Title
High School Choice and Academic Performance in Mexico City

Author
Dustan, Andrew Duane

Publication Date
2014

Peer reviewed|Thesis/dissertation

High School Choice and Academic Performance in Mexico City

by

Andrew Duane Dustan

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Agricultural and Resource Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Alain de Janvry, Co-chair
Professor Elisabeth Sadoulet, Co-chair
Professor Ethan Ligon
Professor Frederico Finan

Spring 2014

High School Choice and Academic Performance in Mexico City

Abstract

High School Choice and Academic Performance in Mexico City

by

Andrew Duane Dustan

Doctor of Philosophy in Agricultural and Resource Economics

University of California, Berkeley

Professor Alain de Janvry, Co-chair

Professor Elisabeth Sadoulet, Co-chair

Mexico City's public high schools use a competitive, choice-based assignment system to allocate students to seats. Similar mechanisms are used in many countries and cities throughout the world. This dissertation explores two aspects of Mexico City's schools: the importance of peer effects on school choice behavior and the consequences of being admitted to a "better" school.

Students have incomplete information about the schools in their choice set, which may make choice difficult. Chapter 1 argues that while new information about a school allows students to update their beliefs about student-school match quality, which may make students more or less likely to choose the school, it also acts through channels that strictly increase demand for the school. Two such channels are a reduction in uncertainty facing risk-averse students and a direct effect of information on returns to attending that school. Peer networks, then, influence choice by providing students with information about some schools but not others. The expected effect of peer-provided information on demand for the peer's school is thus positive. This hypothesis is tested using exogenous variation in older peers' school assignment generated by the allocation mechanism. The average effect of a peer signal on the probability of choosing both the peer's school and observably similar schools is positive, consistent with information increasing expected utility on average. An alternative explanation, that students simply want to go to school with their peers, does not explain the empirical findings. The results suggest that incomplete information has a large impact on school choice even in a relatively information-rich environment, and that social networks partially overcome this problem while encouraging selection into schools attended by peers.

Chapter 2, which is joint work with Alain de Janvry and Elisabeth Sadoulet, explores an important and high-profile question in Mexico City: is there an academic benefit to elite high school admission? Winning a seat in an elite high school both promises modest rewards and imposes substantial risks on many students. We find that admission raises end-of-high school test scores by an average of 0.11 standard deviations for the marginal admittee. On the other hand, elite school admission in Mexico City increases the probability

of high school dropout by 8.5 percentage points. Students with weaker middle school grades experience a much larger rise in dropout probability as a result of admission, suggesting that the additional dropout risk is a result of increased academic rigor. We introduce a new "penalized imputation" method to show that the effect on exam scores is robust to accounting for differential dropout.

Chapter 3 explores the effect of marginal admission to a school with higher-ability peers on dropout probability and exam scores, extending some of the results of Chapter 2 to the full set of high schools that fill up during the assignment process. The average impact of admission on dropout and exam scores is negligible for students who barely score high enough to be admitted. One possible explanation for this finding is that relative ability matters for academic performance, which is consistent with the empirical finding larger jumps in peer ability due to admission predict greater increases in probability of dropout. Motivated by this empirical fact, a simple model of school choice accounting for incomplete information about one's own ability and the dependence of academic performance on relative ability is presented. The model shows that under these two conditions, the optimal choice strategy is much more complicated than under standard models of school choice.

To Tiffany, Hannah, and Katie

Contents

List of Figures

List of Tables

Acknowledgments

I am grateful to my advisors, Alain de Janvry and Elisabeth Sadoulet, for their many years of guidance and support. Their doors have been open to me since the day that I arrived at Berkeley and I am glad to have abused their time and patience. Our many meetings, email exchanges, and Friday workshop discussions have been invaluable. Their financial support has made it possible to pursue this dream. I look forward to many more years of collaboration.

Many other faculty members were also instrumental in my academic and professional development. Fred Finan struck a balance between unreserved criticism and reassuring positivity that I hope I can find with my own students one day. Ethan Ligon pushed me to be a logically consistent thinker and to take models seriously, which I have come to appreciate more each year. Jeremy Magruder was a source of excellent, constructive insights, as well as a reassuring and calming influence through the trials of graduate school. Michael Anderson taught my favorite course at Berkeley and helped me in his office hours for many years after. Sofia Villas-Boas and Max Auffhammer volunteered excellent guidance on all aspects of the job market.

Several professors sparked my interest in economics and helped build a foundation for understanding it. Robin Grier and Bob Reed made the field seem so interesting that I decided to pursue it beyond my undergraduate studies. Carlos Lamarche and Bill Even introduced me to econometrics and the process of academic research. I owe much to their teaching, patience, and encouragement.

Thanks to all of my fellow ARE students, whose friendship and camaraderie enriched this experience considerably. In particular, my officemates Lydia Ashton, Diana Ngo, and Tiffany Shih were constant sources of sanity, support, and laughter through the years. Siwei Cao, Guojun He, and Stanislao Maldonado have been great friends to me since math camp. I am grateful to have gotten to know Manuel Barron, Marshall Burke, and Kyle Emerick better as we faced the development job market together. I am also indebted to (among many others) Ben Crost, Erick Gong, and Alex Solis for being great role models.

The staff at ARE have been indispensable in my time here. Diana Lazo, Gail Vawter, and Carmen Karahalios were unfailingly competent and always kind, even when I forgot deadlines or projected general confusion about administrative affairs. Gary Casterline provided great IT support.

None of this dissertation would have been possible without the support of many outstanding public servants in Mexico. Miguel Székely and Rafael de Hoyos provided the initial opportunity for this research and Araceli Pais, Elizabeth Monroy, Vicente García, and many others supported me in my initial visits to the Secretariat of Public Education. Roberto Peña furnished the data necessary for this dissertation and generously shared his deep knowledge of the topic with me.

My family members deserve special thanks for their contributions. Mom and Dad, thank you for putting your children first in good times and hard times. Matt and Tim, thanks for being good examples (most of the time) and for your encouragement along the way.

Most of all, thank you to my girls for all of your support. Tiffany, your unconditional kindness and understanding are rare in this world. I am lucky to have found someone to put up with so much for so long and am excited for our next adventure. Hannah, you spent so many late nights and early mornings sitting on my knee at the computer that you should probably be credited as a co-author on this dissertation. Your energy and enthusiasm always put a smile on my face after a tough day. Katie, you are still small, but you have already brought so much joy to my life. I can't imagine life without any one of you.

Chapter 1

Peer Networks and School Choice under Incomplete Information

1.1 Introduction

Many education systems allow students and their families some degree of choice in which school they will attend.[1] A key rationale for choice policies is that they allow students to leverage their private information about student-school match quality—the interaction between school attributes and student preferences—by choosing the school that best caters to their own preferences and constraints. Students have incomplete information about schools, however, which may have profound effects on choice behavior. For example, Hoxby and Avery (2012) observe that low-income high-achievers in the United States rarely apply to selective colleges, a phenomenon that they attribute partially to uncertainty about how well selective colleges would suit them. But these students tend to pay little and perform well at selective colleges, highlighting the possibility that incomplete information results in privately suboptimal educational decisions.

How does incomplete information about schools affect students' choices? If we think of the student as a Bayesian learner, then he uses information about a particular school in two ways. First, new information allows him to update his expectation of match quality with that school. This channel has been studied extensively in the school choice literature, reviewed below. Second, and so far unstudied, is that information makes the student's belief about match quality more precise. If students are risk-averse, the uncertainty-reducing value of information makes students more likely to choose schools about which they are well-informed. Given the choice between two schools with identical expected match quality, the risk-averse student will choose the school where his belief is more precise. In addition to

[1] School choice has been analyzed by Abdulkadiroglu et al. (2012) and Dobbie and Fryer (2011) for the United States, Clark (2010) for the United Kingdom, Ajayi (2012) for Ghana, Lucas and Mbiti (2012) for Kenya, de Hoop (2012) for Malawi, Jackson (2010) for Trinidad and Tobago, Pop-Eleches and Urquiola (2013) for Romania, Lai et al. (2011) and Zhang (2012) for China, and de Janvry et al. (this dissertation) for Mexico.

its value for updating expectations, some information may be beneficial once the student is actually attending that school because it tells the student how to behave optimally there. I will refer to such information as "productive knowledge." Uncertainty reduction and building of productive knowledge are both channels through which new information strictly increases the expected utility from attending a school.

If the quantity of information that a student has about each school is an important determinant of choice, then the student's social network may be a crucial determinant of choice behavior because it provides information about some schools but not others. Students may learn about a particular school through interactions with older peers already attending that school. Consequently, beliefs about match quality will be systematically more precise and productive knowledge about a school will be higher where the peer network is denser. This implies (on average) stronger preferences for schools attended by older peers, even if peers do not have a direct positive effect on match quality. Hoxby and Avery's (2012) observation regarding the application behavior of low-income high-achievers may be partially explained by a dependence of choice on information from peer networks, as they find that such students often "have only a negligible probability of meeting a... schoolmate from an older cohort who herself attended a selective college."[2]

This paper shows empirically how school-specific information originating from the peer network affects school choice. To show how information should affect choice in the presence of risk aversion and returns to productive knowledge, I extend a standard school choice model by incorporating features from the literature on experience goods and word-of-mouth information. The model generates clear hypotheses about the effect of schools attended by older peers on the student's own choice of school. These hypotheses are then taken to student-level data from Mexico City's public high school choice system. The system's assignment mechanism, described in detail below, allows for causal identification of the peer effect because it generates exogenous variation in the schools attended by older peers. Both OLS evidence and estimates from a discrete choice model of school choice are consistent with the model's hypotheses. I am able to rule out competing explanations for the empirical findings, in particular pure preference for going to the same school as older peers. Thus the empirical evidence points to students relying on information obtained from peers to overcome incomplete information about match quality and/or build productive knowledge about particular schools.

Existing empirical literature on school choice under incomplete information does not incorporate risk aversion or returns to productive knowledge into student preferences. Hastings et al. (2009) provide a model of school choice where students trade off academic quality with attributes such as proximity. In their model, risk-neutral students optimize with respect to expected quality without regard for the precision of this belief. Empirical studies on the effect of information provision on school choice do not model risk aversion, either. Hastings and Weinstein (2008), for example, demonstrate that providing information on test score aggregates to low-income families in the United States increased the likelihood of choosing

[2] Hoxby and Avery (2012) p. 2.

high-performing schools. Related studies by Koning and van de Wiel (2010) in the Netherlands and Friesen et al. (2012) in Canada come to similar conclusions, while Mizala and Urquiola (2013) find no effect of publishing a quality measure in Chile. In each of these cases, the information was enriched for all schools simultaneously, which did not induce between-school variation in the amount of information available to students.

In contrast to previous studies, the students in my empirical setting learn about some schools but not others. Because of this within-student variation in the relative quantity of information known about each school, it is useful to extend the standard school choice model by allowing risk aversion with respect to match quality and returns to productive knowledge. To model risk aversion, I incorporate elements from models of learning about experience goods. In particular, Roberts and Urban (1988) and Erdem and Keane (1996) model potentially risk-averse consumers as having incomplete information about a consumption good's quality (or bundle of attributes).[3] Consumers have unbiased priors about quality and Bayesian update these priors when they are exposed to an advertisement or word-of-mouth information. For a given level of expected quality, consumers prefer goods where this belief is more precise, such that advertising and word-of-mouth information increase demand on average. My model is similar to these, but word-of-mouth information comes from older peers. Students have an unbiased but noisy prior about match quality between themselves and each school. They receive signals about match quality with schools attended by older peers, and use these signals to update their beliefs. While some students update their expectation of match quality upward and others update downward, these changes cancel out when averaging over all students and schools.[4] The reduction in uncertainty, however, increases expected utility for all students, so that on average a peer signal increases demand for the school. Obtaining more productive knowledge about a school from peers also has an unambiguously positive impact on expected utility from that school. The positive average effect of peer signals on demand is the model's key testable hypothesis.

Hypotheses about the effect of peer learning on school choice are difficult to test without exogenous variation in the schools attended by peers, due to the well-known reflection problem put forth in Manski (1993 and 1995). Students have similar preferences to their peers and share some of the same constraints, so observing a student choosing the same school as older members of his peer group is not necessarily indicative of learning from peer networks. The sociology and education literatures have instead studied the effect of social learning on school choice in a qualitative framework. Most notably, Ball and Vincent (1998) find that, for primary schools in the United Kingdom, parents use their social networks (the "grapevine" as they call it) to obtain specific, detailed information about schools and their likely fit for their own children.[5] The economics literature has so far been limited to carefully documenting correlations, as in Hoxby and Avery (2012).

[3]Many other studies relating to experience goods have used similar models, for example Johnson and Myatt (2006) and Crawford and Shum (2005).

[4]This relies on CARA utility and normally distributed prior beliefs and signals, discussed in Section 1.3.

[5]Ceja (2006) finds qualitative evidence that older siblings are an important source of information for Chicana students as they apply to college in the United States.

Mexico City's school assignment mechanism generates exogenous variation in the school assignment of a student's older peers, which can be used to test the model's hypotheses about the effect of peer signals on choice. Public high schools in Mexico City use a unified choice-based allocation system where assignment priority is determined on the basis of an exam score. This generates a regression discontinuity design where, given a group of older peers who want to attend a certain school, some students score barely high enough to be admitted and others score barely too low and must attend another school. The variation in peer assignment resulting from the discontinuity is used to identify both the OLS and discrete choice models. While the OLS methods give clean, easily-interpreted evidence for the social learning model of school choice, the discrete choice model is important because it directly tests the model's hypotheses about the effect of information on expected utilities while also giving an interpretation of the effects in terms of a marginal willingness to travel (or pay). The data span twelve years of admissions cycles and contain rich information about the choices, demographics, and assignment of each participant. Combining the students' names, locations, and demographic information, I match students with a certain kind of peer that is both identifiable in the data and is expected to be an important peer in a student's network: the older sibling. Thus the relationship of interest in this paper is the effect of an older sibling's admission outcome on the younger student's choice of schools.

The empirical results show that students prefer schools attended by their older siblings. Using the estimates from the discrete choice model, I find that students are willing to increase their round-trip commute by an average of 4.8 km per day in order to attend a school to which the older sibling was admitted, which is valued at $561 over the course of high school. This effect is not driven by the obvious explanation that it is convenient or beneficial for the family to have two children attending the same school. Having an older sibling admitted to a particular high school increases the revealed preference for that school, even when the siblings are far enough apart in age that the older sibling no longer attends high school. Furthermore, having an older sibling admitted to a school increases revealed preference for other campuses belonging to the same school subsystem, within which individual schools throughout the city share many attributes such as curriculum and vocational orientation. This suggests that students generalize the knowledge obtained about about a peer's school when evaluating all other schools within the same subsystem. There is also evidence, although not strictly causal, that revealed preference for a school increases much more when the older sibling experiences a positive academic outcome there. Taken together, these results support the view that students prefer schools where they have more information, and use the information from their peer network to update beliefs about match quality.

The policy prescriptions for addressing incomplete information in school choice depend critically on the source of uncertainty. I will show that in the empirical context of Mexico City, uncertainty about match quality is unlikely to come primarily from an inability to observe basic school attributes such as peer quality and academic rigor. Students already have access to information about peer quality, and each school subsystem has a well-known reputation regarding its curriculum and academic level. Furthermore, differences in costs across schools are not a first-order concern in this setting, in contrast to the context of

United States higher education studied in Hoxby and Avery (2012) and Hoxby and Turner (2013).[6] Rather, uncertainty appears to originate from incomplete information about more specific or idiosyncratic elements of match quality. When students do not know basic school attributes, an easy solution may be to distribute official information about school attributes. But resolving uncertainty about idiosyncratic match quality and transferring productive knowledge about a school requires more personalized information, meaning that the peer network may be more useful than official efforts and that policymakers must find innovative ways to address this individual-specific source of uncertainty.

The remainder of the paper proceeds as follows. Section 1.2 explains the public high school choice system in Mexico City, showing that it provides a good context in which to empirically examine school choice under incomplete information. Section 1.3 sets forth a simple model of school choice under incomplete information, concluding with testable hypotheses about the expected effect of new information. Section 1.4 explains the data and how they will be used to test the model. Section 1.5 gives the OLS method and results, while Section 1.6 lays out the discrete choice model and corresponding results. Section 1.7 provides validity checks for the empirical design and Section 1.8 concludes with policy recommendations.

1.2 High school choice in Mexico City

This section explains Mexico City's public high school choice system. In addition to providing context for the empirical exercise, it explains the assignment mechanism that is the basis for exogneous school assignment of peers and highlights some features of the system that induce students to reveal their true school preferences.

1.2.1 The COMIPEMS assignment mechanism

Prior to 1996, the ten major public high school subsystems in Mexico City controlled their own independent admissions processes. Students applied to schools in one or more of these subsystems, waited to learn where they had been admitted, and then withdrew from all schools except their most-preferred one. In an effort to increase both the efficiency and transparency of this process, the subsystems formed the Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS) in 1996. Each year, COMIPEMS runs a unified, competitive admissions process that assigns students across Mexico City's public high schools on the basis of students' preferences and the results of a standardized exam.

[6]Hoxby and Turner (2013) show that providing semi-personalized information about college costs to low-income, high-achieving high school students in the United States has a large, positive impact on revealed preference for selective colleges. They interpret this as proof of upward-biased beliefs about selective college costs, but uncertainty and risk aversion could be a complementary channel.

The COMIPEMS admissions process is as follows.[7] In late January, students in ninth grade—the final year of middle school—receive informational materials about the admissions process. These materials include a list of all of their "educational options," which in most cases are schools but can also be specific tracks within schools, such as specific vocational education tracks in a technical school. Students then fill out a registration form, demographic survey, and list of up to 20 educational options, ranked in order of their preference. These forms must be submitted in late February or early March, depending on the student's family name. In June of that year, students take a standardized exam consisting of 128 multiple-choice questions, covering both subject-specific material from the public school curriculum and more general mathematical reasoning and language areas.

In July, the assignment process is carried out by the Centro Nacional de Evaluación para la Educación Superior (CENEVAL).[8] First, the school subsystems report the maximum number of seats available to incoming students. Second, all students who did not successfully complete middle school or scored below 31 of 128 points are discarded. Third, all remaining students are ordered by their exam score, from highest to lowest. Fourth, a computer program proceeds sequentially down the ranked list of students, assigning each student to his highest-ranked option that still has a seat remaining.[9] The process continues until all students are assigned, with the exception of students who scored too low to enter any of their listed options. Later in July, the assignment results are disseminated to students. Through 2011 this primarily happened in the form of a printed gazette sold at newsstands, although a system that sends personalized results via text message has become more popular over time.[10] At that time, students who were eligible for assignment but were left unassigned during the computerized process because they scored too low for any of their choices may choose a schooling option from those with seats remaining.

1.2.2 Student decision making under the COMIPEMS mechanism

Students have considerable information about basic school attributes when they choose schools, but this information is generic rather than individually tailored. The subsystem membership of each school is known with certainty, and each subsystem has a well-formed public perception. There are two "elite," university-affiliated subsystems: the Universi-

[7]The timing of each step is given for the 2011 competition, although this may change slightly from year to year. The assignment rules were different in 1996 and 1997, but those years are not considered in this paper so they are not discussed.

[8]CENEVAL is independent of COMIPEMS and its constituent school subsystems. This process is carried out by computer in the presence of representatives from all subsystems and external auditors from a large international accountancy firm.

[9]In the instance that two or more students have the same score and highest-ranked available option, but there are fewer remaining seats than the number of tied students, the assignment process pauses and representatives from the corresponding subsystem must decide to either admit all tied students or none of them.

[10]The gazette was replaced in 2012 with an electronic version.

dad Nacional Autónoma de México (UNAM) and the more technically-focused Instituto Politécnico Nacional (IPN). These are universally understood to be highly competitive, relatively rigorous, prestigious high schools that fill their student capacities before almost all non-elite options. Non-elite subsystems include those with traditional academic curricula and technical subsystems providing academic coursework combined with vocational training for careers such as auto repair and bookkeeping. Even within a subsystem, official information about school-level academic quality is available. Past cutoff scores—the score of the student admitted to the school's final seat—for each school have been available on the COMIPEMS web site since 2005, and this site is actually browsed by many students because it allows them to easily complete most of the registration process online.[11] Cutoff score and the mean score of admitted students are almost perfectly correlated, so students have access to an excellent proxy for mean peer ability. The combination of subsystem reputations and information about peer quality ensures that students are at least somewhat informed about general school attributes, though they may lack more specific details that affect the idiosyncratic match between the student and school.

Students in Mexico City often construct their rankings in the following way, similar to how United States students choose colleges (see Hoxby and Avery (2012), for example).[12] First, they decide whether they would like to attend a high school in either or both of the two elite subsystems. If a student decides to apply within either or both subsystems, he lists some number of elite schools as his top choices. There are 30 elite schools (16 IPN and 14 UNAM), meaning that even within an elite subsystem, students face a wide variety of options. Following the elite schools, if any, he lists various non-elite schools (from about 600 options in most years), which offer a better chance of admission.

Two aspects of the COMIPEMS assignment mechanism make the student's ranking quite informative about true preferences. First, the mechanism is equivalent to the deferred acceptance algorithm proposed by Gale and Shapley (1962), so it induces truth-telling by students.[13] In particular, under such mechanisms it is never optimal to list a less-preferred school before a more-preferred school, regardless of the limit on how many options can be listed. Second, the ability to rank up to twenty options means that few students actually fill up their entire preference sheet; students generate a satisfactory choice portfolio without confronting the space constraint.[14] There is no strategic disadvantage to choosing a school at which the student has a small *ex ante* probability of admission, both because the number of options allowed is high and because the assignment algorithm does not punish students

[11] Approximately 80% of the students analyzed in this sample went through the selection process in 2005 or later.

[12] I thank Roberto Peña Resendíz and advisers at the Subsecretariat of High School Education for insight into this process.

[13] See Dubins and Freedman (1981) and Roth (1982). This particular mechanism is referred to as a student-proposing deferred acceptance mechanism, which is discussed in Abdulkadiroglu and Sonmez (2010).

[14] Choosing the optimal portfolio of schools is a complex problem if listing choices is costly (e.g. time cost or opportunity cost due to a limited number of allowed choices), as mentioned by Ajayi (2012) and explored in depth by Chade and Smith (2005).

for ranking unattainable schools.

1.3 Model of school choice

This section extends a model of school choice from Hastings et al. (2009) by incorporating incomplete information, risk aversion, productive knowledge, and learning from peers. In my model, the utility from attending each school is uncertain because of incomplete information about student-school match quality. Risk-averse students revise their beliefs about utilities by receiving informative signals about match quality from peers. The setup is similar to models of consumer demand for experience goods, in particular Roberts and Urban (1988) and Erdem and Keane (1996), where consumers are uncertain about product quality and revise their beliefs due to word-of-mouth or informative advertising. Students also gain productive knowledge about schools from their peers, which allows them to obtain higher utility from attending the peer's school. This latter advantage can be thought of in a similar way to the effect of learning on technology adoption, as in Foster and Rosenzweig (1995). In this case, students are unsure of how to use the school "technology" to build human capital but learn from peers about how to do so optimally.

This model produces testable hypotheses about how students react to new information about specific schools. First, the model predicts that the average impact of new information on same-school expected utility is positive. This is a prediction about the average effect of new information over all students and schools in the population, not a prediction that the average effect will be positive for each school. Second, the model predicts that the impact of new information depends on how positive or negative the signal was. Finally, these effects are predicted to apply, to a lesser degree, to other schools that are observably similar to the school about which the information was received.

1.3.1 General setup

The student's problem is to maximize expected utility by choosing one school to attend from his choice set. Here I abstract from the problem of portfolio construction and focus on the first choice. This is reasonable if one thinks that the first listed option is the student's most-preferred school, a modest assumption given the large number of options that a student is allowed to list in order to diversify and choose safety schools.

Student i's utility from school $j \in J$ is a function of student-school match quality:

$$U_{ij} = U\left(\boldsymbol{X_{ij}\beta_i}\right) = U\left(\left(\boldsymbol{\bar{X}_j} + \boldsymbol{\widetilde{X}_{ij}}\right)\boldsymbol{\beta_i}\right)$$

where match quality is expressed as the sum of student-school attributes in the vector $\boldsymbol{X_{ij}}$ weighted by the student-specific vector of preference parameters $\boldsymbol{\beta_i}$. The attribute vector is decomposed into two terms: $\boldsymbol{\bar{X}_j}$ is the average level in the population and $\boldsymbol{\widetilde{X}_{ij}}$ is the student-specific deviation from this level. An example of a student-school attribute is academic fit, which is on average higher at some schools than others, but also has a student-specific component that depends on how well the school caters to the student's particular learning style and ability level.

The student knows the relative weights $\boldsymbol{\beta_i}$ he puts on each attribute. If he also knows $\boldsymbol{X_{ij}}$, and if he is risk-neutral with respect to match quality, so that $U(\boldsymbol{X_{ij}\beta_i}) = \boldsymbol{X_{ij}\beta_i}$, this model is nearly identical to the one in Hastings et al. (2009). In that case, the student chooses school j if it provides the highest match quality out of all schools in the choice set: $\boldsymbol{X_{ij}\beta_i} > \boldsymbol{X_{ik}\beta_i}\ \forall k \neq j \in J$.[15]

1.3.2 Incomplete information about match quality

Incomplete information about match quality is modeled by making it so that the student imperfectly observes student-school attributes. He does not observe $\boldsymbol{\bar{X}_j}$ or $\boldsymbol{\widetilde{X}_{ij}}$, but he knows the distributions from which each is drawn:

$$\boldsymbol{\bar{X}_j} \sim \mathcal{N}\left(\boldsymbol{\bar{X}^0_j}, \boldsymbol{\Sigma}_{\boldsymbol{\bar{X}_j}}\right), \quad \boldsymbol{\widetilde{X}_{ij}} \sim \mathcal{N}\left(\boldsymbol{\widetilde{X}^0_{ij}}, \boldsymbol{\Sigma}_{\boldsymbol{\widetilde{X}_{ij}}}\right).$$

For simplicity of exposition, the covariance matrices $\boldsymbol{\Sigma}_{\boldsymbol{\bar{X}_j}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\widetilde{X}_{ij}}}$ are assumed to be diagonal, and $\boldsymbol{\bar{X}_j}$ and $\boldsymbol{\widetilde{X}_{ij}}$ are assumed to be mean independent. Thus $\boldsymbol{X_{ij}}$ is distributed normally with mean $\boldsymbol{X^0_{ij}} = \boldsymbol{\bar{X}^0_j} + \boldsymbol{\widetilde{X}^0_{ij}}$ and diagonal covariance matrix with $(\ell,\ell)^{\text{th}}$ entry $1/\tau^0_{\ell ij}$.[16]

Because $\boldsymbol{X_{ij}}$ is unknown, a risk-neutral student chooses j if it maximizes *expected* match quality: $\mathrm{E}_0[\boldsymbol{X_{ij}\beta_i}] > \mathrm{E}_0[\boldsymbol{X_{ik}\beta_i}]\ \forall k \neq j \in J$, where the 0 subscript indicates that the expectation is formed solely on the basis of the match quality distributions. Incomplete information about match quality (in particular, about mean quality $\boldsymbol{\bar{X}_j}$) is sufficient to predict the results from Hastings and Weinstein (2008), where giving information about school-level average test scores to students increased the weight that students placed on test scores when choosing schools.[17]

1.3.3 Risk aversion and returns to productive knowledge

I now introduce two channels through which information will positively affect expected utility: returns to productive knowledge and risk aversion with respect to match quality.

I parameterize the returns to productive knowledge in a simple way, adding a term $r_j(n_{ij})$ to the utility function, where n_{ij} is the level of i's knowledge about school j. The marginal return to knowledge is strictly positive so that $r'_j > 0$. Examples of productive knowledge

[15]Hastings et al. (2009) do not explicitly model uncertainty, but they do say that uncertainty about an attribute would lead to a lower effective weight being placed on it.

[16]I assume that for any two schools j and k, $\boldsymbol{X_{ij}}$ and $\boldsymbol{X_{ik}}$ are mean independent.

[17]Intuitively, students were choosing on the basis of both signal and noise about test scores, and the information intervention allowed students to choose on the basis of a stronger signal.

are knowing which teachers are the best to take or being aware of an after-school tutoring program.

Allowing the student to be risk-averse will address a troubling result from the risk-neutral model. Risk neutrality implies that the relative precision with which match quality is known does not affect choice. That is, presented with a choice between two schools of equal expected match quality but where one's match is known with complete certainty and the other with uncertainty, the student will be indifferent between them. A risk-averse student will prefer the school where match quality is known with certainty.

To model risk aversion, I allow utility to be concave in match quality. Following Roberts and Urban (1988), I use exponential utility:

$$U_{ij} = -exp\left\{-\rho \boldsymbol{X_{ij}}\boldsymbol{\beta_i} + r_j\left(n_{ij}\right)\right\}$$

where ρ, the coefficient of risk aversion, is assumed to be positive. Due to exponential utility and the joint normal distribution of $\boldsymbol{X_{ij}}$, expected utility from school j in the absence of additional information can be written in terms of the mean and variance (or precision) of the prior distribution of match quality, as well as the return to productive knowledge:[18]

$$\begin{aligned} U^*_{0ij} &= \mathrm{E}_0\left[\boldsymbol{X_{ij}}\boldsymbol{\beta_i}\right] - \frac{\rho}{2} Var\left(\boldsymbol{X_{ij}}\boldsymbol{\beta_i}\right) + r_j\left(n^0_{ij}\right) \\ &= \boldsymbol{X^0_{ij}}\boldsymbol{\beta_i} - \frac{\rho}{2}\sum_{\ell} \frac{\beta^2_{\ell i}}{\tau^0_{\ell ij}} + r_j\left(n^0_{ij}\right). \end{aligned} \tag{1.1}$$

where $\beta^2_{\ell i}/\tau^0_{\ell ij}$ is the variance of the distribution of match quality from attribute ℓ. The student optimizes with respect to both the mean and variance of match quality, so schools are now "penalized" when beliefs about them are noisier. He also values productive knowledge. He chooses the school j that provides the highest expected utility of all available schools: $U^*_{0ij} > U^*_{0ik}\ \forall k \neq j \in J$.

1.3.4 Effect of peer information

When student i's peer attends school j, she gives two pieces of information. First, she provides productive knowledge about school j, so that the new level of knowledge is higher: $n^1_{ij} > n^0_{ij}$. Second, the student improves on his prior belief about match quality by receiving informative signals about student-school attributes $\boldsymbol{X_{ij}}$. This information comes in the form of an unbiased, noisy signal about each attribute:

$$\boldsymbol{P_{ij}} = \boldsymbol{X_{ij}} + \boldsymbol{\varepsilon_{ij}}, \quad \boldsymbol{\varepsilon_{ij}} \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma_{P_{ij}}}\right),$$

[18] The full expression for expected utility is $\mathrm{E}_0\left[U_{ij}\right] = -exp\left\{-\rho\left(\boldsymbol{X^0_{ij}}\boldsymbol{\beta_i} - \frac{\rho}{2}\sum_\ell \frac{\beta^2_{\ell i}}{\tau^0_{\ell ij}} + r_j\left(n^0_{ij}\right)\right)\right\}$, but since this is strictly monotonically increasing in the terms in braces, this is equivalent to optimizing with respect to equation 1.1.

where $\boldsymbol{\Sigma}_{\boldsymbol{P_{ij}}}$ is diagonal with entries $1/\tau^P_{\ell ij}$. The signals received are about student-school attributes for student i, not the peer.[19] The idea is that social interactions with the peer allow i to learn more about the school and infer something about how much he will benefit from different aspects of it.

The student uses this new information to update his expected utility from attending school j. Because the prior and signal are both distributed normally and because the covariance matrix for each is diagonal, the form of the posterior distribution of each student-school attribute is simple:

$$X^1_{\ell ij} \sim \mathcal{N}\left(\frac{\tau^0_{\ell ij} X^0_{\ell ij} + \tau^P_{\ell ij} P_{\ell ij}}{\tau^0_{\ell ij} + \tau^P_{\ell ij}},\ \frac{1}{\tau^0_{\ell ij} + \tau^P_{\ell ij}}\right)$$

The posterior distribution of each attribute is a precision-weighted average of the prior and signal. The expected utility from j is now

$$U^*_{1ij} = \widehat{\boldsymbol{X}^{\boldsymbol{1}}_{\boldsymbol{ij}}} \boldsymbol{\beta_i} - \frac{\rho}{2} \sum_{\ell} \frac{\beta^2_{\ell i}}{\left(\tau^0_{\ell ij} + \tau^P_{\ell ij}\right)} + r_j\left(n^1_{ij}\right) \tag{1.2}$$

where $\widehat{\boldsymbol{X}^{\boldsymbol{1}}_{\boldsymbol{ij}}}$ is the mean of the posterior distribution of $\boldsymbol{X}^{\boldsymbol{1}}_{\boldsymbol{ij}}$. To see how the peer signals affected expected utility, compare equations 1.1 and 1.2:

$$U^*_{1ij} - U^*_{0ij} = \left(\widehat{\boldsymbol{X}^{\boldsymbol{1}}_{\boldsymbol{ij}}} - \boldsymbol{X}^{\boldsymbol{0}}_{\boldsymbol{ij}}\right) \boldsymbol{\beta_i} + \frac{\rho}{2} \sum_{\ell} \frac{\beta^2_{\ell i} \tau^P_{\ell ij}}{\tau^0_{\ell ij}\left(\tau^0_{\ell ij} + \tau^P_{\ell ij}\right)} + \left(r_j\left(n^1_{ij}\right) - r_j\left(n^0_{ij}\right)\right) \tag{1.3}$$

The change in expected utility comes from three sources. The first term is the change in expected match quality. This quantity may be positive or negative depending on the content of the peer signal. Students may learn that the school is a better or worse match for them than they had guessed. The second term is the change in expected utility resulting from the lower variance in the posterior distribution of match quality. This quantity is unambiguously positive. The increased knowledge about match quality works in the school's favor because the risk-averse student is now more certain about how good the match is. The third term is the change in the utility from productive knowledge, which is also positive.

This result gives rise to two testable hypotheses, derived in the appendix:

Hypothesis 1: *The expected effect of peer information on U^*_{ij}, taken over all students i and schools j, is positive:* $\mathrm{E}_{ij}\left[U^*_{1ij} - U^*_{0ij}\right] > 0$.

This is the key testable hypothesis of the model that distinguishes it from models without channels through which information strictly increases expected utility. It says that, on average, receiving peer information about a school increases the expected utility from attending there. Intuitively, the signal is sometimes better than the student's prior belief and sometimes it is worse, but the average effect on expected match quality is zero. On the

[19]This is in contrast with Roberts and Urban (1988), in which only quality for the peer is observed.

other hand, the reduction in uncertainty about match quality and the increase in productive knowledge always work in the school's favor. Note that the expected effect may be positive for certain schools and negative for others, because mean quality $\bar{\boldsymbol{X}}_{\boldsymbol{j}}$ is drawn from a random distribution. This hypothesis is about the expected effect over all schools.

Hypothesis 2: *All else equal, the change in expected utility from j depends positively on how favorable the peer signal about match quality from j was:* $\frac{\partial\left(U^*_{1ij}-U^*_{0ij}\right)}{\partial \boldsymbol{P_{ij}\beta_i}} > 0$.

This hypothesis simply says that when the student receives a relatively good (i.e. high) signal about the match quality from a school, he is more likely to choose that school than if he had received a relatively bad (low) signal.

1.3.5 Shared attributes across schools

Students may know that the level of an attribute is shared across schools. In the empirical setting studied here, schools are divided into subsystems that share important attributes such as curriculum and vocational orientation. In this case, learning about one school in the subsystem also yields useful information about all other schools in the same subsystem. (Likewise, productive knowledge about one school might be applicable to other schools in the subsystem. I will not model this because it is now obvious that this channel will operate identically to the learning about shared attributes channel.) In order to model the shared attributes in a simple way, we can maintain all prior assumptions of the model and additionally assume that for school j in subsystem s, match quality is expressed as $\boldsymbol{X_{ijs}\beta_i} + \mu_{is}$, where $\mu_{is} \equiv \bar{\mu}_s + \widetilde{\mu}_{is}$. The average component of subsystem match quality is distributed $\bar{\mu}_s \sim \mathcal{N}\left(\bar{\mu}^0_s, \sigma^2_s\right)$ and the student-specific component is distributed $\widetilde{\mu}_{is} \sim \mathcal{N}\left(\widetilde{\mu}^0_{is}, \eta^2_{is}\right)$, and $1/\tau^\mu_{is} \equiv \sigma^2_s + \eta^2_{is}$. In addition to the signal $\boldsymbol{P_{ij}}$ about unshared attributes, the student receives a signal about the shared attribute:

$$q_{is} = \mu_{is} + \xi_{is}, \quad \xi_{is} \sim \mathcal{N}\left(0, 1/\tau^q_{is}\right).$$

When the student receives a signal about school j in subsystem s, she can update her expected utility from a different school k in the same subsystem:

$$U^*_{1iks} - U^*_{0iks} = \left(\widehat{\mu}^1_{is} - \mu^0_{is}\right) + \frac{\rho}{2}\frac{\tau^q_{is}}{\tau^\mu_{is}\left(\tau^\mu_{is} + \tau^q_{is}\right)} \tag{1.4}$$

where $\widehat{\mu}^1_{is}$ is the mean of the posterior distribution of the shared attribute and μ^0_{is} is the mean of the prior. This assumption of a shared attribute produces two additional hypotheses, derived in the appendix:

Hypothesis 3: *The expected effect of peer information on the expected utility from any other school in the same subsystem is positive: indexing the peer's school by j and fixing another school k_j in j's subsystem s_j,* $\mathrm{E}_{ij}\left[U^*_{1ik_js_j} - U^*_{0ik_js_j}\right] > 0$.

On average, receiving a signal about a school increases the expected utility from attending other schools in the same subsystem. The intuition is the same as for Hypothesis 1. Surprises about the match quality from j's subsystem are also surprises about the match quality for all other schools in the subsystem. The surprises cancel out when we average across all schools and students. There is always a reduction in uncertainty about match quality from j's subsystem, which increases expected utility from attending schools in the subsystem.

Hypothesis 4: *Suppose the student receives a peer signal about school j in subsystem s. All else equal, the change in expected utility from school k in subsystem s depends positively on how favorable the peer signal about subsystem match quality was:* $\frac{\partial\left(U^*_{1iks}-U^*_{0iks}\right)}{\partial q_{is}} > 0$.

The more positive a surprise to the match quality for j's subsystem, the larger is the increase in expected utility from other schools in the same subsystem.

1.4 Data and sample construction

Testing the school choice model requires student-level choice data that include a measure of signals received from older peers. This section proposes one such measure that exists in the Mexico City data before describing the data set and sample construction in more detail. The sample construction is key because it forms the basis of the regression discontinuity design.

1.4.1 Siblings as peers

The administrative data used in this paper do not contain any explicit information on peer network structure. Moreover, since middle schools in Mexico City are quite large and neighborhoods are not geographically isolated, neither can be used to construct a useful proxy for the student's network. The data do, however, allow for the identification of siblings within a family, which is useful for a number of reasons. First, older siblings are almost surely members of the student's peer network. Second, the strength of the peer relationship is likely to be very strong on average, compared to most classmates and neighbors. Third, the constant interactions between siblings within the home make it probable that the student learns a significant amount about the details of the school attended by his older sibling and how that school might fit his own tastes, which is the mechanism by which the social learning model proposes that peers affect school choice. Thus, the older sibling presents an attractive solution to the lack of social network information in administrative data and is a good candidate for identifying the informational role of peers with sufficient statistical power to make sharp inference.

Siblings are different from other peers in a way that may cause some concern when generalizing sibling-derived effects to those of the broader network. Perhaps students find it useful to attend the same school as an older sibling because they want to commute together or derive some social benefit. These issues are addressed in two ways. First, the analysis is sometimes limited only to sibling pairs where the age gap is large enough that the older sibling no longer attends high school by the time the younger one enrolls, so that there is zero or limited direct logistical or social benefit to attending the same school. This will be explained in more detail below. Second, this direct benefit does not exist when the student does not attend the sibling's school, but rather a different school in the same subsystem; any effect on subsystem preference cannot come through this channel.

1.4.2 Data description

This paper uses administrative data compiled by COMIPEMS for twelve admissions cycles, from 2000 to 2011. For each student who registered for the exam, the database contains basic demographic information including the student's full name, date of birth, phone number, address, and a unique middle school identifier along with the grade point average attained there; the full list of up to 20 ranked school preferences; a context survey, completed by the student, including information about parental education, family composition, and other topics; and assignment results, including the student's exam score and the school assigned during the computerized allocation process. The analysis is limited to students who were in middle school at the time they took the exam rather than re-taking in subsequent years and where the older sibling attended a public middle school.

To measure whether the older sibling graduated or dropped out of high school (a proxy for whether the peer signal transmitted to the younger sibling was good or bad), the COMIPEMS database is merged via national ID number (CURP) with a database from the national 12$^{\text{th}}$ grade exam, called the ENLACE Media Superior. This exam is only given to students who are on track to graduate at the end of the academic year, so it is a good proxy for graduation. Unfortunately, this exam was only administered starting in the spring of 2008, and the database used in this paper contains results from 2008 to 2010, corresponding to students taking the COMIPEMS exam in 2005-2007. Thus the part of the analysis using this graduation data is limited to younger siblings of these cohorts, which limits sample size. The larger and more demanded of the two elite subsystems, the UNAM, does not administer the ENLACE exam so graduation data is missing for students assigned there. This further limits the sample size when the graduation measure is used.

The demographic information is used to match siblings with each other in the following way. First, potential siblings are identified if they have the same paternal and maternal family names and either 1) have the same phone number or 2) live in the same postal code and attend the same middle school. From this pool of potential matches, sibling pairs are discarded if 1) the students state that they have different numbers of siblings; 2) the students do not report a birth order that makes them the closest siblings in the family (e.g. first- and second-born);[20] 3) the students were born fewer than nine months apart or more than five years apart, the latter because it is unlikely that consecutive births five or more years apart represent a true match; or 4) the older student took the exam after the younger one. If

[20] This is done so that the estimated effect of older sibling of admission does not include an indirect effect through the influence on a middle sibling's behavior

one student matches with two potential older siblings, the match based on the shared phone number is used.

This matching process locates 267,748 sibling pairs in a population of 2,127,375 students.[21] Columns 1 and 2 of Table 1.1 give a description of demographic, academic, and school choice variables for the full sample of students and for the matched older siblings (since they are the basis for sample selection), respectively. The matched older siblings, on average, have more educated parents and are modestly higher achievers in terms of both grade point average and COMIPEMS exam score (about 1/5 standard deviations higher in each case). The average student ranks 9 or 10 school choices, which is similar across samples. About 2/3 of students select a school in one of the two elite subsystems as their first option, but fewer than 1/4 are admitted to one. Elite admission is higher for older siblings, consistent with their higher exam scores. On average, students choose a school over 7km away as their first option, measured as a straight line from the center of the student's home postal code to the school.[22] Siblings are, on average, 2.5 years apart and have fairly similar school preferences: 34 percent of sibling pairs select the same school as their first choice. Only 45% took the ENLACE exam, similar to the official graduation rate in Mexico City. This proportion is 10 percentage points higher for older siblings, a gap that drops below 6 percentage points when controlling for older sibling observables (not shown).

Why are the matched older siblings generally higher achievers in middle and high school? One explanation is that the matching process, which relies on siblings having the same phone number or attending the same middle school, finds families in more stable living situations. Such families probably have higher-achieving children. Another is that ability and preference for schooling are correlated within a family, so that families with a high-achieving older sibling are more likely to have the younger sibling decide to undertake high school studies and thus take the COMIPEMS.

1.4.3 Overview of empirical strategy and sample definition

Testing the social learning model with this sibling data requires an exogenous source of variation in the school assignment of the older sibling. The COMIPEMS school assignment mechanism provides such variation because, conditional on the older sibling's ranking of schools, his assignment depends solely on his exam score. This permits the use of a regression discontinuity (RD) design, similar to those used in prior work investigating the academic effects of school assignment in exam-based allocation regimes.[23] The basic idea behind this design is to define, for each school, the sample of older siblings who were either marginally

[21]This is a reasonably high success rate when considering that younger siblings in the early years of the sample and older siblings in the late years of the sample do not have their corresponding siblings take the exam during the sample period.

