The Impact of College Admissions Policies on the Academic Effort of High School Students*

Nicolás Grau Veloso†

Abstract

This paper empirically evaluates the effects of college admissions policies on high school students' academic effort. I build a rank-order tournament model where high school students decide their level of effort and whether or not to take the college admissions test, considering how those decisions affect their future university admissions chances. Using administrative Chilean data for the 2009 college admissions process, I structurally estimate the parameters of the model. Two affirmative action policies are simulated: (a) SES-quota system, which imposes the population's socioeconomic group (SES) distribution for each university; (b) increasing the weight of high school GPA in the admission final score. These simulations support the claim that affirmative action in college admission may boost the amount of academic effort exerted by high school students. I also find that while increasing the weight of high school GPA is more effective in boosting students' academic effort in high school, the SES-quota system is more efficient in allocating the best students to the best universities.

JEL *classification*: C38; C51; C54; D04; I23; I24.

Keywords: college admissions; affirmative action; high school student effort; structural estimation; factor models; ex-ante policy evaluation.

*I thank Diego Amador, Sina Ates, Garth Baughman, Flavio Cunha, Hanming Fang, Tzuo Hann Law, Andrés Hojman, Daniel Hojman, Zehao Hu, Tomas Rau, Minchul Shin, Petra Todd, Juan Pablo Torres-Martinez and Kenneth I. Wolpin for helpful discussions, as well as participants at the European Meeting of the Econometric Society (2013) and at seminars at University of Pennsylvania, Universidad de Chile, Universidad Católica de Chile, and Universidad Diego Portales. I also thank the Ministry of Education of Chile, DEMRE, Paola Bordón, and José Miguel Benavente for providing me with the data used in this paper. Raimundo Smith and Bastián Gómez provided outstanding research assistance for this project. Funding from the Centre for Social Conflict and Cohesion Studies (CONICYT/FONDAP/15130009) is gratefully acknowledged. All errors are mine.

†Department of Economics, Universidad de Chile. Diagonal Paraguay 257, of 1504, Chile. (Email: nicolas.grau@gmail.com).

1 Introduction

There is a continuing debate about how to reduce socio-economic and racial segregation in universities. To this end, many countries have affirmative action programs, intended to increase college admission rates for targeted populations (e.g., specific races or SES). In general, existing evaluations of these programs focus on the application rates of students benefiting from affirmative action, and the academic performance of those who are admitted.[1] However, the existing evaluations generally assume high school student behavior to be exogenous, which overlooks the potential impacts of these programs on the motivation of high school students.

Given that high school students may consider the impact of their effort levels on their university admissions chances and react to different admissions policies accordingly, the goal of this paper is to empirically address the effect of college admissions on high school student academic effort in response to policy changes. In particular, I estimate the structural relationship between college admissions policies, which determine the probabilities of being admitted by different universities, and the student's effort decision in high school. Then I use the estimated model to simulate the effect of different affirmative action policies on the academic effort of high school students.

I model the college admissions process and high school behavior in a static fashion, where students decide how much academic effort to make during high school, and whether to take the national college admissions test, which is mandatory for college applicants. The exerted effort positively impacts the expected performance in high school and on the college admissions test. In the model, there are different universities, each one offering two majors: scientific and humanities. Future payoff (after college graduation) depends on the university quality and the major studied. Because in the model universities have fixed and exogenous amount of seats for each major, the admissions process works as a rank-order tournament, such that the access of each student to universities depends on her individual performance relative to the performance of all the other students.[2] Admissions policies are based on a linear combination of high school grades and test scores that form a final score such that the equilibrium of the model is characterized by a minimum final score to be admitted in each major/university, named final-score cutoff. Intuitively, this final-score cutoff vector has a similar role as prices in a Walrasian equilibrium, in the sense that its value is set such that the number of students admitted to each university is equal to its number of seats, and that conditional on this vector every student is making an optimal decision about taking the test and about how much

[1] See, for example, Arcidiacono (2005), Bowen and Bok (1998), Card and Krueger (2005), Epple et al. (2006), and Long (2004). A summary of the literature before 2000 can be found in Holzer and Neumark (2000); for a more recent survey, see Arcidiacono et al. (2016).

[2] To read more about the theoretical implications of rank order tournament, see Lazear and Rosen (1981).

academic effort to make.

The model is estimated using Chilean administrative data for the 2009 college admissions process. The database, which has $146,319$ observations, contains individual information such as the scores of tests taken between 8th and 12th grade, measures of academic effort (e.g., attendance and GPA) and learning skills, and characteristics of families and of primary and secondary schools. Moreover, the database also includes Ministry of Education's data from tax declarations which links individual post graduation wages to students' scores on the university admission test. The exceptional features and richness of this database are crucial in estimating the structural relationship between high school students' effort and their probabilities of admission to universities with different qualities.

The model estimation is carried out in two stages. In the first stage, I estimate all the parameters of the test production function by two-stage least squares, since I have more than one measure for the endogenous variable (i.e., high school student effort). In the second stage, using some parameters estimated in the first stage, I estimate the utility parameters, the distribution of the unobserved learning skills, and the parameters of the measurement equations by a maximum likelihood procedure. I follow this approach mainly because most of the parameters are estimated in the first stage, leaving just a few parameters to be estimated in the second stage, which is more time consuming. Because from data I observe the final-score cutoffs (the equilibrium object), in the estimation the model can be approached as a single agent problem, which simplifies estimation and makes this approach robust to any multiple equilibria potential problem. Notice, however, that in all my simulation exercises, to study the fit of the model or to study the effect of counterfactual experiments, I do need to solve the equilibrium of the rank-tournament model, by finding the final-score cutoffs of equilibrium. Overall, the simulation of the estimated model fits most of the data features very well.

To study what the impact of college admissions policies is on the academic effort of high school students, two types of affirmative action policies are simulated. The first one is a SES-quota system, which imposes the population's SES distribution for each university. In the second policy experiment, I simulate what happens if the GPA weight is increased, which in practice implies that there is an increased probability of attending better universities for those students attending low income high schools. This is due to the fact that while the high school GPA of each student is to some extent relative to her classmates, the national test scores are relative to the student's national cohort and therefore capture the difference in high school quality, which is highly correlated with income.

There are several lessons from these counterfactual experiments. (1) Average effort significantly increases as opportunities are equalized across different socioeconomic groups (SES). (2) This leads to a moderate improvement in high school students' performances, which is relatively impactful for some groups. (3) Although the effects on performance are moderate, the evidence supports the idea that modeling effort and the decision to take the admission test are important in order to anticipate what would happen with the main features of the college admissions system (e.g., student allocation).(4) The highest change in exerted effort comes from those students who also change their decision about taking the college admissions test. (5) Neither of these policies importantly increases the percentage of students taking the national college admissions test, which is consistent with the fact that in this policy implementation there are winners and losers. However, there are relevant variations in who is taking such a test; in particular, this percentage increases for low-income students and those who have a higher level of learning skills. (6) Although increasing the GPA weight is more effective on boosting students' academic effort in high school, the SES-quota system is more capable of allocating the best students to the best universities, conditional on delivering the same universities' socioeconomic composition.

It is worth mentioning that there is nothing in the model that ensures that any design of an affirmative action policy would positively impact the average of high school academic effort. In the model, the incentives for exerting effort are highly non-linear. In one extreme case, students exert the highest effort when their expected final scores are very close to a final-score cutoff. In the other extreme, students exert the lowest level of effort when effort does not change their college admission performance, either because they have no chance to be admitted to a university or because they have very high probability of being admitted to a top university, without making any additional effort. Yet these two extreme cases are more probable when students have very different backgrounds. In this context, the affirmative action policies studied in this paper have a positive impact on average effort because they make students to compete in more homogeneous groups without breaking the connection between admission probabilities and effort.

Regarding the literature, there are four empirical papers that, in the cotenxt of affrmative action policies, take students' behavior in high school as endogenous, as I do in this paper.[3] The first three, Cotton et al. (2014), Domina (2007) and Ferman and Assunçâo (2011) present some reduced form estimations that address how changes in affirmative action policies may change students' behavior in high school.[4] In the fourth paper, which is the closest to my research, Hickman (2013) models the behavior of U.S. high school students as a function of their future chances of being admitted to different universities.

In particular, Cotton et al. (2014) using experimental data from a high school tournament

[3] In a related paper, Hastings et al. (2012) show how motivation can change students' exerted effort, in particular that the opportunity to attend a better high school has positive and significant effects on both student attendance and test scores. Theoretically and motivated by U.S. legal changes, a series of papers, e.g., Chan and Eyster (2003); Fryer et al. (2008); and Hickman (2011) have focused on how the prohibition of explicit consideration of race in the admissions process may be quite inefficient if colleges still have some preferences toward minorities. Another interesting theoretical paper is Gall et al. (2013).

[4] In the case of Cotton et al. (2014) they base their analysis on a theoretical paper, namely, Bodoh-Creed and Hickman (2014).

that resembles affirmative action policies, show how affirmative action not only may promote racial diversity on college campuses, but may also narrow achievement gaps between advantage and disadvantage groups by motivating higher levels of pre-college human capital investment on the part of under-represented minority students. In the same line, Domina (2007), using panel data for Texan high schools between 1993 and 2002, shows evidence that Texas' post-Hopwood higher education policies boosts high school students' academic engagement at public schools.[5] Opposing this is Ferman and Assunçâo (2011), who used difference-in-difference techniques and quasi-experimental data from Brazilian secondary education, where political forces abruptly imposed an admissions quota for two of Rio de Janeiro's top public universities. They estimate that the quota altered incentives, thus producing a 5.5% decrease in standardized test scores among the favored group, widening the achievement gap by 25%.

These studies show how different ways of increasing the admissions probabilities of the most segregated groups may have different impacts on high school students' behavior. However, a structural approach is required in order to know which admissions policies accomplish an efficient combination of increased diversity and correct incentives. To address this issue, Hickman (2013) uses U.S. data to structurally estimate a model of college admissions where the admissions test is an endogenous variable, using empirical tools borrowed from auction literature.[6] One of his main findings is that current affirmative action policies narrow the achievement and enrollment gaps, but a color blind system results in higher academic achievement in the overall student population. His other finding is that the quota system prohibited by U.S. law is superior to both of the other policies in three dimensions: it produces the highest academic performance; it substantially narrows the achievement gap; and, by design, it closes the enrollment gap completely.

Beyond technicalities, the main differences between my paper and Hickman (2013) are: (1) Given that I have data for the student regardless if she did or did not take the college admissions test, I can see how different admissions rules change the number of people who apply to college, whereas his approach is conditional on admission. Furthermore, it turns out that in my estimation and, hence, in my simulations this decision plays a central role. (2) Given that I observe measures of effort and a set of variables which determine the student performance in my data, the impact of the effort decision is established in a more transparent way, and it is possible to compare the magnitude of the effort's effect with that of the other determinants. Note, however, the differences in our approaches are

[5] These policies include a guarantee that all students who finished in the top 10% of their high school class will be admitted to their chosen public university. Cullen et al. (2013) show that, after the implementation of this Texan policy, among the subset of students with both motive and opportunity for strategic high school choice, at least 5% enroll in a different high school to improve the chances of being in the top 10%.

[6] The model is described in detail in Hickman (2011).

mainly motivated by different access to data and the particular traits in the institutional design of the two educational systems (American and Chilean).

In this context, my paper has three main contributions. First, by modeling how universities' admissions systems can change incentives to exert more effort from high school students, this paper shows that increasing equality in opportunities may lead to a boost in this average effort. Moreover, given that the results of this paper are that the most significant changes in academic effort come from those students who also change their decision about to take the university admission test, this paper represents a relevant improvement with respect to the previous literature that studies the effort of high school students conditional on attending college (e.g., Hickman (2013)). Second, to the best of my knowledge, this is the first paper that estimates a rank-order tournament with heterogeneous ability contestants. Indeed, Vukina and Zheng (2007) present the first estimation of a structural model of an empirically observed rank-order tournament as a strategic game with private information and homogeneous ability contestants. As the authors posit, the structural estimation of rank-order tournament games with heterogeneous ability contestants is cumbersome as this assumption results in equilibrium strategies that are nonsymmetric. In the case of my model, the complexity of the problem is simplified by assuming that there exists a finite set of individual types, with a continuum mass of students of each type, realistic assumption in the context of the Chilean admission system. Third, the paper exploits the interaction between economic theory and factor analysis models in the identification and estimation of the model.

The paper proceeds as follows. Section 2 details the features of the model. Section 3 describes the Chilean college admissions process, explaining the main features of the data. Section 4 discusses the empirical implementation of the model and proves the identification of the model's parameters. Section 5 presents the estimation procedure. In Section 6, the model fit is discussed along with other aspects of the estimation results. Section 7 describes the counterfactual experiments results. Finally, Section 8 concludes and discusses future research.

2 The Model

The aim of this model is to capture in a rank-tournament framework how college admissions policies may affect the effort exerted by high school students. The basic idea of the model is that students compete for the (fixed and pre-determined) slots in the best universities through their performance in high school GPA and admission test scores. Because for each individual what matters is her relative performance with respect to the other students, in principle the rank-tournament models with heterogeneous agents are very difficult to solve. However, I simplify the solution of the model by assuming that there exists a finite set of individual types, with a continuum mass of student of each type. This assumption has two important consequences: 1) students can anticipate without uncertainty what will be the minimum score to be admitted at each university tier, and 2) there is not a strategic consideration in students' optimization. As It will be clear when I describe the admission system to college in Chile, this is a reasonable assumption in the context where this model is implemented.

In this model students have two decisions to make: whether or not to take the college admissions test, a necessary input for university admittance; and how much effort to make during high school.[7] The exerted effort is denoted by the continuos variable, e, and to take the college admissions test is denoted by the binary variable, $TCAT$, such that $TCAT$ is equal to one is she takes the test and to zero, otherwise. These are modelled as two simultaneous decisions. The exerted effort positively impacts expected high school and college admissions test performance.[8] For those students who decide to take the college admissions test, admissions policies consider both high school grades and the test score, such that higher measures lead to admittance by better universities.

Each student i is characterized by her individual characteristics and the characteristics of the high school she is attending. One of the individual characteristics is her preference over two group of college majors: scientific or humanities. All these variables, including the preference for a major, are exogenously determined. Because the distinction is not relevant and to simplify the exposition, throughout the paper I do not distinguish between individual and high school characteristics, and I refer to all them as individual characteristics. Individuals characteristics are denoted by $\{X, Z, \lambda\}$: X is a vector including all the observable characteristics but the preference for a major, which is variable Z; λ is the student learning skill, a scalar variable that is unobservable by the econometrician, but known by the student. As presented below, λ is an important variable that impacts all the tests and scores, and the cost of taking the admission test. Thus, it is an unobserved variable that allows to have correlation among tests scores and individual decisions, conditional on observables. Although from the model's perspective it does not make any difference what is and is not observed by the econometrician, since it is relevant in the empirical section, I introduce this notation in the model description to keep all notations consistent across the paper.

It is assumed that there is a finite space of individual characteristics (i.e., X, Z, λ are

[7] This paper is connected to the literature that models discrete-continuous choices. See, for instance, Hanemann (1984), Dubin and McFadden (1984), Chintagunta (1993), and Dillon and Gupta (1996).

[8] There is a debate about the impact of student academic effort on student performance. For example, Schuman et al. (1985) report four different major investigations and several minor ones over a decade, none of which were very successful in yielding the hypothesized substantial association between the amount of study and GPA. Such an unexpected result is, from different angles, contradicted by Eckstein and Wolpin (1999), Eren and Henderson (2008), Rau and Durand (2000), Stinebrickner and Stinebrickner (2004), and Stinebrickner and Stinebrickner (2008). Related to this literature is the difficulty of having a proper model for cognitive production function. In this regard, Todd and Wolpin (2007) find the most support for the value-added models, particularly if those models include some lagged input variables (see also Todd and Wolpin (2003)).

discrete) where all of the students who share the same characteristics belong to the same student type. Let $t(i) \in \{1, 2, ..., T\}$ denote the type of student i; such that, $\forall\ i, j \mid t(i) = t(j)$: $X_{t(i)} = X_{t(j)}$, $Z_{t(i)} = Z_{t(j)}$, and $\lambda_{t(i)} = \lambda_{t(j)}$. The mass of type t students is denoted by m_t. As discussed bellow, the finite space with a continum of students assumption implies that even though this is a rank-tournament model, there is not any strategic consideration in students' behavior.

There are $N-1$ university tiers, each one offering two majors: scientific and humanities. Because they have different quality levels, each university n implies some specific future pay-off for each major z, $\{R_1^z, R_2^z, ..., R_N^z\}$, such that $R_{n+1}^z > R_n^z\ \forall\ n, z$ and R_1^z is the pay-off for those who prefer major z, but they were not admitted to any college (because they did not try or their final score was too low). Since individual types include the information about the preferred major, $R_n^{Z(t(i))}$ denotes the payoff that individual i obtains if she attends university n.

Each university/major (n/z) has a fixed and exogenous amount of seats S_n^z ($S_1^z > 0$ is the residual: the mass of students preferring major z who are not admitted to any college, i.e., $\sum_t m_t^z = \sum_{\delta=1}^N S_\delta^z$). Given these slots, the admissions process works as a rank-order tournament in which students decide their effort (e_i) and whether or not to take the college admissions test $(TCAT_i)$, taking into account the effort cost, the test's fixed cost $(FC_i \sim N(\overline{FC}_{t(i)}, \sigma_{fc}^2))$, how much they value future payoffs, and their chances of being admitted by each university in their prefered major. Notice that fixed cost's normality implies negative "cost" for some students. This assumption simplifies the equilibrium solution. That said, the existence of negative costs for some students is equivalent of the cost being equal to zero, namely, all of them will take the admission test.

Let FS_i be the type i college admissions final score, such that:

$$FS_i = P_{pm} * PM_i + P_{pv} * PV_i + P_g * GPA_i, \tag{1}$$

where PM_i, PV_i and GPA_i are the math test, the verbal test, and the high school GPA, respectively; whereas P_{pm}, P_{pv} and P_g are the associated weights. The production function of these tests are:

$$PM_i = \beta_0^{pm} + X_{t(i)}\beta_1^{pm} + e_i\beta_2^{pm} + \lambda_{t(i)}\beta_3^{pm} + \varepsilon_i^{pm}, \tag{2}$$

$$PV_i = \beta_0^{pv} + X_{t(i)}\beta_1^{pv} + e_i\beta_2^{pv} + \lambda_{t(i)}\beta_3^{pv} + \varepsilon_i^{pv}, \tag{3}$$

$$GPA_i = \beta_0^{g} + X_{t(i)}\beta_1^{g} + e_i\beta_2^{g} + \lambda_{t(i)}\beta_3^{g} + \varepsilon_i^{g}. \tag{4}$$

$(\varepsilon_i^{pm} \varepsilon_i^{pv} \varepsilon_i^{g}) \sim N(0, \Gamma)$, $E[\varepsilon_i^k | X_{t(i)}, \lambda_{t(i)}] = 0$, $\forall\ k \in \{pm, pv, g\}$, and

$$\Gamma = \begin{bmatrix} \sigma_{pm}^2 & 0 & 0 \\ 0 & \sigma_{pv}^2 & 0 \\ 0 & 0 & \sigma_g^2 \end{bmatrix}.$$

Thus, λ is a single-dimensional unobserved variable, which allows for –conditional on X and e– correlated errors across the performance equations for GPA, PM, and PV, despite the assumption of the independence of their shocks.

Given the number of people who actually take the college admissions test, the seats offered by each university/major, and the final score distribution of those students, the vector r_z ($\{r_2^z, r_3^z, ..., r_N^z\}$) represents the final minimum score needed to be admitted by each university tier in major z. Throughout the paper, I denote this vector as the *final-score cutoff*. Hence, the students who are going to university n in major z are those who prefer major z and have a final score greater than or equal to r_n^z and smaller than r_{n+1}^z. The former inequality is given by the admissions rule, whereas the latter is due to utility maximization, since it is always suboptimal for students to attend university n for major z when their final score is greater than r_{n+1}^z.

Importantly, vector r^z is what characterize the equilibrium in the market of major z. As it is formalized later, an equilibrium requires that conditional on this vector r^z, all students preferring major z are optimizing their effort decision and their decision about taking the admission test. Moreover, this vector r^z has to be the result of these optimally students' decision.

I assume that students have lexicographic preferences. In particular, those who prefer a major in science (humanities) will not major in humanities (science) no matter what the payoff. This implies that scientific and humanities majors work as separated markets. Therefore, an equilibrium for each major can be defined in a isolated way. Notice that this is a simplification in respect to reality where although students may have strong preferences about majors, they may apply to a less preferred major if by doing so they can attend a better university. I consider this separated market by major assumption in order to have a tractable solution to my model.

The utility function, for those who choose to not take the college admissions test, is given by:

$$U_{t(i)}^0(e) = R_1^{Z(t(i))} + \theta_1 GPA(e_{t(i)}) - \theta_2^{t(i)} \frac{e^2}{2}, \tag{5}$$

For those who decide to take the college admissions test, the utility is:

$$U^1_{t(i)}(e) = \sum_{n=1}^{N} R_n^{Z(t(i))} 1(r_n^{Z(t(i))} \leq FS(e_{t(i)}) < r_{n+1}^{Z(t(i))}) + \theta_1 GPA(e_{t(i)}) - FC_i - \theta_2^{t(i)} \frac{e^2}{2}, \quad (6)$$

where $r_1 = -\infty$ and $1(A)$ is an indicator function that takes the value of 1 when A is true and 0 otherwise. In both utilities the parameter associated with future pay-off is normalized to one and θ_1 and $\theta_2^{t(i)}$ represent the importance of high school student performance and the cost of effort, respectively. The cost of effort is quadratic and heterogeneous across individuals.

There are two considerations to be made about students' utility function. On one hand, students make their effort decision before the realization of the GPA, PM, and PV shocks. Therefore, they maximize expected utility. The only private information used in students' decisions is the value of FC_i. The distributions of GPA, PM, PV and FC are common knowledge. On the other hand, all information about the other students that each one needs in order to make her effort decision are the values of r. Moreover, due to the fact that each student anticipates the behavior of other students and that there is a continuum of individuals of each type, the value of the vector r is predicted without uncertainty, even though the individual final score is a random variable. The latter reveals the importance of the assumption on the continuum of individuals of each type, namely, due to this assumption the decision of each student does not have strategic considerations.

2.1 Student's Problem

Because at the time of their decision making, all the students of the same type only differ in the cost of taking the test conditional on taking the test, students of the same type exert the same level of effort. Hence, given a vector $r^{z(t)}$ (*i.e.,* the final-score cutoff for major z) , the optimization problem for those students of type t who do and do not take the national college admissions test can be written as, respectively:[9]

$$\max_{e\geq 0} E[U^0_t(e)] \Leftrightarrow \max_{e\geq 0} \left\{ \theta_1 b_1 e - \theta_2^t \frac{e^2}{2} \right\},$$

$$\max_{e\geq 0} E[U^1_t(e)] \Leftrightarrow \max_{e\geq 0} \left\{ \sum_{n=1}^{N-1} \left(R_n^{z(t)} - R_{n+1}^{z(t)} \right) \Phi \left(\frac{r_{n+1}^{z(t)} - a_1 e - a_{0t}}{\sigma_\eta} \right) + \theta_1 b_1 e - \theta_2^t \frac{e^2}{2} \right\}. \quad (7)$$

[9] Φ denotes the standard normal distribution function. In both cases, I do not include the elements of the expected utility that are not a function of effort.

