Peer Networks and School Choice under Incomplete Information

Andrew Dustan*
Department of Agricultural and Resource Economics
University of California, Berkeley

October 30, 2013

Preliminary. Please do not distribute.

Abstract

Many school systems give students and their families some degree of choice in which school they attend. This choice is difficult because students have incomplete information about schools. This paper 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 twelve years of detailed student-level data from Mexico City's public high school choice system, 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 peer 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.

---

*email: adustan@berkeley.edu. I thank Alain de Janvry, Elisabeth Sadoulet, and Fred Finan for their guidance and Ben Faber, Jeremy Magruder, and Ted Miguel for helpful comments. Seminar participants at UC Berkeley provided useful feedback. Roberto Peña Reséndiz supplied the COMIPEMS data and offered excellent insight into the school selection process, and Rafael de Hoyos afforded access to the ENLACE data. This research was funded in part by grants from the Institute of Business & Economic Research and the Weiss Family Fellowship through the Center for Effective Global Action.

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 characteristics 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 its value for updating expectations, some information may be beneficial once the student is actually attending that school, because it tells him how to use the school in an optimal way. 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.

[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. (2013) for Mexico.

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 school 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 having incomplete information about schools and relying on information obtained from peers to reduce this uncertainty.

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 characteristics 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 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 in-

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

formation 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 characteristics).[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 in the population.[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).

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

[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 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.

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. The data span twelve years of admissions cycles and contain rich information about the choices, demographics, and assignment of each participant. Combining the students' name, location, 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, and 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 characteristics 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 characteristics 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. Rather, uncertainty appears to originate from incomplete information about more specific or idiosyncratic elements of match quality (or a lack of productive knowledge). When students do not know basic school attributes, an easy solution may be to distribute official information about school characteristics. But resolving uncertainty about idiosyncratic match quality and transferring productive knowledge 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 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 3 sets forth a simple model of school choice under incomplete information, concluding with testable hypotheses about the expected effect of new information. Section 4 explains the data and how they will be used to test the model. Section 5 gives the OLS method and results, while Section 6 lays out the discrete choice model and corresponding results. Section 7 provides validity checks for the empirical design and Section 8 concludes with policy recommendations.

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.

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.[6] 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.

The COMIPEMS admissions process is as follows.[7] In late January, students in ninth

[6]The discussion in this section draws on de Janvry et al. (2013).

[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

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 Media 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.

2.2 Student decision making under the COMIPEMS mechanism

Students have considerable information about basic school characteristics 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 Universidad

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.

Nacional Autónoma de México (UNAM) and more technically-focused Instituto Politécnico Nacional (IPN). These are universally understood to be highly competitive, relatively rigorous, prestigious high schools with cutoff scores that exceed those of 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 for each school have been available on the COMIPEMS website 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 generic school characteristics, though they may lack information about specific characteristics that affect the idiosyncratic match between the student and school.

There are more than twenty educational options available to students (indeed, more than 600 in some years), and in any case listing many options can become very tedious, so students omit most available options from their lists. 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). Students in Mexico City often construct their rankings in the following way, similar to how United States students choose colleges (see Hoxby and Avery (2013), 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 32 elite schools (16 in each), 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 that offer a better chance of admission. It is common to observe a "safety" school listed near the bottom of the rankings in order to avoid the possibility of being left unassigned.

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

[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.

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 only a minute percentage of students actually fill up their entire preference sheet; students generate a satisfactory choice portfolio without confronting the space constraint. 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 for ranking unattainable schools.

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 of 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 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 the signal was. Finally, these effects are predicted to apply, to a lesser degree, to other schools that are observably similar to the school for which the information was received.

[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).

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 $\boldsymbol{\widetilde{X}_j}$ is the student-specific deviation from this level. An example of a student-school attribute is academic quality, 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 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 $U\left(\boldsymbol{X_{ij}\beta_i}\right) = \boldsymbol{X_{ij}\beta_i}$, 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: $\boldsymbol{X_{ij}\beta_i} > \boldsymbol{X_{ik}\beta_i}\ \forall k \neq j \in J$.[14]

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

[14] 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.

$1/\tau^0_{\ell ij}$.[15]

Because $\boldsymbol{X_{ij}}$ is uncertain, a risk-neutral student chooses j if it maximizes *expected* match quality: $\mathrm{E}_0\left[\boldsymbol{X_{ij}\beta_i}\right] > \mathrm{E}_0\left[\boldsymbol{X_{ik}\beta_i}\right]\ \forall k \neq j \in J$, 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 raised the weight that students placed on test scores when choosing schools.[16]

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 $r_j\left(n_{ij}\right)$ 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 and decreasing so that $r'_j > 0$ and $r''_j < 0$. Examples of productive knowledge 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:

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

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 can be written in

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

[16] 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.

terms of the mean and variance (or precision) of the prior distribution of match quality:[17]

$$\begin{aligned} U^*_{0ij} &= \mathrm{E}_0\left[\boldsymbol{X_{ij}}\boldsymbol{\beta_i}\right] - \frac{\rho}{2}Var\left(\boldsymbol{X_{ij}}\boldsymbol{\beta_i}\right) + r_j\left(n_{ij}\right) \\ &= \boldsymbol{X^0_{ij}}\boldsymbol{\beta_i} - \frac{\rho}{2}\sum_{\ell}\frac{\beta^2_{\ell i}}{\tau^0_{\ell ij}} + r_j\left(n^0_{ij}\right). \end{aligned} \tag{1}$$

where $\beta^2_{\ell i}/\tau^0_{\ell ij}$ is the variance of the distribution of match quality from characteristic 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: $U^*_{0ij} > U^*_{0ik}\ \forall k \neq j \in J$.

3.4 Effect of peer information

When student i's peer attends school j, she provides two pieces of information. First, she provides productive knowledge about school j, so that the new level of knowledge is now 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:

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

where $\boldsymbol{\Sigma_{P_{ij}}}$ is diagonal with entries $1/\tau^P_{\ell ij}$. The signals received are about student-school attributes for student i, not the peer.[18] 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 characteristic is simple:

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

[17]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.