[22]Postal codes are very geographically specific in Mexico City. Students in the sample belong to more than 2,800 postal codes.

[23]See Pop-Eleches and Urquiola (2013), Abdulkadiroglu et al. (2012), Dobbie and Fryer (2011), Clark (2010), Jackson (2010), de Hoop (2012), and de Janvry et al. (this dissertation).

(barely) admitted or marginally rejected from that school, and then compare the school choices of the younger siblings in the marginally admitted and marginally rejected groups. The rest of this subsection gives the procedure for defining the "marginal" sample of older siblings for use in the RD analysis, followed by a comparison of this sample with the full set of older siblings.

The assignment process results in hard cutoff scores for each school that filled all of its seats and thus had to reject some students; this cutoff is equal to the lowest score among all admitted students. Define this cutoff as c_j for school j. (The cutoff score for a given school varies across years, but for notational simplicity in the present discussion I assume there is only one year of data.) If school k is ranked before j on student i's preference list, including if j is unlisted, we write $k \succ j$. Denote the student's exam score as s_i. Then marginal students for school j are those who:

1. listed school j as a choice;
2. had a score sufficiently close to j's cutoff score to be in a small window around the cutoff, where the window size is determined by a preselected bandwidth w: $-w \leq s_i - c_j < w$;[24]
3. scored too low to be admitted to any more-preferred school: $s_i < c_k, \forall k \succ j$;

This marginal group includes students who were rejected from j ($s_i < c_j$) and those who were admitted ($s_i \geq c_j$). Note that a student may belong to more than one school's sample.

Unless further restrictions are imposed, the sample has one undesirable yet subtle characteristic. Some students rank a school k ahead of j, where k has a cutoff score slightly above j. When this difference is smaller than the bandwidth, so that $c_j < c_k < c_j + w$, students with $s_i \geq c_k$ are missing from j's sample (because they were admitted to k) but those with $s_i < c_k$ are not (because they were rejected from k). Thus there is a sudden drop in the density of students at c_k, and the missing students probably have different unobservable characteristics than those who have the same score but remain in the sample because they did not choose k. This non-smooth change in unobservables violates the assumption of the RD design that unobservable characteristics are a smooth function of s_i.[25] A solution to this is to add one more restriction that excludes students listing a "just above the cutoff" school: marginal students for school j are those who:

4. have no more-preferred school in the $[c_j, c_j + w)$ half-window: $c_k \notin [c_j, c_j + w), \forall k \succ j$.

This condition ensures, for a student satisfying assumptions 1 through 3, that the only way his score affects inclusion in the sample is whether the score falls within the selected window. There is a disadvantage to this solution, particularly for large bandwidths. It omits students

[24]The second inequality is strict because the score variable is discrete, so this definition includes w score values too low to be admitted and w values high enough to be admitted.

[25]Abdulkadiroglu et al. (2011) recognize this problem as well, although they do not confront as closely-spaced cutoffs in their data.

listing more-preferred schools with cutoffs slightly above j's, which both reduces sample size and results in estimates of treatment effects that exclude this subsample of students. It also means that a student who is in the sample for a small bandwidth may leave the sample for a larger bandwidth, which is not the typical behavior as bandwidth increases. These are necessary sacrifices in order to satisfy the assumptions of the RD design, but the results are robust to ignoring condition 4 so these trade-offs are perhaps not very important.

One more restriction is placed on the sample, not to fulfill the assumptions of RD but to ease interpretation of the treatment effect. Any student who would be unassigned to any school for one or more scores within the window is omitted. Such students did not list any school with a cutoff score equal to or less than the lowest score in the window. We do not know if the unassigned students later chose a school from those that did not fill up or if they did not enroll at all. Our focus is on the effect of a sibling being assigned to one school or another, rather than getting into any school or going unassigned. This restriction is another reason that large bandwidth samples are less representative of the actual sample of students near a cutoff.

Comparing columns 2 and 3 of Table 1.1, we see that this restricted RD sample is quite similar to the full sample of older siblings. While the difference in means is statistically significant for all but one variable, the magnitudes of the differences are negligible for the demographic variables. There are larger differences in the school choice variables: older siblings in the RD sample are more likely to request and be assigned to schools in the elite subsystems. This is because all elite schools are oversubscribed, so students requesting them are more likely to end up near a cutoff. Older siblings are about as likely to have the same first choice as their younger siblings in the RD sample (37% probability) as in the full sibling sample (34%).

1.5 Ordinary least squares analysis

While the hypotheses derived from the social learning model are related to expected utility and are most naturally tested in a random utility discrete choice model, the OLS analysis provides well-identified, easily-interpreted evidence about the effect of sibling assignment on school choice. Much of the logic from the OLS analysis will be applied when estimating the discrete choice model as well. It is important to note that while I have put forth two possible channels through which sibling admission can affect younger siblings' choices (increased precision of beliefs and increased productive knowledge), I cannot disentangle these channels empirically. Both channels increase expected utility from a school, which is the criterion on which the student is choosing.

1.5.1 Method

For all regressions in this paper, exam score is centered to be 0 at the school's cutoff score, which (now acknowledging that there are many years of data) is different in each year t:

$\tilde{s}_{ijt} \equiv s_i - c_{jt}$. The basic parametric RD specification for a single school j in year t is as follows:

$$y_{ijt} = \delta_{jt}\text{admit}_{ijt} + f_{1jt}\left(\tilde{s}_{ijt}\right) + \text{admit}_{ijt} f_{2jt}\left(\tilde{s}_{ijt}\right) + \varepsilon_{ijt}$$

where y_{ijt} is the outcome of interest, admit_{ijt} is a dummy variable for whether $\tilde{s}_{ijt} \geq 0$, $f_{1jt}\left(\tilde{s}_{ijt}\right)$ and $f_{2jt}\left(\tilde{s}_{ijt}\right)$ are polynomials in exam score approximating the unobservables that vary with score, and ε_{ijt} is an error term. In our case, y_{ijt} is an outcome for the younger sibling, such as choosing school j as his first option, while the explanatory variables are from the older sibling, since it is the admission outcome of the latter that is hypothesized to affect the choices of the former. The parameter δ_{jt} is the local average treatment effect of the older sibling's admission to j in year t on the younger sibling's outcome for older siblings close to the cutoff, compared to the counterfactual in which the older sibling is rejected from j and admitted to the most-preferred school that would actually accept him.

There are many schools and many exam years, so it is necessary to combine the information from all oversubscribed schools in order to make statements about the average effect of admission. To do this, I stack the samples of all oversubscribed school-years and estimate the RD regression jointly. It would be preferable to include different functions f_{1jt} and f_{2jt} for each school or school-year, similar to Abdulkadiroglu et al. (2012) who include different functions for each school. But the very large number of schools makes this infeasible in most specifications, so I include only one set of polynomials, as in Pop-Eleches and Urquiola (2013).[26] Now including fixed effects for cutoff school j and older sibling's exam year t, the stacked specification is:

$$y_{ijt} = \delta\text{admit}_{ijt} + f_1\left(\tilde{s}_{ijt}\right) + \text{admit}_{ijt} f_2\left(\tilde{s}_{ijt}\right) + \mu_j + \eta_t + \varepsilon_{ijt} \tag{1.5}$$

The parameter δ is now the local average treatment effect of admission across all school-years.[27]

Lee and Card (2008) explain that when the running variable (here, exam score) is discrete, non-parametric RD methods are unsuitable. This is because there is no concept of moving infinitely close to the cutoff—to compare outcomes above and below the cutoff, it is necessary to impose a parametric form that allows extrapolation from the discrete point of support closest to the cutoff. Thus the OLS portion of this paper uses only parametric linear regressions with varying bandwidths and polynomial degrees to show the robustness of the results. Bandwidths of 3, 5, and 10 points (about 1/6, 1/4, and 1/2 of a standard

[26] An exception is when only schools from the elite systems are being considered. In that case there are only 30 schools, all with large sample sizes, so separate polynomials can be fit for each school and robustness can be assessed. Indeed, including separate polynomials has almost zero effect on the treatment effect or its standard error.

[27] Abdulkadiroglu et al. (2012) note that if the f functions were allowed to vary by school-year, then $\hat{\delta}$ would be a variance-weighted average of the $\hat{\delta}_{jt}$'s. Since the f functions do not vary here, δ is not numerically identical to the variance-weighted average, but the quantity estimated can be thought of similarly.

deviation in the population exam score distribution, respectively) are used.[28] Because there are few points of support, it is important to choose a polynomial order that fits the data adequately without overfitting. Following Lee and Lemieux (2010), I select the polynomial order that minimizes the Akaike Information Criterion (AIC).[29] Lee and Card (2008) show that when the running variable is discrete, standard errors should be clustered at the level of the running variable. Since this results in few clusters in the present application, the wild cluster bootstrap from Cameron et al. (2008) is used to obtain p-values for the coefficients of interest under the null hypothesis of zero effect.

1.5.2 Average effect of older sibling admission on school choice

The OLS RD estimates give consistent causal evidence that students are more likely to apply to a school as their first choice if an older sibling was admitted there. This evidence is robust to the choice of bandwidth and the parameterization of the running variable. The estimated impact of older sibling admission on first choice demand for the cutoff school is presented in Table 1.2 for several choices of polynomial order and bandwidth. For example, the estimated effect of older sibling admission is 6.8 percentage points in the 3-point linear specification. This estimate is large compared to the corresponding sample average of 19% choosing the cutoff school. The sibling admission-first choice relationship is illustrated graphically in Figure 1.1.

Bandwidth selection has implications beyond the usual bias-efficiency trade-off, as explained in Section 1.4.3. Recall that, for a given bandwidth, the sample only includes students who 1) would be admitted to the cutoff school for every point value above the cutoff and within the bandwidth, and 2) would not be left unassigned to any school for any point value below the cutoff and within the bandwidth. As a result, larger bandwidths exclude a significant proportion of students. A bandwidth of 3 only excludes 29% of students, almost entirely due to the "not unassigned below the cutoff" restriction. A bandwidth of 10 excludes 58% of students, with most of the additional exclusion driven by the "no other schools above the cutoff" restriction. Because estimates based on smaller bandwidths use a more representative sample of students near cutoffs, they are preferred when sample size allows for reasonably precise inference. The remaining tables report estimates based on bandwidths of 3, 5, and 10 points, while figures use a bandwidth of 5. When discussing point estimates, I

[28]Note that because of the sample selection resulting from condition 4 above, it is not possible to use a bandwidth selection algorithm (e.g. cross-validation) that gives the optimal bandwidth. This is because increasing the bandwidth causes some observations to drop out of the sample, as explained in section 1.4.3, so that the "optimal" bandwidth given by a cross-validation procedure might be undesirable if it selects many students out of the sample, reducing the representativeness of the sample. The empirical estimates will illustrate this issue.

[29]Lee and Card (2008) and Lee and Lemieux (2010) suggest another goodness-of-fit test that compares each polynomial specification to specifications that also include dummy variables for each point of support. Joint significance of the dummies implies that a higher-order polynomial may provide a better fit. Lee and Lemieux (2010) caution, however, that for small bandwidths this test is ineffective at ruling out high-order polynomials, which is a concern in this application. Hence the AIC is used here.

will focus on the 5 point bandwidth results unless otherwise noted. The point estimates of interest are of the same sign and almost always similar in magnitude, regardless of bandwidth choice.

The sibling admission effect is persistent across schools in both elite and non-elite subsystems. Table 1.3 and corresponding Figure 1.2 divide the cutoff schools into elite and non-elite groups. Column 1 reproduces the average effect on first choice demand for all cutoff schools. Column 2 shows that the admission effect for elite schools is 9.5 percentage points compared to the sample mean of 32%. Column 3 reports the effect for non-elite cutoff schools, where the dependent variable is a dummy for whether the cutoff school was the younger sibling's first non-elite choice. This is because most students choose an elite school as their first choice, so that most adjustment in non-elite preferences takes place lower in the choice list. The effect is 10.1 percentage points, compared to an average of 25%.

The admission effect for most schools is positive. Figure 1.3 shows the distribution of estimated admission coefficients, obtained by estimating the RD specification separately for each school while using a bandwidth of 5 points. Panel a gives the distribution of admission effects on first choice demand for elite schools only, which have large corresponding sample sizes (between 1,555 and 12,112 students) and thus fairly precise estimated effects. All but two of the 30 schools have positive estimated effects of admission, suggesting that the risk-reducing and productive knowledge-increasing effects of admission dominate for most schools. Panel b gives the distribution of the effect on first non-elite choice for non-elite cutoff schools with 50 or more observations in their respective RD samples. Here, estimation error overstates the variance of the distribution substantially, such that the estimated effect of admission is negative for 21% of schools. To account for this, I estimate the true variance of the coefficients, following Aaronson et al. (2007).[30] Performing this correction and assuming a normal distribution of admission coefficients, it is estimated that 8% of non-elite schools have a negative admission effect. Hence it appears that the expected utility-increasing channels dominate for most schools.[31]

The effect of older siblings' school assignment on demand does not appear to be driven by a direct effect of sibling presence on match quality. The most obvious channel through which sibling assignment could affect match is if attending school together was convenient for the student or parent, for example in traveling to and from school or attending the same school functions. But the estimated effect of admission is similar between siblings who are close enough in age to attend high school at the same time (two or fewer years apart) and siblings who are too far apart in age to attend contemporaneously. Table 1.4 shows this result.

[30] This is done by subtracting the average estimation error from the variance of the estimated coefficients: $\mathrm{E}\left[\widehat{\boldsymbol{\delta}}'\widehat{\boldsymbol{\delta}}\right] - \mathrm{E}\left[\left(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}}\right)'\left(\boldsymbol{\delta} - \widehat{\boldsymbol{\delta}}\right)\right]$.

[31] One explanation for this result, even in the absence of productive learning, is that most of the uncertainty about a school comes from imprecise beliefs about idiosyncratic match quality rather than about the school's average level of attributes. In the former case, a similar proportion of students receiving a signal from a particular school would find out that it is better for them than expected and others would find out it is worse, while in the latter case, the surprise to beliefs for students within one school would be highly correlated and would thus result in negative demand effects for some schools.

Column 1 reproduces the results for all siblings. Columns 2 and 3 report the admission effect separately for siblings who are 1-2 and 3-5 years apart, respectively. These results are shown graphically in Figure 1.4. Also reported in Table 1.4 is the estimated difference in admission effects between these two groups. This difference is small and statistically insignificant for all choices of bandwidth: the largest estimated difference in admission effects is -1.2 percentage points compared to the average effect of 7.7 percentage points. Thus it does not appear that students choose their siblings' schools simply because they want to attend the same school contemporaneously.

Furthermore, the effect of admission is not confined to demand for the older sibling's school. Consistent with Hypothesis 3, admission leads to the student ranking additional schools from the same subsystem, other than the older sibling's exact school. Table 1.5, accompanied by Figure 1.5, shows that this is the case. The sample definition here is different than in the previous analysis because it only considers students who would leave their subsystem if rejected from the cutoff school, and relaxes the "no other schools above the cutoff" restriction to "no schools from other subsystems above the cutoff" instead. The counterfactual to admission to the threshold school in this case is admission to a school in a different subsystem.

Column 1 shows that when the older sibling is admitted to a school, the younger sibling ranks on average .22 more schools in the same subsystem, excluding the older sibling's school. Columns 2 and 3 show, again, that the estimated effects are almost identical for the closely-spaced and far-apart sibling samples. The admission effect on subsystem demand and the persistence of admission effects for students far apart in age cannot be explained by a direct effect of sibling presence on match quality.

1.5.3 Effect of good versus bad surprises on school choice

The model predicts that a signal's impact on expected utility (and thus demand) depends on the sign and magnitude of the surprise to match quality. Surprises to match quality are unobserved by the econometrician, but one available proxy is an indicator for whether the older sibling graduates from high school or not. The logic for using this proxy is as follows. One contributor to dropout is a bad match between student and school. That is, there are students who will drop out from some schools but not others. Siblings are often similar in their preferences and abilities, so if the older sibling experiences a negative surprise to match quality (proxied by dropout), this suggests to the younger sibling that the school may not be a good match for him either. If younger siblings of dropouts and graduates have the same prior beliefs about match quality of the cutoff school, so that dropout represents only a *surprise* to match quality and not a reflection of the prior, then comparing the admission effect for graduates versus dropouts gives an indication of whether younger siblings change their demand in response to new information about match quality. The assumption that dropout proxies only for a surprise to match quality, rather than proxying for prior beliefs, will be examined at the end of this section.

Any estimates of differential admission effects with respect to dropout must be treated as suggestive rather than rigorously causal, because dropout is not randomly assigned (indeed, if it were, it would have no informational content for the student). Consider the following equation that will be estimated:

$$\begin{aligned} y_{ijt} = \delta \text{admit}_{ijt} + f_1\left(\tilde{s}_{ijt}\right) + \text{admit}_{ijt} f_2\left(\tilde{s}_{ijt}\right) + \mu_j + \eta_t + \\ \text{graduate}_{ijt}\left\{\alpha \text{admit}_{ijt} + g_1\left(\tilde{s}_{ijt}\right) + \text{admit}_{ijt} g_2\left(\tilde{s}_{ijt}\right) + \nu_j + \varphi_t\right\} + \varepsilon_{ijt}, \end{aligned} \tag{1.6}$$

where $\widehat{\alpha}$ is equivalent to the result from estimating the simple RD equation separately for graduates and dropouts and then taking the difference of the estimated admit coefficients. The dependent variable could be either of those used above: selecting the cutoff school as the first choice, or number of other schools chosen in the same subsystem. If dropout were randomly assigned, then $\widehat{\alpha}$ would give the additional effect of admission when the older sibling graduates. The problem arises when $cor\left(\text{graduate}_{ijt} \times \text{admit}_{ijt}, \varepsilon_{ijt}\right) \neq 0$, so that students who are *differentially* more or less likely to drop out when admitted to the cutoff school are systematically more or less likely to be emulated, or have family characteristics that affect the likelihood of choosing the cutoff school.

The empirical analysis addresses the potential issue of endogenous heterogeneous effects in three ways. First, it considers multiple samples and argues that the pattern in the results is consistent with the social learning model in which positive surprises increase demand for a school more than negative ones. Second, it controls for the older sibling's middle school grade point average, which is a significant predictor of high school dropout, and its interactions with admission and exam score. Finally, it may be that the sibling admission effect is heterogeneous with respect to the school's graduation rate or other school characteristics, not the sibling's individual graduation outcome. To control for this, separate admission coefficients are estimated for each cutoff school so that the admission-graduation interaction term gives the estimated heterogeneity due to sibling dropout conditional on cutoff school characteristics.

Keeping in mind the caveats associated with using graduation status to proxy for a surprise to match quality, as well as the data limitations in using the graduation data, Table 1.6 shows that the admission effect is heterogeneous with respect to older sibling dropout. Sample size is a problem, due to the fact that graduation data only exist for older siblings from the 2005-2007 cohorts and that graduation outcomes are missing for students at UNAM schools. This necessitates inclusion of all sibling pairs 1 to 5 years apart in age. A sibling one year below his older sibling still has most of an academic year to learn about his sibling's school, since school begins in the early fall and preference listings are not due until February or March. Although graduation has not occurred yet for the siblings who are 1 or 2 years apart, in Mexico City most dropout occurs in the first or second year and it should be apparent early on whether match quality was good or bad.

The effect of admission on same-school demand is higher when the older sibling graduates, consistent with Hypothesis 2. Column 1 gives the differential effect of admission on application to the cutoff school with respect to graduation status. The coefficient of interest is on "admission $\times$ older sib graduation" which is the additional impact of admission when the sibling graduates instead of dropping out. On average, admission has a 6.1 percentage point higher impact on first choice preference for the cutoff score when the older sibling graduates.[32] The differential effect is illustrated in Figure 1.6, Panel a. Column 2 controls for an interaction between admission and older sibling GPA while estimating each uninteracted admission coefficient separately. The estimates remain very similar (with the exception that the differential effect declines for the 10-point bandwidth and is no longer significant).

In order to explore the issue of endogenous differential dropout, column 3 estimates the impact on the first non-elite choice of students whose siblings were at the threshold of a non-elite school. This, in part, addresses the possibility that students whose older siblings are more able to graduate in the cutoff school are more likely to choose better schools. In particular, we might worry that older siblings able to graduate from elite schools are from families with high academic expectations who push the younger sibling to apply as well. Focusing on the non-elite preferences of students with siblings at non-elite cutoffs, we are likely to mitigate this confounding factor to some degree. The differential effect here is large, 11 percentage points compared to a sample mean of 24%. Graphical results are in Figure 1.6, Panel b. Adding controls in column 4, the estimates remain almost identical in magnitude and statistical significance.

The evidence for Hypothesis 4, which predicts heterogeneity in the effect on demand for other schools in the same subsystem, is weaker. Columns 5 and 6 show the differential impact on the number of other schools selected in the cutoff school's subsystem, restricting the sample to cases where the older sibling is at the margin of a subsystem. Both with and without controls, the differential effect is only statistically significant for the 10-point bandwidth. The (insignificant) estimated differential effect when including controls is .25 additional schools selected in the cutoff subsystem. Figure 1.6, Panel c illustrates this relationship. Thus, between the same-school and weaker subsystem effects, it appears that younger siblings react to signals from siblings with "good" and "bad" outcomes differently, learning about match quality and updating their choice behavior accordingly.

Is the dropout measure proxying for the prior belief about match quality rather than a surprise due to admission? That is, did admitted dropouts simply have lower match quality and know it even before admission? To examine this possibility, Table 1.7 reproduces Table 1.6 except that it estimates the effect of admission of the *younger* sibling on the *older* sibling's choices, allowing the effect to vary with the younger sibling's graduation outcome. If the differential effect is large and positive, as in the case of the older sibling effect, then graduation is capturing the prior belief about match quality, which is correlated within the family. The estimated differential effects in Table 1.7 are all smaller than their counterparts in Table 1.6, and all are statistically insignificant. Only for the 5-point bandwidth estimates of subsystem demand are the point estimates at least half as large as their Table 1.6 counterparts. These

[32]The estimates imply a smaller average effect of admission than did previous tables. This is because the UNAM cutoff schools are missing from the sample, and much of the admission effect on first choice demand comes from the elite UNAM and IPN subsystems.

results are not consistent with the graduation of one sibling serving primarily as a proxy for the prior belief of the other sibling.

1.6 Discrete choice model of school choice

In this section, the basic RD design is extended to a discrete choice model of school choice. This approach has two advantages over the OLS methods above. First, it directly tests the social learning model's hypotheses regarding the effect of peer signals on students' expected utilities. Testing these hypotheses gives more insight into the substitution patterns exhibited by students in response to new information than the OLS analysis does. In particular, the OLS evidence regarding the effect of peer signals on subsystem preferences is not tied directly to the model's hypotheses, while the discrete choice results are. Second, it allows for a natural parameterization of the impact of a peer signal: the change in willingness to travel to that school or another school in the same subsystem, which with further assumptions can then be translated into a willingess to pay measure.

1.6.1 Method

The expected utility formulation from the theoretical model in Section 1.3 can be used as the basis for a reduced form discrete choice model of school choice. Writing expected utility for younger siblings with no signal (equation 1.1) or with a signal (equation 1.2), we have:

$$U^*_{0ij} = \boldsymbol{X^0_{ij}\beta_i} - \frac{\rho}{2}\sum_{\ell}\frac{1}{\tau^0_{\ell ij}} + r_j\left(n^0_{ij}\right)$$

$$U^*_{1ij} = \widehat{\boldsymbol{X}^{\boldsymbol{1}}_{\boldsymbol{ij}}}\boldsymbol{\beta_i} - \frac{\rho}{2}\sum_{\ell}\frac{1}{\tau^0_{\ell ij}+\tau^P_{\ell ij}} + r_j\left(n^1_{ij}\right).$$

Using one equation to write a younger sibling's expected utility in either state, either with a signal from an admitted older sibling or without, we have:

$$U^*_{ij} = \alpha_{ij} + \delta_{ij}\text{admit}_{ij}$$

so that $\alpha_{ij} = U^*_{0ij}$ and $\delta_{ij} = U^*_{1ij} - U^*_{0ij}$. To estimate this model with a discrete choice framework, we can write:

$$U^*_{ij} = \delta\text{admit}_{ij} + \varepsilon_{ij}$$

where the error term ε_{ij} captures heterogeneity in the mean of the prior, its variance, initial productive knowledge, and in the peer admission effect δ_{ij} about its mean. The estimated effect of admission $\widehat{\delta}$ is biased for the same reason as in the OLS specification: a student with a sibling admitted to j probably had a more favorable prior about j than a student without an admitted sibling, due to correlated preferences and constraints within the family.

To address this bias, I apply the principles from the OLS RD design to the utility specification. The sample is restricted to students whose older siblings were close to a cutoff and specifies as the choice set all schools from which the student could choose in his exam year. To model the fact that students will have a higher preference for the older sibling's cutoff school, I add an indicator variable $cut_{ij} = 1$ when i's sibling is in school j's cutoff sample and 0 otherwise. Utility from the cutoff school is allowed to vary with respect to the older sibling's COMIPEMS exam score:

$$U_{ij}^{*} = \theta cut_{ij} + \delta \left(cut_{ij} \times \text{admit}_i \right) + f_1 \left(\tilde{s}_i \right) \text{cut}_{ij} + f_2 \left(\tilde{s}_i \right) \left(\text{cut}_{ij} \times \text{admit}_i \right) + \gamma \text{dist}_{ij} + \varepsilon_{ij}$$

where $admit_i = 1$ when the older sibling scores high enough for admission to her cutoff school, 0 otherwise, f_1 and f_2 are functions of centered exam score, and $dist_{ij}$ is the distance between student and school. Allowing expected utility to be higher or lower for cutoff schools (through θ) and for this expected utility to vary around the cutoff, δ captures only the discontinuous jump in expected utility caused by the peer crossing the cutoff and being admitted. Translating this into an easily interpreted effect, $-\delta/\gamma$ gives the average marginal willingness to travel to the cutoff school due to sibling admission.

A weakness of this specification is that it implicitly assumes that the counterfactual to admission was that the older sibling did not go to school anywhere, meaning no information was received at all. In reality, rejection from the school above the cutoff implies admission to another school below the cutoff. To model this, I define the variable $\text{below}_{ij} = 1$ when j is the school that the older sibling would or did attend upon scoring too low for admission to the cutoff school.[33] Then the specification can be expanded to include the effect of having a sibling admitted to the school below the cutoff:

$$\begin{aligned} U_{ij}^{*} = \theta \text{cut}_{ij} + \delta \left(\text{cut}_{ij} \times \text{admit}_i \right) + f_1 \left(\tilde{s}_i \right) \text{cut}_{ij} + f_2 \left(\tilde{s}_i \right) \left(\text{cut}_{ij} \times \text{admit}_i \right) \\ + \underline{\theta} \text{below}_{ij} + \underline{\delta} \left(\text{below}_{ij} \times (1 - \text{admit}_i) \right) + \underline{f}_1 \left(\tilde{s}_i \right) \text{below}_{ij} + \underline{f}_2 \left(\tilde{s}_i \right) \left(\text{below}_{ij} \times (1 - \text{admit}_i) \right) \\ + \gamma \text{dist}_{ij} + \varepsilon_{ij} \end{aligned}$$

The interpretation of $\underline{\delta}$ is analogous to δ: the average effect of admission to the "below" school on the marginal expected utility from attending there.

Incorporating subsystems into the model is straightforward. The social learning model predicts that on average, older sibling admission increases the marginal expected utility from that school's subsystem. The goal, then, is to allow marginal expected utilities to vary with older sibling admission while addressing the bias from family members having correlated preferences for subsystems. The RD approach works here as well. If there are M subsystems, let $X_j^1, \ldots X_j^M$ be dummy variables equal to 1 if school j belongs to the corresponding subsystem and 0 otherwise. Define cutsub_{ij} equal to 1 if j belongs to the

[33] Under this specification, the sample is limited to students whose older siblings had only one school below the cutoff and within the bandwidth, i.e. rejection from the cutoff school could only result in admission to a single school, no matter how many points were lost within the bandwidth. Otherwise the "below" school is not uniquely defined.

cutoff school's subsystem and belowsub_{ij} equal to 1 if j belongs to the "below" school's subsystem, 0 otherwise. Incorporating these variables into the RD specification, we have:

$$\begin{aligned} U_{ij}^* = \theta \text{cut}_{ij} + \delta \left(\text{cut}_{ij} \times \text{admit}_i\right) + f_1\left(\tilde{s}_i\right) \text{cut}_{ij} + f_2\left(\tilde{s}_i\right)\left(\text{cut}_{ij} \times \text{admit}_i\right) \\ + \underline{\theta} \text{below}_{ij} + \underline{\delta}\left(\text{below}_{ij} \times \left(1 - \text{admit}_i\right)\right) + \underline{f}_1\left(\tilde{s}_i\right) \text{below}_{ij} + \underline{f}_2\left(\tilde{s}_i\right)\left(\text{below}_{ij} \times \left(1 - \text{admit}_i\right)\right) \\ \sum_{\ell=2}^{M} X_j^{\ell}\left(\pi^{\ell} + \eta^{\ell} \text{cutsub}_{ij} + \underline{\eta}^{\ell} \text{belowsub}_{ij}\right) + \text{cutsub}_{ij}\left[\phi \text{admit}_i + h_1\left(\tilde{s}_i\right) + h_2\left(\tilde{s}_i\right) \text{admit}_i\right] + \\ \text{belowsub}_{ij}\left[\underline{\phi}\left(1 - \text{admit}_i\right) + \underline{h}_1\left(\tilde{s}_i\right) + \underline{h}_2\left(\tilde{s}_i\right)\left(1 - \text{admit}_i\right)\right] + \gamma \text{dist}_{ij} + \varepsilon_{ij}. \end{aligned} \tag{1.7}$$

This specification includes subsystem dummy variables (π^{ℓ}) and allows for marginal expected utilities to vary depending on whether the cutoff school belongs to j's subsystem (through subsystem-specific effects η^{ℓ}), the older sibling's centered exam score (h_1 and h_2), and whether the peer was admitted to the cutoff school (ϕ, the coefficient of interest, with $-\phi/\gamma$ the average marginal willingness to travel to a school in the cutoff school's subsystem due to sibling admission). The corresponding underlined coefficients are all analogous except that they apply to the subsystem of the school attended by the older sibling if she scores below the cutoff (with $\underline{\phi}/\gamma$ being the average marginal willingness to travel to a school in the "below" subsystem due to sibling admission).

If ε_{ij} is well-approximated by an i.i.d. extreme value type I distribution, then the parameters of this model can be estimated with a conditional logit. But the model implies that preferences for subsystems are heterogeneous in the population, inducing a correlated error structure. It is more appropriate to estimate a nested logit where subsystems are the nests, so that idiosyncratic preferences may be correlated within a subsystem and thus the restrictive independence of irrelevant alternatives assumption need not apply across nests.[34] Defining V_{ij} as containing all terms in equation 1.7 except ε_{ij}, the contribution to the log-likelihood function from each student i choosing school k in subsystem m is:

$$L_i = log\left(\frac{e^{V_{ik}/\lambda}\left(\sum_{j:X_j^m=1} e^{V_{ij}/\lambda}\right)^{\lambda-1}}{\sum_{\ell=1}^{M}\left(\sum_{p:X_j^{\ell}=1} e^{V_{ip}/\lambda}\right)^{\lambda}}\right),$$

where $1-\lambda$ is a measure of how correlated the error terms are for alternatives in the same subsystem. The model is estimated by maximizing this log-likelihood with standard MLE methods.

The test of Hypothesis 1 is whether $\widehat{\delta} > 0$ and $\widehat{\underline{\delta}} > 0$ (sibling admission to a school increases, on average, expected utility from attendance) and the test of Hypothesis 3 is whether $\widehat{\phi} > 0$ and $\widehat{\underline{\phi}} > 0$ (sibling admission to a school increases, on average, marginal

[34]Train (2009) points out that the nested logit is analogous (but not identical) to a mixed logit with random coefficients for each nest. This allows us to obtain some of the flexibility and enhanced realism of a mixed logit model without the computational burden of estimating a mixed logit on such a large data set.

expected utility from attending schools in the same subsystem). The tests of Hypotheses 2 and 4 are whether $\widehat{\delta}$ and $\underline{\widehat{\delta}}$, and $\widehat{\phi}$ and $\underline{\widehat{\phi}}$, are greater when the sibling graduated than when she did not. These can be tested by estimating the nested logit based on equation 1.7, including interactions of every covariate with a dummy variable equal to 1 when the older sibling graduated and 0 otherwise. Denoting each interaction term with a g superscript, the test of Hypothesis 2 is that $\widehat{\delta^g} > 0$ and $\underline{\widehat{\delta^g}} > 0$, while the test for Hypothesis 4 is that $\widehat{\phi^g} > 0$ and $\underline{\widehat{\phi^g}} > 0$.

1.6.2 Results

Estimating the discrete choice model yields direct evidence for each of the hypotheses of the social learning model. Table 1.8 provides selected estimated parameters from the nested logit specification in equation 1.7, estimated for a bandwidth of 5 with a piecewise-linear control function. An additional covariate, the mean COMIPEMS exam score of students admitted in the previous year, is also included. This is to explain some of the variance in within-subsystem preference. The sample in column 1 consists of all students from the RD sample who are 1) within the 5-point bandwidth, 2) have only one counterfactual school above the cutoff and within the bandwidth, and 3) have only one counterfactual school below the cutoff and within the bandwidth.[35]

Admission to the cutoff school increases expected utility from that school, consistent with Hypothesis 1. Similarly, rejection from the cutoff school (and thus admission to the school below the cutoff) increases the expected utility from the school below the cutoff. We can interpret these as the average marginal effect of sibling admission on willingness to travel (WTT) to that school by taking the ratio of the admission coefficient to the distance coefficient. This calculation gives an increase in WTT of 1.8 km (.247/.138) for the school above the cutoff and 3.0 km (.410/.138) for the school below the cutoff. The model does not demand this asymmetry, but it does permit it. If the younger sibling has a less precise prior on match quality for the school below the cutoff, then the peer signal will be weighted more highly and thus the average change in expected utility will be higher. This is plausible; the older sibling, whose information set is correlated with that of her younger sibling, has already ranked this school as less preferred than the school above the cutoff. One of the possible reasons for this is greater uncertainty about match quality, in addition to differences in expected match quality.

The evidence also supports Hypothesis 3: when the older sibling is on the margin between one *subsystem* and another, admission to the subsystem above the cutoff increases WTT to all other schools in that subsystem by 1.5 km (.213/.138). Admission to the system below the cutoff increases WTT to all other schools in the below subsystem by 1.7 km (.241/.138). These results support the OLS findings, which could only provide suggestive evidence on the

[35]While the correlated error structure induced by the discrete running variable (Lee and Card (2008)) is still a concern in this case, unclustered analytic standard errors are reported. Thus the reported standard errors are too large. The proper procedure for bootstrapping standard errors for MLEs with few clusters is still an open question.

subsystem effect. The total change in WTT for the cutoff school when the student is at the boundary of a subsystem is obtained by summing the effect of admission to the school with the effect for admission to the subsystem: $1.8 + 1.5 = 3.3$ km for the school above the cutoff and $3.0 + 1.7 = 4.7$ km for the school below. The intra-nest correlation parameter λ is .44, where 1 would indicate no heterogeneity in preference for subsystems (i.e. the conditional logit).

Column 2 restricts the sample to students 3 to 5 years apart, so that students do not attend high school at the same time. The estimated effects of admission decline slightly, but the coefficients of interest remain strongly significant.

Evidence for heterogeneous effects of admission with respect to graduation is provided in column 3. There is strong, albeit still suggestive, evidence for heterogeneous same-school effects (Hypothesis 2) and somewhat weaker evidence for heterogeneous subsystem effects (Hypothesis 4). The coefficients of interest are those giving the differential effect of admission by graduation status, labeled "Graduated $\times$ admission." The effect of admission on WTT to the school above the cutoff is 3.1 km (.472/.152) higher when the older sibling graduates, proxying for a positive surprise to match quality. Similarly, the WTT effect for the school below the cutoff is 3.7 km (.558/.152) higher when the sibling graduates, although these two effects are not statistically distinguishable from each other. The point estimates also suggest heterogeneous subsystem effects, although the estimates are less precise: admission to the school above the cutoff increases WTT to all schools in that subsystem by 2.2 km (.328/.152) more when the sibling graduates (p-value=.12), while the differential effect for the subsystem below the threshold is 2.9 km (.446/.152, p-value=.05).

Taken together, the results provide consistent evidence for the social learning model of school choice and agree with the OLS findings. Peer signals increase, on average, the expected utility from the peer's school and schools in the same subsystem. Better surprises to expected match quality (proxied here by sibling graduation) result in a larger increase in demand for the school and its subsystem.

1.6.3 Magnitude of estimated effects

While the OLS estimates gave the effects of sibling admission on choice probabilities and the discrete choice estimates gave effects on willingness to travel, additional assumptions will allow for further interpretation of the effect sizes. Taking the average WTT effect between the schools above and below the cutoff (1.8 km and 3.0 km, respectively), we have a 2.4 km average increase in WTT due to sibling admission. But students must travel both to and from school, so this measure should be doubled to 4.8 km/day. Students in Mexico have 195 instructional days per year, so the annualized effect on WTT is $4.8 * 195 = 936$ km/year. Translating this measure to travel time is difficult because students travel using a combination of subway, private bus, driving, and walking. Assuming that the average speed of travel over these modes during rush hour in Mexico City is 10 km/hour, then students are willing to spend $936/10 = 93.6$ additional hours per year traveling as a result of sibling admission.

The time cost estimate can be translated into a willingness to pay (WTP) estimate as well. According to the National Survey of Occupation and Employment (ENOE), the average urban teen wage is \$2/hour. Taking this as the average valuation of time for students in the estimation sample, the change in WTP is \$187/year. High school is three years long in Mexico City, so the total effect on WTP is $\$187 * 3 = \$561$. This is likely to be a conservative estimate because the WTT measures are in terms of straight line distance rather than true commuting distance, and because traveling farther may require paying an additional bus fare of about \$.50/day. As a point of comparison, median self-reported family income in the 2011 COMIPEMS student demographic survey is \$360/month. Hence it appears that the effect of sibling admission on demand is quite significant.

1.6.4 Alternative explanations

The discrete choice and OLS results are consistent with the social learning model of school choice. Are there alternative models that could explain these findings? The simplest candidate, already mentioned, is that students want to go to school with their peers (in this case, siblings). But the effect of older sibling admission on same-school preference persists when the siblings are different enough in age that they do not attend school at the same time. Nor can it be only that older peers introduce the younger students to their school's social network, because this does not explain why admission to one school in a subsystem increases demand for other schools in the same subsystem. And correlated preferences within the family cannot explain the results, because the RD design has explicitly accounted for preferences by limiting the analysis to narrow windows around the cutoffs and controlling for unobserved characteristics with polynomials in COMIPEMS score.

One might wonder whether having a sibling at a school or subsystem increases the salience of that option for students, so that they are more likely to think of that school or subsystem when writing down their preferences. But the process of school selection is one in which students have ample time to consider their options, and the stakes of their decisions are high. It is thus difficult to believe that salience is the driving factor behind the large observed effects. Also, the cutoff schools being analyzed had already been chosen by the older sibling, so it is likely that the younger sibling is aware of the school's existence whether the older sibling was admitted or rejected. Furthermore, students react differently depending on whether the sibling drops out, even though both outcomes make that school salient to the student.

Finally, could it be that younger students set expectations for their school assignment by observing their older siblings, and then choose accordingly? That is, do students who see a sibling rejected from elite schools decide that they should not even apply to them? It is unlikely. Cutoff scores for schools are public information, and students almost certainly learn their siblings' COMIPEMS exam scores. Students who just missed a cutoff are well aware of it. It is doubtful that a student would see his sibling miss admission by one point and then decide he has no chance at admission himself. Moreover, there is no penalty to applying to a high-cutoff school, so even a discouraged student has no reason not to try. Combined with the result that the effects persist even in non-elite, lower-cutoff schools, these arguments cast doubt on such an explanation.

1.7 Validity checks

This section presents two standard checks for the validity of the RD design. Both provide evidence that the design produces valid inference.

The first check is a visual inspection of whether the density of the running variable (centered COMIPEMS score) suddenly increases or decreases as it crosses the cutoff, as suggested by McCrary (2008). This might occur if the younger siblings of rejected students were less likely to apply to high school, for example if rejected students were more likely to drop out of school and younger siblings followed that example. Another, less likely possibility is that admission induces behavior that makes it impossible to match siblings to each other, such as changing their phone number or middle school. Figure 1.7 shows the density for a bandwidth of 5 for the RD sample of older siblings (corresponding to column 1 of Table 1.3). There is no clear change in density across the threshold, and indeed the density is nearly uniform over this domain. It does not seem that admission to the cutoff school has any effect on high school application behavior or matching success.[36]

The second check is to repeat the RD OLS regression, this time using exogenous student characteristics as the dependent variables. Imbens and Lemieux (2008) propose this as a way of verifying that exogenous characteristics do not suddenly change at the cutoff (which would call into question whether the endogenous variable would be balanced in the absence of a treatment effect). In order to jointly test that the admission coefficient is zero for all tested exogenous characteristics of the older sibling, seemingly unrelated regression (Zellner (1962)) is used.[37] Table 1.9 shows the results of these regressions for each of the chosen bandwidths. Only one of the admission coefficients is statistically significant at the 10% level and in no specification are the admission coefficients jointly significantly different from zero. The point estimates are quite precise as well, ruling out even fairly small covariate imbalances. Thus both checks yield support for the validity of the RD design.

1.8 Conclusion

This paper finds strong evidence for a model of school choice in which peer networks play an important role in overcoming incomplete information about match quality and building productive knowledge. Having an older sibling at a particular school increases revealed preference not only for that school, but also for other schools in the same subsystem. This

[36]While McCrary (2008) presents a formal test for a jump in density at the threshold, his non-parametric approach does not apply well to the present case of a discrete running variable with relatively few points of support. Nevertheless, the visual evidence is compelling in this case.

[37]Unclustered standard errors are reported, meaning that (in expectation) the null hypothesis of no effect is rejected too often.

relationship persists even when the sibling is no longer in attendance, showing that it does not result from a direct benefit from contemporaneously attending the same school. More positive surprises result in more positive effects on demand, consistent with learning about match quality.

What policy lessons can be taken from this result? For example, while selective school application has not been a particular focus of this paper (and indeed, elite school application rates are quite high in Mexico City), we may wonder what these results suggest for policy-makers hoping to encourage such behavior in other contexts. One lesson is that aggregate school-level information is not a perfect substitute for the more subjective, individually-tailored information that students currently obtain from their networks. Match quality for the average student may already be known in the population, but idiosyncratic match quality is uncertain and cannot be ascertained from high-level data. Providing individualized information on match quality, some of which might also be ex post productive, is not a trivial task for individual schools (as in the case of colleges) or public school systems. One approach already being undertaken at the tertiary level is to deploy the school's alumni network to connect with prospective students, providing them with personalized information through informal meetings and repeated electronic communication. But, as pointed out in Hoxby and Turner (2013), such labor-intensive interventions are expensive. Recruitment offices also have a role to play, if they can provide the kinds of individual-specific information desired by students. Public school systems face the challenge of providing individualized information about all member schools. Furnishing printed material containing data beyond school-level aggregates is one way to begin.

The findings offer a mixed appraisal of school choice mechanisms. On the negative side, it appears that the correlation observed by Hoxby and Avery (2012) is indeed causal, at least in this context. Students with a low concentration of peers attending a particular school or set of schools are less likely to apply there, when under full information they might do so. But this is also an endorsement of school choice, because it acknowledges a key rationale for its existence: students have access to a wealth of relevant information, some from their peer networks, that administrators cannot hope to internalize themselves. School choice allows students to put all of this information to work in the matching process. Creative policies that augment the information set of students in disadvantaged peer networks may help to retain the positive features of choice mechanisms while lowering the informational barriers that reduce their effectiveness.

1.9 Appendix: Derivation of model hypotheses

The following are derivations of the hypotheses presented in sections 1.3.4 and 1.3.5.

Hypothesis 1: $\mathrm{E}_{ij}\left[U^*_{1ij} - U^*_{0ij}\right] > 0$.

