The Impact of College Admissions Policies on The Performance of High School Students*

Nicolás Grau Veloso†
University of Pennsylvania

Abstract

This paper empirically evaluates the effects of college admissions policies on high school student performance. To this end, I build a model where high school students decide their level of effort and whether to take the college admissions test, taking into consideration how those decisions may affect their future university admissions chances. Using Chilean data for the 2009 college admissions process, I structurally estimate the parameters of the model in order to study the implications of two types of counterfactual experiments: (a) a SES-Quota system, which imposes the population's SES distribution for each university; (b) increasing the high school GPA weight. The results from these exercises support the claim that increasing the level of equal college opportunities may boost the amount of effort exerted by high school students. Specifically, I find that: (1) average effort significantly increases as opportunities are equalized across different socioeconomic groups. (2) There is a moderate improvement in high school student performance, which is relatively important for certain groups. (3) The highest reactions in terms of exerted effort come from those students who also change their decision about taking the college admissions test. (4) Neither of these policies increases the percentage of students taking the national test for college admissions, 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 higher level of learning skills. (5) Because the SES-Quota system uses the existing information more efficiently, it implies a more efficient student allocation to equalize opportunities.

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 am very grateful to Kenneth I. Wolpin and Petra Todd for their comments and encouragement. I also thank Diego Amador, Sina Ates, Garth Baughman, Flavio Cunha, Hanming Fang, Tzuo Hann Law, Andrés Hojman, Zehao Hu, Minchul Shin, Holger Sieg, and Juan Pablo Torres-Martinez for helpful discussions, as well as participants in the Seminars at Universidad de Chile, Universidad Católica de Chile, Universidad Diego Portales, and empirical micro Lunch at the University of Pennsylvania for their comments. I thank the Ministry of Education of Chile, DEMRE, and José Miguel Benavente for providing me with data used in this paper. All errors are mine.

†160 McNeil Building, 3718 Locust Walk, Philadelphia, PA 19104-6297, USA. (Email: ngrau@sas.upenn.edu).

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 the probability of college admissions for targeted populations (e.g. of a particular race or family income). 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] Since the existing evaluations generally assume high school student behavior to be exogenous, a missing part of this discussion is how high school students may consider the impact of their effort levels on their university admissions chances and react to different admissions policies accordingly.[2]

To fill this gap, this paper addresses empirically the effect of college admissions on high school student effort and performance 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 effort decision in high school.[3] I address this question using Chilean data for the 2009 college admissions process, whose features and richness particularly suit the question raised in this research. In the absence of changes in college admissions policies, I use the estimated model to perform some *ex-ante* policy evaluation experiments.

I model the college admissions process and high school behavior in a static fashion, where students make two decisions: whether or not to take the national test which is necessary for college admissions, and their academic effort during high school. The exerted effort positively impacts the expected performance in high school and on the national test for college admissions. For those students who decide to take the college admissions test, admissions policies are based on a linear combination of high school grades and the test scores, such that higher values lead to admission at better universities. Hence, the admissions process works as a tournament in

[1]For instance, in a interesting paper, Arcidiacono (2005) structurally estimates the effects of removing admission preferences and financial aid race-based advantages on African American earnings and educational choices. A similar approach where factors such as applications costs, geography, and supply-side competition play a role -relative to the costs of high-school academic achievement- is Epple, Romano, and Sieg (2006). Other related research includes Bowen and Bok (1998), Card and Krueger (2005) and Long (2004). A summary of the literature before 2000 can be found in Holzer and Neumark (2000).

[2]Theoretically and motivated by U.S. legal changes, a series of papers, e.g., Chan and Eyster (2003); Fryer, Loury, and Yuret (2008); and Hickman (2011) have focused on how the prohibition of explicit consideration of race in the admissions process may be quite inefficient if the colleges still have some preferences toward minorities. Below, I discuss the literature that empirically addresses the impact of affirmative action on student behavior.

[3]It is an empirical question whether student effort impacts student performance. In this paper, the parameters which drive the relationship between these two things in the model are estimated. Schuman, Walsh, Olson, and Etheridge (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)).

which students decide their effort and whether or not to take the college admissions test, taking into account the effort cost, the national test's fixed cost, how much they value future pay-offs, and their chances of being admitted to a better university.[4] Because this is a tournament (i.e., the amount of university seats are fixed), any admissions policy implies winners and losers.[5] Yet it is relevant to study who are the winners (or losers) and to find out if there are any policies that raise the total average high school performance.

