Elite schools and the formation of expectations of returns to education: evidence from Mexico City

Ricardo Estrada and Jérémie Gignoux*

March 2013

Abstract

This study examines whether, and in what ways, the attendance of elite high schools schools affect students' expectations of the earnings returns to higher educational attainments. We use panel data for a sample of individuals who attend public senior-secondary schools of Mexico City between 2005 and 2008/9 and exploit the variations in admission to more selective schools generated by a centralized and competitive exam-based allocation process of students into public senior-high schools. We find that admission to an elite high school increases substancially both learning achievements and the earnings and returns expected from a college education. The results are consistent with students valuing the higher cognitive, and possibly non-cognitive, skills that they acquire in those elite schools. At the opposite, we find no higher effects on the expectations of students from more disadvantaged family background, suggesting that an imperfect information channel is not explaining our findings. Also, the lack of evidence of effects of admission to elite high schools on the earnings expected with only a high-school diploma tends to discard that the observed changes in expectations are driven by reputation effects in the labor market (i.e. a signaling channel). These findings bring evidence that the schooling environment provided by elite institutions can have benefits that extend beyond gains in scholastic achievements.

JEL: D84, I24, J71.

Keywords: elite high schools, beliefs formation, earnings expectations, returns to education, Mexico.

*Estrada: Paris School of Economics (PSE), 48 boulevard Jourdan, 75014 Paris, restrada@pse.ens.fr. Gignoux: PSE and French National Institute for Research in Agronomy (INRA), 48 boulevard Jourdan, 75014 Paris, gignoux@pse.ens.fr. We are grateful to François Bourguignon, Denis Cogneau, Julie Cullen, Adeline Delavande, Julien Grenet, Marc Gurgand, Rafael de Hoyos, Katia Kaufmann, Sylvie Lambert for helpful comments on earlier drafts. Insightful comments were also received at seminars at the Paris School of Economics and Mannheim University, and at the ECINEQ and LACEA conferences. We thank the staff at the Mexican Ministry of Education for kindly providing us access to data.

1 Motivation

Subjective expectations of the returns to a higher educational attainment are a key determinant of individuals' demand for education. Standard human capital theory predicts that schooling decisions depend on the monetary benefits (together with non-monetary ones) that youth expect from a higher attainment. Some studies have modeled schooling decisions based on assumptions on the way these beliefs are formed; for instance, while some authors assume perfect foresight and rational expectations, others have assumed that individuals' expectations depend on the actual earnings of older cohorts or specific reference groups (e.g. Dominitz and Manski (1996); Manski (1993)). Instead of relying on assumptions, a more recent strand of studies have elicited individuals' expectations, documented the consistency of the obtained measures (Delavande, Giné, and McKenzie 2011), and verified that subjective earnings expectations are closely associated with choices regarding further schooling (Attanasio and Kaufmann 2009; Hotz, Arcidiacono, and Kang 2010). However, little is known on the formation of subjective expectations of the financial returns to higher educational attainments. Beside a couple of studies describing the differentials in earnings expectations and evaluating the quality of the information possessed by different groups of youth (Betts 1996; Zafar 2011), little empirical evidence exists on the formation of expectations.

A specific set of questions concerns the extent to which earnings expectations are influenced by schooling environments. Are these expectations formed at adolescent ages and influenced by schooling environments or shaped early by families (and mayble also neighborhoods of residence)? Are expectations adjusted as youth attain higher scholastic learning achievements or build some non-cognitive skills such as self-confidence? Can imperfect information on earnings opportunities with higher attainments be corrected when attending specific schools and interacting with different groups of peers?

In this paper, we tackle these questions by examining the effects of attending an elite school system on the earnings expectations of a sample of Mexico City high school students. Using a natural experiment from the exam-based admission process to public high schools of the Federal District (FD) of Mexico, we examine the extent to which schooling environments shape earnings expectations. We find evidence that being admitted to an elite high school increases substancially the earnings that students expect with a college education.

There are several ways in which attending an elite school might affect a youth' beliefs on earnings with higher attainments. First, following human capital theory, if these are also better schools, more selective schools could provide higher levels of skills to their students, which, assuming students believe that wages equal the marginal product of labor and a minimum consistency in belief formation, should in turn increase their earnings expectations. Note that a proper definition of skills would include both scholastic achievement and non-cognitive skills. Second, drawing on signaling theory, students might assess the graduation from (or attendance of) a an elite school as a signal of higher productivity. The reputation of a selective school could increase the expected earnings either directly in the labor market or by facilitating the access to more selective universities. Third, if some students have imperfect information on labor market opportunities, better schools may provide more precise information about the distribution of earnings by educational attainment and dilute the over- or underestimation in earnings expectations.[1]

While several of those alternative models of the effects of elite schools on earnings expectations could be relevant in the context under study, some of our results suggest that higher expectations do not stem from the signaling effect of attending a reputed high school or from better access among disadvantaged students to information on labor market opportunities, but rather from the acquisition of some cognitive skills, and maybe also non-cognitive ones, that increase the value from attending college.

Understanding the effects of schools on expected earnings benefits from higher attainments is important for several reasons. First, in some contexts, notably in developing countries, and among youth from poorer backgrounds, the demand for further schooling, in particular at college levels, might be inefficiently low due to incorrect and underestimated expectations of earnings returns (Jensen 2010). In other contexts, youth could make the opposite mistake and overestimate those returns. While such imperfect information and the associated orientation mistakes can be costly, some school environments might provide better information and help students make better choices regarding further studies. Second, some school environments might help students build some specific non-cognitive skills, such as self-confidence or future orientedness, that are complementary to scholastic achievements in determining later attainments and labor outcomes. In all cases, identifying the schools and inputs inducive of higher or more adequate expectations, which may not be the same ones that improve learning achievements, could help find ways to improve schooling decisions at higher levels. It could also allow identifying a source of potentially inefficient educational inequalities, and help design interventions that address those.

In this paper, for examining the effects of attending a Mexico City elite high school on beliefs of the financial returns to attending college in the context of public high schools, we exploit data from a panel of students followed from application and entry into high schools (at 15 year olds for most) to graduation (at 18 years old). We begin by describing the patterns of future earnings and returns to college subjectively expected (5 years after high school graduation) by students and their variations by students and schools characteristics. The observed differentials indicate higher earnings expectations for students with higher learning achievements and more advantaged backgrounds give a sense of the internal validity of the data (the latter could stem either from some information or non-cognitive skills provided by wealthier families or some complementarities with their resources), give a sense of the internal validity of the data. However, while expected earnings with college increase with learning achievements, students in the sample do not perceive positive returns to cognitive skills when holding a high school degree. For external validity, we

[1] Note that students may also have imperfect information on their own ability, and that being exposed to higher achieving peers may also affect this information. Besides, more selective schools may not be the ones which convey the most correct information (or most increase students' non-cognitive skills).

compare earnings expectations to the earnings observed and returns estimated for older cohorts in the Mexican employment survey, and find that many students have relatively high expectations with respect to the earnings of those older cohorts. Regarding the differentials across schools, we observe that, even after conditioning on learning achievements at graduation, expectations of the earnings and returns to college education are both higher and less variable in elite senior secondary schools.

For causal inference, we exploit the allocation of students into high schools based on a common exam and a regression discontinuity (RD) design analysis, and estimate the effects of attending a selective elite high school on expected earnings and returns. Our RD design estimates depart from the approach followed in previous studies, notably Abdulkadiroglu, Angrist, and Pathak (2011); Pop-Eleches and Urquiola (2011), by identifying the effects of admission to the IPN system rather than to a specific IPN school - we do this by considering the cutoff for admission to the IPN elite schools system rather than a specific school. We find that admission to an elite high school increases substancially both learning achievements and the earnings and returns expected from a college education, with point estimates of effects of elite high school admission of about .2 standard deviations on scholastic achievements and about 19 percent points on the expected college premium. This is consistent with students valuing the higher cognitive, and possibly non-cognitive, skills that they acquire in those elite schools. At the opposite, we find no higher effects on the expectations of students from more disadvantaged family background, suggesting that the information mechanism (or other mechanisms affecting in different ways students from richer and poorer backgrounds) is not explaining our results. Also, the lack of evidence of effects of admission to elite high schools on the earnings expected with only a high-school diploma tends to discard that the observed changes in expectations are driven by the reputation of high schools on the labor market (i.e. a signaling mechanism).

For getting a sense of the economic significance of our findings, note that the results of Attanasio and Kaufmann (2009) on the relationship between earnings expectations and attainments, based on an observational study among a similar sample of Mexican students, suggest that a one percent point increase in the expected college premium (compared to a high school degree) is associated with a .2 to .4 percent point higher probability (higher for boys) of attending college. This is probably an overestimate, but interpreting this estimate causally would imply that the increase in earnings expectations that comes with the attendance of an elite school could translate into a up to 4 to 8 points higher probablity of college attendance.

This paper is related to three strands of the literature. First, a group of recent empirical studies have examined the effects of admission into a more selective or higher performing school. Several of these have used data similar to the one we are using (panels of students allocated into schools through a common exam) and examined the effects on learning achievements. The results are mixed: while several studies find positive effects, notably in the context of poor or intermediate-income countries - Pop-Eleches and

Urquiola (2011) in Romania, Jackson (2010) in Trinidad and Tobago, Hastings and Weinstein (2008) in Malawi, and Ding and Lehrer (2007) in China -, Abdulkadiroglu, Angrist, and Pathak (2011) find that gains from attending very selective high schools (i.e. for students at the upper end of the achievement distribution) are not large in the United States (although some positive effects are found for ethnic minorities).[2] Also in the US, Angrist, Pathak, and Walters (2011) and Dobbie and Roland G. Fryer (2011) find substantial gains of attendance of charter schools by students at lower end of achievement distribution, which suggests that the effects of schools on learning achievements may be nonlinear. Other studies have investigated the attendance of better schools induced by specific interventions. For instance, Angrist, Bettinger, and Kremer (2006) examine a voucher program in Colombia and find significant impacts of winning a voucher though a lottery on graduation rates and other measures of academic performance. Hastings and Weinstein (2008) analyze the effects of attending a better school in the US, induced by information provided (randomly) on the choice of schools, and also find positive effects on academic achievements. Thus there is evidence that, in some contexts and for some students, attendance of elite high schools can increase learning achievements. We build on this literature and argue that better or more selective schools might, beyond current learning, generate other benefits, notably in terms of information regarding future learning and labor market opportunities. The evidence on these other effects, which matter for long-term educational and professional outcomes, remains limited.

Second, a small number of studies have examined the differentials in expectations on the monetary returns to higher educational attainments. Betts (1996) thus documents the quality of the information and beliefs of college students on wages, in a sample of undergraduates from University of California at San Diego, and finds that beliefs vary with personal background and year and field of study. Similarly Zafar (2009, 2011) documents the differentials in expected returns to college education across males and females and majors in a sample of undergraduates at Northwestern University and finds that expectations of both educational outcomes and earnings drive differentials in schooling choices. Delaney, Harmon, and Redmond (2011) similarly describe the differentials in those expectations by family background among Irish university students and find persisting gaps by socio-economic status even after controlling for students outcomes and personal traits. Thus, while there are studies that document the gaps in expectations by family background, their determinants, in particular the influence of school environments, has not been examined.[3][4]

Third, other studies have examined the extent to which the information on the earnings returns to

[2] In a paper in progress, Dustan, De Janvry, and Sadoulet (2012) use the same dataset used in this study and also find significant effects of admisssion to IPN elite schools on learning achievements.

[3] There exists a few theoretical studies though. For instance, Roemer and Wets (1994) propose a theoretical model of the effect of spatial segretation on beliefs formation and the distribution of schooling and income based on learning on own ability.

[4] There are also some related studies of non-cognitive abilities and perceptions of own abilities. For instance, Chevalier, Gibbons, Thorpe, Snell, and Hoskins (2009) find, in a descriptive study on a sample of British university students, that self-confidence and academic self-perception are strongly correlated with schooling decisions, that students are often too much self-confident on their academic opportunities (although some groups are too pessimistic), and that less competent students have poorer judgements.

higher attainments affects educational decisions. Hotz, Arcidiacono, and Kang (2010) find that sorting into majors of a sample of US college students is based in part on their expected returns (they also find evidence of learning as students of some majors and upper grades have more precise expectations). Jensen (2010), based on experimental data from the Dominican Republic, finds that providing information on returns to education affects schooling decisions even at basic education levels.[5] While those studies show that expectations matter for later choices, our study investigates the formation of those expectations.

The paper is organized as follows. In section 2, we present the system of Mexico City elite high schools under study and discuss the way attending one of these schools can affect students' expectations of earnings with higher attainments. In section 3, we present the earnings expectations data, check their internal and external validity, and describe the differentials in expectations between students of elite and regular high schools. In section 4, we discuss the process of allocation of students into high schools based on RD design. In section 6, we discuss the obtained results and robustness checks. Section 7 concludes.

2 Mexico City elite high schools

We begin by presenting the system of elite high schools of Mexico City under study, and the way the learning environment these schools provide differs from the one of other high schools, and might affect earnings expectations.

2.1 IPN schools

We consider the effects of admission into any of the 16 high schools of the National Polytechnic Institute ("Instituto Politécnico Nacional" or IPN), a system of elite senior secondary schools of Mexico City. There are ten institutional systems of senior public high schools in Mexico City Metropolitan Area.[6] Together with high schools of the "Universidad Nacional Autonoma de México" (UNAM), IPN schools are the most selective.[7] IPN is also the main public technological higher education center in Mexico, and IPN high schools are administratively associated with that higher education institution. They provide general senior secondary-level education with a specific scientific and technical background, but admission in IPN high schools does not grant access to IPN higher education institute.