Proof: Equation 1.3 gives the expected change, over all students and schools, in expected utilities when a signal is received. The increase in productive knowledge r_j clearly increases expected utility, so I suppress the r_j terms here. This expectation is:

$$\begin{aligned}
\mathrm{E}_{ij}\left[U^*_{1ij} - U^*_{0ij}\right] &= \mathrm{E}_{ij}\left[\left(\widehat{\boldsymbol{X}}^{\mathbf{1}}_{\boldsymbol{ij}} - \boldsymbol{X}^{\mathbf{0}}_{\boldsymbol{ij}}\right)\boldsymbol{\beta_i} + \frac{\rho}{2}\sum_{\ell}\frac{\beta^2_{\ell i}\tau^P_{\ell ij}}{\tau^0_{\ell ij}\left(\tau^0_{\ell ij} + \tau^P_{\ell ij}\right)}\right] \\
&= \mathrm{E}_{ij}\left[\widehat{\boldsymbol{X}}^{\mathbf{1}}_{\boldsymbol{ij}}\boldsymbol{\beta_i}\right] - \mathrm{E}_{ij}\left[\boldsymbol{X}^{\mathbf{0}}_{\boldsymbol{ij}}\boldsymbol{\beta_i}\right] + \frac{\rho}{2}\sum_{\ell}\mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}\left(\tau^0_{\ell ij} + \tau^P_{\ell ij}\right)}\right].
\end{aligned}$$

From the definition of $\widehat{\boldsymbol{X}}^{\mathbf{1}}_{\boldsymbol{ij}}$:

$$\begin{aligned}
\mathrm{E}_{ij}\left[\widehat{\boldsymbol{X}}^{\mathbf{1}}_{\boldsymbol{ij}}\boldsymbol{\beta_i}\right] &= \mathrm{E}_{ij}\left[\sum_{\ell}\beta_{\ell i}\frac{\tau^0_{\ell ij}X^0_{\ell ij} + \tau^P_{\ell ij}P_{\ell ij}}{\tau^0_{\ell ij} + \tau^P_{\ell ij}}\right] \\
&= \sum_{\ell}\left\{\mathrm{E}_{ij}\left[\frac{\tau^0_{\ell ij}}{\tau^0_{\ell ij} + \tau^P_{\ell ij}}\beta_{\ell i}X^0_{\ell ij}\right] + \mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij} + \tau^P_{\ell ij}}\beta_{\ell i}P_{\ell ij}\right]\right\} \\
&= \sum_{\ell}\left\{\mathrm{E}_{ij}\left[\frac{\tau^0_{\ell ij}}{\tau^0_{\ell ij} + \tau^P_{\ell ij}}\beta_{\ell i}X^0_{\ell ij}\right] + \mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij} + \tau^P_{\ell ij}}\beta_{\ell i}\left(X_{\ell ij} + \varepsilon_{\ell ij}\right)\right]\right\} \\
&= \sum_{\ell}\left\{\mathrm{E}_{ij}\left[\frac{\tau^0_{\ell ij}}{\tau^0_{\ell ij} + \tau^P_{\ell ij}}\beta_{\ell i}X^0_{\ell ij}\right] + \mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij} + \tau^P_{\ell ij}}\beta_{\ell i}\left(X_{\ell ij}\right)\right]\right\} \\
&= \sum_{\ell}\left\{\mathrm{E}_{ij}\left[\frac{\tau^0_{\ell ij}}{\tau^0_{\ell ij} + \tau^P_{\ell ij}}\beta_{\ell i}X^0_{\ell ij}\right] + \mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij} + \tau^P_{\ell ij}}\beta_{\ell i}X^0_{\ell ij}\right]\right\} \\
&= \sum_{\ell}\mathrm{E}_{ij}\left[\beta_{\ell i}X^0_{\ell ij}\right] \\
&= \mathrm{E}_{ij}\left[\boldsymbol{X}^{\mathbf{0}}_{\boldsymbol{ij}}\boldsymbol{\beta_i}\right].
\end{aligned}$$

Substituting this result back into the original equation, we have:

$$\begin{aligned}
\mathrm{E}_{ij}\left[U^*_{1ij} - U^*_{0ij}\right] &= \mathrm{E}_{ij}\left[\widehat{\boldsymbol{X}}^{\mathbf{1}}_{\boldsymbol{ij}}\boldsymbol{\beta_i}\right] - \mathrm{E}_{ij}\left[\boldsymbol{X}^{\mathbf{0}}_{\boldsymbol{ij}}\boldsymbol{\beta_i}\right] + \frac{\rho}{2}\sum_{\ell}\mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}\left(\tau^0_{\ell ij} + \tau^P_{\ell ij}\right)}\right] \\
&= \mathrm{E}_{ij}\left[\boldsymbol{X}^{\mathbf{0}}_{\boldsymbol{ij}}\boldsymbol{\beta_i}\right] - \mathrm{E}_{ij}\left[\boldsymbol{X}^{\mathbf{0}}_{\boldsymbol{ij}}\boldsymbol{\beta_i}\right] + \frac{\rho}{2}\sum_{\ell}\mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}\left(\tau^0_{\ell ij} + \tau^P_{\ell ij}\right)}\right] \\
&= \frac{\rho}{2}\sum_{\ell}\mathrm{E}_{ij}\left[\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}\left(\tau^0_{\ell ij} + \tau^P_{\ell ij}\right)}\right] > 0
\end{aligned}$$

where the inequality holds because the τ and ρ terms are all positive by definition. ■

Hypothesis 2: All else equal, $\frac{\partial\left(U^*_{1ij}-U^*_{0ij}\right)}{\partial \boldsymbol{P_{ij}\beta_i}} > 0$.

Proof: Treating $\boldsymbol{\beta_i}$ and $\boldsymbol{X^0_{ij}}$ as fixed:

$$\frac{\partial\left(U^*_{1ij}-U^*_{0ij}\right)}{\partial \boldsymbol{P_{ij}\beta_i}} = \sum_{\ell}\frac{\partial\left(U^*_{1ij}-U^*_{0ij}\right)}{\partial P_{\ell ij}\beta_{\ell i}} = \sum_{\ell}\frac{\partial U^*_{1ij}}{\partial P_{\ell ij}\beta_{\ell i}}$$
$$= \sum_{\ell}\frac{\partial}{\partial P_{\ell ij}\beta_{\ell i}}\left(\beta_{\ell i}\frac{\tau^0_{\ell ij}X^0_{\ell ij}+\tau^P_{\ell ij}P_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}}\right) = \sum_{\ell}\frac{\tau^P_{\ell ij}}{\tau^0_{\ell ij}+\tau^P_{\ell ij}} > 0$$

where the inequality holds because the τ terms are positive. ■

Hypothesis 3: indexing the peer's school by j and fixing another school k_j in j's subsystem s_j, $\mathrm{E}_{ij}\left[U^*_{1ik_js_j}-U^*_{0ik_js_j}\right] > 0$.

Proof: This is almost identical to the proof for Hypothesis 1, except that the student is only receiving information about the shared attribute μ_{is}. From equation 1.4, again excluding the effect of productive knowledge, the expectation of the change in expected utility from any other school in the same subsystem is:

$$\mathrm{E}_{ij}\left[U^*_{1ik_js_j}-U^*_{0ik_js_j}\right] = \mathrm{E}_{ij}\left[\left(\widehat{\mu}^1_{is}-\mu^0_{is}\right)\right] + \mathrm{E}_{ij}\left[\frac{\rho}{2}\frac{\tau^q_{is}}{\tau^\mu_{is}\left(\tau^\mu_{is}+\tau^q_{is}\right)}\right].$$

Using the steps from the proof of Hypothesis 1, we have that $\mathrm{E}_{ij}\left[\widehat{\mu}^1_{is}\right] = \mathrm{E}_{ij}\left[\mu^0_{is}\right]$. So:

$$\mathrm{E}_{ij}\left[\left(\widehat{\mu}^1_{is}-\mu^0_{is}\right)\right] + \mathrm{E}_{ij}\left[\frac{\rho}{2}\frac{\tau^q_{is}}{\tau^\mu_{is}\left(\tau^\mu_{is}+\tau^q_{is}\right)}\right]$$
$$= \mathrm{E}_{ij}\left[\left(\mu^0_{is}-\mu^0_{is}\right)\right] + \mathrm{E}_{ij}\left[\frac{\rho}{2}\frac{\tau^q_{is}}{\tau^\mu_{is}\left(\tau^\mu_{is}+\tau^q_{is}\right)}\right]$$
$$= \mathrm{E}_{ij}\left[\frac{\rho}{2}\frac{\tau^q_{is}}{\tau^\mu_{is}\left(\tau^\mu_{is}+\tau^q_{is}\right)}\right] > 0$$

where the inequality holds because the τ and ρ terms are all positive. ■

Hypothesis 4: Suppose that schools j and k are in the same subsystem s. Then all else equal, $\frac{\partial\left(U^*_{1iks}-U^*_{0iks}\right)}{\partial q_{is}} > 0$.

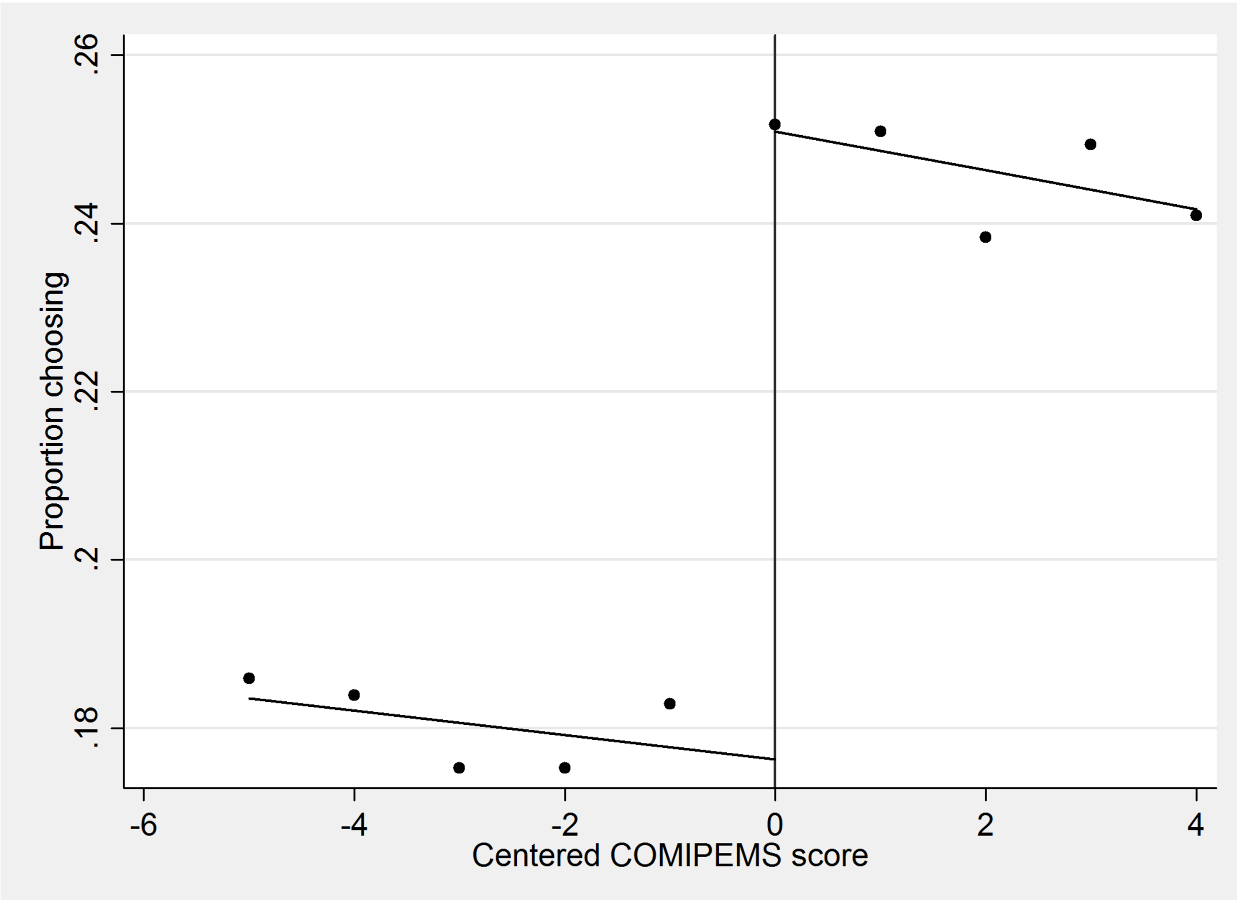
Proof: Treating μ^0_{is} as fixed:

$$\frac{\partial\left(U^*_{1iks}-U^*_{0iks}\right)}{\partial q_{is}} = \frac{\partial U^*_{1iks}}{\partial q_{is}} = \frac{\partial\widehat{\mu}^1_{is}}{\partial q_{is}} = \frac{\partial}{\partial q_{is}}\left(\frac{\tau^\mu_{is}\mu^0_{is}+\tau^q_{is}q_{is}}{\tau^\mu_{is}+\tau^q_{is}}\right) = \frac{\tau^q_{is}}{\tau^\mu_{is}+\tau^q_{is}} > 0$$

where the inequality holds because the τ terms are positive. ■

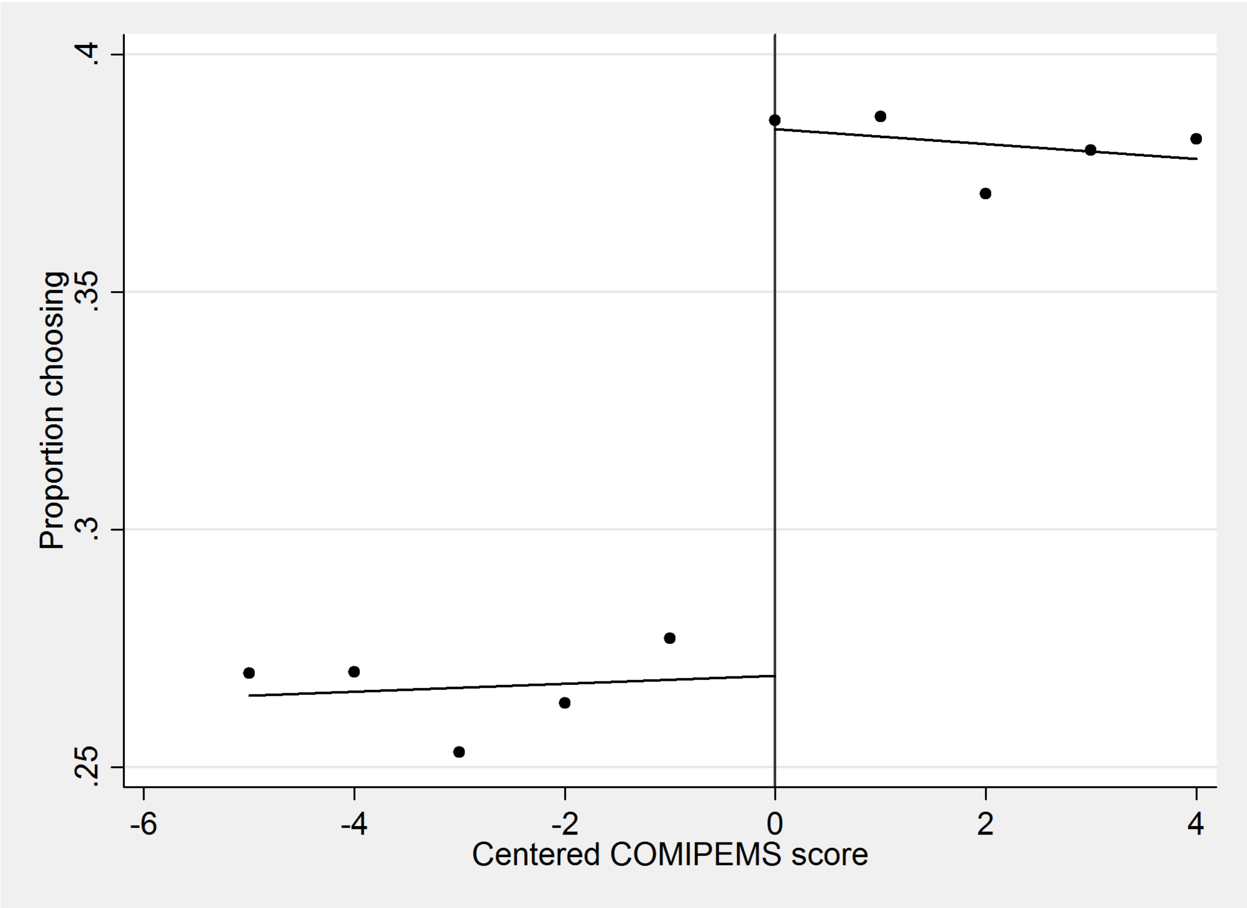
1.10 Figures

Figure 1.1: Effect of older sibling admission on younger sibling's first choice preference for same school



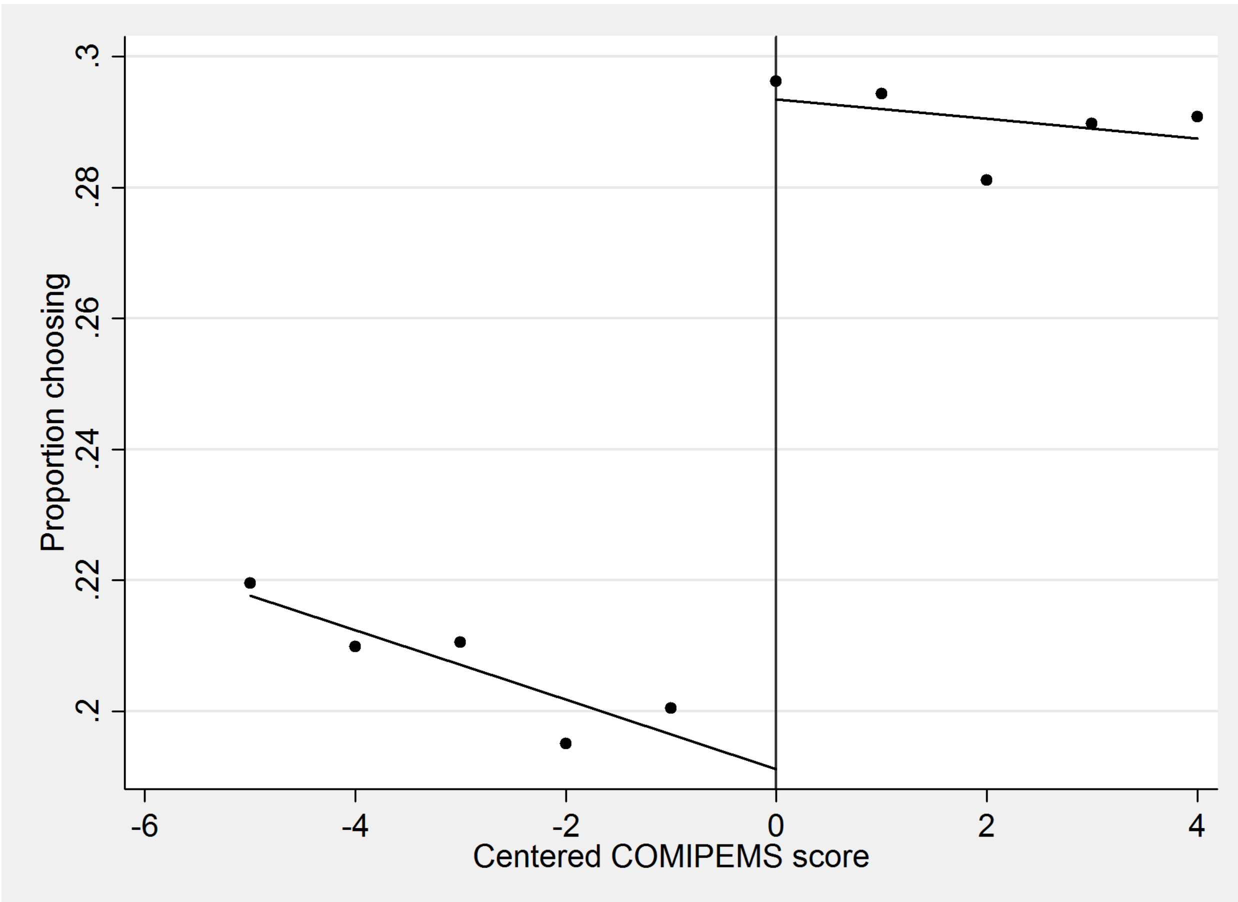
Variable on vertical axis is proportion of students listing their older siblings' cutoff school as their first choice. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

Figure 1.2a: Effect of older sibling admission, elite cutoffs: effect on first choice



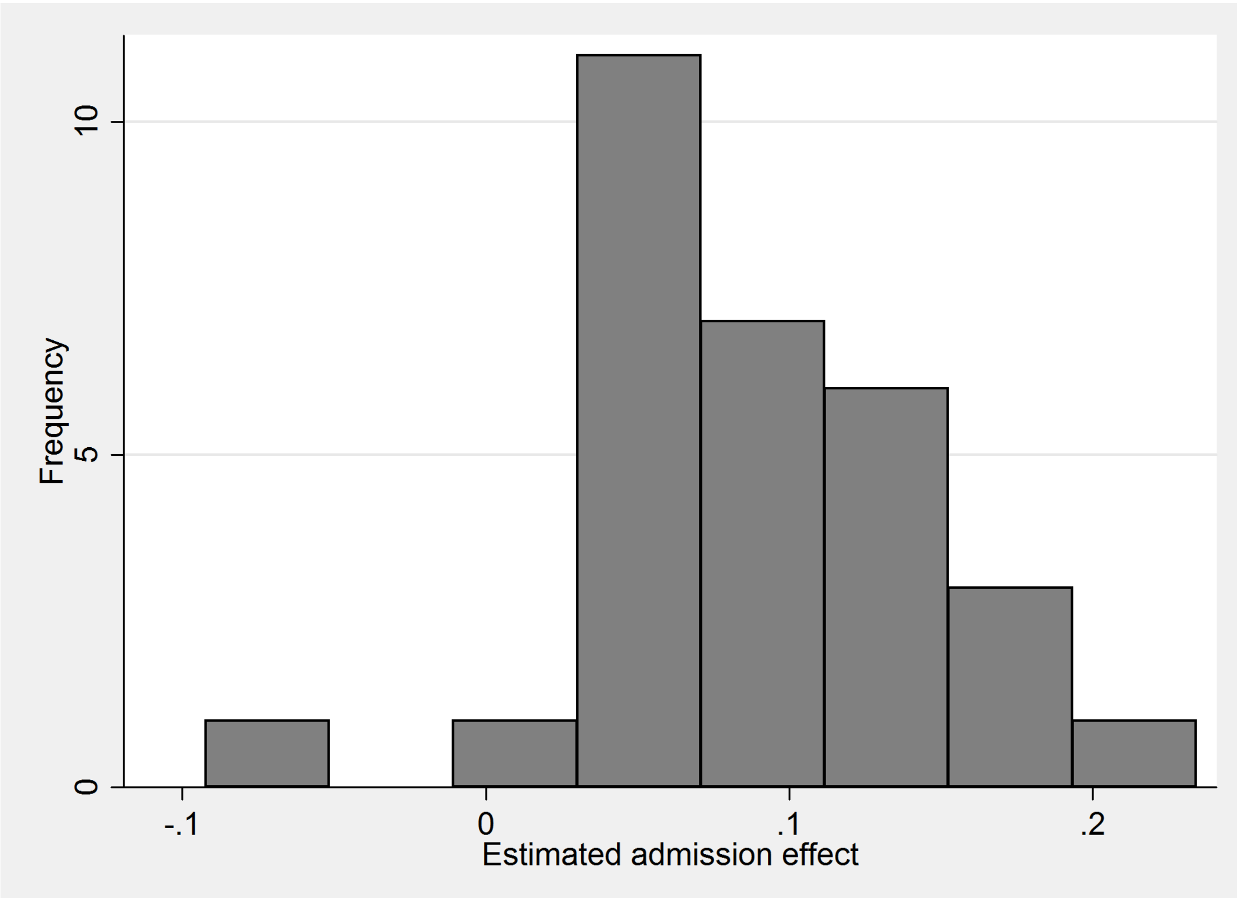
Variable on vertical axis is proportion of students listing their older siblings' cutoff school as their first choice or first non-elite choice. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

Figure 1.2b: Effect of older sibling admission, non-elite cutoffs: effect on first non-elite choice



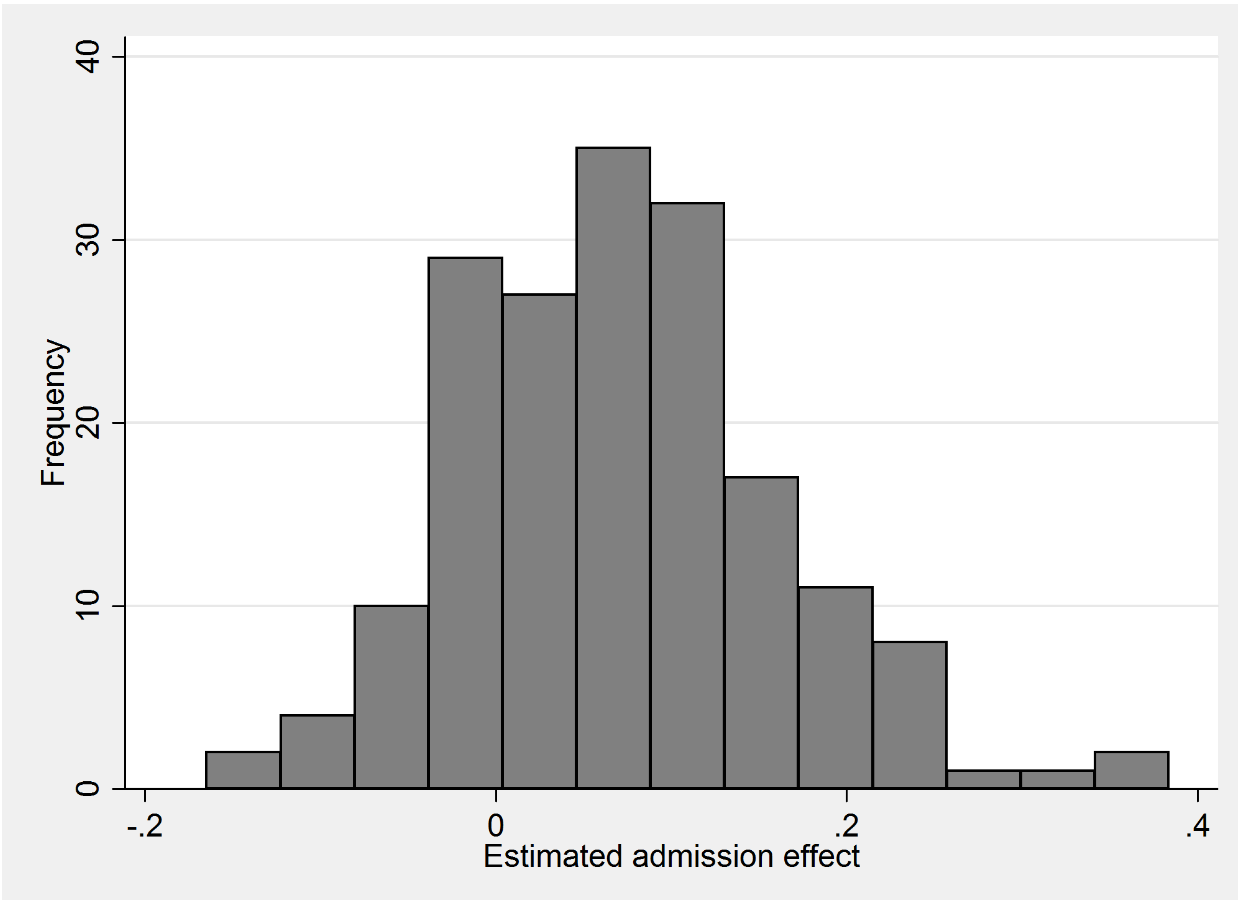
Variable on vertical axis is proportion of students listing their older siblings' cutoff school as their first choice or first non-elite choice. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

Figure 1.3a: Distribution of estimated admission coefficients, elite cutoffs: effect on first choice



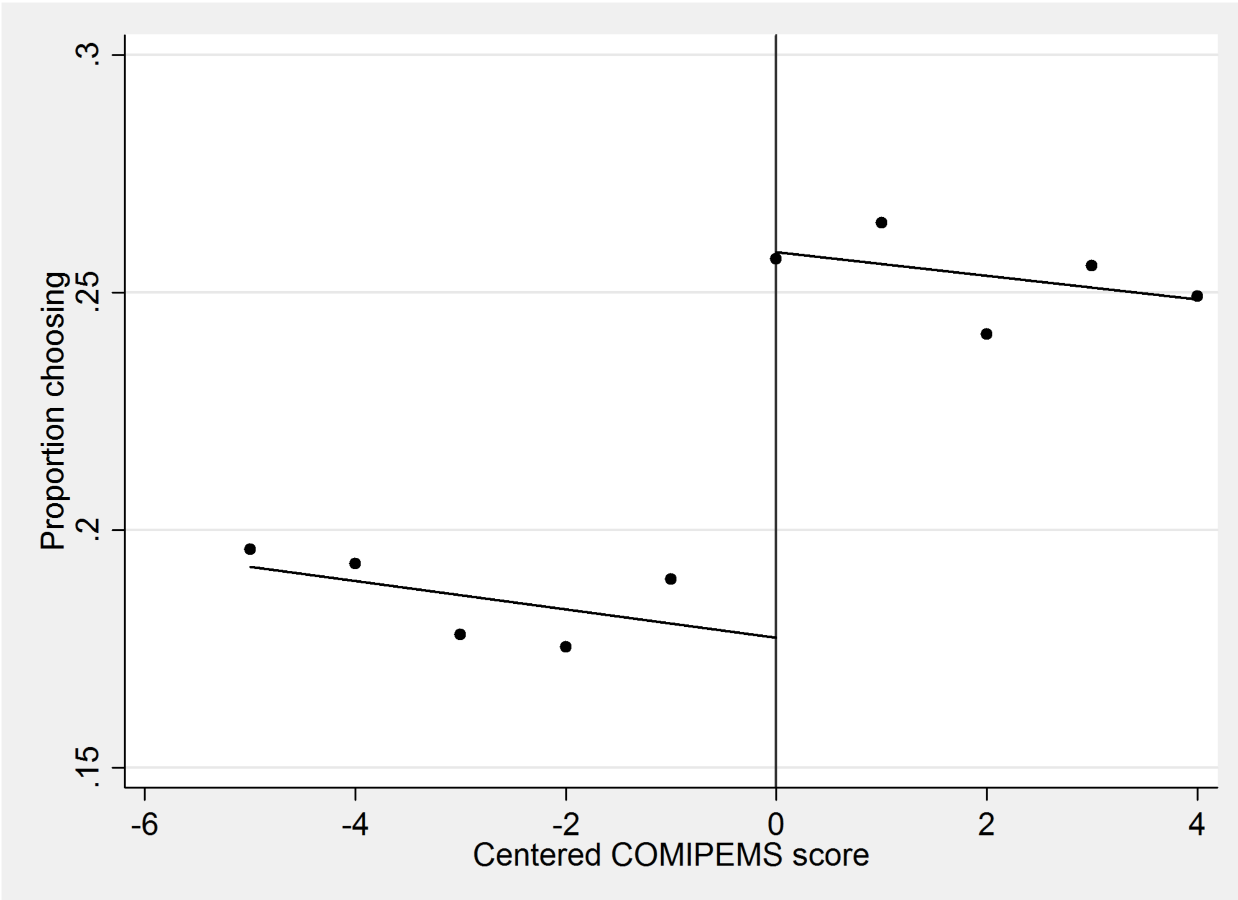
Histogram is of estimated coefficients on older sibling admission, estimated from separate regressions for each cutoff school with bandwidth of 5. Coefficients are for elite cutoff schools.

Figure 1.3b: Distribution of estimated admission coefficients, non-elite cutoffs: effect on first non-elite choice



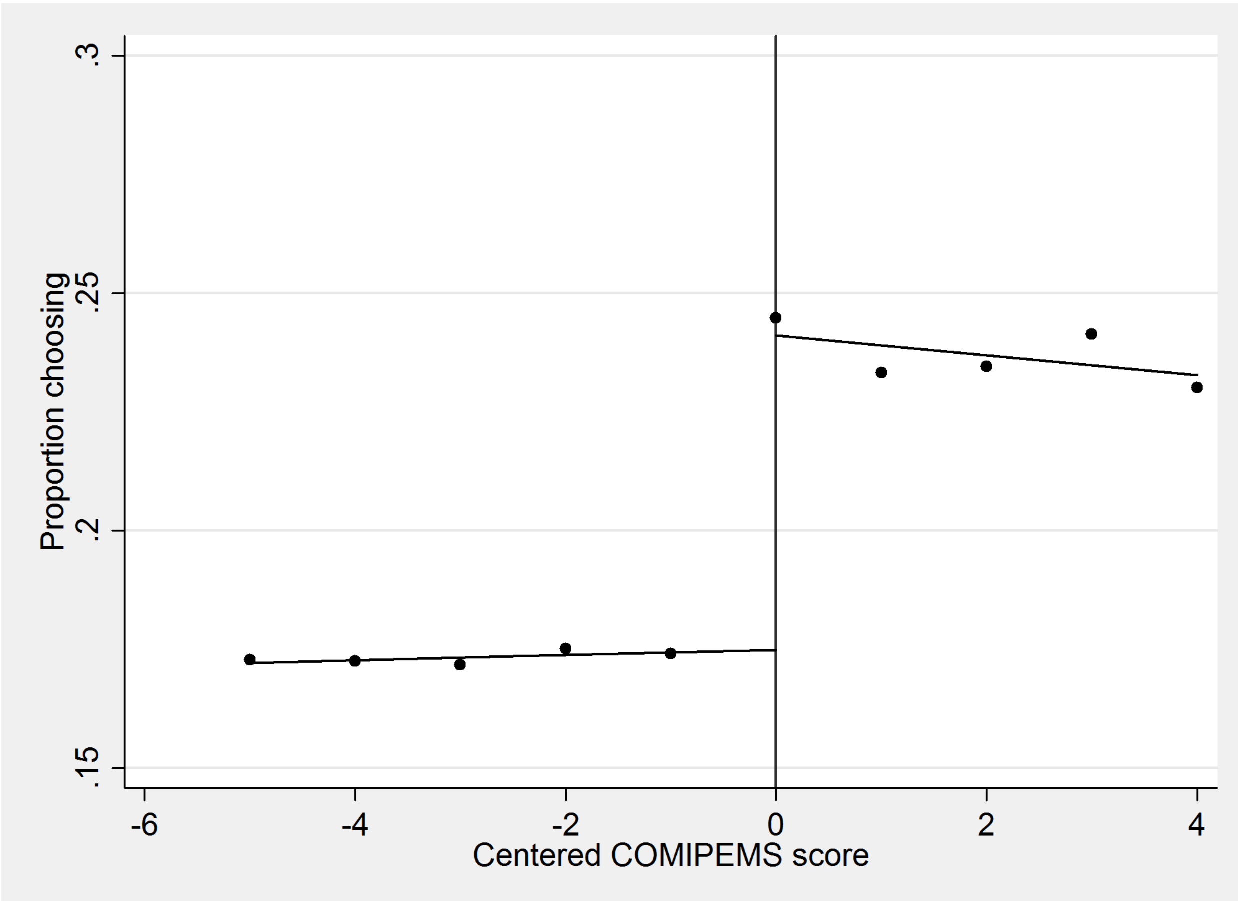
Histogram is of estimated coefficients on older sibling admission, estimated from separate regressions for each cutoff school with bandwidth of 5. Coefficients are for non-elite cutoff schools with at least 50 observations.

Figure 1.4a: Effect of older sibling admission, siblings 1 or 2 years apart



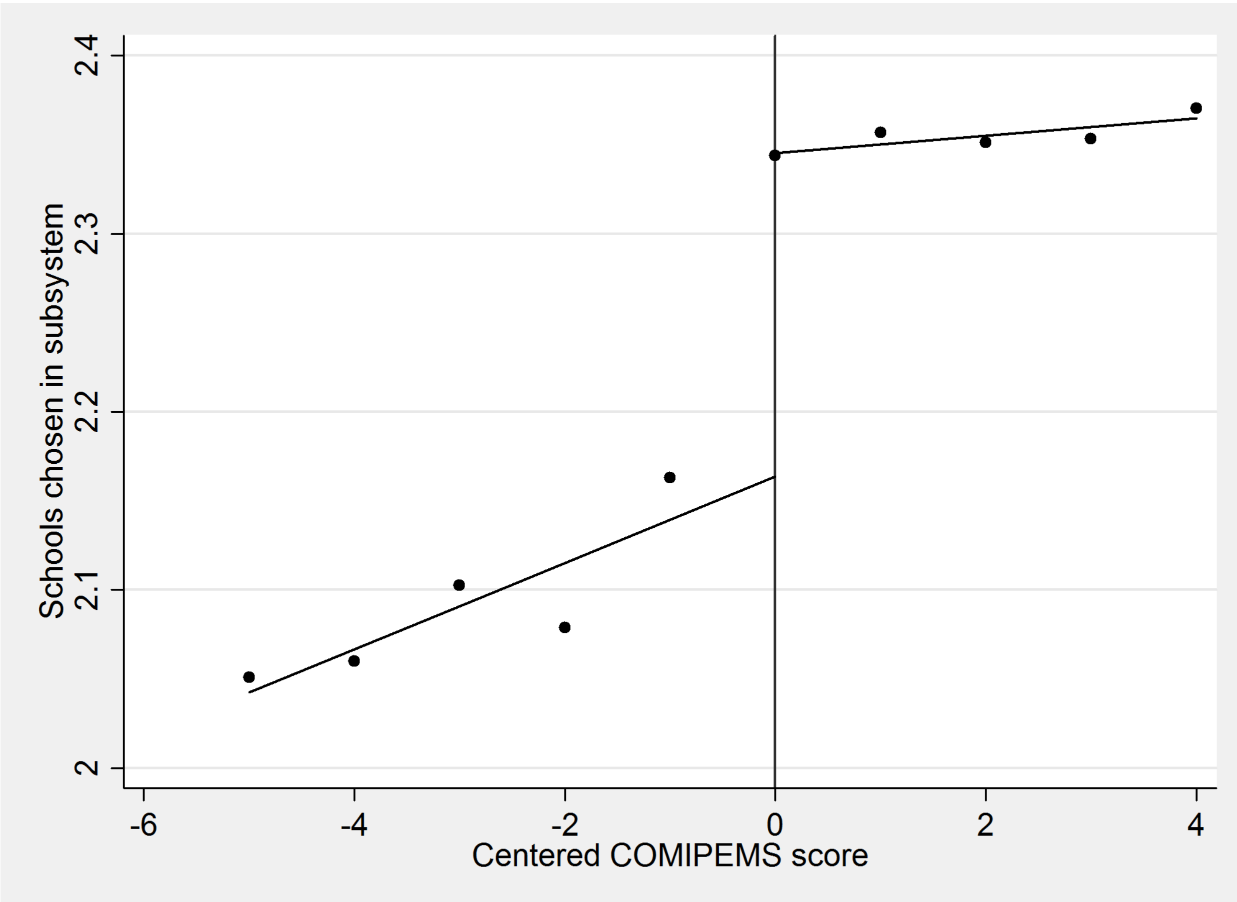
Variable on vertical axis is proportion of students listing their older siblings' cutoff school as their first choice. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

Figure 1.4b: Effect of older sibling admission, siblings 3 to 5 years apart



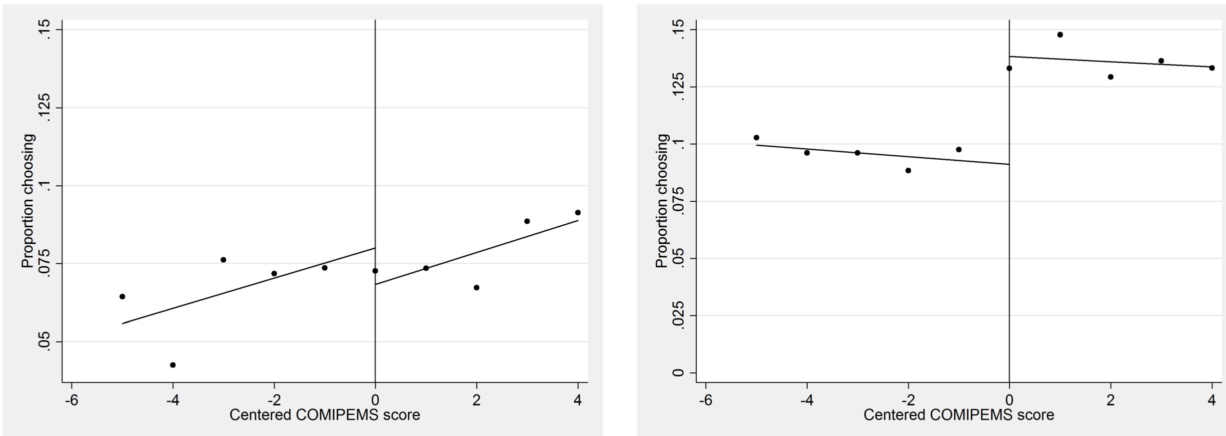
Variable on vertical axis is proportion of students listing their older siblings' cutoff school as their first choice. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

Figure 1.5: Effect of older sibling admission on number of other schools selected in the cutoff school's subsystem



Variable on vertical axis is number of schools selected in the subsystem to which the older sibling's cutoff school belongs, excluding the cutoff school. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

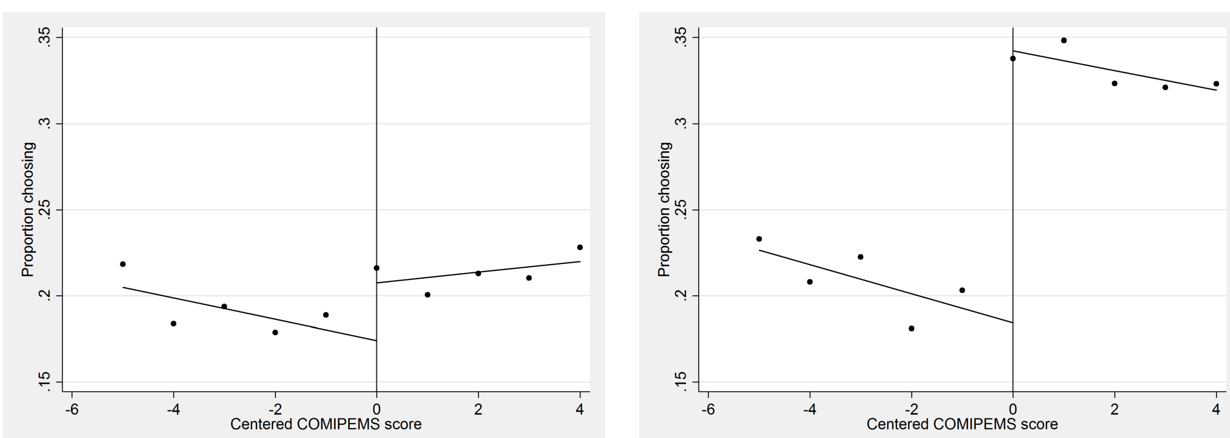
Figure 1.6a: Effect of older sibling admission, by graduation outcome: effect on first choice



(i) Older sibling did not graduate (ii) Older sibling graduated

Variable on vertical axis is proportion of students listing their older siblings' cutoff school as their first choice. Cutoff schools from the UNAM subsystem are excluded because there is no proxy for graduation available from them. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

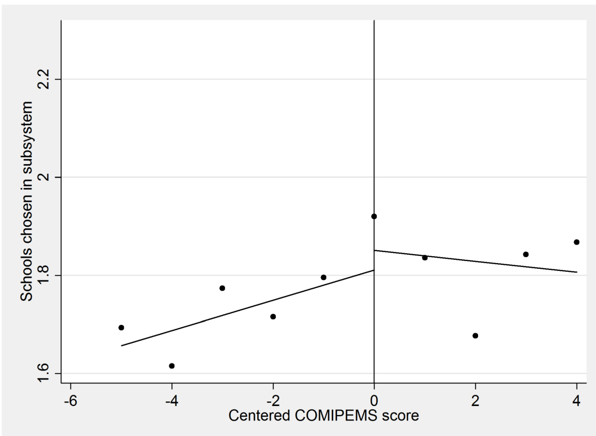
Figure 1.6b: Effect of older sibling admission, by graduation outcome: effect on first non-elite choice



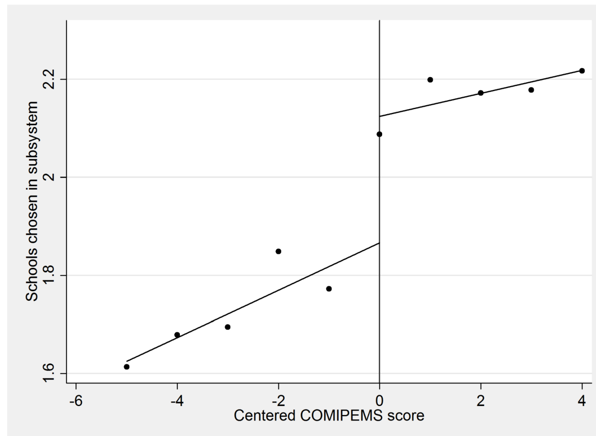
(i) Older sibling did not graduate (ii) Older sibling graduated

Variable on vertical axis is proportion of students listing their older siblings' cutoff school as their top non-elite choice. Only non-elite cutoff schools are included. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

Figure 1.6c: Effect of older sibling admission, by graduation outcome: effect on number of other schools selected in the cutoff school's subsystem



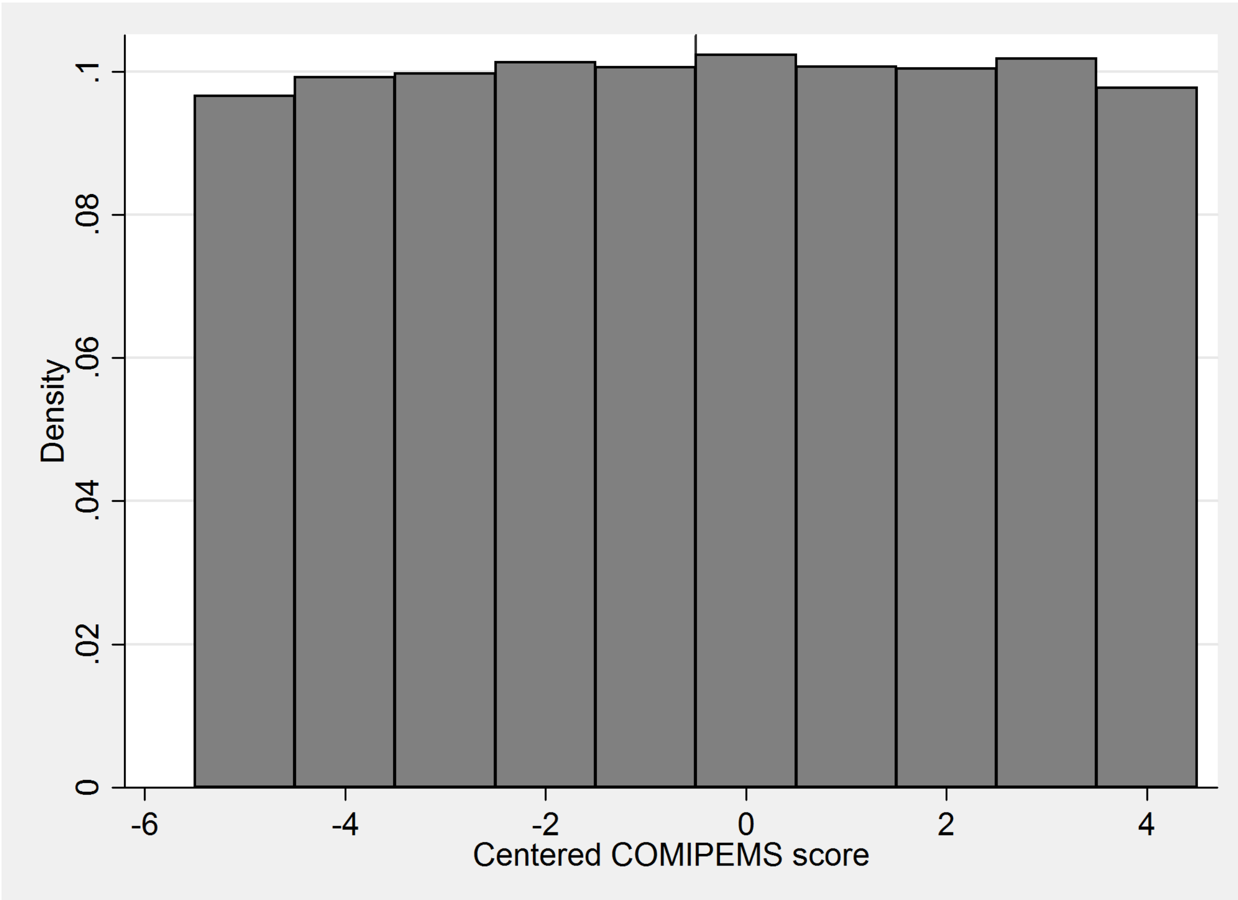
(i) Older sibling did not graduate



(ii) Older sibling graduated

Variable on vertical axis is number of schools selected in the subsystem to which the older sibling's cutoff school belongs, excluding the cutoff school. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

Figure 1.7: Density of centered COMIPEMS score about discontinuity



Histogram is of COMIPEMS score for students near a cutoff. Scores are centered so that they are 0 at the cutoff score.