The database, which has $146,319$ observations, is built using five sources of information: (1) PSU, the national test for college admissions; (2) RECH, the Ministry of Education's data, which includes GPA and attendance information for all high school students; (3) SIMCE 2004 and 2006, a nation-wide test taken by all 14- and 16-year old students. This source provides information about student performance, measures of effort and learning skills, and characteristics of their families and of primary and secondary schools. (4) *Futuro laboral*, Ministry of Education's data from tax declarations which links individual wages to majors and universities. Finally, (5) admissions requirements, data from each university that includes the test's weights for the final score definition and the final cutoff scores (the minimum score for admittance) for each major. While the first three sources are linked through an individual ID, the last two can be merged to link final-score cutoff with future payoffs.

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.

The simulation of the estimated model fits most of the data features reasonably well. In particular, it successfully fits the unconditional and conditional test distributions, and the probability of taking the national test for college admissions across different groups, where both are endogenous variables in the model. Moreover, the simulated final-score cutoff (i.e., the minimum weighted average score for being admitted in each university) replicates data patterns. In the case of exerted effort, both the correlation between the effort measures and the simulated effort and the signs of the factor loadings of the effort measurement equations go in the right direction; both are positive. However, the share of total variance due to estimated effort is quite small for the effort measurement equations. I discuss to what extent this is a drawback and present some evidence that this issue is mainly due to the quality of the measures

[4] There is vast literature, with mixed evidence, to study the impact of college and its quality on future earnings, e.g., Brewer, Eide, and Ehrenberg (1999); Dale and Krueger (2002); Dale and Krueger (2002); James, Alsalam, Conaty, and To (1989); and (with Chilean data) Reyes, Rodríguez, and Urzúa (2013). It is worth noting that while the literature has focused its attention on how to control for the student and college selection, this is not necessarily relevant in my approach because the important feature in my model is not how much students are actually going to earn, rather what they believe is the impact of attending different universities on their future earnings.

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

of the effort as opposed to shortcomings of the model.

Two policies (counterfactual exercises) are simulated in this paper, intended to equalize opportunities. The first one is a SES-Quota system, implying that for each university the SES distribution is the same as the one in the population. In the second policy experiment, I simulate what happens if the GPA weight is increased, which in practice implies that the probability of attending better universities for those students who attend low income high schools is increased. 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 counterfactuals. (1) Average effort significantly increases as opportunities are equalized across different socioeconomic groups. (2) This leads to a moderate improvement in high school students' performances, which is relatively important 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 PSU 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 increases the percentage of students taking the national test for college admissions, 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 higher level of learning skills. (6) Because the SES-Quota system uses the existing information more efficiently, it implies a more efficient student allocation to equalize opportunities.[6]

There are few papers that take students' behavior in high school as endogenous, as I do in this work. Here I summarize three of them.[7] The first two, Domina (2007) and Ferman and Assunçâo (2011) present some reduced form estimations that address how changes in affirmative action policies change students' behavior in high school. In the third paper, which is the closest to my research, Hickman (2010) models the behavior of U.S. high school students as a function of their future chances of being admitted to different universities.

In particular, Domina (2007), using panel data for Texas high schools between 1993 and 2002, shows evidence that Texas' post-Hopwood higher education policies boosts high school students' academic engagement at public schools.[8] 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 pro-

[6]Here, efficiency means to allocate students with respect to their expected GPA and PSU test.

[7]To the best of my knowledge, there are not any others. In a related paper, Hastings, Neilson, and Zimmerman (2012) show how motivation can change the exerted effort of the students, in particular that the opportunity to attend a better high school has positive and significant effects on both student attendance and test scores.

[8]Among other things, Texas' post-Hopwood higher education 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.

ducing a 5.5% decrease in standardized test scores among the favored group, a 25% widening of the achievement gap.

There are two considerations worth pointing out. These studies tell us something about how different ways of increasing the admissions probabilities of the most segregated groups may have different impacts on high school student behavior. However, a structural approach is required in order to have some idea about which admissions policies accomplish an efficient combination of diversity and correct incentives.