IPN high schools are accessed through a centralized selection process shared with most public high schools of Mexico City and based on a common exam (and explicit school choices) (we describe this process in section 4). IPN schools are more demanded and thus more selective than other senior high schools. Among students applying to the centralized system for application to public senior high schools of

[5] A related literature exists on school choices; for instance Hastings and Weinstein (2008) find using experimental data that providing information on quality of schools also affects application choices.

[6] Data from the Mexican Ministry of Education indicate that, during school year 2005-2006, 18.7% of senior high-school students of the Federal District of Mexico were enrolled in a private school.

[7] Our data does not allow us to investigate the effects of attending a UNAM school because UNAM students do not take the test at high school completion and thus drop out of the panel.

the FD for school year 2005-2006, the cohorts among whom our data was collected, 19,042 were admitted to IPN schools and 210,468 to other public high schools (including 34,625 to UNAM schools).

Admission to an IPN school modifies the schooling environment and experience of students in several ways. Table 1 gives descriptive statistics for school inputs provided to IPN and non-IPN students (based on information from the Mexican census of schools, see below).[8] Being selected among the best applicants, IPN students, beside being themselves higher performers, tend to have school peers with higher achievements and, because family background is positively related to achievements and preferences, from more privileged backgrounds. Students' average test-score at the (public) high school entry exam is almost 1.7 standard deviations higher in IPN schools than in non-IPN schools, and the dispersion of achievements is lower, with an average school-level standard deviation of test-scores of .53 (unconditional) standard deviation against .62 in other high schools. Besides, the descriptive results presented in section 3 show that youth attending IPN schools, and thus also peers of an IPN student, have higher expectations than youth in other schools of the wages they would achieve with a college education.

IPN students are also likely to receive better teaching, as teachers in those schools tend to be more qualified and experienced, and more of other school inputs. Table 1 shows that more teachers have a college degree in IPN schools (87%) than in other schools (82%), and teachers of IPN schools are more often in a full-time position (29%) than teachers in other schools (14%).[9] Furthermore, class sizes tend to be smaller in IPN schools, with an average of 39.8 students per class against 42.4 in non-IPN schools, and IPN schools provide a much better access to computers, with 3.6 students per computer in IPN schools against 11.4 in non-IPN schools. Attendance of an IPN high school could also bring credentials that could be valued on the labor market.

In addition, although we do not have information on this, some IPN schools may also engage in some activities to provide information on the returns to higher education. In particular, given their institutional relations with IPN higher education institutes, the administrative and teaching personnel may have information on the job opportunities and wages, notably in the scientific and manufacturing sectors, that could be communicated to their students.

2.2 School inputs and expected returns to education

When compared to other high schools, IPN elite high schools provide to their students a set of enhanced inputs, and several of those inputs are likely to affect expectations of the earnings that can be obtained with higher educational attainments. Three main theoretical frameworks can be resorted to for thinking about the effects of elite school environments on expected earnings and returns to further education.

First, better teaching, interactions with higher performing peers or other school inputs could improve

[8]The group of non-IPN students excludes UNAM students.

[9]However, the teaching provided in IPN schools might not necessarily benefit all students as teachers may target specific (likely majoritarian) groups of students (Duflo et al. 2010).

learning achievements, which, following a classical human capital model and under some conditions (notably that scholastic learning is viewed as a direct determinant of future earnings or as an input for obtaining a more valued higher education diploma), is likely to translate into higher earnings expectations. These school inputs could also enhance students' non-cognitive skills that also enter in the formation of earnings expectations, such as self-confidence and future-orientedness. However, attending a more selective school could also have adverse effects on skill acquisition for students who could get marginalized either because of lower relative scores (that could make the teaching in those schools less suitable for them) or because they belong to a socio-economic minority in the school they attend (e.g. from poor neighborhoods). If the later effects are not necessarily positive, the main implication of a human capital theoretical model of the formation of earnings expectations is that the effects of attending an elite high school on the earnings expected with higher educational attainments should be positively associated with its effects on the accumulation of some skills (scholastic and cognitive or non-cognitive). Note however that the effects of attending an elite schools on the expected returns to higher levels are less predictable as they would depend on the respective effects on the earnings with different educational attainments; one can expect though that the returns to most skills increase with higher educational attainments.

Second, following an imperfect information model, the attendance of an elite high school could provide access to more precise information on labor market opportunities and the returns to higher educational attainments, either through interactions with peers from more priviledged backgrounds or through informational activities at school. Several studies find that youth from disadvantaged have imperfect information and tend to underestimate the returns to further schooling (e.g. see Jensen (2010) for a sample of school-age children in the Dominican Republic). These effects would be expected to be larger for students from more disadvantaged backgrounds who have been argued to lack such information due to the lack of representativity of their reference groups. So, if attendance of an elite schools helps access better information and correct over-pessimistic beliefs, one should observe higher effects on the earnings expectations of those more disadvantaged students. However, there are also studies finding that students can, in other contexts, be over-optimistic about those returns (e.g. Zafar (2009, 2011) for American college students), which would make the later prediction less obvious. Independently from the direction of the bias, access to better information on job market opportunities would likely make earnings expectations less noisy, so that those expectations could be less dispersed among elite schools students.

Third, in line with a signaling model, the higher reputation of IPN schools, and the diploma they award, might grant access to better-paid jobs. In general it could also grant easier access to a better or more selective higher education institution, but this effect is likely to be limited in this setting, as IPN elite high school graduates must, as other applicants, take competitive exams to gain admission into the most prestigious public universities. If signaling is driving the effects of attending an elite high school on earnings expectations, one would anticipate to see students from those elite schools expect higher earnings with either a college or a high school diploma. The effect on the returns from a college education is again less obvious: it could be positive if the high school and college diploma are complements in producing signaling, but negative if the signal attached to the attended high school matters less for college graduates.

We will return to these predictions in the empirical analysis below, which will provide some insights on the relevant theoretical mechanisms at play.

3 Earnings expectations

We now turn to describing earnings expectations data and their differentials by family background and attended schools.

3.1 Data

Our data on students' expectations of earnings comes from a survey answered by a sample of students in the last year of high school when taking the national Enlace achievement test for the evaluation of high schools. The sample is representative of Enlace-test takers of Mexico City, and the survey gathers information on students' backgrounds and schooling experience. (We provide more details on this Enlace survey below.)

Two questions elicit the information about the earnings expected with given educational attainments. Youth are asked questions about expected future earnings under two scenarios: that they terminate their schooling after completing high school and that they continue their studies and obtain a university degree. The answers are given using a pre-codified set of brackets. The questions are the following:

1. *"If you do not obtain a university degree, what monthly income do you expect to have on average in 5 years from now?"*

2. *"If you obtain a university degree, what monthly income do you expect to have on average in 5 years from now?"*

The earnings brackets for both questions are: i) \$4,000 or less; ii) \$4,001 to \$7,000; iii) \$7,001 to \$10,000; iv) \$10,001 to \$15,000; v) 15,001 to \$20,000; and vi) more than \$20,000.[10]

These questions are in line with those used in the growing literature on the measurement of subjective expectations (see Attanasio, 2009; Delavande, Giné, and McKenzie, 2011 for surveys). However, the collection and analysis of data on expected earnings is recent and rises several concerns.

In the context of developing countries, a typical concern is the capacity of a population with low levels of literacy and numeracy to answer subjective expectations questions in a meaningful way. This should not be a problem in our case as we are dealing with individuals who have reached the final year

[10]In 2008, 7.5 Mexican pesos (\$) were equivalent to 1 US dollar in terms of purchasing power parity (OECD).

of senior-secondary education. Moreover, in another study of high school students in Mexico, Attanasio and Kaufmann (2009) use more complex questions, that demand a basic understanding of probability concepts, to elicit subjective expectations of earnings, and verify that the obtained answers are broadly consistent with observed wage data and with the expected variation in the expectations across individuals' observed characteristics. One may still wonder from what age youth are able to form expectations of their likely future wages, and whether those expectations are sufficiently informed. Given that 18 year-olds can consider entering the labor market, they should have already gathered information on earnings. Again, Attanasio and Kaufmann (2009) also use a sample of that age and verify the consistency in the information they provide. Thus, there should not be major obstacles to eliciting subjective expectations of earnings among 18 year-olds of Mexico FD.

The subjective expectations questions used in the Enlace survey have limitations, though, that must be addressed. In specific, the survey omits the probability questions necessary to elicit the full distribution of the individual's expected future earnings, collects the answers on expected earnings using categorical ordinal variables, and asks about the expectations of future earnings at a younger age of reference than other studies.

The Enlace survey asks about the "average" expected earnings, so that we can only obtain for each individual a point estimate of his expected earnings or college premium. The evidence emerging from the literature is that the elicitation of probability distribution of future variables is preferred to the elicitation of point estimates, as the former allow for a more precise estimation of the subjective mean. Delavande, Giné, and McKenzie (2011) argue that it is difficult to interpret if the responses to "what do you expect?" questions refer to the mean, median or another measure of the center of the expectations probability distribution. Using a survey of Sri Lankan micro enterprises, they compare a simple expectation measure, obtained through "what do you expect?" questions, to different central tendency measures obtained after eliciting the full probability distribution of expectations by asking about probabilities of several bins. They find that the simple expectation measure tends to be closer to the median than to the mean, and predicts future outcomes with less precision than the mean from the full probability distribution. They attribute this lower prediction power notably to the larger sensitivity of the simple expectation measure to the presence of outliers, a concern that the use of wage brackets should lessen in our data.

The use of an ordinal scale to input the answers demands additional assumptions to construct an implied college premium. Some rule is necessary to translate the qualitative information into a quantitative measure. We return to this below.

Finally, while we obtain expectations of earnings five years from the moment of the survey, that is for a future date when individuals will be on average about 23 years old, precedent studies, such as Dominitz and Manski (1996), query for expected earnings at ages 30 and 40, and Attanasio and Kaufmann (2009) at age 25. As the typical duration of college education in Mexico is four to five years, there is almost no space for gaining labor experience after graduation in the question for expected earnings conditional on university attainment. If the earnings of college graduates grow at a higher rate with experience, the elicited college premium might underestimate the subjective expectations of the long-term returns to university education.

3.2 Expectations and wage data

For checking their validity, a description of the main patterns of our measures of expected earnings is thus required. Figure 1 gives the distributions of responses for both questions in the whole sample of respondents in Mexico City metropolitan area public schools. As expected, the distribution of expected earnings is more right-skewed in the university attainment scenario than in the scenario with only senior-secondary education.

Although there is no reason for the distribution of students' expected earnings in five years and the earnings of current workers to be strictly similar - as they concern earnings of different cohorts, at different dates and with different selection patterns - the comparison of the two distributions is a useful device to gain insight about the external validity of the data (Attanasio, 2009). After all, current students likely obtain information about their potential earnings in part from the observation of earnings obtained by current workers.

We use the Mexican National Employment Survey (ENOE) from the second quarter of 2008 as information on earnings of current workers. Figure 2 shows the distributions of earnings by schooling level, conditional on employment, of individuals 23 to 30 years-old who live in the Federal District or the State of Mexico. Mere eyeballing suggests that students' earnings expectations tend to be over-optimistic when compared to earnings data. This is particularly true in the college graduation scenario.

One can also compare the expected college premia implied from the expectations data and observed in the actual earnings data. To construct the subjective expected college premium, we assume that each discrete earnings category corresponds to the mean of the two values that define each bracket. For the first and last bracket, which do not have an obvious interval, we assume that the brackets are, respectively, [$3,000 - $4,000] and [$20,000 - $27,000]. Then, the implied college premium, or expected returns to college, is given by the difference between expected log college earnings and expected log senior high school earnings: $expected\ returns = log\ \ (expected\ earnings \mid college) - log\ \ (expected\ earnings \mid high\ school)$.

Table 2 shows the summary statistics for the implied college premium derived from the expectations data and the college premium observed in the earnings data. The former is almost twice larger than the later in a rough comparison (see columns 3 and 6). However, the difference between the two estimates is notably smaller when we control for experience and gender in the earnings data. Moreover, the magnitude of the two estimates gets closer when we use hourly earnings to estimate the college premium in the earnings data. The difference between the two estimates of returns is similar in sign and magnitude to the one found in Attanasio and Kaufmann (2009), who use census information to validate a college premium elicited from the full distribution of subjective expected earnings from a survey on 15 to 25 year old beneficiaries of the program Oportunidades in urban areas.

Thus, we do not find evidence that youth tend to under-estimate the benefits from higher education. At the contrary, their reported expected earnings benefits from college attainment seem high when compared to earnings data for current workers with college education in the region. Students seem also over-optimistic with respect to future earnings with a high school diploma and, hence, the college premium elicited from the subjective expectations is only slightly larger than the one estimated, as a reference, using earnings data.

3.3 Differentials across family backgrounds

Turning to internal validity, Table 3 reports the partial correlations of earnings expectations and implied college premium with a set of covariates (using a specification in logarithms).

Individuals with larger endowments of cognitive skills at the time of entry into high school, measured by scores at the Comipems entry exam, have on average larger expectations of earnings benefits from college education (by 6% for a 1 sd increase in t-score). Junior high school GPA are also associated with higher expectations of returns to college but not with earnings (the conditioning on the entry score makes the interpretation of these parameters difficult though).

