European University Institute

MAX WEBER PROGRAMME FOR POSTDOCTORAL STUDIES

EUI WORKING PAPERS

MWP 2016/02
Max Weber Programme

The Effect of the Increasing Demand for Elite Schools on Stratification

Ricardo Estrada

European University Institute
Max Weber Programme

The Effect of the Increasing Demand for Elite Schools on Stratification

Ricardo Estrada

EUI Working Paper **MWP** 2016/02

ISSN 1830-7728

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu
cadmus.eui.eu

Abstract
I use detailed applications data to document a case in which, contrary to prevailing concerns, increasing school stratification by ability co-existed with stable stratification by family income: Mexico City public high schools. To understand this puzzle, I develop a model that shows that the effect of an overall increase in the demand for elite schools on school stratification by family income is a horse race between the correlations of family income and ability, and family income and demand. My empirical analysis reveals an initial (and decreasing) demand gap by family income that explains the observed stability in stratification.

Keywords
School choice, stratification, elite schools, aspirations gap.

JEL codes
I21, I24, D59

Via dei Roccettini, 9, 1-50014 San Domenico di Fiesole, Italy, ricardo.estrada@eui.eu. *I am indebted to Marc Gurgand for detailed comments. I am grateful to Yann Algan, Axelle Arquié, François Bourguignon, Caterina Calsamiglia, Andrew Clark, Julien Combes, Antoni-Italo De Moragas, Adeline Delavande, Francisco Ferreira, Jérémie Gignoux, Julien Grenet, Andrea Ichino, Claire Lebarz, Hugo Ñopo, Arthur Silve, Olivier Tercieux and participants at seminars at the Paris School of Economics, the conferences EWMES Bocconi University and NEUDC Boston University, and the Max Weber Economics Group at EUI for insightful comments. Financial support from CONACYT and the Max Weber Programme at EUI is gratefully acknowledged.*

Ricardo Estrada
Max Weber Fellow 2014-2015, 2015-2016

1 Introduction

A growing number of parents and students seem to put a high value on education at elite institutions, pushing up their academic selectivity. For example, the number of college applicants in the United States has doubled since the early 1970s, while the number of school seats in the most selective institutions has barely grown (Bound, Hershbein and Long, 2009). In consequence, the most selective colleges are substantially more selective today than before. The average student in the top one percent of colleges went from scoring at the 90th percentile of the SAT/ACT tests in 1962 to scoring at the 98th percentile in 2007 (Hoxby, 2009).

As academic selectivity increases, there are growing concerns that the correlation between family income and ability, an empirical regularity, leads to more stratification by family income and reduced social mobility. To make the story worse, higher admission bars might not be the only obstacles between low-income students and elite schools. Several studies have documented that, even conditional on academic ability, low-income students tend to make less ambitious schooling choices, both at the high school and the college level (see, for example, Hastings, Kane and Staiger, 2006 and Hoxby and Avery, 2012 on the US).

In this paper, I use detailed applications data to document a case in which, contrary to the above concerns, increasing school stratification by ability co-existed with stable stratification by family income: Mexico City public high schools. As Figure 1 shows, school stratification by academic ability increased sharply during the 2001-2010 period (as measured by the within- and between-school variance of incoming students' exam scores).[1] In contrast, school stratification by family income remained stable. To understand this puzzle, I develop a simple model to study the effect of an exogenous change in the demand for elite schools on stratification by family income.

Admission to public high schools (grades 10-12) in Mexico City is centralized and depends solely on school capacities, student test scores and student preferences over schools.[2] A key feature of this school system is the prominence of two sets of schools: the National Polytechnic Institute and the National Autonomous University of Mexico (IPN and UNAM, respectively, using their Spanish acronyms). Both the IPN and UNAM schools are managed by leading universities and offer better school inputs than the other public schools. Furthermore, as I show in the empirical section, the demand for admission and the academic selectivity of both types of school saw a large increase during the period of analysis, while their capacity remained roughly constant.

If admissions are based on ability, an increase in the demand for elite schools leads to higher selectivity and stratification by ability – holding constant school seats. The effect on stratification by family income is less straightforward, though. Hence, I develop a simple model

[1] All statistics presented are estimated from micro-data of the applications to the centralized admission system (Comipems) for the 2001-2010 period. I describe the data in detail below.

[2] The assignment mechanism ranks students according to their score in a common exam and then allocates students to schools, following their stated preferences.

to investigate how an exogenous change in demand affects stratification by family income. In this model, students (who are either high or low income, and high or low ability) can apply to a (selective) elite school or stay at a neighborhood school. Admissions are based on ability. I allow for positive correlations between family income and ability, and family income and demand for the elite school. I study first the effect of an overall change in demand, e.g. one that does not vary by family income or ability.

The model shows that an overall increase in the demand for elite schools has an ambiguous effect on stratification by income if there is a demand gap (conditional on ability, high-income students are more likely to apply to elite schools than low-income students). The intuition is twofold. On the one hand, as demand increases, high-income students (who tend to be higher ability) are more likely to gain admission to elite schools. But, on the other hand, as elite schools admit more high-ability students (both from high and low-income status) there are fewer seats available for low-ability, high-income students, who before were disproportionately benefiting from the reduced demand of more able students. In other words, the effect of a change in demand on stratification is a horse race between the correlations of family income and ability, and family income and demand. Less surprisingly, a reduction in the demand gap has the effect of reducing stratification by family income.

For the empirical analysis, I use time variation in demand, stability in the supply of elite-school seats and the rule-based allocation of students to schools to identify the causal effect of an increase in the demand for elite schools on admission outcomes. The matching between students and schools depends solely on student test scores in a common exam and submitted schooling preferences, on the one hand, and the school seats, on the other. Hence, holding constant test scores (position at the priority queue) and the supply of seats at elite schools, any change in admission outcomes *must* come from a change in demand. Because of the rule-based identification, the results presented here do not depend on the specific factors driving the observed change in demand.

First, I show that the generalized increase in demand translated into dissimilar admission outcomes along the ability (priority) distribution. As demand increases, students with high test scores are more likely to be admitted to elite schools, at the expense mainly of students in the middle of the priority distribution (as those in the bottom have, from the start, low admission rates).

My empirical analysis then reveals the existence of initial (and decreasing) demand (and admission) gaps by family income. In the year 2001, conditional on the exam score, low-income students were 15 percentage points less likely to select at least one elite school in their choice list than their more privileged counterparts, and 20 percentage points less likely to select an elite school as their first choice. This gap was roughly constant along the ability distribution and, given the allocation rules, it translated into an admission gap. High-income students were more, or at least equally, likely to be assigned to elite schools than low-income students along the priority queue. Then, after 2001, the demand gap by family income went down. For the top of the priority distribution, the reduction between 2001 and 2010 was around 2/3 of the initial gap for students who select at least one elite school and around 3/4 for those who select an elite school as their first choice.

As a consequence of both the increasing demand and the decreasing demand gap, there is a clear decline in the admission gap by family income after 2001, particularly in the middle and top of the priority distribution. The reduction in the admission gap in the top of the ability distribution amounts to roughly 2/3 of the initial gap. Among high-achievers, the admission gap decreases because more low-income students demanded and gained admission to elite schools, while among middle-achievers, the admission gap decreases because fewer high-income students could gain elite-school admission.

The main contribution of this paper is to the literature on school stratification. In an influential paper, Epple and Romano (1998) analyze how competition between public and private schools (with school vouchers) and household preferences for peer quality leads to school stratification by ability and, to a lesser degree, by family income. I describe a similar stratification pattern in a very different context: a one-sided market with centralized admission and rule-based allocation of students to public schools.

There are several recent papers that develop rich theoretical models to explain the prevalence of school stratification. Important works along this line include MacLeod and Urquiola (2015) on stratification in competitive markets; and Calsamiglia, Martínez-Mora and Miralles (2015) and Avery and Pathak (2015) on stratification in school choice systems. Because of analytical complexity, these studies model students (households) as uni dimensional (making ability and income interchangeable). My results highlight the relevance to study stratification of modeling family income and academic ability as two separate dimensions.

