hauts-de-Seine, France). She graduated with honors (*mention très bien*) from Université Pierre-et-Marie-Curie, Paris-6 in June 2009 with a B.S. (*Licence*) in Mathematics. In June 2009, she also earned at B.A. in Political Science from SciencesPo Paris. In 2012, she earned an M.A. in Economics and Public Policy from SciencesPo Paris. She graduated with a Ph.D. in Economics from Duke University in May 2018. She will be joining the Economics faculty of the University of Pennsylvania in the Summer of 2018.