1.11 Tables

Table 1.1: Summary statistics for full, sibling, and regression discontinuity samples

	(1)	(2)	(3)		
	All students	Older siblings	Older siblings in any RD sample[a]	p-value for equality of (1) and (2)	p-value for equality of (2) and (3)
Male	0.46	0.46	0.45	0.00	0.00
Maximum of mother's and father's education (years)	10.17	10.68	10.84	0.00	0.00
	(3.45)	(3.33)	(3.33)		
Number of siblings	2.20	2.19	2.16	0.12	0.00
	(1.45)	(1.15)	(1.12)		
Birth order (1 is first-born)	2.17	1.61	1.58	0.00	0.00
	(1.35)	(0.96)	(0.94)		
Hours studied per week	4.92	5.16	5.24	0.00	0.00
	(3.21)	(3.30)	(3.31)		
Middle school grade point average (of 10)	8.07	8.23	8.27	0.00	0.00
	(0.82)	(0.83)	(0.80)		
Number of schools ranked	9.28	9.23	9.89	0.00	0.00
	(3.73)	(3.76)	(3.79)		
Elite school as first choice	0.63	0.66	0.75	0.00	0.00
Assigned to an elite school	0.22	0.30	0.41	0.00	0.00
Comipems examination score	62.34	66.50	66.84	0.00	0.00
	(18.92)	(18.83)	(14.88)		
Distance from student's home to first choice school (km)	7.52	7.59	7.98	0.00	0.00
	(6.05)	(5.96)	(5.96)		
Distance from student's home to first non-elite choice school (km)	5.86	5.69	5.84	0.00	0.00
	(5.14)	(4.88)	4.90		
Grade year difference between siblings		2.51	2.51		0.25
		(1.25)	(1.25)		
Siblings chose same first choice school		0.34	0.37		0.00
Graduated high school (for non-UNAM students in 2005-2007 cohorts)	0.45	0.55	0.57	0.00	0.00
Observations	2,127,375	267,748	81,434		

Note. Standard deviations in parentheses.

[a] The RD sample is the set of older siblings who meet the RD sample definition for a bandwidth of 5.

Table 1.2: Effect of older sibling admisson on younger sibling's first choice preference for same school

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Bandwidth	3	4	5	6	7	8	9	10
Admission coefficient								
Order of polynomial:								
1	0.068*	0.075*	0.077***	0.082***	0.088***	0.091***	0.092***	0.094***
	(.10)	(.06)	(.00)	(.01)	(.00)	(.00)	(.00)	(.00)
2	0.074*	0.064*	0.069***	0.073***	0.079***	0.080***	0.090***	0.096***
	(.06)	(.06)	(.01)	(.00)	(.00)	(.00)	(.00)	(.00)
3		0.085	0.061	0.074**	0.071***	0.073**	0.075***	0.074***
		(.24)	(.13)	(.03)	(.01)	(.02)	(.00)	(.00)
Mean of dependent variable	0.194	0.204	0.214	0.225	0.234	0.245	0.257	0.265
Proportion of observations lost to sample selection	0.294	0.341	0.393	0.446	0.488	0.525	0.557	0.583
AIC-optimal polynomial order	1	1	1	1	1	2	2	3
R^2 for AIC-optimal model	0.200	0.200	0.206	0.215	0.219	0.223	0.226	0.228
Observations	61,938	76,990	88,222	96,191	103,193	108,978	113,741	118,177

Note. Dependent variable is 0/1 for whether the younger sibling chose the school above the cutoff, i.e. the school attended by the older sibling if he scores at or above the cutoff score. Regressions include piecewise polynomial terms of the order indicated in centered test score, fixed effects for cutoff school, older sibling's exam year, and number of years between siblings' exam years. Admission coefficients correspond to the bandwidth given in the column header and the polynomial order given in the corresponding row. AIC-polynomial order is the polynomial order that minimizes the Akaike Information Criterion for that bandwidth. **Bootstrapped p-values accounting for clustering at the centered score level are in parentheses.**

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.3: Effect of older sibling admission on younger sibling's preference for same school, disaggregated by type of cutoff school

	(1)	(2)	(3)
	All cutoff schools	Elite cutoff schools: dependent variable is first choice	Non-elite cutoff schools: dependent variable is first non-elite choice
Bandwidth=3			
Admission coefficient	0.068*	0.096***	0.101**
	(.07)	(.00)	(.02)
Mean of dependent variable	0.19	0.30	0.24
R^2	0.20	0.15	0.13
Observations	61,938	32,436	29,492
AIC-optimal polynomial order	1	1	1
Bandwidth=5			
Admission coefficient	0.077***	0.095*	0.101***
	(.00)	(.10)	(.00)
Mean of dependent variable	0.21	0.32	0.25
R^2	0.21	0.15	0.13
Observations	88,222	47,021	41,188
AIC-optimal polynomial order	1	2	1
Bandwidth=10			
Admission coefficient	0.074***	0.097***	0.112***
	(.00)	(.00)	(.00)
Mean of dependent variable	0.23	0.39	0.28
R^2	0.27	0.16	0.12
Observations	118,177	65,353	52,814
AIC-optimal polynomial order	3	3	2

Note. Dependent variable is 0/1 for whether the younger sibling chose the school above the school cutoff, i.e. the school attended by the older sibling if he scores at or above the cutoff score. Regressions include piecewise polynomial terms in centered test score, fixed effects for cutoff school, older sibling's exam year, and number of years between siblings' exam years. Results are reported for the polynomial order that minimizes the Akaike Information Criterion. **Bootstrapped p-values accounting for clustering at the centered score level are in parentheses.**

*** p<0.01, ** p<0.05, * p<0.1.

Table 1.4: Effect of older sibling admission on younger sibling's preference for same school, disaggregated by age difference of siblings

	(1)	(2)	(3)
	All siblings	Siblings 1-2 years apart	Siblings 3-5 years apart
Bandwidth=3			
Admission coefficient	0.068*	0.072**	0.064**
	(.07)	(.04)	(.04)
Mean of dependent variable	0.19	0.20	0.19
R^2	0.20	0.20	0.21
Observations	61,938	34,849	26,844
AIC-optimal polynomial order	1	1	1
Difference in admission coefficient between 3-5 years apart and 1-2 years apart samples			-0.007
			(.37)
Bandwidth=5			
Admission coefficient	0.077***	0.068**	0.071**
	(.00)	(.05)	(.02)
Mean of dependent variable	0.21	0.199	0.22
R^2	0.21	0.221	0.21
Observations	88,222	49,634	38,355
AIC-optimal polynomial order	1	2	1
Difference in admission coefficient between 3-5 years apart and 1-2 years apart samples			-0.012
			(.31)
Bandwidth=10			
Admission coefficient	0.074***	0.093***	0.097***
	(.00)	(.00)	(.00)
Mean of dependent variable	0.23	0.24	0.23
R^2	0.27	0.27	0.26
Observations	118,177	66,776	51,103
AIC-optimal polynomial order	3	1	1
Difference in admission coefficient between 3-5 years apart and 1-2 years apart samples			0.004
			(.71)

Note. Dependent variable is 0/1 for whether the younger sibling chose the school above the school cutoff, i.e. the school attended by the older sibling if he scores at or above the cutoff score. Regressions include piecewise polynomial terms in centered test score, fixed effects for cutoff school, older sibling's exam year, and number of years between siblings' exam years. Results are reported for the polynomial order that minimizes the Akaike Information Criterion. Difference in admission coefficients between 3-5 year apart and 1-2 year apart samples is from a fully interacted joint regression at the AIC-optimal polynomial order. **Bootstrapped p-values accounting for clustering at the centered score level are in parentheses.** *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 1.5: Effect of older sibling admission on other of schools chosen in cutoff subsystem

	(1)	(2)	(3)
	All siblings	Siblings 1-2 years apart	Siblings 3-5 years apart
Bandwidth=3			
Admission coefficient	0.24**	0.20*	0.29
	(.05)	(.08)	(.14)
Mean of dependent variable	2.22	2.16	2.31
R^2	0.18	0.19	0.20
Observations	31,474	17,630	13,518
AIC-optimal polynomial order	1	1	1
Difference in admission coefficient between			0.07
3-5 years apart and 1-2 years apart samples			(.25)
Bandwidth=5			
Admission coefficient	0.22***	0.19***	0.27***
	(.01)	(.00)	(.01)
Mean of dependent variable	2.23	2.16	2.32
R^2	0.18	0.18	0.19
Observations	48,153	27,105	20,771
AIC-optimal polynomial order	1	1	1
Difference in admission coefficient between			0.07
3-5 years apart and 1-2 years apart samples			(.31)
Bandwidth=10			
Admission coefficient	0.24***	0.23***	0.26***
	(.00)	(.00)	(.00)
Mean of dependent variable	2.26	2.19	2.35
R^2	0.17	0.17	0.18
Observations	76,901	43,442	33,140
AIC-optimal polynomial order	1	1	1
Difference in admission coefficient between			0.02
3-5 years apart and 1-2 years apart samples			(.44)

Note. Dependent variable is the number of schools selected by the younger sibling that belong to the cutoff school's subsystem, excluding the cutoff school. Only students at the threshold of a subsystem are included, i.e. rejection from the cutoff school implies attending a different subsystem. Regressions include piecewise polynomial term in centered test score, fixed effects for cutoff school, older sibling's exam year, and number of years between siblings' exam years. Results are reported for the polynomial order that minimizes the Akaike Information Criterion. Difference in admission coefficients between 3-5 year apart and 1-2 year apart samples is from a fully interacted joint regression at the AIC-optimal polynomial order. **Bootstrapped p-values accounting for clustering at the centered score level are in parentheses.**
*** p<0.01, ** p<0.05, * p<0.1.

Table 1.6: Differential effect of older sibling admission on school choice by graduation outcome

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: student chose cutoff school[a]				Dependent variable: number of other schools selected in cutoff subsystem[b]	
	All schools: first choice		Non-elite schools: first non-elite choice		All schools	
Bandwidth=3						
Admission coefficient	0.003	(One per school)	0.047*	(One per school)	0.274	(One per school)
	(.43)		(.07)		(.14)	
Admission × older sib graduation	0.041*	0.042*	0.090	0.115*	0.005	0.014
	(.09)	(.08)	(.12)	(.09)	(.96)	(.74)
Admission × older sib GPA		-0.017		0.009		-0.028
		(.20)		(.62)		(.51)
Mean of dependent variable	0.09	0.09	0.23	0.23	1.87	1.87
R^2	0.13	0.14	0.17	0.21	0.26	0.30
Observations	13,925	13,925	10,902	10,902	6,655	6,655
AIC-optimal polynomial order	1	1	1	1	1	1
Bandwidth=5						
Admission coefficient	-0.010	(One per school)	0.034**	(One per school)	0.130	(One per school)
	(.15)		(.02)		(.26)	
Admission × older sib graduation	0.061*	0.059**	0.110***	0.109***	0.193	0.245
	(.06)	(.03)	(.01)	(.00)	(.24)	(.11)
Admission × older sib GPA		-0.009		0.02		-0.133
		(.70)		(.18)		(.27)
Mean of dependent variable	0.10	0.10	0.24	0.24	1.87	1.87
R^2	0.12	0.14	0.19	0.19	0.24	0.27
Observations	19,153	19,153	15,142	15,142	9,947	9,947
AIC-optimal polynomial order	1	1	1	1	1	1
Bandwidth=10						
Admission coefficient	0.019**	(One per school)	0.024	(One per school)	0.133	(One per school)
	(.02)		(.14)		(.21)	
Admission × older sib graduation	0.032*	0.019*	0.097***	0.097***	0.270**	0.331**
	(.09)	(.09)	(.00)	(.00)	(.03)	(.02)
Admission × older sib GPA		0.002		0.017		-0.03
		(.19)		(.28)		(.89)
Mean of dependent variable	0.13	0.13	0.27	0.27	1.93	1.93
R^2	0.13	0.14	0.15	0.17	0.22	0.24
Observations	24,954	24,954	19,378	19,378	14,887	14,887
AIC-optimal polynomial order	1	1	1	1	1	1

Note. Sample excludes all students at an UNAM school cutoff or who would attend an UNAM school if rejected from the cutoff school, since the UNAM schools have no graduation data available. Specifications include piecewise polynomial terms in centered test score, fixed effects for cutoff school, older sibling's exam year, and number of years between siblings' exam years; all are fully interacted with the graduation dummy. Columns 2, 4, and 6 also include one admission coefficient per cutoff school, de-meaned older sibling's GPA, a first-order piecewise polynomial in the interaction between older sibling's GPA and centered test score, and the interaction between GPA and graduation. Results are reported for the polynomial order that minimizes the Akaike Information Criterion. **Bootstrapped p-values accounting for clustering at the centered score level are in parentheses.**

*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

[a] Dependent variable is 0/1 for whether the younger sibling chose the school above the school cutoff, i.e. the school attended by the older sibling if he scores at or above the cutoff score.

[b] Dependent variable is the number of schools selected by the younger sibling that belong to the cutoff school's subsystem, excluding the cutoff school. Only students at the threshold of a subsystem are included, i.e. rejection from the cutoff school implies attending a different subsystem.

Table 1.7: Placebo: differential effect of younger sibling admission on older sibling's school choice by graduation outcome

	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: student chose cutoff school[a]				Dependent variable: number of other schools selected in cutoff subsystem[b]	
	All schools: first choice		Non-elite schools: first non-elite choice		All schools	
Bandwidth=3						
Admission coefficient	-0.012	(One per	-0.018	(One per	-0.18	(One per
	(.47)	school)	(.34)	school)	(.15)	school)
Admission × older sib graduation	0.011	0.025	0.023	0.017	0.114	0.225
	(.41)	(.22)	(.47)	(.57)	(.23)	(.57)
Admission × older sib GPA		-0.029		0.012		0.056
		(.15)		(.40)		(.11)
Mean of dependent variable	0.10	0.10	0.23	0.20	1.72	1.72
R^2	0.15	0.17	0.17	0.23	0.25	0.30
Observations	11,297	11,297	8,947	8,947	5,426	5,426
AIC-optimal polynomial order	1	1	1	1	1	1
Bandwidth=5						
Admission coefficient	-0.007	(One per	-0.038	(One per	-0.018	(One per
	(.91)	school)	(.30)	school)	(.84)	school)
Admission × older sib graduation	0.003	0.012	0.041	0.054	0.098	0.178
	(.76)	(.72)	(.42)	(.14)	(.63)	(.50)
Admission × older sib GPA		-0.021		-0.019		-0.099
		(.41)		(.43)		(.24)
Mean of dependent variable	0.11	0.11	0.24	0.24	1.73	1.73
R^2	0.15	0.17	0.16	0.18	0.24	0.28
Observations	15,745	15,745	12,620	12,620	8,177	8,177
AIC-optimal polynomial order	1	1	1	1	1	1
Bandwidth=10						
Admission coefficient	0.002**	(One per	-0.013	(One per	0.024	(One per
	(.59)	school)	(.64)	school)	(.45)	school)
Admission × older sib graduation	0.009	0.004	0.037	0.036	0.08	0.068
	(.75)	(.87)	(.11)	(.13)	(.77)	(.93)
Admission × older sib GPA		0.003		-0.008		0.099
		(.64)		(.89)		(.11)
Mean of dependent variable	0.14	0.14	0.27	0.27	1.82	1.82
R^2	0.16	0.17	0.16	0.18	0.20	0.23
Observations	19,851	19,851	15,635	15,635	11,974	11,974
AIC-optimal polynomial order	1	1	1	1	1	1

Note. Sample excludes all students at an UNAM school cutoff or who would attend an UNAM school if rejected from the cutoff school, since the UNAM schools have no graduation data available. Specifications include piecewise polynomial terms in centered test score, fixed effects for cutoff school, older sibling's exam year, and number of years between siblings' exam years; all are fully interacted with the graduation dummy. Columns 2, 4, and 6 also include one admission coefficient per cutoff school, de-meaned older sibling's GPA, a first-order piecewise polynomial in the interaction between older sibling's GPA and centered test score, and the interaction between GPA and graduation. Results are reported for the polynomial order that minimizes the Akaike Information Criterion. Bootstrapped p-values accounting for clustering at the centered score level are in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

[a] Dependent variable is 0/1 for whether the older sibling chose the school above the school cutoff, i.e. the school attended by the older sibling if he scores at or above the cutoff score.

[b] Dependent variable is the number of schools selected by the older sibling that belong to the cutoff school's subsystem, excluding the cutoff school. Only students at the threshold of a subsystem are included, i.e. rejection from the cutoff school implies attending a different subsystem.

Table 1.8: Nested logit estimates of school choice model

		(1)	(2)	(3)
Variable	Interacted with	All siblings	Siblings 3-5 years apart	All siblings
School above cutoff	Constant	0.579***	0.562***	0.770***
		(0.022)	(0.033)	(0.107)
	Admission (score ≥ 0)	0.247***	0.216***	-0.215
		(0.026)	(0.039)	(0.129)
	Graduated			-0.135
				(0.133)
	Graduated × admission			0.472***
				(0.162)
School below cutoff	Constant	0.711***	0.628***	0.505***
		(0.030)	(0.047)	(0.152)
	Admission (score ≥ 0)	-0.410***	-0.340***	-0.024
		(0.042)	(0.064)	(0.190)
	Graduated			0.370**
				(0.178)
	Graduated × admission			-0.558**
				(0.238)
Subsystem of school	Fixed effects	YES	YES	YES
above cutoff	**Admission (score ≥ 0)**	0.213***	0.183***	-0.055
		(0.046)	(0.071)	(0.160)
	Graduated × fixed effects			YES
	Graduated × admission			0.328
				(0.212)
Subsystem of school	Fixed effects	YES	YES	YES
below cutoff	**Admission (score ≥ 0)**	-0.241***	-0.166***	0.046
		(0.051)	(0.081)	(0.173)
	Graduated × fixed effects			YES
	Graduated × admission			-0.446*
				(0.229)
Distance to school	Constant	-0.138***	-0.136***	-0.152***
		(0.002)	(0.002)	(0.003)
	Graduated			0.008***
				(0.002)
Mean COMIPEMS score	Constant	0.033***	0.032***	0.027***
of school		(0.001)	(0.001)	(0.001)
	Graduated			0.003***
				(0.001)
	Intra-nest correlation	0.444***	0.439***	0.456***
	parameter (λ)	(0.006)	(0.008)	(0.010)
	Students	64,486	28,332	14,655
	Student-school observations	39,587,950	17,361,084	9,121,816

Note. Results are from a nested logit model, with subsystem as the nest, for all students within 5 points of the cutoff of a school, where there is only one counterfactual school above the cutoff and one below the cutoff within the 5 point bandwidth. Specifications include dummy variables for school subsystem and interactions of these dummy variables with 1) an indicator for whether the school above the cutoff belongs to that subsystem and 2) an indicator for whether the school below the cutoff belongs to that subsystem. Also included are first-order piecewise polynomials in school above cutoff, school below cutoff, subsystem above cutoff, and subsystem below cutoff. In column 3, every variable is interacted with the "graduated" dummy variable. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 1.9: Balance of exogenous characteristics at cutoff

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	Parental education	Male	Hours studied per week	Middle school GPA	Number of siblings	Birth order
Bandwidth=3						
Admission coefficient	-0.018	-0.012	0.053	0.005	0.014	0.013
	(0.055)	(0.008)	(0.057)	(0.013)	(0.019)	(0.017)
Mean of dependent variable	10.89	0.440	5.251	8.267	2.156	1.580
SD of dependent variable	3.28	0.50	3.30	0.79	1.12	0.94
R^2	0.112	0.086	0.080	0.186	0.068	0.035
Observations	60,299	60,299	60,299	60,299	60,299	60,299
p-value for test that all admission coefficients = 0	0.75		AIC-optimal polynomial	order	1	
Bandwidth=5						
Admission coefficient	-0.049	-0.004	-0.014	-0.009	0.025*	0.019
	(0.044)	(0.007)	(0.045)	(0.010)	(0.015)	(0.013)
Mean of dependent variable	10.86	0.445	5.264	8.274	2.155	1.581
SD of dependent variable	3.30	0.50	3.31	0.80	1.12	0.94
R^2	0.113	0.090	0.081	0.187	0.060	0.030
Observations	85,795	85,795	85,795	85,795	85,795	85,795
p-value for test that all admission coefficients = 0	0.52		AIC-optimal polynomial	order	1	
Bandwidth=10						
Admission coefficient	0.014	-0.008	0.030	-0.004	0.016	0.016
	(0.037)	(0.006)	(0.037)	(0.008)	(0.013)	(0.011)
Mean of dependent variable	10.79	0.457	5.251	8.284	2.160	1.582
SD of dependent variable	3.32	0.50	3.30	0.81	1.13	0.95
R^2	0.122	0.102	0.083	0.203	0.059	0.029
Observations	114,856	114,856	114,856	114,856	114,856	114,856
p-value for test that all admission coefficients = 0	0.39		AIC-optimal polynomial	order	1	

Note. Dependent variable is given in the column header. Seemingly unrelated regression is used to jointly estimate all columns. Specifications include piecewise polynomial terms in centered test score, fixed effects for cutoff school, older sibling's exam year, and number of years between siblings' exam years. Results are reported for the polynomial order that minimizes the Akaike Information Criterion. Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Chapter 2

Flourish or Fail? The Risky Reward of Elite High School Admission in Mexico City

Joint work with Alain de Janvry and Elisabeth Sadoulet

2.1 Benefits and Risks of Attending an Elite School

Families often have some choice in where their children attend school, and all else equal, most families prefer a school of higher academic quality (see, e.g., Hastings, Kane, and Staiger 2009). Attending a "better" school, as defined by peer ability or school resources, is usually thought to benefit students academically. For example, a better-funded school is able to afford more and better educational inputs, and a student may benefit from working with high-achieving and highly motivated peers. But there is also a risk to attending a better school, particularly if doing so means that the student is closer to the bottom of the school-specific ability distribution. The difficulty level of the coursework may prove too much for the student to handle. Teachers may teach mostly to the top of the class, leaving behind those who enter the school with a weaker academic background.[1] Students experiencing such challenges may fail to complete their education at all, which is probably a much less desirable outcome than graduating from a lower-quality school.

This chapter quantifies the trade-off between academic benefit and dropout risk facing students admitted to a subset of Mexico City's elite public high schools. Mexico City is ideal for this exercise for three reasons. First, there are large perceived disparities in public high school quality, with a well-identified group of "elite" schools standing above all others. This gives a natural definition of what an "elite" (or "better") school is. Second, nearly all public high schools in the city participate in a unified merit-based admissions system, using a standardized exam and students' stated preferences to allocate all students across

[1]Duflo et al. (2011) elaborate on the potential benefits and drawbacks of ability tracking.

schools. This mechanism allows us to credibly identify the impact of elite school admission on dropout probability and end-of-high school exam scores. Third, Mexico is characterized both by a high secondary school dropout rate and a significant estimated return to high school education, so the risk of dropping out is a first-order issue facing students. In our sample, about half of students who are assigned to a high school do not take the exit exam three years later. At the same time, young men with a high school diploma have 24% higher wages than those who only completed middle school (Campos-Vazquez 2013). Though this is not a causal statement, it is suggestive that dropping out has a real cost for students.

A simple regression discontinuity design, made possible by the assignment mechanism, is used to discover whether students experience a change in dropout probability and in exam scores as a result of admission to an elite school, using their most-preferred non-elite school that would admit them as the counterfactual. We find that there is a clear trade-off for most marginally admitted students. Admission to an elite school raises the probability of high school dropout by 8.5 percentage points, compared to an average probability of 46%. Along with this substantial increase in dropout probability, elite school admission also results in an average gain of 0.11 standard deviations on the 12^{th} grade standardized exam, which apparently comes entirely from gains in math. Students with lower middle school grade point averages experience larger increases in dropout probability, but there is no evidence that they experience a smaller boost in their exam scores from elite admission. We introduce and carry out a procedure that evaluates the potential bias in the exam score effect due to differential dropout with respect to observable and unobservable characteristics, and find that the effect is quite robust.

While a structural treatment of student preferences is not the subject of this chapter, we present reduced-form evidence showing that students with lower performance in middle school choose elite schools less often, compared to neighboring high-performing students with the same entrance exam score. The chapter's main findings offer one explanation for this result. Weak students may understand that elite school admission is a double-edged sword: while the expected academic benefit for graduates is positive, the increased chance of leaving high school without a diploma makes applying to an elite school a risky choice.

Most previous studies on the effects of elite high school admission have focused on the impact on exam scores. Such studies typically analyze cases of merit-based admission systems, and use a sharp or fuzzy regression discontinuity design to estimate the effect of elite school admission on outcomes. Most have found zero or modest effects: Clark (2010) in the United Kingdom, Abdulkadiroglu et al. (2011) in Boston and New York, and Lucas and Mbiti (2013) in Kenya all find zero or negligible impacts from elite high schools while Jackson (2010) and Pop-Eleches and Urquiola (2013) find a modest benefit of admission to high schools with higher-scoring peers in Trinidad and Tobago and Romania, respectively. Zhang (2012) exploits a randomized lottery for elite Chinese *middle* schools to show that elite admission has no significant impact on academic outcomes. Beyond the zero effect on exam scores, Dobbie and Freyer (2011) find that the New York elite high schools do not have an appreciable effect on long-run outcomes such as SAT score or college graduation.[2]

In a much different study, Duflo et al. (2011) randomly assigned Kenyan schools into a tracking regime where they divided their first grade classes by student ability. They find that while tracking is beneficial, there is no evidence that being in a class with better peers is the mechanism through which these benefits are manifested. We note that in the case of admission to competitive elite schools, admission results both in a more able peer group as well as a different schooling environment with resources, management, and culture that may be quite different from other public schools. Thus the effect of elite school admission is a reflection of both the peer and institutional channels, which regression discontinuity designs such as the present one cannot effectively disentangle.[3]

The literature on the relationship between school quality and student dropout is sparser. Recent studies have mostly focused on the impacts of specific aspects of quality, randomly varying one aspect to see if it increased school participation, which differs from the concept of dropout in that reduced participation may not result in permanently abandoning schooling while dropout usually does. For example, Glewwe, Ilias, and Kremer (2010) find no effect of a teacher incentive pay scheme on student participation in Kenyan public primary schools. More related to our study, de Hoop (2011) estimates the impact of admission to competitive, elite public secondary schools on dropout in Malawi. He finds that admission decreases dropout. This could be due to increased returns from an elite education inducing students to attend, or because the elite schools provide a more supportive environment. Our findings provide a stark contrast to these results, although in a much different economic and social context.

The rest of the chapter is organized as follows. Section 2.2 gives a detailed overview of the Mexico City high school admissions system. Section 2.3 sets forth the method for identifying the effects of admission on outcomes. Section 2.4 describes the data and Section 2.5 gives the empirical results and several validity checks. Section 2.6 uses the results to rationalize revealed preference for elite schools. Section 2.7 concludes.

2.2 Mexico City public high school system and student enrollment mechanism

Beginning in 1996, the ten public high school subsystems in Mexico's Federal District and various municipalities in the State of Mexico adopted a competitive admissions process. This

[2]Estrada and Gignoux (2014) use a similar empirical strategy to ours with one year of COMIPEMS data and a separate survey (administered in a subsample of high schools) to estimate the effect of elite school admission on subjective expectations of the returns to higher education, finding that admission leads to higher expected returns.

[3]Further studies on the impact of specific aspects of school quality on test scores include Dearden, Ferri, and Meghir (2002), Newhouse and Beegle (2006), Gould, Lavy, and Paserman (2004), Hastings, Kane, and Staiger (2006), Hastings and Weinstein (2008), Cullen, Jacob, and Levitt (2005 and 2006), and Lai et al. (2010).

consortium of schools is known as the Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS). COMIPEMS was formed in response to the inefficient high school enrollment process at the time, in which students attempted to enroll in several schools simultaneously and then withdrew from all but the most-preferred school that had accepted them. The goal of COMIPEMS was to create a unified high school admissions system for all public high schools in the Mexico City metropolitan area that addressed such inefficiencies and increased transparency in student admissions.

Any student wishing to enroll in a public high school must participate in the COMIPEMS admissions process. In February of the student's final year of middle school (grade nine), informational materials are distributed to students explaining the rules of the admissions system and registration begins. As part of this process, students turn in a ranked list of up to twenty high schools that they want to attend.[4] In June of that year, after all lists of preferred schools have been submitted, registered students take a comprehensive achievement examination. The exam has 128 multiple-choice questions worth one point each, covering a wide range of subject matters corresponding to the public school curriculum (Spanish, mathematics, and social and natural sciences) as well as mathematical and verbal aptitude sections that do not correspond directly to the curriculum.

After the scoring process, assignment of students to schools is carried out in July by the National Center of Evaluation for Higher Education (Ceneval), under the observation of representatives from each school subsystem and independent auditors. The assignment process is as follows. First, each school subsystem sets the maximum number of students that it will accept at each high school. Then, students are ordered by their exam scores from highest to lowest. Any student who scored below 31 points or failed to complete middle school is disqualified from participating. Next, a computer program proceeds in descending order through the students list, assigning each student to her highest-ranked school with seats remaining when her turn arrives. If by the time a student's turn arrives, all of her selected schools are full, she must wait until after the selection process is complete and choose from the schools with open spots remaining. This stage of the allocation takes place over several days, as unassigned students with the highest scores choose from available schools on the first day and the lowest scorers choose on the final days.

In some cases, multiple students with the same score have requested the final seats available in a particular school, such that the number of students outnumbers the number of seats. When this happens, the representatives in attendance from the respective school subsystem must choose to either admit all of the tied applicants, exceeding the initial quota, or reject all of them, taking fewer students than the quota. The number of offered seats and the decisions regarding tied applicants are the only means by which administrators determine student assignment to schools; otherwise, assignment is entirely a function of the students'

[4]Students actually rank programs, not schools. For example, one technical high school may offer multiple career track programs. A student may choose multiple programs at the same school. For simplicity we will use the term "school" to refer to a program throughout. No elite school has multiple programs at the same school, so this distinction is unimportant for the empirical analysis.

reported preferences and their scores. Neither seat quotas nor tie decisions offer a powerful avenue for strategically shaping a school's student body.[5]

At the end of the final year of high school (grade twelve), students who are currently enrolled take a national examination called the Evaluación Nacional de Logro Académico en Centros Escolares (ENLACE), which tests students in Spanish and mathematics. This examination has no bearing on graduation or university admissions and the results have no fiscal or other consequence for high schools. It is a benchmark of student and school achievement and progress. There are two elite school subsystems, each affiliated with a prestigious national university. One of these subsystems, the Universidad Nacional Autónoma de México, refuses to administer the ENLACE exam and is legally able to do so because of its autonomous status. The other elite subsystem, the Instituto Politécnico Nacional (IPN), administers the ENLACE, as do all of the other public school subsystems. Because the ENLACE data provide the dependent variable for our analysis, only IPN schools are examined in this chapter.

2.3 Regression discontinuity design and sample definition

The goal of this chapter is to determine how much (marginal) admission to an IPN school changes students' probability of dropout and their end-of-high school exam scores. Put another way, the econometric challenge is to estimate the effect on academic outcomes from admission to a school in an IPN subsystem instead of admission to the student's most-preferred non-elite choice, holding constant COMIPEMS score and all student characteristics, observed and unobserved.

The COMIPEMS assignment mechanism permits a straightforward strategy for identifying the causal effect of IPN school admission on outcomes, through a sharp regression discontinuity (RD) design. Each school that is oversubscribed (i.e., with more demand than available seats) accepts all applicants at or above some cutoff COMIPEMS exam score, and rejects all applicants scoring below that cutoff.[6] Whether or not a student wanting to attend a particular school is actually admitted is determined entirely by whether or not she is above or below the cutoff score, giving a sharp discontinuity in the probability of admission (from 0 to 1) when the student reaches the cutoff.

We are interested in the comparison between attending an IPN school and a non-elite school, so we restrict attention to the set of students who would attend an IPN school for COMIPEMS scores above some cutoff and would attend a non-elite school for COMIPEMS

[5]The only obvious case would be to drastically under-report available seats at a school to reduce enrollment. But setting an artificially low seat quota and planning to accept students up to a level close to "true" capacity in the event of a tie either results in the school being under-enrolled (if there are too many tied students to accept) or enrolled near the level that would prevail with the true quota reported and all ties rejected.

[6]The cutoff is set implicitly by the student who gets the final seat in that school.

scores below that cutoff. For each such student i in year t, there is a "cutoff school" indexed by j with cutoff score C_{jt}. Denoting student i's score by c_i, the cutoff school is the one whose cutoff score serves as the lower boundary of IPN schools: for all $c_i \geq C_{jt}$ (and within a selected bandwidth, discussed below), an IPN school is attended, and for all $c_i < C_{jt}$, a non-elite school is attended. This cutoff school is unique for each student; among the set of IPN schools listed before any non-elite option, it is the school with the lowest cutoff score.

Following Abdulkadiroglu et al. (2012), we use a stacked parametric RD design that estimates a single average admission effect over all cutoff schools while allowing for separate polynomials in the COMIPEMS score for each cutoff school. The estimating equation is:

$$Y_{ijt} = \delta \text{admit}_i + g_{1j}\left(c_i - C_{jt}\right) + \text{admit}_i g_{2j}\left(c_i - C_{jt}\right) + \mu_j + \eta_t + \varepsilon_{ijt} \tag{2.1}$$

where Y_{ijt} is the outcome of interest (dropout or ENLACE exam score) g_{1j} and g_{2j} are school-specific polynomial terms, $c_i - C_{jt}$ (the "centered" COMIPEMS score) is the difference between i's COMIPEMS score and the cutoff school's cutoff score, and $\text{admit}_i = 1$ if $c_i - C_{jt} \geq 0$. Note that if a student scores high enough above j's cutoff, she may get into a more-preferred IPN school with a higher cutoff score. This does not compromise the research design because the admitting school is still elite. The parameter of interest is δ, the local average treatment effect of being admitted to an IPN school instead of a non-elite school (Imbens and Lemieux 2008). This is an intention-to-treat effect since students do not necessarily attend the school to which they were admitted. But in practice, compliance is almost perfect. Of those in the RD sample who take the 12^{th} grade exam, 99.8% of the students rejected from the IPN subsystem take the exam in a non-elite school, while 96.1% of students admitted to an IPN school take the exam in an IPN school.

The running variable, centered COMIPEMS score, is discrete, which necessitates a parametric approach to the RD design (Lee and Card 2008). Per Lee and Card's (2008) suggestion, we cluster our standard errors at the level of the centered score in order to account for specification error in the polynomials. Because there are few clusters, wild-cluster bootstrapped p-values are computed rather than analytical clustered standard errors (Cameron et al. 2008). We use the Akaike Information Criterion to select the appropriate polynomial order, which in practice is always 1.

Choosing the optimal bandwidth is not straightforward for two reasons, both related to sample selection. First, we exclude any student who would go completely unassigned for one or more point values within the bandwidth. This is because we want to compare outcomes due to IPN vs. non-elite admissions, without considering the effects of being left unassigned. Second, we mentioned that of the two elite subsystems, only the IPN administers the ENLACE exam (while the UNAM does not). For this reason, we restrict the sample to students for whom no score within the selected bandwidth would result in admission to an UNAM school. Because of the criteria, some observations drop out of the sample as the bandwidth grows, while other observations are added. Thus a standard bandwidth selection procedure is inappropriate here, and we instead present results for a variety of bandwidths.

An advantage of the RD design is that it does not require any assumptions about the decision-making process by which students choose schools and whether their rankings of schools truly represent revealed preferences. Conditional on COMIPEMS score, the admitted and rejected students near a school's cutoff have the same expected characteristics, including school preferences. Even if students are trying to choose strategically or making mistakes in their selections, this behavior should not differ by admissions outcome near the cutoff. We can thus remain agnostic on the issue of the distribution of student preferences and the factors that influence them.

2.4 Data description

The data used in this chapter come from two sources, both obtained from the Subsecretariat of Secondary Education of Mexico: the registration, scoring, and assignment data for the 2005 and 2006 COMIPEMS entrance examination processes, and the scores from the 2008, 2009, and 2010 12th grade ENLACE exams.[7] The COMIPEMS dataset includes all students who registered for the exam, with their complete ranked listing of up to twenty high school preferences, basic background information such as middle school grade point average and gender, exam score out of 128 points, and the school to which the student was assigned as a result of the assignment process. It also includes student responses to a multiple choice demographic survey turned in at the time of registration for the exam.

The ENLACE dataset consists of exam scores for all students who took the test in Spring 2008 (the first year that the 12th grade ENLACE was given), 2009, or 2010. The scores for both the math and Spanish sections are reported as a continuous variable, reflecting the weighting of raw scores by question difficulty and other factors. We normalize the scores by subtracting off the year-specific mean score for all examinees in public high schools within the COMIPEMS geographic area and dividing by the year-specific standard deviation from this same sample. The ENLACE scores are matched with the 2005 and 2006 COMIPEMS-takers by using the Clave Única de Registro de Población (CURP), a unique identifier assigned to all Mexican citizens. Matching is performed by name and date of birth if no CURP match is found. The matching rate of ENLACE takers to their COMIPEMS scores is nearly 100% and will be discussed further in Section 2.5.4.

The IPN schools are highly-demanded.[8] For every seat available in an IPN school, 1.9 students list an IPN school as their first choice. Every IPN school is oversubscribed. Compared to the non-elite schools, the IPN's student body has higher COMIPEMS exam scores (74.9 points vs. 58.7), grade point (8.24/10 vs. 7.98/10), parental education (10.7 years vs. 9.7), family income (4,634 pesos/month vs. 3,788), and ENLACE exam score (0.52 normalized score vs. -0.12). While we do not have data on this point, it is widely accepted that IPN schools receive more funding on a per-student basis than non-elite schools.

[7]The 2010 data is used in order to match students from the 2006 COMIPEMS cohort who took four years to complete high school instead of three.

[8]Students selecting an UNAM school (the other elite subsystem) as their first choice must take a version of the entrance exam written by UNAM, which is advertised to be equivalent to the standard version in content and difficulty. We include a dummy variable for exam version in all regressions.

We limit the sample to applicants who graduated from a public middle school in Mexico City in the year that they took the COMIPEMS exam. Summary statistics for this sample and the subsample consisting only of students located near the cutoff of IPN admission are in Table 2.1.[9] Students near the IPN cutoff (column 2) are substantially different from the full sample (column 1), as we would expect since students near the IPN cutoff both selected at least one IPN school and scored high enough to be close to an IPN cutoff. They are more likely to be male, have more educated parents and higher incomes, better grades, and COMIPEMS scores that are more than half a standard deviation above the sample mean. These students score 0.36 standard deviations above the full sample average on the ENLACE exams.[10]

It is clear from Table 2.1 that many COMIPEMS exam takers do not take the ENLACE. We will present evidence in Section 2.5.4 that this is almost entirely due to student dropout rather than some other feature of the data. For the moment, we treat non-taking as dropout and show in Table 2.2 that dropout is predicted both by academic ability and IPN admission. Column 1 shows that, in the cross-section, COMIPEMS exam score and middle school grade point average (GPA) are negatively correlated with dropout. Particularly striking is the GPA coefficient, showing that a one standard deviation (.82) increase in GPA predicts a 15 percentage point decrease in dropout probability. Parental education and family income are both negatively correlated with dropout as well, but the magnitude of the coefficients is very small compared to those of COMIPEMS and GPA. Column 2 adds high school fixed effects and shows that these relationships are very similar within a high school. Column 3 shows that, conditional on listing an IPN school as one's first choice, dropout is much higher for students admitted to IPN schools than for those admitted to non-elite schools. This correlation does not have a causal interpretation, however, because unobservable student attributes could affect both selection into an IPN school and dropout probability. The next section uses the RD design to establish the causal IPN admission-dropout relationship.