Even when controlling for achievements at entry, more favorable family resources are associated with higher expectations of earnings with college and corresponding premium. Thus, youth with a senior high school graduate parent expect higher earnings with college (by 4%) and college premium (2%), youth with a white-collar parent also expect higher college premium (2%), youth from more wealthy families expect higher earnings with college (this variable is not normalized), and the effects of parental occupation on earnings expected with college and of wealth on the expected premium are positive but not statistically significant. At the opposite, youth from indigenous origin tend to expect lower earnings with either high school (2%) or college (4%), and a lower college premium (2%, not statistically significant at a 5% level). Similarly, attendance of a private junior secondary school is associated with higher expectations of earnings with high school (4%) and college (8%) and of college premiums (3%).

Female students, while they could potentially anticipate some discrimination on the labor market, tend to have higher expectations of earnings with college (3%) and of the corresponding premium (8%). This pattern suggests that females might consider that a college diploma should decrease the discrimination they would face in the labor market.

Overall, the differentials by family background confirm the expected positive effects of better backgrounds (due either to complementarities between family resources and educational investments or to access to different information) and provide evidence of the internal validity of our measures of earnings expectations. There is also evidence of gender differentials.

3.4 Differentials across attended systems of high schools

We now turn to differentials across attended schools. Figures 3a and 3b plot the conditional means of expected earnings, given learning achievements, with respectively a high school degree and a college diploma separately for the two samples of students of IPN and other (non-elite school) students, and Figure 3c plots the corresponding conditional means of expected earnings returns from attending college (in log).

Several patterns emerge. The first striking result, in Figure 3a, is that, even at similar levels of learning achievements at high school completion, IPN students tend to have higher average expectations of earnings with a college diploma than non-elite school students. The conditional means are taken for each decile of the distribution of achievements. The differentials are clear for all 5 upper deciles, and earnings expections of IPN students are 10 to 20% higher than those of non-IPN students. The lower and upper bounds of 95% confidence intervals for expected earnings are shown on the graphs, and the differentials are statistically significant at least from the median. We also observe that IPN students in the left tail of the distribution of achievements have higher expectations, but there are few IPN students with low achievements so that the corresponding conditional mean expectations are very imprecisely estimated. In addition, earnings expectations tend to increase slightly with scholastic achievement. Second, as shown in Figure 3b, IPN students tend to have only sligthly higher expectations of earnings with a high-school degree than non-IPN students, with a difference of 5 to 10%, and the 95% confidence intervals do not reject that those expectations are equal; thus much smaller or no differences in earnings expected if leaving school after obtention of a high school degree are observed. In consequence of the two previous results, as shown in Figure 3c (for expected college premium computed as the difference of the logs of earnings with college and high school), IPN students, again after conditionning on achievements, tend to expect higher returns of attending college than non-IPN student, with a difference of 5 to 15% marginally significant in the right part of the distribution of achievements. Besides, while the slopes of the conditional expectation functions depicted in Figure 3b for earnings with college follow an upward trend, the ones in Figure 3a for earnings with high school are flat, indicating that students believe that the return to scholastic achievement in the labor market for high school graduates is small or insignificant.

Figure 4 shows similar graphs for the dispersion, measured by the coefficient of variation (CV) of the earnings expectations of students in IPN and non-IPN schools conditional on learning achievements (for each decile). Figure 4a shows that expectations of earnings with college tend to be less dispersed among students of IPN schools, with a value of the CV lower by about .005 at all deciles except the fourth. Earnings expectations tend to be also less dispersed at higher levels of achievements. Figures 4b and 4c show smaller gaps in the dispersion of expected earnings with high school and college premium, although the dispersion of expected college premium is lower among IPN students at most deciles.

Table 4 reports OLS regression estimates, similar to those in Table 3, to test whether the differentials, between students of different school systems, in conditional mean expectations given achievements persist after controlling for family and individual observable characteristics. The outcomes are again logarithms of expected earnings and returns to college. The gaps in expectations between IPN and non-elite school schools students remain with the controls, and the estimates indicate that IPN students tend expect higher earnings both with college (by 8%) and with high school (by 2%) diploma, and also higher returns to college (by 6%).

Overall, these descriptive results indicate that students of IPN schools expect to achieve higher earnings if attending college than students of non-elite high schools at similar levels of scholastic achievements, and also that IPN students have less noisy expectations than their non-IPN counterparts. This striking pattern suggests that the attendance of an IPN-school does affect expectations of earnings with higher educational attainments, and that some of these effects occur independently from learning achievements.

However, these patterns could also be driven by the selection of students with higher expectations into elite high schools. For instance, due to unobserved family background characteristics (e.g. family environments inductive of a higher motivation of youth, or providing inputs complementary to higher education diploma for obtaining higher earnings) correlated with the likelihood to attend IPN schools. To address these concerns and provide rigorous evidence on the causal effects of elite schools, the next sections use the local natural experiments generated by the process of assignment of students into schools for identifying the causal effects of admission into an elite IPN high school.

4 Data

4.1 The Comipems high schools application system

The Metropolitan Commission of Public Senior-secondary Educaton Public Institutions (Comipems) is the consortium of nine of the ten systems of public high schools of the Mexico City metropolitan area.[11]

[12] Since 1996, Comipems manages applications and admissions to public high schools of the area through a single process based on a common exam. This process selects and allocates students into schools based on the orientation choices they have expressed in advance, their scores at the exam, and the numbers of slots available in each school. It is based on the student-proposing deferred acceptance algorithm (Pathak, 2011).

More specifically, it operates in the following way. First, after registering to the process, but before

[11] A recent system of high schools administered by the FD government and targeted to low achieving studentf does not belong to Comipems.

[12] The Metropolitan area includes the Federal District and 22 municipalities from the neighbouring State of Mexico.

taking the exam, applicants formulate and rank their preferred orientation choices. They can select up to 20 options. As some schools offer more than one track, students actually submit a list of preferred tracks. (We use the term school as synonymous of track, though, because most of the schools of our interest, those with selective admission and notably IPN schools, have only one track at the time of admission.)

Second, all applicants simultaneously take a common competitive exam. All applicants with at least 31 correct answers out of 128 questions in the exam are allowed to register in a Comipems school, but not all gain admission into their most preferred option.[13]

In a third step, the applications of each candidate are considered in the order of their ranking at the exam, and the seats at the most demanded schools are given to the students who achieved the highest grades. As schools fill in, students' next orientation choices are considered. In 2005, only one third of applicants were granted admission into their first-choice school. In practice, the Mexican Center of Evaluation of Education (Ceneval), the institution in charge of the assignment process, uses a computer program that ranks students according to their exam score and proceeds to allocate individuals into their preferred choices, following the ranking they submit, until predetermined school capacity constraints bind.

Finally, students who only chose schools which happened to be too selective with respect to their test scores, i.e. who miss the cutoffs for all their listed orientation choices, are allowed to register in the schools with remaining slots in a second-stage allocation process. In 2005, slightly less than 10% of the admitees were granted a seat at this second stage.

Note that, in this system, schools are free to determine the number of seats they open, but afterwards simply apply the admission list produced by Ceneval based on students stated choices and exam ranks. Hence, no complex choice strategies would help students increase their probabilities of being admitted into their most preferred choices, so that they have an incentive to list their chosen schools in a way that is consistent with their true preferences (Pathak (2011)).

The combination of the institutional setting and students' stated preferences produces a set of admission thresholds which are determined ex-post and cannot be predicted beforehand. As students tend to prefer schools providing better learning environments, schools with higher admission thresholds offer on average better school inputs. As we discuss below, these admission thresholds generate some variation in the allocation of students into schools that can be considered as locally random and exploited for identifying the causal effects of admission to more selective high schools.

[13] Applicants who list a school from the UNAM system must take an exam version designed by this institution, while all the other students take an exam design designed by Ceneval. Both exams are designed to be equivalent in level of difficulty. We do not have information to suppose that some students might prefer taking one version of the test to strategically increase the probability of gaining admission into one of their most preferred choices.

4.2 Datasets

The data we use matches information from the 2005 Comipems admission process to information both from the 2008 and 2009 National achievement test (Enlace) of third graders of all high schools and from the 2008 and 2009 versions of a questionnaire survey of a representative sample of students who took the same Enlace test. It thus forms a panel dataset in which students are followed from application in 2005 up to graduation from high schools in 2008 or 2009 (unless they drop out before - we document this in depth below). We also match to this panel of students information on the schools they attend from the 2005 version of the annual census of senior secondary schools carried out by the Secretary of Education (called "Formato 911").

The main outcomes of interest are measured in 2008 and 2009, as the cohort of 2005 Comipems applicants graduate from high schools. The Mexican Evaluation of Scholastic Achievement of Educational Institutions (Enlace) is a national standardized exam taken, since 2008, by students in the last year of senior- secondary education (12th grade).[14] The purpose of the examination is to evaluate schools, and the educational system as a whole, and, hence, it has no bearing for students on graduation or university admissions. However, Enlace results are widely reported by media outlets and non-governmental organizations in Mexico, and are used as the principal input for the creation of school league tables. The publicity provides school agents with incentives to perform better and makes Enlace a medium-stake test.[15]

In parallel to the Enlace exam and at the same date, the Federal Ministry of Education conducts a complementary survey among a random sample of students, gathering information on their individual characteristics, family background, and schooling experience.

All four datasets (for Comipems 2005, Enlace 2008 and 2009 tests and surveys, and 2005 census of schools) can be matched at the individual level, and this was done by the Ministry of Education ('Secretaria de Educacion Publica') data administration teams. The Comipems and Enlace datasets are matched using, in this order, the national population identification code (CURP), combinations of CURP and name when the former is incomplete, and the name and birth date when there are missing values for the CURP. The matching between the Enlace exam and the survey results is straightforward using the exam identification code (available for all observations in both datasets). The school census information is recovered for the specific schools students attend using school identifiers. We return to sample sizes below.

As for the information provided by those different datasets, the Comipems database includes the submitted ordered listing of orientation choices (tracks and schools), the score at the entry exam, and the assignment outcome for all applicants, and also contains some family background information from a questionnaire attached to the registration form. The Enlace dataset contains the exam scores for all

[14]Enlace exams are also taken by students at other grades.
[15]students enrolled in UNAM schools do not take Enlace and, hence, are not considered in our analysis.

3rd graders. The scores for Mathematics and Spanish language sections are reported, and we aggregate the two scores in one variable which is normalized by exam cohort with mean 0 and standard deviation 1 for the sample of matched 2005 comipems applicants. The additional survey conducted at Enlace exam taking provides information on students' background and experience in high school, expected wages conditional on schooling levels, and aspirations to pursue further education. Finally, the school census data provides information on the characteristics of schools, in particular: class size, teachers' profile and information technology infrastructure.

4.3 Samples

The Comipems process concerns all applicants to public senior-secondary schools in the Mexico City Metropolitan area. Excluding students admitted to UNAM schools, 220,659 eligible applicants took the Comipems exam in 2005 (another 24,999 took the exam, but did not complete their junior-secondary education on time). 195,802 (88.7%) applicants were allocated to 620 tracks. 2.3% of applicants did not achieve the minimum required score in the exam to access senior secondary education and 9.0% were not admitted to any of their listed options and did not use the second-stage process to select a school with remaining seats. The pool of applicants includes students currently enrolled at junior-secondary schools in the Federal District (34.0%) and in the participating localities from the State of Mexico (39.9%), but also youth who graduated in previous years (23.7%), who attend schools in other localities (1.7%) and students of special adult-education institutions (0.6%).

For estimating the effects of IPN admission, we focus on a restricted sample of 2005 Comipems applicants a) who complete at the end of the academic year 2004-2005 in a junior high school of the Federal District and b) who applied to at least one IPN school and were either admitted or rejected from such school.

We exclude returning-to-school and special adult-education students because they have lower chances of completing senior-secondary schooling and likely different responses to IPN admission. Similarly, as 15 of the 16 IPN high schools are located within the Federal District, we omit students from the neighbouring State of Mexico. These restrictions reduce the sample to 75,137 applicants.

The restriction to applicants of IPN schools is required because our identification strategy compares the outcomes of admitted and rejected IPN applicants and it reduces the sample to 28,378 applicants at the Comipems exam (baseline).

Only a share of those applicants (47.7% of all Comipems applicants and 51.4% of those allocated to a Comipems school in 2005) can be matched to the Enlace test in 2008 or 2009 (more on attrition below). This leads to samples of 13,538 students at Enlace test-taking, and 8,845 students (31.2%) at the Enlace complementary survey conducted among a random sub-sample of the test-takers. Note that some of the applicants can change school afterward, so that while 90.5% complete a public high school, 9.5% of them are observed in a private high school when taking the Enlace test.[16]

In practice, as we explain below, our estimators only use observations "close" to the IPN admission cutoffs.

4.4 Attrition

Dropping out from school (with a prominent role), imperfect matching, and Enlace turnout rates are the suspects behind the attrition between the end of junior and senior secondary school in our sample.

First, dropping out from senior high school is a wide spread phenomena in Mexico. In 2010, only 45% of 25-34 year-old population in the country had completed a senior-secondary education. The same year, the completion rate among all Mexican students admitted in high school three years before was 52% (OECD, 2011). As dropping out is correlated with family background, completion rates should be higher among students in Mexico City and notably IPN students. According to administrative statistics, the ratio of graduates in 2007 (whatever the year of admission) to new entrants in 2004 was 62% in IPN high schools. Among our sample of 2005 Comipems applicants, 56.9% of the 7,067 students admitted to an IPN school are observed in the Enlace dataset at the end of high school (52.3% in 2008 and 4.6% in 2009), and, respectively, 44.6% of the 21,311 students marginally rejected from the IPN system (38.1% in 2008 and 6.6% in 2009). This suggests that dropping out explains the bulk of the attrition between entry and completion of high school and our 2005 and 2008/09 data. However, notably for students allocated to non-IPN schools, some other factors could have contributed to the observed attrition.