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

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

$$U_{1ij}^{*}=\widehat{\boldsymbol{X}_{\boldsymbol{ij}}^{\mathbf{1}}}\boldsymbol{\beta_i}-\frac{\rho}{2}\sum_{\ell}\frac{\beta_{\ell i}^{2}}{\left(\tau_{\ell ij}^{0}+\tau_{\ell ij}^{P}\right)}+r_j\left(n_{ij}^{1}\right). \tag{2}$$

where $\widehat{\boldsymbol{X}_{\boldsymbol{ij}}^{\mathbf{1}}}$ is the mean of the posterior distribution of $\boldsymbol{X}_{\boldsymbol{ij}}^{\mathbf{1}}$. To see how the peer signals affected expected utility, compare equations 1 and 2:

$$U_{1ij}^{*}-U_{0ij}^{*}=\left(\widehat{\boldsymbol{X}_{\boldsymbol{ij}}^{\mathbf{1}}}-\boldsymbol{X}_{\boldsymbol{ij}}^{\mathbf{0}}\right)\boldsymbol{\beta_i}+\frac{\rho}{2}\sum_{\ell}\frac{\beta_{\ell i}^{2}\tau_{\ell ij}^{P}}{\tau_{\ell ij}^{0}\left(\tau_{\ell ij}^{0}+\tau_{\ell ij}^{P}\right)}+\left(r_j\left(n_{ij}^{1}\right)-r_j\left(n_{ij}^{0}\right)\right) \tag{3}$$

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 other hand, the reduction in uncertainty about match quality always works in the school's favor. Note that the expected effect may be positive for certain schools and negative for others, because mean quality $\bar{\boldsymbol{X}}_{\boldsymbol{i}}$ 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.

3.5 Shared characteristics across schools

Students may know that the level of a characteristic is shared across schools. In the empirical setting studied here, schools are divided into subsystems that share important characteristics 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 characteristics channel.) In order to model the shared characteristics in a simple way, we can maintain all prior assumptions of the model and additionally assume that for school j in subsystem s, $\boldsymbol{X_{ijs}} = \bar{\boldsymbol{X}}_{\boldsymbol{js}} + \widetilde{\boldsymbol{X}}_{\boldsymbol{ijs}} + \mu_{is}$, where $\mu_{is} \equiv \bar{\mu}_s + \widetilde{\mu}_{is}$. The average component of subsystem match quality is distributed $\bar{\mu}_s \sim \mathcal{N}\left(\bar{\mu}_s^0, \sigma_s^2\right)$ and the student-specific component is distributed $\widetilde{\mu}_{is} \sim \mathcal{N}\left(\widetilde{\mu}_{is}^0, \eta_{is}^2\right)$, and $1/\tau_{is}^{\mu} \equiv \sigma_s^2 + \eta_{is}^2$. In addition to the signal $\boldsymbol{P_{ij}}$ about unshared characteristics, the student receives a signal about the shared attribute:

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

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:

$$U_{1iks}^{*} - U_{0iks}^{*} = \left(\widehat{\mu}_{is}^{1} - \mu_{is}^{0}\right) + \frac{\rho}{2} \frac{\tau_{is}^{q}}{\tau_{is}^{\mu}\left(\tau_{is}^{\mu} + \tau_{is}^{q}\right)} \tag{4}$$

where $\widehat{\mu}_{is}^{1}$ is the mean of the posterior distribution of the shared characteristic and μ_{is}^{0} 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 other schools in the same subsystem is positive: indexing the peer's school by j and all other schools in j's subsystem s_j by k_j,* $\mathrm{E}_{ij}\left[\mathrm{E}_{k_j}\left[U_{1ik_js_j}^{*} - U_{0ik_js_j}^{*}\right]\right] > 0$.

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. 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.

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.

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.

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.[19] 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

[19]For more details on the ENLACE and how it relates to graduation, see de Janvry et al. (2013).

do not report a birth order that makes them the closest siblings in the family (e.g. first- and second-born), 3) the students were born fewer than nine months apart or more than five years apart, or 4) the older student took the exam after the younger one. If 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.[20] Columns 1 and 2 of Table 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.[21] 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.

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 school assignment of the older sibling. The COMIPEMS school assignment

[20] 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.

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

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.[22] The basic idea behind this design is to define, for each school, the sample of older siblings who were either marginally (barely) admitted or marginally rejected from that school, and then compare the school choices between 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: $-w \leq s_i - c_j < w$;[23]
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

[22] 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. (2013).

[23] 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.

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.[24] 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 listing more-preferred schools with cutoffs slightly above j's, which both reduces sample size and results in estimates of the 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 normal 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. This is because 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, 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 be imitated by their younger siblings in the RD sample (37% probability) as in the full sibling sample (34%).

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

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 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.

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:

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

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).[25] Now including fixed effects for cutoff school and older sibling's exam year, the stacked specification is:

$$y_{ijt} = \delta admit_{ijt} + f_1(\tilde{s}_{ijt}) + admit_{ijt} f_2(\tilde{s}_{ijt}) + \mu_j + \eta_t + \varepsilon_{ijt} \tag{5}$$

The parameter δ is now the local average treatment effect of admission across all school-years.[26]

While non-parametric approaches are common in papers using RD, Lee and Card (2008) explain that when the running variable (here, exam score) is discrete, non-parametric 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 part 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 deviation in the population exam score distribution, respectively) are used.[27] 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).[28] 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.

[25] An exception is when only schools from the elite systems are being considered. In that case there are only 32 schools, most 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.

[26] 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.

[27] 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 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.

[28] 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.

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 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.

Bandwidth selection has implications beyond the usual bias-efficiency trade-off, as explained in Section 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 graphs use a bandwidth of 5.

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 3 shows this result. 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 2. Also reported in Table 3 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 more schools from the same subsystem, beyond the effect of choosing the older sibling's exact school. Table 4, accompanied by Figure 3, 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. Thus 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 .36 more schools in the same subsystem. This cannot be accounted for by the .068 to .077 effect of admission on probability of listing the same school reported in column 1 of Table 2. Even if the entire same-school effect were due to students adding the sibling's school as the top choice, without removing any other schools in the same subsystem, the effect on total schools chosen in the subsystem would only be .068 to .077. Thus it must be that admission causes siblings, on average, to list additional schools in the same subsystem as a result of sibling admission. 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.

The effect on subsystem demand also suggests that younger siblings are learning about idiosyncratic match quality, not simply generic characteristics, of subsystems. Column 4 of Table 4 reports estimates of the admission effect on subsystem demand, restricting the sample to those at the margin of one of the two elite subsystems. Despite the fact that both of these systems have universally strong reputations for quality, older sibling admission leads to a large and significant increase in the number of schools selected in the same subsystem. One possible explanation for this result is that even though students know that the elite subsystems offer high match quality on average, there is substantial residual uncertainty about the student-specific component of match and that students respond to a decrease in this uncertainty with increased demand. Another explanation is that the student gains substantial productive knowledge from his old sibling about elite schools.