2.5 Effects of elite school admission

This section uses the RD strategy outlined in Section 2.3 to estimate the effect of marginal admission to an IPN school on the probability of dropping out of high school and, conditional on taking the ENLACE exam, on the exam score obtained. Because we lack individual-level data on graduation, taking the ENLACE exam is used as a proxy for graduation. Only students on track to graduate at the end of the school year are registered to take the exam. We present evidence in Section 2.5.4 that this is a good proxy, in particular that schools do

[9] The size of the window for being considered "near the cutoff" is 10 points above or below the respective IPN school's cutoff score.

[10] There is no binding test score ceiling for either exam. Score ceilings present a problem for academic gains because there is no way for students with the highest score to demonstrate progress. The COMIPEMS exam intentionally avoids a ceiling in order to sort students during assignment.

not strategically administer this exam. Thus the only sample used from this point forward is that of students near an IPN school's cutoff, as defined in Section 2.3.

2.5.1 Probability of dropout

Marginal admission to an IPN school has a large, significant positive impact on the probability of dropout. Figure 2.1, Panel a illustrates this graphically, centering students' scores about their school-specific cutoff score and plotting the dropout rate in a 10 point window around the threshold.[11] Table 2.3 confirms this finding, reporting the average effect of admission on dropout estimated using the regression discontinuity design for bandwidths of 4, 7, and 10 points. Results are quite similar across bandwidths, so for brevity we only discuss those using a bandwidth of 7. Column 5, which excludes any additional covariates, estimates that the probability of dropping out increases by 8.53 percentage points, compared to the mean probability of 46.1%. Adding covariates—middle school GPA, parental education, family income, gender, hours studied per week in middle school, a normalized index of responses to questions about parental effort and involvement in schooling, and employment—in column 6 decreases this estimate slightly to 8.18 percentage points. Column 7 adds interactions between the covariates and admission in order to explore whether the admission effect is heterogeneous with respect to student characteristics. The empirical specification is:

$$\begin{aligned} dropout_{ijt} &= \delta \text{admit}_i + g_{1j}\left(c_i - C_{jt}\right) + \text{admit}_i g_{2j}\left(c_i - C_{jt}\right) \\ &+ \Sigma_k x_{ik}\left[\gamma_{jk} + \alpha_{1k}\left(c_i - C_{jt}\right) + \text{admit}_i \alpha_{2k}\left(c_i - C_{jt}\right) + \theta_k \text{admit}_i\right] \\ &+ \mu_j + \eta_t + \varepsilon_{ijt} \end{aligned} \tag{2.2}$$

where i indexes the student, j indexes the cutoff school, k indexes the covariates, and x_{ik} is the value of the covariate for student i. In words, this specification has cutoff school and COMIPEMS exam year fixed effects, an admission effect, separate polynomials in COMIPEMS score for each cutoff school, and for each covariate, a separate level effect for each cutoff school, an interaction between the covariate and COMIPEMS score that varies on either side of the threshold, and an interaction between the covariate and admission. The coefficients of interest are the θ_k's, which show whether the average effect of marginal admission is different for students with different levels of the covariate.

The effect of IPN admission on dropout is strongly heterogeneous with respect to middle school GPA. All else equal, students with lower GPAs experience a larger increase in probability of dropout, as shown graphically in Figure 2.1, Panels b and c. To interpret this differential effect, consider that the standard deviation of GPA in this sample is 0.74, the effect for a student with the mean GPA is 8.24 percentage points, and that $\widehat{\theta}_{GPA}$ is -9.06. Then a student with a GPA one standard deviation below the mean experiences a

[11] While we cannot make causal interpretations regarding the slope parameters, the positive score-dropout relationship among rejected students is consistent with students who miss the cutoff by a large margin attending schools that are easier to graduate from than those who barely miss the cutoff. The negative slope among admitted students is expected, as students with better scores are more likely to graduate.

$8.24+(0.74\times 9.06) = 14.94$ percentage point effect of admission on dropout probability. Only students with very high GPAs, at the 86th percentile of the sample or above, are predicted to have a negative effect of admission on this probability.[12] The results for other student characteristics are not statistically different from zero when using the unadjusted standard errors, although the wild-cluster bootstrapped p-values indicate statistical significance in some cases. This is unusual behavior and indicates that the bootstrap is not performing as expected for the coefficients on variables that have little intra-cluster correlation. The two coefficients on which we focus, the effect of admission and the differential effect with respect to middle school GPA, are highly significant in both the unclustered and bootstrapped approaches. We take a conservative approach with all other coefficients and do not attempt to make inference where one approach gives significant results and the other does not.

It is possible to predict for each student, on the basis of observables, the differential probability of dropout induced by admission simply by summing the $\widehat{\theta}_k \times x_{ik}$'s. Doing this, we find that 90% of students are predicted to have a higher chance of dropout due to IPN admission. This is not inconsistent with the IPN's academic demands increasing the odds of school dropout for all admitted students. Rather, all students may want more strongly to stay in school if they are admitted to an elite school (causing a decrease in dropout probability), with the rigor of the IPN schools more than offsetting this impact for all but the best-prepared students.

Finally, column 8 allows the admission effect to vary by cutoff school and reports the coefficients for the heterogeneous admission effects. The purpose of this is to see if the heterogeneous effects persist even for students at the same cutoff, rather than the result being driven by cutoff schools with large effects on dropout and many students with a low GPA. The estimated differential effect with respect to GPA increases slightly.

These results make clear that dropout is systematically related to IPN admission and its interaction with students' academic preparation. Students admitted to an IPN school are on average more likely to drop out and thus less likely to take the ENLACE, such that even after conditioning on COMIPEMS score, IPN admittees taking the ENLACE have higher middle school GPAs. To show this, we estimate the following equation for each of the student characteristics:

$$x_{ijtk} = \phi_k \text{admit}_i + h_{1jk}\left(c_i - C_{jt}\right) + \text{admit}_i h_{2jk}\left(c_i - C_{jt}\right) + \mu_{jk} + \eta_{tk} + \varepsilon_{ijtk} \tag{2.3}$$

If x_k is balanced across the threshold, then $\widehat{\phi}_k$ should be close to zero. Table 2.4, Panel a gives estimates at the time of assignment (prior to dropout), where we expect balance. Of the seven covariates tested, only hours studied per week is found to change discontinuously at the threshold. When estimating the equations jointly and performing a joint test for discontinuities, we fail to reject the null hypothesis of no discontinuity. Panel b, however,

[12]One might wonder if middle school GPA is a good proxy for student academic performance or if it could reflect characteristics of the middle school itself. To explore this possibility, we re-estimated the model while also including the mean GPA of the student's middle school and its interactions as covariates. The results are nearly identical. It seems that GPA is a good proxy for academic performance, even across middle schools.

shows that within the sample of ENLACE takers middle school GPA is unbalanced (about 1/7 standard deviations higher for admitted students) as well as parental education, family income, and hours studied. This differential dropout, due entirely to the effect of IPN admission, may bias upward estimates of the IPN admission effect on ENLACE exam scores if the additional dropout is among the students who would have the lowest ENLACE scores. We will assess how severe this bias would have to be in order to push the estimate of the admission effect, presented in the next section, to zero.

2.5.2 ENLACE exam performance

We now turn to the effect of IPN admission on the ENLACE exam score. We first ignore the differential dropout issue raised in the previous section, and then propose a way to account for it in the next subsection. Using all observed scores, Figure 2.2, Panel a suggests that there is a significant, positive effect of IPN admission on ENLACE score. Panels b and c show that this effect comes entirely from improved math scores. This result is unsurprising, as IPN schools focus heavily on mathematics, engineering, and the sciences in their curriculum (implied by the "Politécnico" in its name). Table 2.5 reports the regression discontinuity results for these relationships. Again, the results are robust to the choice of bandwidth, although the results for a bandwidth of 4 are noisy. We will again focus on the results for a bandwidth of 7. Column 5 of Panel a, without covariates, gives a highly statistically significant admission effect of 0.11 standard deviations on the exam. Adding covariates in Column 6, the coefficient remains stable. Columns 7 and 8 add interactions between admission and the covariates, but we fail to reject that there are no differential impacts. Panel b suggests that this effect comes entirely from gains in math scores, between .18 and .21 standard deviations depending on the specification for the 7 point bandwidth. The effect on Spanish scores, shown in Panel c, is indistinguishable from zero. This is perhaps unsurprising, given the IPN's focus on math, science, and engineering.

2.5.3 Imputation from conditional quantiles

In order to assess the potential bias from dropout on the estimated exam score effect, we propose a method to impute "penalized" scores to students who were induced to drop out either by admission to or rejection from an elite school. The idea behind this method is to assume that the rejected dropouts would have had better ENLACE scores than their rejected, non-dropout peers with identical observable characteristics, and the admitted dropouts would have had worse scores than their observationally identical admitted non-dropout peers. The full procedure is explained in the Appendix. The intuition behind the procedure is given here.[13]

[13]There are other methods for assessing or bounding the bias due to sample attrition, but they are not well-suited to this application. Lee (2009) shows how to derive sharp bounds of a treatment effect under random assignment and attrition, but the differential attrition in our case is large enough that the resulting bounds are uninformative. Under the assumption of a normally distributed dependent variable, Angrist et

We saw above that the probability of dropping out of school was higher among the students marginally admitted to IPN schools than their marginally rejected counterparts. The effect of admission on dropout, for a bandwidth of 10, was estimated to be $\widehat{\Delta} = 9.1\%$ of the marginally admitted student population. Suppose for the moment that this effect is homogeneous and thus identical for all students. We can then classify the students in three groups: those who would never drop out regardless of admission, those who will always drop out, and those who are induced to drop out by admission. The concern is related to this third group because it is observed among the rejected but not among the admitted. By virtue of the discontinuity design, the shares of the three categories of students are identical on both sides of the threshold (after controlling for the function of COMIPEMS score), and hence the share of induced dropouts is equal to Δ, the difference in dropout rates among the admitted and rejected. As we do not know which of the admitted dropouts are in this group, the idea is to impute a low grade to all of the admitted dropouts, but to weigh these observations with imputed scores by $\Delta/(\pi + \Delta)$, where $(\pi + \Delta)$ is the dropout rate among the admitted. This is equivalent to assigning an imputed score only to a share Δ of the admitted students. There is no need to impute scores to the rejected dropouts since, under the assumption that admission only increases the probability of dropout, they would have dropped out if they had been admitted. This method thus avoids imputing scores to the very large number of dropouts among the admitted and rejected whose behavior is unrelated to admission.

Next, we allow for some heterogeneity among students. The probability π_i of dropping out if rejected is specific to a student (we will use an estimated function of covariates and COMIPEMS score $\pi(COMIPEMS_i, X_i)$), as is the impact of admission on dropout $\Delta_i = \Delta(COMIPEMS_i, X_i)$. We can then apply the rule described above for all individuals with $\Delta_i > 0$. In addition, there may be some students that will drop out of school if they are rejected while staying in school if admitted. This is the group with $\Delta_i < 0$. For these students, the concern is the excess dropout among the rejected. We thus apply a high score to all of these rejected dropout students, and weigh the imputed score by $-\Delta_i/\pi_i$, where π_i is their probability of dropout if rejected.

Which low score should be applied to each admitted dropout, and which high score should be applied to each rejected dropout? Recognizing student heterogeneity here as well, we use conditional quantile regressions to define high or low scores as observed among the non-dropouts with similar covariates and admission status.

We now summarize the method:

1. There are four groups of students: those who would never drop out regardless of admission, those who would always drop out, those who are induced to drop out by admission, and those who are induced to drop out by rejection.

al. (2006) recover unbiased estimates of a treatment effect using a tobit procedure that censors all outcomes below a low threshold and counts missing values as censored observations. It is unclear how to apply this method to the RD case because the censoring point would have to vary with respect to the covariates, in particular the running variable.

2. Predict conditional dropout probability if rejected $\pi_i = \pi\left(\text{COMIPEMS}_i, X_i\right)$ and impact of admission on dropout $\Delta_i = \Delta\left(\text{COMIPEMS}_i, X_i\right)$.

3. Use conditional quantile regression to impute "low" ENLACE scores (m^{th} conditional quantile) for admitted dropouts with positive predicted differential dropout due to elite admission ($\Delta_i > 0$), and "high" scores ($1 - m^{th}$ conditional quantile) for rejected dropouts with negative predicted differential dropout due to elite admission ($\Delta_i < 0$):

$$\widehat{\text{ENLACE}}_{ijt} = \begin{cases} Q_m\left(\text{ENLACE}|\text{COMIPEMS}_i, X_i, \text{admit}_i = 1\right), & \text{if } \Delta_i > 0, \text{admit}_i = 1, \text{drop}_i = 1. \\ Q_{1-m}\left(\text{ENLACE}|\text{COMIPEMS}_i, X_i, \text{admit}_i = 0\right), & \text{if } \Delta_i < 0, \text{admit}_i = 0, \text{drop}_i = 1. \end{cases}$$

4. Assign non-zero weights ω_i to dropouts with imputed ENLACE scores according to the magnitude of their differential dropout, such that the weighted observations represent the sizes of the two groups that were induced to drop out as a result of admission or rejection. This is $\omega_i = \Delta_i / \left(\Delta_i + \pi_i\right)$ for admitted dropouts with $\Delta_i > 0$ and $\omega_i = -\Delta_i/\pi_i$ for rejected dropouts with $\Delta_i < 0$. Assign a weight of $\omega_i = 1$ to non-missing ENLACE scores and $\omega_i = 0$ to those who dropped out but did not have higher predicted dropout probability due to their admission outcome. The result of this is a smooth density across the admissions threshold and balance of covariates across the threshold, as would be the case in a no-differential dropout scenario.

5. Perform the weighted ENLACE score regression, including both the non-dropouts with their true scores and the dropouts with their imputed scores.

6. If the point estimate of the admission effect on ENLACE score is still positive (or, alternatively, if the confidence interval of this estimate does not contain zero), repeat the process while imputing a lower quantile for admitted students and a higher quantile for the rejected students. Stop when the point estimate is zero (or the confidence interval contains zero).

This procedure is performed for the composite ENLACE score and the math score, but not for the Spanish score since the point estimate of the admission effect is negative. Figure 2.3, Panels a and b illustrate how the treatment effect and its 95% confidence interval (based on unadjusted standard errors) change as we impute increasingly extreme scores for dropouts. For the point estimate of the effect of admission on the composite score to be zero, students induced to drop out by admission would have to be on average in the 14^{th} percentile of the conditional distribution of observed scores for admitted students with the same covariates, while at the same time the students induced to drop out by rejection would have to be in the 86^{th} percentile of the conditional distribution of observed scores for rejected students. For math, these numbers would have to be more extreme, in the 7^{th} and 93^{th} percentiles, respectively. The effect on composite ENLACE score becomes insignificant at the 5% level when the 30^{th} and 70^{th} percentiles are imputed, respectively; there, the point estimate is 0.07. The effect on math scores becomes insignificant when the 13^{th} and 87^{th} percentiles are imputed, where the point estimate is 0.07. That is, differential dropout would have to be among students who are quite low-performing in comparison to non-dropout peers with the same observable characteristics and admissions outcomes. In particular, the effect on math scores appears very robust to the influence of dropout.

2.5.4 Validity checks

Here we present three validity checks to address potential concerns with the results. First, support for the validity of the regression discontinuity design is given. Second, the dropout results are shown to hold for a subsample of students with relatively low travel time to IPN schools, in order to show that the result is not due to increased commuting distance. Third, support is given for the assertion that the dropout-related results in this chapter are indeed due to IPN students leaving school at a higher rate, rather than a data issue.

There is no a priori reason to think that the regression discontinuity design might be invalid. Because the school-specific cutoff scores are determined in the process of the computerized assignment process, monitored by school subsystem representatives and independent auditors, there is no opportunity for student scores to be manipulated in order to push particular students from marginal rejection to marginal admission. Nevertheless, Figure 2.4 provides graphical evidence of the design's validity, showing the distribution of COMIPEMS scores of students near each IPN school cutoff normalized by subtracting off the threshold-specific cutoff score. While the histogram is fairly coarse due to the discreteness of the score, there is no visual evidence for a jump in the density of COMIPEMS score to one side of the cutoff or the other.[14]

There may be some concern that the differential dropout result comes from students having to travel farther to reach an IPN school. In particular, all but one of the IPN schools is located in the Federal District, while about half of COMIPEMS takers reside in the surrounding State of Mexico. To address this potential issue, we repeat the dropout exercise from Table 2.3 while restricting the sample to students living in the Federal District. The estimated admission effects, shown in Table 2.8, are noisier and somewhat smaller than those of the full sample, but remain large and in most cases statistically significant. For example, the estimated effect for the 7 point bandwidth without covariates falls from to 8.5 to 7.6 percentage points, while the effect with covariates falls from 8.2 to 5.8 percentage points. This result, combined with the differential effect with respect to GPA, suggests that dropout is induced by an academic issue rather than a prohibitively long commute.

Finally, there is substantial evidence that the difference in ENLACE taking rate between students admitted to and rejected from the IPN is due to students dropping out of school, rather than a data problem or rate at which 12^{th} graders in IPN schools take the ENLACE

[14]The test for a discontinuity in the density of the running variable, proposed by McCrary (2008), does not apply well to the case where the running variable has few points of support.

exam. The difference cannot be due to a lower rate of success in matching ENLACE takers from IPN schools to their COMIPEMS score. Of all ENLACE takers admitted to the IPN in the full sample, 99% are matched successfully to their COMIPEMS score. Another possibility that we can dismiss is that the IPN is selectively administering the exam to its best 12th graders. Although the ENLACE is taken at the end of the school year, schools must report the full roster of students in their final academic year to the Secretariat of Education so that all of those students can be programmed to take the exam. The ratio of actual exam takers to those programmed in the fall is nearly identical between the IPN and non-IPN schools (81%). Thus differential exam taking would have to be sufficiently premeditated to 1) fail to register low-ability students in the Fall and 2) systematically prevent the unregistered students from showing up at the exam. The exam is given by proctors from outside of the school. Administrators who run the ENLACE express doubt that a school subsystem would go through this trouble, especially when considering that ENLACE scores are not used to allocate resources or to incentivize or punish educators. Finally, because the ENLACE dataset used in this chapter includes years 2008 through 2010, it captures COMIPEMS takers from 2005 who took four or five years to graduate, and COMIPEMS takers from 2006 who took four years to graduate, instead of the standard three years. The differential exam taking rate, then, cannot be explained by students taking longer to graduate in the IPN schools but not dropping out.

As with any study using a regression discontinuity approach, there may be some skepticism in extrapolating the effects for marginal students to the rest of the sample. This would be a particular concern if there were few students near the margin compared to the total population of IPN students. The nature of the assignment mechanism, however, tends to bunch students near the cutoff of the school to which they are admitted, since a modestly higher score would often lead to admission to a preferred school. Similarly, many of the students admitted to the IPN *subsystem* are only a few points away from rejection to a non-IPN school. In fact, 34% of students admitted to an IPN school are within 7 COMIPEMS points of falling out of the IPN subsystem, while more than half are within 12 points of the boundary. The standard deviation of COMIPEMS score in the full sample is 17.95 and the within-school standard deviation for IPN students is 7.19, implying that a significant portion of IPN students are not far from the margin of the IPN subsystem.

2.6 Preference for elite schools

Students with lower GPAs are less likely to apply to elite schools. The findings in this chapter offer one way of rationalizing this empirical regularity. Students with a weak academic background face a less desirable dropout risk-academic reward trade-off and may respond rationally by choosing to avoid it altogether. This should be particularly true for students who are likely to gain admission to an elite school only at the margin.

To show that conditional on COMIPEMS score, high-achieving students are more likely to list an elite school as their first choice, the following local linear regressions are estimated for all observations within a 2-point bandwidth of each COMIPEMS point value c:

$$\text{elite}_{imtc} = \alpha_{mtc} + \beta_c \text{COMIPEMS}_i + \theta_c \text{GPA}_i + \varepsilon_{imtc}, \tag{2.4}$$

where elite_{imt} is a dummy variable equal to 1 if student i in year t from municipality/delegation m chose an elite school as her first choice, and GPA_i is middle school GPA. The municipality/delegation of residence of the student is added to control for the possible unequal geographic access to elite schools. The parameters of interest are the θ_c's, which measure the marginal effect (though not a causal relationship) of GPA on elite school preference only for students with COMIPEMS_i near c. Figure 2.5 graphs these coefficients and shows that for all values of COMIPEMS score above 70 points, i.e., that are high enough to gain admission to the least-competitive elite school, a higher GPA is correlated with higher rates of elite school preference. At a COMIPEMS score of 80, students with a 9.0 GPA are 15 percentage points more likely to select an elite school than those with a 7.0 GPA. This is a large difference, indicating that among students living in the same municipality or delegation and with the same possibility of admission to elite schools as a result of their COMIPEMS score, those with a lower GPA are much less likely to list an elite school as a first choice. The less favorable risk-reward tradeoff facing these students offers one way to explain this result.

2.7 Discussion

This chapter used Mexico City's high school allocation mechanism to identify the effects of admission to a subset of its elite public schools, relative to their non-elite counterparts. At least for marginally admitted students, elite schools present an important trade-off. Admission is found to significantly raise the probability of dropping out of school for the vast majority of marginal admits, with an average increase of 8.5 percentage points. Students with relatively low middle school GPAs are especially affected (e.g. a student with a GPA one standard deviation below the RD sample mean experiences a 14.9 percentage point increase in dropout probability), suggesting that elite schools are too challenging for some students and they either fail out or elect to leave school. On the other hand, elite admission appears to positively affect student test scores, increasing end-of-high school exam scores by 0.12 standard deviations under the assumption that dropout does not bias the estimated effect. Allowing for bias due to differential dropout lowers this estimate, but the results are fairly robust to assumptions about the severity of the bias. In particular, students' math scores seem to improve significantly with attending an elite school. The fact that this trade-off is, in expectation, worse for those from weaker academic backgrounds offers one possible explanation for the lower rate at which qualified students with low GPAs apply to elite high schools.

The existence of this trade-off between dropout probability and academic benefit highlights an important educational policy issue in Mexico. The current configuration of the high school education system does not facilitate lateral transfers of students between school subsystems, which are run by numerous entities at the local, state, and national level. Students who find that their current school is a bad fit cannot easily switch to a school that balances academic rigor, curriculum, and other characteristics to their taste, unless they drop out of school entirely and attempt to begin elsewhere with zero credits. The recently begun Comprehensive High School Education Reform (RIEMS) represents an attempt to rectify this by imposing a (partial) common curriculum. Such rigidity in the current system may explain why the academic benefit-dropout tradeoff is so strong in this chapter in comparison to studies in other countries. Our result highlights the value of flexibility in choice-based admissions systems so that the consequences of a "bad" choice can be mitigated, provided that lateral transfers to more competitive schools are not allowed as a means of gaming the current system.

2.8 Appendix: Method for assessing bias induced by differential dropout

In this appendix, we set forth a method for assessing the bias due to differential dropout induced by admission to an elite school, accounting for the heterogeneity of this dropout effect in the population of students. This procedure is in the spirit of previous bias-assessment and bias-bounding procedures and has some methodological similarities, but there are key differences. Lee (2005) trims the upper or lower part of the outcome distribution for treated (or untreated) observations, leading to sharp upper and lower bounds on the estimated treatment effect. When dropout from the sample is substantial, as it is here, these bounds can be wide. Still, Lee's approach leads to tighter bounds than worst-case bounds such as in Horowitz and Manski (1995). We take a less conservative approach that allows us to see how bad the bias must be in order to find a point estimate of zero effect, rather than assuming extreme outcomes for dropouts and then seeing if the resulting bounds contain zero or not. This is more in the spirit of Altonji et al. (2005), although our focus is on addressing dropout through imputation of outcomes rather than making assumptions about the correlation between error terms in the treatment and outcome equations for students with observed outcomes.

2.8.1 Basic setup

To understand why dropout (not taking the ENLACE exam) may induce bias in the estimated effect of admission on ENLACE score, first consider the case of one elite school with randomly assigned admission, where there are no covariates. There are two stages, one where it is determined whether the student drops out ($drop_i = 1$) and then the stage where

ENLACE score is observed for those who do not drop out:

$$\text{drop}_i = \begin{cases} 1, & \text{if } \text{drop}_i^* = \pi + \Delta\text{admit}_i + \nu_i > 0. \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{ENLACE}_i = \begin{cases} \alpha + \delta\text{admit}_i + \varepsilon, & \text{if } \text{drop}_i = 0. \\ - \ , & \text{otherwise.} \end{cases}$$

where π is the dropout rate among rejected students, Δ is the effect of elite admission on dropout probability, δ is the effect of elite admission on ENLACE score, and ν_i and ε_i are error terms. The problem is that if $\text{E}[\nu|\text{admit}] \neq 0$ and $cor(\varepsilon, \nu) \neq 0$, then the estimated effect of admission on score $(\widehat{\delta})$ will be biased. The following procedure will make no assumptions about $\text{E}[\nu|\text{admit}]$ or $cor(\varepsilon, \nu)$, but rather see how severe the effects of differential dropout must be, in particular how poorly (or how well) the rejection- or admission-induced dropouts must do compared to the ENLACE takers, in order to attribute the entire treatment effect to this bias.

We begin by imposing a monotonicity assumption: admission may not increase the probability of taking for some students and decrease it for others. This is satisfied by assuming a homogeneous treatment effect of admit_i on drop_i^*, as in the setup above.

2.8.2 Decomposition of mean score

First, suppose that admission increases the probability of dropout, so $\Delta > 0$. The hypothetical average ENLACE score, regardless of whether the exam is actually taken, is decomposed separately for rejected and admitted groups as follows:

$$\overline{\text{ENLACE}}^R = \underbrace{\frac{1}{n_1^R} \sum_{i:\nu_i < -\pi-\Delta} \text{ENLACE}_i}_{\text{(observed)}} + \underbrace{\frac{1}{n_2^R} \sum_{\substack{i:\nu_i < -\pi, \\ \nu_i > -\pi-\Delta}} \text{ENLACE}_i}_{\text{(observed)}} + \underbrace{\frac{1}{n_3^R} \sum_{i:\nu_i > -\pi} \text{ENLACE}_i}_{\text{(not observed)}}$$

$$\overline{\text{ENLACE}}^A = \underbrace{\frac{1}{n_1^A} \sum_{i:\nu_i < -\pi-\Delta} \text{ENLACE}_i}_{\text{(observed)}} + \underbrace{\frac{1}{n_2^A} \sum_{\substack{i:\nu_i < -\pi, \\ \nu_i > -\pi-\Delta}} \text{ENLACE}_i}_{\textbf{(not observed)}} + \underbrace{\frac{1}{n_3^A} \sum_{i:\nu_i > -\pi} \text{ENLACE}_i}_{\text{(not observed)}}$$

where n_1^R is the number of students who were rejected from the elite school and would take the exam regardless of admissions outcome, n_2^R is the number of rejected students who take the exam when rejected but would not when admitted, n_3^R is the number of rejected students who did not take the exam (and would not have if admitted), and n_1^A, n_2^A, and n_3^A indicate the number of students in the corresponding groups for those students who are admitted to the elite school.

The first sum in each group is the set of students who take the exam regardless of admission status, so their scores are always observed. The final sum is over students who never take the exam, so their scores are never observed. The middle sum is the set of students who take the exam if rejected but not if they are admitted. This is analogous to the "compliers" in an IV design, where compliance is dropping out and having no score observed. The bias in $\widehat{\delta}$ comes from including the scores of compliers in the rejected group but not the admitted group.

Of course, the set of compliers in the admitted and rejected groups is unknown, but under randomization its size is not. How big is the set of missing compliers in the admitted group? To answer this, consider the following expressions for the count of observed exam scores as a proportion of all students in the group:

$$\frac{N_{obs}^R}{N^R} = \frac{n_1^R}{N^R} + \frac{n_2^R}{N^R} = 1 - \pi$$

$$\frac{N_{obs}^A}{N^A} = \frac{n_1^A}{N^A} = 1 - (\pi + \Delta)$$

2.8.3 Defining weights for dropouts

Because of randomized admission, we know that $\frac{n_1^R}{N^R} = \frac{n_1^A}{N^A}$ and $\frac{n_2^R}{N^R} = \frac{n_2^A}{N^A}$. It follows that $\frac{n_2^A}{N^A} = \Delta$, meaning that we are "missing" $n_2^A = \Delta N^A$ compliers in the admitted group. The goal of this procedure is to add these "missing" admitted dropouts back into the sample with increasingly low imputed scores until their addition causes the estimated admission effect to be zero. We will do this by weighting the imputed scores of all $N^A - n_1^A$ dropouts such that the equivalent of n_2^A of them are added. The proper weight is given by:

$$\omega_i = \frac{n_2^A}{(N^A - n_1^A)} = \frac{\Delta}{(\pi + \Delta)}$$

This weight can be estimated easily, as Δ and π are estimated in the dropout prediction equation. All admitted and rejected students who took the exam have $\omega_i = 1$ and all admitted students without a test score have $\omega_i = 0$.

If $\Delta < 0$, then the result is derived in the same way, and $\omega_i = \frac{-\Delta}{\pi}$ is applied for dropouts in the rejected group and $\omega_i = 0$ for dropouts in the admitted group.

2.8.4 Imputing scores for the missing observations

Imputation of scores for the admitted students can be done by quantile regression. In the simple case of randomization with no covariates, the equation for this is:

$$\widehat{\text{ENLACE}}_i = Q_m\left(\text{ENLACE}|\text{admit}_i = 1\right)$$

where Q_m is the quantile function giving the m^{th} quantile of the observed score distribution among the admitted students. If the imputation is for rejected students, the conditional quantile is taken among the rejected students.

2.8.5 Estimating the admission effect on scores including imputed observations

Estimation of the admission effect proceeds as it would without the imputed observations, with two obvious differences: the imputed observations are included and the observations are weighted. If the resulting $\widehat{\delta}$ is still positive, then the conditional quantile is decreased (or increased, if the imputation is for rejected students) and the exercise is carried out again until the selected quantile is sufficiently low (high) that the admission effect is zero.

2.8.6 Adding covariates

We have seen that the predicted probability of dropout depends on covariates, and that the effect of admission on dropout also depends on covariates. In fact, the predicted effect of admission on dropout is negative for some students and positive for others. Here the procedure is extended to allow for covariates, so the dropout equation is redefined as:

$$\widehat{\text{drop}}_i = \begin{cases} 1, & \text{if } \text{drop}_i^* = \widetilde{\pi} + \widetilde{\Delta}\text{admit}_i + X_i\gamma + (X_i \times \text{admit}_i)\,\theta + \nu_i > 0. \\ 0, & \text{otherwise.} \end{cases}$$

The randomization assumption is retained, while the monotonicity assumption is relaxed slightly: *conditional on covariates* X_i, admission may not increase the probability of taking for some students and decrease it for others. Again, this is satisfied if we assume a treatment effect that is homogeneous conditional on observables, as presented in the equation above.

For notational convenience, define $\pi_i \equiv \widetilde{\pi} + X_i\gamma$ as the predicted probability of dropout if the student is rejected, conditional on covariates. Also define $\Delta_i \equiv \widetilde{\Delta} + X_i\theta$ as the change in dropout probability due to admission for a student with covariate values X_i.

The decomposition of mean ENLACE scores is almost identical to the no-covariate case, except that the admission effect Δ is replaced by Δ_i and the baseline dropout π is replaced by π_i. Now there are some observations with $\Delta_i > 0$ and some with $\Delta_i < 0$:

$$\overline{\text{ENLACE}}^R =$$
$$1(\Delta_i > 0) \times$$
$$\left[\frac{1}{n_1^R} \overset{\text{(observed)}}{\sum_{\substack{i:\nu_i<-\pi_i,\\ \nu_i<-\pi_i-\Delta_i}}} \text{ENLACE}_i + \frac{1}{n_2^R} \overset{\textbf{(not observed)}}{\sum_{\substack{i:\nu_i>-\pi,\\ \nu_i<-\pi-\Delta}}} \text{ENLACE}_i + \frac{1}{n_3^R} \overset{\text{(not observed)}}{\sum_{\substack{i:\nu_i>-\pi,\\ \nu_i>-\pi-\Delta}}} \text{ENLACE}_i \right]$$
$$+1(\Delta_i < 0) \times$$
$$\left[\frac{1}{n_1^R} \overset{\text{(observed)}}{\sum_{\substack{i:\nu_i<-\pi_i,\\ \nu_i<-\pi_i-\Delta_i}}} \text{ENLACE}_i + \frac{1}{n_2^R} \overset{\text{(observed)}}{\sum_{\substack{i:\nu_i<-\pi,\\ \nu_i>-\pi-\Delta}}} \text{ENLACE}_i + \frac{1}{n_3^R} \overset{\text{(not observed)}}{\sum_{\substack{i:\nu_i>-\pi,\\ \nu_i>-\pi-\Delta}}} \text{ENLACE}_i \right]$$

$$\overline{\text{ENLACE}}^A =$$
$$1(\Delta_i > 0) \times$$
$$\left[\frac{1}{n_1^A} \overset{\text{(observed)}}{\sum_{\substack{i:\nu_i<-\pi_i,\\ \nu_i<-\pi_i-\Delta_i}}} \text{ENLACE}_i + \frac{1}{n_2^A} \overset{\text{observed}}{\sum_{\substack{i:\nu_i>-\pi,\\ \nu_i<-\pi-\Delta}}} \text{ENLACE}_i + \frac{1}{n_3^A} \overset{\text{(not observed)}}{\sum_{\substack{i:\nu_i>-\pi,\\ \nu_i>-\pi-\Delta}}} \text{ENLACE}_i \right]$$
$$+1(\Delta_i < 0) \times$$
$$\left[\frac{1}{n_1^A} \overset{\text{(observed)}}{\sum_{\substack{i:\nu_i<-\pi_i,\\ \nu_i<-\pi_i-\Delta_i}}} \text{ENLACE}_i + \frac{1}{n_2^A} \overset{\textbf{(not observed)}}{\sum_{\substack{i:\nu_i<-\pi,\\ \nu_i>-\pi-\Delta}}} \text{ENLACE}_i + \frac{1}{n_3^A} \overset{\text{(not observed)}}{\sum_{\substack{i:\nu_i>-\pi,\\ \nu_i>-\pi-\Delta}}} \text{ENLACE}_i \right]$$

The unobserved sets of students for which scores need to be imputed are indicated in bold. By the same derivation as the no-covariate case but with the covariates and their interactions included, we derive the following set of weights:

1. Admitted, dropped out, $\Delta_i > 0$ (increased dropout probability due to admission): $\omega_i = \frac{\Delta_i}{\pi_i + \Delta_i}$
2. Rejected, dropped out, $\Delta_i < 0$ (increased dropout chance due to rejection): $\omega_i = \frac{-\Delta_i}{\pi_i}$
3. Did not drop out: $\omega_i = 1$
4. Otherwise: $\omega_i = 0$.

This is the same as the no-covariate case except that it allows students in both the rejected and admitted groups to be weighted up, depending on the sign of the conditional differential dropout probability Δ_i.

The rest of the process is the same as the no-covariate case but with one change: imputation is done via quantile regression, now conditional on the full set of covariates X_i, but imputing a low quantile for the admitted students and a high quantile for the rejected students. Here, we impute the m^{th} conditional quantile for the admitted students and $1-m^{\text{th}}$ conditional quantile for the rejected students:

$$\widehat{\text{ENLACE}}_i = \begin{cases} Q_m\left(\text{ENLACE}|X_i, admit_i = 1\right), & \text{if } \Delta_i > 0, \text{admit}_i = 1, \text{drop}_i = 1. \\ Q_{1-m}\left(\text{ENLACE}|X_i, \text{admit}_i = 0\right), & \text{if } \Delta_i < 0, \text{admit}_i = 0, \text{drop}_i = 1. \end{cases}$$

2.8.7 For which set of students is the admission effect estimated?

If differential dropout were only predicted to be positive for admitted students, then the imputation exercise would allow us to estimate the (penalized) admission effect for the group of students who do not drop out if rejected (regardless of whether they drop out if admitted). But here we have both students who are more likely to drop out when admitted and students who are more likely to drop out when they are rejected. So this exercise is performed for the group of students who are not "always-quitters" – students for whom admission *and/or* rejection would lead to taking the ENLACE. This is not a commonly-used group in the treatment effects literature, but it has some appeal. It can be thought of as the whole group of students for whom we can conceive of comparing outcomes between groups of schools – we should never compare on the basis of students who will always drop out, but we may indeed want to include in the comparison students who drop out in one group of schools but not the other, as well as those who always stay in school.

2.8.8 Extension to regression discontinuity

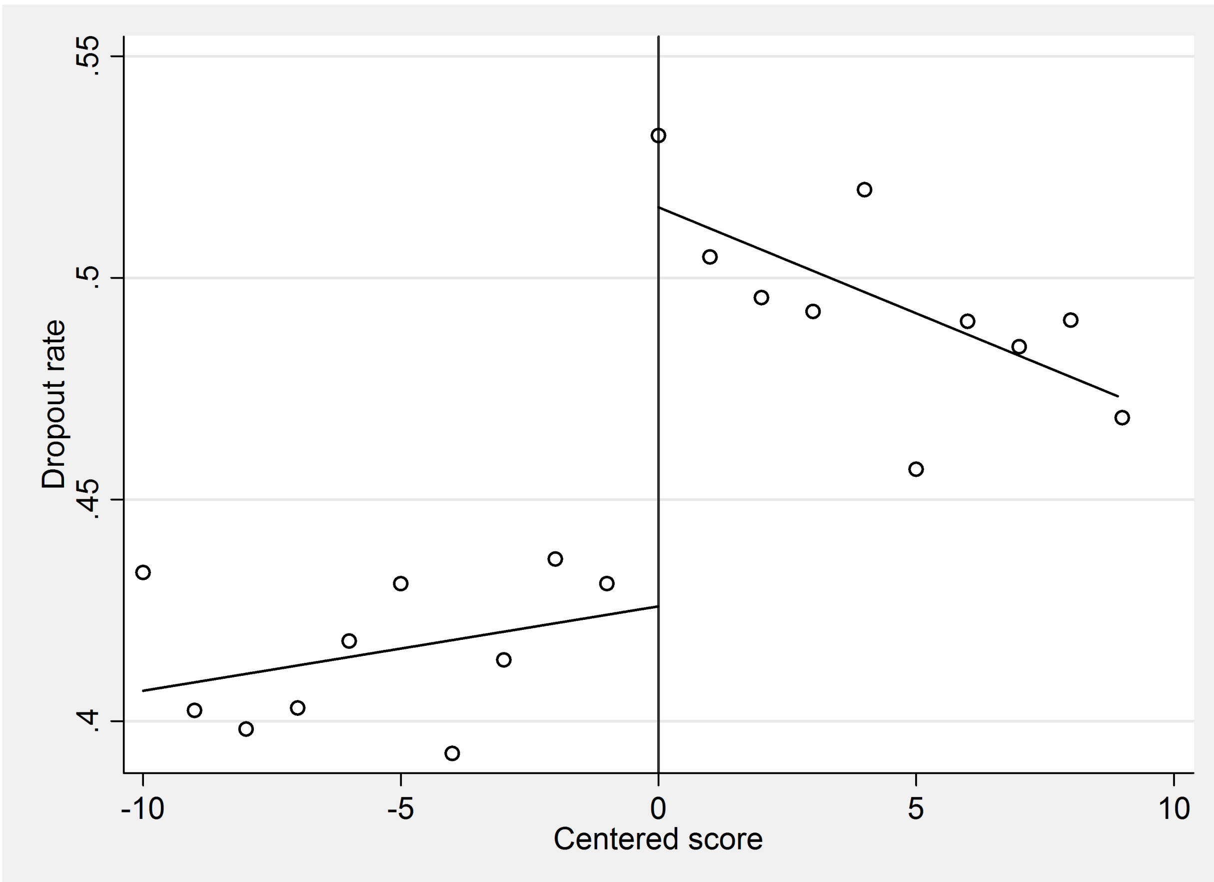
The previous sections assumed randomization into treatment. To apply the same procedure to regression discontinuity, we use the assumption that in a sufficiently small window about the threshold, treatment is as good as randomly assigned conditional on a properly-specified function of the running variable (Imbens and Lemieux 2008). Thus we can simply include a function of COMIPEMS score (normalized to zero at the cutoff score) in the dropout equation and in the ENLACE score equation. We also include interactions between the de-meaned covariates and COMIPEMS score, to allow the possibility that the covariates' influence varies with COMIPEMS score:

$$\begin{aligned} \text{drop}_i^* &= \widetilde{\pi} + \widetilde{\Delta}\text{admit}_i + \beta_1 \text{COMIPEMS}_i + \beta_2\left(\text{COMIPEMS}_i \times \text{admit}_i\right) + X_i\gamma \\ &+ \left(X_i \times \text{admit}_i\right)\theta + \left(X_i \times \text{COMIPEMS}_i\right)\phi_i + \left(X_i \times \text{COMIPEMS}_i \times \text{admit}_i\right)\phi_2 + \nu_i \end{aligned}$$

The rest of the procedure is the same, since $\Delta_i \equiv \widetilde{\Delta} + X_i\theta$ is still the difference in ENLACE taking probability due to admission. The predicted probability of dropout given rejection, p_i, can be estimated in the same way as before, but including the COMIPEMS terms; likewise for the imputation of the conditional quantiles.