To address this issue, Hickman (2010) 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 auctions literature.[9] One of his main findings is that current affirmative action policies narrow the achievement gap and the enrollment gap, 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. Importantly, he does not, nor do I, have data from before and after some policy change, and thus he uses the structure of the model to perform ex ante policy evaluation. Yet, his and my paper are complementary and are the first attempts to structurally estimate the relationship between college admissions system and high school student behavior.

Beyond technicalities, the main differences between my paper and Hickman (2010) are: (1) My theoretical approach does not impose a distinct univeristy type for each admitted student. (2) 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. (3) Finally, 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. Yet, the differences in our approaches are mainly motivated by different access to data and the particular traits in the institutional design of the two educational systems (American and Chilean).

My paper has three main contributions. First, it empirically shows how high school student effort would react to different college admissions policies, establishing that increasing the level of equal opportunities leads to a boost in the average effort. Second, it estimates a rank-tournament with heterogeneous ability contestants.[10] Third, the paper exploits the interaction between economic theory and factor analysis models in the identification and estimation of the model, and in the analysis of the results.

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

[10] Vukina and Zheng (2007) present the first attempt to estimate a structural model of an empirically observed rank-order tournament as a strategic game with private information. 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.

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 how college admissions policies may affect the effort exerted by high school students. Students have two decisions to make: whether or not to take the college admissions test, a necessary input for university admittance; and they must decide how much effort to make during high school. The exerted effort positively impacts expected high school and college admissions test performance. 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.

The college admissions test scores and GPA production technologies are functions of high school and student characteristics. To have a tractable problem, it is assumed that there is a finite space of individual and school characteristics. Thus, let $i \in \{1, 2, ..., M\}$ denote the student-school type; the vectors of observed and unobserved individuals characteristics of student type i are given by $\{x_i, \lambda_i\}$, whereas the mass of those students is denoted by m_i.[11]

There are $N-1$ university types, each one offering the same major.[12] But, because they have different quality levels, each university implies some specific future pay-off $\{R_1, R_2, ..., R_N\}$, such that $R_{n+1} > R_n \ \forall\ n$ and R_1 is the pay-off for those who were not admitted to college (because they did not try or their final score was too low).[13]

Each university n has a fixed and exogenous amount of seats S_n ($S_1 > 0$ is the residual: the mass of students who are not admitted to any college, i.e., $\sum_i m_i = \sum_{\delta=1}^{N} S_\delta$). Hence, the admissions process works as a 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 ($FC_i \sim N(\bar{FC}, \sigma^2_{fc})$), how much they value future pay-offs, and their chances of being admitted by each university.

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

[11] Although from the model's perspective it does not make any difference what is and is not observed by the econometrician, I introduce this notation in the model description to keep the same notation throughout the paper.

[12] This is something that is possible to relax given my data (although it is challenging in terms of the model). In fact, I can have people with different interests and universities teaching different majors, which will create different markets. However, in this model and in my current empirical specification I do not assume such heterogeneity.

[13] To keep a tractable specification, in this model I am not considering individual heterogeneity in future pay-off and in credit constraints. Using Chilean data, Urzua and Rau (2012) show evidence of the impact of short-term credit constraint on dropouts.

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

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:

$$PM_i = \beta_0^{pm} + x_i\beta_1^{pm} + e_i\beta_2^{pm} + \lambda_i\beta_3^{pm} + \varepsilon_i^{pm}, \tag{2}$$

$$PV_i = \beta_0^{pv} + x_i\beta_1^{pv} + e_i\beta_2^{pv} + \lambda_i\beta_3^{pv} + \varepsilon_i^{pv}, \tag{3}$$

$$GPA_i = \beta_0^{g} + x_i\beta_1^{g} + e_i\beta_2^{g} + \lambda_i\beta_3^{g} + \varepsilon_i^{g}. \tag{4}$$

$\varepsilon_i^k \sim N(0, \sigma_k^2)$, $\varepsilon_i^k \perp\!\!\!\perp \varepsilon_i^{k'} \forall\ k \neq k'$ and $E[\varepsilon_i^k | x_i, \lambda_i] = 0$, $\forall\ k \in \{pm, pv, g\}$.