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. The surprise to match quality is unobserved, but one proxy for it is an indicator for whether the older sibling graduated from high school or not. The logic for using this measure 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 drops out, this suggests to the younger sibling that the school may not be a good match for him. The older sibling had a prior belief about her graduation probability between 0 and 1, so dropping out or graduating are both "surprises" for her and for the younger sibling.

Any estimates of the differential effects of dropout and graduation 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:

$$\begin{aligned} y_{ijt} = \delta admit_{ijt} + f_1\left(\tilde{s}_{ijt}\right) + admit_{ijt} f_2\left(\tilde{s}_{ijt}\right) + \mu_j + \eta_t + \\ graduate_{ijt}\left\{\alpha admit_{ijt} + g_1\left(\tilde{s}_{ijt}\right) + admit_{ijt} g_2\left(\tilde{s}_{ijt}\right) + \nu_j + \varphi_t\right\} + \varepsilon_{ijt}, \end{aligned} \tag{6}$$

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 schools chosen in the same subsystem. If dropout were randomly assigned, then this would give the differential effect of admission with respect to graduation. The problem arises when $cor\left(graduate_{ijt} \times admit_{ijt}, \varepsilon_{ijt}\right) \neq 0$, so that students who are *differentially* 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, considers multiple samples and argue 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 serves as a proxy for how likely the student is to be emulated, 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, not the sibling's individual graduation outcome. To control for this, older sibling admission and exam score are both interacted with school graduation rate so that the estimated heterogeneity is from sibling dropout conditional on the dropout rate.

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 5 shows that the admission effect is quite 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 × graduation" which is the additional impact of admission when the sibling graduates instead of dropping out. The point estimates are positive and statistically significant for each reported bandwidth, suggesting that on average admission has a 3.2 to 6.1 percentage point higher impact on first choice preference for the cutoff score when the older sibling graduates.[29] The differential effect is illustrated in Figure 4, Panel A. Column 2 controls for interactions between admission and older sibling GPA and the cutoff school's dropout rate. 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,

[29]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.

ranging form 9 to 11 percentage points compared to a sample mean of about 25%. Graphical results are in Figure 4, Panel B. Adding controls in column 4, the estimates remain almost identical in magnitude and statistical significance.

Heterogeneity also exists in the effect on demand for schools in the same subsystem, consistent with Hypothesis 4. Columns 5 and 6 show the differential impact on the number of schools selected in the cutoff school's subsystem, restricting the sample to cases where the older sibling is at the margin of a subsystem. Without controls, the differential effect is only statistically significant for the 10-point bandwidth. With controls, the point estimates are similar and statistically significant for the 5-point and 10-point bandwidths. For these bandwidths, the estimated differential effect is .37 to .38 additional schools selected in the cutoff subsystem. Figure 4, Panel C illustrates this relationship. Thus, between the same-school and subsystem effects, it appears that younger siblings react to signals from siblings with "good" and "bad" outcomes differently, learning about match quality from their peers and updating their choice behavior accordingly.

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 a school in the same subsystem.

6.1 Method

The expected utility formulation from the theoretical model in Section 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) or with a signal (equation 2), we have:

$$U^*_{0ij} = \boldsymbol{X^0_{ij}}\boldsymbol{\beta_i} - \frac{\rho}{2}\sum_{\ell}\frac{1}{\tau^0_{\ell ij}}$$

$$U_{1ij}^{*}=\widehat{\boldsymbol{X}_{\boldsymbol{ij}}^{\mathbf{1}}}\boldsymbol{\beta_i}-\frac{\rho}{2}\sum_{\ell}\frac{1}{\tau_{\ell ij}^{0}+\tau_{\ell ij}^{P}}.$$

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:

$$U_{ij}^{*}=\alpha_{ij}+\delta_{ij}admit_{ij}$$

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:

$$U_{ij}^{*}=\delta admit_{ij}+\varepsilon_{ij}$$

where the error term ε_{ij} captures heterogeneity in the mean of the prior, its variance, 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 older siblings who are close to a cutoff, adding 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:

$$U_{ij}^{*}=\theta cut_{ij}+\delta\left(cut_{ij}\times admit_{i}\right)+f_{1}\left(\tilde{s}_{i}\right)cut_{ij}+f_{2}\left(\tilde{s}_{i}\right)\left(cut_{ij}\times admit_{i}\right)+\gamma dist_{ij}+\varepsilon_{ij}$$

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 j due to peer 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 $below_{ij}=1$ when j is the school to which the older sibling would or did attend upon scoring too low for admission to the cutoff school.[30] Then the specification can be expanded to include the effect of having a sibling admitted to the school below the cutoff:

$$\begin{aligned} U^*_{ij} = \theta cut_{ij} + \delta \left(cut_{ij} \times admit_i\right) + f_1\left(\tilde{s}_i\right) cut_{ij} + f_2\left(\tilde{s}_i\right)\left(cut_{ij} \times admit_i\right) \\ + \underline{\theta} below_{ij} + \underline{\delta}\left(below_{ij} \times \left(1 - admit_i\right)\right) + \underline{f}_1\left(\tilde{s}_i\right) below_{ij} + \underline{f}_2\left(\tilde{s}_i\right)\left(below_{ij} \times \left(1 - admit_i\right)\right) \\ + \gamma dist_{ij} + \varepsilon_{ij} \end{aligned}$$

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, due to reduced uncertainty about student-subsystem match.[31] 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 $X_j^1, ...X_j^M$ 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 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:

$$\begin{aligned} U^*_{ij} = \theta cut_{ij} + \delta \left(cut_{ij} \times admit_i\right) + f_1\left(\tilde{s}_i\right) cut_{ij} + f_2\left(\tilde{s}_i\right)\left(cut_{ij} \times admit_i\right) \\ + \underline{\theta} below_{ij} + \underline{\delta}\left(below_{ij} \times \left(1 - admit_i\right)\right) + \underline{f}_1\left(\tilde{s}_i\right) below_{ij} + \underline{f}_2\left(\tilde{s}_i\right)\left(below_{ij} \times \left(1 - admit_i\right)\right) \\ \sum_{\ell=2}^{M} X_j^\ell \left(\pi^\ell + \eta^\ell cutsub_{ij} + \underline{\eta}^\ell belowsub_{ij}\right) + cutsub_{ij}\left[\phi admit_i + h_1\left(\tilde{s}_i\right) + h_2\left(\tilde{s}_i\right) admit_i\right] + \\ belowsub_{ij}\left[\underline{\phi}\left(1 - admit_i\right) + \underline{h}_1\left(\tilde{s}_i\right) + \underline{h}_2\left(\tilde{s}_i\right)\left(1 - admit_i\right)\right] + \gamma dist_{ij} + \varepsilon_{ij}. \end{aligned} \quad (7)$$

This specification includes subsystem fixed effects (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

[30] 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.