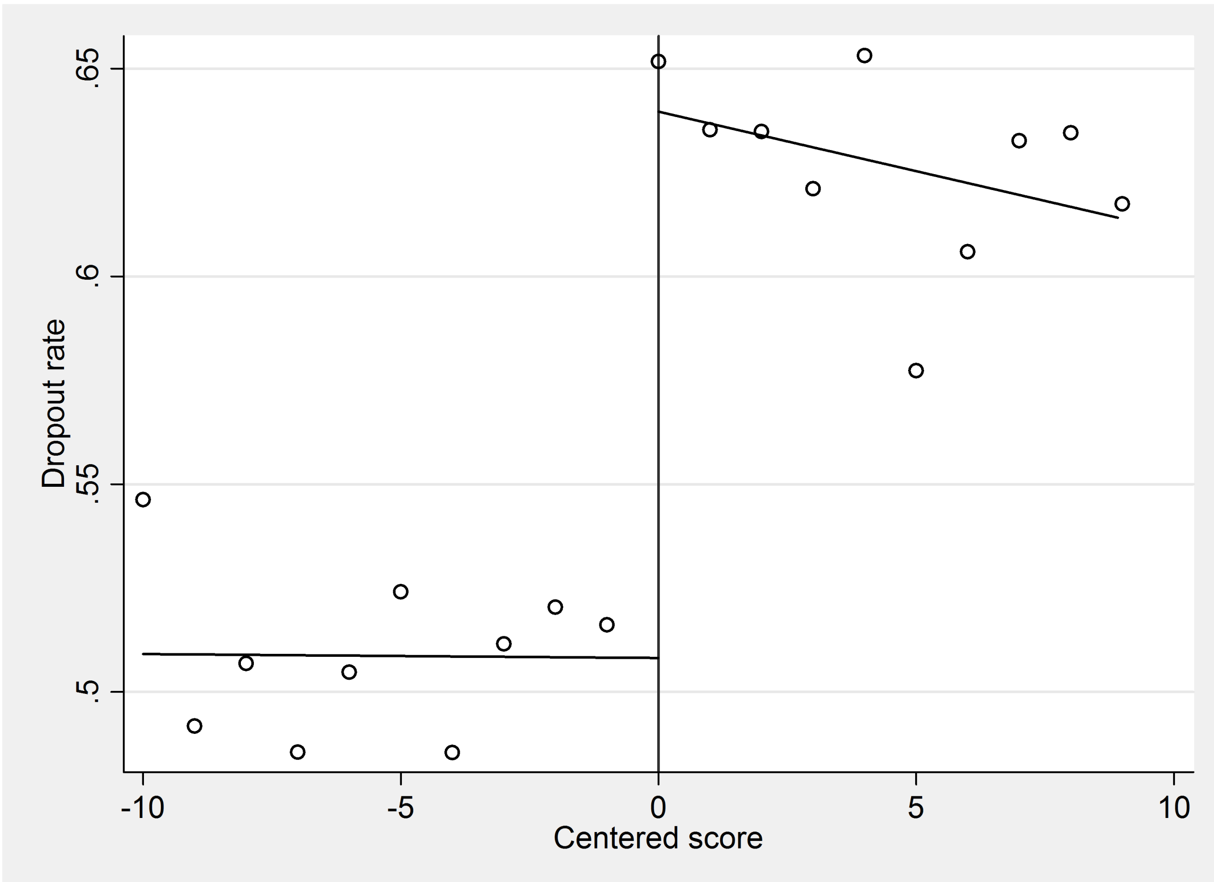
2.9 Figures

Figure 2.1a: Dropout rate for students near IPN system cutoff: all students



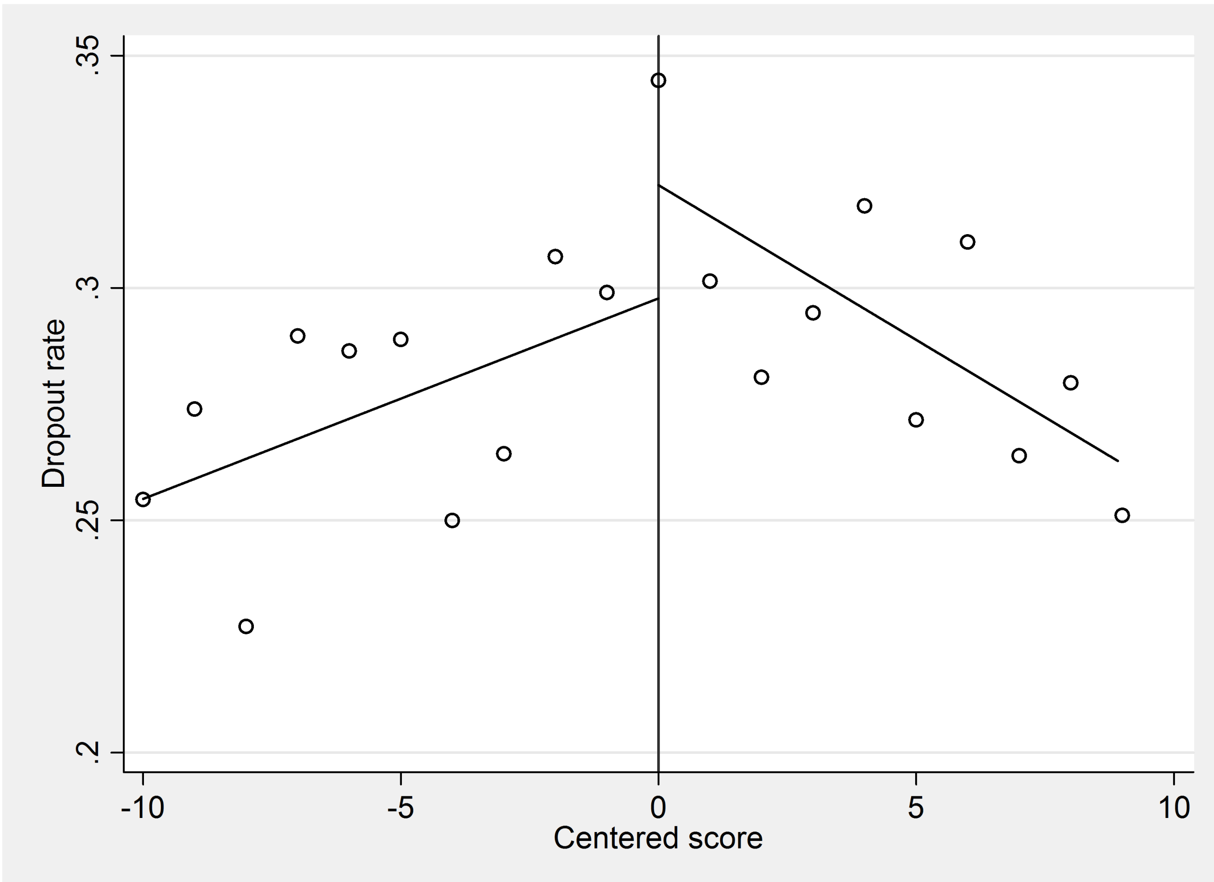
Scatterplot is of mean dropout rate vs. centered COMIPEMS score, where dropout has been de-meaned by regressing dropout on a set of cutoff school fixed effects and middle school GPA and using the residuals. Lines represent a separate linear fit on each side of the admissions cutoff.

Figure 2.1b: Dropout rate for students near IPN system cutoff: students with middle school GPA below 8.5



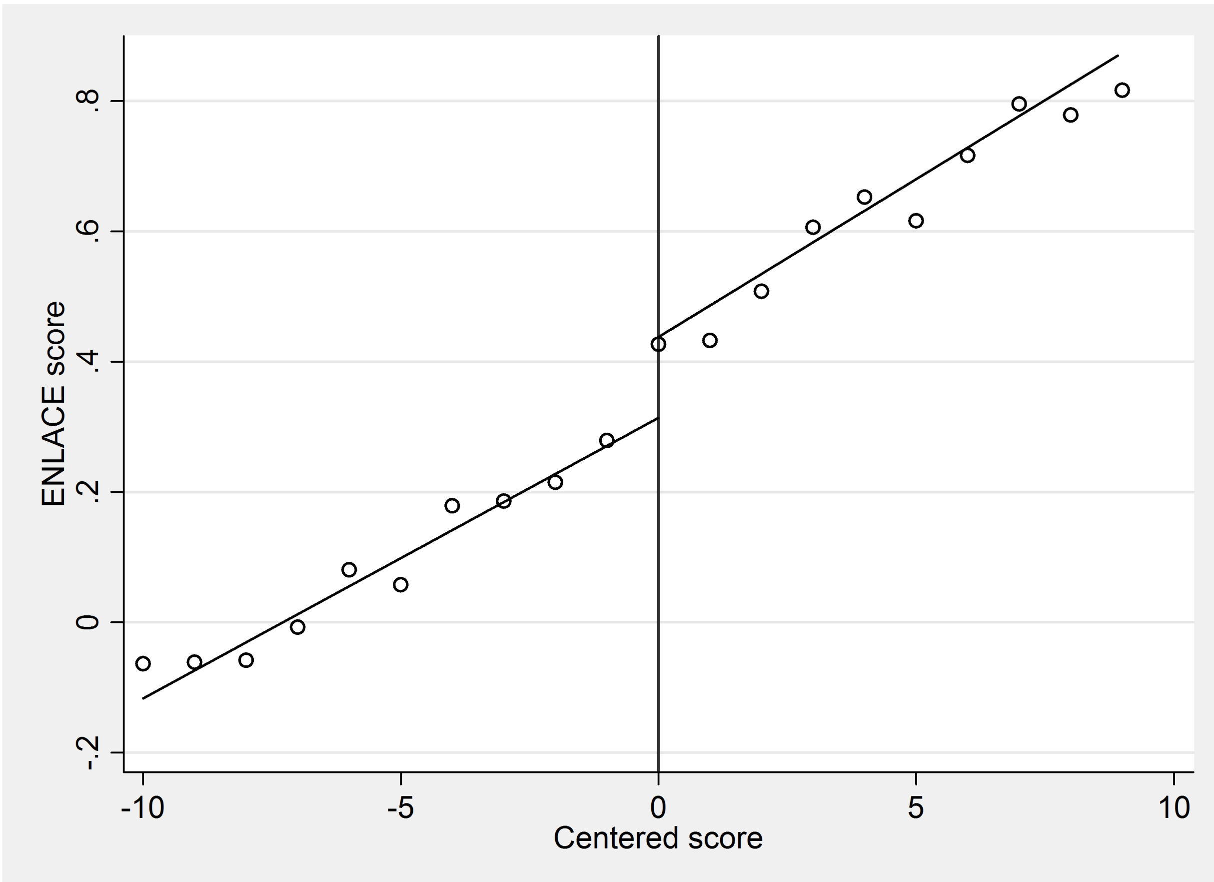
Scatterplot is of mean dropout rate vs. centered COMIPEMS score, where dropout has been de-meaned by regressing dropout on a set of cutoff school fixed effects and middle school GPA and using the residuals. Lines represent a separate linear fit on each side of the admissions cutoff.

Figure 2.1c: Dropout rate for students near IPN system cutoff: students with middle school GPA of at least 8.5



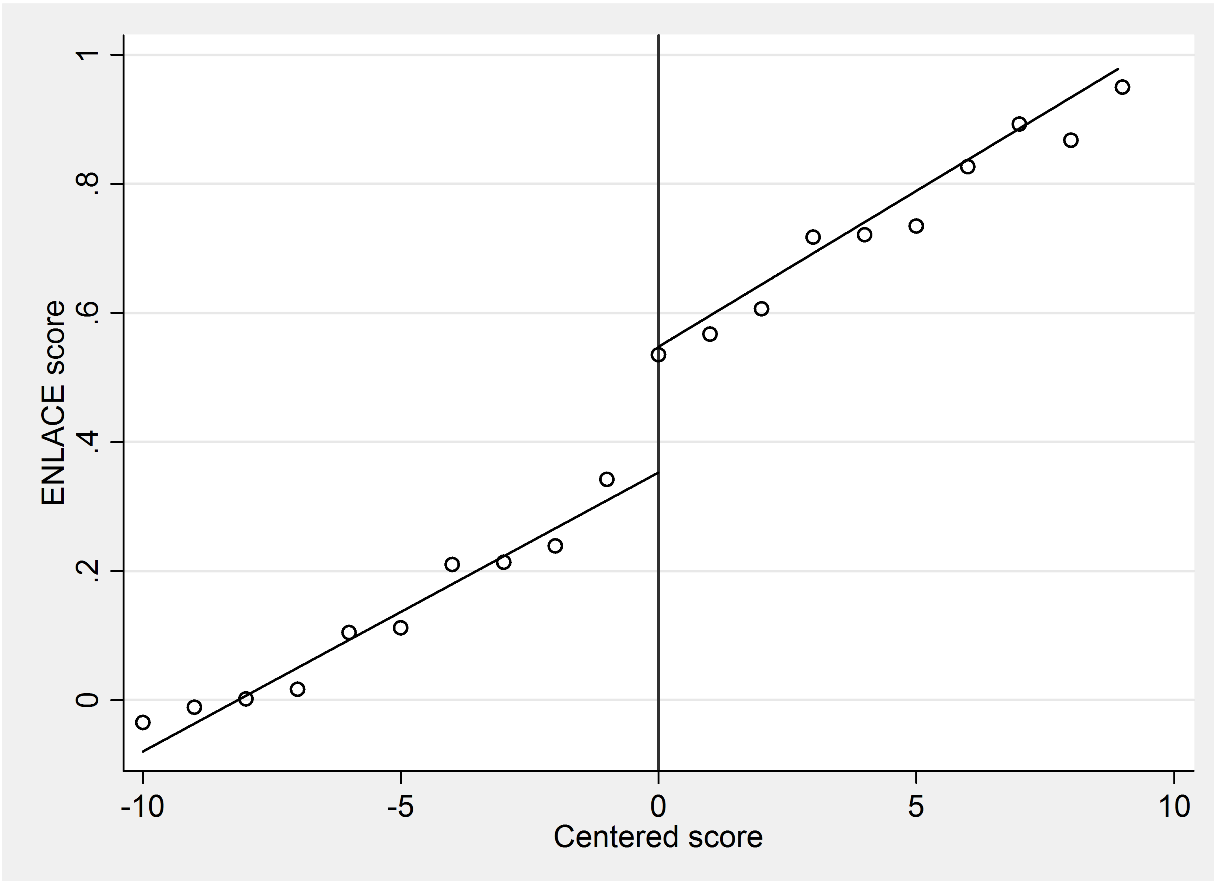
Scatterplot is of mean dropout rate vs. centered COMIPEMS score, where dropout has been de-meaned by regressing dropout on a set of cutoff school fixed effects and middle school GPA and using the residuals. Lines represent a separate linear fit on each side of the admissions cutoff.

Figure 2.2a: ENLACE performance for students near IPN system cutoff: composite score (math and Spanish)



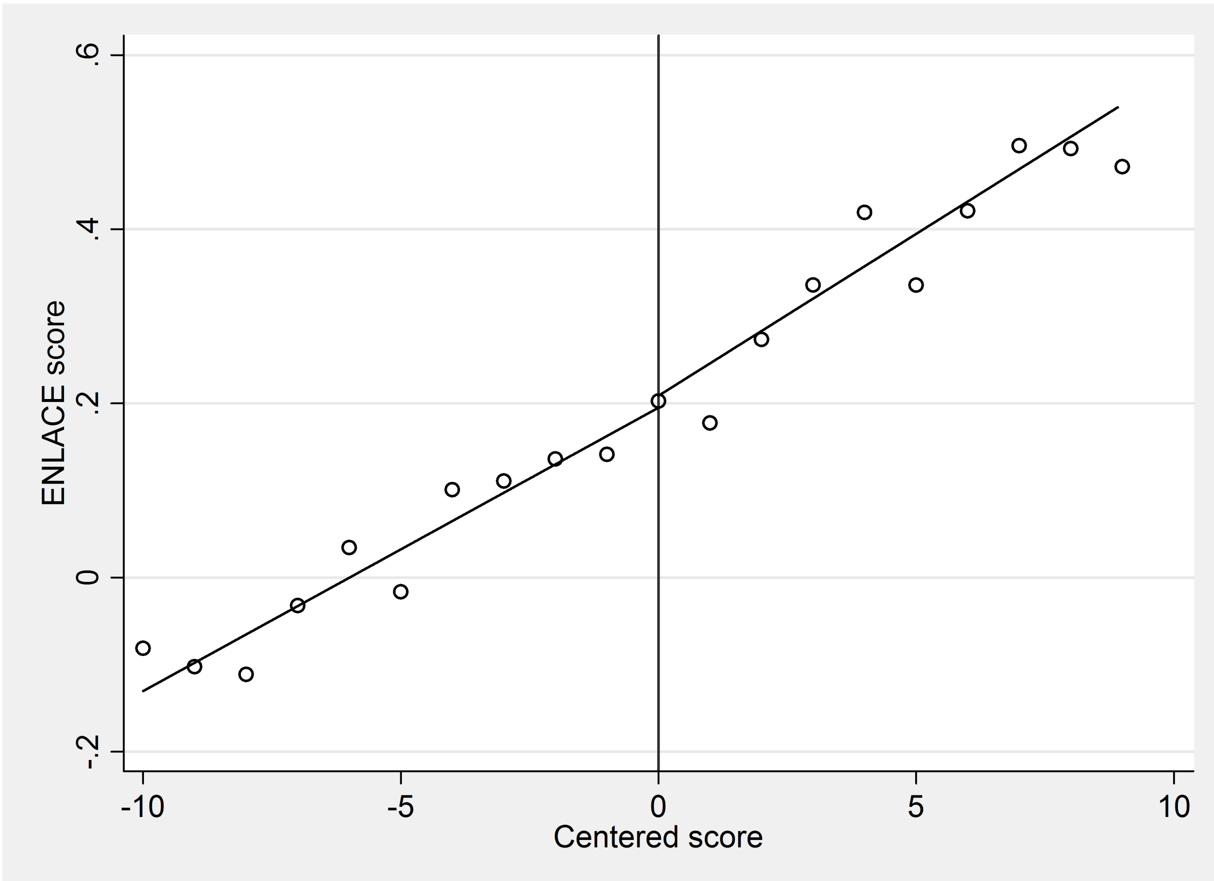
Scatterplot is of mean ENLACE score vs. centered COMIPEMS score, where ENLACE score has been de-meaned by regressing ENLACE score on a set of cutoff school fixed effects and middle school GPA and using the residuals. Lines represent a separate linear fit on each side of the admissions cutoff.

Figure 2.2b: ENLACE performance for students near IPN system cutoff: math score



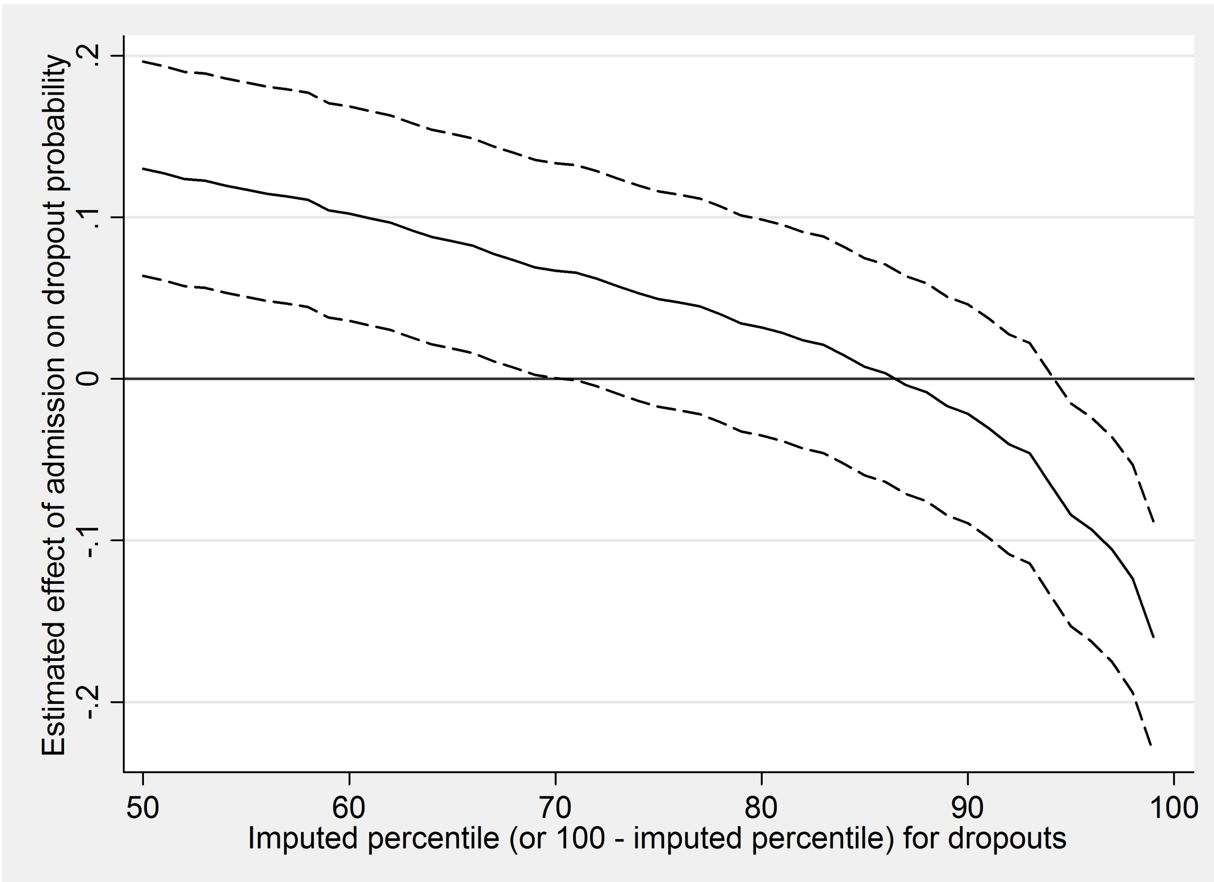
Scatterplot is of mean ENLACE score vs. centered COMIPEMS score, where ENLACE score has been de-meaned by regressing ENLACE score on a set of cutoff school fixed effects and middle school GPA and using the residuals. Lines represent a separate linear fit on each side of the admissions cutoff.

Figure 2.2c: ENLACE performance for students near IPN system cutoff: Spanish score



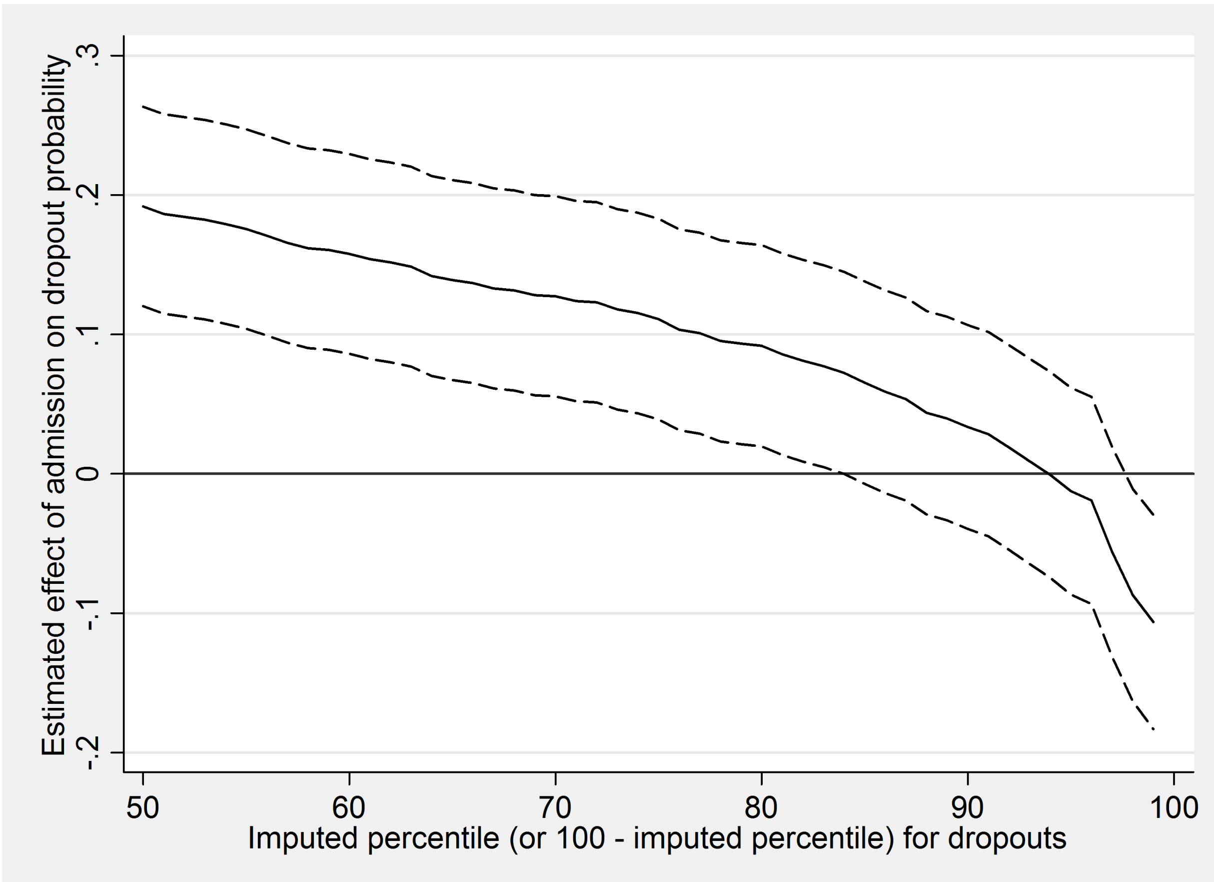
Scatterplot is of mean ENLACE score vs. centered COMIPEMS score, where ENLACE score has been de-meaned by regressing ENLACE score on a set of cutoff school fixed effects and middle school GPA and using the residuals. Lines represent a separate linear fit on each side of the admissions cutoff.

Figure 2.3a: Estimated effect of admission on ENLACE score under different penalized imputations for dropouts: composite score (math and Spanish))



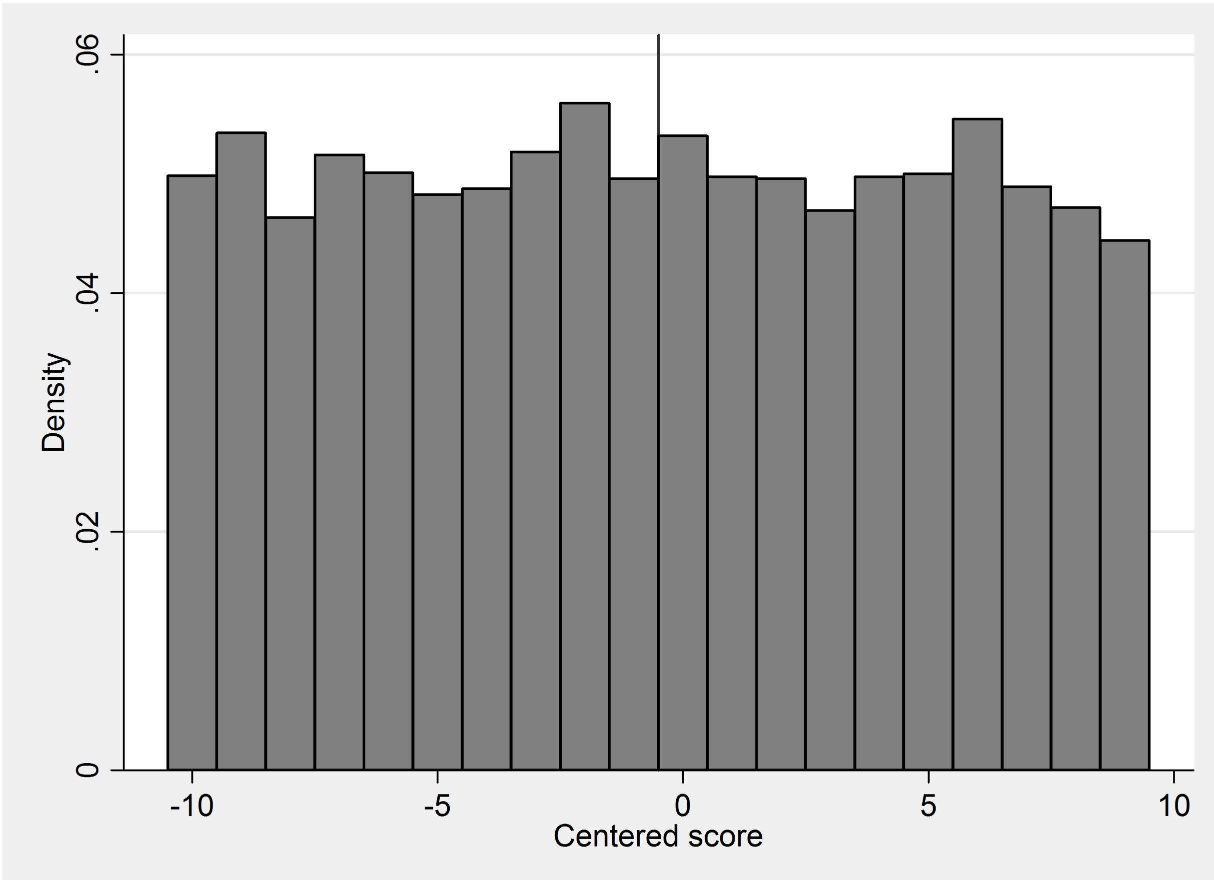
Solid line is the estimated admission coefficient from the RD specification corresponding with column 9 in Table 2.5a, where the missing ENLACE scores have been imputed using the procedure in the appendix with the percentile given on the x-axis. The imputed percentile corresponds to dropouts in the rejected group and 100 – imputed percentile corresponds to dropouts in the admitted group. Dashed lines are 95% confidence intervals from unadjusted standard errors.

Figure 2.3b: Estimated effect of admission on ENLACE score under different penalized imputations for dropouts: math score



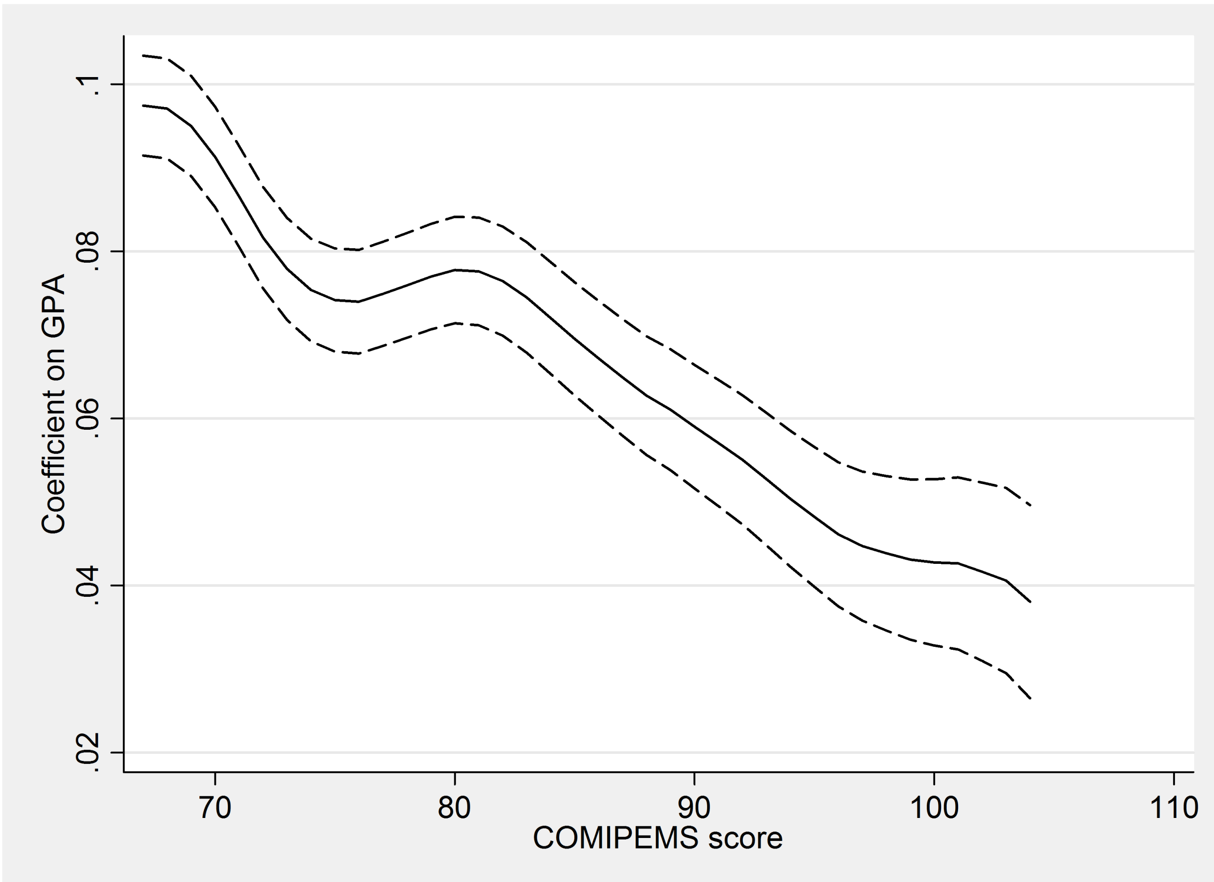
Solid line is the estimated admission coefficient from the RD specification corresponding with column 9 in Table 2.5b, where the missing ENLACE scores have been imputed using the procedure in the appendix with the percentile given on the x-axis. The imputed percentile corresponds to dropouts in the rejected group and 100 – imputed percentile corresponds to dropouts in the admitted group. Dashed lines are 95% confidence intervals from unadjusted standard errors.

Figure 2.4: Density of student scores around IPN system cutoffs



Histogram shows the density of centered COMIPEMS score for the 10-point regression discontinuity sample.

Figure 2.5: Partial correlation of middle school GPA with elite school first-choice preference



Solid line is a smoothed line through the $\widehat{\theta}_c$ coefficients from estimating: $\text{elite}_{imt} = \alpha_{mtc} + \beta_c \text{COMIPEMS}_i + \theta_c X_i + \varepsilon_{imtc}$, where elite_{imt} is a dummy variable equal to 1 if student i in year t from municipality/delegation m chose an elite school as her first choice, and X_i is middle school GPA. The lines represent the partial correlation between X_i and elite school preference for different COMIPEMS score values. Dashed lines are the 95% confidence intervals for the estimated $\widehat{\theta}_c$'s.

2.10 Tables

Table 2.1: Characteristics of students eligible for assignment

	(1)	(2)	
	All students	Students within 10 points of IPN cutoff	p-value for equality of (1) and (2)
Male	0.46	0.63	0.00
Maximum of mother's and father's education	10.18	10.61	0.00
	(3.35)	(3.12)	
Family income (thousand pesos/month)[a]	4.22	4.45	0.00
	(3.35)	(3.18)	
Hours studied per week	5.19	5.41	0.00
	(3.26)	(3.24)	
Index of parental effort[b]	0.00	0.05	0.00
	(1.00)	(0.97)	
Student is employed	0.04	0.04	0.38
Middle school grade point average (of 10)	8.10	8.24	0.00
	(0.82)	(0.74)	
Number of schools ranked	9.31	10.25	0.00
	(3.59)	(3.66)	
Elite school as first choice	0.64	1.00	0.00
COMIPEMS examination score	63.74	74.55	0.00
	(17.95)	(7.90)	
Dropped out (only for students assigned to a non-UNAM school)	0.50	0.45	0.00
ENLACE examination score (for those who took the exam)[c]	-0.01	0.34	0.00
	(0.99)	(0.80)	
Observations	354,581	11,978	

Note. Standard deviations in parentheses.

[a] Average 2005-2006 exchange rate was 10.9 pesos/dollar.

[b] The parental effort index is constructed by averaging the scores (1-4 ordinal scale) for 13 questions about parental effort and involvement from the survey filled out at the time of COMIPEMS registration. The survey asked "How often do your parents or adults with whom you live do the following activities?" for activities such as "help you with schoolwork" and "attend school events." The measure is normalized to have mean zero and standard deviation of 1 in the sample of all students.

[c] The normalized ENLACE examination score is constructed by subtracting off the year-specific mean score for all examinees in public high schools within the COMIPEMS geographic area and dividing by the year-specific standard deviation from this same sample.

Table 2.2: Correlates of high school dropout

	(1)	(2)	(3)
COMIPEMS score	-0.24***	-0.29***	-0.31***
	(0.01)	(0.01)	(0.01)
Middle school GPA	-18.38***	-17.68***	-18.53***
	(0.14)	(0.14)	(0.14)
Parental education (years)	-0.35***	-0.48***	-0.37***
	(0.03)	(0.03)	(0.03)
Family income (thousand pesos/mo)	-0.07**	-0.17***	-0.08**
	(0.04)	(0.03)	(0.04)
Male	-0.03	-0.22	-0.43**
	(0.21)	(0.22)	(0.21)
Hours studied per week	-0.23***	-0.30***	-0.24***
	(0.03)	(0.03)	(0.03)
Parental effort index	-1.31***	-1.12***	-1.31***
	(0.10)	(0.10)	(0.10)
Employed	7.58***	7.39***	7.58***
	(0.48)	(0.48)	(0.48)
Exam year 2006	2.24***	2.05***	2.39***
	(0.19)	(0.21)	(0.19)
Admitted to IPN school			7.86***
			(0.44)
IPN school as first choice			-0.29
			(0.33)
Admitted high school fixed effects	NO	YES	NO
Observations	233,359	233,359	233,359
R^2	0.12	0.16	0.12
Mean of dependent variable	48.84	48.84	48.84

Note. Dependent variable is dropout*100.
Sample excludes students admitted to an UNAM high school, since these schools do not proctor the ENLACE exam used as a proxy for graduation. Also excluded are students who were not assigned to any school during the allocation process.
Huber-White robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.3: Regression discontinuity estimates of effect of IPN admission on dropout

	Bandwidth: 4				Bandwidth: 7				Bandwidth: 10			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Admitted to IPN	7.860	7.693	7.901	(One per	8.533	8.176	8.239	(One per	9.071	9.264	9.156	(One per
	(2.628)	(2.754)	(2.754)	cutoff)	(2.042)	(2.133)	(2.138)	cutoff)	(1.810)	(1.880)	(1.890)	cutoff)
	[0.00]***	[0.02]**	[0.01]***		[0.01]***	[0.00]***	[0.00]***		[0.00]***	[0.00]***	[0.00]***	
Middle school GPA * Admitted			-7.640	-8.028			-9.064	-9.937			-8.104	-9.254
			(4.123)	(4.175)			(3.156)	(3.202)			(2.767)	(2.813)
			[0.08]*	[0.08]*			[0.01]***	[0.01]***			[0.03]**	[0.03]**
Parental education * Admitted			-0.515	-0.689			-0.924	-0.977			-0.881	-0.946
			(2.415)	(2.469)			(1.121)	(1.131)			(0.709)	(0.712)
			[0.15]	[0.07]*			[0.04]	[0.03]			[0.06]*	[0.08]*
Family income * Admitted			-0.874	-0.829			-0.857	-0.848			-0.743	-0.779
			(0.998)	(1.003)			(0.755)	(0.757)			(0.668)	(0.670)
			[0.05]**	[0.03]**			[0.07]*	[0.08]*			[0.07]*	[0.04]**
Male * Admitted			5.031	7.419			3.971	4.578			4.669	4.997
			(6.067)	(6.275)			(4.699)	(4.893)			(4.144)	(4.339)
			[0.33]	[0.40]			[0.52]	[0.48]			[0.22]	[0.33]
Hours studied per week * Admitted			0.297	0.244			0.274	0.234			0.030	0.019
			(0.869)	(0.872)			(0.679)	(0.681)			(0.601)	(0.602)
			[0.76]	[0.78]			[0.80]	[0.79]			[0.98]	[0.99]
Parental effort index * Admitted			-1.424	-2.254			0.462	0.088			1.175	0.960
			(2.996)	(3.019)			(2.328)	(2.336)			(2.047)	(2.053)
			[0.48]	[0.42]			[0.84]	[0.96]			[0.56]	[0.64]
Employed * Admitted			-1.765	-1.870			-4.903	-4.357			-7.498	-7.016
			(14.043)	(14.104)			(10.791)	(10.808)			(9.615)	(9.621)
			[0.78]	[0.80]			[0.28]	[0.31]			[0.04]**	[0.07]*
Student covariates	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES
AIC-optimal polynomial order	1	1	1	1	1	1	1	1	1	1	1	1
Observations	6,174	5,208	5,208	5,208	9,590	8,113	8,113	8,113	11,978	10,138	10,138	10,138
R^2	0.030	0.146	0.153	0.155	0.018	0.127	0.131	0.132	0.018	0.128	0.132	0.133
Mean of dependent variable	45.93	45.12	45.12	45.12	46.07	45.40	45.40	45.40	45.47	44.91	44.91	44.91

Note. Dependent variable is dropout*100.

All regressions include cutoff fixed effects, cutoff-specific controls for COMIPEMS score and COMIPEMS score * admitted), a dummy variable for taking the 2006 COMIPEMS exam, and a dummy variable for whether the UNAM exam was taken. Columns (2-4), (6-8), and (10-12) include covariates interacted with cutoff fixed effects. Columns (3-4), (7-8), and (11-12) include controls for (COMIPEMS score * covariate) and (COMIPEMS score * covariate * admitted) for each of the covariates. Each of the covariates is de-meaned.

Unadjusted standard errors are in parentheses. Wild cluster bootstrapped p-values for the null hypothesis of the coefficient equaling zero are in brackets, where the cluster is the centered COMIPEMS score.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.4: Balance of covariates before and after assignment

Panel a. At time of assignment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable	Middle school GPA	Parental education	Family income (thousand pesos/mo)	Male	Hours studied per week	Parental effort index	Employed
Admitted to IPN	-0.008	0.139	0.099	-0.008	0.222	0.000	0.001
	(0.026)	(0.120)	(0.123)	(0.017)	(0.124)	(0.037)	(0.008)
	[0.67]	[0.10]*	[0.40]	[0.51]	[0.04]**	[0.98]	[0.70]
Observations	11,978	10,842	10,728	11,978	10,836	10,914	10,508
R^2	0.065	0.021	0.014	0.110	0.019	0.007	0.004
Mean of dependent variable	8.238	10.61	4.453	0.626	5.415	0.0488	0.0406
S.D. of dependent variable	0.741	3.121	3.180	0.484	3.236	0.968	0.197
p-value, joint significance of admission coefficients[a]							0.40

Panel b. After dropout

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable	Middle school GPA	Parental education	Family income (1000 pesos/mo)	Male	Hours studied per week	Parental effort index	Employed
Admitted to IPN	0.102	0.360	0.273	-0.034	0.336	0.006	-0.002
	(0.035)	(0.163)	(0.169)	(0.023)	(0.172)	(0.050)	(0.010)
	[0.01]***	[0.06]*	[0.06]*	[0.12]	[0.05]**	[0.92]	[0.60]
Observations	6,532	5,968	5,907	6,532	5,951	5,992	5,769
R^2	0.093	0.025	0.015	0.117	0.020	0.011	0.009
Mean of dependent variable	8.451	10.75	4.523	0.583	5.626	0.105	0.0328
S.D. of dependent variable	0.729	3.115	3.208	0.493	3.281	0.950	0.178
p-value, joint significance of admission coefficients[a]							0.00

Note. All regressions include cutoff fixed effects, cutoff-specific controls for COMIPEMS score and COMIPEMS score * admitted), a dummy variable for taking the 2006 COMIPEMS exam, and a dummy variable for whether the UNAM exam was taken. Results are reported for a bandwidth of 10 points. AIC-optimal bandwidth of 1 is used for all regressions.

Unadjusted standard errors are in parentheses. Wild cluster bootstrapped p-values for the null hypothesis of the coefficient equaling zero are in brackets, where the cluster is the centered COMIPEMS score.

*** p<0.01, ** p<0.05, * p<0.1

[a] p-value is from chi-square test of joint equality to zero of "Admitted to IPN" coefficients in columns 1-7. The equations are jointly estimated with seemingly unrelated regression.