Where:

$$
\begin{aligned}
a_{0t} &= P_{pm} * (\beta_0^{pm} + X_t\beta_1^{pm} + \lambda_t\beta_3^{pm}) + P_{pv} * (\beta_0^{pv} + X_t\beta_1^{pv} + \lambda_t\beta_3^{pv}) + P_g * (\beta_0^{g} + X_t\beta_1^{g} + \lambda_t\beta_3^{g}),\\
a_1 &= P_{pm} * \beta_2^{pm} + P_{pv} * \beta_2^{pv} + P_g * \beta_2^{g},\\
b_{0t} &= \beta_0^{g} + X_t\beta_1^{g} + \lambda_t\beta_3^{g},\\
b_1 &= \beta_2^{g},\\
\eta_i &= P_{pm} * \varepsilon_i^{pm} + P_{pv} * \varepsilon_i^{pv} + P_g * \varepsilon_i^{g},\\
\sigma_\eta^2 &= Var(\eta_i).
\end{aligned}
$$

The final effort decision for students i, who belong to type t ($i.e.$ $t(i) = t$) is $\hat{e}_i = (1 - TCAT_i) * \hat{e}_t^0 + TCAT_i * \hat{e}_t^1$, where $TCAT_i$ is the decision of individual i on taking the test:

$$
TCAT_i = \begin{cases} 1 & \text{if } \max_{e\geq 0} E[U_i^1(e)] \geq \max_{e\geq 0} E[U_i^0(e)], \\ 0 & \text{if } \max_{e\geq 0} E[U_i^1(e)] < \max_{e\geq 0} E[U_i^0(e)]. \end{cases} \tag{8}
$$

Notice that $E[U_{t(i)}^1(e)] + FC_i > E[U_{t(i)}^0(e)]$ for any individual and for any level of effort, therefore without the test's fixed cost, all students would take the test.

The solution of the problem for student i, who belongs to type t and prefers major $z(t)$, is characterized by the following first order conditions:

For those who do not take the college admissions test:

$$
\hat{e}_t^0 = \frac{\theta_1 b_1}{\theta_2^t}. \tag{9}
$$

For those who take the college admissions test:[10]

$$
\hat{e}_t^1 = \frac{1}{\theta_2^t}\left[\sum_{n=1}^{N-1}\left(R_{n+1}^{z(t)} - R_n^{z(t)}\right)\phi\left(\frac{r_{n+1}^{z(t)} - a_1\hat{e}_t^1 - a_{0t}}{\sigma_\eta}\right)\frac{a_1}{\sigma_\eta} + \theta_1 b_1\right], \tag{10}
$$

$\Rightarrow$

$$
TCAT_i = \begin{cases} 1 & \text{if } D_t \geq FC_i \\ 0 & \text{if } D_t < FC_i, \end{cases} \tag{11}
$$

[10] ϕ denotes the standard normal density.

$$
D_t = \left(\sum_{n=1}^{N-1} \left(R_n^{z(t)} - R_{n+1}^{z(t)} \right) \Phi \left(\frac{r_{n+1}^{z(t)} - a_1 \hat{e}_t^1 - a_{0t}}{\sigma_\eta} \right) \right) + (R_N^{z(t)} - R_1^{z(t)})
$$
$$
+ \theta_1 b_1 (\hat{e}_t^1 - \hat{e}_t^0) - \theta_2^i \frac{(\hat{e}_t^1)^2 - (\hat{e}_t^0)^2}{2}.
$$

It is direct that $E[U_t^0]$ is strictly concave. A sufficient condition for strict concavity of $E[U_t^1]$ is given by $(R_N^{z(t)} - R_1^{z(t)}) a_1^2 \phi(1) < \sigma_\eta^2 \theta_2^t$, $\forall t$ (Appendix A.1). The intuition of this condition is that in order to have only one local maximum for students' maximization, it is required that the impact of effort on expected utility is not too high relative to the cost of effort and the variance of final score. When this condition is fulfilled, as it happens given the estimated parameters of this model, the solution to (10) is unique and $\hat{e}_t^1$ is continuous in r, which is always the case for $\hat{e}_t^0$. This continuity condition is important for the equilibrium analysis. That said, it is possible to ensure existence of a solution for the student's problem even without this condition, as I prove by Lemma 1.

Lemma 1: Given a vector $r^{z(t)}$, the student's problem (7) has at least one interior solution.

Proof: Lets define the function f such that:

$$
f(e) = e - \frac{1}{\theta_2^t} \left[\sum_{n=1}^{N-1} \left(R_{n+1}^{z(t)} - R_n^{z(t)} \right) \phi \left(\frac{r_{n+1}^{z(t)} - a_1 e - a_{0t}}{\sigma_\eta} \right) \frac{a_1}{\sigma_\eta} + \theta_1 b_1 \right].
$$

In this context, proving the existence of one interior solution is equivalent to showing that there exists at least one value of e that is not equal to zero, such that $f(e) = 0$. To do so, I first notice that $f(e)$ is continuous and well defined for any value of e. On one hand, because ϕ is a density (*i.e.* cannot be negative), $\theta_1 b_1 > 0$, and $\theta_2^t > 0$; it is true that $f(0) < 0$. On the other hand, because $\phi(x)$ is bounded by one, for any value of x, it is also true that $\lim_{e \to \infty} f(e) = -\infty$. Therefore, by continuity, there exists at least one value of e, not equal to zero, such that $f(e) = 0$. Moreover, if we define $\underline{e}$ as $\underline{e} \equiv \theta_1 b_1 = \lim_{r \to -\infty} f(\hat{e}_t^1)$ and $\overline{e}$ as $\overline{e} \equiv \frac{a_1}{\sigma_\eta} \left(R_{n+1}^t - R_n^t \right) + \theta_1 b_1$, we know that it should be the case that $\hat{e}_t^1 \in [\underline{e}, \overline{e}]$. ■

2.2 Equilibrium

Let $\tilde{m}_t$ be the mass of students of type t who take the college admissions test, then:

$$
\tilde{m}_t = m_t \Phi \left(\frac{D_t - \overline{FC}_t}{\sigma_{fc}} \right)
$$

As we stated above, even though students just observe their own fixed cost realization, this mass can be predicted without uncertainty by the students due to the continuum of individuals in each type.

An equilibrium in the market of major z is given by a set of vectors $\{\hat{e}_t^0\}_{t\in\{t|z(t)=z\}}$, $\{\hat{e}_t^1\}_{t\in\{t|z(t)=z\}}$ and $\hat{r}^z$, such that:

- Given $\hat{r}^z$, $\forall\ i \mid t(i)\ \in \{t|z(t) = z\}$:
  - $\hat{e}_t^0 = \frac{\theta_1}{\theta_2^t} b_1$,
  - $\hat{e}_t^1 = \frac{1}{\theta_2^t}\left[\sum_{n=1}^{N-1}\left(R_{n+1}^{z(t)} - R_n^{z(t)}\right)\phi\left(\frac{\hat{r}_{n+1}^z - a_1\hat{e}_t^1 - a_{0t}}{\sigma_\eta}\right)\frac{a_1}{\sigma_\eta} + \theta_1 b_1\right]$,
  - $\hat{D}_{t(i)} = (E[U_i^1(\hat{e}_t^1, \hat{r}^z)] + FC_i) - E[U_i^0(\hat{e}_t^0)]$.
- $\forall\ n = 1, ..., N-1$:

$$\sum_{\delta=n+1}^{N} S_\delta = \sum_t \tilde{m}_t\left[1 - \Phi\left(\frac{\hat{r}_{n+1}^z - \hat{e}_t^1 a_1 - a_{0t}}{\sigma_\eta}\right)\right]$$
$$= \sum_t m_t \Phi\left(\frac{\hat{D}_t - \overline{FC}_t}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{\hat{r}_{n+1}^z - \hat{e}_t^1 a_1 - a_{0t}}{\sigma_\eta}\right)\right].$$

Notice that in this setup the vector $\hat{r}^z$ has a similar role as prices in a Walrasian equilibrium, in the sense that its value is set such that the number of students admitted to each university is equal to its number of seats.

Lemma 2: If $\forall\, i \mid t(i) = t:\ (R_N^{z(t)} - R_1^{z(t)}) < \left(\frac{\sigma_\eta}{a_1}\right)^2 \frac{\theta_2^t}{\phi(1)}$ and $\sum_t m_t \Phi\left(\frac{(R_N^{z(t)} - R_1^{z(t)}) - \overline{FC}_t}{\sigma_{fc}}\right) > \sum_{\delta=2}^{N} S_\delta$, there exists at least one equilibrium in market of z major.

Proved in Appendix A.1.

The sufficient conditions for existence have clear interpretations. On one hand, the first condition implies that the effort decision cannot be overly important for the final score determination (given by the ratio $\frac{a_1}{\sigma_\eta}$) and that the differences in the future pay-offs cannot be overly relevant (given by $\frac{1}{\theta_2^t}(R_N^{z(t)} - R_1^{z(t)})$). Hence, to be sure about the equilibrium existence requires that the impact of the effort on the utility is moderate. On the other hand, the second condition is more innocuous and establishes that the national test's fixed cost cannot be too large in comparison with future pay-offs. Otherwise, even when all the elements of r are close to $-\infty$, there are not enough students taking the national test to fill all of the seats offered by each university.

Lemma 3: In the case where $N = 2$, the equilibrium is unique when it exists.

Proved in Appendix A.1.

Although there is not a proof for $N > 2$, in Appendix A.1 I present a result which limits the potential extent of multiple equilibria. In particular, it narrows the possibility of having *high* and *low* effort equilibria. In addition to that, in the empirical implementation of the model, I solve the equilibrium with very different starting points for r, characterizing *high* and *low* effort equilibria, and in all these cases the algorithm converges to the same value of r.

It is worth mentioning that the potential lack of uniqueness is not an issue in the identification of the model's parameters. In fact, to calculate the likelihood function it is only necessary to solve the student's problem as opposed to the equilibrium (the values of r). The latter is not calculated in the estimation given that I observe the final-score cutoff (r) in the data. Thus, in the case of having more than one equilibrium, the estimation procedure selects the one that the students actually played. The usefulness of narrowing the potential extent of multiple equilibria is for counterfactual experiments.

In sum, this model characterizes a rank-order tournament with heterogeneous contestants. As it is clear from the first order conditions of the student problem, the effort decision is type specific, namely, it depends on X, Z, and λ. It is also unique when the first condition of Lemma 2 is fulfilled. Furthermore, the decision about taking the test can be different for students belonging to the same type, given that this decision depends on the realization of the admission test' fixed cost. Thus, even though two students of the same type (*i.e.* same X, Z, and λ) would exert the same level of effort in the case that they make the same decision about the test, they can have different level of exerted effort to the extent that one of them decide to take the test and the other one does not. Given these effort and admission test decisions, which are very heterogeneous, and the other (exogenous) components that determine the final score, students are sorted across universities by their final scores.

3 The Chilean System for College Admissions and Description of the Data

In the Chilean educational system, students can continue their studies after high school at different tertiary institutions, some *selective* (the best and most prestigious universities) and some not so (including some universities and technical institutions). In 2009, 29% of 18 to 25 year-olds were attending some type of tertiary institution.[11]

[11] CASEN 2009 (Chilean survey for socioeconomic characteristics).

The Chilean university system is highly structured: after knowing their final admissions score (a linear combination of high school GPA and admission test scores), students apply for a particular major at a particular university. They can apply for more than one major/university combination. Each university has an admission quota, a fixed number of seats, for each major.

As considered in the model, to be admitted to a selective university, the student must take a national college admissions test (PSU); the math and verbal sections are mandatory while certain majors require additional tests. Most of the selective universities have an explicit formula to calculate the final score (different weights for the PSU scores and GPA are considered). Thus selection is simply based on the final score ranking.

For the 2009 admissions process, among the $212,656$ students who finished high school, $56,437$ (27%) did not take the college admissions test and $156,219$ (73%) did. Note that the national test can be taken yearly and those who change majors must retest, so a percentage of those taking the college admissions test finished their secondary studies more than one year before. In this paper, I only use data for those students who finished high school in 2008, and those who didn't repeat any grades between 2004 and 2008. For the cohort, those students represent 84.5% ($179,725$ of $212,656$) of the total.

There are four sources of information in this paper; the first three are linked through an individual ID.[12]

- **PSU**: the national test for college admissions. These are census data provided by the DEMRE (Department of Educational Evaluation, Measurement and Recording).
- **RECH**: Ministry of Education's data. It includes information for all high school students including annual average attendance for each student, their GPA, and all high schools in which each student was enrolled. There is an identification number for each high school that can be used to link this RECH data with many other sources of high school information (including SIMCE data).
- **SIMCE 2004 and 2006**: Nation-wide tests taken by students in eighth grade (14 years old) and 10th grade (16 years old). These tests are designed to measure the quality of the system, are public information, and do not have any direct consequences for the tested students. During the week of the test, parents are surveyed to characterize students' families. From that survey, I have information on students; performance, some proxy measures of effort and learning skills, and characteristics of their families, primary and secondary schools.

[12] The Ministry of Education of Chile has all individual information with RUT (Chilean national ID), but for confidentiality reasons this data is given to the researchers with a new ID, which is useful to link the different data bases provided by the Ministry, but does not link with other databases at an individual level.

- **Futuro Laboral:** this is a project of the Chilean Ministry of Education that follows individuals for the first few years after graduation from higher education programs. The panel dataset matches tax returns – provided by Chile's Internal Revenue Service – with transcripts including students' majors and colleges, and their score on the PAA (*Prueba de admision universitaria*). The PAA was the national test for college admissions until 2002, when it was replaced by the PSU (they are very similar and have the same scale). The database only considers those who graduated from universities. Income information is available between the years 1996 and 2005, and contains individuals in Chile who had positive earnings between those years, including even those who were exempt from taxes. I use 2001 graduating class.[13]

  The wage measured in the sample is the annual income received from jobs and services and does not include self-employment income. The final sample consists of 19,870 individuals. Notice that these individuals were in the labor market when those in the sample used to estimate the model were attending the end of primary education and the start of secondary education. To the extent that my model seeks to understand the academic effort exerted by students during their secondary education (between 2005 and 2008), it makes sense they were observing wages between 2001 and 2005 as an estimation of their potential future wages based on major and university attended.

  From this source of information, I can estimate, using individual data, the relationship between the national test for college admissions and wage profile.

The final database, that put the first three sources of information together, contains $146,319$ observations, where the difference between this number and $179,725$ (who did not repeat any grade between 2004 and 2008) is mainly for two reasons: (1) lack of data for the 2004 SIMCE for some students, and/or (2) lack of socioeconomic information for some students. The vector X (which determines the test score production functions) includes: mother's education, father's education, the primary and high schools type attended by the students, dummies for rural condition of the primary and secondary schools, and the primary and high schools SES. There are five SES categories, which are defined by the Ministry of Education of Chile, such that the higher is the SES number of the school, the wealthier are its students. In Appendix B there is a description and statistics of the variables considered in this paper, showing the differences between the population and the estimation sample. Although the differences between them are not economically important, in general they are statistically significant. In particular, the estimation sample has more students coming from higher socieconomic groups.

[13] This database is a sub sample of the database used by Braga and Bordon (2017). They study whether employers use university prestige as a signal of workers' unobservable productivity.

Note that in this empirical implementation all independent variables that determine students' effort decision and the their decision about taking admission test are discrete. Having only discrete independent variables implies that I can group students who share the same characteristics, defining *student types*. This feature of my database is consistent with the model, and its assumption about a mass of students of each type. Moreover, it helps to speed up the estimation, given that effort decisions, more precisely $\hat{e}_i^0$ and $\hat{e}_i^1$, are the same for all students belonging to the same type. In concrete, there are $4,484$ types (grouping $146,319$ students), whose average size is 32.63, with a maximum size of 828.

Few decisions should be made to adjust data contents to model simplifications. In the model, universities differ in quality and there are two majors: scientific and humanities. To proxy the major preference, in the case of the *Futuro Laboral* database, I define those who score higher in the math than in the verbal PSU, as scientific major students, and those scoring higher in verbal as humanities major students. Then, I assume that for each major there are eight university tiers and one residual (for those who prefer major z and either do not take the college admissions test or have a final score below r_1^z). To define these tiers, I group students according their preferences in majors and calculate the quintiles of the final scores, dividing the fourth quintile into two and fifth into three. In other words, I divide the groups of students using the percentiles 20, 40, 60, 70, 80, 87, and 93; of the final score. To have finer groups at the end of the final score distribution, is to capture the more steeply relationship between wages and final score values which is observed as the final scores increase. Then, R_n^z, $n \in \{2, 3, ..., 9\}$, is calculated as the mean of the monthly income a few years after graduating for students who preferred major z and whose final score belongs to tier n. R_1^z is calculated as the mean of the monthly income of the students who preferred major z and whose final score is below 450, which in Chile is the minimum final score to apply to selective universities.

The calculation of r_n^z and S_n^z, $n \in \{1, 2, ..., 9\}$, uses the estimation sample ($146,319$ observations). Since in this sample there is information on the performance on both verbal and math in the 8th grade standardized test (the SIMCE), I use the relative performance in this test to define the major preferences for these $146,319$ students. Moreover, I define the weights of the different tests on the final score function by approximating what is observed in real life, such that in scientific major the weights are 0.4 for the math PSU, 0.2 for the verbal PSU, and 0.4 for GPA, and in humanities major the weights are 0.2 for the math PSU, 0.4 for the verbal PSU, and 0.4 for GPA. Then to calculate S_n^z, $n \in \{2, 3, ..., 9\}$, I take all of the students who prefer major z and whose final score is above 450 points and consider the same percentiles used to calculate R^z: 20, 40, 60, 70, 80, 87, and 93. S_1^z is the residual and can be also calculated as the total number of students who prefer major z and whose final score is below 450 or who decided not to take the admission test. Finally, r_n^z is calculated as the minimum final score for those students who attend university n.

Table 1 summarizes all these calculations, showing the estimated relationship between the national test for college admissions and wage profile. Note the relevant difference between the wage profile of the students who perform better in math (*i.e.,* scientific major), versus those who perform better in verbal (*i.e.,* humanities major). This table also shows the more steeply relationship between wages and final score values as the final score increases. In fact, even though higher university tiers have fewer students, the increases in payoffs from one tier to another one are similar to the increases between the first tiers, which include more students.

Notice that income data are used in estimation primarily as a means of extracting ordinal information on college quality. In this context, observed and unobserved student characteristics determine payoffs only indirectly through improving high school grades (and therefore college placement). Regarding this, there is a vast literature, with mixed evidence, that studies the impact of college and its quality on future earnings, e.g., Brewer et al. (1999); Dale and Krueger (2002); Dale and Krueger (2002); James et al. (1989); and (with Chilean data) Reyes et al. (2013). It is worth noting that while the literature has focused its attention on how to control for student and college selection, this is not relevant in my approach because the important feature in my model is not how much students are actually going to earn, but their beliefs on the impact of different universities on their future earnings.

Table 1: Universities' Payoffs as Function of Students' Final Scores

University (Scientific major)	R: payoff	r: final-score cutoff	S: seats
1 (Not admitted to college)	0.59	0	Residual
2	0.90	450	10,565
3	1.10	499	10,643
4	1.23	541	10,616
5	1.33	585	5,304
6	1.43	610	5,308
7	1.57	640	3,555
8	1.69	665	3,549
9	1.81	701	3,514
University (Humanities major)	**R**	**r**	**S**
1 (Not admitted to college)	0.67	0	Residual
2	0.69	450	7,340
3	0.83	482	7,338
4	0.89	515	7,372
5	1.05	551	3,667
6	1.16	573	3,684
7	1.27	601	2,451
8	1.42	625	2,467
9	1.64	661	2,438

Note: The payoffs are monthly wages and are measured in thousands of US dollars. The final score (which determines the cutoff) is a linear combination of test scores and high school GPA: $0.4 * (Math\ PSU) + 0.2 * (Verbal\ PSU) + 0.4 * GPA$, in the case of scientific major, and $0.2 * (Math\ PSU) + 0.4 * (Verbal\ PSU) + 0.4 * GPA$, en the case of humanities major. These variables have a mean of 550 and an standard deviation of 100 points.

4 Empirical Specification and Identification

Following the factor model literature, I assume that there are three unobserved variables for which I have measures (i.e., proxies): λ_i (learning skills), e_i^p (student effort at primary school), and e_i^h (student effort at secondary school). The last is modeled in the paper, while the first two are treated as unobserved heterogeneity. Regarding this, I take advantage of the panel data in order to have measures of learning skills and student effort at primary school before the decisions modeled in this paper are made, namely, high school effort and whether to take the admission test.

A key element of the model are the learning skills, which are assumed to be scalar, time invariant, and independent of x. This is an unobserved variable that impacts all of the tests results and scores, and the cost of taking the admission test (discussed below). Then, as usual in the structural estimation approach, this unobserved variable allows me to have correlation among tests, scores, and students' decisions, conditional on observables. The innovation in this paper is that the identification of the distribution of this unobserved variable takes insights from structural estimation approach and factor model literature.

Besides the functions that determine the final score, I consider several measures and tests, which, in the context of latent variables, are useful to identify the parameters of interest: the final score determinants, i.e., 2009 national test for college admissions (PM, the math test; and PV the verbal test) and high school GPA; the earlier standardized test scores (SIMCEs, 2004 and 2006); and some direct measures of primary and high school effort (Me^p and Me^h, respectively), along with measures for the unobserved learning skills ($M\lambda^p$). The direct learning skills measures ($M\lambda^p$) are: whether the student repeated at least one primary school grade; and the student's answer to a set of questions about her *ability to understand hard subjects, self-confidence in performing well in exams, determination to set learning goals*.[14] The direct primary school effort measures (Me^p) are attendance in 8th grade; and self-reported perception about her effort at 8th grade including *exerts effort at hard subjects, try hard to learn, looks for additional information, math and study intensity.* The direct high school effort measures (Me^h) are: attendance at 10th grade and self-reported perception about her effort at 8th grade, such as: *Exert effort, Study at home, use textbook and calculators at home.* Statistics of all these measures are summarized at Appendix B.

To emphasize, the direct measures (Me^p, Me^h, and $M\lambda^p$) are assumed to be functions of the latent variable that they measure, and of exogenous variables. In other words, this direct measures are only function of one latent variable at the same time, which is an exclusion restriction that ensures identification. Thus, for example, learning skills does not affect high school attendance directly, but through its effect on academic effort.

[14]In the context of the papers. Cunha et al. (2010) and Heckman et al. (2006), these learning skill variables would be closer to non-cognitive skills given the measures available.

That said, it should be noticed that I also have other measures such as the standardized high school test scores, which are functions of two latent variables (high school effort and student learning skills).