[31] Recall that this is an average effect. For some students the signal is a negative surprise for expected utility and for some it is a positive surprise. The average is positive because of the reduced uncertainty about marginal match quality and the increase in productive knowledge.

$-\phi/\gamma$ 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.[32] Defining V_{ij} as containing all terms in equation 6 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 $\underline{\widehat{\delta}} > 0$ (sibling admission to a school increases, on average, expected utility from attendance) and the test of Hypothesis 3 is whether $\widehat{\phi} > 0$ and $\underline{\widehat{\phi}} > 0$ (sibling admission to a school increases, on average, marginal 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 larger when the sibling graduated than when she did not. These can be tested by estimating the nested logit based on equation 6, 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 $\underline{\widehat{\phi^g}} > 0$.

[32] 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.

6.2 Results

Estimating the discrete choice model yields direct evidence for each of the hypotheses of the social learning model. Table 6 provides selected estimated parameters from the nested logit specification in equation 8, 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.[33]

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 1.8km for the school above the cutoff and 3.0km 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 raises WTT to all other schools in that subsystem by 1.5km. Rejection (and hence admission to the system below the cutoff) raises WTT to all other schools in the below subsystem by 1.7km. These results support the OLS findings, which could only provide suggestive evidence on the 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 is .44,

[33] While the correlated error structure induced by the discrete running variable (Lee and Card (2008)) is still an issue here, 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.

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 in 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 × admission." The effect of admission on WTT to the school above the cutoff is 3.1km 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 3.7km 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 2.2km more when the sibling graduates (p-value=.12), while the differential effect for the subsystem below the threshold is 2.9km (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.

6.3 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 the 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 raises 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.

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 of middle school. Figure 5 shows the density for a bandwidth of 5 for the RD sample of older siblings (corresponding to column 1 of Table 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 quality.[34]

The second check is to repeat the RD 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. Table 7 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.

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 build productive knowledge. Having an older sibling at a particular school increases revealed preference not only for that school, but also for schools in the same subsystem. This 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, as predicted by the social learning model.

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 policymakers hoping to encourage such behavior in other contexts. One lesson is that aggregated 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 net-

[34]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 quite compelling in this case.

work to connect with prospective students, providing them with personalized information through informal meetings and repeated electronic communication. 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 data beyond school-level aggregates is one way to begin; arranging school visits including individual meetings with students and faculty for those in disadvantaged networks may be another positive approach.

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 and are less likely to apply, 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.

References

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

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

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

Ball, Stephen J., and Carol Vincent. 1998. "'I Heard It on the Grapevine': 'Hot' Knowledge and School Choice." British Journal of Sociology of Education 19(3): 377-400.

Bowen, William G., Matthew M. Chingos, and Michael S. McPherson. 2011. *Crossing the Finish Line: Completing College at America's Public Universities.* Princeton University Press.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2008. "Bootstrap-Based Improvements for Inference with Clustered Errors." Review of Economics and Statistics 90(3): 414-427.

Ceja, Miguel. 2006. “Understanding the Role of Parents and Siblings as Information Sources in the College Choice Process of Chicana Students.” Journal of College Student Development 47(1): 87-104.

Chade, Hector, and Lones Smith. 2006. “Simultaneous Search.” Econometrica 74(5): 1293-1307.

Clark, Damon. 2010. “Selective Schools and Academic Achievement.” The B.E. Journal of Economic Analysis & Policy (Advances) 10(1), Article 9.

Crawford, Gregory S., and Matthew Shum. 2005. “Uncertainty and Learning in Pharmaceutical Demand.” Econometrica 73(4): 1137-1173.

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

de Janvry, Alain, Andrew Dustan, and Elisabeth Sadoulet. 2013. “Flourish or Fail? The Risky Reward of Elite High School Admission in Mexico City.” Unpublished working paper.

Dobbie, Will, and Roland G. Fryer, Jr. 2011. “Are High-Quality Schools Enough to Increase Achievement Among the Poor? Evidence from the Harlem Children’s Zone.” American Economic Journal: Applied Economics 3(3): 158-187.

Dubins, L. E., and D. A. Freedman. 1981. “Machiavelli and the Gale-Shapley Algorithm.” The American Mathematical Monthly 88(7): 485-495.

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

Foster, Andrew D., and Mark R. Rosenzweig. 1995. “Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture.” Journal of Political Economy 103(6): 1176-1209.

Friesen, Jane, Mohsen Javdani, Justin Smith, and Simon Woodcock. 2012. “How Do School ’Report Cards’ Affect School Choice Decisions?” Canadian Journal of Economics 45(2): 784-807.

Gale, David, and Lloyd S. Shapley. 1962. “College Admissions and the Stability of Marriage.” The American Mathematical Monthly 69(1): 9-15.

Hastings, Justine S., Thomas J. Kane, and Douglas O. Staiger. 2009. “Heterogeneous Preferences and the Efficacy of Public School Choice.” NBER Working Paper No. 12145.

Hastings, Justine, and Jeffrey Weinstein. 2008. “Information, School Choice, and Academic Achievement: Evidence from Two Experiments.” Quarterly Journal of Economics 123(4): 1373-1414.

Hoxby, Caroline M., Avery, Christopher. 2013. “The Missing ’One-Offs’: The Hidden Supply of High-Achieving, Low Income Students." NBER Working Paper No. 18586.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." Journal of Econometrics 142(2): 615-635.

Jackson, Kirabo. 2010. "Do Students Benefit From Attending Better Schools?: Evidence From Rule-based Student Assignments in Trinidad and Tobago." The Economic Journal 142(549): 1399-1429.

Johnson, Justin P., and David P. Myatt. 2006. "On the Simple Economics of Advertising, Markting, and Product Design." American Economic Review 96(3): 756-784.