This paper also relates to a strand of studies interested in why low-income students express a lower demand for selective schools and colleges. In addition to document this so-called aspirations gap, economists have turned to investigate issues of imperfect information about peer quality (Hastings and Weinstein, 2008) and allocation rules (Ajayi, 2013) in high schools, or about the costs and benefits to colleges (Hoxby and Turner, 2015; Hastings, Neilson, Ramirez and Zimmerman, 2015; Pallais, 2015). I contribute to this literature by studying how the interaction between an increasing demand for elite schools and a demand gap by family income shapes schools stratification.

2 Stratification by Family Income

If the allocation of students to schools depends on ability and student choices, an overall increase in the demand for elite schools leads to higher selectivity and stratification by ability. The effect on stratification by family income is less straightforward, though. In this section, I develop a simple model to study the effect of an exogenous change in demand for elite schools on stratification by family income.

2.1 Model set up

Suppose there is a mass 1 of students I and a mass 1 of school seats Q. School seats belong to either an elite or a neighborhood school. Let q represent the share of seats belonging to the elite school.

Students can be either high or low income, and either high or low ability. Let w define the proportion of students who are high income. Among high-income students, let a be the share who are high ability and $1-a$ the share who are low ability. Among low-income students, let $a-s$ be the share who are high ability and $1-(a-s)$ the share who are low ability. s captures the ability gap among low-income students, which is the difference in the share of high and low-income students who are high ability. I am interested in a context in which there is a positive correlation between family income and ability, so s is positive and has a support $s \in (0,a)$. Summing up, the proportion of students by category is:

- high-income, high-ability students are $HIHA = aw$
- high-income, low-ability students are $HILA = (1-a)w$
- low-income, high-ability students are $LIHA = (a-s)(1-w)$
- low-income, low-ability students are $LILA = (1+s-a)(1-w)$

Let d be the share of high-income students who apply for admission to the elite school; and $d-g$ be the share of low-income students who apply for admission to the elite school. g captures the demand gap for the elite school among low-income students, which is the difference in the share of high- and low-income students who demand admission to the elite school. In line with the previous restriction, $g \in (0,d)$. Using the superscript ^{D} to denote applicants, the proportion of students by category who apply to the elite school is:

- $HIHA^D = daw$
- $HILA^D = d(1-a)w$
- $LIHA^D = (d-g)(a-s)(1-w)$
- $LILA^D = (d-g)(1+s-a)(1-w)$

I am interested in elite schools which are selective, and by selective I mean that they have a binding admission quota. Hence, I will focus on values for d and g that make total applications to the elite school larger than the seats at this school. Hence, $HIHA^D + HILA^D + LIHA^D + LILA^D > q$.

Students are allocated to the elite school using ability as a priority. First, high-ability applicants are assigned to the elite school. If the number of high-ability applicants is larger than the number of elite seats, applicants are chosen for admission at random. Then, if this is not the case, low-ability applicants are assigned to the elite school. If the number of low-ability applicants to the elite school is larger than the number of available seats (after high-ability admission), applicants are chosen for admission at random. Students who are not assigned to the elite school are assigned to the neighborhood school.

The ability-based priority implies that if the number of high-ability applicants is larger than the capacity at the elite school, then no low-ability student gains elite admission. In this case, an increase in demand does not have any effect on stratification by ability (as both before and after only high-ability students are admitted to the elite school). As I am interested in changes in demand that increase stratification by ability, I will focus on the case in which there are less high-ability applicants than elite school seats $\left(HIHA^D + LIHA^D < q\right)$.

There are always more low-ability applicants than remaining seats at elite schools (because of the binding admission quota imposed). Low-ability applicants to the elite school are admitted with a probability (α) equal to the number of available seats over the number of low-ability applicants. Using the superscript ^{E} to denote assigned to the elite school, the proportion of students admitted to the elite school is:

- $HIHA^E = HIHA^D = daw$
- $HILA^E = \alpha d(1-a)w$
- $LIHA^E = LIHA^D = (d-g)(a-s)(1-w)$
- $LILA^E = \alpha(d-g)(1+s-a)(1-w)$

And the probability that low-ability applicants are admitted to the elite school (α) is:

$$\alpha = \frac{q - HIHA^D - LIHA^D}{HILA^D + LILA^D} \tag{1}$$

2.2 The Effect of Demand on Stratification

Let school stratification by family income be positive if the share of high-income students allocated to the elite school is larger than the ratio of elite over total school seats $\left(\frac{HIHA^E + HILA^E}{HIHA + HILA} > q\right)$. Note that this inequality holds true because of the positive relationship between family income and ability ($a > a - s$) and the demand for the elite school ($d > d - g$). In other words, high-income students are overrepresented at elite schools. Hence, the sign of the change in stratification is the same as the sign of the change in the share of high-income students admitted to the elite school. For simplicity, call this share γ.

The share of high-income students allocated to the elite school is (γ):

$$\begin{aligned}\gamma &= \frac{daw + \alpha d(1-a)w}{aw + (1-a)w} \\ &= da + \alpha d(1-a)\end{aligned} \tag{2}$$

I am interested in the sign of the partial derivative of γ with respect to d, to sign the effect of an exogenous change in demand for elite schools on stratification by family income.

$$\frac{\partial \gamma}{\partial d} = a + \alpha(1-a) + \frac{\partial \alpha}{\partial d} d(1-a) \tag{3}$$

$$\frac{\partial \alpha}{\partial d} = \frac{(-HIHA - LIHA)(HILA^D + LILA^D) - (q - HIHA^D - LIHA^D)(HILA + LILA)}{(HILA^D + LILA^D)^2} \tag{4}$$

Note that the first two terms on the right-hand side of equation 3 capture the direct (choice) effect of a change in demand on the allocation of high-income students to the elite school. Both terms have a positive sign. The first term corresponds to the choice effect for high-income, high-ability students, and depends on the proportion of high-ability (a) among high-income students. The second term corresponds to the choice effect for high-income, low-ability students, and depends both on the share of these among high-income students $(1-a)$ and the probability that a low-ability student is admitted to the elite school (α). Finally, the third term captures the indirect (crowd-out) effect of the change in demand, which is negative. This crowd-out effect is the change in the probability of admission for low-ability applicants as demand increases (this effect does not concern high-ability applicants because they have guaranteed admission).

Summing up, the effect of a change in demand on stratification by family income depends on the magnitude of the direct (choice) effect versus the magnitude of the indirect (crowd-out) effect. To understand better how demand shapes stratification, it can also be useful to see how the share of low-income students allocated to the elite school $\left(\frac{LIHA^E + LILA^E}{LIHA + LILA}\right)$ varies as demand changes. For simplicity, call this share γ'. Analogously to equations 2 and 3:

$$\begin{aligned} \gamma' &= \frac{(d-g)(a-s)(1-w) + \alpha(d-g)(1+s-a)(1-w)}{(a-s)(1-w) + (1+s-a)(1-w)} \\ &= (d-g)(a-s) + \alpha(d-g)(1+s-a) \end{aligned} \tag{5}$$

$$\frac{\partial \gamma'}{\partial d} = (a-s) + \alpha(1+s-a) + \frac{\partial \alpha}{\partial d}(d-g)(1+s-a) \tag{6}$$

Now, let's compare the effect of a change in demand over elite school admission of high- and low-income students by looking at equations 3 and 6. In this comparison, it is the role of the correlations between family income and ability (ability gap, s) and family income and demand (demand gap, g) that stands out. Both correlations affect the admission chances of all low-ability students in a similar way through the success parameter α (see equations 1 and 4). However, both correlations also shape differently the admission chances of both groups as equation 6 shows.

First, given a change in demand, a larger correlation between family income and ability (s) decreases the allocation of low-income students to the elite school – in other words, it increases school stratification by family income. This is not surprising. A larger s, implies that there are fewer high-ability, low-income, students, who have priority for elite admission (admission chances equal to one 1 conditional on application); and more low-ability, low-income students, who if they apply to the elite school enter into the admission lottery (with probability $\alpha \in (0,1)$).

Second, and maybe this is less evident, given a change in demand, a larger correlation between family income and demand (g) increases the allocation of low-income students to the elite school – and decreases school stratification by family income. This result is explained because a larger correlation between family income and demand (g) decreases, for low-income students, the (negative) crowd-out effect of an increase in demand. In other words, the crowd-out effect is not a problem for (low-ability) students who do not apply to the elite school, and these are more common among low-income students (because g>0).