Let i denote the individual, j the measure, and J the total number of measures that I have for each unobserved variables, the empirical implementation is characterized by the following equations:

Final Score Determinants:

$$PM_i = \beta_0^{pm} + x_i^h \beta_1^{pm} + e_i^h \beta_2^{pm} + \lambda_i \beta_3^{pm} + \varepsilon_i^{pm}, \quad \forall i \ s.t. \ TCAT_i = 1, \tag{12}$$

$$PV_i = \beta_0^{pv} + x_i^h \beta_1^{pv} + e_i^h \beta_2^{pv} + \lambda_i \beta_3^{pv} + \varepsilon_i^{pv}, \quad \forall i \ s.t. \ TCAT_i = 1, \tag{13}$$

$$GPA_i^h = \beta_0^{gh} + x_i^h \beta_1^{gh} + e_i^h \beta_2^{gh} + \lambda_i \beta_3^{gh} + \varepsilon_i^{gh}. \tag{14}$$

High school performance and effort measurements:

$$SIMCE_{ji}^h = \beta_0^{sjh} + x_i^h \beta_1^{sjh} + e_i^h \beta_2^{sjh} + \lambda_i \beta_3^{sjh} + \varepsilon_i^{sjh}, \quad j \in \{verbal, math\}, \tag{15}$$

$$Me_{ji}^h = x_i^{ejh} \beta_1^{ejh} + e_i^h \alpha^{ejh} + \varepsilon_i^{ejh}, \quad j \in \{1, ..., J_{eh}\} \ J_{eh} \geq 2. \tag{16}$$

Primary school performance, learning skill and effort measurements:

$$SIMCE_{ji}^p = \beta_0^{sjp} + x_i^p \beta_1^{sjp} + e_i^p \beta_2^{sjp} + \lambda_i \beta_3^{sjp} + \varepsilon_i^{sjp}, \quad j \in \{verbal, math, natural\ science, social\ science\}, \tag{17}$$

$$GPA_i^p = \beta_0^{gp} + x_i^p \beta_1^{gp} + e_i^p \beta_2^{gp} + \lambda_i \beta_3^{gp} + \varepsilon_i^{gp}, \tag{18}$$

$$Me_{ji}^p = x_i^{ejp} \beta_1^{ejp} + e_i^p \alpha^{ejp} + \varepsilon_i^{ejp}, \quad j \in \{1, ..., J_{ep}\} \ J_{ep} \geq 2, \tag{19}$$

$$M\lambda_{ji}^p = x_i^{\lambda jp} \beta_1^{\lambda jp} + \lambda_i \alpha^{\lambda jp} + \varepsilon_i^{\lambda jp}, \quad j \in \{1, ..., J_\lambda\} \ J_\lambda \geq 2. \tag{20}$$

As emphasized above, it is assumed that all the ε_is are normally and independently distributed, with one exception. The identification strategy requires that at least one measurement to be a linear function of each unobservable. Because all learning skills measures are binaries, I assume a linear probability model for $M\lambda_1^p$. In consequence, $\varepsilon_i^{\lambda 1p}$ is not normal distributed. Notice that independency implies that, conditional on observables, the correlation across equations is only given by the unobserved skill heterogeneity.

Following factor analysis literature, I normalize $\alpha^{e1h} = \alpha^{e1p} = \alpha^{\lambda 1p} = 1$. As Cunha and Heckman (2008) stress, because the tests only contain ordinal information, it is more appropriate to anchor the scale of the latent factors using measures with an interpretable metric, such as the ones used in this paper. They are a binary variable that takes the value of 1 if the student had repeated at least one year and 0 otherwise (measuring learning skills), the attendance rate for the last year of primary school (measuring primary school academic effort), and the mean of school attendance over the four years of secondary school (measuring secondary school academic effort). Using attendance as a measure of effort is a common practice. see for example Hastings et al. (2012). For simplicity, I also assume that $x_i^{\lambda jp}$, x_i^{ejp} and x_i^p do not have elements in common, and the same for x_i^{ejh} and x_i^h.

As described in the model section, the effort cost ($\theta_2^{t(i)}$) is individual specific. This allows different effort decisions among students who are not taking the college admissions test, remember that $\hat{e}_t^0 = \frac{\theta_1}{\theta_2^{t(i)}} b_1$. In the SIMCE 2004, the students are asked about how much do they like to study math and language. The possible answers are: strongly agree, agree, disagree, and strongly disagree. Given that few people choose the last category, I use three values: 1 if the student strongly agrees, 2 if the student agrees, and 3 if the student disagrees or strongly disagrees. Then, I define a variable *LikeStudy* with three possible values. It takes a value of one if a student strongly agrees that she likes to study math and language, two if a student agrees that she likes to study math or language, and three if a student disagrees or strongly disagrees that she likes to study math and language. Given this, the cost of effort parameter is defined as:

$$\theta_2^{t(i)} = exp(\theta_2^0 + \theta_2^1 * 1(LikeStudy_{t(i)} = 2) + \theta_2^2 * 1(LikeStudy_{t(i)} = 3)).$$

Which implies that $\theta_2^{t(i)}$ is equal to $exp(\theta_2^0)$ when the student strongly agrees with the statement: *I enjoy the study of math and language.*

The mean for the cost of taking the test ($\overline{FC}$) is also individual specific. Given the structure of the model, a natural approach is to make this mean a function of the unobserved learning skill variable (λ). Moreover, this approach gives more flexibility to the model to fit the data, since it allows a – conditional on observables – correlation between high school GPA and the decision to take the admission test. Concretely, the mean cost is equal to:

$$\overline{FC}_{t(i)} = \overline{FC}_0 + \overline{FC}_1 * 1(\lambda_{t(i)}\ Group\ 2) + \overline{FC}_2 * 1(\lambda_{t(i)}\ Group\ 3) + \overline{FC}_3 * 1(\lambda_{t(i)}\ Group\ 4).$$

4.1 Identification

To the extent that the final goal of this paper is to perform counterfactuals related to the college admissions process, the objects which must be identified for this analysis are $\{\beta^{pm}, \beta^{pv}, \beta^{gh}\}$, $\{Var(\varepsilon_i^{pm}), Var(\varepsilon_i^{pv}), Var(\varepsilon_i^{gh})\}$, $\{\theta, \overline{FC}(\lambda), \sigma_{fc}, \sigma_\eta\}$ and the distribution of λ. The identification strategy has three steps. First, I identify the final score's expectation and variance.Notice that if $Var(\varepsilon_i^{pm})$, $Var(\varepsilon_i^{pv})$ and $Var(\varepsilon_i^{gh})$ are identified, then σ_η is also identified. Second, I non-parametrically identify the distribution of learning skills. Third, I identify the utility parameters from different moments of the effort measurements.

Before proceeding with the proof, I list all the assumptions that it requires. Although all these conditions have been already mentioned; they are stated here to provide a big picture about what supports identification. Each time that an assumption is needed, it is noted in parentheses.

INDME: All the errors of the measurement equations are independent of each other and of any other error in the model.

INDPF: All the errors for the production functions of the components of the final score are independent of each other and of any other error in the model. In other words, conditional on X and e, the correlation across different tests is only driven by the unobserved learning skill characteristic (λ).

AtL2M: There are at least two measures for each unobserved variable (*i.e.*, effort at primary school, effort at high school, and learning skill).

TPfL: The test production functions are linear in the parameters.

EXAD: The effort decision and the decision about taking the admission tests are made before the realization of the test production functions' shocks.

INDL: λ is independent of all the errors in the model.

SCOF: The objective function for those students who do take the admission test is strictly concave, which can be ensured by $\forall\ i:\ \theta_1(R_N - R_1) < \left(\frac{\sigma_\eta}{a_1}\right)^2 \frac{1}{\phi(1)}$.

Step 1, the final score's expectation and variance:

Let $T_i \in \{PM_i, PV_i, GPA_i^h\}$. By **(TPfL)** and **(AtL2M)**, it is clear that:

$$T_i = \beta_0^T + x_i^h \beta_1^T + \beta_2^T (Me_{1i}^h - x_i^{e1h} \beta_1^{e1h}) + \beta_3^T (M\lambda_1^p - x_i^{\lambda 1p} \beta_1^{\lambda 1p}) - (\beta_2^T \varepsilon_i^{e1h} + \beta_3^T \varepsilon_i^{\lambda 1p}) + \varepsilon_i^T$$

Thus, defining $\delta_i^T = \varepsilon_i^T - (\beta_2^T \varepsilon_i^{e1h} + \beta_3^T \varepsilon_i^{\lambda 1p})$, it is possible to construct the following moment conditions:[15] $E[\delta_i^T | x_i^h] = 0$, $E[\delta_i^T | x_i^{e1h}] = 0$, $E[\delta_i^T | x_i^{\lambda 1p}] = 0$, $E[\delta_i^T | Me_{2i}^h] = 0$ and $E[\delta_i^T | M\lambda_{2i}^p] = 0$, where the last two moment conditions are possible because of **(AtL2M)**. From these moment conditions, β^T, β^{e1h} and $\beta^{\lambda 1p}$ are identified. Therefore, $\{\beta^{pm}, \beta^{pv}, \beta^{gh}\}$ are also identified. This also implies that all the parameters involved in a_{0i}, a_1 and b_1 are identified.

Given that $\{\beta^{pm}, \beta^{pv}, \beta^{gh}\}$ are identified, $\{var(\delta_i^{pm}), var(\delta_i^{pv}), var(\delta_i^{gh})\}$ are also identified. Moreover, to show the identification of $\{var(\varepsilon_i^{pm}), var(\varepsilon_i^{pv}), var(\varepsilon_i^{gh})\}$ notice that:

$$cov(T_i - \beta_0^T - x_i^h \beta_1^T - \beta_2^T (Me_{1i}^h - x_i^{e1h} \beta_1^{e1h}) - \beta_3^T (M\lambda_1^p - x_i^{\lambda 1p} \beta_1^{\lambda 1p}), Me_{1i}^h - x_i^{e1h} \beta_1^{e1h}) = cov(\varepsilon_i^T - (\beta_2^T \varepsilon_i^{e1h} + \beta_3^T \varepsilon_i^{\lambda 1p}), e_i^h + \varepsilon_i^{e1h}) = -\beta_2^T var(\varepsilon_i^{e1h}).$$

$$cov(T_i - \beta_0^T - x_i^h \beta_1^T - \beta_2^T (Me_{1i}^h - x_i^{e1h} \beta_1^{e1h}) - \beta_3^T (M\lambda_1^p - x_i^{\lambda 1p} \beta_1^{\lambda 1p}), M\lambda_1^p - x_i^{\lambda 1p} \beta_1^{\lambda 1p}) = cov(\varepsilon_i^T - (\beta_2^T \varepsilon_i^{e1h} + \beta_3^T \varepsilon_i^{\lambda 1p}), \lambda_i^p + \varepsilon_i^{\lambda 1p}) = -\beta_3^T var(\varepsilon_i^{\lambda 1p}).$$

These results imply that $var(\varepsilon_i^{e1h})$ and $var(\varepsilon_i^{\lambda 1h})$ are identified, and consequently $var(\varepsilon_i^{pm})$, $var(\varepsilon_i^{pv})$, and $var(\varepsilon_i^{gh})$ are also identified, where $cov(\varepsilon_i^T - (\beta_2^T \varepsilon_i^{e1h} + \beta_3^T \varepsilon_i^{\lambda 1p}), e_i^h) = 0$ is because of **(EXAD)** and $cov(\varepsilon_i^T - (\beta_2^T \varepsilon_i^{e1h} + \beta_3^T \varepsilon_i^{\lambda 1p}), \lambda_i^p) = 0$ is because of **(INDL)**. Along the same lines, by **(INDME)**, it is proven that $cov(\varepsilon_i^{e1h}, \varepsilon_i^{e1h}) = 0$ and $cov(\varepsilon_i^{e1h}, \varepsilon_i^{\lambda 1p}) = 0$; and by **(INDME)** or **(INDPF)**, it is ensured that $cov(\varepsilon_i^T, \varepsilon_i^{e1h}) = 0$ and $cov(\varepsilon_i^T, \varepsilon_i^{\lambda 1p}) = 0$, $\forall\ T$.

Step 2, the distribution of learning skills and high school student's effort:

The nonparametric identification of $f(\lambda)$ and $f(e^h|x)$ can be proven following an analysis similar to Cunha and Heckman (2008). First, proceeding in a similar fashion as before, with two measures for each latent variable **(AtL2M)**, it is possible to identify $\{\beta_0^{sjp}, \beta_1^{sjp}, \beta_2^{sjp}, \beta_3^{sjp}, \beta_1^{e1p}\}$ for any $j \in \{verbal, math, natural\ science, social\ science\}$. Hence, defining $\widehat{SIMCE}_{ji}^p = (SIMCE_{ji}^p - \beta_0^{sjp} - x_i^p \beta_1^{sjp} - \beta_2^{sjp}(Me_{1i}^p - x_i^{e1p} \beta_1^{e1p}))\frac{1}{\beta_3^{sjp}}$ and $\widehat{\varepsilon}_i^{sjp} = (\varepsilon_i^{sjp} - \beta_2^{sjp} \varepsilon_i^{e1p})\frac{1}{\beta_3^{sjp}}$, it follows that:

[15]By **(EXAD)**, because the decisions are taken before the shocks' realization, such decisions are independent of the errors; hence, when $T_i \in \{PM_i, PV_i\}$: $E[\delta_i^T | Me_{2i}^h, M\lambda_{2i}^p, x_i^h, x_i^{e1h}, x_i^{\lambda 1p}, TCAT_i = 1] = E[\delta_i^T | Me_{2i}^h, M\lambda_{2i}^p, x_i^h, x_i^{e1h}, x_i^{\lambda 1p}]$. Thus, there is no selection bias.

$$\widehat{SIMCE}^p_{ji} = \lambda_i + \hat{\varepsilon}^{sjp}_i$$
$$M\lambda^p_{ji} - x^{\lambda 1p}_i \beta^{\lambda 1p}_1 = \lambda_i + \varepsilon^{\lambda 1p}_i$$

Therefore, because $\hat{\varepsilon}^{sjp}_i$ and $\varepsilon^{\lambda 1p}_i$ are independent of each other and with respect to λ_i (due to **(INDME)** and **(INDL)**), the distribution of λ is identified (Cunha and Heckman (2008)). This result comes from Kotlarski's Theorem. Moreover, the identification can be achieved under much weaker conditions regarding measurement errors. Indeed, independence is not necessary, see Cunha et al. (2010).

Along the same lines, it is possible to prove the nonparametric identification of $f(e^h|x)$. As in the case of $f(\lambda)$, the identification of $f(e^h|x)$ rests on the existence of two measures of e^h.

Step 3, the parameters of the utility function:

Once the distribution of λ is identified, it is possible to identify the utility parameters.[16] First, notice that if $Me^h_{1i} - \varepsilon^{e1h}_i = e_i$, where e_i is the effort of individual i. Let $e = g(x_i, a(\lambda_i), b, \theta_1, \theta^t_2, \sigma_\eta)$ be the implicit relation resulting from the first order condition (10), which can be characterized as a function due to assumption **(SCOF)**. Thus, based on the first order conditions and for any t, it is the case that:

$E[Me^h_{1i} - \varepsilon^{e1h}_i | TCAT_i = 0, t(i) = t] = \frac{\theta_1}{\theta^t_2} b_1$,

$E[Me^h_{1i} - \varepsilon^{e1h}_i | TCAT_i = 1, t(i) = t] = \int_\lambda g(x_i, a(\lambda), b, \theta_1, \theta^t_2, \sigma_\eta) f(\lambda | TCAT = 1, t) d\lambda$.

It should be noticed that the strength and flexibility of this proof is given by the fact that the distribution of λ (and $e^h|x$) are nonparametrically identified.

Finally, the identification of $\overline{FC}(\lambda)$ and σ_{fc} comes from

$$Pr(TCAT_i = 1 | D_i(\lambda_i, x_i), \overline{FC}(\lambda_i), \sigma_{fc}) = \Phi\left(\frac{D_i(\lambda_i, x_i) - \overline{FC}(\lambda_i)}{\sigma_{fc}}\right)$$

$$\Rightarrow \int_\lambda Pr(TCAT_i = 1 | D_i(\lambda, x_i), \overline{FC}(\lambda), \sigma_{fc}) f(\lambda) d\lambda = \int_\lambda \Phi\left(\frac{D_i(\lambda, x_i) - \overline{FC}(\lambda)}{\sigma_{fc}}\right) f(\lambda) d\lambda.$$

Indeed, by varying $D_i(\lambda, x_i)$, it is possible to obtain the value of σ_{fc}, because Φ is a monotonic function. Moreover, by fixing $D_i(\lambda, x_i)$ for any particular value, it is possible to find the value of $\frac{\overline{FC}(\lambda)}{\sigma_{fc}}$, which implies the identification of $\overline{FC}(\lambda)$.

[16] The identification of $f(e^h|x)$ gives another way to identify the utility parameters.

5 Estimation

The estimation is carried out in two steps. In the first step, following the identification analysis presented above and the standard approach to dealing with measurement error in independent variables (both effort and learning skills), I can consistently estimate all the parameters of the test equations ((12), (13), (14), (15) and (17)) by a two-stage least square. In the first stage, one measure of each latent variable is regressed against all the other measures: high school attendance rate (the average between 9th and 12th grade) against the other measures of high school academic effort; primary school attendance rate (8th grade) against the other measures of primary school academic effort; and grade repetition before 8th grade against all the other learning skills measures. In the second stage, all the scores that measure the primary (secondary) education student performance are regressed against X (*e.g.,* gender, socioeconomic groups, high school or primary school type) and the projection of the (high) primary school academic effort and learning skill that come from the first stage.

In the second step, using relevant parameters from the first step as inputs, I estimate the utility parameters, the distribution of the unobserved learning skills, and the parameters of the measurement equations by maximum likelihood procedure. I follow this approach mainly because most of the parameters are estimated in the first step, which only takes a few seconds, leaving just a few parameters to be estimated in the second step. This is a large time gain, given that in each iteration the model needs to be solved (which takes around 15 seconds for each set of parameters). In terms of numbers, 161 parameters are estimated in the first step, whereas 85 are estimated in the second step.

Let Ω_s be the set of parameters estimated in the s step ($s \in \{1, 2\}$, $\Omega = \{\Omega_1, \Omega_2\}$). The estimation procedure for the second step is the following:

- Guess the initial values for all the parameters, Ω_2^0 (this includes the parameters of the learning skills distribution).
- Given Ω_2^0, r, R, and X, find the effort decision for each student. There are two features of this procedure that speed up this calculation. First, given that the final score cutoff is observed, the equilibrium conditions are not required.Indeed, because I only need to calculate the student's problem, conditional on r, the estimation method used is maximum likelihood as opposed to simulated maximum likelihood. Second, the first order conditions of the student' problem, which lack a closed form solution, should only be solved for the $4,484$ student types.
- Calculate the likelihood function.
- Continue with a new guess until finding the Ω_2 that maximizes the likelihood function.[17]

[17]This is done using the derivative free solver, *HOPSPACK* in Fortran and *fminsearch* in Matlab.

There are some features of this procedure that are worth highlighting. The distribution of unobserved learning skills is approximated by a discrete distribution of four groups. This approach has two advantages: first, it is consistent with the model, in which there is a mass of students for each type and where the unobserved groups is one of the dimensions in which types differ. Second, it speeds up the estimation, because the student optimization has to be solved just once per student type in each iteration ($4,484 * 4$ instead of $146,319 * 4$ times). Although some of the parameters that are estimated in the second step can also be estimated in the first step (e.g., the factor loadings as shown in the identification argument), I prefer estimating those parameters in the second step to give to the model a better chance of fitting the data (since the model is only solved in the second step). Additionally, the distribution of the unobserved primary school effort is not estimated. Instead, I calculate the projection of one of the continuous measures of that effort on its other measures and then replace the primary school effort by that projection. Finally, when I have missing data in one of the measures (high school effort or learning skills), I assume that it is random and don't consider the contribution to the likelihood of this measure for such a student; I don't have to drop the entire data point.

To have a clear picture of the likelihood function, in Appendix C, I describe the contributions of different data to the likelihood.

6 Results

The first step estimation results are presented in Appendix D.1 (Tables 4, 5, and 6). Some aspects of these estimations are worth mentioning. First, for the OLS regressions where the dependent variable is either high school effort or learning skills and the rest of the measures are independent variables, the magnitudes, signs, and statistical significance are as expected. Although in some cases the r squared is fairly small, the instruments are not weak.[18] Second, Table 8, in Appendix D.1, shows that the estimated parameters of the equations that determine the effort decision (math PSU, verbal PSU, and high school GPA), are all as expected in terms of statistical significance, magnitudes, and signs.

Finally, the second stage OLS for the primary education performance presents some problems (Table 7). Indeed, the effect of effort (predicted with instruments) on SIMCE scoress is not in the expected direction. Nevertheless, the effect is in the expected direction for the GPA equation, and in both cases, the effect is statistically significant. Furthermore, the effect of the predicted learning skills is positive and large in magnitude in all equations (the parameters are negative because the variables are ordered from greater to fewer skills). It is worth mentioning that these problematic signs do not have

[18]The F statistics are: 16.99 (Primary School Attending Regression), 103.19 (Secondary School Attending Regression), and 58.09 (Repetitions Linear Probability Regression).

any relevant consequences for the model estimation and simulation. This is the case because, in my estimation, the primary school standardized test scores are only useful in identifing the distribution of learning skills ($f(\lambda)$), the signs of which are as expected. To be more specific, the primary education performance measures are important because they give information about students' learning skill before they choose their effort in secondary school.

The parameters estimated in the second step are shown in Appendix D.2 (Table 9). As in the first step, the vast majority of the estimated parameters have the expected sign. In particular, all the parameters of the utility function have the right sign and are statistically significant. Another aspect to highlight is that all parameters that multiply the latent effort decision in each measurement equation (*i.e.,* the factor loadings) are positive and statistically significant.

The standard errors reported in Table 9 were calculated using the approximation of the hessian given by the mean of the outer product of the scores, as if all the parameters were estimated by MLE.[19] Thus, the scores used are the derivatives with respect to the 246 parameters, which implies that the fact that the model is estimated in two steps does not have any consequence for the estimation of the standard errors. The intuition of this procedure is that, given the consistent parameter estimates, one Newton-Raphson step would yield consistent, efficient, and asymptotically normal parameter estimates. Thus applying the outer-product approximation would provide standard errors for those parameter estimates.

Given the non-linear relationship between the parameters and the model's outputs, the best way to assess the relevance of parameter magnitudes is through model fit analysis and counterfactual experiments.

6.1 Model Fit

To study how well this model fits the data, I simulate it using the estimated parameters. Due to the size of the database, I only have one simulation per student. The computational algorithm to solve the equilibrium of the model works as follows: (1) Draw the individual cost of taking the PSU and the individual shocks for PSU tests and GPA. (2) Guess an initial value for the final-score cutoff r^0 of each major. (3) Given r^0, the parameters of the model and the cost of taking the PSU, calculate the optimal effort and optimal decision to take the PSU for each student. (4) Given the shocks and effort decisions, calculate the new final-score cutoff (r^1), which solves the equilibrium conditions. (5) Stop if – for each major – this new r^1 is close enough to r^0 ($\max_{n\in\{2,\ldots,N\}} |r^0_n - r^1_1| < \epsilon$);

[19]This method to estimate standard error is based on asymptotic theory; in particular, it uses the idea that, under general conditions, $\sqrt{N}(\hat{\Omega} - \Omega) \xrightarrow{p} N(0, A^{-1})$, where $A^{-1} = E[H_i(\Omega)]$ is estimated as $\hat{A} = \frac{1}{N}\sum_{i=1}^{N} s_i(\Omega)s_i(\Omega)'$, such that s_i is the score: $s_i = \frac{\partial l_i}{\partial \Omega}$.

otherwise, restart from point (2), with r^1 as the new guess. Notice that to order to get convergence, the shocks of step one are simulated only once.