Second, we estimate that matching errors (due to incorrect or imprecise information on CURP, name and/or birth date) and Enlace exam turnout could have contributed to overall attrition in our sample by up to 6 points. In the 2008 Enlace-exam database, there are 13,453 students observed in Comipems schools who are not matched in the Comipems data. If all of them were eligible applicants from the 2005 Comipems process, matching errors would explain 6 points of attrition. However, some of those observations should correspond to students who did not take the Comipems exam in 2005, e.g. students admitted in 2004 or before or who migrated into Mexico City after completing the first year of senior-secondary education in another region (those school changes are possible).

Third, some 3rd grade students, while still enrolled, could have not shown up at the Enlace exam. Exam participation might be affected by random absenteeism or, more worrisome, some principals, looking to achieve a higher mean Enlace score in their school, could have encouraged, or at least not detered, lower performing students to skip school the day of the exam. If we approximate the Enlace exam turnout rate by the ratio of students who took the Enlace exam in Comipems schools in 2008 over the number of students enrolled in grade 12th reported in the school census at the beginning of the 2007-2008 academic year, the average Enlace turnout rate is estimated at 90%. Although drop-outs during the year could lead

[16] Private schools do not belong to Comipems and manage admission decisions independently.

to overestimate slightly the Enlace turnout, this suggests that incomplete exam turnout could explain up to 5 points of attrition.

To summarize, those back of the envelop calculations suggest that the attrition rate of 49.9% between the beginning and end of senior secondary education in our sample can be roughly decomposed in the following way: at least 38% due to school drop-outs, up to 6% to matching errors, up to 5% to incomplete Enlace turnout, and 3% due to other factors (such as more than one grade repetition). Beyond the overall composition of attrition, we are interested in verifying that there is no relationship between the treatment of interest, i.e. admission to an IPN school, and attrition, measured by Enlace exam taking. Before that, we need to present our identification strategy.

The sample used for these estimations consists in all students who were either admitted or rejected from an IPN school, and contains 28,378 observations, before the bandwidths restrictions.

5 Empirical strategy

Our empirical strategy is based on a Regression Discontinuity (RD) design, and provides estimates of the effects of admission to a school of the IPN system.

5.1 Allocation process and discontinuities in admission to an IPN school

To explicit it, consider a student i who was admitted to an IPN school $S^{IPN} \in \Theta^{IPN}$, where Θ^{IPN} is the set of IPN schools (call him a treated student). Assume this student expressed a set of school choices $C = \{S^1, ..., S^K = S^{IPN}, ..., S^L = S^{OTH}, ..., S^{20}\}$, with school S^{IPN} ranked as Kth choice, and obtained a score s at the Comipems entry exam. A set of admission cutoffs corresponding to this set of choices $\{c^1, ..., c^K, ..., c^L, ..., c^{20}\}$ was generated by the admission process. Note that the order of student's choices can differ from the schools selectivity order, so that some less selective schools can be preferred to more selective ones (student i's preferences for closer schools, or other school features, might generate this type of ranking). Then, if student i is admitted to school S^{IPN}, we must have:

$$\begin{aligned} s &< c_k, \quad k = 1, ..., K-1 \\ s &\geq c_K \end{aligned}$$

where c_K is the admission cutoff score of IPN school S^{IPN}.

Now, there are other students who stated identical choices C but achieved lower scores and were thus not admitted to school S. Consider one such student, denoted j, who also expressed choices C but achieved a lower Comipems score $s' < s$, and was not accepted to school S^{IPN} but to a school $S^{OTH} \notin \Theta^{IPN}$ which happens not to be an IPN school and was ranked further, at Lth rank, in the choice set (call him a comparison student). We have:

$$\begin{aligned} s' &< c_k, \quad k = 1, ..., K-1 \\ s' &< c_K \\ s' &< c_k, \quad k = K+1, ..., L-1 \\ s' &\geq c_L \end{aligned}$$

where c_L is the admission cutoff score of non-IPN school S^{OTH}.

If the difference in scores between the two students is not large, student j can provide a counterfactual to student i for measuring the effects of admission in a school of the IPN system, as both expressed the same choices, but student j achieved a slightly lower score and was thus rejected from the IPN school S^{IPN} and admitted to the non-IPN school S^{OTH}.

Indeed, for any IPN admittee with school choices C and a given exam score s, one can define a cutoff score below which a student with similar stated choices would be allocated to a non-IPN school.[17] This cutoff score is obtained by decreasing progressively the score starting from s, and considering the next choices. For this, if i was admitted to school S^{IPN} ranked Kth among its stated choices, define $C_{ls}(C, s) = \{S'_{K+1}, ..., S'_{K_{max}}\} \subset C$ as the subset of choices corresponding to schools which are ranked further (not preferred) and less selective than school S^{IPN} (the more selective ones are not accessible given the student's score s). Some of those can be IPN schools, but in general students will have put less selective non-IPN schools among those next choices. Two cases are possible. If the first of those choices S'_{K+1} corresponds to a non-IPN school, then the cutoff for admission in any IPN school is school S^{IPN} admission cutoff c_K: had he achieved a score below c_K, student i would have been admitted to non-IPN school S'_{K+1}. Now, if the next choices, begining from S'_{K+1}, are also IPN schools, consider all choices $C_{ls,IPN}(C, s) = \{S'_{K+1}, ..., S'_{K_{max}}\} \subset C_{ls} \cap \Theta^{IPN}$ before the first non-IPN school $S'_M \notin \Theta^{IPN}$; then the student would be admitted to one of the IPN schools $S'_{K+1}, ..., S'_{M-1}$ as long as his score is higher than the cutoff of the less selective in this set. So the relevant cutoff for IPN system admission is the admission cutoff of the less selective IPN schools which appear in the choice set before the first accessible non-IPN school. Thus, one can define for each IPN admittee with stated choices C and exam score s (admitted to school $S^{IPN}(C, s)$), an IPN admission cutoff as:

$$c^{down}(C, S^{IPN}(C, s)) = Min \quad \{c_k(S^k), \quad S^k \in C_{ls,IPN}(C, s)\}$$

Similarly, for any student with school choices C (including a preferred IPN school) and a given exam score s, who was not admitted to an IPN school S^{OTH} (or $S^{OTH}(C, s)$) with admission cutoff c_L, one

[17] In practice, we don't require treatment and comparison students to have stated the exact same choices, but can compare students with choices that are sufficiently similar to lead to the same IPN admission cutoffs.

can define a cutoff score above which the student would be admitted to an IPN school. This cutoff is obtained by increasing his score starting from s and considering the options stated before in the choice set, and corresponds to the admisson cutoff to the least selective of those preferred IPN school. Denoting $C_{ms,IPN}\left(C,s\right)$ the set of more selective IPN schools appearing before in the choice set, the IPN admission cutoff is thus now defined as:

$$c^{up}(C,S^{OTH}\left(C,s\right))=Min\ \ \left\{c_k\left(S^k\right),\ \ S^k\in C_{ms,IPN}\left(C,s\right)\right\}$$

And we can define an IPN admission cutoff, corresponding to a given score, for all students as:

$$\begin{aligned} c(C,s) &= c^{down}(C,s,S^k\left(C,s\right))\ \ if\ \ S^k\left(C,s\right)\in\Theta^{IPN} \\ &= c^{up}(C,s,S^k\left(C,s\right))\ \ if\ \ S^k\left(C,s\right)\notin\Theta^{IPN} \end{aligned}$$

Note that the IPN admission cutoffs defined above are local in the sense that they depend on stated choices and achieved scores.

Now, there will be students with the stated choices C sufficiently similar to lead to the same cutoffs and slightly different scores (in the neighborhood of a given score s) above and below the IPN admission cutoff, such that those with scores just above get admitted to an IPN school and the others not. Table 5 describes the obtained IPN admission cutoffs for the sample of IPN applicants and the schools to which students are allocated in case they score below the cutoffs. This provides a clear picture of the treatment and control. The IPN admission cutoffs are in average 75.5 (with a standard deviation of 7.37) and range from 66 to 99.

Thus, the process generates discontinuities in the relation between the students' score and their admission to an IPN school: given their choices and their scores being in a given neighborhood, the probability that a student is admitted to an IPN school will jump from 0 to 1 at the IPN admission cutoffs. Most importantly, the IPN admission cutoffs are determined by the entire admission process, including the choices and scores of all applicants, so that, although applicants can certainly influence their Comipems score through effort, they cannot know and/or precisely determine ex-ante the cutoffs and the relative position of their scores with respect to those; this would indeed require both total control over their own score and perfect knowledge and anticipation of other applicants' scores and school choices.

To verify this empirically, Figure 5 plots the distribution of the distance of students' scores from the IPN admission cutoffs. The cutoffs for IPN applicants vary with their stated choices and scores and we consider the distance to the different IPN admission cutoffs for the aggregate sample. No discontinuities are apparent at the cutoff (zero) in that distribution, which confirms local randomness: students who score close to the cutoffs are unable to precisely predict those and manipulate their scores and position to the cutoffs.

For another empirical test, Figure 6 plots the means of a series of observable characteristics of IPN applicants by distance to the IPN admission cutoff; none of those graphs exhibits any clear discontinuity: students scoring just below and above the IPN admission cutoffs have similar characteristics (we check this using estimation techniques below). This provides complementary evidence that admission is locally random and that IPN applicants are unable to influence the allocation process and self-select into treatment.

Although, at the IPN admission cutoffs, the probability that a student is admitted to an IPN school will jump from 0 to 1 (students are admitted to one and only one school), the probability of actually attending an IPN school will also be discontinuous at the IPN admision cutoffs but may not be complete above the cutoff for two reasons. First, some students may drop out before or during senior-secondary schooling. Selective attrition correlated with IPN admission would be particularly problematic. It would reflect in a discontinuity, at the admission cutoff, in the distribution of the Comipems score among students completing senior-secondary education. Figure 7 plots the distribution of the distance of students' scores from the IPN admission cutoffs for the sample of completers: the distribution is very similar to the one at baseline (in Figure 5) indicating no differential attrition in our sample.

Second, some students can decide not to go to the public school they were allocated to through the Comipems system and attend some private school, and there are also a small number of students who change school after the first or second year of senior high school. For documenting this, Figure 8 plots the probability of taking Enlace in an IPN school, conditional on taking the exam in 2008 or 2009, by distance to the IPN admission cutoff. The probability jumps from zero to about .95 at the cutoff. Thus, although some will drop-out (see below), the great majority of IPN admittees who complete senior-secondary schooling do attend an IPN school up until completion.

Students who scored above the cutoffs and were admitted to an IPN high school will then be exposed to a different school environment and will benefit from different school inputs than students who were rejected. Figure 9 plots the averages of a series of school characteristics of by distance to the cutoff and shows that students scoring just below and above the cutoffs will experience rather different schooling environments during the following high school years. In particular, they tend to be taught by more qualified and experienced teachers, in smaller class sizes, and have access to more computers, and probably other school inputs.

5.2 RD estimates

The discontinuities in allocations to school systems described above make it possible to identify the effects of admission to a school of the IPN system using a RD design strategy. The variations in treatment near the cutoffs are as good as random because individuals are unable to precisely manipulate their scores or allocation, and thus can serve to unveil a causal effect (Lee and Lemieux, 2010; Imbens and Lemieux, 2007). As applicants cannot locally self-select into the IPN system, the expected potential outcomes if they attended an IPN school or not (i.e. with and without treatment) as a function of exam scores will be continuous at the cutoff, and a discontinuity in the outcomes can be attributed to the admission to different school systems. For each subsample of students with a similar IPN admission cutoff, the effects of admission into the IPN system will then be given by the discontinuities in later outcomes at the cutoff. The discontinuities thus provide a set of local experiments which allow to identify the effects of IPN admission at different admission scores.

As students with scores above the cutoffs are admitted with certainty to an IPN school, the setting is one of sharp regression discontinuity. We can thus estimate the effects of IPN admission at any cutoff using local linear regressions. In practice, for a subsample of students with the same IPN admission cutoff c (as defined above), we restrict the sample to observations with scores in an interval $[c-h; c+h]]$ around the cutoff and estimate the average effect of IPN admission on an outcome Y_i using a single regression. Denoting $d_i = s_i - c$ the distance between a student score and the IPN admission cutoff (the forcing variable), and $W_i = 1\{s_i \geq c\} = 1\{d_i \geq 0\}$ an indicator for his admission to the IPN system, the model to be estimated writes:

$$Y_i = \alpha + \tau W_i + \beta d_i + \gamma d_i.W_i + \epsilon_i \tag{1}$$

where βd_i and $\beta d_i + \gamma d_i.W_i$ are distinct linear control functions for the slopes of the relationship between scores and outcome Y_i on the left and right-hand sides of the admission cutoff c . The discontinuity in the outcome is obtained by extrapolating those slopes at the cutoff and taking the difference between the left and right-hand extrapolations. The parameter τ captures this discontinuity, and OLS estimates of (1) are consistent for this parameter.