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

Lai, Fang, Elisabeth Sadoulet, and Andrew Dustan. 2009. "The Adverse Effects of Parents' School Selection Errors on Academic Achievement: Evidence from the Beijing Open Enrollment Program." Economics of Education Review 28(4): 485-496.

Lai, Fang, Elisabeth Sadoulet, and Alain de Janvry. 2011 "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(1): 123-53.

Lee, David S., and David Card. 2008. "Regression Discontinuity Inference with Specification Error." Journal of Econometrics 142(2): 655-674.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." Journal of Economic Literature 48(2), 281-355.

Lucas, Adrienne M., and Isaac M. Mbiti. 2012. "The Determinants and Consequences of School Choice Errors in Kenya." American Economic Review: Papers and Proceedings 102(3): 283-288.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Refection Problem." Review of Economic Studies 60(3): 531-42.

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

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." Journal of Econometrics 142(2): 698-714.

Mizala, Alejandra, and Miguel Urquiola. 2013. "School Markets: The Impact of Information Approximating Schools' Effectiveness." Journal of Development Economics 103: 313-335.

Moretti, Enrico. 2011. "Social Learning and Peer Effects in Consumption: Evidence from Movie Sales." Review of Economic Studies 78(1): 356-393.

Pallais, Amanda. 2009. "Small Differences that Matter: Mistakes in Applying to College." Revise and resubmit, Journal of Labor Economics.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a Better School: Effects and

Behavioral Responses." American Economic Review, forthcoming.

Roberts, John H., and Glen L. Urban. 1988. "Modeling Multiattribute Utility, Risk, and Belief Dynamics for New Consumer Durable Brand Choice." Management Science 34(2): 167-185.

Roth, Alvin E. 1982. "Incentive Compatibility in a Market With Indivisible Goods." Economics Letters 9(2): 127-132.

Smith, Jonathan, Matea Pender, and Jessica Howell. 2013. "The Full Extent of Student-College Academic Undermatch." Economics of Education Review 32: 247-261.

Train, Kenneth E. 2009. *Discrete Choice Methods with Simulation.* Cambridge University Press.

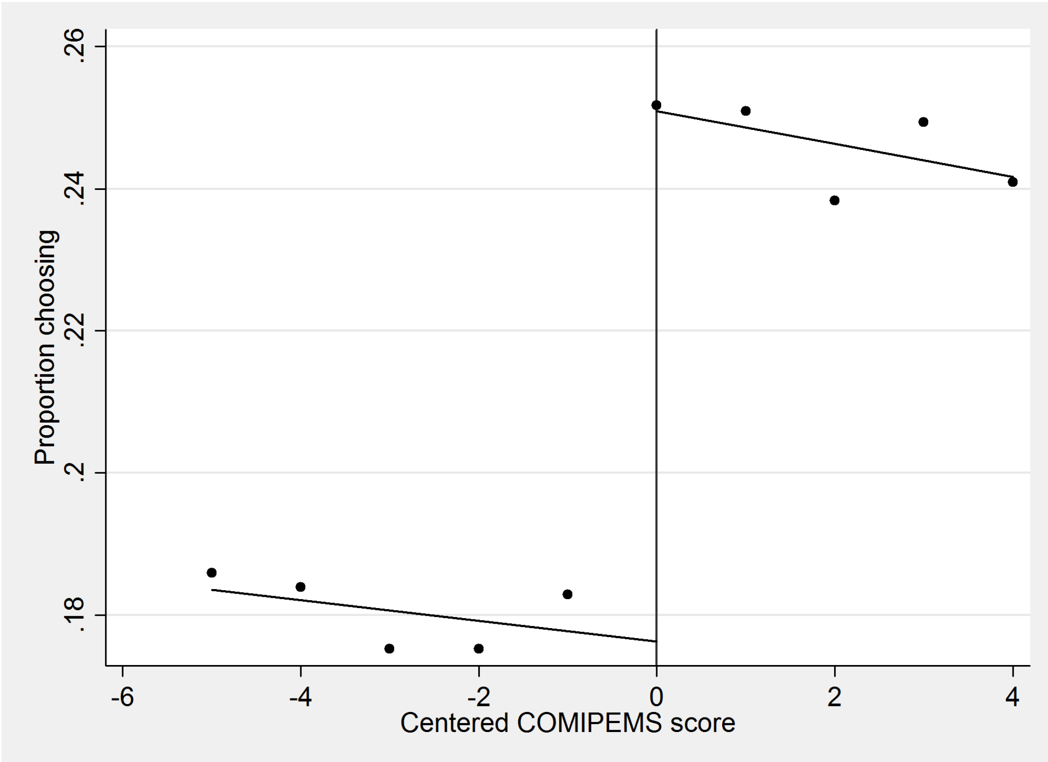
Zellner, Arnold. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." Journal of the American Statistical Association 57(298): 348-368.

Zhang, Hongliang. 2012. "The Mirage of Elite Schools: Evidence from Lottery-based School Admissions in China." Unpublished working paper, Chinese University of Hong Kong.

Appendix: Derivation of model hypotheses

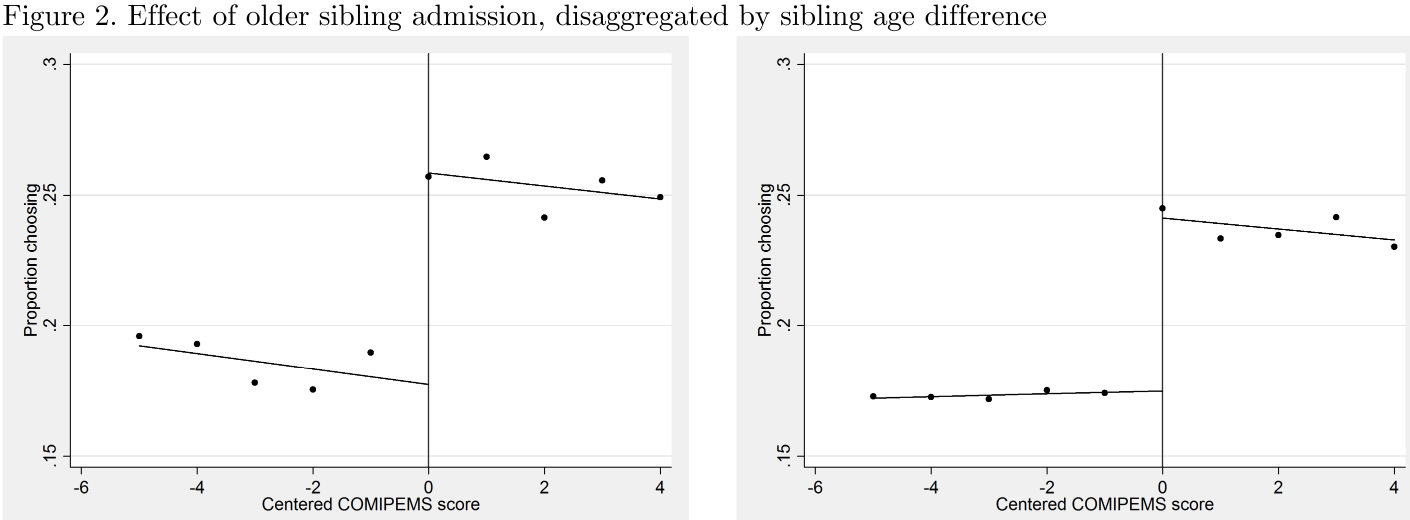
Under revision.