To complement the theoretical results that limit – but not eliminate – the possibility of multiple equilibria, I simulate the model considering very different initial guesses for r. In particular, let r_{data} be the guess I consider in the rest of the paper when I simulate the model, Table 10, in Appendix D.3, presents the r to which the algorithm converges by starting from the following initial guesses: $\{(0.4+\alpha)*r_{data}\}_{\alpha\in\{0.1,0.2,...,1,1.1,...,1.3\}}$. The idea of this exercise is to look for the simultaneous existence of *worse* and *better* equilibria. Since it is similar for both majors, this table only presents the simulations for scientific major. As can be seen, at least numerically, the existence of multiple equilibria does not seem to be a relevant issue. The equilibrium reached starting from very different – extreme – initial guesses are almost equal.

Although in the estimation procedure only the student's problem is solved, because the final-score cutoff comes from data, to simulate the model implies finding the equilibrium r vector for each major. Thus, the first element to consider in model fit analysis is how close the simulated r_n are to the cutoffs coming from the data. In this regard, Figure 7, in Appendix D.3, shows that for each major the simulated vector r captures the trend and magnitudes of the data fairly well.

Though the model shows a good fit in all the aspects of the data, given that the goal of this paper is to study how different college admissions policies may affect high school students' behavior, I focus my attention on the model fit for those tests that are relevant in the admissions process, along with the student test decision. Figure 1 shows that the model replicates the test distribution observed in the data. The discrepancies in the case of high school GPA are because the data is discrete and there are agglomerations in some grades, something that cannot be replicated by the model. Moreover, Table 2 shows that the model is able to replicate student performance across different groups relatively well. Appendix D.3 contains Figures 8 and 9, which show the model fit of the densities for the remaining tests (2004 and 2006), all of them showing a good fit.

Furthermore, the simulated model also fits the data patterns with regard to the fraction of students taking the PSU across different groups, which is important because one of the two decisions considered in my model is whether to take the national admissions test. Indeed, Figure 2 shows how the simulation of the model replicates this fraction, particularly the patterns, and, with some discrepancies, the magnitudes across gender, socioeconomic status of high school, maternal and paternal education, and high school categories (public, private subsidized, and private non-subsidized).[20] Moreover, Table

[20]The Chilean educational system has three school types: (i) Public or municipal schools are run by 345 municipalities which receive a per-student subsidy from the central government. (ii) Private-voucher schools; these are independent religious or secular institutions that receive the same per student subsidy as public schools. (iii) Private unsubsidized schools are also independent, but receive no public funding.

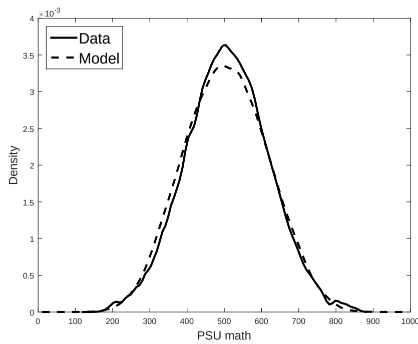
Table 2: Model Fit by Different Groups

	PSU math		PSU verbal		GPA	
	Model	Data	Model	Data	Model	Data
All	508	505	505	502	537	538
Female	494	493	500	499	550	550
Male	525	518	511	505	523	524
SES 1	423	414	422	411	516	509
SES 2	452	449	453	451	517	513
SES 3	517	525	515	524	542	547
SES 4	581	586	573	579	570	579
SES 5	640	636	626	622	611	614
F wo college	490	486	488	484	529	528
F w college	597	599	589	590	591	596
Public	475	472	472	471	530	530
Private Sub	503	500	502	499	530	531
Private non Sub	637	633	622	619	607	611

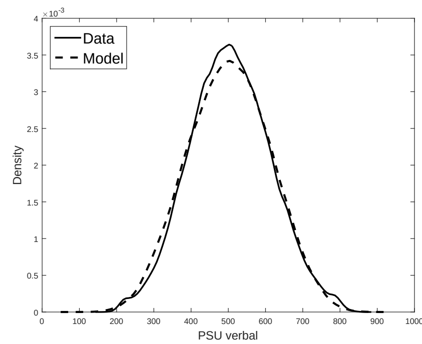
11, in appendix D.3, shows the model fit at individual level on the decision about taking the admission test but, which is much more demanding than the previous analysis at aggregate level. However, the table presents a very high fraction of coincidences for different groups, between 54 and 96 percent.

Figure 1: Model Fit in Tests Determining Final Score

(a) PSU math



(b) PSU verbal



(c) GPA in high school

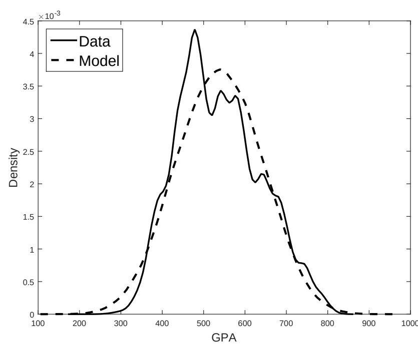
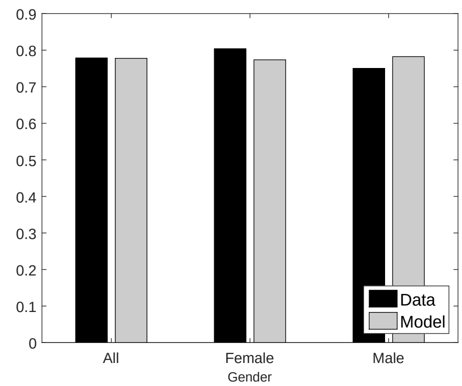
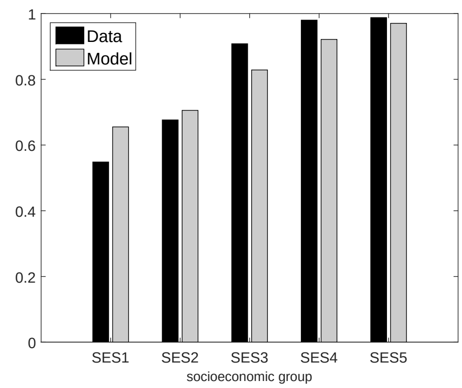


Figure 2: Fraction of the Students Taking the PSU by Groups

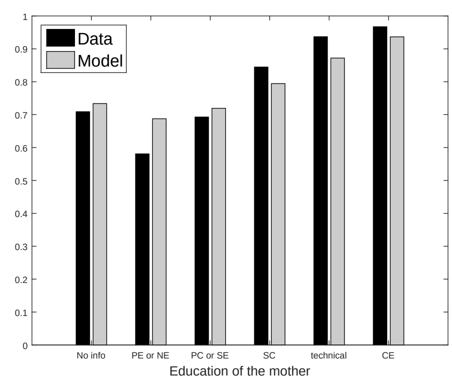
(a) By gender



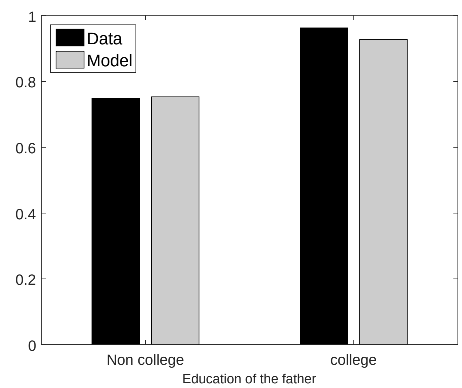
(b) By high school SES



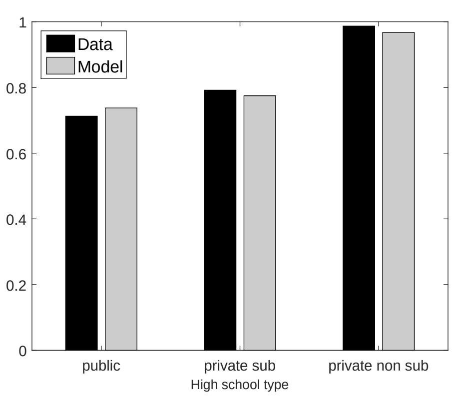
(c) By mother's education



(d) By father's education



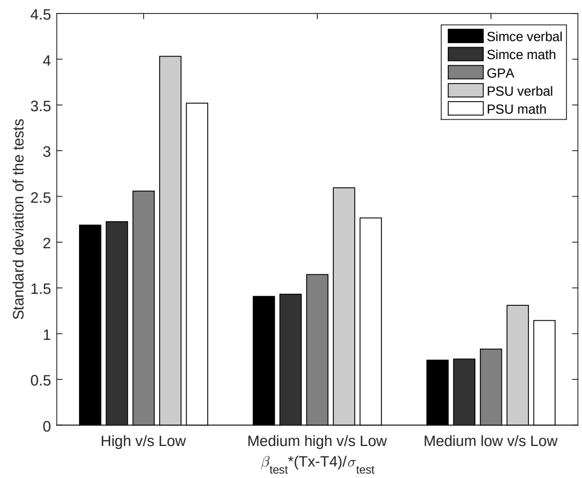
(e) By high school categories



6.2 Learning Skills

As usual in structural estimations, discrete unobserved groups improve the fit of the model. Besides its contribution to fitting the data, these unobserved groups turned out to be a relevant driver of the observable differences in test scores. As Figure 3 shows, the impact of these learning skills groups on tests are between 0.5 and 1.5 standard deviations (medium-low versus low group), 1 and 2.5 standard deviations (medium-high versus low group) and 2 and 4 standard deviations (high versus low group).[21]

Figure 3: The Impact of Learning Skills Groups on Tests



7 Counterfactual Experiments

7.1 Impact of the quota by Socioeconomic Group Admission System

To study the impact of affirmative action policies on effort, test scores, and the probability of taking the college admissions test, I begin by simulating the model under a quota by socioeconomic group admission system (SES-quota system), which imposes that, for each university tier, the SES distribution is the same as the population. In other words, if, in the whole system, $x\%$ of the students attend high schools with socioeconomic group i (SES i), then there should be $x\%$ of students from each high school type in each university tier. In practice, the way to simulate this counterfactual is by

[21] In all the simulations and counterfactual experiments performed in this paper, the probabilities of each learning skills group are calculated by Bayes' rules, *i.e.*, $\pi^i_{t|x} = \frac{\pi_t L_i(\Omega|Type_\lambda=t)}{\sum_\tau \pi_\tau L_i(\Omega|Type_\lambda=\tau)}$.

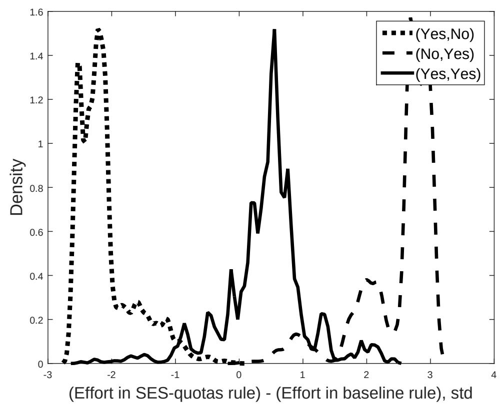
having a tournament within each socioeconomic group and major (keeping the weights constant for each PSU test and GPA), such that the seats available for students attending high schools of socioeconomic group g in university tier n and major z is equal to $S_n^z * \left(\frac{\#students\ SES\ g\ preferring\ major\ z}{\#students\ preferring\ major\ z}\right)$, in which case there are five vectors r for each major (one for each socioeconomic group). Notice that beyond the seats offered by universities for each socioeconomic group, the only aspect that change between the baseline scenario and the SES-quota system counterfactual is that in the latter instead of having two markets, one for scientific and one for humanities major, there are 10 markets, one for each combination of socioeconomic group and major offered. In other words, the properties of the equilibrium (*i.e.,* existence and the restrictions about multiplicity) are exactly the same across these 12 markets. In consequence, the algorithm to calculate each of these 12 equilibria is also the same.

Figure 10 (Appendix E.1) shows how the SES-quota system makes important changes to universities' socioeconomic composition. Because this is a tournament, where the seats and "prizes" are fixed, there are winners and losers. In particular, the figure reveals that the fraction of students admitted to the top universities who belong to SES 1, SES 2, and SES 3 increases, at the expense of higher socioeconomic groups (SES 4 and 5).

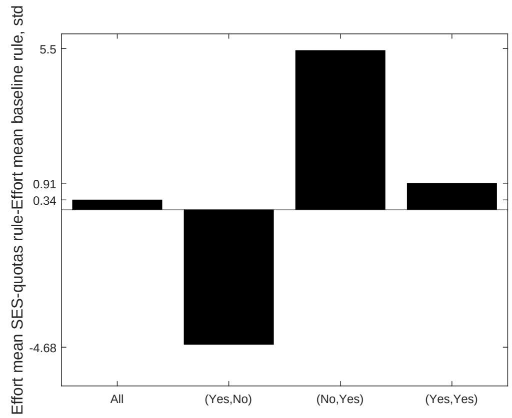
To see how changes in students' opportunities may affect their behavior in high school, Figure 4 shows that the SES-quota implementation increases the average effort of high school students by 0.34 standard deviations. Furthermore, these plots show the importance of the interaction between the two student decisions (i.e. effort and taking the PSU), in the sense that the greatest changes in effort come from those students who also change their decision about taking the college admissions test. Indeed, for those students who were not taking the test in the baseline simulation, but do take it once the admission process is changed, the increase in average effort is 5.5 standard deviations. The opposite occurs for those who change from taking to not taking the test (-4.68 standard deviations). However, even for those students who do take the college admissions test in both scenarios, there is an important increase in average effort (0.91 standard deviations). Because by construction there are no changes for those who do not take the college admissions test in both scenarios, I do not show this simulation.

Figure 4: The Impact on Effort of Quota by SES (SES-quota Rule versus Baseline Rule, in std.)

(a) Effort densities



(b) Effort means



Note: ({Yes,No},{Yes,No}) stands for (Whether the students were taking the PSU in baseline scenario,Whether the students are taking the PSU in counterfactual scenario).

Given the linear form of the tests' production function, the effect of this affirmative action policy on tests is a linear function of the effect on effort. In particular, Figure 11, in Appendix E.1, presents the numbers for those students who attend SES 1 or 2 high schools, where the average PSU score (math and verbal) increases by around 0.04 standard deviations and the average high school GPA by around 0.08. The opposite occurs for socioeconomic groups 4 and 5. In all cases, these moderate effects more than double for those who change their PSU decision. Figure 11 also shows that, even though the magnitudes of these changes are small, there is an important effect on the average final score at each university, decreasing the final score in the higher-ranked universities, and doing the opposite in the lower-ranked universities. This reveals an important trade-off between equal opportunity and efficiency, at least for the policies studied in this paper. In the sense that, even though the SES-quota system increases effort exerted by students, that positive outcome does not fully compensate the fact that – in order to do so – the SES-quota system requires to importantly break the sorting of students across universities by their final score.

As pointed out above, admissions rules also affect the test-taking decision. As my model considers, because the test is costly, students take it if they have a fair chance of being admitted to a good university. Indeed, Figure 12 (Appendix E.1) shows that the implementation of the SES-quota system increases (decreases) PSU participation by about $5-16$ percentage points for socioeconomic groups 1 and 2 (3, 4, and 5). Interestingly, for the entire population, the increase and decrease in the PSU participation almost cancel each other out. There are two reasons behind this result, whose effects interact. The first explanation is that the new admissions policy does not change the total number of seats offered by universities. Hence, no matter how opportunities are distributed across socioeconomic groups change, it is always the case that for each student who now has access to the university system, there is another student who loses it. Notice that the one-to-one relationship between losers and winners is only an *ex-post* phenomenon, because before the realization of the production test score shocks (*ex-ante*) there could be many students fighting for the same seat. The second reason is that due to the relative small size of the (estimated) variances of the production test score shocks, in both scenarios (baseline and couterfactual), the vast majority of the students who decide do not take the admission tests, have negligible probabilities of being admitted in any university in case of taking those tests. Thus, for many students the ex-ante is very similar to the ex-post scenario.

In terms of policy analysis, it is not only relevant how many students change their behavior, but also who those students are. The empirical approach followed in this paper allows for such an analysis. In particular, the second plot of Figure 12 (Appendix E.1), shows that, when the SES-Quota system is introduced, the new PSU-takers are noticeably more skilled (i.e. are a higher learning skill type) than those who decide to abandon the admissions process, i.e. do not take the PSU.

From the previous analysis, it is clear that effort is quite elastic to changes in college admissions rules. However, given the estimated parameters of the tests' production functions, these effort reactions do not imply changes of the same magnitudes for student performance. In other words, the estimated model requires large changes in college admissions rules in order to have substantial variations in high school students' performance. In this context, it is pertinent to ask how relevant this is to the model of effort. In this regard, I compare how the final-score cutoff and the admissions at each university would change, given the described counterfactual experiments, in two scenarios: (1) with optimal effort (i.e., simulating the model), and (2) with fixed effort (i.e., the effort exerted in the baseline scenario). The results plotted in Figure 13 (Appendix E.1) show that there is an important difference between the optimal effort's final-score cutoffs and the fixed effort's final-score cutoffs, given the implementation of the SES-Quota system. For example, in the case of the final-score cutoffs for SES 1 and 2, the difference between these two scenarios ranges from 0.2 to 1 standard deviations. Moreover, about 30% of the students are admitted to a different university tier in these two scenarios, a relevant figure given that I assume only eight university tiers.

7.2 Impact of the Increasing GPA weight on Final Score

In the second counterfactual experiment, I simulate what happens if the GPA weight is increased, which in practice implies that an increased probability of attending better universities for students from low income high schools. This is because, while the high school GPA of each student is, to some extent, relative to that of her classmates, the national test scores are relative to the student's national cohort. Therefore the latter captures differences in high school quality, which is highly correlated with income.

Figure 14, in Appendix E.2, shows that increasing the GPA weight from 0.4 (the baseline) to 0.5 leads to a moderate increase in the fraction of students attending top universities who come from low and medium income high schools (SES 1, 2, and 3). Figure 15 shows that, as expected, this change increases when the new GPA weight is 0.7, in which case the fraction of the students admitted to the top university tiers who belong to SES 1 and 2 increases substantially. Additionally the fraction of the students admitted to the top university tier who belong to SES 1 is also increased. All these increments are at the expense of higher socioeconomic groups (SES 4 and 5). Notice that the SES-Quota system presented in this paper is a more aggressive affirmative action policy than changing GPA weights.

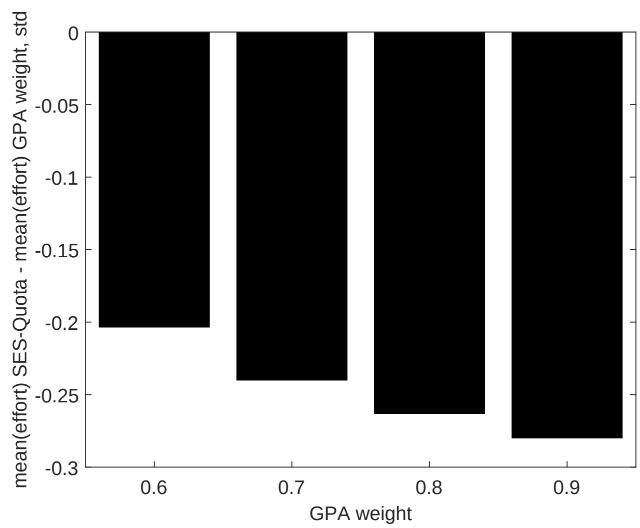
Regarding the impact of effort, Figure 16, in Appendix E.2, shows that the changes in GPA weight imply increases in students' average effort from 0.3 to 1.2 standard deviations, depending on the magnitude of the weight's change. However, this result is very influenced by changes in the decision to take the admission exam, since for those students who take the PSU in both the baseline and in the counterfactual, the increase in effort is always lower than 0.3 standard deviations.

7.3 Comparing the Efficiency of these Admission Rules

Let define an efficient student allocation, where, conditional on any particular socioeconomic composition in each university, the best students (i.e., with the highest expected final score) attend the best universities. Given that definition, I discuss which college admissions rule, presented in this paper, leads to the most efficient student allocation. To do so, I first simulate the estimated model for different GPA weights and calculate the resulting socioeconomic composition among universities from each of these exercises. Then I impose these quotas in the SES-quota system. As a result, I can compare outcomes of the two policy experiments while having the same socioeconomic composition in both cases.

As Figure 5 shows, increasing the GPA weight lead to a higher boost in average effort than for the SES-quota system. This is mainly because the estimated effort marginal productivity is much higher in the GPA production function than in the production functions of the two PSU tests.

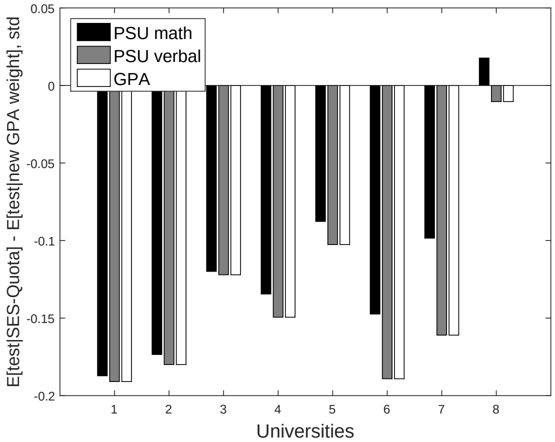
Figure 5: Average Effort: SES-quota versus Changing GPA's Weight



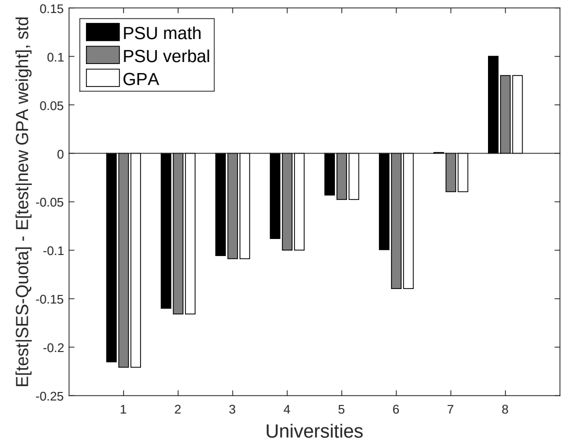
However, this does not mean that increasing the GPA weight is the preferred system to achieve equal opportunity. Instead, Figure 6 shows that the higher the GPA weight, the larger the advantage of the SES-quota system, in terms of expected PSU test scores and GPA of those admitted to the top universities. As the GPA weight increases, the GPA shock becomes more relevant in the admissions process, while in the SES-Quota system, the same equal opportunity achievement is reached by keeping the weights of the PSU tests and GPA constant, 0.4 and 0.3, respectively. Therefore the latter keeps the weights of each shock constant, which attenuates the risk of admitting a bad student to a good university due to one extremely positive shock (the three shocks are independent), in this case extremely GPA positive shock. In sum, when a policy maker pretends a relevant affirmative action policy, in spite its lower capacity to boost student academic effort, the SES-Quota system implies a more efficient student allocation, because it is able to achieve any level of equal opportunities by making better use of the existing information.