Equation 6 shows that the effect of an increase in demand over the elite school on stratification by family income is a horse race between the correlations between family income and ability, on the one hand, and family income and demand, on the other. Intuitively, as demand increases, high-income applicants (who tend to be also higher ability) are more likely to be admitted to the elite school, but as the elite school receives more high-ability students (both high and low-income), there are fewer seats available for low-ability, high-income students, who before were disproportionately benefiting from the reduced demand of more able students.

Less surprisingly, an increase(reduction) in the demand gap has the effect of increasing (reducing) stratification by family income:

$$\frac{\partial \gamma}{\partial g} = \frac{\partial \alpha}{\partial g} d(1-a) \tag{7}$$

$$\frac{\partial \alpha}{\partial g} = \frac{(LIHA)(HILA^D + LILA^D) - (q - HIHA^D - LIHA^D)(-LILA)}{(HILA^D + LILA^D)^2} \tag{8}$$

The crowd-out effect is now positive (see equation 8) because an increase in the demand gap means that there are fewer (low-income) applicants to the elite school. Finally, the total derivative of stratification with respect to exogenous changes in the (overall) demand (d) and the demand gap (g) is:

$$\begin{aligned} df(\gamma) &= \frac{\partial \gamma}{\partial d} dd + \frac{\partial \gamma}{\partial g} dg \\ &= \left[a + \alpha(1-a) + \frac{\partial \alpha}{\partial d} d(1-a) \right] dd + \left[\frac{\partial \alpha}{\partial g} d(1-a) \right] dg \end{aligned} \tag{9}$$

So, an increasing (overall) demand coupled with a decreasing demand gap (g) have an ambiguous effect on stratification by family income. As before, the sign of the change in stratification depends on the relative magnitude of the direct choice effects $(a + \alpha(1-a))$ versus the crowd-out effects. Though, in this case, the crowd-out effect is larger than when only the overall demand (d) changes.

3 Mexico City public high schools

3.1 Elite school systems: IPN and UNAM

There are two sets of schools that stand out among public high schools (grades 10-12) in Mexico City: National Polytechnic Institute and National Autonomous University of Mexico (IPN and UNAM, respectively by their Spanish acronyms). Both IPN and UNAM schools are managed by leading universities, offer better school inputs and are by far the more selective.

IPN is also the main public technological higher education institution in Mexico, while UNAM is the oldest and largest public university. IPN high schools provide general senior secondary-level education with a scientific and technical background, but attendance at an IPN high school does not grant access to the IPN higher education institute. In contrast, students at UNAM high schools follow a generalist curriculum and have automatic admission to an UNAM higher education campus, although with some restrictions in the choice of college major.

The IPN system has 16 high schools with a total incoming class of around 21,600 students, while the 14 UNAM high schools enroll 34,200 new students per year (around 8% and 12%, respectively, of Comipems applicants in 2010). Table 1 presents simple averages of school characteristics across school systems (based on information from the Comipems application system and the Mexican census of schools). Students admitted to IPN and UNAM schools tend to be more academically able and to come from a more privileged background. They score 1.5 SD higher on average on the entry exam than students in the other schools and are around twice as much more likely to have at least one senior-secondary educated and white-collar parent. There are also clear differences in terms of graduation from a private junior high school, an indication of socioeconomic status, and per capita household income.

IPN and, particularly, UNAM schools also have a set of improved inputs with respect to the other public schools, at least in terms of class size and teacher credentials (see again Table 1). Furthermore, there is evidence that IPN schools translate their better inputs into better student outcomes. Using the variation in admission outcomes generated by the centralized allocation from students to schools, Estrada and Gignoux (2014) find a large effect of admission to the IPN system on students' expected earnings with a college education and learning achievement in Mathematics. Studying the effects of admission in other equally selective schools in the city, but not with better school inputs, the authors also find that these gains are not driven by exposure to more highly selected peers.

Summing up, IPN and UNAM schools offer admitted students exposure to academically stronger and more privileged peers, and school inputs that can generate better student outcomes. In the case of UNAM, they also offer secure admission to one of the leading public universities in the country. Hence, there is no surprise that, from the beginning of the period under study, the demand for admission into IPN and UNAM schools is high.

3.2 The Comipems admission process

Since 1996, the Metropolitan Commission of Public Senior-Secondary Education Institutions (Comipems) has centralized admission into nine of the ten systems of public high schools (grades 10-12) in the Mexico City metropolitan area, which comprises the Federal District and 22 municipalities from the neighboring State of Mexico.[3] The admission process is based on students' scores in a common test and explicit schooling choices, and the number of slots available in each school. The allocation of students to schools uses the serial dictatorship algorithm and works as follows:

First, applicants submit a ranked set with up to 20 schooling options (from 536 options in 2010). Students actually submit a list of preferred tracks as some schools offer more than one track. However, most schools, including IPN and UNAM schools, have only one track at the time of admission. In practice, students receive, at their current school, an application package in January of their last year of junior high school (grade 9). Applicants must turn in their registration form (with their choice list) in February or March of the same year. Second, applicants take a common standardized exam in the last weekend of June. All applicants with at least 31 correct answers out of 128 questions in the exam are allowed to register in a Comipems school (around 97% in the 2001-2010 period).[4]

In parallel, schools submit their number of available seats. Schools do not submit any priority criteria over students.

Next, students are allocated to schools. A computer program places students in a queue according to their test score, where the student with the highest score gets the first position in the queue, the student with the second highest score obtains the second position and so on. Students with the same test score obtain the same priority. In the order of the queue, the algorithm proceeds to allocate individuals to the school with available seats, which the students themselves had ranked the highest in their choice list. So, the first student obtains her top choice, the second student obtains her top choice among schools with available seats, and so on. The algorithm stops when the application from the last student in the queue is processed. In case of ties, schools decide to either accept or reject all students involved in the tie.[5]

[3] A recent system of high schools administered by the Federal District government and targeted to low-achieving students does not belong to Comipems.

[4] Applicants who list a school from the UNAM system as their first choice must take an exam version designed by this institution, while all the other students take an exam design designed by Ceneval (the institution in charge of the assignment process). Both exams are designed to be equivalent in level of difficulty. I 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 to one of their most preferred choices.

[5] IPN and UNAM schools require a minimum junior high school GPA of 7/10 for admission, in contrast to the 6/10 necessary for graduation from this schooling level and for admission to the other Comipems schools. In consequence, the allocation algorithm ignores the applications to IPN and UNAM schools from applicants with a

Finally, students who choose only schools which happened to be too selective with respect to their test scores, i.e. who miss admission for all their listed choices, can register in the schools with remaining slots in a second-stage process. All IPN and UNAM schools fill their seats at the first stage.

The serial dictatorship algorithm makes it a dominant strategy for students to submit their true preferences over schools (see Pathak, 2011 and Balinski and Sonmez, 1999) given hedonic preferences and no constraints in the ranking of schooling options. Preferences are hedonic if students care only about the school they are assigned independently of the other students who are assigned there. Comipems limit students to rank a maximum of 20 options, but this is hardly binding in practice. Applicants submitted 9.1 schooling options on average and only 2.5% listed 20 during the 2001-2010 period.

4 Data

4.1 The Comipems dataset

My main data set comprises the micro-data with the applications made to the Comipems system during the 2001-2010 period. The database includes, for all applicants, the submitted choice lists, the score at the entry exam, the assignment outcome in the first stage of the allocation process and the junior high school (JHS) attended. I normalize the exam score by exam cohort with mean 0 and standard deviation 1. In addition, I have basic family background information from a questionnaire attached to the registration form (e.g. parental occupation and education, and household income and size). The response to the background questionnaire is optional and there is some non-response that I discuss below.

Students report household income in the background questionnaire selecting one of 15 income brackets. I assume that each discrete income category corresponds to the mean of the two values that define each bracket. For the last bracket, which does not have an obvious interval, I make an assumption about the upper value (only .08 percent of students selected this income bracket). I then convert the values to pesos at December 2010 prices and construct a measure of per capita household income using the reported household size in the same questionnaire. I generate a dummy variable to categorize students as high- and low-income (above and below the cohort median of the per capita household income).[6]

The number of total applicants increased steadily from 207,662 in 2001 to 276,581 in 2010. I focus the analysis of the individual admission outcomes on students ranked in the first 200,000 positions in the admission priority queue. I do it to gain comparability across years and because

GPA below 7/10 (5% of elite school applicants in 2001-2010).