Figure 1. Effect of older sibling admission on younger sibling's first choice preference for same school



Note. 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 2. Effect of older sibling admission, disaggregated by sibling age difference

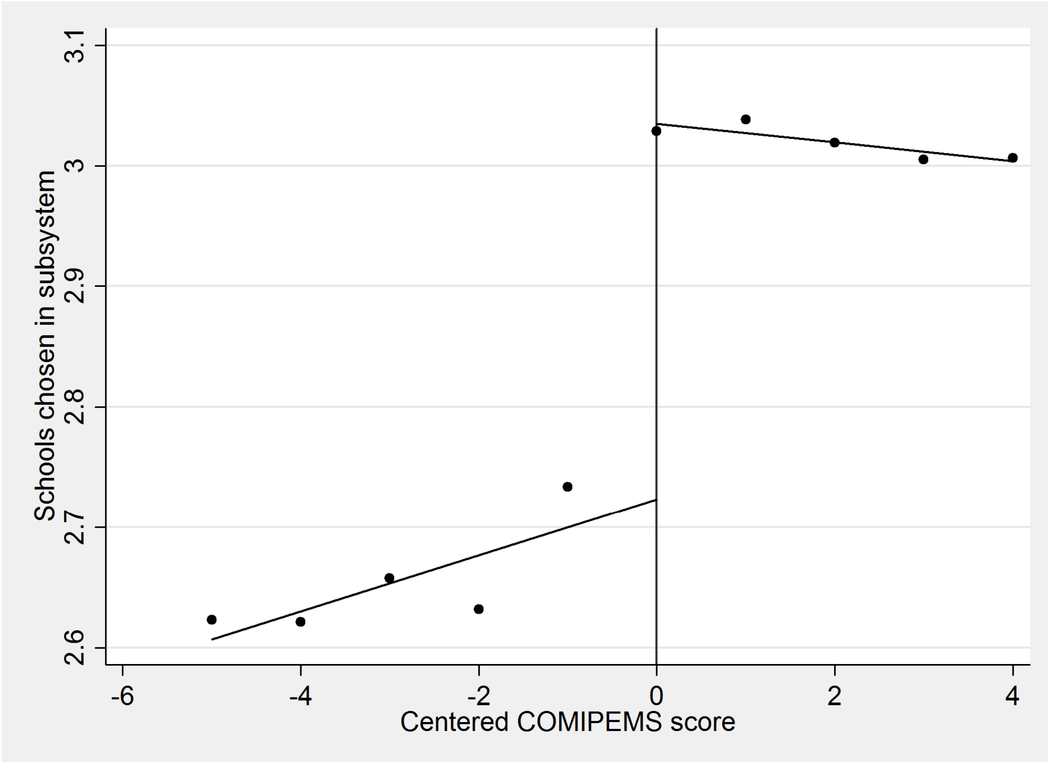


(a) Siblings 1 or 2 years apart

(b) Siblings 3 to 5 years apart

Note. 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.

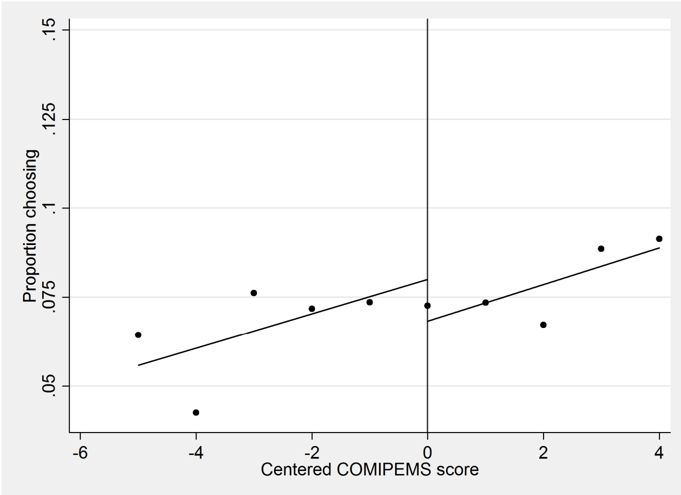
Figure 3. Effect of older sibling admission on number of schools selected in the cutoff school's subsystem



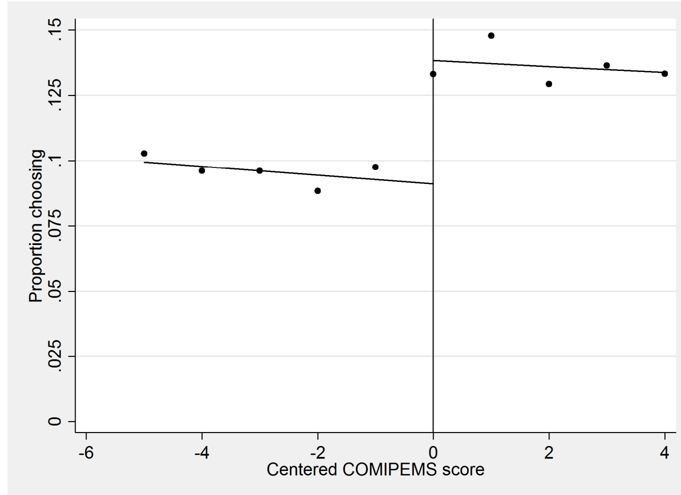
Note. Variable on vertical axis is number of schools selected in the subsystem to which the older sibling's cutoff school belongs. Variable on horizontal axis is older sibling's COMIPEMS exam score, centered to be 0 at the corresponding cutoff score. Fitted lines are from a linear fit.