Figure 6: Expected Tests and GPA: SES-quota versus Changing GPA's Weight

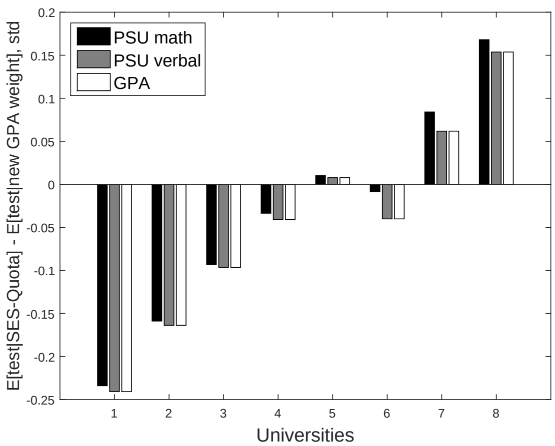
(a) GPA weight = 0.6



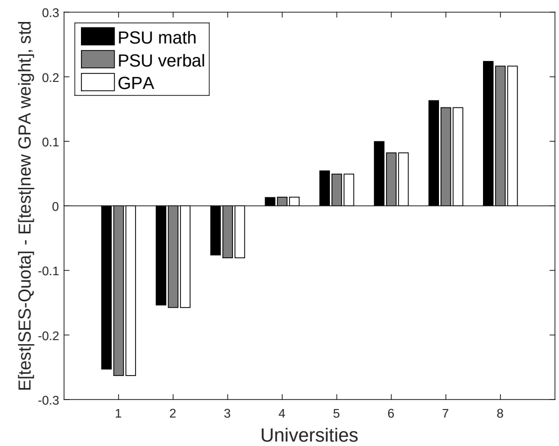
(b) GPA weight = 0.7



(c) GPA weight = 0.8



(d) GPA weight = 0.9



8 Conclusion

The first lesson from this paper is that it is qualitatively and quantitatively important to consider how a college admissions system may impact high school students' behavior. Specifically, the results of this paper support the claim that affirmative action in access to college may boost the average of the academic effort exerted by high school students. Moreover, this paper sheds some light on which admissions system could be optimal in the sense of inducing an efficient student allocation conditional on delivering the desired change in universities' socioeconomic composition.

In particular, the most important results of this paper are the following: (1) Average academic effort in high school is increased by 0.34 standard deviations under a SES-quota system. (2) This leads to a moderate improvement in high school students' performance. (3) Modeling effort and the decision to take the college admissions test are important for assessing the effect of affirmative action policies on students' allocation across universities. (4) The largest change in exerted effort comes from those students who also change their decision about taking the college admissions test. (5) Neither of the affirmative action policies simulated in the paper increase – by a significant amount – the percentage of students taking the national test for college admissions, which is consistent with the fact that there are the still the same number of winners and losers. However, there are relevant variations in who is taking the test; in particular, this percentage increases for low-income students and those who have a higher level of learning skills. (6) Although increasing the GPA weight is more effective in boosting academic effort in high school, the SES-quota system is more capable in terms of allocating the best students to the best universities, conditional on delivering the same universities' socioeconomic composition.

The intuition of the efficiency in allocating students in the SES-quota system is that, within each socioeconomic group, students are ranked based on all of the available information about their ability. Thus, the probability of misallocating a student based on a noisy part of a particular piece of information (as the other policy puts on high school GPA) is attenuated. This insight is particularly relevant when the allocation efficiency of the two admission rules are compared conditional on implementing an intensive affirmative action policy, because the weight on GPA (and on its shock) to deliver this intensity is closer to one.

In my model, life-time payoffs do not depend on individual characteristics, due to the fact that I don't have this information in the database which links tax records to admission test scores. This is a restrictive assumption, however its effect on policy evaluation could be both ways, incrementing or attenuating the effects. In Chile, as everywhere, students coming from low socioeconomic backgrounds have lower life-time payoffs, conditional on attending the same universities and majors. Thus, if the estimated parameters were constant in this new database context, this would attenuate the positive effects of affirmative action policies, to the extent that these policies would give more opportunities to students whose prizes for exerting effort are lower. However, in this new database context the estimated parameters should change, since the parameters would have to help the model to rationalize the fact that in the data even though low socioeconomic students have less monetary incentive to exert effort, they exert not that different academic effort relative to wealthier students.

Furthermore, to answer the questions discussed in this paper, it would have been best to have data before and after an admissions policy change. This ideal data would make it easier to capture the effects of admissions rules on high school students' performance. In the absence of such data, structural estimations allow for *ex ante* policy evaluation. Yet, even with the ideal data, the structural approach used in this paper would be needed in order to study the effect of several policies. The current paper is one of the first steps in studying the structural relationship between high school student effort and the probabilities of admission to a good university. The ability to estimate the structural relationship between high school academic effort and probability of admission to various universities is a powerful tool to discuss what should be the main features of an admission system that seeks to find an optimal combination of efficiency and equal opportunity.

A Existence and uniqueness

A.1 Existence in the market of major z

Lemma 2: If $\forall\, i \mid t(i) = t:\ (R_N^{z(t)} - R_1^{z(t)}) < \left(\frac{\sigma_\eta}{a_1}\right)^2 \frac{\theta_2^t}{\phi(1)}$ and $\sum_t m_t \Phi\left(\frac{(R_N^{z(t)} - R_1^{z(t)}) - \overline{FC}_t}{\sigma_{fc}}\right) > \sum_{\delta=2}^{N} S_\delta^{z(t)}$, there exists at least one equilibrium in market of z major.

Proof: To prove the lemma, I show that the conditions for the Brouwer fixed point theorem are satisfied. Let $G_n(r^{z(t)}) = r_n^{z(t)} - \sum_{\delta=n}^{N} S_\delta^{z(t)} + \sum_t m_t \Phi\left(\frac{D_t(r^{z(t)}) - \overline{FC}_t}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r_n^{z(t)} - \hat{e}_t^1(r) a_1 - a_{0t}}{\sigma_\eta}\right)\right]$, where $r = (r_2^{z(t)}, r_3^{z(t)}, ..., r_N^{z(t)}) \in \mathbb{R}^{N-1}$, then I define the vector-value function $G(r^{z(t)})$ as:[22]

$$G(r^{z(t)}) = \begin{bmatrix} G_2(r^{z(t)}) \\ G_3(r^{z(t)}) \\ . \\ . \\ . \\ G_N(r^{z(t)}) \end{bmatrix},$$

$$G(r^{z(t)}) = \begin{bmatrix} r_2^{z(t)} - \left(\sum_{\delta=2}^{N} S_\delta^{z(t)} - \sum_t m_t \Phi\left(\frac{D_t(r^{z(t)}) - \overline{FC}_t}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r_2^{z(t)} - \hat{e}_t^1(r^{z(t)}) a_1 - a_{0t}}{\sigma_\eta}\right)\right]\right) \\ r_3^{z(t)} - \left(\sum_{\delta=3}^{N} S_\delta^{z(t)} - \sum_t m_t \Phi\left(\frac{D_t(r^{z(t)}) - \overline{FC}_t}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r_3^{z(t)} - \hat{e}_t^1(r^{z(t)}) a_1 - a_{0t}}{\sigma_\eta}\right)\right]\right) \\ . \\ . \\ . \\ r_N^{z(t)} - \left(\sum_{\delta=N}^{N} S_\delta^{z(t)} - \sum_t m_t \Phi\left(\frac{D_t(r^{z(t)}) - \overline{FC}_t}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r_N^{z(t)} - \hat{e}_t^1(r^{z(t)}) a_1 - a_{0t}}{\sigma_\eta}\right)\right]\right) \end{bmatrix}.$$

Hence, proving existence for the equilibrium is equivalent to showing the existence of a fixed point for $G(r^{z(t)})$. In order to fulfill the Brouwer fixed point theorem's conditions, the vector-valued function $G: M \to M$ should be continuous and M non-empty, and a compact and convex subset of some Euclidean space $\mathbb{R}^{N-1}$.

Given that the effort decision of any student is bounded by $[\min_t\{\underline{e}_t\}, \max_i\{\bar{e}_t\}]$ it is clear that:[23]

[22] $\hat{e}_t^1(r^{z(t)})$ stands for the optimal effort decision for all students of type t who decide to take the college admissions test given the vector of cutoff scores $r^{z(t)}$.

[23] Given that $\lim_{r^{z(t)} \to -\infty} \Phi\left(\frac{r_n^{z(t)} - a_1 \hat{e}_t^1(r^{z(t)}) - a_{0t}}{\sigma_\eta}\right) = 0$, $\lim_{r^{z(t)} \to \infty} \Phi\left(\frac{r_n^{z(t)} - a_1 \hat{e}_t^1(r^{z(t)}) - a_{0t}}{\sigma_\eta}\right) = 1$, and

$$r^{z(t)} \to \infty \Rightarrow \sum_t m_t \Phi\left(\frac{D_t(r^{z(t)}) - \overline{FC}_t}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r_n^{z(t)} - e_t^1(r^{z(t)})a_1 - a_{0t}}{\sigma_\eta}\right)\right] \to 0,$$

$$\begin{aligned} r^{z(t)} \to -\infty \Rightarrow & \sum_t m_t \Phi\left(\frac{D_t(r^{z(t)}) - \overline{FC}_t}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r_n^{z(t)} - e_t^1(r^{z(t)})a_1 - a_{0t}}{\sigma_\eta}\right)\right] \\ & \to \sum_t m_t \Phi\left(\frac{(R_N^{z(t)} - R_1^{z(t)}) - \overline{FC}_t}{\sigma_{fc}}\right) > \sum_{\delta=2}^{N} S_\delta^{z(t)}. \end{aligned}$$

Then, taking any small number $\varepsilon > 0$, it is true that:

$$\begin{aligned} &\forall\ n:\ r^{z(t)} \to \infty \Rightarrow G_n(r^{z(t)} + \varepsilon * \vec{1}) - G_n(r^{z(t)}) \to \varepsilon > 0 \\ &\forall\ n:\ r^{z(t)} \to -\infty \Rightarrow G_n(r^{z(t)} - \varepsilon * \vec{1}) - G_n(r^{z(t)}) \to -\varepsilon < 0 \end{aligned}$$

Therefore, there exist two vectors $\underline{r}$ and $\bar{r}$ such that $\forall\ r^{z(t)} < \bar{r} \Rightarrow G(r^{z(t)}) < G(\bar{r}) < \bar{r}$ and $\forall\ r^{z(t)} > \underline{r} \Rightarrow G(r^{z(t)}) > G(\underline{r}) > \underline{r}$.[24] Hence, I can define the set $M = \{r^{z(t)} \in \mathbb{R}^{N-1},\ \underline{r} \leq r^{z(t)} \leq \bar{r}\}$. This set is compact, convex, and not empty.[25]

To show that $G(r^{z(t)})$ is continuous it is sufficient to prove that $\forall t\ \hat{e}_t(r^{z(t)})$ is continuous.[26] Moreover, applying the Berge's maximum theorem and considering the fact that the effort decision of any student is bounded by $[\min_t\{\underline{e}_t\}, \max_t\{\bar{e}_t\}]$ (compact set), a sufficient condition for the continuity of $e_t^1(r^{z(t)})$ is that the objective function for those students who decide to take the college admissions test is strictly concave.

Taking the derivative to the first order condition (10), it follows that:

$\lim_{r^{z(t)} \to \infty}(\hat{e}_t^1(r^{z(t)}) - \hat{e}_t^0) = 0$: as $r^{z(t)} \to -\infty$

$$\begin{aligned} D_t(r^{z(t)}) = &\left(\sum_{n=1}^{N-1}\left(R_n^{z(t)} - R_{n+1}^{z(t)}\right)\Phi\left(\frac{r_n^{z(t)} - a_1\hat{e}_t^1(r^{z(t)}) - a_{0t}}{\sigma_\eta}\right)\right) + (R_N^{z(t)} - R_1^{z(t)}) \\ & + \theta_1 b_1(\hat{e}_t^1(r^{z(t)}) - \hat{e}_t^0) - \frac{1}{\theta_2^i}\frac{(\hat{e}_t^1(r^{z(t)}))^2 - (\hat{e}_t^0)^2}{2} \to (R_N^{z(t)} - R_1^{z(t)}). \end{aligned}$$

[24] Because there exist $\bar{r}$ such that $\forall r^{z(t)} > \bar{r}$: $\sum_t m_t \Phi\left(\frac{D_t(r^{z(t)}) - \overline{FC}_t}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r_n^{z(t)} - \hat{e}_t^1(r^{z(t)})a_1 - a_{0t}}{\sigma_\eta}\right)\right] < \sum_{\delta=n}^{N} S_\delta^{z(t)}$, and $\underline{r}$ such that $\forall r^{z(t)} < \underline{r}$: $\sum_t m_t \Phi\left(\frac{D_t(r^{z(t)}) - \overline{FC}_t}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r_n^{z(t)} - \hat{e}_t^1(r^{z(t)})a_1 - a_{0t}}{\sigma_\eta}\right)\right] > \sum_{\delta=n}^{N} S_\delta^{z(t)}$.

[25] To ensure the non-emptiness, it is possible to pick $\underline{r} < 0$ and $\bar{r} > 0$.

[26] If $\hat{e}_t^1(r^{z(t)})$ is continuous then $D_t(r^{z(t)})$ is also continuous.

$$\frac{\partial^2 E[U_t^1(e)]}{\partial e^2} = \sum_{n=1}^{N-1} \left(R_{n+1}^{z(t)} - R_n^{z(t)}\right) \left(\frac{r_{n+1}^{z(t)} - a_1 e - a_{0t}}{\sigma_\eta}\right) \phi \left(\frac{r_{n+1}^{z(t)} - a_1 e - a_{0t}}{\sigma_\eta}\right) \left(\frac{a_1}{\sigma_\eta}\right)^2 - \theta_2^t.$$

But because the first term can not be bigger than $(R_N^{z(t)} - R_1^{z(t)}) \left(\frac{a_1}{\sigma_\eta}\right)^2 \phi(1)$, then[27]

$$(R_N^{z(t)} - R_1^{z(t)}) a_1^2 \phi(1) < \sigma_\eta^2 \theta_2^t \Rightarrow \frac{\partial^2 E[U_t^1(e)]}{\partial e^2} < 0.$$

Moreover, $G(r^{z(t)})$ is well defined for any $r^{z(t)}$ because, as shown above, for any $r^{z(t)}$ there exist optimal efforts for those who take the college admissions test ($\hat{e}_t^1(r^{z(t)})$) and for those who do not ($\hat{e}_t^0$). ■

Uniqueness

Lemma 3: In the case where $N = 2$, the equilibrium is unique when it exists.

Proof: The lemma is proved by contradiction. In particular, assuming there are two equilibria $\{\hat{r}, \hat{e}\}$ and $\{\hat{r}', \hat{e}'\}$, where, without loss of generality $r' > r$,[28] from the equilibrium definition, it is directly shown that:

$$S = \sum_t m_t \Phi \left(\frac{\hat{D}_t - \overline{FC}_t}{\sigma_{fc}}\right) \left[1 - \Phi \left(\frac{r - \hat{e}_t a_1 - a_{0t}}{\sigma_\eta}\right)\right],$$
$$S = \sum_t m_t \Phi \left(\frac{\hat{D}_t' - \overline{FC}_t}{\sigma_{fc}}\right) \left[1 - \Phi \left(\frac{r' - \hat{e}_t' a_1 - a_{0t}}{\sigma_\eta}\right)\right]. \quad (21)$$

To get the contradiction I proceed in two steps. First, I show that the statement: $\forall \hat{r}' > \hat{r}, t: \ \Phi\left(\frac{\hat{r}' - \hat{e}_t' a_1 - a_{0t}}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\hat{r} - \hat{e}_t a_1 - a_{0t}}{\sigma_\varepsilon}\right) > 0$, is a sufficient condition to get the desired contradiction. Second, I show that this statement is true regardless of the continuity of effort in r.

Step 1:

Let $\Pi_{0i} = \max_e E[U_i^0(e)]$ and $\Pi_{1i}(r) = \max_e E[U_i^1(e)]$, then $D_{t(i)} = \Pi_{1i}(r) + FC_i - \Pi_{0i}$.[29] Due to the envelope theorem, taking the derivative to $D_{t(i)}$ with respect to r we get:[30]

[27] The function $x\phi(x)$ is maximized at $x = 1$.

[28] Notice since $N = 2$, $\hat{r}$ and $\hat{r}'$ are scalars. S is the amount of seats offered by the only university. For simplicity I suppress the individual index z.

[29] The value function for those who do not take the college admissions test does not depend on r.

[30] Here, I am assuming that effort is continuous in r (if that is the case, the value function is differentiable), but in step 2 I also show that $\Pi_{1i}(\hat{r}) > \Pi_{1i}(\hat{r}')$ when effort is not continuous in r.

$$\frac{\partial D_{t(i)}}{\partial r} = \frac{\partial \Pi_{1i}(r)}{\partial r} = \frac{R_1^t - R_2^t}{\sigma_\eta}\phi\left(\frac{r - \hat{e}_t a_1 - a_{0t}}{\sigma_\eta}\right) < 0$$

$$\Rightarrow \hat{D}_{t(i)} > \hat{D}^{'}_{t(i)} \Rightarrow \left(\frac{\hat{D}_{t(i)} - \overline{FC}_t}{\sigma_{fc}}\right) > \left(\frac{\hat{D}^{'}_{t(i)} - \overline{FC}_t}{\sigma_{fc}}\right).$$

Therefore, from equations (21) and the later inequality, it is directly shown that:

$$\sum_t m_t \Bigg(\Phi\left(\frac{\hat{D}_{t(i)} - \overline{FC}_t}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{\hat{r} - \hat{e}_t a_1 - a_{0t}}{\sigma_\eta}\right)\right] - \Phi\left(\frac{\hat{D}^{'}_{t(i)} - \overline{FC}_t}{\sigma_{fc}}\right)$$
$$\left[1 - \Phi\left(\frac{\hat{r}' - \hat{e}^{'}_t a_1 - a_{0t}}{\sigma_\eta}\right)\right]\Bigg) = 0,$$

$$\Rightarrow \sum_t m_t \Bigg(\left[1 - \Phi\left(\frac{\hat{r} - e_t a_1 - a_{0t}}{\sigma_\eta}\right)\right] - \left[1 - \Phi\left(\frac{\hat{r}' - e^{'}_t a_1 - a_{0t}}{\sigma_\eta}\right)\right]\Bigg)$$
$$\Phi\left(\frac{\hat{D}^{'}_{t(i)} - \overline{FC}_t}{\sigma_{fc}}\right) < 0,$$

$$\Rightarrow \sum_t m_t \left(\Phi\left(\frac{\hat{r}' - \hat{e}^{'}_t a_1 - a_{0t}}{\sigma_\eta}\right) - \Phi\left(\frac{\hat{r} - e_t a_1 - a_{0t}}{\sigma_\eta}\right)\right) < 0.$$

because $m_t > 0\ \forall t$, this last inequality contradicts that $\forall \hat{r}' > \hat{r}, t:\ \Phi\left(\frac{\hat{r}' - \hat{e}^{'}_t a_1 - a_{0t}}{\sigma_\varepsilon}\right) - \Phi\left(\frac{\hat{r} - \hat{e}_t a_1 - a_{0t}}{\sigma_\varepsilon}\right) > 0$. ■

Step 2:

I prove this inequality in two steps. First, I prove it for those r where the effort decision is continuous. Then I show the inequality when the effort decision is not continuous in r.

Case 1: effort decision is continuous in r:

First, it should be considered that the second order condition when $N = 2$ implies that:[31]

$$\frac{a_1(R_2 - R_1)}{\sigma_\eta}\left(\frac{r - \hat{e}a_1 - a_0}{\sigma_\eta}\right)\phi(r)\frac{a_1}{\sigma_\eta} < \theta_2.$$

[31] For simplicity, I suppress the individual sub-index and denote $\phi\left(\frac{r - \hat{e}a_1 - a_0}{\sigma_\eta}\right)$ as $\phi(r)$.

Moreover, taking a derivative of the first order condition (10), when $N = 2$ implies:

$$\frac{\partial \hat{e}}{\partial r} = \frac{1}{\theta_2}\left[\frac{(R_2 - R_1)a_1}{\sigma_\eta}\left(\frac{r - a_1\hat{e} - a0}{\sigma_\eta}\right)\phi(r)\left(\frac{1 - a_1\frac{\partial \hat{e}}{\partial r}}{\sigma_\eta}\right)\right],$$
$$\Rightarrow \frac{\partial \hat{e}}{\partial r} < \frac{1}{\theta_2}\left[\frac{\sigma_\eta\theta_2}{a_1}\left(\frac{1 - a_1\frac{\partial \hat{e}}{\partial r}}{\sigma_\eta}\right)\right] = \frac{1}{a_1}\left(1 - a_1\frac{\partial \hat{e}}{\partial r}\right).$$

Therefore, it should be the case that $\left(1 - a_1\frac{\partial \hat{e}}{\partial r}\right) > 0$. Because if $\frac{\partial \hat{e}}{\partial r}$ is negative, the result is direct and if $\frac{\partial \hat{e}}{\partial r}$ is positive, the inequality delivers the result.

This result allows me to get the desired inequality:

$$\frac{\partial \Phi\left(\frac{r - \hat{e}a_1 - a_0}{\sigma_\eta}\right)}{\partial r} = \phi\left(\frac{r - \hat{e}a_1 - a_0}{\sigma_\eta}\right)\left(1 - \frac{\partial \hat{e}}{\partial r}a_1\right)\frac{1}{\sigma_\eta} > 0.$$

Therefore, it follows that $\frac{\partial \Phi\left(\frac{r - \hat{e}a_1 - a_0}{\sigma_\eta}\right)}{\partial r} > 0$. ■

Case 2: effort decision is discontinuous at r:[32]

Without loss of generality, assume there are two different effort decisions which are optimal at r ($\hat{e}_h > \hat{e}_l$). Defining $\Pi_x = (R_1 - R_2)\Phi\left(\frac{r - \hat{e}_x a_1 - a_0}{\sigma_\eta}\right) + R_2 + \theta_1(b_0 + b_1\hat{e}_x) - \theta_2\frac{\hat{e}_x^2}{2}$, $x = l, h$ (the value function for each local equilibrium) and applying the envelope theorem implies:

$$\frac{\partial \Pi_l}{\partial r} - \frac{\partial \Pi_h}{\partial r} = \frac{(R_2 - R_1)}{\sigma_\eta}\left[\phi\left(\frac{r - \hat{e}_l a_1 - a_0}{\sigma_\eta}\right) - \phi\left(\frac{r - \hat{e}_h a_1 - a_0}{\sigma_\eta}\right)\right]. \tag{22}$$

Moreover, from the first order conditions it is directly shown that:

$$\hat{e}_h - \hat{e}_l = \frac{a_1(R_2 - R_1)}{\theta_2\sigma_\eta}\left[\phi\left(\frac{r - \hat{e}_l a_1 - a_0}{\sigma_\eta}\right) - \phi\left(\frac{r - \hat{e}_h a_1 - a_0}{\sigma_\eta}\right)\right]$$
$$\Rightarrow \frac{\partial \Pi_l}{\partial r} - \frac{\partial \Pi_h}{\partial r} = \frac{\theta_2(\hat{e}_h - \hat{e}_l)}{a_1} > 0 \tag{23}$$

[32]Given that the discontinuity is possible only for those who take the college admissions test, for this proof I assume all take the college admissions test.

Therefore, by (23) I proved that increasing r leads to some jump in the global optimal effort from high local optimal effort to low local optimal effort, which ensured that $\forall \hat{r}' > \hat{r}$ such that the effort decision is not continuous at r for students type t, then $\Phi\left(\frac{\hat{r}'-\hat{e}'_t a_1 - a_{0t}}{\sigma_\eta}\right) - \Phi\left(\frac{\hat{r}-\hat{e}_t a_1 - a_{0t}}{\sigma_\eta}\right) > 0$. ■

In the case where $N > 2$, as in this paper, it can be established that $\sum_{n=2}^{N} \frac{\partial \overline{G_m}}{\partial r_n} < 0 \; \forall m$, where $\overline{G_m} = G_m - r_m$. This result implies that if $\overline{G(\hat{r})} = 0$ (i.e., $\hat{r}$ is an equilibrium), then $\hat{r}' = \hat{r}(a+1)$ where $a \neq 0$, cannot be an equilibrium.[33] Loosely speaking, this means that if there is an equilibrium denoted by $\hat{r}$, the further $\hat{r}'$ departs from $\hat{r}$ the harder it is to have $\hat{r}'$ as another equilibrium.