[6]The 2001 background questionnaire does not include a question for household size. I predict household size for 2001 applicants using the parameters estimated in an OLS regression of household size on entry exam score, female status, private JHS attendance, number of submitted choices, share of elite schools and the selection of a elite school as top ranked option in the 2002-2010 sample.

students with a position in the admission queue above 200,000 have barely a chance of being admitted to an elite school. These restrictions lead to a data set with 2,024,493 observations and 1,647,078 observations when the no-response to background questions is taken into account.

I consider all admitted applicants in the first stage of the admission process for the analysis of school stratification. Because of missing information, I omit the students enrolled in the second stage of the process. Around 15% of the total applicants missed admission to all their listed schooling choices during the 2001-2010 period. Not all students registered at one school with available spots during the second stage, though. In 2005, the only year for which I have information about second-stage admission, less than half did so (47%).

4.2 Descriptive statistics

Table 2 presents descriptive statistics for the sample of students used for the analysis of admission outcomes. Column 1 reports means and standard deviations (in parenthesis) for all students, while column 2 is restricted to students who responded to the three questions in the background questionnaire used for the analysis. As we can observe, the differences between the two samples are very small and the non-response, which accounts for 19% of students, does not compromise the empirical analysis.

Half of students are female (50%) and 7% come from a private junior high school (an indication of high socioeconomic status). Students select on average 8.9 schooling options and only 2.4% list 20 options (the maximum number allowed). The preference for elite schools is clear. 77% of students select at least one elite school and 65% have one elite school as their most preferred choice. On average, elite schools account for 44% of schooling options in students' choice lists. Students are assigned on average to their 3rd/4th more preferred schooling option (3.4) and only 28% gain admission to an elite school (the figure would be lower if we consider the omitted students at the end of the priority queue). 10% miss admission to all their listed choices and go to the second round assignment.

In terms of family background, 42% of students have at least one parent with a senior high-school education and 32% at least one parent with a white-collar occupation. Per capita family income is on average \$1,312 pesos per month (in December 2010), which amounts to around \$175 US dollars, adjusting for purchasing power parity.

Columns 3 and 4 give the same descriptive statistics for students above (column 3) and below (column 4) the cohort median of per capita household income. The large correlations between family income and academic achievement, and family income and demand for (and admission to) elite schools stand out. High-income students (column 3) have on average an entry exam score larger by .37 SD than low-income students. Also, high-income students are more likely to select at least one elite school (by 12 percentage points), to rank an elite school as their more preferred school (by .17 percentage points) and to have a larger share of elite schools in their ranked set (by 16 percentage points). Consequently, high-income students are almost twice as much more likely to gain admission to an elite school than low-income students. As expected, high-income students are more likely to have white-collar and SHS educated parents, and to graduate from a private JHS. Actually, only 2% of low-income students graduated from a private JHS versus 12% of high-income students.

Summing up, the descriptive statistics presented here indicate that there is a large demand for elite schools, and that higher income students tend to obtain higher entry scores, to have a higher demand for elite schools and to be more often admitted to them.

5 Empirical analysis

5.1 Increasing demand

I first document a secular increase in demand for elite schools in the 2001-2010 period. Figure 2 shows the evolution of the expressed demand for elite (IPN and UNAM) schools from 2001 to 2010. All four indicators reported are consistent with an increasing trend in demand. For example, the share of applicants who select at least one elite school goes up from 70% in 2001 to 80% in 2010 and the share that selects an elite school as first option from 54% to 71%. Similarly, the number of elite schools in students' choice list went from 2.8 in 2001 to 4.6 in 2010 (and the share from 36% to 48%).

Second, I study if the change in demand for elite high schools varies according to the test score (priority) distribution. Panels A and B in Figure 3 show local means for two demand indicators over the priority queue (in vintiles) for three selected years (2001, 2005 and 2010). To complement this, panels C and D in Figure 3 report the regression coefficients for within-vintile changes with respect to 2001 (with confidence intervals at the 95% level). Results show that there is an economically and statistically significant increase in the demand for elite schools and that this change is not restricted to some segments of the priority (ability) distribution. The change in demand is increasing in ability, but it shifts up even at the bottom of the ability distribution.

The vector of reported coefficients captures an increase in demand that can be the result of different factors (e.g. changing schooling preferences or applicant characteristics). The results presented below do not depend on what is driving the change in demand though, because the identification strategy is based on the rule-based matching of students to schools.

5.2 Identification of the demand effect

I use time variation in demand, the stability in the supply of elite-school seats during the period and the rule-based allocation of students to schools to identify the causal effect of an increase in the demand for elite schools on admission outcomes and school stratification.

The capacity of elite schools during this period was roughly constant. The number of students admitted to elite schools only moved from 52, 500 in 2001 to 55,890 in 2010, leading to stability in the share of students in the top 200,000 positions in the priority queue allocated to elite schools.

I use the following model to estimate changes in admission outcomes within-priority-vintiles:

$$y_{ic} = \beta_0 + \beta_{1C}C'_{ic} + \beta_{2P}P'_{ic} + \beta_{3CP}C'_{ic} * P'_{ic} + e_{ic} \tag{10}$$

where y_{ic} denotes the outcome of student i from exam year (cohort) c, C' is a vector of dummies for years 2002...2010 (2001 is the reference year), P' is a vector of dummies for 2...20 priority vintiles, C'*P' is a vector of interactions among dummies for years 2002...2010 and 2...20 priority vintiles, β_{3PC} is the associated vector of parameters and e_{ic} is a disturbance term. In all estimations, standard errors are clustered at the attended junior high school level.

The allocation of students to schools depends solely on the priority queue and submitted schooling preferences, on the one hand, and the school seats, on the other. Hence, holding constant the position at the priority queue and the supply of seats at elite schools, any change in admission outcomes *must* come from a change in demand expressed in the choice lists. In this framework, the vector $\beta_{1C} + \beta_{3CP}$ captures the (overall) causal effect of a changing demand on admission outcomes of students with the same priority profile.

I estimate the following extension to equation 10 to obtain estimates of changes in an outcome gap within-priority-vintiles between high and low-income students:

$$y_{ic} = \beta_0 + \beta_{1C}C'_{ic} + \beta_{2P}P'_{ic} + \beta_{3CP}C'_{ic} * P'_{ic} + \beta_4 w_{ic} + \beta_{5C}C'_{ic} * w_{ic} + \beta_{6P}P'_{ic} * w_{ic} + \beta_{7CP}C'_{ic} * P'_{ic} * w_{ic} + e_{ic} \tag{11}$$

where w_{iy} is an indicator that turns on if student i from cohort c is a high-income student and is otherwise off, $C'_{ic} * w_{ic}$ is a vector of interactions between year dummies and income status, $P'_{ic} * w_{ic}$ is a vector of interactions between priority-vintile dummies and income status, and $C'_{ic} * P'_{ic} * w_{ic}$ is a vector of triple interactions among year dummies, priority-vintile dummies and the income status dummy, and all the rest is the same as before.

I am interested in the vector of parameters $\beta_{5C} + \beta_{7CP}$, which capture the change with respect to 2001 in the within-priority-vintile gap in outcome y between high and low-income students.

5.3 Admission outcomes

Panel A in Figure 4 presents the conditional expectation function of the probability of admission to an elite school on the position in the priority queue (in vintiles) for the years 2001, 2005 and 2010. In all cases, it is possible to observe a monotonically increasing relationship between elite-school admission and priority position (academic ability). Students first in the priority queue are more likely to be allocated to elite schools than students later in the queue. Yet, there are important differences in the intensity of this relationship across years.

In 2001, admission to an elite school is almost a linear function of the position in the priority queue. Although at different levels, there are applicants across practically all the support of the priority distribution who are able to obtain a spot at an elite school. As some high-priority students do not apply to elite schools, see again Panel A in Figure 3, not all high-priority students are allocated to elite schools, which leaves some school seats available for students with a lower admission priority.

In contrast, there are practically no students in the bottom-half of the priority queue who are allocated to elite schools in 2005. In the upper-half of the queue, some students in the vintiles 7-10, and 11, are still able to gain elite-school admission, but fewer than before. This is the result from a clear increase in the rate of admission in the top 6 vintiles. The same process continues to 2010.