Figure 4. Effect of older sibling admission, by graduation outcome

Panel A. Effect on first choice



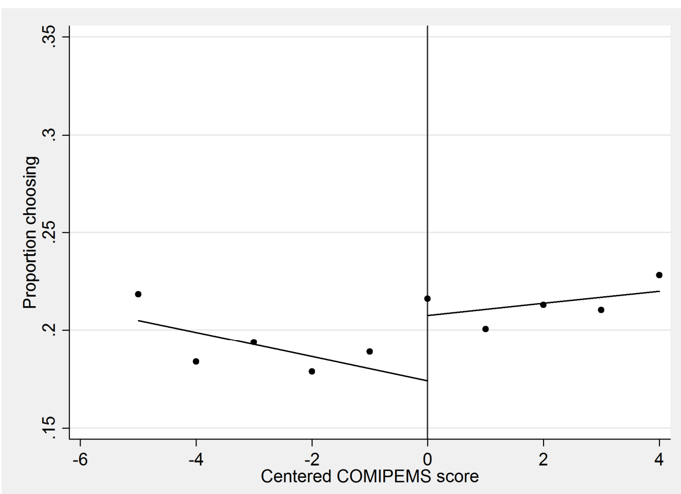
(a) Older sibling did not graduate



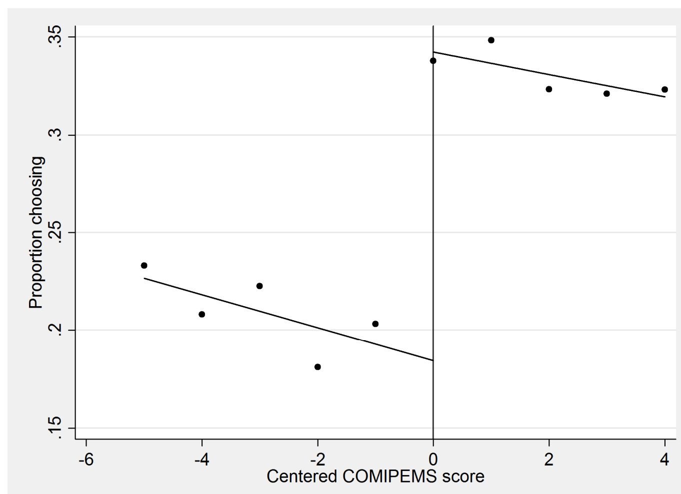
(b) Older sibling graduated

Note. 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.

Panel B. Effect on first non-elite choice



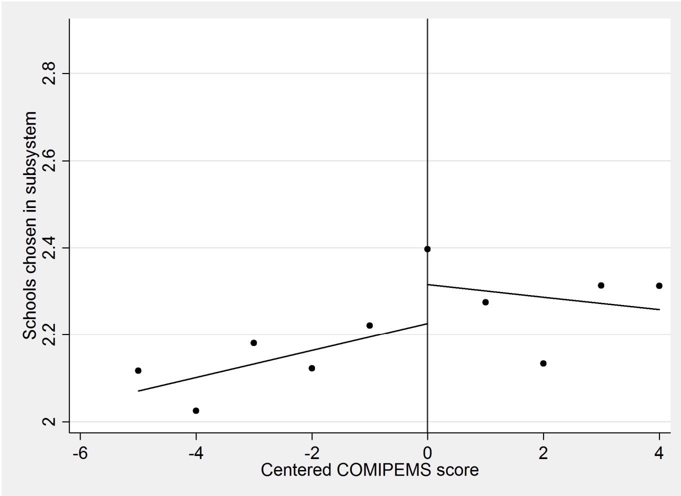
(a) Older sibling did not graduate



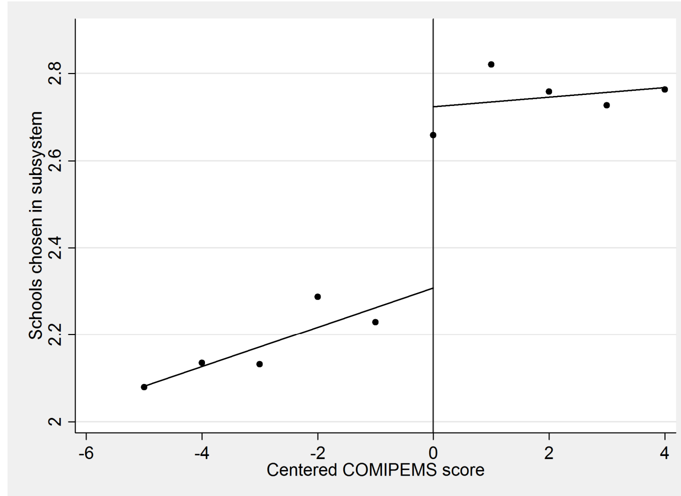
(b) Older sibling graduated

Note. 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.