To prove the statement, I proceed in two steps.[34] First, it is proved that $\sum_{n=2}^{N} \frac{\partial \overline{G_m}}{\partial r_n} < -\sum_t m_t \Phi\left(\frac{\hat{D}_t - \overline{FC}_t}{\sigma_{fc}}\right) \frac{\phi_t(r)}{\sigma_\eta} \left[1 - a_1 \sum_{n=2}^{N} \frac{\partial \hat{e}_t^1}{\partial r_n}\right]$.[35] Second, I show that $1 - a_1 \sum_{n=2}^{N} \frac{\partial \hat{e}_t^1}{\partial r_n} > 0 \; \forall t$.

To get the first result, notice that:

$$\forall n \neq m: \; \frac{\partial \overline{G_m}}{\partial r_n} = \sum_t m_t \phi\left(\frac{\hat{D}_t - \overline{FC}_t}{\sigma_{fc}}\right) [1 - \Phi_t(r_m)] \frac{\partial \hat{D}_t}{\partial r_n} \frac{1}{\sigma_{fc}} + \sum_t m_t \Phi\left(\frac{\hat{D}_t - \overline{FC}_t}{\sigma_{fc}}\right) \phi_t(r_m) \frac{\partial \hat{e}_t^1}{\partial r_n} \frac{a_1}{\sigma_{fc}} < \sum_t m_t \Phi\left(\frac{\hat{D}_t - \overline{FC}_t}{\sigma_{fc}}\right) \phi_t(r_m) \frac{\partial \hat{e}_t^1}{\partial r_n} \frac{a_1}{\sigma_{fc}}.$$

$$\frac{\partial \overline{G_m}}{\partial r_m} = \sum_t m_t \phi\left(\frac{\hat{D}_t - \overline{FC}_t}{\sigma_{fc}}\right) [1 - \Phi_t(r_m)] \frac{\partial \hat{D}_t}{\partial r_m} \frac{1}{\sigma_{fc}} - \sum_t m_t \Phi\left(\frac{\hat{D}_t - \overline{FC}_t}{\sigma_{fc}}\right) \frac{\phi_t(r_m)}{\sigma_{fc}} \left[1 - \frac{\partial \hat{e}_t^1}{\partial r_m} a_1\right] < -\sum_t m_t \Phi\left(\frac{\hat{D}_t - \overline{FC}_t}{\sigma_{fc}}\right) \frac{\phi_t(r_m)}{\sigma_{fc}} \left[1 - \frac{\partial \hat{e}_t^1}{\partial r_m} a_1\right],$$

where both inequalities are driven by the fact that $\frac{\partial \hat{D}_t}{\partial r_m} < 0$. From these two inequalities the first result follows:

[33] It would be better to show that this is true even when the increase (or decrease) is not proportional across score cutoffs. However, i could stablish that result.

[34] For simplicity the result is shown for the case where G is continuous.

[35] $\phi_t(r_m) = \phi\left(\frac{r_m - \hat{e}_t^1 a_1 - a_{0t}}{\sigma_\eta}\right)$ and $\Phi_t(r_m) = \phi\left(\frac{r_m - \hat{e}_t^1 a_1 - a_{0t}}{\sigma_\eta}\right)$.

$$\sum_{n=2}^{N} \frac{\partial \overline{G_m}}{\partial r_n} < -\sum_t m_t \Phi\left(\frac{\hat{D}_t - \overline{FC}_t}{\sigma_{fc}}\right)\frac{\phi_t(r_m)}{\sigma_{fc}} + \sum_{n=2}^{N}\sum_t m_t \Phi\left(\frac{\hat{D}_t - \overline{FC}_t}{\sigma_{fc}}\right)\phi_t(r_m)\frac{\partial \hat{e}_t^1}{\partial r_n}\frac{a_1}{\sigma_{fc}},$$
$$= -\sum_t m_t \Phi\left(\frac{\hat{D}_t - \overline{FC}_t}{\sigma_{fc}}\right)\frac{\phi_t(r_m)}{\sigma_{fc}}\left[1 - a_1 \sum_{n=2}^{N}\frac{\partial \hat{e}_t^1}{\partial r_n}\right].$$

To establish the second result, I take the derivative of the first order condition respect to r_m:

$$\frac{\partial \hat{e}_t^1}{\partial r_m} = \frac{1}{\theta_2^t}\sum_{n=1}^{N-1}(R_{n+1}^t - R_n^t)\left(\frac{r_n - \hat{e}_t^1 a_1 - a_{0t}}{\sigma_\eta}\right)\phi_t(r_n)\left(\frac{a_1}{\sigma_\eta}\right)^2 \frac{\partial \hat{e}_t^1}{\partial r_m} -$$
$$\frac{1}{\theta_2^t}(R_m^t - R_{m-1}^t)\left(\frac{r_m - \hat{e}_t^1 a_1 - a_{0t}}{\sigma_\eta}\right)\phi_t(r_m)\frac{a_1}{\sigma_\eta^2}$$
$$= \frac{-\frac{1}{\theta_2^t}(R_m^t - R_{m-1}^t)\left(\frac{r_m - \hat{e}_t^1 a_1 - a_{0i}}{\sigma_\eta}\right)\phi_i(r_m)\frac{a_1}{\sigma_\eta^2}}{1 - \frac{1}{\theta_2^t}\sum_{n=1}^{N-1}(R_{n+1}^t - R_n^t)\left(\frac{r_n - \hat{e}_i^1 a_1 - a_{0i}}{\sigma_\eta}\right)\phi_i(r_n)\left(\frac{a_1}{\sigma_\eta}\right)^2}$$

$$\Rightarrow 1 - a_1 \sum_{n=1}^{N-1}\frac{\partial \hat{e}_t^1}{\partial r_n} = \frac{1}{1 - \frac{1}{\theta_2^t}\sum_{n=1}^{N-1}(R_{n+1}^t - R_n^t)\left(\frac{r_n - \hat{e}_t^1 a_1 - a_{0t}}{\sigma_\eta}\right)\phi_t(r_n)\left(\frac{a_1}{\sigma_\eta}\right)^2} > 0,$$

where the inequality is because the denominator is positive due to the second order condition of student maximization. ■

B Variables: Descriptions and Statistics

Table 3: Independent Variables

	Full Sample			Estimation Sample			Statistics		
Variable	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Diff.	Stat.[0]	P-Val.
Gender[36]	0.48	0.50	212656	0.47	0.50	146319	0.01	3.15	0.00
Mother's Education									
No Information[37]	0.28	0.45	212656	0.16	0.37	146319	0.11	52.59	0.00
Incomplete Primary[37]	0.11	0.31	212656	0.12	0.32	146319	-0.01	-6.24	0.00
Primary[37]	0.21	0.41	212656	0.24	0.42	146319	-0.03	-12.38	0.00
Secondary[37]	0.23	0.42	212656	0.26	0.44	146319	-0.04	-17.21	0.00
Technical Post-Secondary[37]	0.10	0.30	212656	0.11	0.32	146319	-0.02	-9.36	0.00
College[37]	0.09	0.28	212656	0.10	0.30	146319	-0.02	-9.72	0.00
Father's Education									
College[37]	0.12	0.32	212656	0.14	0.35	146319	-0.02	-11.96	0.00
Primary School Type									
Public[37]	0.50	0.50	174883	0.49	0.50	146319	0.00	1.11	0.27
Subsidized Private[37]	0.41	0.49	174883	0.41	0.49	146319	0.00	0.02	0.99
Private[37]	0.09	0.29	174883	0.10	0.29	146319	-0.00	-1.47	0.14
High School Type									
Public[37]	0.41	0.49	212656	0.39	0.49	146319	0.02	7.18	0.00
Subsidized Private[37]	0.51	0.50	212656	0.52	0.50	146319	-0.01	-4.15	0.00
Private[37]	0.08	0.28	212656	0.09	0.29	146319	-0.01	-3.92	0.00

[0] This column corresponds to the statistic from an equality of means test between the Estimation Sample and the Full Sample. If binary, a two-tailed z-test of proportions is performed. If else, a two-tailed test t-test is performed.

[36] 1 if male, 0 if female

[37] Binary variable

Table 3 – *Continued from previous page*

	Full Sample			Estimation Sample			Statistics		
Variable	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Diff.	Stat.[0]	P-Val.
Primary School SES[38]									
SES 1[37]	0.08	0.27	174883	0.08	0.27	146319	-0.00	-0.48	0.63
SES 2[37]	0.30	0.46	174883	0.29	0.45	146319	0.01	2.47	0.01
SES 3[37]	0.36	0.48	174883	0.36	0.48	146319	0.00	0.18	0.86
SES 4[37]	0.18	0.39	174883	0.18	0.39	146319	-0.00	-1.07	0.29
SES 5[37]	0.09	0.28	174883	0.09	0.29	146319	-0.00	-1.65	0.10
High School SES[38]									
SES 1[37]	0.17	0.38	212656	0.16	0.37	146319	0.01	5.37	0.00
SES 2[37]	0.39	0.49	212656	0.37	0.48	146319	0.01	6.09	0.00
SES 3[37]	0.25	0.43	212656	0.26	0.44	146319	-0.01	-3.87	0.00
SES 4[37]	0.11	0.31	212656	0.12	0.33	146319	-0.01	-4.99	0.00
SES 5[37]	0.08	0.27	212656	0.09	0.28	146319	-0.01	-4.08	0.00
School Location									
Rural Primary School[37]	0.10	0.31	174883	0.11	0.31	146319	-0.00	-0.58	0.56
Rural High School[37]	0.04	0.20	212528	0.04	0.19	146319	0.00	1.96	0.05
Primary School Scores									
Math Simce Score	266.65	48.99	173481	268.64	48.50	145236	-1.99	-11.47	0.00
Verbal Simce Score	265.54	49.08	174394	267.56	48.34	145944	-2.02	-11.70	0.00
Natural Science Simce Score	269.40	49.31	174680	271.28	48.84	146177	-1.88	-10.80	0.00
Social Science Simce Score	264.08	48.02	173195	265.98	47.60	145011	-1.90	-11.14	0.00
GPA Score	5.84	0.53	181551	5.87	0.52	146319	-0.03	-17.21	0.00
High School Scores									
Math Simce Score	263.06	63.75	185994	268.59	62.68	146041	-5.53	-25.02	0.00
Verbal Simce Score	262.22	51.16	186060	266.25	50.49	146083	-4.03	-22.66	0.00
Math PSU Score	497.68	111.34	156243	508.20	110.70	113946	-10.52	-24.33	0.00
Spanish PSU Score	495.48	110.10	156243	505.05	108.85	113946	-9.57	-22.72	0.00

[38] SES refers to one of the five School Socioeconomic Groups

Table 3 – *Continued from previous page*

Variable	Full Sample			Estimation Sample			Statistics		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N	Diff.	Stat.[0]	P-Val.
Takes PSU[37]	0.73	0.44	212656	0.78	0.42	113946	-0.04	-19.86	0.00
GPA Score	520.32	102.82	212517	537.39	100.93	146319	-17.07	-49.25	0.00
Primary School Effort Measures Me^p									
Exerts effort at hard subjects [39]	1.73	0.82	171092	1.72	0.81	146319	0.01	2.11	0.03
Try hard to learn[39]	2.08	1.01	172021	2.08	1.00	146319	0.00	0.06	0.95
Looks for additional information [39]	1.53	0.72	171270	1.52	0.72	146319	0.00	1.82	0.07
Attendance, 8th Primary Grade [39]	95.23	6.48	181551	95.71	3.87	146319	-0.48	-25.07	0.00
Spanish Study Intensity [39]	2.60	0.73	171123	2.60	0.72	146319	-0.00	-0.24	0.81
Math Study Intensity [39]	2.54	0.80	170971	2.53	0.80	146319	0.00	0.85	0.39
High School Effort Measures Me^h									
Attendance[40]	92.61	5.74	212656	93.78	3.86	146319	-1.17	-67.94	0.00
Exerts Effort [39]	0.26	0.44	174481	0.27	0.44	137532	-0.01	-3.78	0.00
Uses Calculator[39]	3.70	0.83	139176	3.71	0.83	111366	-0.01	-3.84	0.00
Uses Textbooks[39]	4.00	0.80	143604	4.02	0.79	114742	-0.01	-4.42	0.00
Studies at Home[39]	4.17	0.82	115287	4.18	0.82	92329	-0.02	-4.16	0.00
Learning Skills Measures $M\lambda^p$									
Able to understand hard subjects[39]	1.90	0.82	172281	1.89	0.81	145938	0.01	1.73	0.08
I Trust I can excel at exams[39]	1.60	0.74	171849	1.59	0.74	145768	0.01	2.49	0.01
Can set learning goals[39]	1.33	0.60	171511	1.33	0.59	145581	0.01	2.67	0.01
Can avoid poor marks[39]	1.56	0.80	171065	1.56	0.80	145261	0.01	1.76	0.08
Repeated at least 1 grade[37]	0.08	0.27	171536	0.07	0.25	144128	0.01	3.62	0.00

[39] Can be integers between 1 and 4 (1 is always or almost always, 2 is often, 3 is ocasionally and 4 is never or almost never)
[40] Percentage of attendance during the 2nd High School grade

The definition of SES (socioeconomic groups) was made by the Ministry of Education using cluster analysis and four variables: a) father's years of education, b) mother's years of education, c) monthly family income (declared), and d) an index of student vulnerability.

To characterize student families I only use information from the 2006 SIMCE . Using 2004 information would have lead to an increased loss of data.

C Likelihood

Let $T_i = \beta_0^T + x_i^h\beta_1^T + e_i^h\beta_2^T + \lambda_i\beta_3^T + \varepsilon_i^T$, such that

$$T_i \in \{PSUM_i, PSUV_i, GPA_i^h, SIMCE_{math,i}^h, SIMCE_{verbal,i}^h\}.$$

Given that conditional on λ_i, x_i^h and e_i^h, the ε_i are independent across tests, the contribution of the individual i's test to the likelihood is given by:

If $T_i \in \{PSUM_i, PSUV_i\}$:

$$f(T_i|x_i^h, e_i^h, \lambda_t, \Omega) = \left[\phi\left(\frac{T_i - \beta_0^T - x_i^h\beta_1^T - e_i^{1h}\beta_2^T - \lambda_t\beta_3^T}{\sigma_{\varepsilon^T}}\right)\frac{1}{\sigma_{\varepsilon^T}}\right] \; if \; TCAT_i = 1$$

$$Pr(TCAT|x_i^h, e_i^h, \lambda_t, \Omega) = \Phi\left(\frac{D_i - \overline{FC}}{\sigma_{fc}}\right)^{TCAT_i}\left(1 - \Phi\left(\frac{D_i - \overline{FC}}{\sigma_{fc}}\right)\right)^{1-TCAT_i} \tag{24}$$

If $T_i \in \{GPA_i^h, SIMCE_{math,i}^h, SIMCE_{verbal,i}^h\}$:

$$f(T_i|x_i^h, TCAT_i, e_i^h, \lambda_t, \Omega) =$$

$$\phi\left(\frac{T_i - \beta_0^T - x_i^h\beta_1^T - e_i^{1h}\beta_2^T - \lambda_t\beta_3^T}{\sigma_{\varepsilon^T}}\right)^{TCAT_i}\phi\left(\frac{T_i - \beta_0^T - x_i^h\beta_1^T - e_i^{0h}\beta_2^T - \lambda_t\beta_3^T}{\sigma_{\varepsilon^T}}\right)^{1-TCAT_i}\frac{1}{\sigma_{\varepsilon^T}}$$

$$F_i(high\ school\ tests \mid Type_\lambda = t) = \prod_{T_i} f(T_i|x_i^h, e_i^h, \lambda_t, \Omega) \tag{25}$$

Similarly, the contributions to the likelihood of high school effort measures are described by:[41]

[41] For simplicity the effort measurements are assumed to be continuous, but in the estimation I use ordered probit specifications.

$$f(Me^h_{ji}|x^{ejh}_i,e^h_i,TCAT_i,\Omega) = \left[\phi\left(\frac{Me^h_{ji} - x^{ejh}_i\beta^{ejh}_1 - e^{1h}_i\alpha^{ejh}}{\sigma_{\varepsilon^{ejh}}}\right)\frac{1}{\sigma_{\varepsilon^{ejh}}}\right]^{TCAT_i}$$

$$\left[\phi\left(\frac{Me^h_{ji} - x^{ejh}_i\beta^{ejh}_1 - e^{0h}_i\alpha^{ejp}}{\sigma_{\varepsilon^{ejh}}}\right)\frac{1}{\sigma_{\varepsilon^{ejh}}}\right]^{1-TCAT_i}, \; j \in \{1, ..., J_{eh}\}$$

$$F_i(high\ school\ effort\ measures) = \prod_j f(Me^h_{ji}|x^{ejh}_i, e^h_i, TCAT_i, \Omega) \tag{26}$$

Along the same lines, the contributions to the likelihood of the unobserved learning skill measures are described by:[42]

$$f(M\lambda^p_{ji}|x^{\lambda jp}_i, \lambda_t, \Omega) = \phi\left(\frac{M\lambda^p_{ji} - x^{\lambda jp}_i\beta^{\lambda jp}_1 - \lambda_t\alpha^{\lambda jp}}{\varepsilon^{ejh}}\right)\frac{1}{\sigma_{\varepsilon^{ejh}}}, \; j \in \{1, ..., J_\lambda\}$$

$$F_i(learning\ skill\ measures \mid Type_\lambda = t) = \prod_j f(M\lambda^p_{ji}|x^{\lambda jp}_i, \lambda_t, \Omega) \tag{27}$$

Let $T_i = \beta^T_0 + x^h_i\beta^T_1 + e^p_i\beta^T_2 + \lambda_i\beta^T_3 + \varepsilon^T_i$, such that

$T_i \in \{GPA^p_i, SIMCE^p_{math,i}, SIMCE^p_{verbal,i}, SIMCE^p_{socialscience,i}, SIMCE^p_{naturalscience,i}\}$.

Given that, conditional on λ_i, x^h_i and e^h_i, the ε_i are independent across tests, the contribution to the likelihood is given by[43]:

$$f(T_i|x^p_i, e^p_i, \lambda_t, \Omega) = \phi\left(\frac{T_i - \beta^T_0 - x^h_i\beta^T_1 - (\widehat{Me}^p_{1i} - x^{e1p}_i\beta^{e1p}_1)\beta^T_2 + \lambda_t\beta^T_3}{\sigma_{\varepsilon^T}}\right)\frac{1}{\sigma_{\omega^T}}$$

$$F_i(primary\ school\ tests \mid Type_\lambda = t) = \prod_{T_i} f(T_i|x^p_i, e^p_i, \lambda_t, \Omega) \tag{28}$$

---

[42] Again, these measures are assumed to be continuous, but in the estimation I use ordered probit specifications.

[43] $\widehat{Me}^p_{1i} = \widehat{\delta}_1 + \sum_{m=2}^{J_{ep}} Me^p_{mi}\widehat{\delta}_m$ and $\omega^T_i = \varepsilon^T_i - \varepsilon^{e1p}_i\beta^T_2$, where the $\widehat{\delta}s$ are the OLS coefficients.

Therefore, the likelihood contribution for the ith individual is:

$$L_i(\Omega) = log\Bigg(\sum_t F_i(high\ school\ tests \mid Type_\lambda = t)\ F_i(high\ school\ effort\ measures \mid Type_\lambda = t)$$
$$F_i(learning\ skill\ measures \mid Type_\lambda = t)\ F_i(primary\ school\ tests \mid Type_\lambda = t)\ \pi_t \Bigg) \quad (29)$$

D Results

D.1 First stage parameters

Table 4: Primary School Attending Regression

Variable	Primary School Attendance
Exert effort at hard subjects: *often*	-0.133***
	(0.0256)
Exert effort at hard subjects: *occasionally*	-0.214***
	(0.0360)
Exert effort at hard subjects: *never or almost never*	-0.218***
	(0.0787)
Try hard to learn: *often*	0.0586**
	(0.0258)
Try hard to learn: *occasionally*	0.115***
	(0.0284)
Try hard to learn: *never or almost never*	0.124***
	(0.0393)
Look for additional information: *Often*	-0.0500*
	(0.0263)
Look for additional information: *occasionally*	-0.123***
	(0.0422)
Look for additional information: *never or almost never*	-0.280***
	(0.107)
Language study intensity: *some days a week*	0.0140
	(0.0473)
Language study intensity: *only for exams*	-0.0481
	(0.0497)
Spanish study intensity: *never*	-0.380***
	(0.0672)
Math study intensity: *some days a week*	0.0277
	(0.0405)
Math study intensity: *only for exams*	-0.0286
	(0.0432)
Math study intensity: *never*	-0.0208
	(0.0573)
Constant	95.82***
	(0.0435)
Observations	146,319
R-squared	0.002
F statistic	16.99

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. In the case of the question whether she exerts effort at hard subjects, the omitted category is is *always or almost always*. In the case of the question about language study intensity, the omitted variable is *every or almost every day*.

Table 5: Grade Retention Before 8th Grade (Linear Probability Model)

Variable	Grade Retention before 8th grade
Able to understand hard subjects: *often*	0.00174
	(0.00154)
Able to understand hard subjects: *occasionally*	0.0290***
	(0.00224)
Able to understand hard subjects: *never or almost never*	0.0307***
	(0.00581)
Trust about skills to do homework and exams: *often*	0.00425**
	(0.00166)
Trust about skills to do homework and exams: *occasionally*	0.0212***
	(0.00283)
Trust about skills to do homework and exams: *never or almost never*	0.0287***
	(0.00947)
Can set learning goals: *often*	0.00502**
	(0.00196)
Can set learning goals: occasionally	0.0172***
	(0.00404)
Can set learning goals: *never or almost never*	0.0183
	(0.0132)
Can avoid poor marks: *often*	0.000835
	(0.00166)
Can avoid poor marks: *occasionally*	0.0142***
	(0.00300)
Can avoid poor marks: *never or almost never*	0.00883**
	(0.00416)
Constant	0.0532***
	(0.00107)
Observations	141,916
R-squared	0.006
F statistic	58.09

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1. In the case of all the questions, the omitted category is *always or almost always*.

Table 6: High School Attending Regression

Variable	High School Attendance
Persistence in exerting effort (Parents' opinion)	0.152***
	(0.0278)
Do homework at the proper space at home: *almost never*	0.111
	(0.0819)
Do homework at the proper space at home: *frequently*	0.371***
	(0.0791)
Do homework at the proper space at home: *almost always*	0.498***
	(0.0798)
Use of calculator to study at home: *almost never*	0.376***
	(0.0771)
Use of calculator to study at home: *frequently*	0.837***
	(0.0778)
Use of calculator to study at home: *almost always*	0.873***
	(0.0803)
Read textbooks at home: *almost never*	-0.635***
	(0.102)
Read textbooks at home: *frequently*	-0.353***
	(0.0397)
Read textbooks at home: *almost always*	-0.0721**
	(0.0312)
Constant	93.14***
	(0.105)
Observations	83,366
R-squared	0.013
F statistic	103.2

Note: Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$. In the case of all the questions, the omitted category is *never*.