Panel B in the same graph presents the regression coefficients from an OLS estimation of equation 10. The coefficients correspond to changes within-vintile in the probability of elite-school admission between 2001 (the year of reference) and 2005 and 2010. Again, it is visually clear, and highly statistically significant, that students in the bottom and middle of the priority queue reduced drastically their admission chances between 2001 and 2005 and 2010. The largest drop is for students in the middle of the queue, who were able to obtain admission before because of the relatively low demand for elite schools among top-priority students.

The empirical results are consistent with simple predictions derived from the assignment rules. As the allocation of students to schools is based solely on students' preferences (expressed through the choice lists) and position in the queue (fully determined by their exam score), a roughly constant increase in demand for elite schools translates in dissimilar admission chances along the ability distribution. A larger demand increases admission chances a priori, but actual admission outcomes depend on the preferences of those first in the queue. Student with high scores translate their higher demand into higher admission rates to elite schools at the expense of middle and bottom-priority students.

5.4 Demand and admission gaps

Panels A and B in Figure 5 present the local means of two demand indicators along the priority queue (in vintiles) by income status for the year 2001. The existence of a (conditional on ability) demand gap is clear. Conditional on the exam score, low-income students are less likely to select at least one elite school in their choice list by around 15 percentage points than more privileged youth and to select an elite school as their first choice by around 20 percentage points. The demand gap is roughly constant along the ability distribution, although the gap in the second indicator is smaller at the very top of the distribution (first vintile).

Given the rules to allocate students to schools, the (conditional) demand gap generates mechanically a (conditional) admission gap. Panel A in Figure 6 shows the corresponding conditional expectation function for elite-school admission by income status in 2001. As expected, high-income students are more, or at least equally, likely to be assigned to elite schools than low-income students at each vintile of the priority queue. The gap in the bottom-half of the priority queue is small or negligible, as only a small part of these students gain elite-school admission. The gap is increasing though along the priority distribution; from the top-five quantile and on, it stabilizes at around 20 percentage points, a similar amount to the gap in selection of an elite school as first choice.

Panels C and D in Figure 5 present the point estimates for the change in the demand gap in 2010 and 2005 with respect to 2001–with confidence intervals at the 95% level. As we can see, the demand gap decreased after 2001, although the confidence intervals for several point estimates corresponding to 2005 are marginally insignificant at the 95% level. The estimation is relatively imprecise and the confidence intervals overlap, but it seems that the reduction in the demand gap is larger for the top-half of the ability distribution and continued between 2005 and 2010. For the top vintiles of the priority queue, the reduction between 2001 and 2010 amounted to around 10 percentage points for students who select at least one elite school and around 15 percentage points for those who select an elite school as first choice. The reduction represents around 2/3 and 3/4, respectively, of the initial gap.

Panel B in Figure 6 reports the coefficients of interest for changes in the admission gap. In line with the previous results, there is a clear decline in the admission gap after 2001, particularly between 2001 and 2005 (the decline is less evident later). The reduction in the admission gap is zero for the very bottom priority vintiles; it then increases along the priority queue. Comparing 2001 and 2010, the reduction in the admission gap for the top priority-vintiles is of around 12-15 percentage points, or roughly 2/3 of the initial gap. This is an important reduction in the inequality of outcomes for students of different socioeconomic background, but with a similar ability profile.

In line with the theoretical analysis, the reduction in the (conditional) admission gap is the result of both: 1) a reduction in the conditional demand gap (Panels C and D in Figure 5) and 2) the interaction between the initial admission gap (Panel A in 6) and a crowding-out effect (Figure 3 Panel A) due to a general increase in the demand for elite schools (Panels C and D in Figure 3). In other words, among high-achievers, the admission gap decreases as more low-income students demand and gain admission to elite schools, while among low and middle-achievers, the admission gap decreases as less high-income students in these groups can achieve elite-school admission.

6 Conclusions

In this paper, I study the effect of the increasing demand for elite schools on school stratification by family income in a system in which the allocation of students to schools depends solely on student scores in a common exam, student explicit preferences over schools and school capacities. Contrary to common concerns, I show that higher school stratification by ability does not translate mechanically into higher stratification by family income. I develop a simple model that shows that the effect of an exogenous change in demand over elite schools on stratification depends on the relative strength of the correlations between family income and ability, on the one hand, and family ability and demand, on the other. Initial (and decreasing) demand gaps into elite schools by family income explain why increasing school stratification by academic ability co-existed with stable stratification by family income in Mexico City's public high schools during the 2001-2010 period.

The framework presented here can inform policy makers interested in implementing policies that increase demand for specific sets of schools (by, for example, removing informational or financial barriers), but who might be concerned that the increase in demand lead to higher stratification by family income. More generally, the large equalization process documented in this paper gives some optimistic evidence on the dynamic effects of school choice on stratification. The presented results are of special interest for school systems with clear patterns of vertical differentiation, like those in countries as diverse as the US, France, Chile, Mexico and Ghana.

References

Ajayi, Kehinde, "School Choice and Educational Mobility: Lessons from Secondary School Applications in Ghana," Working Paper December 2013.

Avery, Christopher and Parag A. Pathak, "The Distributional Consequences of Public School Choice," NBER Working Papers 21525, National Bureau of Economic Research, Inc September 2015.

Balinski, Michel and Tayfun Sonmez, "A Tale of Two Mechanisms: Student Placement," *Journal of Economic Theory*, January 1999, *84* (1), 73–94.

Bound, John, Brad Hershbein, and Bridget Terry Long, "Playing the Admissions Game: Student Reactions to Increasing College Competition," *Journal of Economic Perspectives*, Fall 2009, *23* (4), 119–46.

Calsamiglia, Caterina, Francisco Martínez-Mora, and Antonio Miralles, "School Choice Mechanisms, Peer Effects and Sorting," Discussion Papers in Economics 15/01, Department of Economics, University of Leicester January 2015.

Epple, Dennis and Richard E Romano, "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects," *American Economic Review*, March 1998, *88* (1), 33–62.

Estrada, Ricardo and Jérémie Gignoux, "Benefits to elite schools and the formation of expected returns to education: Evidence from Mexico City," PSE Working Papers halshs-00951763, HAL February 2014.

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

_ **, Christopher A. Neilson, Anely Ramirez, and Seth D. Zimmerman**, "(Un)Informed College and Major Choice: Evidence from Linked Survey and Administrative Data," NBER Working Papers 21330, National Bureau of Economic Research, Inc July 2015.

_ **, Thomas J. Kane, and Douglas O. Staiger**, "Parental Preferences and School Competition: Evidence from a Public School Choice Program," NBER Working Papers 11805, National Bureau of Economic Research, Inc December 2006.

Hoxby, Caroline M., "The Changing Selectivity of American Colleges," *Journal of Economic Perspectives*, 2009, *23* (4), 95–118.

_ **and Christopher Avery**, "The Missing "One-Offs": The Hidden Supply of High-Achieving, Low Income Students," NBER Working Papers 18586, National Bureau of Economic Research, Inc December 2012.

_ **and Sarah Turner**, "What High-Achieving Low-Income Students Know about College," *American Economic Review*, May 2015, *105* (5), 514–17.

MacLeod, W. Bentley and Miguel Urquiola, "Reputation and School Competition," *American Economic Review*, 2015, *105* (11), 3471–88.

Pallais, Amanda, "Small Differences that Matter: Mistakes in Applying to College," *Journal of Labor Economics*, 2015, *33* (2), 493–520.

Pathak, Parag, "The Mechanism Design Approach to Student Assignment," *Annual Review of Economics*, Sept 2011, *3*, 513–536.

Tables

Table 1: School Inputs and Students: Means and (SD)

VARIABLES	(1) IPN Schools	(2) UNAM Schools	(3) Other Schools
Panel A: Student Characteristics			
School Entry Score	1.378	1.299	-0.203
	(0.400)	(0.447)	(0.488)
SD of School Entry Score	0.478	0.385	0.639
	(0.0841)	(0.0614)	(0.119)
At least one parent has SHS	0.537	0.629	0.275
	(0.0734)	(0.111)	(0.105)
At least one parent is white collar	0.408	0.494	0.210
	(0.0707)	(0.101)	(0.0815)
Private JHS	0.121	0.195	0.0277
	(0.0504)	(0.110)	(0.0282)
Per capita household income	1,598	1,932	995.4
	(221.0)	(406.0)	(227.3)
Admitted students	1,187	2,470	582.9
	(340.6)	(1,055)	(457.0)
Panel B: School Inputs			
Class Size	39.59	24.65	42.23
	(2.782)	(4.690)	(5.207)
Share Teachers with College	0.856	0.943	0.826
	(0.133)	(0.0480)	(0.122)
Observations	16	14	258

Notes: School-level means and standard deviations (in parentheses) for the sample of students admitted in the first stage process of the Comipems assignment process to IPN high schools (1), UNAM high schools (2) and other Comipems high schools (3). There is missing information for 7 other Comipems schools. Source: school census data (Formato 911) 2005 and COMIPEMS 2005.