Now, the sample consists in students who applied and were admitted or rejected to different IPN schools, so that there are multiple IPN admission cutoffs c, and our primary parameter of interest is the average effect of IPN admission over the different cutoffs. For estimating this parameter, we aggregate all students with different IPN admission cutoffs into a single sample, incorporate cutoff fixed-effects in (1), and estimate the model:

$$Y_{ic} = \alpha + \tau W_i + \delta_c + \sum_c 1\{c\}.\left[\boldsymbol{\beta_c} d_i + \boldsymbol{\gamma_c} d_i.W_i\right] + \epsilon_i \tag{2}$$

where δ_c are fixed effects for the cutoffs relevant for each student, $1\{c\}$ is an indicator that cutoff c is relevant for student i, and $\boldsymbol{\beta_c} d_i$ and$\boldsymbol{\beta} d_{ci} + \boldsymbol{\gamma_c} d_i.W_i$ are now control functions which are specific to each different cutoff. Identification remains within groups of students with the same cutoffs, and the parameter τ now captures the average discontinuity in the outcomes at the cutoffs and is again consistently estimated by OLS.[18]

[18]For checking the robustness of the estimates, we also use a specification with controls for the stated school choices of students. In practice, we specify $\Phi_i = ipn_i + hdemand_i + \sum_{r=1}^{10} \phi_{ir}$, where ipn_i and $hdemand_i$ are, respectively, the share

The choice of the bandwidth derives from a tradeoff between bias and precision: larger bandwidths will increase precision by using more data but may induce more bias by relying more on the linear extrapolation on the two sides of the cutoff. The Imbens-Kalyanaraman (IK) optimal bandwidth (Imbens and Kalyanaraman, 2011) is a solution of this tradeoff based on minimizing the mean squared error; in our data, the IK bandwidth is about 2 and we consider this as our benchmark. We also use two larger bandwidths of 5 and 10 that allow higher precision. For the larger bandwidth of 10, in order to reduce extrapolation error, we replace the linear control function by a quadratic one. In addition, when estimating (2), we cluster the standard errors at the cutoff level to allow for common unobserved shocks at this level.

For those estimates, the sample of 28,378 students who applied and were either admitted or rejected from an IPN school is reduced by the bandwidths around the IPN admission cutoffs. For the three bandwidths of 2, 5 and 10 points, we have samples of respectively 1,396, 3,391 and 7,115 observations, and of 8,845, 481, 1,182 and 2,454 observations when restricting to students taking the Enlace survey.

Several points should be noted on those estimates. First, the estimated parameter τ captures the local effects of IPN admission for marginal IPN admitees. It is the relevant estimator for the effect of a policy change that would consider a marginal increase in the number of available slots in the IPN elite school system, with those slots distributed across IPN schools as the existing ones. Although it captures the effect of admission (students cannot be offered to choose between two Comipems schools after the allocation process), it has an intent-to-treat interpretation because students can opt for a private school, study in another region or leave the education system.

Second, the important feature of our strategy is that we identify the effects of admission to the IPN system rather than to a specific IPN school. We do this by considering the cutoff for admission to the IPN system, which can differ from the cutoff for admission to the specific school to which admittees were allocated, and a control group of students who were not admitted to an IPN school. Our estimates differ here from those proposed by several related studies, notably Abdulkadiroglu, Angrist, and Pathak (2011); Pop-Eleches and Urquiola (2011), who consider school-specific cutoffs and focus on the marginal effects of admission to a more selective school compared to admission to a less selective school (which can be another elite school). This makes the interpretation of the estimated effects more straigthforward: while other studies estimate the effects of admission to more selective elite schools, we estimate the effects of admssion to any school of the elite system.

of IPN and "high demand" schools that student i listed in her stated preferences, and ϕ_{ir} indicates whether student i selected a "high demand" school in position r of her submitted ranking. The Comipems application leaflet marks "high-demand" schools to indicate students that these are options where admission is highly competitive. The bulk of this category is formed by IPN and UNAM schools. The obtained estimates (available upon request from the authors) are very similar to those controls.

5.3 Validity of the RD design at assignment and differential attrition

We follow Lee and Lemieux (2010) to analyze the validity of the regression discontinuity design, in a process analog to assess whether the randomization was carried out properly in a randomized experiment. We have already verified in Figure 5 that there is no sign of discontinuities in the number of students with scores just above the IPN admission cutoffs, which confirms that individuals have no precise control over the assignment process.

We now formally test the local balance of baseline covariates across both sides of the IPN admission cutoffs by using a vector of baseline covariates as dependent variables in our main econometric specification. We expect the coefficients for treatment not to be different from zero if students are not able to sort above the IPN cutoffs. Table 6 gives the results of three sets of RD estimates, using the bandwidth of respectively 2, 5 and 10 in the three columns, for a set of baseline covariates (similar to the one used before) as dependent variables. We also test the joint significance of discontinuities in the full set of covariates using seemingly unrelated regressions (SUR). We do not find any evidence that any groups of students sort above the IPN admission threshold, and the results are consistent across specifications. With one exception, the admission coefficient is not statistically significant at the 10 percent level in either of the 12 regressions reported in each column. The Chi-square test for the discontinuity indicator being zero in all equations takes high p-values of .321, .655 and .845 using the three bandwidths. This is strong evidence in support of the locally random-like variation of assignment to treatment.

We discussed in section 4 the presence of some attrition in our sample, due mainly to dropping out from school, between the beginning and end of senior-secondary education. We thus need to investigate the robustness of the RD design to this attrition at Enlace survey-taking. Figure 7 already provided visual evidence rejecting selective attrition linked to admission into the IPN system. We formally test the presence of such attrition using RD estimates of the effect of IPN admission on the probabilities of taking the Enlace exam and answering the Enlace survey. Results are reported in Table 7. The magnitude of the point estimates is negative but small, of -0.0265 using the bandwidth of 2, and we cannot reject that IPN admission has zero effect on the probability of taking the Enlace exam.

Now, even with no effect of admission into the IPN system on the probability of taking the Enlace exam on average, some heterogenous effects across specific groups could remain possible. This would happen if, for example, IPN schools are relatively better at keeping in school some types of students, but are relatively worse at keeping other types. We thus test for the balance of baseline covariates conditional on taking the Enlace exam. Table 8 gives the results of RD estimates (again for the three bandwidths), together with the SUR joint-significance test, of differences at the cutoff in the same characteristics used to evaluate the balance of covariates at Comipems assignment. With the bandwidth of 2, none of the admission coefficients in the 12 equations estimated is statistically significant at the 10 percent level. With the bandwidth of 5, only one discontinuity, for junior high school gpa, is found statistically significant at the 5 percent level. And, with the bandwidth of 10, there are statistically significant discontinuities in the conditional expectations of junior gpa and attendance of a private junior HS. However, the SUR tests reject the joint significance for the three bandwidths with p-values of respectively .239, .633 and .131. Thus we fail to reject that all discontinuities are jointly equal to zero, and overall students below and above the admission thresholds are not systematically different at Enlace taking.

These checks confirm that discontinuities in outcomes at completion of high school can be causally attributed to the locally exogenous allocation of students, generated by the exam-based allocation process, to schools of the IPN system.

6 Results

We now turn to our main results for the effects of admission into an elite school system on subjective expectations of future earnings conditional on educational attainment and the financial returns to higher education. We also report the effects on learning achievements and the educational attainments students aspire to achieve. Those results provide some insights on the hypotheses formulated in section 2 regarding the formation of expectations, and notably suggest that elite schools enhance students' expected returns from college education either by improving some of their skills or providing specific information.

6.1 Expected college earnings and premium

Figure 10 plots the averages of the main outcomes we investigate, and notably expected earnings returns from a high school and college education and associated college premium. Students admitted to IPN elite schools exhibit consistently higher expectations of earnings with college and returns, with a marked discontinuities at the cutoff. Table 9 presents the RD design estimates of the effects of IPN admission on the implicit earnings premium of attending college derived from the expected earnings with a college and a high school education. We find that IPN admission does increase the expected financial returns to college education. Compared to average expected returns to college of about 75 percent below the cut-off, the point estimates indicate a large effect of about 17/19 percentage points, and are very robust to the bandwidth selection. The statistical significance diminishes as we restrict the number of observations in the sample to the bandwidth of 2, but the estimates with the two larger bandwidths are statistically significant at 1% and the point estimate remains unchanged with the smallest bandwidth, suggesting that bias is negligible in the larger bandwidths estimates.

Table 10 reports the estimates of the effects on the earnings expected with a college and a high school education. Admission into the IPN system increases expectations of earnings associated to a college education. The estimates of the three local linear regressions show large increases from 12 to 21 percent with the larger 21 point estimate for the bandwidth of 2. This effect is to be compared to average expected earnings of about 14,000$ (about 1,900 USD) below the cut-off. The estimates are statistically significant at the five percent level with the three bandwidths. In contrast, we do not find any effect of IPN admission on earnings expectations conditional on staying with a high school education. The magnitude of the point estimates is rather small and changes sign across specifications, from 2 percent with the bandwidth of 2 to -4/-5 percent with the larger ones, and the effects are never statistically significant effect at the 10 percent level.

Thus, our results indicate that IPN admission, and the subsequent attendance of an elite high school, increases the earnings expected with college education, which in turn increases the financial returns expected from attending college. However and rather surprisingly, admission to an elite school does not increase the earnings expected if entering the job market after completing senior-secondary schooling. This suggests that the effects on the expected college premium are not explained by the students' perception of a higher signaling value of the IPN system diploma on the labor market – or at least not on the labor market for high school graduates.

6.2 Scholastic achievements

For understanding the mechanisms through which elite schools affect students' earnings expectations, it is important to examine their effects on learning. Figure 10 again provides some visual evidence of a discontinuity at the cutoff, and the estimates in Table 11 show that admission into the IPN system also increases substantially students' learning achievement at the Enlace high school completion exam. Again, the local regressions show large effects for all three bandwidths, of .19 standard deviations with the bandwidth of 2 and .23-.24 with the larger bandwidths, and the later two are statistically significant at the 1 percent level. The positive effect of IPN admission on exam achievement indicates that this elite school system combines inputs in a way that is meaningful and enhances students' skills, at least scholastic ones.

Regarding the formation of earnings expectations, we outlined that, in a human capital model, we can expect that, assuming that students believe that earnings are positively associated with the marginal product of labor, an increase in scholastic achievement or, more generally in productivity, induces an increase in the subjective expectations of future earnings. The absence of an effect of IPN admission on earnings expected with a high school education seems at odds with this framework. In principle, if higher scholastic achievement translates into higher productivity, we should observe an increase in the earnings expected with both a college and a high school education. However, it is possible that students believe that the labor markets for college and high school graduates reward differently cognitive skills, and this explanation is consistent with the descriptive evidence presented in Figure 3 suggesting that students expect no gains from higher scholastic achievements if begining to work with a high school degree.

Besides, although we do not have any information on those outcomes, the attendance of an elite schools could also enhance some non-cognitive skills of students, such as future-orientedness or self-confidence. Again for those skills to intermediate the effects of IPN admission on earnings expectations, we would need the skills to be relevant only conditional on college attendance.

Thus, we cannot reject that the effects of attendance of elite schools on earnings expectations derive from the acquisition of additional skills, but students would then value those skills only conditional on getting a college education. This would occur if the skills obtained in elite high schools make college attendance more profitable either in terms of learning (complementarities in skills production) or in terms of earnings (complementarities in labor market productivity).

6.3 Aspirations regarding further education

Now, for getting a sense of the long term effects of the increase in earnings expectations associated with admission to an IPN elite school on students trajectories, one would ideally want to observe their later outcomes in college (if they attend it) and on the labor market. Our data does not allow tracking students after high school completion. However, the Enlace survey contains information on the educational attainments students aspire to reach.

The visual evidence in Figure 10 is again suggestive of significant effects. Table 12 gives the results of RD design estimates of the effects of IPN admission on those aspirations for college undergraduate (first three columns) and graduate (last three) degrees. Although the estimates with the smaller bandwidth of 2 shows a positive effect of 8 percentage points of the probability to wish to attain an undergraduate college education, the effect is not statistically significant and is not robust to the larger bandwidths (the point estimate decreases to 2 points), so that there is no strong evidence of an effect on aspirations for undergraduate college degrees. This may be explained by the already high share of students who aspire for a college education, about 90%, below the cut-off. However, there is evidence for some positive effects on the probability to wish to attain a graduate college education, with increases of 10 percentage points with the smaller bandwidth of 2 and about 17-18 points with the two larger bandwidths, the latest two being significant at the 1 percent level. This effect is to be compared to a share of about 60% aspiring to a graduate education. Thus, the effects of admission to an elite IPN school on learning achievements and expectations of the financial gains from a college education seem to translate into higher educational aspirations of students admitted to IPN schools, with them being more likely to aspire for graduate college degrees than their counterparts who were rejected from the IPN schools.

As indicated in the introduction, Attanasio and Kaufmann (2009), in an observational study among a similar sample of Mexican students, find that a one percentage point increase in the expected college premium (compared to a high school degree) is associated with a .2 to .4 percentage point higher probaility (higher for boys) of attending college. This is probably an overestimate, but interpreting this estimate causally would imply that the increase in earnings expectations we found with the attendance of an elite school could translate into a up to 4 to 8 points higher probablity of college attendance. This seems in the lower range of the magnitude but consistent with the effects we obtain on aspirations for a graduate college diploma. This would confirm that the gains in expectations from attending an elite school could explain an economically significant share of the differentials in college attendance.

Hence, although our estimates are of a reduced-form and could incorporate notably the effects of gains in learning achievements, these results suggest that those other benefits, in terms of expectations, of attending an elite school could be long-lasting.

6.4 Heterogenous effects by students' background

We also examined the heterogeneity in the effects of elite school attendance by students' characteristics. There is some evidence that attending an IPN school affects in somewhat different ways girls and boys. Table 13 gives RD design estimates for the samples of girls and boys considered separately, and shows that boys benefit more from attending an IPN school in terms both of learning achievements and expectations of returns to college attendance.