Given the number of people who actually take the college admissions test, the seats offered by each university, and the final score distribution of those students, the vector r ($\{r_1, r_2, ..., r_{N-1}\}$) represents the final minimum score needed to be admitted by each university type. Throughout the paper, I denote this vector as the *final-score cutoff*. Hence, the students who are going to be part of the university n are those who have a final score greater than or equal to r_{n-1} and smaller than r_n. The former inequality is given by the admissions rule, whereas the latter is due to utility maximization.

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

$$U^0(e) = \theta_1 R_1 + \theta_2 GPA(e) - \frac{e^2}{2}, \tag{5}$$

For those who decide to take the college admissions test, the utility is:[14]

$$U^1(e) = \theta_1 \sum_{n=1}^{N} R_n 1(r_{n-1} \leq FS(e) < r_n) + \theta_2 GPA(e) - FC - \frac{e^2}{2}, \tag{6}$$

where $1(A)$ is an indicator function which takes the value of 1 when A is true and 0 otherwise, and θ_1 and θ_2 represent the importance of future pay-offs and the importance of high school student performance, respectively. The cost of effort is quadratic and its parameter is normalized to one.[15]

There are two considerations to be made about students' utility function. On one hand, students make their effort decision before the realization of the shocks (the distributions are common knowledge). For that reason, they maximize expected utility. The only private information used in the student decisions is the value of FC, though the distribution is common

[14] $r_0 = -\infty$.

[15] In the empirical implementation of the model, I allow for some heterogeneity, which does not qualitatively change any outcomes of the model.

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 facts 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 final score is a random variable.

2.1 Partial Equilibrium

Given a vector r, exogenous in a partial equilibrium context, the optimization problem for those who do and do not take the national college admissions test can be written as:[16]

$$\max_{e\geq 0} U_i^0(e) = \max_{e\geq 0}\left\{\theta_1 R_1 + \theta_2(b_{0i} + b_{1i}e) - \frac{e^2}{2}\right\},$$

$$\max_{e\geq 0} U_i^1(e) = \max_{e\geq 0}\left\{\theta_1 \sum_{n=1}^{N-1}(R_n - R_{n+1})\,\Phi\left(\frac{r_n - a_{1i}e - a_{0i}}{\sigma_\eta}\right) + \theta_1 R_N + \theta_2(b_{0i} + b_{1i}e) - FC - \frac{e^2}{2}\right\}. \tag{7}$$

Where:

$$\begin{aligned}
a_{0i} &= P_{pm} * (\beta_0^{pm} + x_i\beta_1^{pm} + \lambda_i^{pm}\beta_3^{pm}) + P_{pv} * (\beta_0^{pv} + x_i\beta_1^{pv} + \lambda_i^{pv}\beta_3^{pv}) + P_g * (\beta_0^g + x_i\beta_1^g + \lambda_i^g\beta_3^g),\\
a_{1i} &= P_{pm} * \beta_2^{pm} + P_{pv} * \beta_2^{pv} + P_g * \beta_2^g,\\
b_{0i} &= \beta_0^g + x_i\beta_1^g + \lambda_i^g\beta_3^g,\\
b_{1i} &= \beta_2^g,\\
\eta_i &= P_{pm} * \varepsilon_i^{pm} + P_{pv} * \varepsilon_i^{pv} + P_g * \varepsilon_i^g.
\end{aligned}$$

Therefore, the decision about taking the test is given by:

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

Lemma 1: Given a vector r, the student's problem (7) has at least one solution.

Proof: When the student does not take the college admissions test ($TCAT = 0$), it is clear that there exists a unique optimal solution, equal to $\theta_2 b_{1i}$. On the other hand, when the student does take the college admissions test ($TCAT = 1$), for any vector r and regardless the level of effort, the marginal revenue of effort is upper bounded by $\bar{e}_i = \theta_1(R_N - R_1) + \theta_2 b_{1i}$ and lower bounded by $\underline{e}_i = \theta_2 b_{1i}$. Thus, because the effort's marginal cost is e, it should be the case

[16] Φ denotes the standard normal distribution function.

that the optimal effort decision for student i belongs to the interval $[\underline{e}_i, \bar{e}_i]$.[17] Given that the objective function is continuous in e and the relevant set is compact, for all i, there is also an optimal solution when $TCAT = 1$. ■

Therefore the partial equilibrium is characterized by the following first order conditions:

For those who do not take the college admissions test:

$$\hat{e}_i^0 = \theta_2 b_{1i}. \tag{9}$$

For those who take the college admissions test:[18]

$$\hat{e}_i^1 = \theta_1 \sum_{n=1}^{N-1} (R_{n+1} - R_n)\, \phi\left(\frac{r_n - a_{1i}\hat{e}_i^1 - a_{0i}}{\sigma_\eta}\right) \frac{a_{1i}}{\sigma_\eta} + \theta_2 b_{1i}, \tag{10}$$

$\Rightarrow$

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

$$\begin{aligned} D_i = {} & \theta_1 \left(\sum_{n=1}^{N-1} (R_n - R_{n+1})\, \Phi\left(\frac{r_n - a_{1i}\hat{e}_i^1 - a_{0i}}{\sigma_\eta}\right)\right) + \theta_1(R_N - R_1) \\ & + \theta_2 b_{1i}(\hat{e}_i^1 - \hat{e}_i^0) - \frac{(\hat{e}_i^1)^2 - (\hat{e}_i^0)^2}{2}. \end{aligned}$$

As pointed out, since U_i^0 is strictly concave, the first order condition is sufficient and the solution in that case is given by $\theta_2 b_{1i}$. A sufficient condition for strict concavity of U_i^1 is given by $\forall\, i:\ \theta_1(R_N - R_1)a_{1i}^2\phi(1) < \sigma_\eta^2,\ \forall i$.[19] When this condition is fulfilled, the solution to (10) is unique and e_i^1 is continuous in r, which is always the case for e_i^0. This continuity is important for the general equilibrium analysis.

It should be noted that the vector $\{\hat{e}_i^0, \hat{e}_i^1\}$ does not vary across students of the same type. However, the final effort decision ($\hat{e}_i = (1 - TCAT_i) * \hat{e}_i^0 + TCAT_i * \hat{e}_i^1$) varies within each type, due to the fact that $TCAT_i$ depends on the fixed cost realization, which is specific to each student.[20]

[17] In fact, any positive effort implies a non-negative probability of attending to any university, thus the optimal effort can not be equal to $\underline{e}$. This means that, for all students, their optimal effort when $TCAT = 1$ is larger than the optimal effort when $TCAT = 0$, i.e., the solution is interior.

[18] ϕ denotes the standard normal density.

[19] This is shown in Appendix A.1.

[20] Therefore, $\hat{e}_i$ could be confusing due to effort heterogeneity within group i, since for some of the students who are type i, this is equal to $\hat{e}_i^0$ and for the others it is equal to $\hat{e}_i^1$. The same holds for $TCAT_i$.

2.2 General Equilibrium

Let $\tilde{m}_i$ be the mass of students of type i who take the college admissions test, then:[21]

$$\tilde{m}_i = m_i \Phi\left(\frac{D_i - \bar{FC}}{\sigma_{fc}}\right)$$

A general equilibrium in this setting is given by a set of vectors $\hat{e}^0$, $\hat{e}^1$ and $\hat{r}$, such that:

- Given $\hat{r}$, $\forall\ i$:
  - $\hat{e}_i^0 = \theta_2 b_{1i}$,
  - $\hat{e}_i^1 = \theta_1 \sum_{n=1}^{N-1} (R_{n+1} - R_n)\, \phi\left(\frac{\hat{r}_n - a_{1i}\hat{e}_i^1 - a_{0i}}{\sigma_\eta}\right) \frac{a_{1i}}{\sigma_\eta} + \theta_2 b_{1i}$,
  - $\hat{D}_i = (U_i^1(\hat{e}_i^1, \hat{r}) + FC_i) - U_i^0(\hat{e}_i^0)$.
- $\forall\ n = 1, ..., N-1$:

$$\sum_{\delta=n+1}^{N} S_\delta = \sum_i \tilde{m}_i \left[1 - \Phi\left(\frac{\hat{r}_n - \hat{e}_i^1 a_{1i} - a_{0i}}{\sigma_\eta}\right)\right] = \sum_i m_i \Phi\left(\frac{\hat{D}_i - \bar{FC}}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{\hat{r}_n - \hat{e}_i^1 a_{1i} - a_{0i}}{\sigma_\eta}\right)\right].$$

Thus, in this setup the vector r 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 in each university is equal to its number of seats.