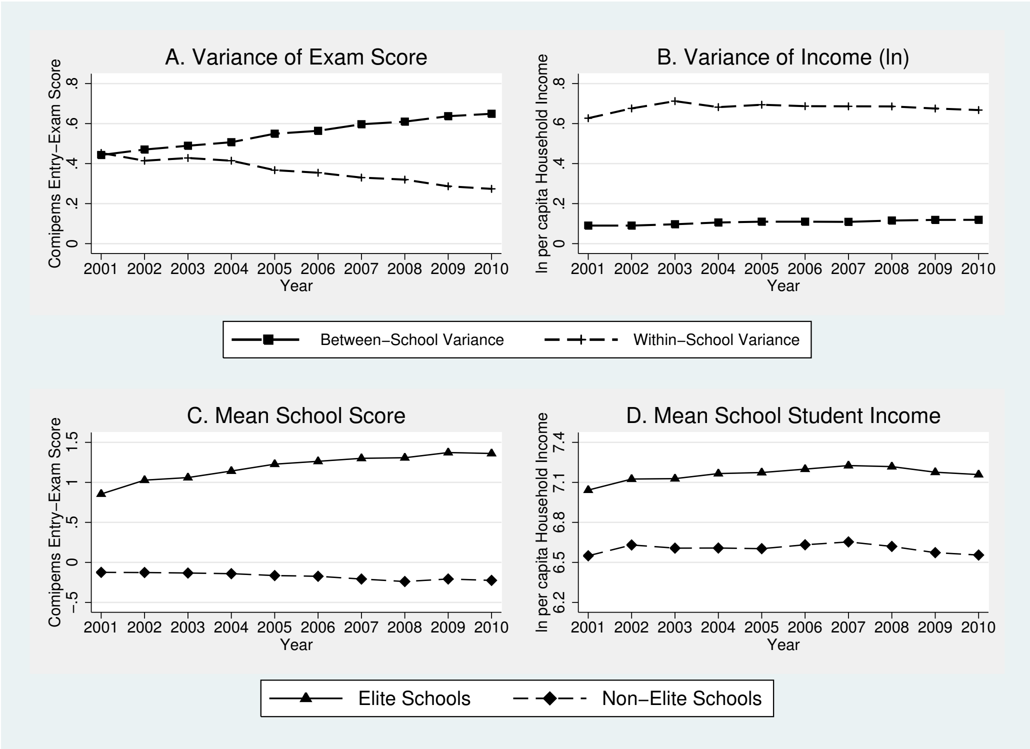
Table 2: Comipems Applicants: Means and (SD)

VARIABLES	(1) Full Sample	(2) Full Sample Minus Non-Response	(3) High-Income	(4) Low-Income
Comipems Score (std)	0.315	0.346	0.525	0.149
	(0.842)	(0.847)	(0.871)	(0.773)
Female	0.497	0.504	0.482	0.528
	(0.500)	(0.500)	(0.500)	(0.499)
Private JHS	0.0703	0.0709	0.120	0.0170
	(0.256)	(0.257)	(0.325)	(0.129)
Number of Chosen Options	8.847	8.908	8.908	8.908
	(3.775)	(3.771)	(3.847)	(3.687)
Submitted 20 options	0.0236	0.0241	0.0261	0.0219
	(0.152)	(0.153)	(0.160)	(0.146)
Selected an Elite School	0.760	0.771	0.829	0.707
	(0.427)	(0.420)	(0.377)	(0.455)
Share of Elite Schools	0.431	0.438	0.515	0.352
	(0.352)	(0.351)	(0.353)	(0.328)
First Choice is Elite School	0.637	0.650	0.730	0.561
	(0.481)	(0.477)	(0.444)	(0.496)
Number of Assigned Option	3.454	3.441	3.509	3.365
	(3.197)	(3.188)	(3.269)	(3.095)
Admitted into Elite School	0.267	0.281	0.365	0.188
	(0.443)	(0.449)	(0.482)	(0.391)
2nd Round Assignment	0.110	0.107	0.116	0.0964
	(0.313)	(0.309)	(0.321)	(0.295)
One parent has at least SHS		0.425	0.595	0.237
		(0.494)	(0.491)	(0.425)
At least one parent is white-collar		0.323	0.489	0.141
		(0.468)	(0.500)	(0.348)
PC Family Income		1,312	2,024	529.9
		(1,160)	(1,206)	(236.3)
Observations	2,024,493	1,647,078	862,737	784,341
Years	2001-2010	2001-2010	2001-2010	2001-2010

Notes: Means and standard deviations (in parentheses) for Comipems applicants in first 200,000 positions in priority queue.. Column A presents statistics for all these applicants and column B for those who reported parental schooling, occupation and household income. Columns C and D split the sample in Column 3 in applicants with a per capita household income above their cohort median (high-income) and below (low-income). Source: COMIPEMS 2001-2010.

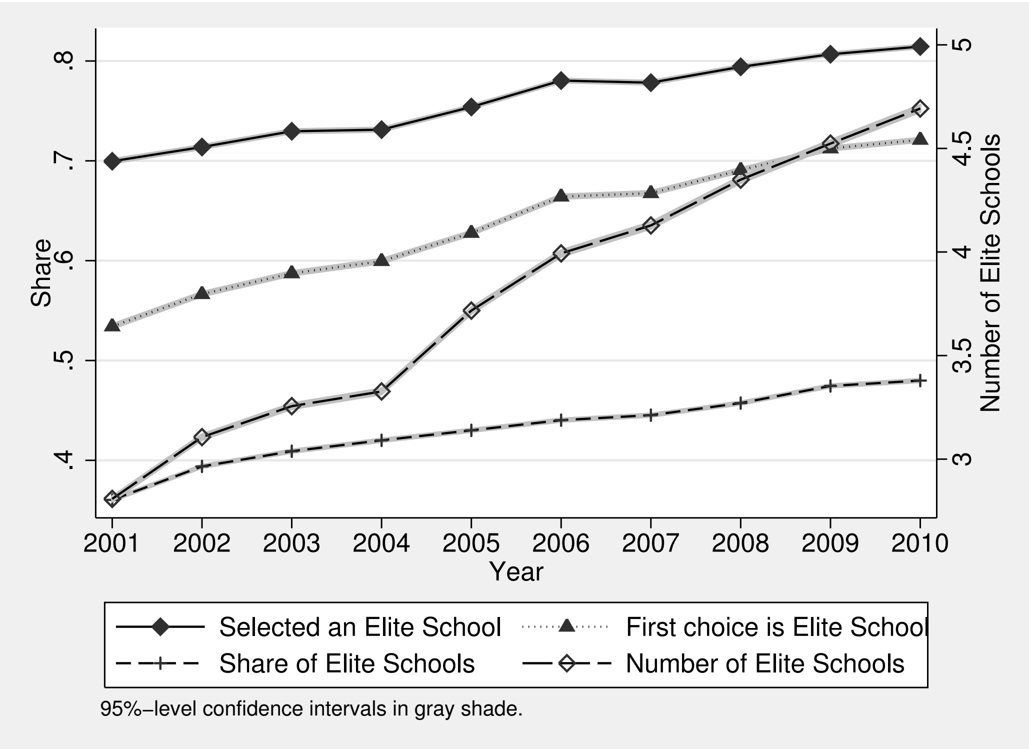
Figures

Figure 1: School Stratification by Academic Ability and Household Income: 2001-2010