Panel C. Effect on number of schools selected in the cutoff school's subsystem



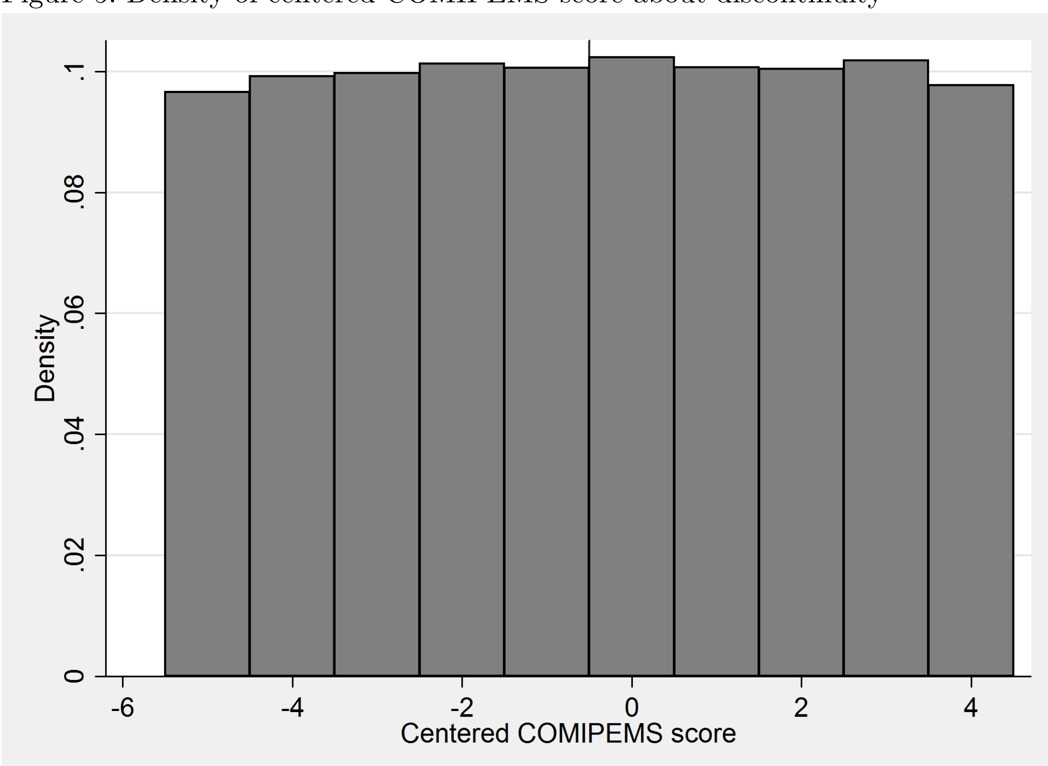
(a) Older sibling did not graduate



(b) Older sibling graduated

Note Variable on vertical axis is number of schools selected in the subsystem to which the older sibling's cutoff school belongs. 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 5. Density of centered COMIPEMS score about discontinuity



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

Table 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 in Section 4.3 for a bandwidth of 5.

Table 2. Effect of older sibling admission 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 3. Effect of older sibling admission on younger sibling's preference for same school, disaggregated

	(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 4. Effect of older sibling admission on number of schools chosen in cutoff subsystem

	(1)	(2)	(3)	(4)	(5)
	All siblings	Siblings 1-2 years apart	Siblings 3-5 years apart	All siblings, elite cutoffs only	All siblings, non-elite cutoffs only
Bandwidth=3					
Admission coefficient	0.35**	0.34*	0.38*	0.42**	0.26*
	(.02)	(.01)	(.07)	(.02)	(.10)
Mean of dependent variable	2.83	2.79	2.92	3.32	2.22
R^2	0.20	0.20	0.22	0.12	0.23
Observations	31,474	17,630	13,518	17,654	13,820
AIC-optimal polynomial order	1	1	1	1	1
Difference in admission coefficient between 3-5 years apart and 1-2 years apart samples			0.03		
			(.30)		
Bandwidth=5					
Admission coefficient	0.36***	0.33***	0.39***	0.47***	0.19
	(.01)	(.00)	(.00)	(.00)	(.13)
Mean of dependent variable	2.84	2.79	2.93	3.29	2.23
R^2	0.19	0.19	0.21	0.13	0.22
Observations	48,153	27,105	20,771	27,695	20,458
AIC-optimal polynomial order	1	1	1	1	1
Difference in admission coefficient between 3-5 years apart and 1-2 years apart samples			0.049		
			(.45)		
Bandwidth=10					
Admission coefficient	0.37***	0.37***	0.37***	0.50***	0.17***
	(.00)	(.00)	(.01)	(.00)	(.01)
Mean of dependent variable	2.88	2.82	2.95	3.25	2.29
R^2	0.18	0.18	0.19	0.12	0.21
Observations	76,901	43,442	33,140	47,166	29,735
AIC-optimal polynomial order	1	1	1	1	1
Difference in admission coefficient between 3-5 years apart and 1-2 years apart samples			-0.010		
			(.92)		

Note. Dependent variable is the number of schools selected by the younger sibling that belong to the cutoff school's subsystem. Only students who are 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 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. 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 5. 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 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	0.003	0.047*	0.043	0.304*	0.282*
	(.31)	(.25)	(.06)	(.22)	(.10)	(.10)
Admission × older sib graduation	0.041*	0.041**	0.090	0.097	0.080	0.141
	(.09)	(.05)	(.13)	(.13)	(.62)	(.64)
Admission × older sib GPA		-0.014		0.021		0.009
		(.20)		(.54)		(.94)
Admission × cutoff school dropout rate		0.113		-0.276		-1.789
		(.25)		(.13)		(.16)
Mean of dependent variable	0.0893	0.0893	0.226	0.226	2.355	2.355
R^2	0.126	0.128	0.174	0.175	0.252	0.254
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	-0.005	0.034**	0.036*	0.175*	0.136
	(.15)	(.54)	(.05)	(.09)	(.09)	(.23)
Admission × older sib graduation	0.061**	0.053**	0.110**	0.106**	0.301	0.383**
	(.03)	(.04)	(.00)	(.01)	(.15)	(.05)
Admission × older sib GPA		-0.008		0.030*		-0.101
		(.12)		(.06)		(.37)
Admission × cutoff school dropout rate		0.210**		-0.134		-0.792
		(.05)		(.36)		(.43)
Mean of dependent variable	0.0975	0.0975	0.243	0.243	2.349	2.349
R^2	0.122	0.124	0.161	0.161	0.231	0.231
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**	0.029**	0.024	0.028*	0.153	0.163*
	(.05)	(.04)	(.16)	(.10)	(.19)	(.07)
Admission × older sib graduation	0.032*	0.017	0.097**	0.092**	0.387**	0.371**
	(.07)	(.36)	(.00)	(.00)	(.03)	(.01)
Admission × older sib GPA		0.003		0.023		0.009
		(.58)		(.25)		(.85)
Admission × cutoff school dropout rate		0.198**		-0.083		0.558
		(.04)		(.21)		(.33)
Mean of dependent variable	0.126	0.126	0.265	0.265	2.402	2.402
R^2	0.129	0.131	0.152	0.152	0.208	0.208
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. Also included are de-meaned older sibling's GPA and cutoff school's dropout rate, first-order piecewise polynomials in the interaction between these variables and centered test score, and the interaction between GPA and graduation. **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. Only students who are at the threshold of a subsystem are included, i.e. rejection from the cutoff school implies attending a different subsystem.

Table 6. Nested logit estimates of school choice model

Variable	Interacted with	(1) All siblings	(2) Siblings 3-5 years apart	(3) 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 7. 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. Standard errors are in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.