Lemma 2: If $\forall\ i:\ \theta_1(R_N - R_1)a_{1i}^2\phi(1) < \sigma_\eta^2$ and $\sum_i m_i \Phi\left(\frac{\theta_1(R_N - R_1) - \bar{FC}}{\sigma_{fc}}\right) > \sum_{\delta=2}^{N} S_\delta$, there exists at least one equilibrium.

Proved in Appendix A.1.

The sufficient conditions for existence have clear interpretations. On one hand, the first condition implies that the effort decision can not be overly important for the final score determination (given by the ratio $\frac{a_{1i}}{\sigma_\eta}$) and that the differences in the future pay-offs can not be overly relevant (given by $\theta_1(R_N - R_1)$). 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 big 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.2.

[21] Again, 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 of each type.

Although there is not a proof for $N > 2$, in Appendix A.2 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.

It is worth mentioning that the potential lack of uniqueness is not an issue in the estimation of the model. In fact, to calculate the likelihood function it is only necessary to solve the partial equilibrium as opposed to the general equilibrium. 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.

3 The Chilean System for College Admissions and Data Description

In the Chilean educational system, students can continue their studies after high school at types of tertiary institutions: the *selective* (the best and most prestigious universities) and the *non-selective* (some universities and technical institutions). In 2009, 29% of 18 to 25 year-olds were attending some type of tertiary institution.[22]

The Chilean university system is highly structured: after knowing their final admissions score (a linear combination of high school GPA and test scores), students apply for a particular college major at a particular university. They can apply for more than one major at any given school. The vast majority of the college courses correspond to the core of the specific major. In other words, other than her college major choice, the student has little agency in choosing the components of her academic training. In this system, each university has an admission quota for each major.

As considered in the model, to be admitted into the selective universities, the student must take a national college admissions test (PSU); math and verbal 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 PSUs and GPA are considered). Thus selection is simply based on the final score ranking. A few selective universities have a less transparent admissions process, but from the data it is possible to see their implicit final score cut-off.

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%) took it.[23] Because the national test can be taken once per year and because those who change majors must retest, 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 who didn't repeat any grades between 2004 and 2008). For the cohort, those

[22] CASEN 2009 (Chilean survey for socioeconomic characterization).

[23] Those are the students who took the college admissions test in December 2008. The academic year is from March to December.

students represent 84.5% (179,725 of 212,656) of the total.[24]

There are five sources of information in this paper; the first three are linked through an individual ID.[25]

- **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. It provides the annual average attendance for each high school 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's information).
- **SIMCE 2004 and 2006**: Nation-wide tests taken by students in the eighth grade of primary school (14 years old) and the second grade of secondary (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 the students performance, some proxy measures of effort and learning skills, and characteristics of their families, primary, and secondary schools.
- **Futuro Laboral:** Ministry of Education's data from tax declarations which link individual wages to major and attended university. This public access database contains some statistics about the distribution of wages for each area of study.[26] In particular, it includes the $10th$, $25th$, $50th$, $75th$ and $90th$ percentiles along with wage means, one and five years after leaving college for each area of study.[27] From this I can infer the average pay-off associated with each university and college major.
- **Admissions requirements:** Data from each university that includes the tests' weights for the final score definition and the final-score cutoffs for each major. It is possible to link this information with the previous wage information.

The final database 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. In Appendix B there is a description of the variables considered is this paper along with some statistics.

[24]In the analysis I need high school students' data, which is not available for students who finished high school before 2008.

[25]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 stops linking with other databases at an individual level.

[26]The definition of area of study is quite fine. In fact, there are 105 areas, which in many cases imply that an area contains only one major.

[27]In www.mifuturo.cl/images/metodologias/nota_metodologica_buscador_empleabilidad_e_ingresos.pdf there is a detailed description of this data (the document is in Spanish).

Something worth highlighting is the fact that all independent variables that determine the effort decision are discrete.[28] This feature of the data implies that I have *student types*, namely, groups of students who share the same characteristics. The existence of types has two positive and important consequences. First, it helps to speed up the estimation, since the effort decisions, more precisely $\hat{e}_i^0$ and $\hat{e}_i^1$, are the same for all students belonging to the same type. Second, it better suits the theory, because the higher the cardinality of each student type (described in Table 1), the assumption of a continuum of agents within each type is closer to the data specification.