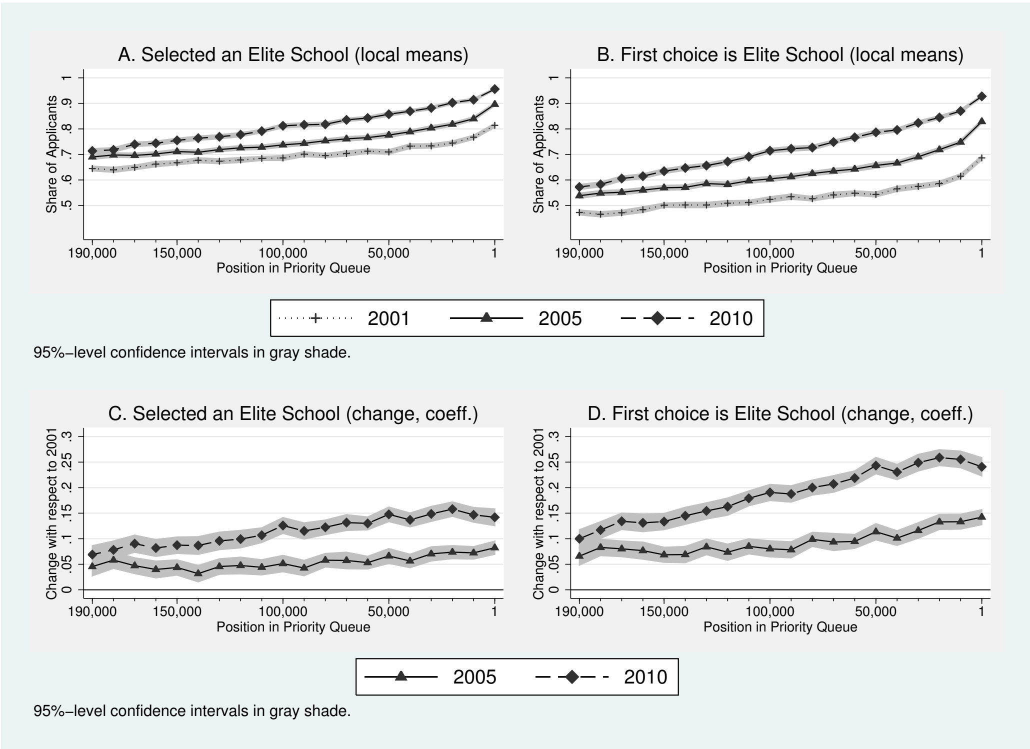
Notes: Panels A and B report, respectively, the between and within-school variance of the assigned applicants' entry-exam score and per capita household income (in natural logarithms). Panels C and D report, respectively, means by elite-school status of school-level means of the assigned applicants' entry-exam score and per capita household income (in natural logarithms). Source: COMIPEMS 2001-2010. Sample: Applicants allocated to a school in the first stage of the Comipems assignment process.

Figure 2: Demand for Elite Schools: 2001-2010



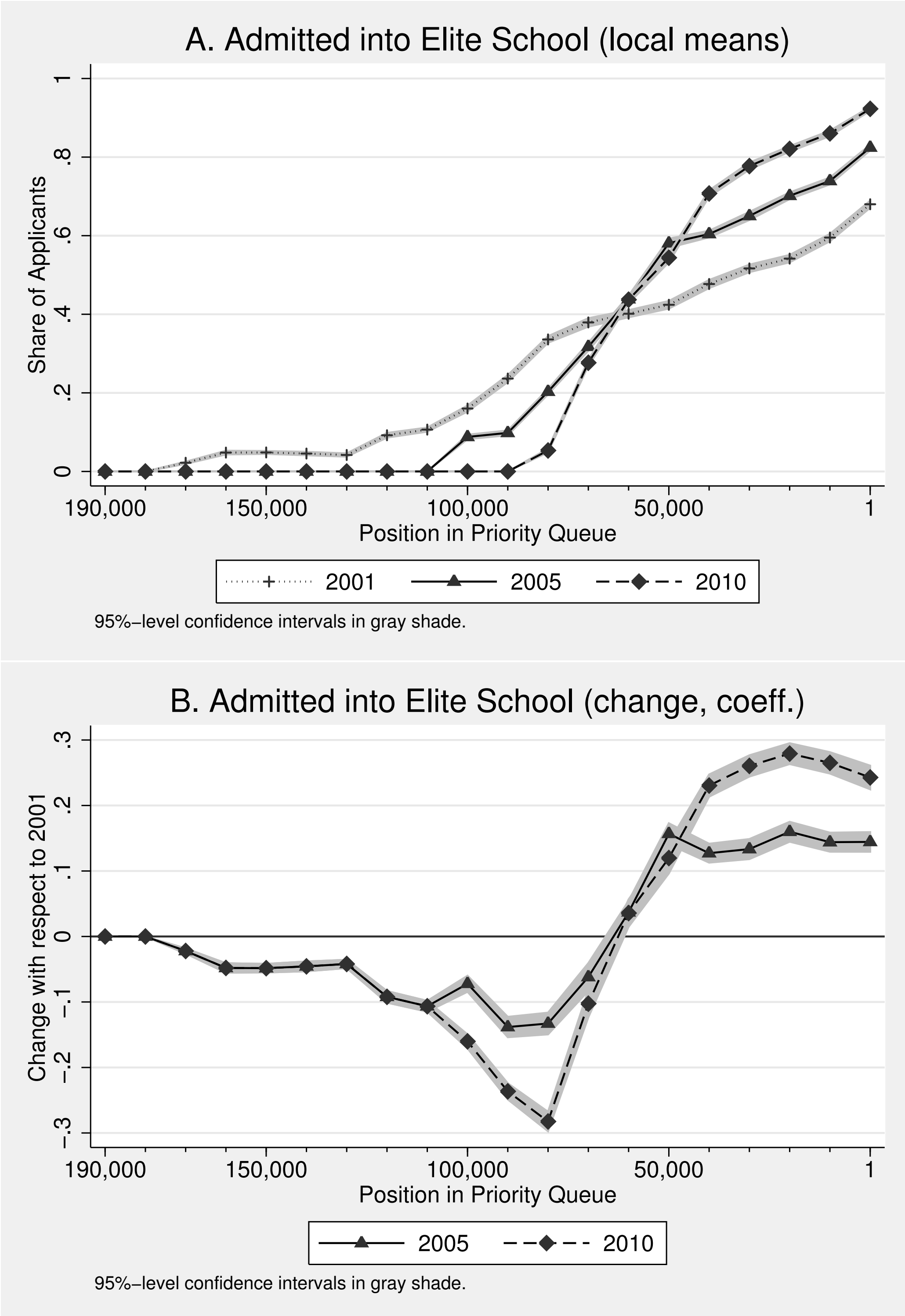
Notes: Conditional means by vintile of Comipems entry-exam score of the following variables: an indicator for whether the applicant selected at least an elite school in her choice list, an indicator for the selection of an elite school as first choice, and the share and number of elite schools in the applicant's choice list. Confidence intervals at the 95-percent level are reported in all cases. Source: COMIPEMS 2001-2010. Sample: Comipems applicants in first 200,000 positions in priority queue.