At the opposite, we do not find any evidence of heterogeneities by family background, and notably that students from more disadvantaged backgrounds exhibit larger increases in earnings expectations than other students when attending an elite school. For instance, Table 14 shows that the effects differ by parental education for none of the considered outcomes: youth with more educated parents benefit in terms of learning achievements, increase their earnings expectations and their aspirations for a graduate college degree as much as those with less educated parents. Similar results (available from the authors) are obtained using other parental background characteristics. This pattern suggests that an information channel, which would reflect in larger effects for students who had less access to information on labor market opportunities while in junior high school, is not key for explaining the gains in expectations.

6.5 Robustness checks

We examine the robustness of our results by adding a vector of baseline covariates as explanatory variables in our econometric model. As in the case of experimental designs with random assignment to treatment and control groups, the inclusion of baseline covariates should not affect the LATE estimates in the absence of selective attrition, the main concern here, or manipulation of treatment assignment (Lee and Lemieux, 2010).

We include controls for a vector of junior high school GPA deciles fixed effects, schooling in a private junior high school, gender, parental education (indicators for whether the father/mother has senior secondary education), and parental occupation (indicators for whether the father/mother has a white collar job). This specification aims in particular to control for junior high school GPA and private junior high school past attendance, the only two variables for which there were some marginally significant (at the

10 percent level) differences between IPN admittees and rejected candidates at follow-up. Intuitively, we are comparing the outcomes of students within the same decile of junior HS GPA or past schooling background that were marginally admitted and rejected from the IPN system. We also include three control variables for the set of school choices stated during the Comipems allocation process (although the local randomness of the IPN admission cutoffs should be orthogonal to those choices).

Table 15 reports the RD design estimates with the controls for the set of outcomes examined above for the bandwidth of 5. We obtain the same picture as before, with statistically significant effects of IPN admission on expected college premium, earnings expectated with a college education, learning achievements and aspirations for a graduate college education, but no effect on earnings expected with an high school HS education. In the same fashion, the magnitudes of the coefficients across the specifications are very similar to the ones reported in Tables 9 to 11. The results from our former RD design estimates thus appear very robust to the inclusion of those controls.

7 Conclusions

This study finds evidence that, in the context of public senior-high schools of Mexico City, admission to an elite school increases substancially not only students' learning achievements, but also the earnings and returns they expect from a college education, with point estimates of effects of elite high school admission of about .2 standard deviations on scholastic achievements and about 19 percentage points on the expected college premium. This finding suggests that attending an elite school can have effects that extend beyond scholastic achievements and affect the value that students attach to higher educational attainements and their beliefs on their returns. As those beliefs have been found by other authors to be important determinants of schooling decisions, this could have long-lasting effects on students educational and professional trajectories. A rough calculation based on the findings of Attanasio and Kaufmann (2009), in an observational study of the relationships beween expected college premium and college attendance among a similar sample of Mexican students, suggests that the effects of elite schools on earnings expectations we found could translate into a up to 4 to 8 points higher probablity of college attendance. Consistently, we find that students who were allocated to elite schools more often aspire to obtain a graduate college degree.

Our results also shed light on the formation of beliefs regarding the returns to college education. They are of a reduced-form nature and may capture various effects of attendance of an elite school, which makes inference on the mechanisms at play more complicated than in controlled interventions such as the provision of information on the returns to education. But the observed patterns allow testing a set of predictions derived from basic theoretical models of the formation of subjective expectations of returns to higher attainments. They are notably consistent, in a simple human capital framework, with students valuing the higher cognitive, and possibly non-cognitive, skills that they acquire in those elite schools, although students seem to value those skills only conditional on getting a college education. At the opposite, we find no higher effects on the expectations of students from more disadvantaged family background, suggesting that an information channel is not determinant, and the lack of evidence of effects of admission to elite high schools on the earnings expected with only a high-school diploma tends to discard that the changes in expectations are driven by the reputation on the labor market (i.e. a signaling mechanism).

On a methodological standpoint, as several other studies (e.g. Abdulkadiroglu, Angrist, and Pathak (2011); Pop-Eleches and Urquiola (2011)), we have identified the causal effects of admission to an elite senior-high school in an RD design strategy by using the exam-based admission process to public high schools of the Federal District (FD) of Mexico and data from a panel of students allocated through that process. However, our RD design strategy departs from previous studies by identifying the effects of admission to the IPN system rather than to a specific IPN school. This allows isolating the effects of admission to an elite school compared to another school system, and makes the interpretation of treatment and of the estimated effects straigthforward.

References

Abdulkadiroglu, A., J. D. Angrist, and P. A. Pathak (2011): "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools," NBER Working Papers 17264, National Bureau of Economic Research, Inc.

Angrist, J., E. Bettinger, and M. Kremer (2006): "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia," *American Economic Review*, 96(3), 847–862.

Angrist, J. D., P. A. Pathak, and C. R. Walters (2011): "Explaining Charter School Effectiveness," NBER Working Papers 17332, National Bureau of Economic Research, Inc.

Attanasio, O. P. (2009): "Expectations and Perceptions in Developing Countries: Their Measurement and Their Use," *American Economic Review*, 99(2), 87–92.

Attanasio, O. P., and K. Kaufmann (2009): "Educational Choices, Subjective Expectations, and Credit Constraints," NBER Working Papers 15087, National Bureau of Economic Research, Inc.

Betts, J. R. (1996): "What Do Students Know about Wages? Evidence from a Survey of Undergraduates," *Journal of Human Resources*, 31(1), 27–56.

Chevalier, A., S. Gibbons, A. Thorpe, M. Snell, and S. Hoskins (2009): "Students' academic self-perception," *Economics of Education Review*, 28(6), 716–727.

Delaney, L., C. Harmon, and C. Redmond (2011): "Parental education, grade attainment and earnings expectations among university students," *Economics of Education Review*, 30(6), 1136–1152.

Delavande, A., X. Giné, and D. McKenzie (2011): "Measuring subjective expectations in developing countries: A critical review and new evidence," *Journal of Development Economics*, 94(2), 151–163.

Ding, W., and S. F. Lehrer (2007): "Do Peers Affect Student Achievement in China's Secondary Schools?," *The Review of Economics and Statistics*, 89(2), 300–312.

Dobbie, W., and J. Roland G. Fryer (2011): "Getting Beneath the Veil of Effective Schools: Evidence from New York City," NBER Working Papers 17632, National Bureau of Economic Research, Inc.

Dominitz, J., and C. F. Manski (1996): "Eliciting Student Expectations of the Returns to Schooling," *Journal of Human Resources*, 31(1), 1–26.

Dustan, A., A. De Janvry, and E. Sadoulet (2012): "The Benefits and Hazards of Elite High School Admission: Academic Opportunity and Dropout Risk in Mexico City," Mimeo, NEUDC conference paper.

Hastings, J. S., and J. M. Weinstein (2008): "Information, School Choice, and Academic Achievement: Evidence from Two Experiments," *The Quarterly Journal of Economics*, 123(4), 1373–1414.

Hotz, V. J., P. Arcidiacono, and S. Kang (2010): "Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals," Discussion paper.

Imbens, G., and K. Kalyanaraman (2011): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79(3), 933–959.

Imbens, G., and T. Lemieux (2007): "Regression Discontinuity Designs: A Guide to Practice," NBER Technical Working Papers 0337, National Bureau of Economic Research, Inc.

Jackson, C. K. (2010): "Do Students Benefit from Attending Better Schools? Evidence from Rule-based Student Assignments in Trinidad and Tobago," *Economic Journal*, 120(549), 1399–1429.

Jensen, R. (2010): "The (Perceived) Returns to Education and the Demand for Schooling," *The Quarterly Journal of Economics*, 125(2), 515–548.

Lee, D. S., and T. Lemieux (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48(2), 281–355.

Manski, C. F. (1993): "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?," in *Studies of Supply and Demand in Higher Education*, NBER Chapters, pp. 43–60. National Bureau of Economic Research, Inc.

Pathak, P. (2011): "The Mechanism Design Approach to Student Assignment," *Annual Review of Economics*, 3, 513–536.

Pop-Eleches, C., and M. Urquiola (2011): "Going to a Better School: Effects and Behavioral Responses," NBER Working Papers 16886, National Bureau of Economic Research, Inc.

Roemer, J.-E., and R.-J.-B. Wets (1994): "Neighborhood Effects on Belief Formation and the Distribution of Education and Income," Discussion paper.

Zafar, B. (2009): "College major choice and the gender gap," Discussion paper.

——— (2011): "How Do College Students Form Expectations?," *Journal of Labor Economics*, 29(2), 301–348.

Table 1: School inputs in IPN and other schools

Variables	Other Schools*	IPN Schools
Comipems Score means	-0.048	1.634
	(0.549)	(0.416)
Comipems Score SD	0.625	0.529
	(0.132)	(0.088)
Clase Size	42.353	39.839
	(5.648)	(2.752)
Students/PC	11.358	3.608
	(11.113)	(1.031)
% Teachers w/College	0.817	0.869
	(0.115)	(0.123)
% Full-time Teachers	0.139	0.286
	(0.167)	(0.095)
Observations	**593**	**16**

Standard deviations in parentheses.
* Does not include UNAM schools.

Table 2: College premium in expectations and earnings data

	Enlace survey			**Employment survey data (ENOE)**				
	Expected log earnings			**Log earnings**			**Regression Coeff.***	
	Senior HS	University	Implied Return	Senior HS	University	Implied Return	Ln(month. wage)	Log hourly wage
Mean	8.802	9.550	0.749	8.324	8.775	0.451	0.608	0.697
SE	(0.00246)	(0.00218)	(0.00227)	(0.0306)	(0.369)	(0.0438)	(0.0608)	(0.0630)
Observations	4,3691	43,616	43,542	277	263	540	540	539

* College premium estimated through the regression $ln(wage_i) = b_0 + b_1\ university_i + b_2\ exp_i + b_3\ exp_i^2 + b_4\ female_i + u_i$

Table 3: Partial correlations analysis of expected college premiums: invidual characteristics

	(1)	(2)	(3)
VARIABLES	ereturns	ewage_hs	ewage_college
Comipems Score	0.0424***	0.0245***	0.0669***
	(0.00359)	(0.00355)	(0.00389)
Junior HS GPA	0.0153***	-0.0161***	-0.000860
	(0.00261)	(0.00339)	(0.00287)
Female	0.0770***	-0.0485***	0.0285***
	(0.00601)	(0.00636)	(0.00650)
Private JHS	0.0335***	0.0448***	0.0783***
	(0.0110)	(0.0105)	(0.0105)
Indigenous origin	-0.0185	-0.0237*	-0.0422***
	(0.0124)	(0.0123)	(0.0110)
At least one parent has senior HS	0.0192***	0.0202***	0.0395***
	(0.00598)	(0.00668)	(0.00515)
At least one parent is white-collar	0.0232***	-0.0154**	0.00774
	(0.00681)	(0.00779)	(0.00588)
Wealth Index	0.0122	0.0243**	0.0365***
	(0.00974)	(0.0114)	(0.0101)
Constant	0.680***	8.813***	9.493***
	(0.00811)	(0.00848)	(0.00927)
Observations	44,064	44,064	44,064
R-squared	0.020	0.007	0.030

Outcomes: logs of expected earnings with high school, college and ratio (returns)
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Standard errors clustered at the school level.

Table 4: Partial regression analysis of expected college premiums: school systems

VARIABLES	(1) ereturns	(2) ewage_hs	(3) ewage_college
IPN student	0.0588***	0.0229*	0.0817***
	(0.0123)	(0.0132)	(0.0127)
Enlace Score	0.0484***	-0.0257***	0.0228***
	(0.00601)	(0.00527)	(0.00455)
Comipems Score	-0.00524	0.0390***	0.0338***
	(0.00598)	(0.00601)	(0.00515)
Junior HS GPA	0.00551**	-0.0137***	-0.00817***
	(0.00259)	(0.00328)	(0.00281)
Female	0.0821***	-0.0485***	0.0337***
	(0.00579)	(0.00618)	(0.00636)
Private JHS	0.0246**	0.0434***	0.0681***
	(0.0109)	(0.0111)	(0.0109)
Indigenous origin	-0.0174	-0.0240*	-0.0414***
	(0.0125)	(0.0124)	(0.0111)
At least one parent has senior HS	0.0170***	0.0200***	0.0370***
	(0.00595)	(0.00663)	(0.00515)
At least one parent is white-collar	0.0219***	-0.0152*	0.00673
	(0.00678)	(0.00777)	(0.00595)
Wealth Index	0.00853	0.0241**	0.0326***
	(0.00970)	(0.0113)	(0.00992)
Constant	0.676***	8.808***	9.485***
	(0.00780)	(0.00868)	(0.00888)
Observations	44,064	44,064	44,064
R-squared	0.026	0.008	0.034

Outcomes: logs of expected earnings with high school, college and ratio (returns)
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Standard errors clustered at the school level.