Table 1: Cardinality of the student types groups

	Mean	Std. Dev.	Min	Max	N
Size of student types	56.36	135.40	1	1447	2596

To be able to estimate the model, a few decisions should be made to adjust the data to model simplifications. First, in the model universities differ only in quality (i.e., there is only one major), and each student has the same ranking for these universities. In this regard and in the empirical implementation of the model, I consider twenty university types, where the first one is the residual (for those who either do not take the college admissions test or have a final score below r_1).

Second, to define R_n and r_n I proceed using the following steps:

- In the admissions process, I assume that all universities only consider GPA, math and verbal PSU scores (i.e., they do not consider the other PSUs). Furthermore, I assume that all universities use the same weights (0.3 for both PSUs and 0.4 for GPA).[29] Thus, I have one final-score cutoff for each student who took the college admissions test.
- I use the information of the fifth year's wages for each area of study.[30]
- I classify each major of the *admissions requirements* database into one of the areas of study of the *Futuro laboral* data base. By doing so, I have the final-score cutoff and the future wages percentiles associated with a particular major (one distribution across universities), for each major/university. Thereafter, I linearly extrapolate the wage percentile information that I have to obtain all deciles for each major.
- In order to have a database containing one final-score cutoff and one future wage for each major/university, I assume a positive monotonic relationship between final-score cutoffs and future payoffs. In particular, for each major/university, I first calculate the decile of that university in the distribution of final-score cutoffs for that particular area of

[28] All of them are discrete by nature. But in order to have this feature in my data, I did not include a few variables that were continuous, e.g., family income (which may have significant measurement error).

[29] They are close to the mode in my data base. They can not be exactly the mode in order to have weights that add to one.

[30] The resulting final-score cutoffs are quite similar if I use first year's wages.

study, and then impute the wage for that decile (from the distribution of wages for that particular area of study), such that the wage's percentile x is merged with the final-score cutoff's percentile x. The outcome is a relationship between the final-score cutoffs and future payoffs that is plotted in Figure 1.

- To group the university-degree points into twenty "university types," I first non-parametrically estimate the relationship between future payoffs and the final-score cutoffs, plotted in Figure 1. This creates a monotonic relationship between these two variables. I then define the groups using cluster analysis, where the universities are grouped by similar future wages.[31]

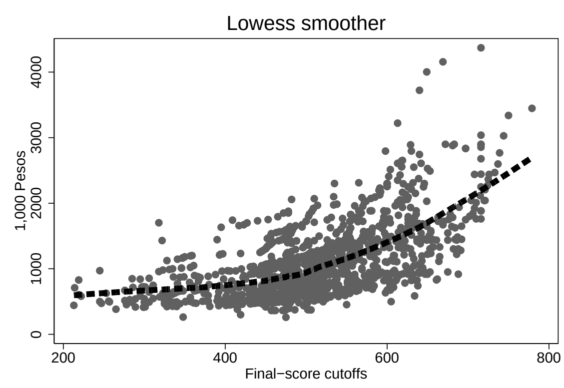
Finally, to define the number of seats for each university type S_n, I calculate how many students, coming directly from high school, had final scores between r_{n-1} and r_n. This means that all my counterfactuals will assume that the share of students who come directly from high school is invariant to policy experiments. In Table 2, the resulting final-score cutoff (r), payoffs (R), and seats available (S) for each university are presented.

Table 2: Universities' payoffs and cutoff scores

University	R	r	S
1	730407	0	Residual
2	813903	437	5114
3	823605	450	2160
4	858348	455	9231
5	887939	476	1869
6	889166	480	1904
7	911408	484	6498
8	954100	498	3738
9	988201	506	1913
10	1007949	510	1881
11	1054916	514	8783
12	1121856	533	6825
13	1175584	548	4107
14	1226456	558	4868
15	1315568	570	9462
16	1428676	596	5180
17	1541462	613	5727
18	1696450	635	6611
19	1966697	669	4356
20	2245443	704	3847

R is in Chilean Pesos. In 2009, one Dollar was 559.67 Pesos.

Figure 1: Imputed fifth year wages and locally weighted regression



[31] I use k-means clustering algorithm.