Figure 3: Demand for Elite Schools by Position in Priority Queue



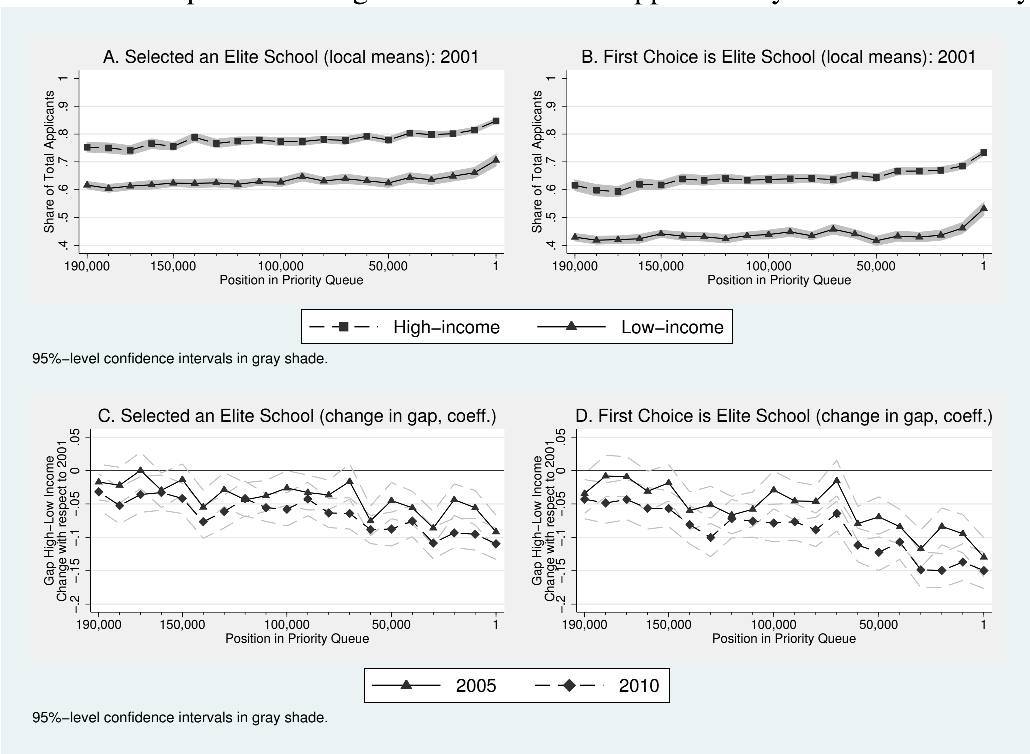
Notes: Panels A and B report, respectively, conditional means by vintile of the entry-exam score of an indicator for whether the applicant selected at least an elite school in her choice list and an indicator for the selection of an elite school as first choice. Panels C and D report regression coefficients for within-entry-exam-vintile changes in the same variables between 2001 and 2005 and 2010. Coefficients are obtained from, separate, OLS regressions of the outcomes on a vector of year dummies, a vector of extry-exam vintiles, a set of full interactions between these two vectors, and a gender indicator, an indicator for graduation from a private junior high school (JHS) and a vector of JHS-municipality dummies. Standard errors are clustered at the attended-JHS level in the regressions. Confidence intervals at the 95-percent level are reported in all cases. Source: COMIPEMS 2001-2010. Sample: Comipems applicants in first 200,000 positions in priority queue.

Figure 4: Admission to Elite Schools by Position in Priority Queue



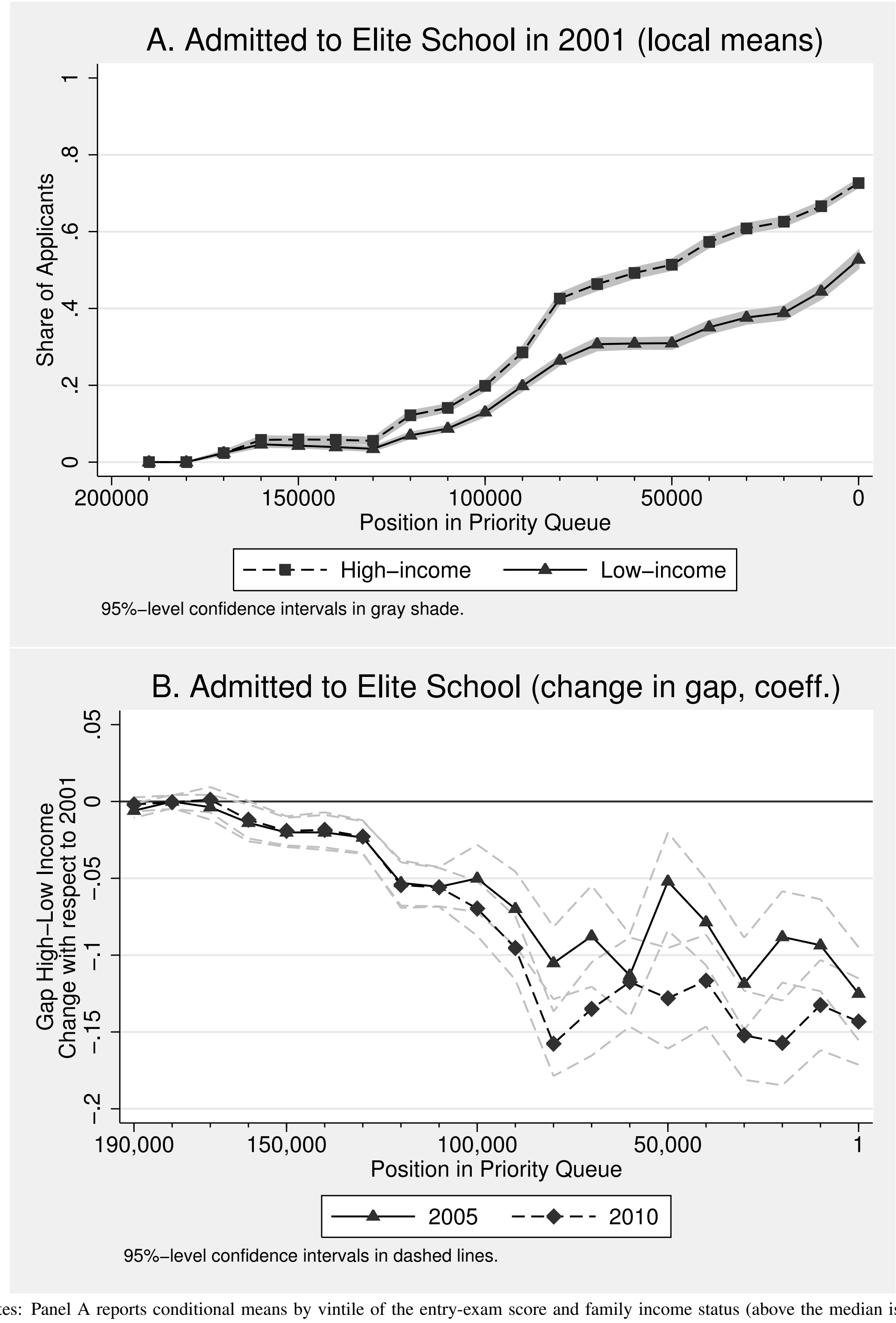
Notes: Panel A reports conditional means by vintile of the entry-exam score of an indicator for whether the applicant was assigned to an elite school. Panels B reports regression coefficients for within-entry-exam-vintile changes in the same variable between 2001 and 2005 and 2010. Coefficients are obtained from an OLS regressions of the outcome on a vector of year dummies, a vector of extry-exam vintiles, a set of full interactions between these two vectors, and a gender indicator, an indicator for graduation from a private junior high school (JHS) and a vector of JHS-municipality dummies. Standard errors are clustered at the attended-JHS level in the regression. Confidence intervals at the 95-percent level are reported in both cases. Source: COMIPEMS 2001-2010. Sample: Comipems applicants in first 200,000 positions in priority queue.

Figure 5: Demand Gap between High and Low-income Applicants by Position in Priority Queue



Notes: Panels A and B report, respectively, conditional means by vintile of the entry-exam score and family income status (above the median is high and below the median is low) of an indicator for whether the applicant selected at least an elite school in her choice list and an indicator for the selection of an elite school as first choice. Panels C and D report regression coefficients for within-entry-exam-vintile changes in the gap between high and low-income applicants in the same variables between 2001 and 2005 and 2010. Coefficients are obtained from, separate, OLS regressions of the outcomes on a vector of year dummies, a vector of extry-exam vintiles, a set of full interactions between these two vectors, an indicator for high-income status and a set of full interactions between this indicator and the previous vectors, and a gender indicator, an indicator for graduation from a private junior high school (JHS) and a vector of JHS-municipality dummies. Standard errors are clustered at the attended-JHS level in the regressions. Confidence intervals at the 95-percent level are reported in all cases. Source: COMIPEMS 2001-2010. Sample: Comipems applicants in first 200,000 positions in priority queue.

Figure 6: Admission Gap between High and Low-income Applicants by Position in Priority Queue



Notes: Panel A reports conditional means by vintile of the entry-exam score and family income status (above the median is high and below the median is low) of an indicator for whether the applicant was assigned to an elite school. Panels B reports regression coefficients for within-entry-exam-vintile changes in the gap between high and low-income applicants in the same variable between 2001 and 2005 and 2010. Coefficients are obtained from an OLS regressions of the outcome on a vector of year dummies, a vector of extry-exam vintiles, a set of full interactions between these two vectors, an indicator for high-income status and a set of full interactions between this indicator and the previous vectors, and a gender indicator, an indicator for graduation from a private junior high school (JHS) and a vector of JHS-municipality dummies. Standard errors are clustered at the attended-JHS level in the regressions. Confidence intervals at the 95-percent level are reported in all cases. Source: COMIPEMS 2001-2010. Sample: Comipems applicants in first 200,000 positions in priority queue.