Table 2.5a: Regression discontinuity estimates of effect of IPN admission on ENLACE scores: composite score (math and Spanish)

	Bandwidth: 4				Bandwidth: 7				Bandwidth: 10			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Admitted to IPN	0.179	0.157	0.193	(One per	0.114	0.100	0.122	(One per	0.137	0.127	0.145	(One per
	(0.050)	(0.055)	(0.057)	cutoff)	(0.039)	(0.042)	(0.044)	cutoff)	(0.034)	(0.037)	(0.039)	cutoff)
	[0.21]	[0.18]	[0.14]		[0.01]***	[0.01]***	[0.01]***		[0.00]***	[0.00]***	[0.00]***	
Middle school GPA * Admitted			-0.158	-0.174			-0.082	-0.104			-0.063	-0.099
			(0.084)	(0.085)			(0.063)	(0.064)			(0.055)	(0.056)
			[0.29]	[0.28]			[0.32]	[0.19]			[0.49]	[0.25]
Parental education * Admitted			0.027	0.026			0.001	-0.001			-0.000	-0.001
			(0.019)	(0.019)			(0.015)	(0.015)			(0.013)	(0.013)
			[0.54]	[0.47]			[0.98]	[0.98]			[0.97]	[0.92]
Family income * Admitted			-0.019	-0.014			0.003	0.006			0.008	0.010
			(0.020)	(0.020)			(0.015)	(0.015)			(0.013)	(0.013)
			[0.19]	[0.28]			[0.67]	[0.50]			[0.47]	[0.40]
Male * Admitted			-0.194	-0.242			0.006	-0.015			0.089	0.069
			(0.119)	(0.125)			(0.092)	(0.096)			(0.079)	(0.083)
			[0.07]*	[0.06]*			[0.97]	[0.89]			[0.27]	[0.46]
Hours studied per week * Admitted			-0.002	-0.000			0.007	0.009			0.002	0.001
			(0.017)	(0.017)			(0.013)	(0.013)			(0.012)	(0.012)
			[0.93]	[0.92]			[0.67]	[0.69]			[0.91]	[0.93]
Parental effort index * Admitted			-0.008	-0.007			-0.004	-0.007			-0.032	-0.030
			(0.061)	(0.061)			(0.047)	(0.047)			(0.040)	(0.041)
			[0.91]	[0.98]			[0.97]	[0.94]			[0.71]	[0.70]
Employed * Admitted			-0.361	-0.263			-0.207	-0.167			-0.023	0.007
			(0.317)	(0.319)			(0.234)	(0.235)			(0.208)	(0.208)
			[0.14]	[0.41]			[0.57]	[0.63]			[0.93]	[0.99]
Student covariates	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES
AIC-optimal polynomial order	1	1	1	1	1	1	1	1	1	1	1	1
Observations	3,338	2,858	2,858	2,858	5,172	4,430	4,430	4,430	6,532	5,585	5,585	5,585
R^2	0.196	0.237	0.244	0.249	0.234	0.267	0.271	0.274	0.291	0.324	0.326	0.329
Mean of dependent variable	0.352	0.355	0.355	0.355	0.339	0.344	0.344	0.344	0.343	0.349	0.349	0.349

Note. Dependent variable is dropout*100.
All regressions include cutoff fixed effects, cutoff-specific controls for COMIPEMS score and COMIPEMS score * admitted), a dummy variable for taking the 2006 COMIPEMS exam, and a dummy variable for whether the UNAM exam was taken. Columns (2-4), (6-8), and (10-12) include covariates interacted with cutoff fixed effects. Columns (3-4), (7-8), and (11-12) include controls for (COMIPEMS score * covariate) and (COMIPEMS score * covariate * admitted) for each of the covariates. Each of the Unadjusted standard errors are in parentheses. Wild cluster bootstrapped p-values for the null hypothesis of the coefficient equaling zero are in brackets, where the cluster is the centered COMIPEMS score.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.5b: Regression discontinuity estimates of effect of IPN admission on ENLACE scores: math score

	Bandwidth: 4				Bandwidth: 7				Bandwidth: 10			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Admitted to IPN	0.247	0.262	0.294	(One per	0.181	0.188	0.208	(One per	0.204	0.198	0.212	(One per
	(0.054)	(0.058)	(0.061)	cutoff)	(0.042)	(0.045)	(0.047)	cutoff)	(0.037)	(0.039)	(0.042)	cutoff)
	[0.11]	[0.11]	[0.15]		[0.00]***	[0.01]***	[0.00]***		[0.00]***	[0.00]***	[0.00]***	
Middle school GPA * Admitted			-0.118	-0.135			-0.038	-0.053			-0.030	-0.056
			(0.089)	(0.091)			(0.067)	(0.068)			(0.059)	(0.060)
			[0.49]	[0.46]			[0.60]	[0.51]			[0.79]	[0.48]
Parental education * Admitted			0.027	0.024			-0.000	-0.002			0.008	0.008
			(0.020)	(0.020)			(0.016)	(0.016)			(0.014)	(0.014)
			[0.67]	[0.66]			[0.97]	[0.96]			[0.81]	[0.71]
Family income * Admitted			-0.016	-0.011			0.002	0.005			-0.005	-0.003
			(0.021)	(0.021)			(0.016)	(0.016)			(0.014)	(0.014)
			[0.56]	[0.68]			[0.91]	[0.71]			[0.75]	[0.82]
Male * Admitted			-0.099	-0.128			0.088	0.093			0.138	0.146
			(0.127)	(0.133)			(0.097)	(0.102)			(0.085)	(0.089)
			[0.13]	[0.22]			[0.29]	[0.31]			[0.01]***	[0.00]***
Hours studied per week * Admitted			-0.003	-0.001			-0.001	-0.000			-0.005	-0.006
			(0.018)	(0.018)			(0.014)	(0.014)			(0.012)	(0.012)
			[0.77]	[0.92]			[0.90]	[0.97]			[0.45]	[0.37]
Parental effort index * Admitted			0.019	0.022			0.018	0.016			-0.016	-0.011
			(0.065)	(0.065)			(0.050)	(0.050)			(0.043)	(0.043)
			[0.72]	[0.53]			[0.59]	[0.63]			[0.65]	[0.66]
Employed * Admitted			0.026	0.104			-0.060	-0.042			-0.003	0.018
			(0.337)	(0.340)			(0.249)	(0.250)			(0.223)	(0.223)
			[0.77]	[0.50]			[0.80]	[0.86]			[0.95]	[0.97]
Student covariates	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES
AIC-optimal polynomial order	1	1	1	1	1	1	1	1	1	1	1	1
Observations	3,338	2,858	2,858	2,858	5,172	4,430	4,430	4,430	6,532	5,585	5,585	5,585
R^2	0.172	0.235	0.240	0.242	0.207	0.257	0.260	0.263	0.262	0.310	0.312	0.315
Mean of dependent variable	0.418	0.423	0.423	0.423	0.403	0.406	0.406	0.406	0.413	0.417	0.417	0.417

Note. Dependent variable is dropout*100.

All regressions include cutoff fixed effects, cutoff-specific controls for COMIPEMS score and COMIPEMS score * admitted), a dummy variable for taking the 2006 COMIPEMS exam, and a dummy variable for whether the UNAM exam was taken. Columns (2-4), (6-8), and (10-12) include covariates interacted with cutoff fixed effects. Columns (3-4), (7-8), and (11-12) include controls for (COMIPEMS score * covariate) and (COMIPEMS score * covariate * admitted) for each of the covariates. Each of the covariates is de-meaned.
Unadjusted standard errors are in parentheses. Wild cluster bootstrapped p-values for the null hypothesis of the coefficient equaling zero are in brackets, where the cluster is the centered COMIPEMS score.
*** p<0.01, ** p<0.05, * p<0.1

Table 2.5c: Regression discontinuity estimates of effect of IPN admission on ENLACE scores: math score

	Bandwidth: 4				Bandwidth: 7				Bandwidth: 10			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Admitted to IPN	0.058	0.001	0.034	(One per	0.012	-0.022	-0.004	(One per	0.028	0.017	0.035	(One per
	(0.058)	(0.063)	(0.066)	cutoff)	(0.045)	(0.049)	(0.051)	cutoff)	(0.040)	(0.043)	(0.045)	cutoff)
	[0.39]	[0.97]	[0.73]		[0.69]	[0.27]	[0.90]		[0.33]	[0.57]	[0.32]	
Middle school GPA * Admitted			-0.164	-0.175			-0.111	-0.136			-0.085	-0.123
			(0.097)	(0.098)			(0.073)	(0.074)			(0.064)	(0.065)
			[0.14]	[0.11]			[0.07]*	[0.04]**			[0.21]	[0.04]**
Parental education * Admitted			0.020	0.023			0.002	0.001			-0.010	-0.011
			(0.022)	(0.022)			(0.017)	(0.017)			(0.015)	(0.015)
			[0.36]	[0.32]			[0.95]	[0.92]			[0.57]	[0.65]
Family income * Admitted			-0.018	-0.013			0.003	0.005			0.021	0.022
			(0.023)	(0.023)			(0.017)	(0.017)			(0.015)	(0.015)
			[0.13]	[0.18]			[0.82]	[0.73]			[0.14]	[0.17]
Male * Admitted			-0.254	-0.312			-0.087	-0.131			0.014	-0.031
			(0.138)	(0.144)			(0.106)	(0.111)			(0.092)	(0.097)
			[0.01]***	[0.01]***			[0.38]	[0.29]			[0.86]	[0.79]
Hours studied per week * Admitted			0.000	0.000			0.015	0.017			0.009	0.009
			(0.020)	(0.020)			(0.015)	(0.015)			(0.014)	(0.014)
			[0.99]	[0.97]			[0.58]	[0.49]			[0.69]	[0.71]
Parental effort index * Admitted			-0.037	-0.038			-0.028	-0.033			-0.044	-0.044
			(0.070)	(0.070)			(0.054)	(0.054)			(0.047)	(0.047)
			[0.93]	[0.86]			[0.76]	[0.81]			[0.70]	[0.67]
Employed * Admitted			-0.701	-0.604			-0.317	-0.261			-0.034	0.002
			(0.364)	(0.366)			(0.269)	(0.270)			(0.241)	(0.241)
			[0.06]*	[0.13]			[0.42]	[0.55]			[0.95]	[0.99]
Student covariates	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES
AIC-optimal polynomial order	1	1	1	1	1	1	1	1	1	1	1	1
Observations	3,341	2,861	2,861	2,861	5,176	4,434	4,434	4,434	6,537	5,590	5,590	5,590
R^2	0.122	0.185	0.194	0.201	0.138	0.194	0.197	0.201	0.164	0.215	0.218	0.221
Mean of dependent variable	0.194	0.192	0.192	0.192	0.185	0.192	0.192	0.192	0.181	0.188	0.188	0.188

Note. Dependent variable is dropout*100.

All regressions include cutoff fixed effects, cutoff-specific controls for COMIPEMS score and COMIPEMS score * admitted), a dummy variable for taking the 2006 COMIPEMS exam, and a dummy variable for whether the UNAM exam was taken. Columns (2-4), (6-8), and (10-12) include covariates interacted with cutoff fixed effects. Columns (3-4), (7-8), and (11-12) include controls for (COMIPEMS score * covariate) and (COMIPEMS score * covariate * admitted) for each of the covariates. Each of the covariates is de-meaned.

Unadjusted standard errors are in parentheses. Wild cluster bootstrapped p-values for the null hypothesis of the coefficient equaling zero are in brackets, where the cluster is the centered COMIPEMS score.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.6: Regression discontinuity estimates of effect of IPN admission on dropout, Federal District students only

	Bandwidth: 4				Bandwidth: 7				Bandwidth: 10			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Admitted to IPN	6.017	3.887	4.498	(One per	7.635	5.760	5.641	(One per	8.896	8.558	8.527	(One per
	(3.849)	(4.152)	(4.174)	cutoff)	(3.025)	(3.212)	(3.227)	cutoff)	(2.777)	(2.944)	(2.973)	cutoff)
	[0.07]*	[0.17]	[0.14]		[0.04]**	[0.02]**	[0.02]**		[0.01]***	[0.04]**	[0.03]**	
Middle school GPA * Admitted			-8.644	-8.529			-13.012	-13.496			-8.633	-8.707
			(6.193)	(6.253)			(4.791)	(4.825)			(4.371)	(4.412)
			[0.08]*	[0.08]*			[0.06]*	[0.03]**			[0.29]	[0.28]
Parental education * Admitted			-2.245	-2.362			-1.218	-1.083			-0.834	-0.831
			(1.503)	(1.518)			(1.183)	(1.199)			(1.078)	(1.091)
			[0.06]*	[0.06]*			[0.09]*	[0.22]			[0.38]	[0.41]
Family income * Admitted			-0.329	0.005			-1.927	-2.003			-2.002	-2.091
			(1.540)	(1.552)			(1.140)	(1.148)			(1.077)	(1.082)
			[0.49]	[0.99]			[0.00]***	[0.00]***			[0.04]**	[0.02]**
Male * Admitted			2.236	3.569			0.795	2.439			3.912	6.251
			(9.004)	(9.296)			(6.981)	(7.269)			(6.392)	(6.716)
			[0.69]	[0.50]			[0.92]	[0.73]			[0.52]	[0.41]
Hours studied per week * Admitted			1.122	0.886			1.670	1.611			0.265	0.266
			(1.300)	(1.306)			(1.007)	(1.010)			(0.933)	(0.936)
			[0.17]	[0.35]			[0.10]*	[0.09]*			[0.80]	[0.76]
Parental effort index * Admitted			-1.237	-1.541			-3.112	-3.572			-0.641	-0.958
			(4.528)	(4.588)			(3.525)	(3.544)			(3.247)	(3.259)
			[0.75]	[0.81]			[0.44]	[0.47]			[0.85]	[0.72]
Employed * Admitted			33.380	32.185			27.436	27.378			21.135	23.215
			(26.692)	(26.847)			(18.664)	(18.762)			(17.052)	(17.098)
			[0.21]	[0.24]			[0.07]*	[0.10]*			[0.10]*	[0.07]*
Student covariates	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES
AIC-optimal polynomial order	1	1	1	1	1	1	1	1	1	1	1	1
Observations	3,034	2,541	2,541	2,541	4,505	3,769	3,769	3,769	5,209	4,362	4,362	4,362
R^2	0.025	0.150	0.162	0.166	0.017	0.140	0.147	0.149	0.017	0.134	0.141	0.144
Mean of dependent variable	50.33	49.67	49.67	49.67	49.86	49.51	49.51	49.51	49.97	50.07	50.07	50.07

Note. Dependent variable is dropout*100.

All regressions include cutoff fixed effects, cutoff-specific controls for COMIPEMS score and COMIPEMS score * admitted), a dummy variable for taking the 2006 COMIPEMS exam, and a dummy variable for whether the UNAM exam was taken. Columns (2-4), (6-8), and (10-12) include covariates interacted with cutoff fixed effects. Columns (3-4), (7-8), and (11-12) include controls for (COMIPEMS score * covariate) and (COMIPEMS score * covariate * admitted) for each of the covariates. Each of the covariates is de-meaned.

Unadjusted standard errors are in parentheses. Wild cluster bootstrapped p-values for the null hypothesis of the coefficient equaling zero are in brackets, where the cluster is the centered COMIPEMS score.

*** p<0.01, ** p<0.05, * p<0.1

Chapter 3

Little Fish, Big Pond? Negative Effects of Marginal School Admission and Consequences for School Choice

3.1 Introduction

Is it always wise for students to attend a "better" or more selective school? Both recent evidence from economics and older results in the education literature show that the effects on academic outcomes are often very small or even negative. One obvious candidate explanation for this finding is that objective measures of school quality, such as an elite reputation or high peer quality, do not matter much for achievement. But there is another possible explanation: for admittees who are academically marginal, the (potentially large) benefits of admission to a better school are attenuated or negated by being at the bottom of the ability distribution in the population at that school. This chapter shows that, for public high school students in Mexico City, the average effect of admission to a school with higher-performing peers has at most a negligible impact on dropout and exam scores. It then shows that when admission is to a school with much higher peer quality than the student's next most-preferred school, the effect of admission on dropout probability is *more positive* than when the the admitting school is similar to the next most-preferred school. The negative relationship between peer quality and dropout probability is consistent with relative ability being important for academic success. Finally, I show that if students account for their position in the ability distribution when choosing schools, the optimal solution to the school choice problem for students in Mexico City (and many other systems throughout the world) is complicated and requires sophistication on the part of the student.

Several recent papers have used regression discontinuity designs to see how admission to an elite school, or one where peers have higher test scores, affects academic outcomes. Such designs are made possible by exam-based admissions systems where some students score barely high enough to enter into a school and others score barely too low and must attend a less-preferred school. Among the studies finding zero or very small effects on test scores are Lucas and Mbiti (2013) and Clark (2010) in elite Kenyan and United Kingdom high schools, respectively, and Abdulkadiroglu et al. (2011) in Boston and New York exam high schools. Jackson (2010) and Pop-Eleches (2013) find small test score effects in Trinidad and Tobago and Romania, respectively, while de Hoop (2011) finds that admission to elite schools in Malawi reduces dropout. At the primary level, Duflo et al. (2011) find no evidence that students at the bottom of the ability distribution of a high-ability *classroom* perform differently from similar students in the top of the ability distribution in a low-ability classroom. Chapter 2 of the present work shows that there is a trade-off in Mexico City's elite high schools, where test scores increase but the risk of dropout rises substantially.

Why does admission to a "better" school sometimes fail to yield academic benefits? Each of the research designs above necessarily focuses on students who are either marginally admitted or rejected from the better school, so the local average treatment effect encompasses two phenomena: first, attending a school with different peers and educational inputs, and second, being at the very bottom of the ability distribution instead of being potentially much higher in the distribution a less selective school. It is possible that more selective schools have large academic benefits but that marginally admitted students have these benefits offset by their location in the ability distribution. This is not a new idea. As long ago as 1890, James (1890; 1983) posited what is now called the "little-fish–big-pond" effect (Marsh and Parker 1984), where people judge their accomplishments not only on their own merits but also in comparison to their peer group. While much of the focus in the education and psychology literatures has been on how peer groups affect self-concept (e.g. Marsh and Hau 2003), others have looked at another aspect of being a "little fish": students at the bottom of the ability distribution in a school may face teaching and grading practices that are too difficult. Duflo et al. (2011) review this literature in some detail, although like the regression discontinuity evidence given above, no consistent conclusions emerge with respect to relative ability and achievement.

The present chapter uses a regression discontinuity design, similar to those in the above-mentioned studies, to show that the average effect of admission to a more selective school on dropout and achievement in Mexico City's public high schools is, at most, very small. Marginally admitted students whose next-best option had much lower-ability peers do not seem to suffer a penalty on their exam scores, while their probability of dropout rises in comparison to marginal admittees whose next-best option had median peer quality similar to the admitted school. This result holds even when allowing the average effect of admission to vary by cutoff school, so that the comparison in outcomes is between students with the same more-preferred school but with different less-preferred schools. While not indisputable evidence for the importance of relative ability for academic outcomes, the fact that the "worse" the school, the better the graduation outcome, is suggestive that this channel is important in determining the net effect of admission. In contrast to Chapter 2 of this dissertation, the present chapter estimates the average admission effect over all schools that fill to capacity instead of focusing exclusively on elite schools. This allows me to show the effect of admission to even when the counterfactual admission outcome is to a school within the same academic subsystem.

The Mexico City public high school admissions competition, like many systems in the world (e.g. Kenya, Ghana, Malawi, Trinidad and Tobago, and Romania), requires students to choose schools before they know the outcome of the entrance exam used to determine assignment priority. This arrangement exacerbates the problem faced by students who are trying to balance school quality or selectivity and the desire to avoid being at the bottom of the ability distribution. Students do not know their ability, so if the entrance exam offers useful information about ability, they would prefer to know their results before choosing. I show via a simple theoretical model that, under this prevalent allocation mechanism, the optimal strategy is not simply to rank schools based on one's prior about his own ability. The student can improve on this strategy by accounting for the fact that his unknown exam score is correlated with his ability and that this score is used for assignment. He ranks schools based on expected utility conditional on scoring high enough for admission, which differs from the unconditional expectation because exam score is correlated with ability. In practice, the student may leave desirable schools off his list in order to avoid having several schools with bunched up cutoff scores. This strategy is not obvious and arguably requires more sophistication on the part of the student than choosing on the basis of unconditional expectations.

The rest of this chapter proceeds as follows. Section 3.2 summarizes Mexico City's public high school assignment system. Section 3.3 explains the regression discontinuity design that will be used to estimate the heterogeneous effects of admission. Section 3.4 describes the data used and Section 3.5 gives the results from the empirical exercise. Section 3.6 shows how, when relative ability matters for utility and exam scores are correlated with ability, students can improve their expected outcome by deviating from a simple ranking of schools by unconditional expected utility. Section 3.7 concludes.

3.2 Context

The mechanics of Mexico City's public high school choice system, run by the Comisión Metropolitana de Instititiciones de Educación Media Superior (COMIPEMS), have been explained at length in the previous two chapters. The following list summarizes the application and assignment process:

1. January of 9$^{\text{th}}$ grade: students receive a catalog of schools (or, for technical schools, school-track combination) from which to choose for high school.

2. February-March: students turn in a list of up to twenty ranked options, along with a demographic survey.

3. June: students take a multiple-choice, standardized exam consisting of 128 questions.

4. July: a computer scores the exams (one point per correct response) and assigns students to schools in the following way:

a) Each of the ten school subsystems with campuses in Mexico City reports the number of seats available for each school (or school-track);

b) The computer orders students from highest-scoring to lowest-scoring;

c) The computer descends through the ordered list, assigning each student to his most-preferred option that still has a seat remaining. The one exception is that for admission to schools in the two elite subsystems, students must have a 7/10 middle school grade point average (GPA) or higher. The GPA is known at the time of assignment so no admission results have to be rescinded later.

Ties are not broken–the school subsystem must decide in each case whether to admit all tied students or none of them. Unassigned students must wait until after the process and choose a school from those that still have seats remaining.

3.3 Empirical method

This section explains the regression discontinuity (RD) design that will be used to estimate the causal impact of admission to a student's more-preferred school on academic outcomes, in comparison to the student's next-most-preferred school that would admit him. Repeating the intuition for this design from the previous two chapters: given that a school filled all of its seats during the assignment process, there is a student (or group of students) who won the final available seat there. This student's score is the school's cutoff score–nobody with a lower score was admitted, and nobody with at least that score was rejected if he "wanted" to attend that school when his turn for assignment arrived. COMIPEMS exam score, then, perfectly predicts admission to a school among the group of students who wanted to attend at the time of assignment. This gives a sharp RD design where the running variable is COMIPEMS exam score and admission to the "cutoff school" is the treatment.

To formalize this method, consider the following parametric estimating equation:

$$Y_{ijt} = \delta \text{admit}_i + g_1\left(c_i - C_{jt}\right) + \text{admit}_i g_2\left(c_i - C_{jt}\right) + \mu_j + \eta_t + \varepsilon_{ijt} \tag{3.1}$$

where Y_{ijt} is the academic outcome (dropout or ENLACE score) of student i near the admission cutoff of school j who took the exam in year t, $admit_i$ is a dummy variable for whether the student scored high enough to be admitted (and thus was admitted, according to the assignment rule), g_1 and g_2 are polynomial functions in "centered" COMIPEMS score (c_i denotes the student's score and C_{jt} gives school j's cutoff score in year t), μ_j and η_t are cutoff school and exam year fixed effects, respectively, and ε_{ijt} is a mean-zero error term. The coefficient of interest is δ, which, provided that the polynomial functions adequately fit the relationship between centered COMIPEMS score and outcome on each side of the cutoff, gives the local average treatment effect of admission to the cutoff school. This is an average over all schools that filled their seats, weighted by the number of marginally admitted and rejected students at each. The counterfactual to cutoff school admission is admission to the next-most-preferred school, which varies both across cutoff schools j and between students i within the same cutoff school.

For each school j, the sample of marginal students to be used in the RD analysis is defined in a way similar to that of Chapter 1, but with one important difference. For convenience, I restate the first four criteria from Chapter 1 here with slightly different notation and a modified version of criterion 4. If school k is ranked before j on student i's preference list, including if j is unlisted, we write $k \succ j$. Denote the student's exam score as c_i. Then marginal students for school j are those who:

1. listed school j as a choice;

2. had a score sufficiently close to j's cutoff score to be in a small window around the cutoff, where the window size is determined by a preselected bandwidth w: $-w \leq c_i - C_{jt} < w$;[1]

3. scored too low to be admitted to any more-preferred school: $c_i < C_{kt}, \forall k \succ j$;

4. would be admitted only to the cutoff school for all scores within the $[C_{jt}, C_{jt} + w)$ half-window: $C_{kt} \notin [C_{jt}, C_{jt} + w), \forall k \succ j$;

5. would be admitted to a single less-preferred school for all scores within the $[C_{jt} - w, C_{jt})$ half-window, i.e. the cutoff score for the next-most-preferred school is not within the half-window: defining ℓ as the school attended by the student if he scores $C_{jt} - 1$, then $C_{\ell t} < C_{jt} - w$.

Condition 5 ensures two things. First, the student is actually admitted to a school (ℓ) if he scores too low for j, rather than being left unassigned. Second, within the selected window, the only two outcomes possible for the student are admission to j and admission to a single school ℓ. This will be important when examining the heterogeneous effects of admission with respect to school characteristics, which will now be discussed.

Beyond the average effect of being admitted to a more-preferred school, one can see how this effect varies with respect to differences between the schools above and below the cutoff. For example, it may be that admission to the cutoff school makes a small difference when the school below the cutoff has students with similar COMIPEMS scores, while the effect may be larger when the two schools are very different on this dimension. The following equation is used to estimate such heterogeneous effects:

$$\begin{aligned} Y_{ijt} = & \delta \text{admit}_i + g_1\left(c_i - C_{jt}\right) + \text{admit}_i g_2\left(c_i - C_{jt}\right) + \\ & \Sigma_k \left[\beta_k \Delta X_{kijt} + h_{k1}\left(\Delta X_{kijt} \times \left(c_i - C_{jt}\right)\right) + \text{admit}_i h_{k2}\left(\Delta X_{kijt} \times \left(c_i - C_{jt}\right)\right) + \right. \\ & \left. \gamma_k \left(\text{admit}_i \times \Delta X_{kijt}\right)\right] + \mu_j + \eta_t + \varepsilon_{ijt} \end{aligned} \tag{3.2}$$

[1] The second inequality is strict because the score variable is discrete, so this definition includes w score values too low to be admitted and w values high enough to be admitted.

where k indexes the covariates that may alter the admission effect and ΔX_{kijt} is the difference in covariate X_k that the student will experience if he is rejected from the cutoff school. In practice, these covariates are median COMIPEMS score of all admitted students, distance from home to the school, and (logarithm of) cohort size. The γ_k coefficients tell us, *conditional on the ΔX_k generated by the student's preferences and exam score*, how the admission effect differs with respect to the change in covariates experienced due to admission. A further specification allows for one admission effect per cutoff school, so that δ is instead δ_j. In this case, the heterogeneity is estimated over students with the same cutoff school but (potentially) different schools below the cutoff.

The heterogeneous effects γ_k are identified under the standard assumptions of the RD design and similar assumptions about the relationship between ΔX_k and Y being well-approximated by polynomials h. But it is still possible that the heterogeneous effects of admission with respect to a particular ΔX_k do not represent a causal relationship between ΔX_k and the admission effect. Perhaps students who place themselves at cutoffs with a large ΔX_k are simply the type of student who experiences a larger (or smaller) effect of rejection from their preferred school, independent of how much they are actually affected by X_k. We might reasonably assume that students who place themselves at cutoffs where ΔX_k is large care about X_k less than those with small ΔX_k, because the latter group has chosen schools in a way that minimizes the change in this characteristic.[2] If true, then γ_k understates the true average contribution of ΔX_k to the admission effect. Still, the estimated heterogeneous effects should not be regarded as strictly causal relationships between changes in school characteristics and the effect of admission.

The discrete running variable, COMIPEMS exam score, necessitates the use of a parametric RD design (Lee and Card 2008). I select the optimal polynomial order on the basis of the Akaike Information Criterion (AIC), as Lee and Lemieux (2010) suggest. Bandwidth selection in this case is complicated because of the criteria used for inclusion in the sample. In particular, criteria 4 and 5 require that only one school be attended for all scores above the cutoff and within the bandwidth, and only one school be attended for all scores below the cutoff and within the bandwidth. This leads to some observations being dropped from the sample when the bandwidth is increased, which negates some of the benefit of increasing sample size and implies that bandwidth selection is no longer the bias-efficiency trade-off seen in most designs. Further, we have reason to favor a small bandwidth in this application because large bandwidths by definition include mostly students who have not chosen schools with bunched cutoff scores. In order to keep a small bandwidth while ensuring a reasonable number of points of support, I include students within 5 points of the cutoff and then show that key results are robust to larger and smaller bandwidths.

[2]Other student characteristics, such as geographic location, could also affect the chosen ΔX_k.

3.4 Data

The data used in this paper come from two sources. The first is student-level data covering all participants in the 2005 through 2007 COMIPEMS competitions. These data include, for each student, the full ranked list of school preferences, the entrance exam score and resulting school assignment, middle school grade point average, and personal information such as address and telephone number. Most students also turned in a demographic survey as well. The second is student-level results from years 2008 through 2010 of the 12^{th} grade ENLACE exam, which is a national exam administered in the spring to students who are expected to graduate at the end of the academic year (June). This exam does not have a bearing on whether the student graduates or not and does not impact funding to schools or allocation of any other resource. The ENLACE data is primarily used in this paper as a proxy for whether the student dropped out of high school or not, but the scores are also used as an outcome variable.

The sample selection rules in the previous section determine which students are included in the RD sample, but the data impose two further constraints. First, the high schools affiliated with the Universidad Nacional Autónoma de México (UNAM) do not administer the ENLACE to its students. Therefore, UNAM schools are omitted as cutoff schools, as well as any student who would attend an UNAM school upon rejection from his cutoff school. Second, some of the Colegio be Bachilleres (Colbach) offered delayed-start programs during the time period of analysis. Such options are omitted as cutoff schools, as well as any student who would be admitted to such an option upon rejection from a cutoff school. Finally, in order to focus on students making the transition to public high school, we limit the sample to students in Mexico City who were currently enrolled at public middle schools when they participated in the COMIPEMS competition. Private school students may be planning to attend private high schools but participate anyway to see if they can be admitted to an elite high school as an additional option.

Table 3.2 describes the full sample of public middle school students who were assigned to a school during the automated COMIPEMS process, as well as the RD sample of students (within five points of an admission cutoff). Students are, in general, quite comparable between the full and RD samples. About half of students are male, the average education level of students' most-educated parent is between middle and high school, and self-reported family income is approximately 4,000 pesos ($367) per month. Students rank about 10 schools on average. While the average COMIPEMS exam score is 66.6/128, among the RD sample it is 59.9. This is due in part to the UNAM students being dropped from the sample, since they have scores that are far above the mean, and due to the fact that very high scorers are unlikely to be near the cutoff score of any school.

Students in the full sample are admitted to schools with a median COMIPEMS score of 66.2, which is similar to the mean student-level score in the population. The standard deviation of 14.6 indicates that average peer ability varies substantially across schools. Students are about 7 km (straight line distance) from their admitted school and are admitted in very large cohorts–the average incoming class is over 1,000 in the full sample and 624.9 in the RD sample. The difference is mostly due to the exclusion of the UNAM schools, which are the largest in the city.

About half of students drop out of school in this sample, as proxied by taking the ENLACE exam. Another three percent take the exam one or two years late. Because we include students who took the COMIPEMS exam in 2007 and only have ENLACE data through 2010, the existence of delayed graduates implies that the dropout proportion is an overestimate (as we miss 2007 entrants who take the exam a year late). But we can also see that this overestimate is very small, as few students graduate late. The normalized ENLACE score, generated by subtracting off the mean score for all students in the area served by COMIPEMS and dividing by the standard deviation of this population, is 0.15 standard deviations lower in the RD sample than in the full sample. This is because the highest-scoring students are less likely to be near an admissions cutoff than other students.

The key heterogeneity explored in this paper is the differential effect of admission with respect to the change in median COMIPEMS score due to admission. Estimating this requires heterogeneity in the difference between the median COMIPEMS scores for the school above the cutoff and below the cutoff. Figure 1 shows that such variation does exist in the RD sample, both for median COMIPEMS score and for other characteristics. As expected for a score-based assignment mechanism, admission to the school above the cutoff almost always results in higher-achieving peers. The difference is often quite large, as shown in Panel a. Admission above the cutoff results in a geographically closer school about half the time (Panel b), and students tend to end up in larger cohorts above the cutoff (Panel c).

3.5 Results

This section presents the key results regarding the effect of admission to the cutoff school on dropout: the estimated effect is close to zero and *increases* in the difference between the median COMIPEMS score of the schools above and below the cutoff. I first present the unconditional effects, followed by the heterogeneity results. I then show, as a supplemental finding, that there are no discernible effects of admission on ENLACE exam performance. The lack of a test score effect, combined with higher risk of dropout when admission results in going to a school with much higher-achieving peers than rejection, suggests that any academic benefits from going to a "better" and more-preferred school are on average negated by the effect of being located at the bottom of the ability distribution within the school.

3.5.1 Dropout

The average effect of marginal admission to the student's cutoff school is very small and statistically indistinguishable from zero. Figure 3.2 shows graphically the relationship between admission to the school above the cutoff and probability of dropping out, controlling for COMIPEMS exam score (normalized to zero at the cutoff score). If anything, the graph suggests a small increase in dropout due to admission. Table 3.2 confirms this finding by presenting estimates of RD equation 3.1. Column 1 includes all cutoff schools and estimates an increase of 0.7 percentage points in dropout probability due to admission, although this is statistically insignificant. The lower bound of the 95% confidence interval is -0.6 percentage points, ruling out even modest beneficial effects of admission on dropout. Chapter 2 showed that admission to an IPN school, compared to a non-IPN school, increased dropout probability substantially. Removing these students from the sample, we find an insignificant point estimate of -0.2 percentage points, and again rule out even modest effects. Columns 3 and 4 repeat this exercise, replacing the dependent variable with a measure of whether the student either delayed in graduation or dropped out. The results are very similar: admission to the more-preferred school does not, on average importantly affect progress through high school. In contrast, middle school grade point average is a strong predictor of dropout.

Behind this near-zero effect, however, there is heterogeneity with respect to the difference in attributes between the schools above and below the cutoff. Table 3.3 presents these results. Column 1 presents estimates from equation 3.2 with a single admission coefficient across all cutoff schools and allows it to vary based on the change in median COMIPEMS score, distance from home, and log cohort size due to admission. We see that admission to the school above the cutoff is worse for dropout the higher-achieving the peers and the farther away the school is in comparison to the school below the cutoff. These differential effects persist when allowing one admission coefficient per cutoff school, so that the heterogeneous effects are estimated from students with the same cutoff school but different schools below the cutoff (and in the case of distance, students who live different distances from the cutoff school). The estimated effects decline but remain statistically significant. The coefficient of 0.33 on $Admit \times (\Delta\ median\ COMIPEMS)$ suggests that for every standard deviation increase in the above-below median COMIPEMS difference (5.8 points), admission increases the probability of dropout by 1.9 additional percentage points.[3] Column 3 adds one dummy variable per subsystem (10 in total), equal to 0 if the schools above and below the cutoff are in the same subsystem, 1 if the above school belongs to that subsystem but the below school does not, and -1 if the opposite is true; interactions between each of these dummy variables and the admission dummy; and interactions between each such dummy variable and a polynomial function of centered COMIPEMS exam score. Even after accounting for changes in subsystem due to admission, the heterogeneity with respect to median COMIPEMS score persists, albeit with a lower coefficient of 0.24.

3.5.2 ENLACE scores

Students experiencing larger gains in peer COMIPEMS exam score fare worse in terms of dropout, but do their ENLACE exam scores increase? Figure 3.3 suggests that there is no ENLACE score gain, on average, from admission.[4] Table 3.4 confirms this result in column 1, ruling out all but the smallest impacts (the upper bound of the 95% confidence interval

[3]The standard deviation of 5.8 points is within-cutoff school, while the unconditional standard deviation is 6.5 points.

[4]This sets aside the issues of attrition, which are addressed at length in Chapter 2.

is 0.03). Columns 2 through 4 replicate the specifications in the previous table, and find no discernible heterogeneous effect of admission with respect to median COMIPEMS exam score when school-specific admission effects are included. The strong dropout-ENLACE trade-off found for admission to IPN schools in Chapter 2 does not appear in the full sample.

3.5.3 Validity checks

Standard checks of the RD design suggest that the design is valid. Figure 3.4 shows the density of students' centered exam scores for the sample of students in the dropout analysis. While McCrary's (2008) test for jumps in the density of the running variable at the cutoff is inappropriate here due to the discreteness of exam score, a visual check of the histogram does not suggest a problem. Table 5 presents tests for balance of the baseline covariates. In no case is the admission coefficient large or statistically significant, and a joint test of significance using seemingly unrelated regression yields a p-value of 0.54. Thus it seems that admitted and rejected students do not differ, after controlling for centered exam score.

Figure 3.5 plots out the estimated coefficient and 95% confidence interval for the main coefficient of interest, the interaction between admission and the difference in median COMIPEMS exam score between the schools above and below the cutoff. The coefficient corresponds to Column 3 in Table 3.3. While the coefficient is statistically insignificant for low bandwidths, it is fairly stable across bandwidths and is always positive. The evidence appears to be robust to bandwidth choice, which is fortunate given the already-discussed inability to choose an optimal bandwidth based on standard tools.

3.6 Implications for optimal school choice

The empirical exercise in this chapter illustrates a possible drawback of marginal admission for a student: being the lowest-ability student in a school may overwhelm the benefit of going to a "better" or more-preferred school. Such a possibility calls into question whether ranking schools in order of ex ante expected utility is the optimal strategy in a student-proposing deferred acceptance (SPDA) mechanism where exam score determines priority (e.g. the COMIPEMS assignment mechanism). While Gale and Shapley (1962) show that this is the dominant strategy under traditional SPDA mechanisms, the difference in COMIPEMS and similar systems is that the exam score actually conveys useful information to the student about expected utilities. The student would prefer to know his exam score before applying so that he might avoid being the lowest-scoring student in a school, but most assignment systems require that applications precede exam administration.

This section presents a highly simplified school choice problem and shows how a student can improve his choice portfolio by deviating from a simple ordering of schools by expected utility. Instead of using his ex ante expected ability to guess where he will fall in the ability distribution within a school, he can use his exam score as a proxy for ability and utilize that information to choose the ex ante optimal school that balances his preferences for school characteristics and his desire to avoid being at the bottom of the ability distribution at his assigned school.[5]

3.6.1 Set-up

Student i's problem is to choose an ordered, possibly incomplete, list of preferences over schools $S = \{1, 2, ..., N\}$. Assignment will be determined by whether the student's score on an exam, c_i, exceeds the cutoff (minimum) score m_j for each school, which the student takes as given.[6] This is a simplification from the true matching problem of assigning students to schools with seat quotas, using the exam score to determine priority. Let $M_i = \{m_{i1}, m_{i2}, ..., m_{iN_i}\}$ index the cutoff scores of the student's ranked schools. The student will be assigned to his most-preferred school where his exam score meets the cutoff: $min\,\{k|m_{ik} \leq c_i\}$.

Utility from a school is determined by school characteristics and their interaction with student characteristics, where some of the latter may only be known with uncertainty. To be concrete, suppose that a student does not know his own academic ability with certainty, and the utility derived from attending each school depends on own academic ability. Denoting school characteristics by X_j, known student characteristics by Z_i, and student ability by a_i, utility from attending school j is defined as:

$$U_{ij} = U\left(X_j, Z_i, a_i\right). \tag{3.3}$$

While ability is unknown, the student knows the distribution from which it is drawn, so $a_i \sim f(\theta_i)$ with known θ_i.

The student's optimal strategy appears obvious: order the schools by expected utility, where expectations are obtained by integrating utility over the ability distribution:

$$E\left[U_{ij}\right] = \int_{-\infty}^{\infty} U\left(X_j, Z_i, v\right) f(v;\theta_i) dv. \tag{3.4}$$

This strategy is the optimal behavior in SPDA mechanisms as shown by Gale and Shapley (1962).

3.6.2 Correlation between ability and exam score

It is likely that the student's exam score is correlated with his ability, so that if the student knew his exam score in advance he might change his ordering of schools. The following shows that, even without having the exam score at the time of ranking schools, the fact that assignment is on the basis of exam score will affect the optimal selection strategy.

[5]The existence of an expected utility-increasing deviation from the usual SPDA strategy suggests that further work is required to determine if and under what conditions COMIPEMS and similar systems are stable and Pareto efficient when ability is unknown and enters into expected utilities.

[6]This is a good approximation to the COMIPEMS system, where cutoff scores are remarkably stable from year to year.

As a simple case, suppose the exam score is a one-to-one function of ability: $c_i = g(a_i)$. What is the optimal decision rule? I propose that the optimal first choice school is not chosen on the basis of the expected utilities in equation 3.4. Instead, the student can do better by realizing that he will only attend a school if his score meets or exceeds the cutoff: $c_i \geq m_j$. Then he should only care about expected utility from a school conditional on having scored high enough for admission:

$$E\left[U_{ij}|m_j \leq c_i\right] = \frac{1}{1 - F\left(g^{-1}\left(m_j\right)\right)} \int_{g^{-1}(m_j)}^{\infty} U\left(X_j, Z_i, v\right) f(v;\theta_i) dv. \tag{3.5}$$

For his second choice, the student can incorporate yet more information: conditional on rejection from his first choice, he knows that $c_i < m_{i1}$. Thus he considers the expected utility from each remaining option with a lower cutoff score,[7] conditional on having scored too low for the first option but high enough to be admitted to the school under consideration:

$$\begin{aligned} &E\left[U_{ij}|m_{1i} \leq c_i \leq m_j\right] = \\ &\frac{1}{F\left(g^{-1}\left(m_{i1}\right)\right) - F\left(g^{-1}\left(m_j\right)\right)} \int_{g^{-1}(m_j)}^{g^{-1}(m_{i1})} U\left(X_j, Z_i, v\right) f(v;\theta_i) dv. \end{aligned} \tag{3.6}$$

For the third and later choices, the same comparison is made, except that for the n^{th} choice, $m_{i(n-1)}$ is used instead of m_{i1} to compute the expected utilities in equation 3.6.

3.6.3 Consequences for choice behavior when relative ability affects utility

What are the implications of the above choice rules when relative ability affects utility? To model this simply, suppose that $U_{ij} = U(X_j, Z_i) + H_j(a_i)$, where $H_j(\cdot)$ is the cumulative density function of ability in school j. Students at the bottom of the ability distribution obtain less utility from a school than students at the top, all else equal. There are two ways in which, qualitatively, the improved choice rules dictate different rankings than the simple ordering with respect to ex ante expected utility.

1. Students will be more "ambitious," at least in their first choice, by being more likely to choose a school where the unconditional expectation is to be at or near the bottom of the ability distribution. This is because his expected position in the ability distribution *conditional on admission* is higher than the unconditional expectation.

2. Students will be less likely to choose schools with cutoff scores that are very close to each other. In the extreme case (and with continuous exam scores), suppose the

[7] Even a trivial utility cost (e.g. time cost) of listing a choice will prevent students from listing a school with a cutoff score greater than those of more-preferred listed options, since it will be impossible to be admitted to the less-preferred school.

student consecutively ranks two schools j and k, where $m_k = m_j - \varepsilon,\ \varepsilon \to 0$. Then $E[H_j(a_i)|m_j > c_i \geq m_k] \to 0$ because the student knows that rejection from j and admission to k means that he barely made it into k and must be at the bottom of the ability distribution. The student thus spaces out cutoffs in order to reduce the probability of being a "little fish in a big pond," *even when the unconditional probability of this occurring is low.*

Both of these behaviors appear suboptimal from the standpoint of an unconditional expected utility maximizer, because they cause students to choose schools that are ex ante too "good" for them or to forego some schools that are excellent matches according to ex ante unconditional expectations but have cutoffs that are too similar to other options.

3.6.4 Implications for revealed preference interpretation of student choices

The dependence of utilities of on (unobserved) ability has implications for our interpretation of students' preference rankings as revealed preferences. Under traditional SPDA mechanisms, even those where the number of choices a student can list are limited, it is never optimal to list a school above another school with higher expected utility. In the COMIPEMS system, this is untrue because decisions are made on the basis of expected utility *conditional on admission.* The result is that we may see low-scoring students choosing somewhat competitive schools and wrongly interpret this as either naive decision-making or an indication that students do not care if they are the lowest-ability student in a school, in which case we may wonder if the "small fish, big pond" problem is acknowledged by students at all

Supposing that students could exhaustively rank all options, the fact that students may condition their n^{th} choice on the $(n-1)^{th}$ previous choices implies that rankings do not give a true ordering of schools by unconditional expected utilities. Students may optimally omit schools from their list in order to avoid "bunching" cutoff scores. It is incorrect to interpret omission from the preference list as an indication that the school was less-preferred than all other schools with lower cutoff scores.

Giving students their exam scores before listing their choices would eliminate each of these problems for the researcher, as well as giving the student more information with which to form expected utilities. The extent to which this issue matters will depend on how much information the exam score actually conveys about ability, as well as how much students actually wish to avoid being at the bottom of the school's ability distribution. If students already know their own ability quite well or if the exam is weakly correlated with ability, then reporting exam scores before listing choices will have little use. Likewise, if position in the ability distribution within a school is unimportant to students, then so too might be revealing exam scores.