Table 5: IPN admission cutoffs and next options

	Asignment Status - First Round								
Cutoff	Col. Bach. No.	Conalep No.	DG Bach. No.	DGETI No.	IPN No.	SE Edomex No.	UAEM No.	2nd. Round No.	Total No.
66	182	104	30	146	811	31	0	293	1597
69	249	61	4	195	593	34	0	287	1423
70	46	18	9	73	183	5	0	83	417
72	62	18	0	63	110	7	0	38	298
73	335	37	0	124	341	21	0	93	951
74	26	7	0	23	46	4	0	11	117
75	106	17	13	120	208	12	0	89	565
76	181	33	17	166	295	12	1	188	893
78	175	30	3	68	114	16	0	55	461
80	31	11	0	20	38	2	0	20	122
81	57	12	3	30	39	0	0	30	171
83	8	7	10	34	39	9	0	12	119
86	0	0	0	3	7	2	0	0	12
88	47	16	3	52	47	3	0	10	178
99	0	0	1	6	7	0	0	0	14
Total	1505	371	93	1123	2878	158	1	1209	7338

Sample Bandwidth: 10 Exam points above/below cutoff.

Table 6: Balance of covariates at baseline assignment to schools

		(1)	(2)	(3)
EQUATION	VARIABLES	[2]	[5]	[10]
jgpa	Admitted	-0.124	-0.0122	0.0396
		(0.0986)	(0.0493)	(0.0525)
jhs_priv	Admitted	-0.0389	-0.0169	-0.00472
		(0.0465)	(0.0246)	(0.0252)
female	Admitted	-0.101*	-0.00378	-0.0140
		(0.0583)	(0.0329)	(0.0363)
nopc_sol	Admitted	0.713	-0.132	-0.144
		(0.511)	(0.307)	(0.313)
shipn	Admitted	-0.0969	-0.205	-0.149
		(0.292)	(0.154)	(0.161)
shhdem	Admitted	0.274	-0.0981	-0.00887
		(0.274)	(0.173)	(0.183)
p_shs	Admitted	0.343	0.0455	0.208
		(0.313)	(0.177)	(0.181)
p_whcol	Admitted	0.388	0.161	0.270
		(0.339)	(0.176)	(0.184)
	Observations	1,396	3,391	7,115
	Control Fn	Linear	Linear	Quadratic
	Cutoff FE	YES	YES	YES
	chi2	9.256	5.928	4.137
	Prob > chi2	0.321	0.655	0.845

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors clustered at the school level

Table 7: RD design estimates: Enlace test-taking

	(1)	(2)	(3)
VARIABLES	[2]	[5]	[10]
Admitted	-0.0280	-0.0235	-0.0397
	(0.0637)	(0.0345)	(0.0356)
Observations	1,396	3,391	7,115
R-squared	0.024	0.016	0.013
Control Fn	2	5	10
Cutoff FE	YES	YES	YES

Robust standard errors in parentheses clustered at the school level.
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Balance of covariates at Enlace test-taking

EQUATION	VARIABLES	(1) [2]	(2) [5]	(3) [10]
jgpa	Admitted	0.0527	0.0700	0.143*
		(0.131)	(0.0727)	(0.0768)
jhs_priv	Admitted	0.0356	0.0487*	0.0563*
		(0.0617)	(0.0263)	(0.0305)
female	Admitted	-0.147	-0.0263	-0.0449
		(0.0990)	(0.0462)	(0.0529)
nopc_sol	Admitted	0.639	-0.245	-0.101
		(0.804)	(0.406)	(0.446)
shipn	Admitted	0.373	-0.00618	0.0411
		(0.305)	(0.179)	(0.187)
shhdem	Admitted	0.477	0.231	0.234
		(0.431)	(0.210)	(0.240)
p_shs	Admitted	0.653	0.133	0.206
		(0.397)	(0.207)	(0.223)
p_whcol	Admitted	0.127	0.106	0.0932
		(0.510)	(0.224)	(0.244)
	Observations	644	1,635	3,518
	Control Fn	Linear	Linear	Quadratic
	Cutoff FE	YES	YES	YES
	chi2	10.39	6.128	12.48
	Prob > chi2	0.239	0.633	0.131

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors clustered at the school level

Table 9: RD design estimates: expected college premium

	(1)	(2)	(3)
Admitted	0.188	0.171***	0.185***
	(0.120)	(0.0588)	(0.0619)
Mean for non-admitted	0.693	0.740	0.773
Observations	480	1,176	2,435
R-squared	0.039	0.042	0.026
Bandwidth	[2]	[5]	[10]
Control Fn	Linear	Linear	Quadratic
Cutoff FE	YES	YES	YES

Outcome: logs of ratio of expected earnings with college and high school
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors clustered at the school level

Table 10: RD design estimates: expected earnings with high school and college education

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ewage_hs	ewage_hs	ewage_hs	ewage_college	ewage_college	ewage_college
Admitted	0.0234	-0.0519	-0.0408	0.211**	0.119**	0.145***
	(0.120)	(0.0570)	(0.0665)	(0.0892)	(0.0471)	(0.0489)
Mean for non-admitted	8.868	8.818	8.793	9.561	9.558	9.566
Observations	480	1,176	2,435	480	1,176	2,435
R-squared	0.031	0.027	0.020	0.038	0.040	0.031
Bandwidth	[2]	[5]	[10]	[2]	[5]	[10]
Control Fn	Linear	Linear	Quadratic	Linear	Linear	Quadratic
Cutoff FE	YES	YES	YES	YES	YES	YES

Outcomes: logarithms of expected earnings with high school and college
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors clustered at the school level

Table 11: RD design estimates: achievements at Enlace exam

	(1)	(2)	(3)
Admitted	0.188	0.240***	0.226***
	(0.132)	(0.0783)	(0.0792)
Mean for non-admitted	0.285	0.244	0.180
Observations	480	1,176	2,434
R-squared	0.145	0.137	0.205
Bandwidth	[2]	[5]	[10]
Control Fn	Linear	Linear	Quadratic
Cutoff FE	YES	YES	YES

Outcome: standardized test-scores at Enlace high-school completion exam
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors clustered at the school level

Table 12: RD design estimates: aspirations of higher educational attainments

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	ecollege_shs	ecollege_shs	ecollege_shs	egrad_shs	egrad_shs	egrad_shs
Admitted	0.0766	0.0175	0.0211	0.0974	0.165***	0.176***
	(0.0588)	(0.0321)	(0.0354)	(0.111)	(0.0482)	(0.0504)
Mean for non-admitted	0.906	0.890	0.892	0.580	0.602	0.607
Observations	480	1,176	2,433	480	1,176	2,433
R-squared	0.116	0.042	0.030	0.092	0.050	0.043
Bandwidth	[2]	[5]	[10]	[2]	[5]	[10]
Control Fn	Linear	Linear	Quadratic	Linear	Linear	Quadratic
Cutoff FE	YES	YES	YES	YES	YES	YES

Outcomes: probability of aspiring to attain a college undergraduate or graduate education
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors clustered at the school level

Table 13: RD design estimates: heterogeneities by gender

VARIABLES	(1) ereturns	(2) ewage_hs	(3) ewage_college	(4) enlace	(5) ereturns	(6) ewage_hs	(7) ewage_college	(8) enlace
Admitted	0.215***	-0.0334	0.182***	0.393***	0.109	-0.0594	0.0496	0.0899
	(0.0818)	(0.0804)	(0.0624)	(0.106)	(0.0764)	(0.0822)	(0.0640)	(0.109)
Observations	635	635	635	635	541	541	541	541
R-squared	0.057	0.030	0.059	0.175	0.054	0.062	0.052	0.151
Female	0	0	0	0	1	1	1	1
Bandwidth	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]
Cutoff FE	YES	YES	YES	YES	YES	YES	YES	YES
Control Fn	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
SE are clustered at the school level

Table 14: RD design estimates: heterogeneities by family background

VARIABLES	(1) ereturns	(2) ewage_hs	(3) ewage_college	(4) enlace	(5) ereturns	(6) ewage_hs	(7) ewage_college	(8) enlace
Admitted	0.129	-0.0260	0.103*	0.273***	0.264***	-0.0935	0.170**	0.219*
	(0.0782)	(0.0730)	(0.0611)	(0.0948)	(0.0925)	(0.0939)	(0.0721)	(0.115)
Observations	664	664	664	664	512	512	512	512
R-squared	0.080	0.055	0.055	0.128	0.080	0.067	0.052	0.177
One parent SHS	0	0	0	0	1	1	1	1
Bandwidth	[5]	[5]	[5]	[5]	[5]	[5]	[5]	[5]
Cutoff FE	YES	YES	YES	YES	YES	YES	YES	YES
Control Fn	Linear	Linear	Linear	Linear	Linear	Linear	Linear	Linear

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
SE are clustered at the school level

Table 15: RD design estimates: robustness to controls for covariates

VARIABLES	(1) ereturns	(2) ewage_hs	(3) ewage_college	(4) enlace	(5) ecollege_shs	(6) egrad_shs
Admitted	0.168***	-0.0567	0.111**	0.212***	0.00720	0.145***
	(0.0577)	(0.0566)	(0.0474)	(0.0776)	(0.0303)	(0.0472)
Observations	1,176	1,176	1,176	1,176	1,176	1,176
R-squared	0.064	0.045	0.057	0.181	0.082	0.105
Bandwidth	[5]	[5]	[5]	[5]	[5]	[5]
Control Fn	Linear	Linear	Linear	Linear	Linear	Linear
Cutoff FE	YES	YES	YES	YES	YES	YES
JHS GPA FE	YES	YES	YES	YES	YES	YES
Baseline Cov	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Robust standard errors clustered at the school level

Figure 1: Distribution of earnings expected with high school and college

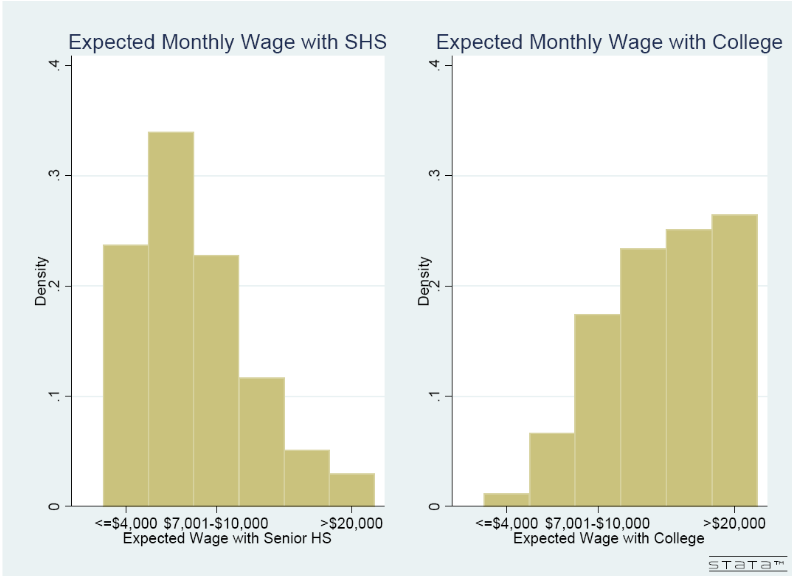

Figure 2: Distribution of observed earnings by educational attainment

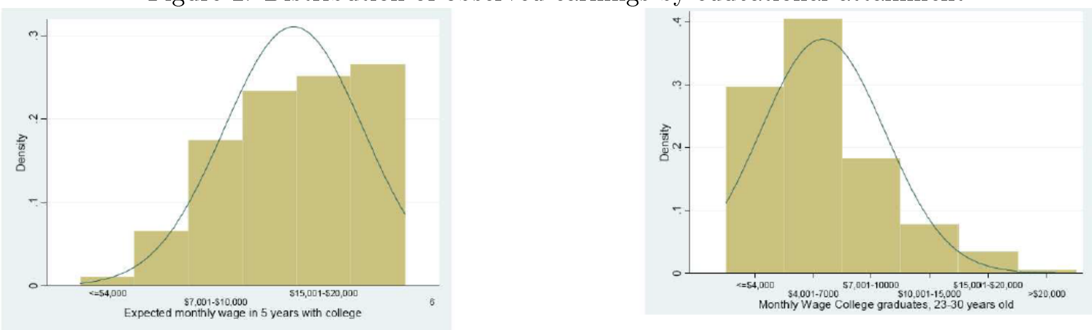

Figure 3: Mean of earnings and returns expectations conditional on learning achievements among IPN and non-IPN graduates

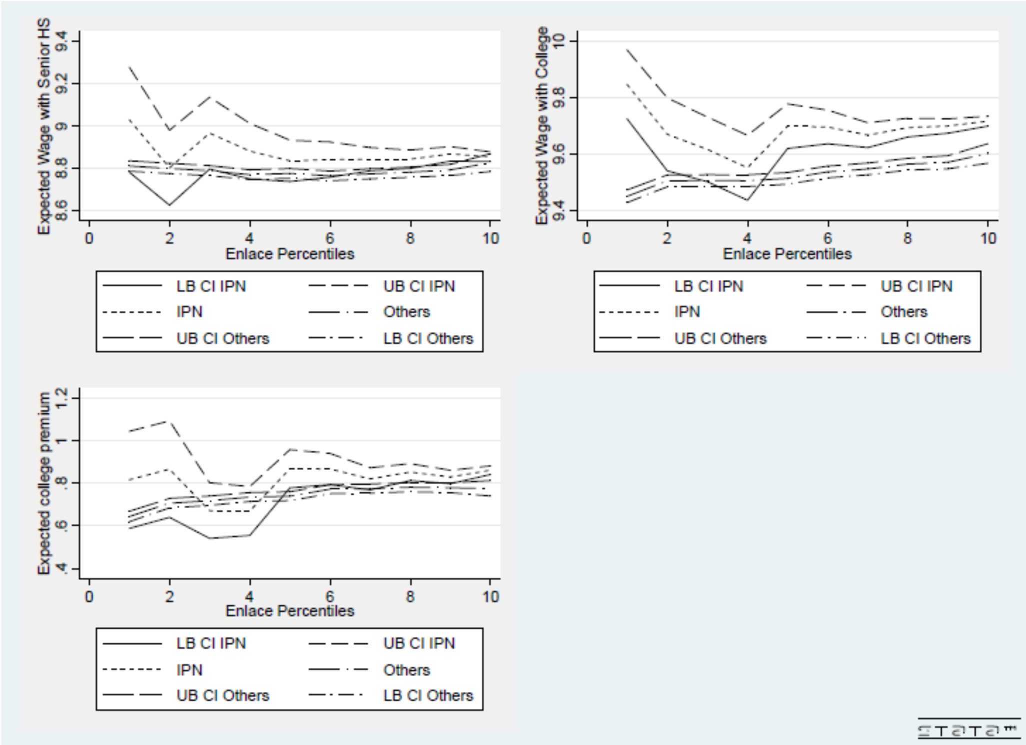

Figure 4: Standard deviation of earnings and returns expectations conditional on learning achievements among IPN and non-IPN graduates