Table 7: Two Stage Least Square for Primary Education Students' Performance

Variables	Verbal SIMCE	Math SIMCE	Natural Science SIMCE	Social Science SIMCE	GPA
Male	-8.655***	9.499***	9.414***	9.565***	-0.164***
	(0.232)	(0.222)	(0.228)	(0.230)	(0.00258)
Rural School	1.562***	1.439***	3.548***	2.933***	0.0627***
	(0.463)	(0.449)	(0.434)	(0.459)	(0.00529)
School $SES = 2$	1.055**	-0.880*	0.0666	0.626	-0.0681***
	(0.529)	(0.513)	(0.488)	(0.526)	(0.00605)
School $SES = 3$	10.99***	8.463***	9.793***	11.44***	-0.0697***
	(0.581)	(0.559)	(0.536)	(0.573)	(0.00652)
School $SES = 4$	28.05***	28.06***	29.29***	30.13***	-0.0390***
	(0.653)	(0.634)	(0.615)	(0.643)	(0.00724)
School $SES = 5$	42.33***	48.86***	46.75***	42.92***	0.0632***
	(1.040)	(1.016)	(1.044)	(1.007)	(0.0114)
Mother's Edu: Incomplete Primary	-6.537***	-5.747***	-6.131***	-5.259***	-0.0186***
	(0.454)	(0.430)	(0.431)	(0.447)	(0.00512)
Mother's Edu: Primary	-1.687***	-1.345***	-2.424***	-1.734***	0.0456***
	(0.380)	(0.359)	(0.367)	(0.374)	(0.00424)
Mother's Edu: Secondary	6.452***	5.097***	5.362***	6.826***	0.138***
	(0.374)	(0.355)	(0.366)	(0.368)	(0.00414)
Mother's Edu: Technical post Secondary	9.035***	7.438***	8.870***	10.36***	0.146***
	(0.467)	(0.452)	(0.469)	(0.462)	(0.00517)
Mother's Edu: College	14.68***	14.04***	16.16***	16.51***	0.209***
	(0.526)	(0.508)	(0.529)	(0.519)	(0.00585)
Faher's Edu: College	6.646***	7.574***	8.040***	6.743***	0.0583***
	(0.410)	(0.401)	(0.420)	(0.407)	(0.00455)
School type: Subsidized Private	1.392***	0.733***	2.320***	2.267***	-0.103***
	(0.289)	(0.276)	(0.283)	(0.287)	(0.00318)
School type: Non Subsidized Private	-0.515	-1.043	1.070	-2.413***	-0.136***
	(0.851)	(0.830)	(0.882)	(0.822)	(0.00923)
Attendance	-6.594***	-13.26***	-8.170***	-7.535***	0.262***
	(0.751)	(0.711)	(0.748)	(0.732)	(0.00813)
Grade Retention before 8th grade	-347.8***	-430.5***	-360.9***	-379.0***	-6.906***
	(5.985)	(5.640)	(5.874)	(5.837)	(0.0652)
Constant	838.8***	1,387***	997.0***	955.6***	-18.65***
	(71.99)	(68.18)	(71.72)	(70.25)	(0.780)
Observations	143,646	142,964	143,889	142,747	144,028
R-squared	0.202	0.278	0.247	0.201	0.143

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Two Stage Least Square for Secondary Education Students' Performance

Variables	Verbal SIMCE	Math SIMCE	Math PSU	Verbal PSU	GPA
Male	-7.136***	11.66***	23.28***	3.641***	-28.31***
	(0.271)	(0.322)	(0.579)	(0.594)	(0.572)
Rural School	-3.703***	-5.968***	-12.84***	-12.15***	-0.624
	(0.763)	(0.927)	(1.768)	(1.869)	(1.650)
School $SES = 2$	10.02***	13.94***	22.80***	24.38***	-4.249***
	(0.443)	(0.530)	(1.058)	(1.091)	(0.935)
School $SES = 3$	29.05***	39.20***	82.08***	77.91***	11.86***
	(0.506)	(0.603)	(1.142)	(1.178)	(1.078)
School $SES = 4$	44.82***	65.23***	134.3***	122.1***	30.27***
	(0.635)	(0.750)	(1.357)	(1.420)	(1.339)
School $SES = 5$	53.15***	79.01***	166.9***	151.2***	57.56***
	(1.289)	(1.443)	(2.538)	(2.630)	(2.818)
Mother's Edu: Incomplete Primary	-4.213***	0.235	-1.655	-9.321***	16.25***
	(1.456)	(1.737)	(3.323)	(3.430)	(2.888)
Mother's Edu: Primary	-0.757	4.018**	2.099	-4.819	14.84***
	(1.430)	(1.704)	(3.238)	(3.348)	(2.830)
Mother's Edu: Secondary	6.147***	11.08***	15.17***	10.40***	24.08***
	(1.427)	(1.699)	(3.221)	(3.329)	(2.824)
Mother's Edu: Technical post Secondary	9.046***	13.78***	20.05***	18.37***	26.28***
	(1.459)	(1.734)	(3.271)	(3.381)	(2.899)
Mother's Edu: College	16.54***	21.63***	37.82***	37.63***	43.86***
	(1.482)	(1.760)	(3.310)	(3.421)	(2.959)
Father's Edu: College	6.973***	9.051***	20.59***	20.08***	16.39***
	(0.451)	(0.528)	(0.880)	(0.913)	(0.975)
School type: Subsidized Private	-2.134***	-2.407***	-14.71***	-11.13***	-13.03***
	(0.315)	(0.378)	(0.693)	(0.710)	(0.657)
School type: Non Subsidized Private	-1.673	-0.909	-1.440	-1.454	-9.691***
	(1.167)	(1.298)	(2.256)	(2.316)	(2.555)
Attendance	3.817***	4.724***	11.52***	10.54***	26.94***
	(0.296)	(0.356)	(0.641)	(0.657)	(0.614)
Grade Retention before 8th grade	-316.3***	-462.2***	-736.7***	-675.2***	-971.5***
	(6.754)	(8.013)	(14.32)	(14.95)	(13.58)
Constant	-90.85***	-185.4***	-611.9***	-512.2***	-1,936***
	(27.87)	(33.50)	(60.30)	(61.84)	(57.73)
Observations	107,632	107,613	86,817	86,817	107,766
R-squared	0.239	0.303	0.426	0.374	0.166

Note: Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

D.2 Second stage parameters

Table 9: Second Stage Parameters: Point Estimates and Standard Errors

Utility								
θ_1	0.253	(0.0080)	θ_2^2 $(LikeStudy = 3)$	0.001	(0.0001)	$\overline{FC}_2$ $(\lambda\ Group = 3)$	-0.261	(0.0054)
θ_2^0	-2.667	(0.0583)	$\overline{FC}_0$	0.195	(0.0055)	$\overline{FC}_3$ $(\lambda\ Group = 4)$	-0.256	(0.0068)
θ_2^1 $(LikeStudy = 2)$	0.0001	(0.0001)	$\overline{FC}_1$ $(\lambda\ Group = 2)$	-0.189	(0.0058)	σ_{fc}	0.316	(0.0034)
Production function of tests								
β_λ^{smp}	-391.1	(4.88)	β_e^{svh}	4.0	(0.60)	β_{const}^{pv}	-474.0	(64.48)
β_λ^{svp}	-349.4	(4.79)	β_λ^{svh}	-488.7	(5.93)	β_e^{gh}	25.5	(0.99)
β_λ^{snp}	-8.2	(0.00)	β_{const}^{svh}	-83.9	(56.74)	β_λ^{gh}	-925.1	(9.68)
β_λ^{ssp}	-7.5	(0.00)	β_e^{pm}	11.3	(0.73)	β_{const}^{gh}	-1789.2	(92.83)
β_λ^{gp}	-5.4	(0.06)	β_λ^{pm}	-1009.1	(10.35)	σ_{pm}	56.3	(0.13)
β_e^{smh}	4.8	(0.73)	β_{const}^{pm}	-563.5	(68.63)	σ_{pv}	53.6	(0.14)
β_λ^{smh}	-589.7	(7.36)	β_e^{pv}	10.6	(0.69)	σ_{gh}	71.0	(0.16)
β_{const}^{smh}	-178.0	(68.74)	β_λ^{pv}	-1100.7	(11.13)			
Measures of student effort at high school								
α_e(effort_p)	0.11	(0.009)	Cut_{sb}^2	3.18	(0.907)	Cut_{sp}^3	4.75	(1.035)
α_{const}(effort_p)	-10.48	(0.883)	Cut_{sb}^3	4.36	(0.907)	β_{ca}^{eh} (effort)	0.04	(0.009)
σ_{atten}	3.82	(0.007)	β_{sp}^{eh} (effort)	0.05	(0.011)	β_{ca}^{eh} (ses_s2)	0.00	(0.010)
β_{sb}^{eh} (effort)	0.04	(0.010)	β_{sp}^{eh} (ses_s2)	0.04	(0.013)	β_{ca}^{eh} (ses_s3)	0.06	(0.012)
β_{sb}^{eh} (ses_s2)	-0.05	(0.011)	β_{sp}^{eh} (ses_s3)	0.11	(0.014)	β_{ca}^{eh} (ses_s4)	0.22	(0.014)
β_{sb}^{eh} (ses_s3)	-0.08	(0.012)	β_{sp}^{eh} (ses_s4)	0.22	(0.017)	β_{ca}^{eh} (ses_s5)	0.32	(0.018)
β_{sb}^{eh} (ses_s4)	-0.01	(0.015)	β_{sp}^{eh} (ses_s5)	0.40	(0.020)	β_{ca}^{eh} (edu_fac)	0.03	(0.011)
β_{sb}^{eh} (ses_s5)	0.398	(0.02)	β_{sp}^{eh} (edu_fac)	0.118	(0.01)	Cut_{ca}^1	1.939	(0.87)
β_{sb}^{eh} (edu_fac)	0.08	(0.012)	Cut_{sp}^1	2.67	(1.034)	Cut_{ca}^2	3.58	(0.872)
Cut_{sb}^1	1.83	(0.907)	Cut_{sp}^2	3.64	(1.034)	Cut_{ca}^3	4.56	(0.872)
Measures and distribution of the learning skill								
α_{ms1}^λ	4.59	(0.072)	α_{ms3}^λ	3.61	(0.078)	λ (Group 1)	0.0002	(0.00002)
Cut_{ms1}^1	0.04	(0.006)	Cut_{ms3}^1	0.96	(0.008)	λ (Group 2)	0.07	(0.001)
Cut_{ms1}^2	1.13	(0.007)	Cut_{ms3}^2	1.93	(0.009)	λ (Group 3)	0.13	(0.001)
Cut_{ms1}^3	2.49	(0.010)	Cut_{ms3}^3	0.96	(0.008)	λ (Group 4)	0.20	(0.002)
α_{ms2}^λ	3.49	(0.072)	α_{ms4}^λ	3.19	(0.067)	π_1	0.25	(0.016)
Cut_{ms2}^1	0.43	(0.007)	Cut_{ms4}^1	0.53	(0.007)	π_2	0.33	(0.016)
Cut_{ms2}^2	1.45	(0.008)	Cut_{ms4}^2	1.44	(0.007)	π_3	0.31	(0.016)
Cut_{ms2}^3	2.68	(0.013)	Cut_{ms4}^3	2.14	(0.008)	π_4	0.10	

Note: the estimated parameters for the unobserved types probabilities are: p_1, p_2 and p_3, where $\pi_1 = \frac{p_1}{p_1+p_2+p_3+1}$. In these cases the reported standard error refer to p_i. Notice that $\theta_2^{t(i)} = exp\left(\theta_2^0 + \theta_2^1 * 1(LikeStudy_{t(i)} = 2) + \theta_2^2 * 1(LikeStudy_{t(i)} = 3)\right)$.

D.3 Model fit

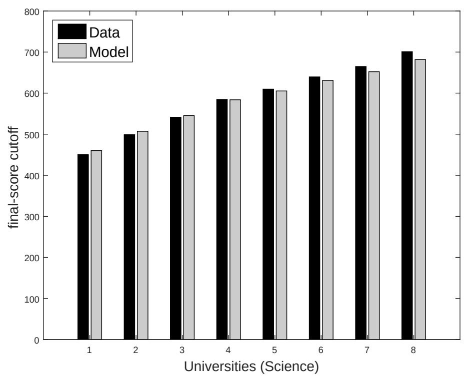
Table 10: Simulated Final-score Cutoffs using Different Initial Guesses

	Initial Guess (r_0)							
	$0.5 * r_{data}$		$0.6 * r_{data}$		$0.7 * r_{data}$		$0.8 * r_{data}$	
University	r_0	r_1	r_0	r_1	r_0	r_1	r_0	r_1
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	225.00	460.11	270.00	460.11	315.00	460.11	360.00	460.11
3	249.40	507.04	299.28	507.03	349.16	507.04	399.04	507.04
4	270.60	545.53	324.72	545.53	378.84	545.53	432.96	545.53
5	292.40	583.82	350.88	583.82	409.36	583.82	467.84	583.82
6	304.90	605.34	365.88	605.34	426.86	605.34	487.84	605.34
7	319.80	631.03	383.76	631.03	447.72	631.03	511.68	631.03
8	332.50	652.21	399.00	652.22	465.50	652.21	532.00	652.21
9	350.50	681.74	420.60	681.73	490.70	681.74	560.80	681.74
	$0.9 * r_{data}$		r_{data}		$1.1 * r_{data}$		$1.2 * r_{data}$	
University	r_0	r_1	r_0	r_1	r_0	r_1	r_0	r_1
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	405.00	460.11	450.00	460.12	495.00	460.12	540.00	460.12
3	448.92	507.04	498.80	507.05	548.68	507.04	598.56	507.04
4	487.08	545.53	541.20	545.55	595.32	545.53	649.44	545.53
5	526.32	583.82	584.80	583.82	643.28	583.82	701.76	583.82
6	548.82	605.34	609.80	605.34	670.78	605.34	731.76	605.34
7	575.64	631.03	639.60	631.02	703.56	631.02	767.52	631.02
8	598.50	652.21	665.00	652.21	731.50	652.21	798.00	652.21
9	630.90	681.74	701.00	681.74	771.10	681.74	841.20	681.74

Note: r_0 refers to the initial guess and r_1 is the final-score cutoff to which the algorithm converges. Moreover, r_{data} is the guess – coming from data – that it is considered when the model is simulated. The table shows the final-scores of equilibrium starting from seven new initial guesses and starting from r_{data}.

Figure 7: Final-score Cutoffs for 2009 University Admissions Process

(a) Scientific major



(b) Humanities major

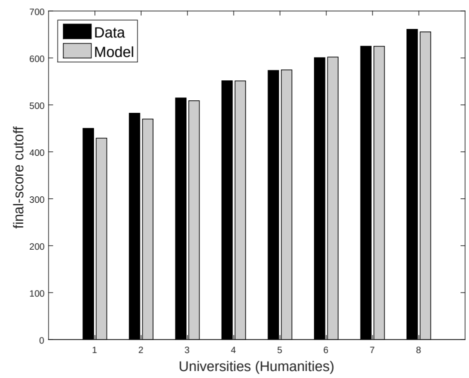
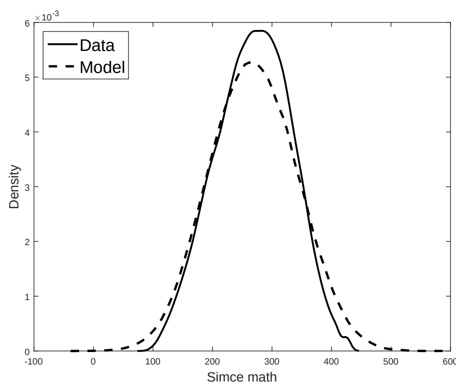


Figure 8: Tests 2006

(a) Simce math 2006



(b) Simce verbal 2006

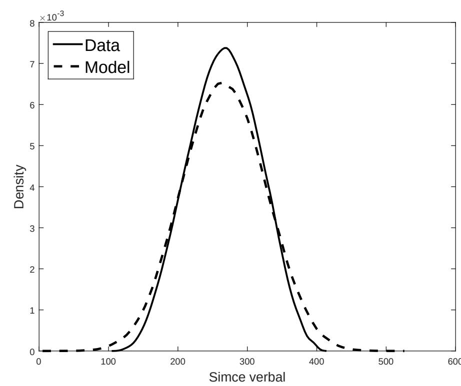
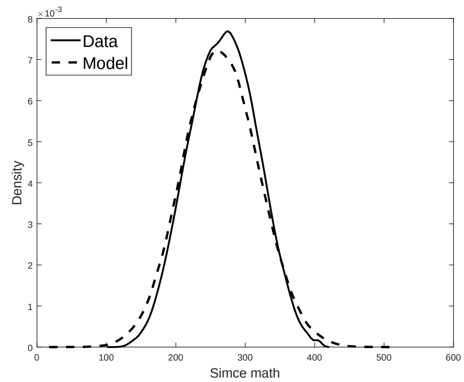
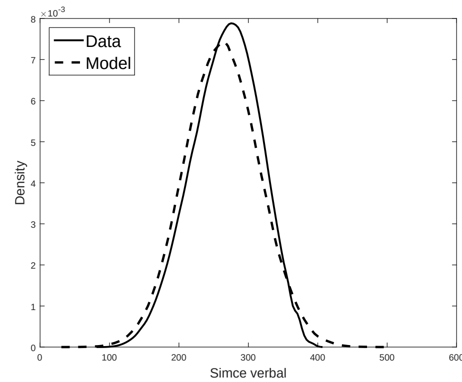


Figure 9: Tests 2004

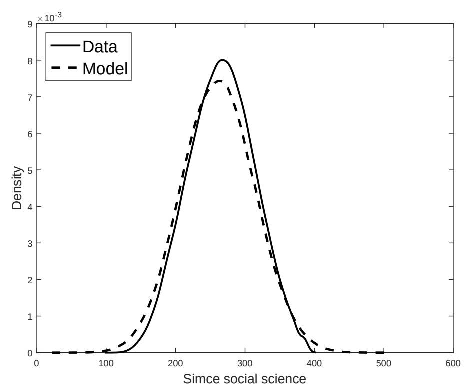
(a) Simce math 2004



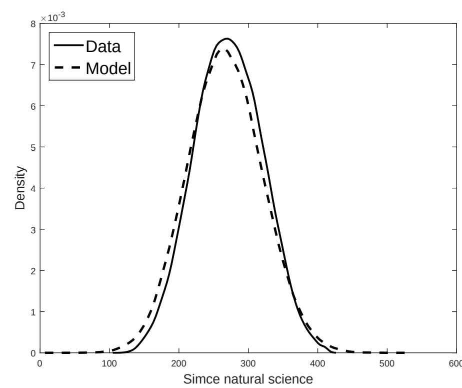
(b) Simce verbal 2004



(c) Simce natural science 2004



(d) Simce social science 2004



(e) GPA 2004

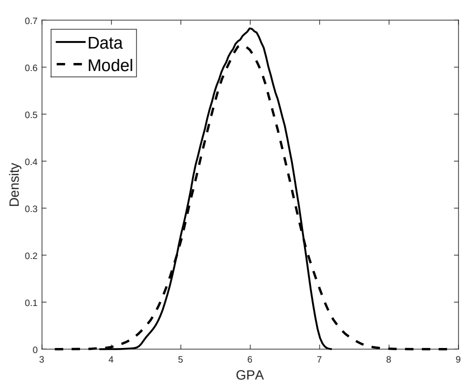


Table 11: Model Fit about the Prediction on Taking the Admission Test

	Fraction of correctly predicting choice	Fraction of students simulated as taking the test
All	0.70	0.78
Female	0.71	0.77
Male	0.70	0.78
SES 1	0.54	0.66
SES 2	0.60	0.71
SES 3	0.78	0.83
SES 4	0.91	0.92
SES 5	0.96	0.97

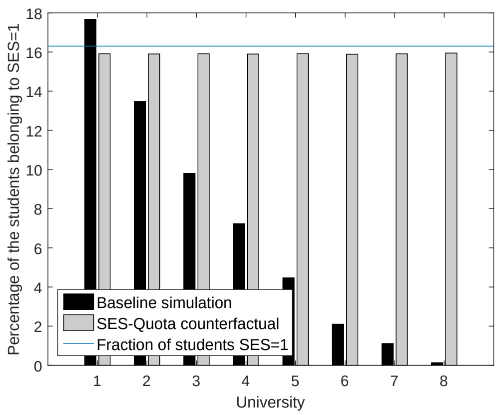
Note: the fraction of correctly predicting choice considers both the students who are simulated as taking the test and in the data they also do so, and the students who are simulated as not taking the test and in the data they neither do so.