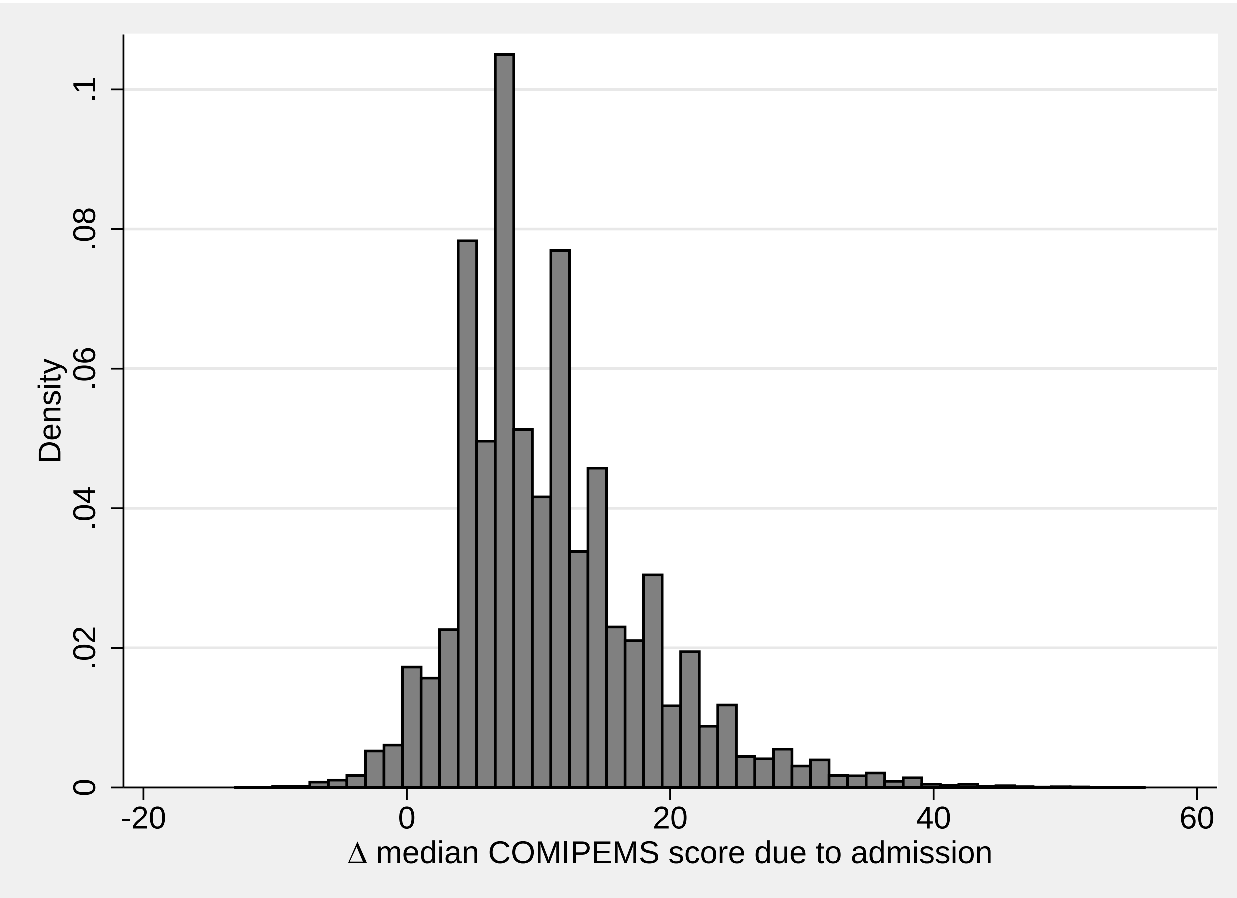
3.7 Discussion

We have seen in this chapter that marginal admission to a school has little or no average effect on dropout probability or test scores, but that this impact is heterogeneous with respect to the difference in median COMIPEMS scores between the admitted and rejected schools. The results suggest that admission to a school where one is at the bottom of the ability distribution can have a negative impact on probability of graduation. This finding is in line with the "big-fish-little-pond" theory in the education literature, although the results here cannot discern between a psychological effect of low relative ability causing dropout and the more mundane possibility that the lowest-ability students in a school are simply less likely to be capable of completing graduation requirements.

Students may take steps to avoid being a "little fish" by spacing out the cutoff scores of the schools that they select, leading to choice behavior different from what would be predicted in a setting where relative ability was unimportant. Revealing the entrance exam score before requiring preferences to be turned in would provide useful information for students. But an obvious fact bears mentioning: someone needs to be the lowest-ability student in each school. Thus it is not clear that the aggregate welfare effects of exam score provision are positive, even if each individual would prefer to have that information before choosing. Further work will attempt to understand the equilibrium that results from the current system and how it compares to the situation where ability is known (with more precision, at least) before school selection.

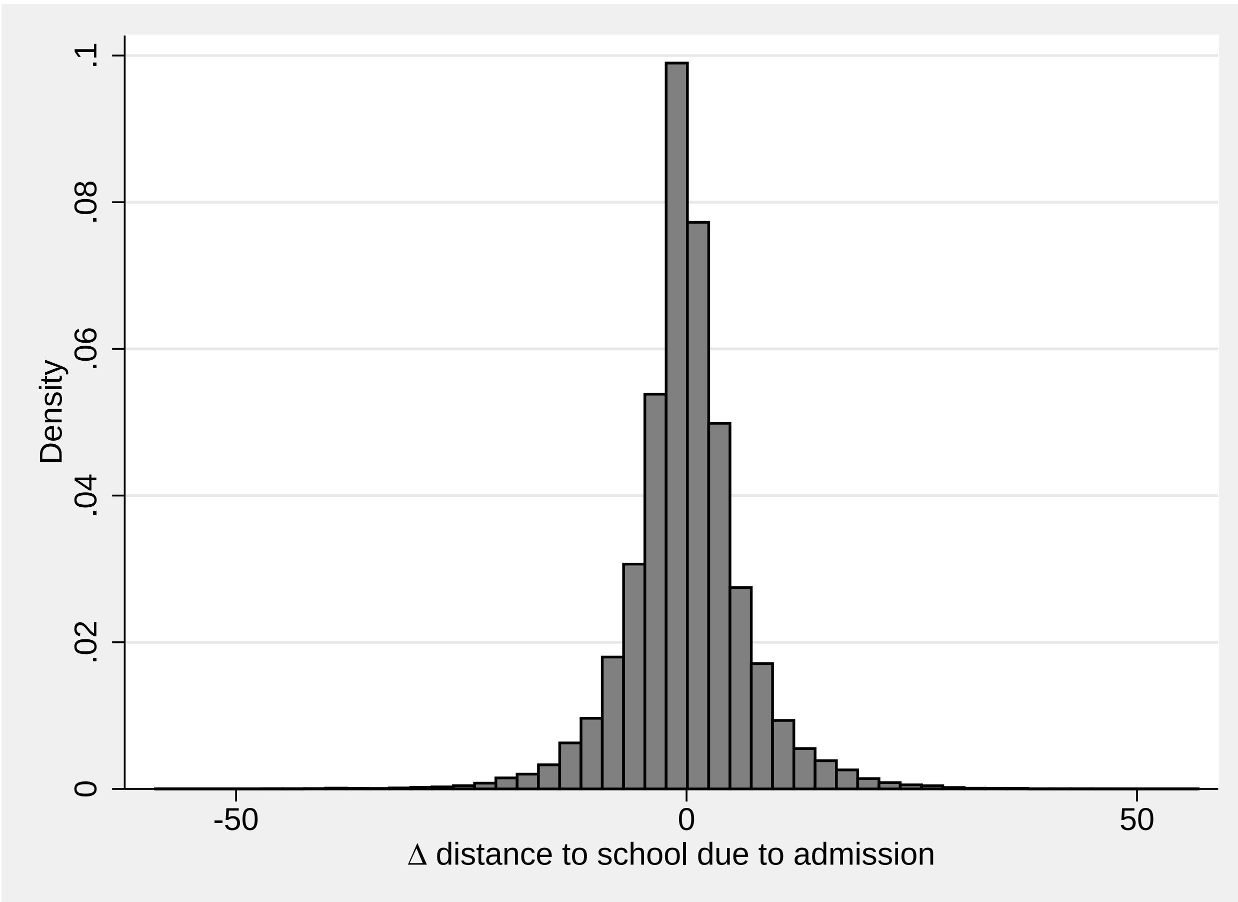
3.8 Figures

Figure 3.1a: Distribution of change in median COMIPEMS score due to admission for regression discontinuity sample



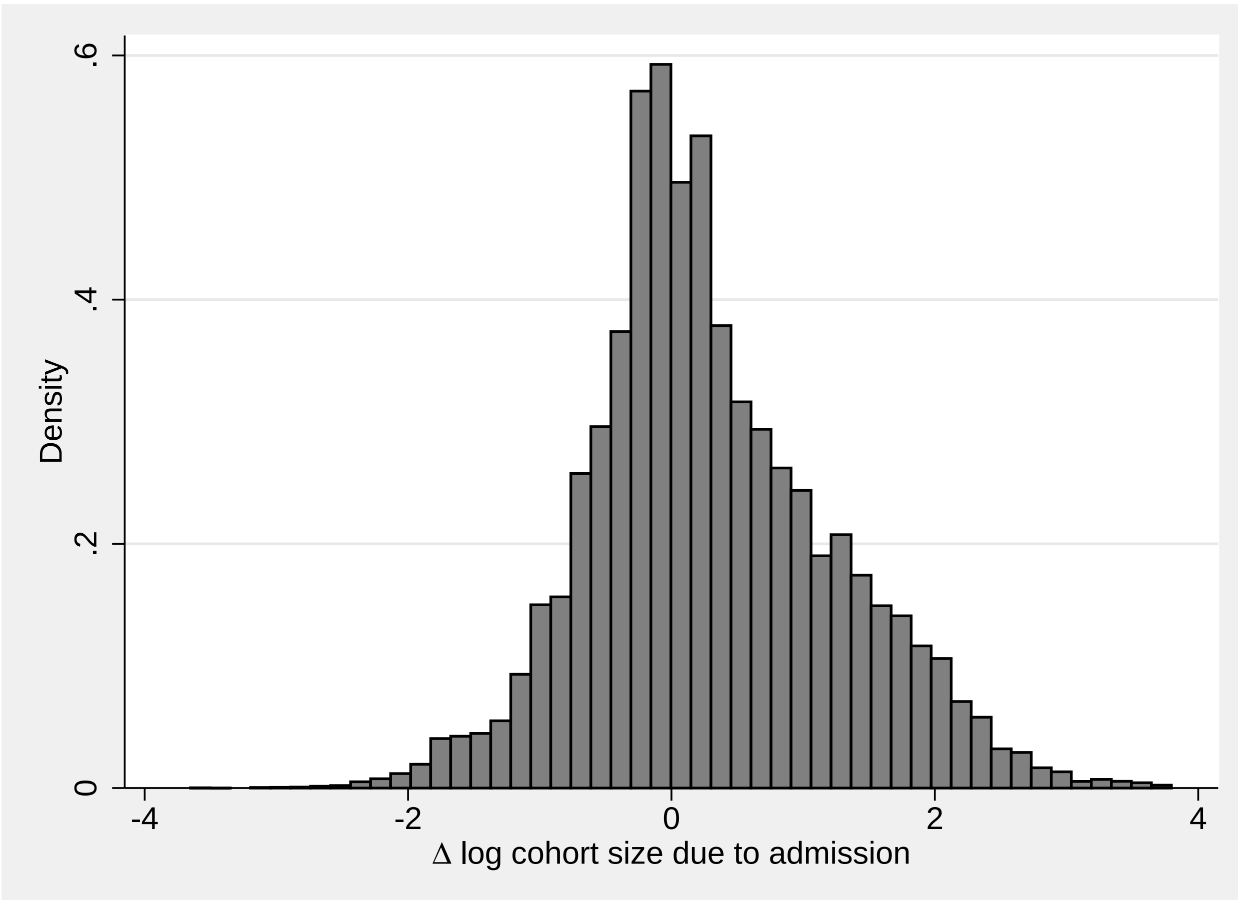
Plotted variable is the difference between the median COMIPEMS exam score in the school attended if the student scores at or above the cutoff score, minus the median score in the school attended if the student scores below the cutoff score.

Figure 3.1b: Distribution of change in home-to-school distance due to admission for regression discontinuity sample



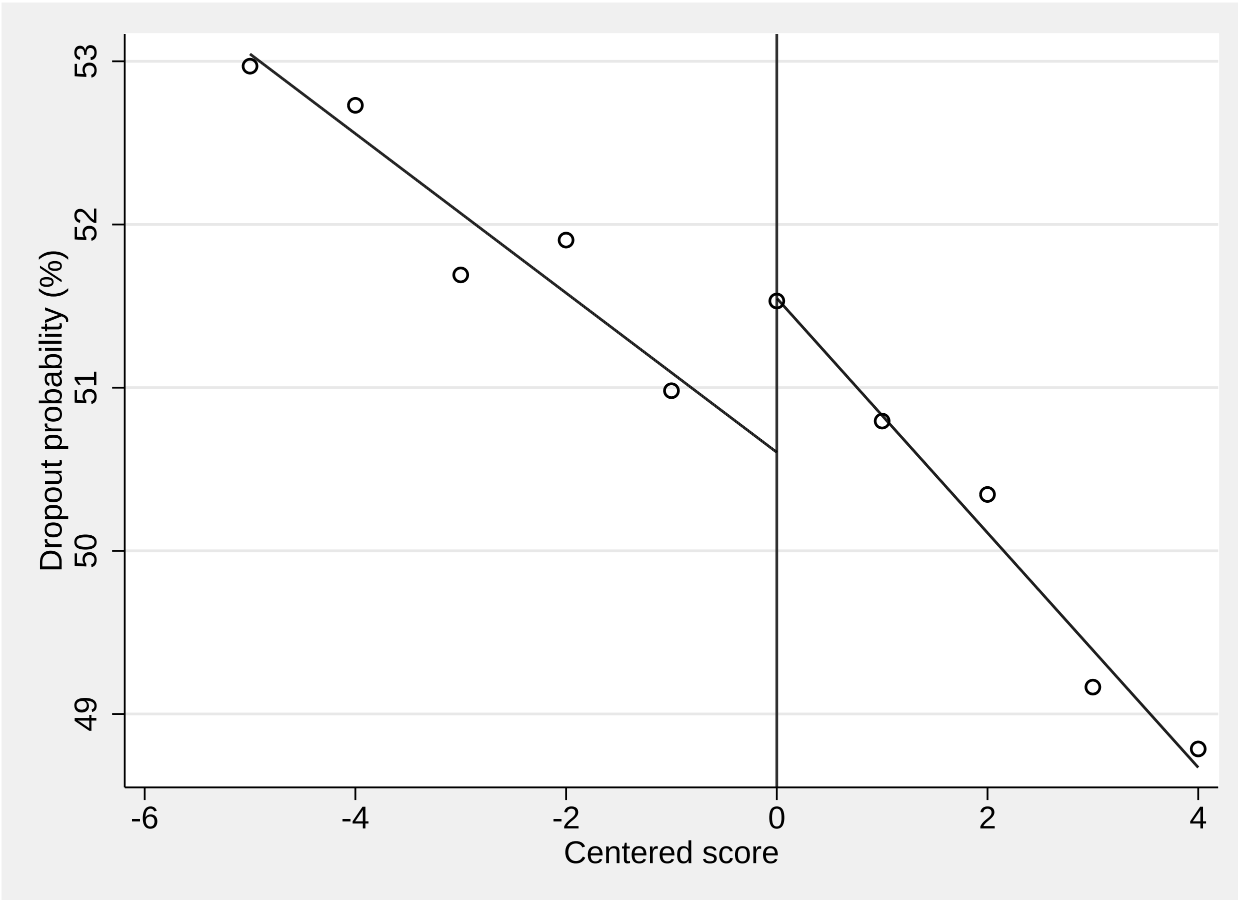
Plotted variable is the difference between the home-to-school distance for the school attended if the student scores at or above the cutoff score, minus the home-to-school distance for the school attended if the student scores below the cutoff score.

Figure 3.1c: Distribution of change in log cohort size due to admission for regression discontinuity sample



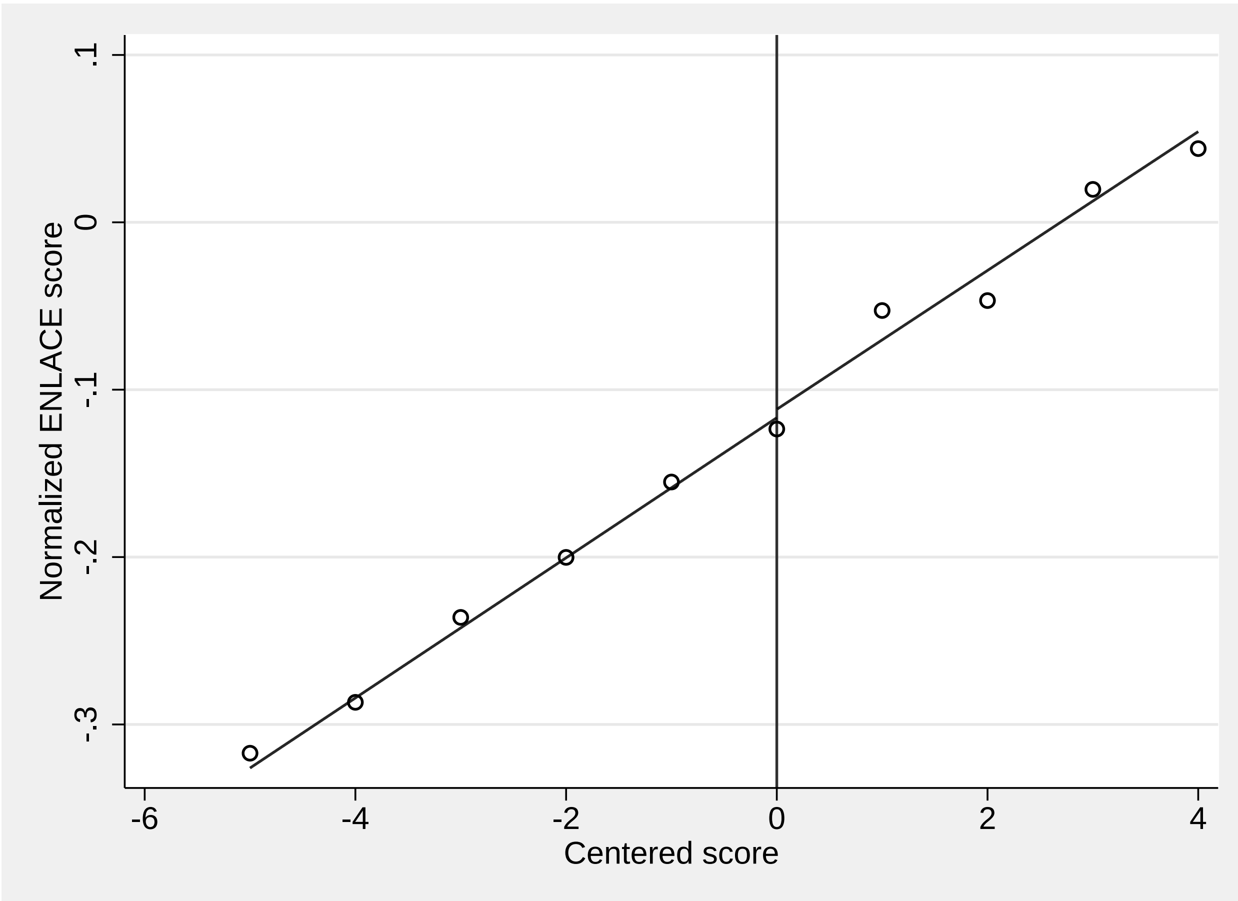
Plotted variable is the difference between the log cohort size in the school attended if the student scores at or above the cutoff score, minus the cohort size in the school attended if the student scores below the cutoff score.

Figure 3.2: Effect of admission to cutoff school on dropout probability



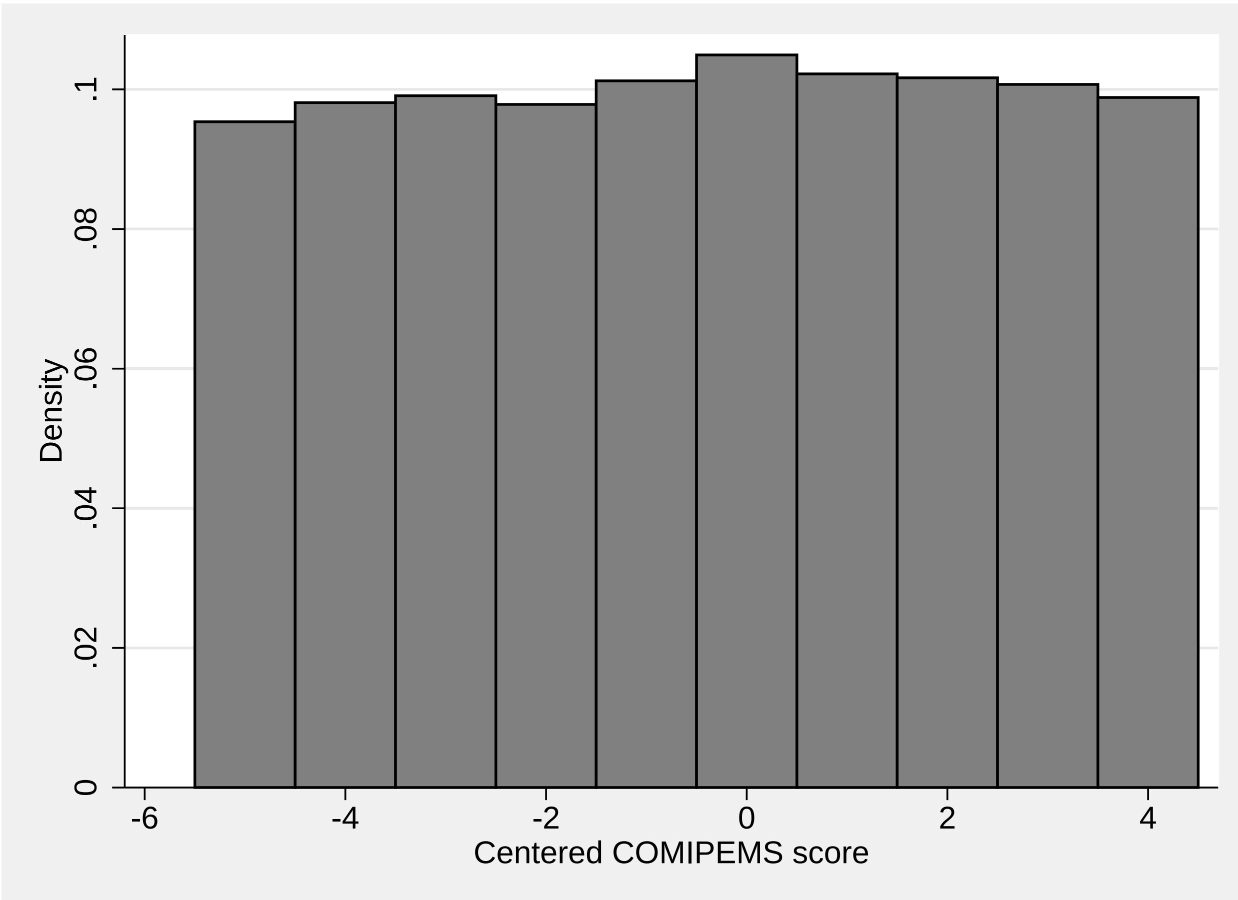
Dependent variable is dropout probability, after de-meaning dropout with respect to school fixed effects and adding the sample mean. Independent variable is the difference between the student's score and the cutoff score of the corresponding school.

Figure 3.3: Effect of admission to cutoff school on ENLACE score



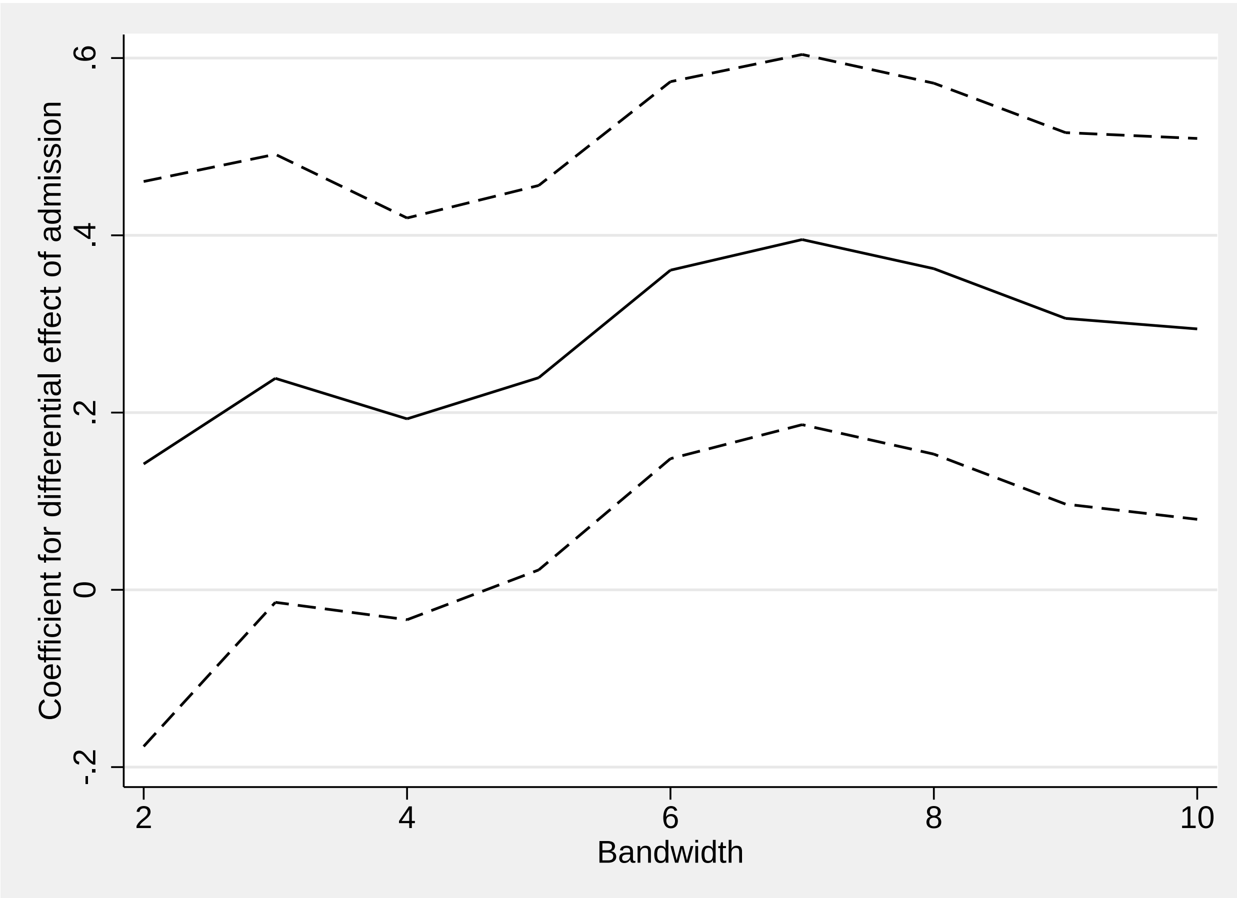
Dependent variable is normalized ENLACE score, after de-meaning dropout with respect to school fixed effects and adding the sample mean. Independent variable is the difference between the student's score and the cutoff score of the corresponding school.

Figure 3.4: Density of centered cutoff score for regression discontinuity sample



Plotted variable is the difference between the student's COMIPEMS exam score and the cutoff score of the corresponding school in the regression discontinuity sample.

Figure 3.5: Estimated differential effect of admission on dropout probability with respect to difference in median COMIPEMS score, for various bandwidths



Solid line is the regression coefficient on Admit $\times$ (Δ median COMIPEMS) for the corresponding bandwidth, estimated from equation 3.2. Dashed lines give the 95% confidence intervals for these estimates.

3.9 Tables

Table 3.1: Summary statistics for students assigned to schools by COMIPEMS

	(1) Full sample	(2) RD sample	p-value for equality of (1) and (2)
Middle school GPA	8.16	8.03	0.00
	(0.82)	(0.78)	
Male	0.46	0.45	0.00
	(0.50)	(0.50)	
Maximum of mother's and father's education	10.37	9.88	0.00
	(3.34)	(3.21)	
Family income (thousand pesos/month)	4.45	3.98	0.00
	(3.46)	(3.11)	
Hours studied per week	5.30	4.88	0.00
	(3.27)	(3.14)	
Student is employed	0.04	0.04	0.00
	(0.19)	(0.20)	
Number of schools ranked	9.97	9.99	0.00
	(3.70)	(3.61)	
COMIPEMS score	66.63	59.86	0.00
	(16.83)	(13.83)	
Median COMIPEMS score of assigned school	66.22	62.72	0.00
	(14.56)	(12.91)	
Distance from home to assigned school (km)	7.40	6.99	0.00
	(6.25)	(6.06)	
Cohort size of assigned school	1095.94	624.86	0.00
	(1092.63)	(554.87)	
Dropped out (did not take ENLACE)	0.49	0.51	0.00
	(0.50)	(0.50)	
Delayed/dropped out	0.52	0.54	0.00
(did not take ENLACE within 3 years)	(0.50)	(0.50)	
Normalized ENLACE score	0.02	-0.13	0.00
	(0.94)	(0.88)	
Observations	676114	89435	

Standard deviations in parentheses. Column 1 includes all students in their final year at a public middle school in Mexico City who were assigned to a school during the automated portion of the assignment process (i.e. did not choose from remaining seats after failing to place at any listed option). Dropout and delay measures do not include students admitted to UNAM schools.

Table 3.2: Regression discontinuity estimates of effect of admission on dropout probability

	(1) Dropout, all cutoffs	(2) Dropout, no IPN cutoffs	(3) Dropout or delay, all	(4) Dropout or delay, no IPN cutoffs
Admitted (scored above cutoff)	0.716	-0.243	0.836	-0.489
	(0.641)	(0.674)	(0.636)	(0.668)
Centered COMIPEMS score	-0.0890	-0.183	0.0418	-0.0379
	(0.158)	(0.166)	(0.157)	(0.165)
Centered COMIPEMS × admitted	-0.273	-0.166	-0.424*	-0.324
	(0.221)	(0.232)	(0.219)	(0.230)
Middle school GPA	-19.43***	-19.23***	-20.75***	-20.53***
	(0.219)	(0.231)	(0.217)	(0.229)
Observations	89435	80807	89435	80807
Adjusted R^2	0.132	0.134	0.141	0.143
Cutoff FEs	Yes	Yes	Yes	Yes
AIC-optimal polynomial order	1	1	1	1

Standard errors in parentheses

Dependent variable in columns 1 and 2 is dropout × 100 so that coefficients are percentage point changes in predicted dropout probability per 1-unit change in the independent variable. Dependent variable in columns 3 and 4 is delay or dropout × 100, which is characterized by either taking the exam after 4+ years or not at all. Columns 2 and 4 exclude students who were at the boundary between an IPN school (above the cutoff) and a non-IPN school (below the cutoff).

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Table 3.3: Regression discontinuity estimates of heterogeneous effects of admission on dropout with respect to changes in student-school attributes

	(1)	(2)	(3)
Admitted (scored above cutoff)	-4.991***		
	(1.203)		
Δ median COMIPEMS score due to admission	-0.270***	-0.188**	-0.165*
	(0.0771)	(0.0793)	(0.0885)
Δ distance to school due to admission	-0.221***	-0.194**	-0.160*
	(0.0830)	(0.0836)	(0.0857)
Δ log cohort size due to admission	-0.239	0.238	1.053
	(0.614)	(0.640)	(0.902)
Admit $\times$ (Δ median COMIPEMS)	0.499***	0.330***	0.239**
	(0.0929)	(0.0993)	(0.111)
Admit $\times$ (Δ distance)	0.283***	0.238**	0.213**
	(0.102)	(0.103)	(0.106)
Admit $\times$ (Δ log cohort size)	0.418	-0.555	-0.616
	(0.725)	(0.810)	(1.133)
Observations	84355	84355	84355
Adjusted R^2	0.135	0.141	0.142
Cutoff FEs	Yes	Yes	Yes
Cutoff $\times$ admit FEs	No	Yes	Yes
Δ subsystem effects	No	No	Yes
AIC-optimal polynomial order	1	1	1

Standard errors in parentheses

Dependent variable is (dropout $\times$ 100) so that coefficients are percentage point changes in dropout probability. The Δ variables are the difference in the respective attribute between the school above the cutoff and the school below the cutoff. All specifications include the student's middle school GPA, the AIC-optimal order of piecewise polynomial in centered COMIPEMS exam score, the uninteracted Δ variables, and interactions between each Δ variable and the AIC-optimal piecewise polynomial in centered COMIPEMS exam score. Column 3 includes one control per subsystem for whether the subsystem of the schools above and below the cutoff are different (=0 if both or neither belong, =1 if school above belongs to that subsystem and the school below does not, =-1 if school below belongs but the school above does not), and interactions between these variables and admission and the AIC-optimal piecewise polynomial in centered COMIPEMS exam score.

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Table 3.4: Regression discontinuity estimates of heterogeneous effects of admission on ENLACE score with respect to changes in student-school attributes

	(1)	(2)	(3)	(4)
Admitted (scored above cutoff)	0.00392	-0.0515**		
	(0.0133)	(0.0258)		
Δ median COMIPEMS score due to admission		0.00223	0.00407**	0.00445**
		(0.00155)	(0.00159)	(0.00176)
Δ distance to school due to admission		0.000734	0.000904	0.000252
		(0.00178)	(0.00180)	(0.00186)
Δ log cohort size due to admission		-0.0112	-0.0134	0.0188
		(0.0129)	(0.0135)	(0.0192)
Admit $\times$ (Δ median COMIPEMS)		0.00495***	0.00117	0.00159
		(0.00189)	(0.00201)	(0.00222)
Admit $\times$ (Δ distance)		-0.000464	-0.000909	-0.000603
		(0.00220)	(0.00223)	(0.00230)
Admit $\times$ (Δ log cohort size)		0.00279	0.0106	-0.0460*
		(0.0153)	(0.0171)	(0.0241)
Observations	43754	41384	41384	41384
Adjusted R^2	0.418	0.419	0.427	0.430
Cutoff FEs	Yes	Yes	Yes	Yes
Cutoff $\times$ admit FEs	No	No	Yes	Yes
Δ subsystem effects	No	No	No	Yes
AIC-optimal polynomial order	1	1	1	1

Standard errors in parentheses

Dependent variable is normalized ENLACE exam score. The Δ variables are the difference in the respective attribute between the school above the cutoff and the school below the cutoff. All specifications include the student's middle school GPA and the AIC-optimal order of piecewise polynomial in centered COMIPEMS exam score. Specifications 2-4 include the uninteracted Δ variables and interactions between each Δ variable and the AIC-optimal piecewise polynomial in centered COMIPEMS exam score. Column 4 includes one control per subsystem for whether the subsystem of the schools above and below the cutoff are different (=0 if both or neither belong, =1 if school above belongs to that subsystem and the school below does not, =-1 if school below belongs but the school above does not), and interactions between these variables and admission and the AIC-optimal piecewise polynomial in centered COMIPEMS exam score.

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Table 3.5: Regression discontinuity tests for balance of baseline covariates

	(1) GPA	(2) Parental ed.	(3) Income	(4) Male	(5) Hours studied	(6) Works
Admitted (scored above cutoff)	-0.0117	-0.0217	-0.0436	0.000487	-0.0415	-0.00421
	(0.00982)	(0.0446)	(0.0442)	(0.00637)	(0.0450)	(0.00297)
Centered COMIPEMS score	0.0205***	0.0543***	0.0468***	0.000377	0.0468***	-0.000296
	(0.00241)	(0.0110)	(0.0109)	(0.00157)	(0.0111)	(0.000730)
Centered COMIPEMS × admitted	-0.00216	-0.0378**	-0.0312**	0.00232	-0.0122	0.000865
	(0.00338)	(0.0154)	(0.0152)	(0.00220)	(0.0155)	(0.00102)
Observations	89435	79638	78475	89435	79393	77076
Adjusted R^2	0.154	0.094	0.066	0.136	0.040	0.004
Cutoff FEs	Yes	Yes	Yes	Yes	Yes	Yes
Mean of dependent variable	8.030	9.884	3.981	0.452	4.875	0.0419
SD of dependent variable	0.775	3.210	3.112	0.498	3.144	0.200

Standard errors in parentheses

Dependent variable is dropout × 100 so that coefficients are percentage point changes in predicted dropout probability per 1-unit change in the independent variable. Joint estimation of columns 1-6 using seemingly unrelated regression and testing for joint significance of the admission coefficients yields a p-value of .54.

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Bibliography

Aaronson, Daniel, Lisa Barrow, and William Sander, "Teachers and Student Achievement in the Chicago Public High Schools," *Journal of Labor Economics* 25 (2007), 95-135.

Abdulkadiroglu, Atila, Joshua Angrist, and Parag Pathak, "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," NBER Working Paper No. 17264 (2011).

Abdulkadiroglu, Atila, and Tayfun Sonmez, "Matching Markets: Theory and Practice" Prepared for the Econometric Society World Congress, China (2010).

Ajayi, Kehinde F., "School Choice and Educational Mobility," Unpublished working paper, Boston University (2012).

Altonji, Joseph, Todd Elder, and Christopher Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy* 113 (2005), 151-184.

Angrist, Joshua, Eric Bettinger, and Michael Kremer, "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review* 96 (2006), 847-862.

Ball, Stephen J., and Carol Vincent, "'I Heard It on the Grapevine': 'Hot' Knowledge and School Choice," *British Journal of Sociology of Education* 19 (1998), 377-400.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics* 90 (2008), 414-427.

Campos-Vázquez, Raymundo M., "Why Did Wage Inequality Decrease in Mexico after NAFTA?," *Economía Mexicana NUEVA EPOCA* 22 (2013), 245-278.

Ceja, Miguel, "Understanding the Role of Parents and Siblings as Information Sources in the College Choice Process of Chicana Students," *Journal of College Student Development* 47 (2006), 87-104.

Chade, Hector, and Lones Smith, "Simultaneous Search," *Econometrica* 74 (2006), 1293-1307.

Clark, Damon, "Selective Schools and Academic Achievement," *The B.E. Journal of Economic Analysis & Policy (Advances)* 10 (2010), Article 9.

Crawford, Gregory S., and Matthew Shum, "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica* 73 (2005), 1137-1173.

Cullen, Julie, Brian Jacob, and Steven Levitt, "The Impact of School Choice on Student Outcomes: An Analysis of the Chicago Public Schools," *Journal of Public Economics* 89 (2005), 729-760.

Cullen, Julie, Brian Jacob, and Steven Levitt, "The Effect of School Choice on Participants: Evidence from Randomized Lotteries," *Econometrica* 74 (2006): 1191-1230.

de Hoop, Jacobus, "Selective Schools and Education Decisions: Evidence from Malawi." Unpublished working paper, Free University of Amsterdam (2011).

Dearden, Lorraine, Javier Ferri, and Costas Meghir, "The Effect of School Quality on Educational Attainment and Wages," *The Review of Economics and Statistics* 84 (2002), 1-20.

Dobbie, Will, and Roland Fryer, Jr., "Exam High Schools and Academic Achievement: Evidence from New York City," NBER Working Paper 17286 (2011).

Dubins, L. E., and D. A. Freedman, "Machiavelli and the Gale-Shapley Algorithm," *The American Mathematical Monthly* 88 (1981), 485-495.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, "Peer Effects and the Impacts of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review* 101 (2011), 1739–74.

Erdem, Tülin, and Michael P. Keane, "Decision-Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets," *Marketing Science* 15 (1996), 1-20.

Estrada, Ricardo, and Jeremie Gignoux, "Benefits to Elite Schools and the Formation of Expected Returns to Education: Evidence from Mexico City," Working Paper 2014-06, Paris School of Economics (2014).

Foster, Andrew D., and Mark R. Rosenzweig, "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy* 103 (1995), 1176-1209.

Friesen, Jane, Mohsen Javdani, Justin Smith, and Simon Woodcock, "How Do School 'Report Cards' Affect School Choice Decisions?," *Canadian Journal of Economics* 45 (2012), 784-807.

Gale, David, and Lloyd S. Shapley, "College Admissions and the Stability of Marriage," *The American Mathematical Monthly* 69 (1962), 9-15.

Glewwe, Paul, Nauman Ilias, and Michael Kremer, "Teacher Incentives," *American Economic Journal: Applied Economics* 2 (2010), 205-227.

Gould, Eric, Victor Lavy, and Daniele Paserman, "Immigrating to Opportunity: Estimating the Effect of School Quality Using a Natural Experiment on Ethiopians in Israel," *Quarterly Journal of Economics* 119 (2004), 489-526.

Hastings, Justine, Thomas Kane, and Douglas Staiger, "Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery," NBER Working Paper No. 12145 (2009).

Hastings, Justine, and Jeffrey Weinstein, "Information, School Choice, and Academic Achievement: Evidence from Two Experiments," *Quarterly Journal of Economics* 123 (2008), 1373-1414.

Horowitz, Joel, and Charles Manski, “Identification and Robustness with Contaminated and Corrupted Data,” *Econometrica* 63 (1995), 281-302.

Hoxby, Caroline M., and Avery, Christopher, “The Missing ’One-Offs’: The Hidden Supply of High-Achieving, Low Income Students,” NBER Working Paper No. 18586 (2012),

Hoxby, Caroline M., and Sarah Turner, “Expanding College Opportunities for High-Achieving, Low Income Students,” SIEPR Discussion Paper No. 12-014 (2013).

Imbens, Guido, and Thomas Lemieux, “Regression Discontinuity Designs: A Guide to Practice,” *Journal of Econometrics* 142 (2008), 615-635.

James, William, *Principles of Psychology*, (Cambridge: Harvard University Press, 1983). Originally published 1890.

Jackson, Kirabo, “Do Students Benefit From Attending Better Schools?: Evidence From Rule-based Student Assignments in Trinidad and Tobago,” *The Economic Journal* 142 (2010), 1399-1429.

Johnson, Justin P., and David P. Myatt, “On the Simple Economics of Advertising, Markting, and Product Design,” *American Economic Review* 96 (2006), 756-784.

Koning, Pierre, and Karen van de Wiel, “Ranking the Schools: How Quality Information Affects School Choice in the Netherlands,” IZA Discussion Paper No. 4984 (2010).

Lai, Fang, Elisabeth Sadoulet, and Alain de Janvry, “The Adverse Effects of Parents’ School Selection Errors on Academic Achievement: Evidence from the Beijing Open Enrollment Program,” *Economics of Education Review* 28 (2009), 485-496.

Lai, Fang, Elisabeth Sadoulet, and Alain de Janvry “The Contributions of School Quality and Teacher Qualifications to Student Performance: Evidence from a Natural Experiment in Beijing Middle Schools,” *Journal of Human Resources* 46 (2011), 123-53.

Lee, David S., “Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects,” *Review of Economic Studies* 76 (2009), 1071-1102.

Lee, David S., and David Card, “Regression Discontinuity Inference with Specification Error,” *Journal of Econometrics* 142 (2008), 655-674.

Lee, David S., and Thomas Lemieux, “Regression Discontinuity Designs in Economics,” *Journal of Economic Literature* 48 (2010), 281-355.

Lucas, Adrienne M., and Isaac M. Mbiti, “The Determinants and Consequences of School Choice Errors in Kenya,” *American Economic Review: Papers and Proceedings* 102 (2012), 283-288.

Marsh, Herbert W., and Kit-Tai Hau, “Big-Fish–Little-Pond Effect on Academic Self-Concept: A Cross-Cultural (26-Country) Test of the Negative Effects of Academically Selective Schools,” *American Psychologist*, 58 (2003), 364-376.

Marsh, Herbert W., and John W. Parker. “Determinants of Student Self-Concept: Is it Better to Be a Relatively Large fish in a Small Pond Even if You Don’t Learn to Swim as Well?,” *Journal of Personality and Social Psychology*, 47 (1984), 213-231.

Manski, Charles F., “Identification of Endogenous Social Effects: The Refection Problem,” *Review of Economic Studies* 60 (1993) 531-42.

Manski, Charles F., *Identification Problems in the Social Sciences*, Harvard University Press (1995).

McCrary, Justin, “Manipulating the Running Variable in the Regression Discontinuity Design: A Density Test,” *American Economic Review* 97 (2008), 318-353.

Mizala, Alejandra, and Miguel Urquiola, “School Markets: The Impact of Information Approximating Schools’ Effectiveness,” *Journal of Development Economics* 103 (2013), 313-335.

Moretti, Enrico, “Social Learning and Peer Effects in Consumption: Evidence from Movie Sales,” *Review of Economic Studies* 78 (2011), 356-393.

Newhouse, David, and Kathleen Beegle, “The Effect of School Type on Academic Achievement: Evidence from Indonesia,” *Journal of Human Resources* 41 (2006): 529-557.

Pallais, Amanda, “Small Differences that Matter: Mistakes in Applying to College.” Revise and resubmit, *Journal of Labor Economics* (2009).

Pop-Eleches, Cristian, and Miguel Urquiola, “Going to a Better School: Effects and Behavioral Responses,” *American Economic Review* 103 (2013), 1289-1324.

Roberts, John H., and Glen L. Urban, “Modeling Multiattribute Utility, Risk, and Belief Dynamics for New Consumer Durable Brand Choice,” *Management Science* 34 (1988), 167-185.

Roth, Alvin E., “Incentive Compatibility in a Market With Indivisible Goods,” *Economics Letters* 9 (1982), 127-132.

Zhang, Hongliang, “The Mirage of Elite Schools: Evidence from Lottery-based School Admissions in China,” Mimeo (2013).