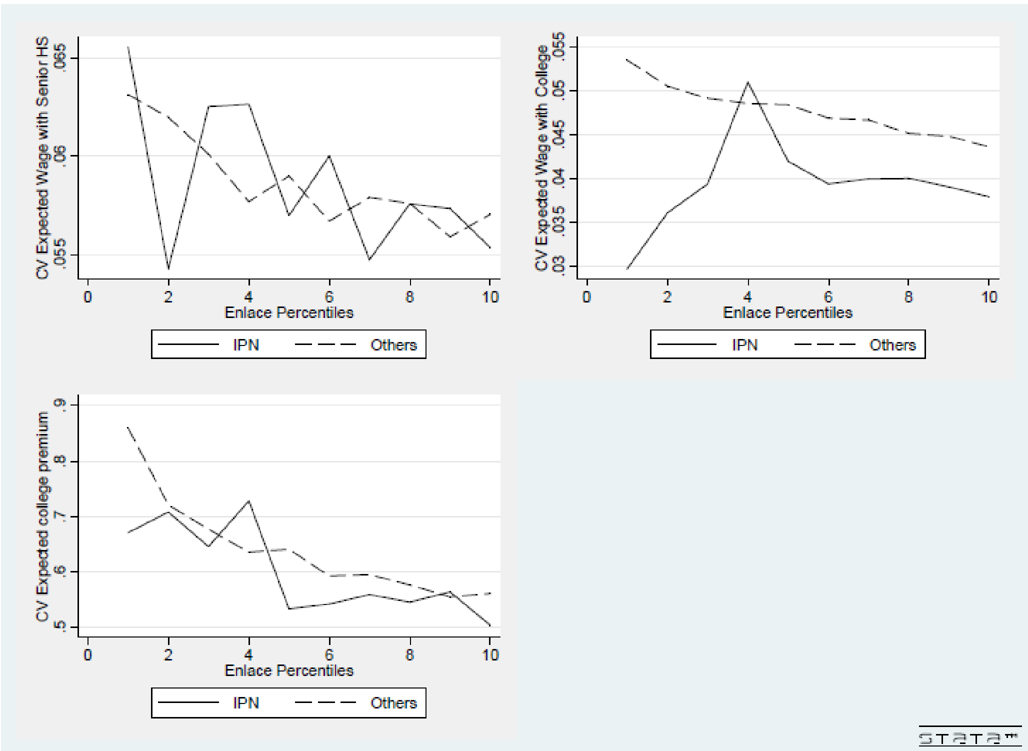

Figure 5: Distribution of the distance of students' scores from IPN admission cutoffs at assignment to schools

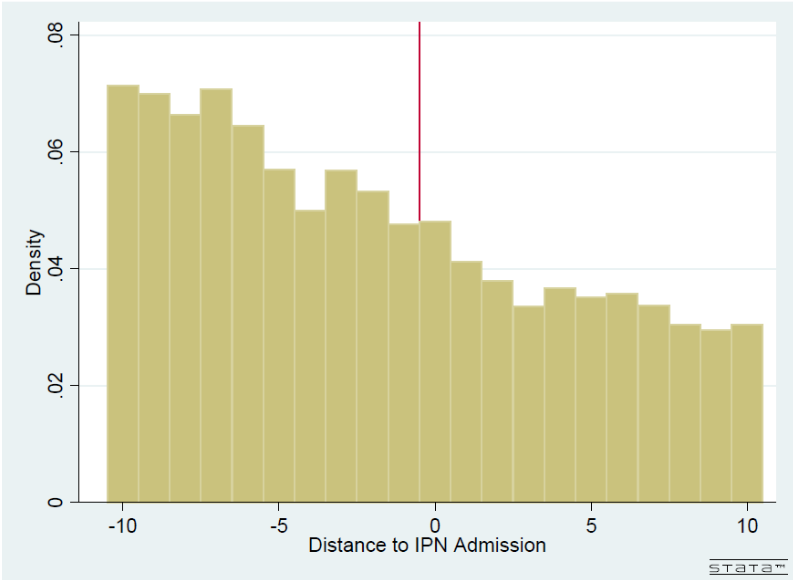

Figure 6: Means of individual characteristics by distance from IPN admission cutoffs

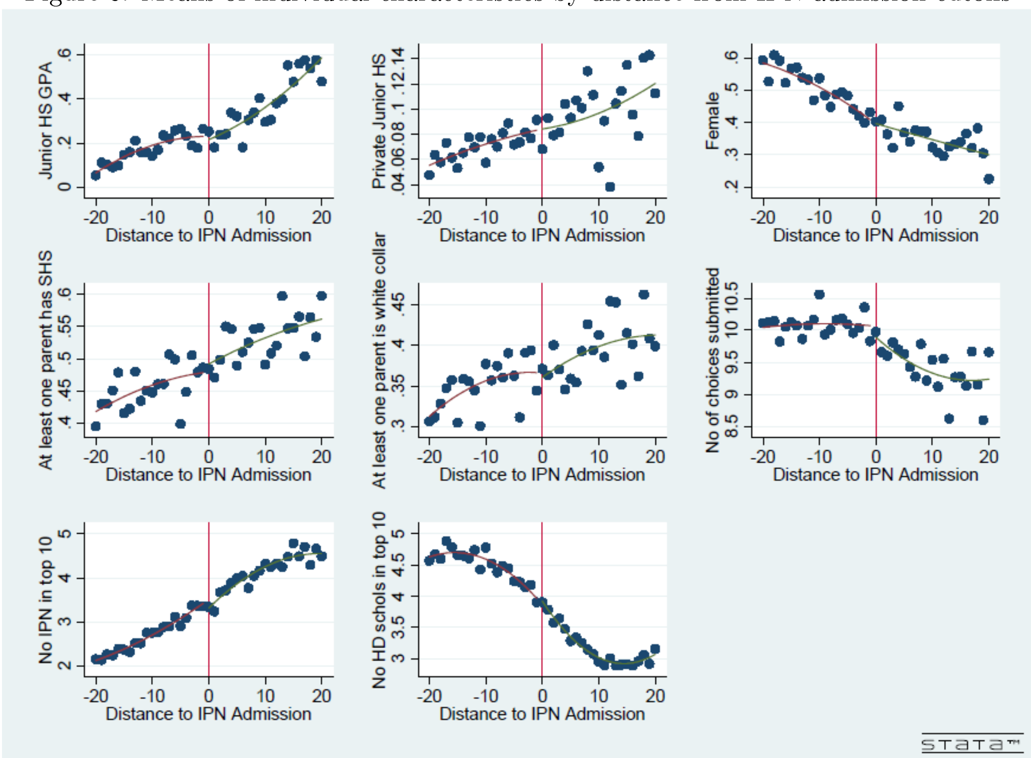

Figure 7: Distribution of the distance of students' scores from IPN admission cutoffs at Enlace test-taking

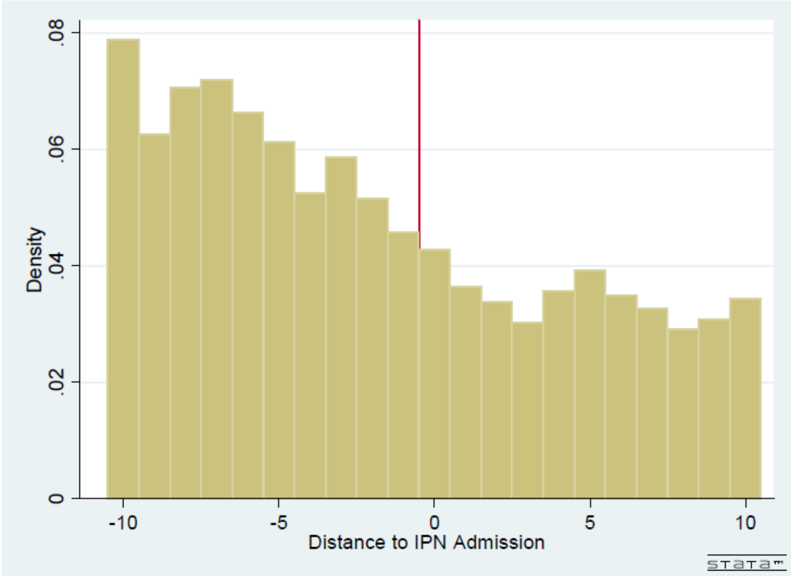

Figure 8: Probability of taking Enlace in an IPN school, conditional on taking the exam, by distance to the IPN admission cutoffs

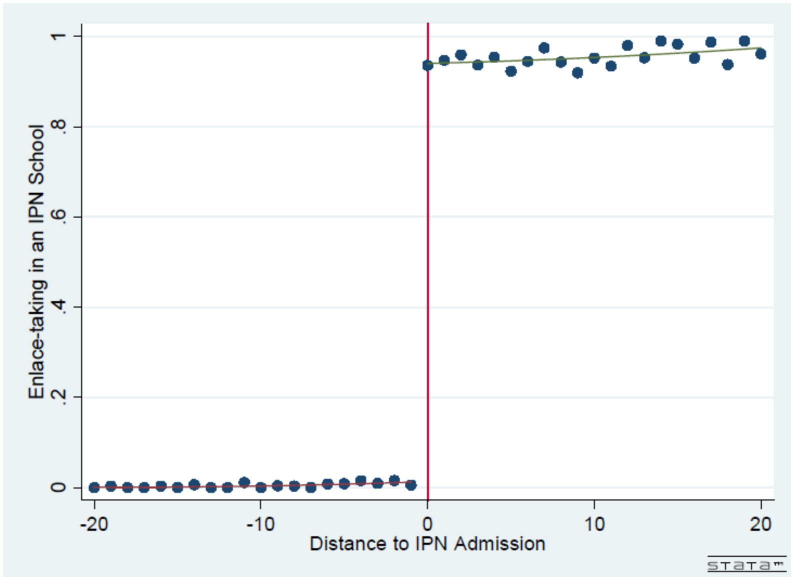

Figure 9: Means of school inputs by distance from IPN admission cutoffs

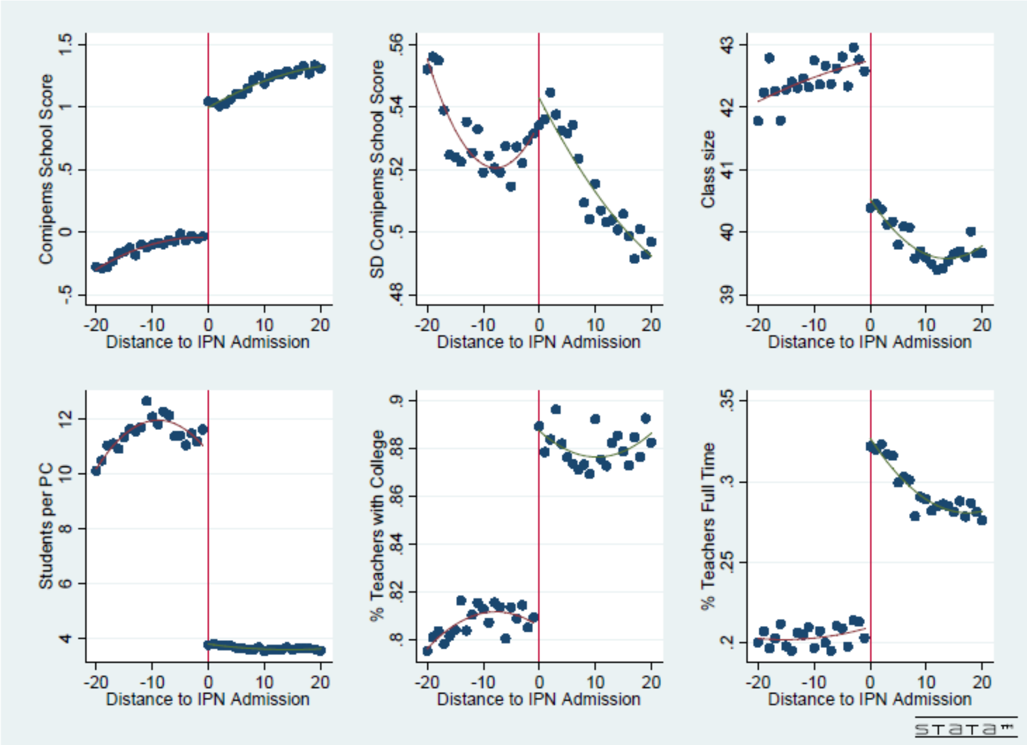

Figure 10: Means of student outcomes by distance from IPN admission cutoffs