E Counterfactual experiments

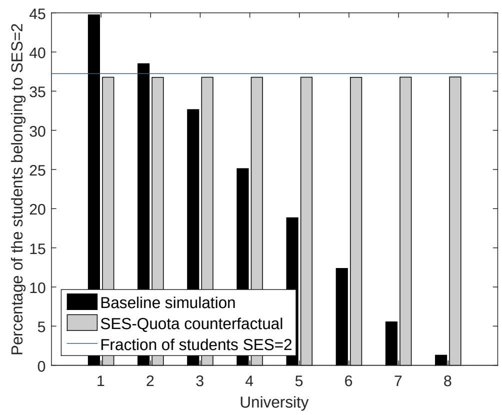
E.1 Impact of the SES-quota System

Figure 10: Impact of SES-quota System on Universities' Socioeconomic Composition

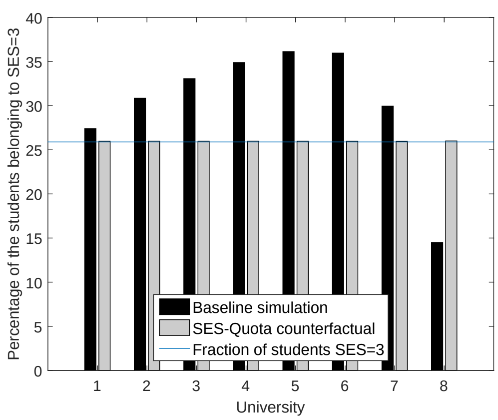
(a) $SES = 1$



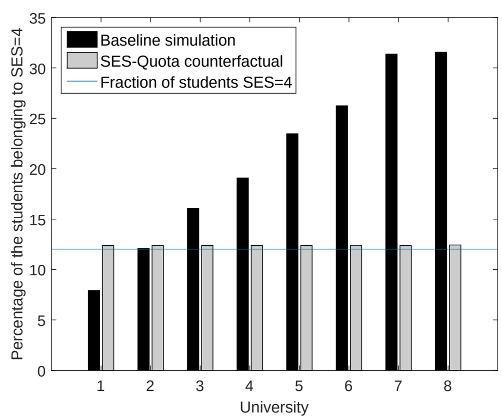
(b) $SES = 2$



(c) $SES = 3$



(d) $SES = 4$



(e) $SES = 5$

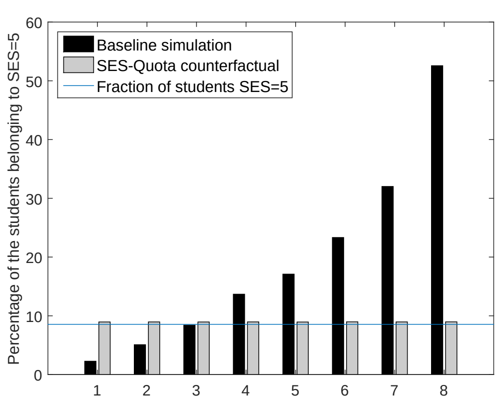
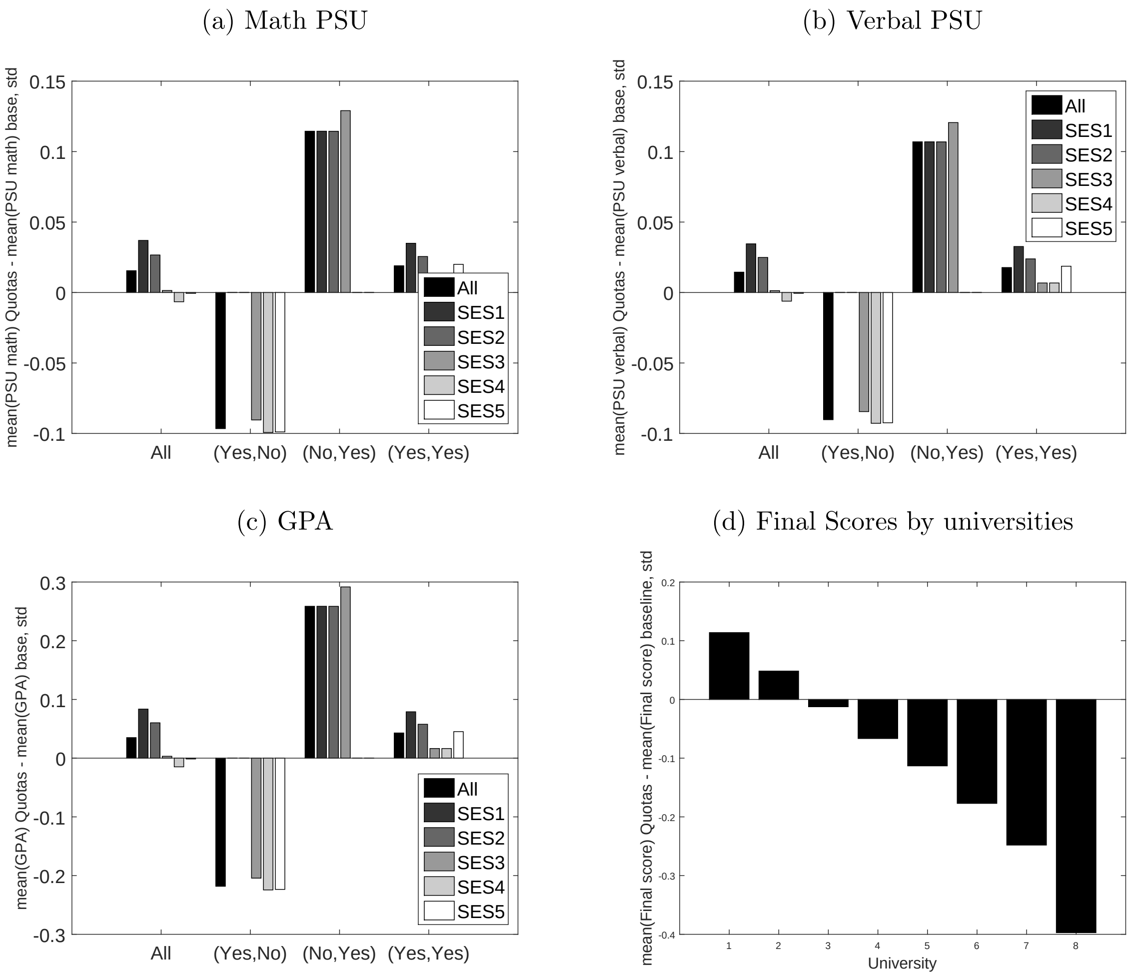


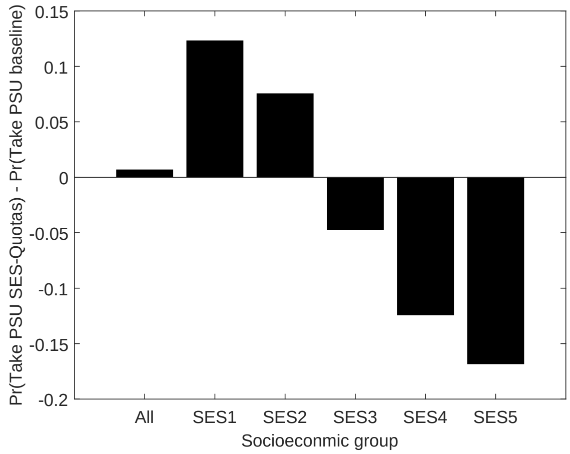
Figure 11: Impact of SES-quota System on Tests by SES and Universities



Note: ({Yes,No},{Yes,No}) stands for (Whether the students were taking the PSU in the baseline scenario,Whether the students are taking the PSU in the counterfactual scenario).

Figure 12: Impact of SES-quota System on who is Taking the PSU

(a) Change in the fraction of students taking the PSU by SES



(b) Impact on the PSU-takers' learning skill distribution

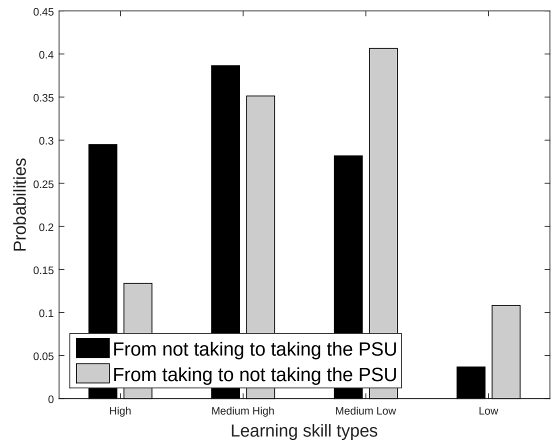
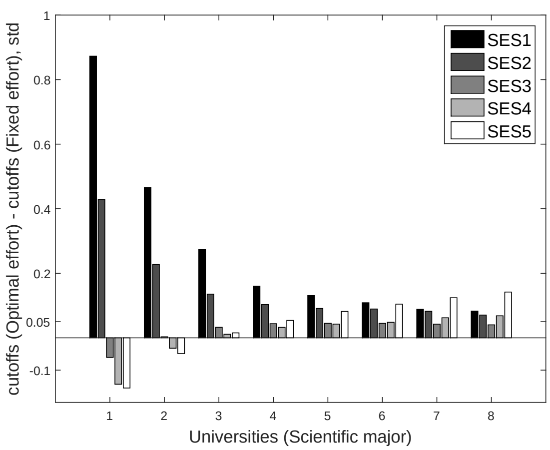
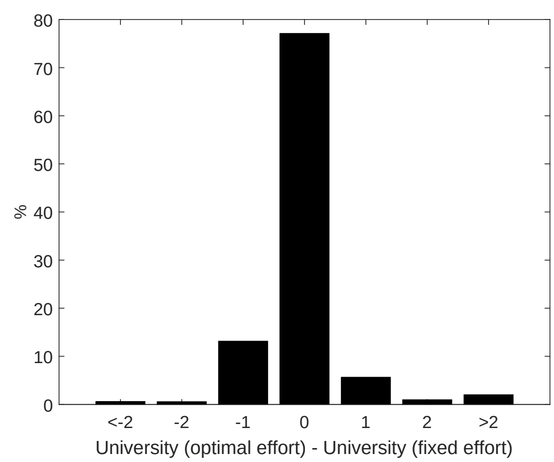


Figure 13: The Impact on Final-score Cutoffs and College Admissions of the Introduction of the SES-quota System, with and without Endogenous Effort (Scientific Major)

(a) Final-score cutoff



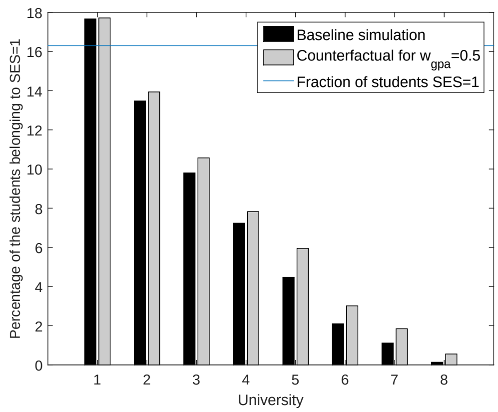
(b) University admission



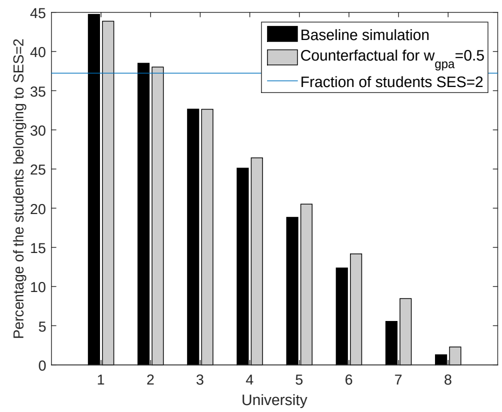
E.2 Impact of Increasing GPA weight on Final Score

Figure 14: Impact of Changing GPA Weight from 0.4 to 0.5 on Universities' Socioeconomic Composition

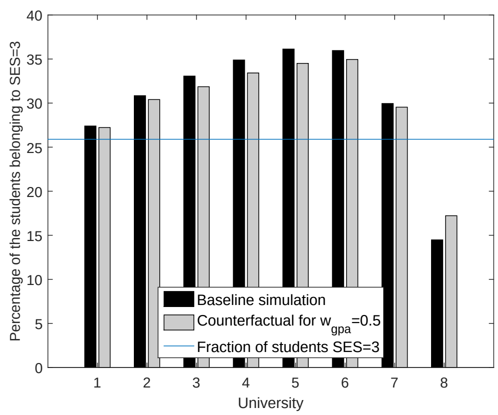
(a) $SES = 1$



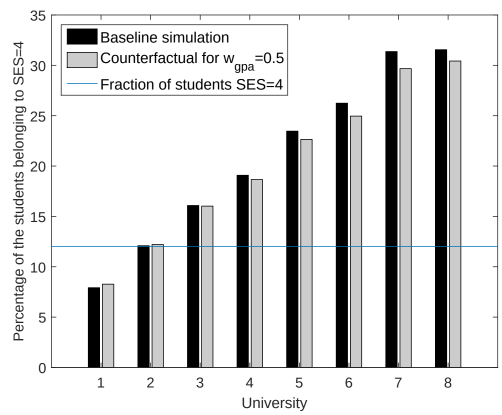
(b) $SES = 2$



(c) $SES = 3$



(d) $SES = 4$



(e) $SES = 5$

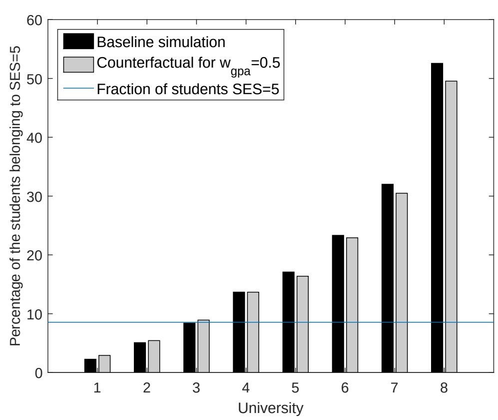
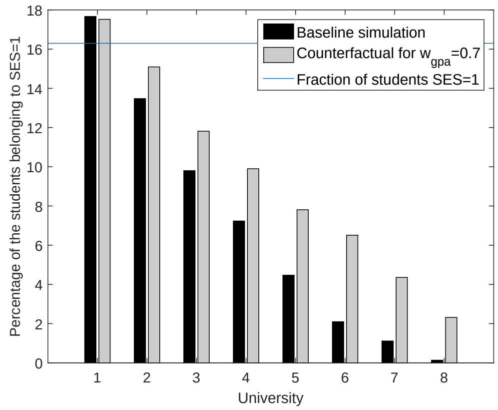
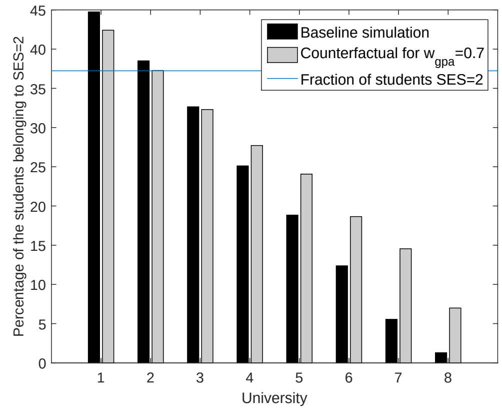


Figure 15: Impact of Changing GPA Weight from 0.4 to 0.7 on Universities' Socioeconomic Composition

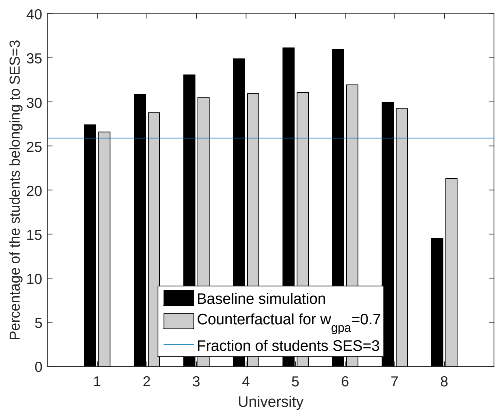
(a) $SES = 1$



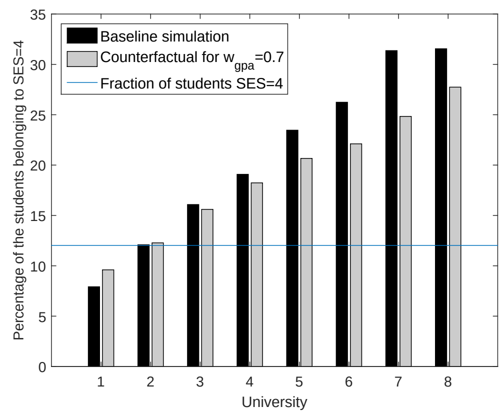
(b) $SES = 2$



(c) $SES = 3$



(d) $SES = 4$



(e) $SES = 5$

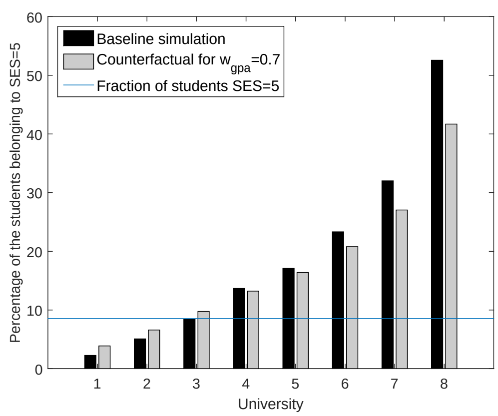
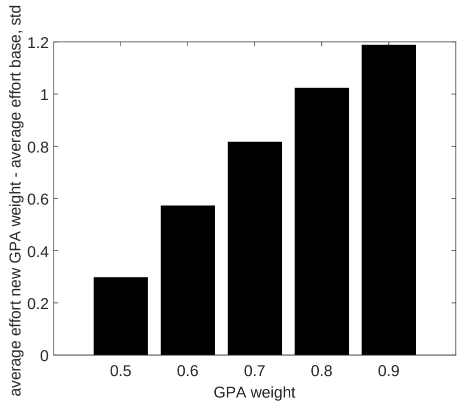
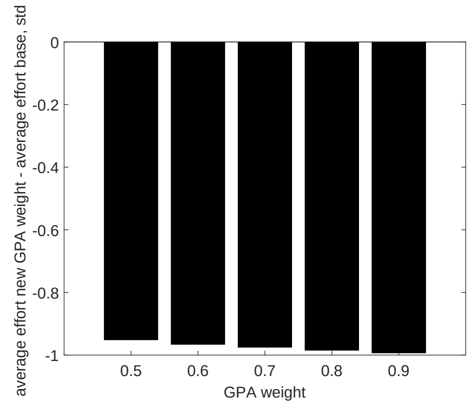


Figure 16: The Impact on Effort of Changing GPA Weight

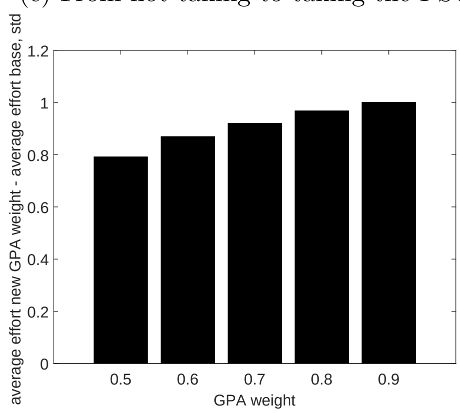
(a) All the students



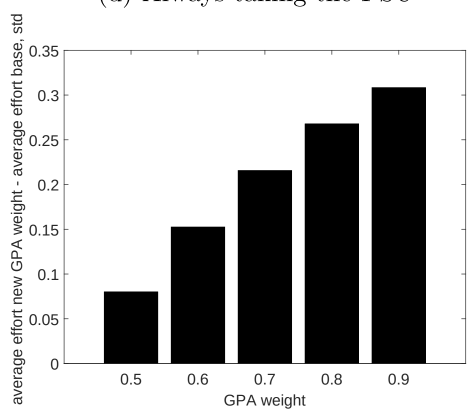
(b) From taking to not taking the PSU



(c) From not taking to taking the PSU



(d) Always taking the PSU



References

Arcidiacono, P. (2005): "Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?" *Econometrica*, 73, 1477–1524.

Arcidiacono, P., M. Lovenheim, and M. Zhu (2016): "Affirmative Action in Undergrate Education," *Anual Review of Economics*, 7, 487–518.

Bodoh-Creed, A. and B. R. Hickman (2014): "Using Auction Theory to Study Human Capital Investment in Assortative Matching Markets: a Look at Affirmative Action in College Admissions," Working paper, University of Chicago Department of Economics.

Bowen, W. G. and D. Bok (1998): *The Shape of The River: Long-term Consequences of Considering Race in College and University Admissions*, Princeton, New Jersey: Princeton University Press.

Braga, B. and P. Bordon (2017): "Employer Learning, Statistical Discrimination and University Prestige," Working paper.

Brewer, D. J., E. R. Eide, and R. G. Ehrenberg (1999): "Does It Pay to Attend an Elite Private College? Cross-Cohort Evidence on the Effects of College Type on Earnings." *The Journal of Human Resources*, 34, pp. 104–123.

Card, D. and A. B. Krueger (2005): "Would the Elimination of Affirmative Action Affect Highly Qualified Minority Applicants? Evidence from California and Texas," *Industrial and Labor Relations Review*, 58, 416–434.

Chan, J. and E. Eyster (2003): "Does Banning Affirmative Action Lower College Student Quality?" *American Economic Review*, 93, 858–872.

Chintagunta, P. K. (1993): "Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households," *Marketing Science*, 12, 184–208.

Cotton, C., B. R. Hickman, and J. P. Price (2014): "Affirmative Action and Human Capital Investment: Evidence from a Randomized Field Experiment," Working Paper 20397, National Bureau of Economic Research.

Cullen, J. B., M. C. Long, and R. Reback (2013): "Jockeying for Position: Strategic High School Choice under Texas' Top ten Percent Plan," *Journal of Public Economics*, 97, 32 – 48.

Cunha, F. and J. J. Heckman (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43.

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78, 883–931.

Dale, S. B. and A. B. Krueger (2002): "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables," *The Quarterly Journal of Economics*, 117, pp. 1491–1527.

Dillon, W. R. and S. Gupta (1996): "A Segment-Level Model of Category Volume and Brand Choice," *Marketing Science*, 15, 38–59.

Domina, T. (2007): "Higher Education Policy as Secondary School Reform: Texas Public High Schools after Hopwood," *Educational Evaluation and Policy Analysis*, 29, 200–217.

Dubin, J. A. and D. L. McFadden (1984): "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption," *Econometrica*, 52, 345–362.

Eckstein, Z. and K. I. Wolpin (1999): "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities," *Econometrica*, 67, 1295–1340.

Epple, D., R. Romano, and H. Sieg (2006): "Admission, Tuition, and Financial Aid Policies in the Market for Higher Education," *Econometrica*, 74, 885–928.

Eren, O. and D. J. Henderson (2008): "The Impact of Homework on Student Achievement," *Econometrics Journal*, 11, 326–348.

Ferman, B. and J. Assunçâo (2011): "Does Affirmative Action Enhance or Undercut Investment Incentives? Evidence From Quotas in Brazilian Public Universities," Working paper, Department of Economics, Massachusetts Institute of Technology.

Fryer, R. G., G. C. Loury, and T. Yuret (2008): "An Economic Analysis of Color-Blind Affirmative Action," *Journal of Law, Economics and Organization*, 24, 319–355.

Gall, T., P. Legros, and A. Newman (2013): "A Theory of Re-match and Aggregate Performance," Discussion paper.

Hanemann, W. M. (1984): "Discrete/Continuous Models of Consumer Demand," *Econometrica*, 52, 541–561.

Hastings, J. S., C. A. Neilson, and S. D. Zimmerman (2012): "The Effect of School Choice on Intrinsic Motivation and Academic Outcomes," Working Paper 18324, National Bureau of Economic Research.

Heckman, J. J., J. Stixrud, and S. Urzua (2006): "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, 24, 411–482.

Hickman, B. R. (2011): "Human Capital Investment, Race Gaps and Affirmative Action: A Game-Theoretic Analysis of College Admissions," Working paper, Department of Economics, University of Chicago.

——— (2013): "Pre-College Human Capital Investment and Affirmative Action: A Structural Policy Analysis of US College Admissions," Working paper, Department of Economics, University of Chicago.

Holzer, H. and D. Neumark (2000): "Assessing Affirmative Action," *Journal of Economic Literature*, 38, 483–568.

James, E., N. Alsalam, J. C. Conaty, and D.-L. To (1989): "College Quality and Future Earnings: Where Should You Send Your Child to College?" *The American Economic Review,Papers and Proceedings of the Hundred and First Annual Meeting of the American Economic Association*, 79, pp. 247–252.

Lazear, E. P. and S. Rosen (1981): "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89, 841–64.

Long, M. C. (2004): "College Applications and the Effect of Affirmative Action," *Journal of Econometrics*, 121, 319–342.

Rau, W. and A. Durand (2000): "The Academic Ethic and College Grades: Does Hard Work Help Students to "Make the Grade"?" *Sociology of Education*, 73, pp. 19–38.

Reyes, L., J. Rodríguez, and S. S. Urzúa (2013): "Heterogeneous Economic Returns to Postsecondary Degrees: Evidence from Chile," NBER Working Papers 18817, National Bureau of Economic Research, Inc.

Schuman, H., E. Walsh, C. Olson, and B. Etheridge (1985): "Effort and Reward: The Assumption That College Grades are Affected by Quantity of Study," *Social Forces*, 63, pp. 945–966.

Stinebrickner, R. and T. R. Stinebrickner (2004): "Time-use and College Outcomes," *Journal of Econometrics*, 121, 243–269.

——— (2008): "The Causal Effect of Studying on Academic Performance," *The B.E. Journal of Economic Analysis & Policy*, 8, 14.

Todd, P. E. and K. I. Wolpin (2003): "On The Specification and Estimation of The Production Function for Cognitive Achievement," *Economic Journal*, 113, F3–F33.

——— (2007): "The Production of Cognitive Achievement in Children: Home, School, and Racial Test Score Gaps," *Journal of Human Capital*, 1, 91–136.

Vukina, T. and X. Zheng (2007): "Structural Estimation of Rank-Order Tournament Games with Private Information," *American Journal of Agricultural Economics*, 89, 651–664.