4 Empirical Specification and Identification

For the empirical implementation, besides the functions that determine the final score, I consider several measures and tests, which are useful to identify the parameters of interest in the context of latent variables. 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. Moreover, I take advantage of the panel data in order to have learning skill measures before the effort decision was made, which is the endogenous variable in my model. The learning skills are assumed to be scalar and time invariant.[32]

I assume λ is independent of x. This assumption is not relevant for the identification argument presented in this paper but it reduces the number of parameters to be estimated. Moreover, as shown below, the results of the estimation seem to support this assumption.

The measures considered are: the final score determinants, i.e., 2009 PSUs (PM, the math test; and PV the verbal test) and high school GPA; the SIMCEs (2004 and 2006); and some direct measures of effort and unobserved learning skills. Hence, 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:

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

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

[32] In the context of the papers. Cunha, Heckman, and Schennach (2010) and Heckman, Stixrud, and Urzua (2006), these learning skills variables would be closer to non-cognitive skills given the measures that I have.

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

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

In this setup, I assume that all the ε_is are normally and independently distributed.[33] Namely, conditional on observables, the correlation across equations is only given by the unobserved skill heterogeneity. In Appendix B, there is a description of the different dependent and independent variables used in the estimation. The following are relevant for the identification analysis:

- $M\lambda_{1i}^{p}$, Me_{1i}^{p} and Me_{1i}^{h} are measures of learning skills, the exerted effort at primary school and the exerted effort at secondary school, respectively. As usual in factor analysis, there are the following normalizations: $\alpha^{e1h} = \alpha^{e1p} = \alpha^{\lambda 1p} = 1$. As will be shown, to ensure identification it is necessary to have at least one measurement being a linear function of each unobservable and one more measurement which does not need to be a linear function of the latent variable.[34] The variables used are: 1) for learning skills, a binary variable that takes the value of 1 if the student had repeated at least one year and 0 otherwise (I use a linear probability model); 2) for the effort exerted at primary school, attendance for the last year of primary school; 3) for the effort exerted at secondary school, the mean of the student attendance over the four years of secondary school.[35]

- 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, as the ones used in this paper.

- In order to gain flexibility, in the estimation, the model specification has an effort cost that is individual specific. This allows different effort decisions among students who are not taking the college admissions test, otherwise $\hat{e}_i^0 = \theta_2 b_2^g$. In this specification, instead of $\frac{e^2}{2}$ the cost of effort is $\exp(\theta_{3i})\frac{e^2}{2}$, where:[36]

$$\begin{aligned}\theta_{3i} = \theta_3^1 1(\mathit{Like\ math} \; = \; 2) + \theta_3^2 1(\mathit{Like\ math} \; = \; 3) + \theta_3^3 1(\mathit{Like\ spanish} \; = \; 2) \\ + \theta_3^4 1(\mathit{Like\ spanish} \; = \; 3).\end{aligned}$$

[33] There is one exception: $\varepsilon_i^{\lambda 1p}$ is not normal because, as specified below, $M\lambda_{1i}^{p}$ is binary and a linear probability model is assumed.

[34] 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}; this just for simplicity.

[35] Using attendance as a measure of effort is a common practice; see for example Hastings, Neilson, and Zimmerman (2012)

[36] In the SIMCE 2004, the students are asked about how much they like to study math and Spanish and the possible answers are: strongly agree, agree, disagree, and strongly disagree. Given that few people choose the last category, I take three values: 1 if the student strongly agrees, 2 if the student agrees, and 3 if the student disagrees or strongly disagrees.

Which implies that θ_{3i} is normalized to zero, and the cost is equal to $\frac{e^2}{2}$, when the student strongly agrees about the statement: *I enjoy the study of math and Spanish.*

In terms of the model characterization, the cost heterogeneity does not imply any relevant changes. In fact, this new specification has the same structure as the previous one, but with new parameters $\tilde{\theta}_{1i} = \theta_1 \exp(-\theta_{3i})$, $\tilde{\theta}_{2i} = \theta_2 \exp(-\theta_{3i})$ and $\tilde{FC}_i \sim N(\bar{FC} \exp(-\theta_{3i}), \sigma^2_{fc} \exp(-2\theta_{3i}))$.

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 $\{\beta^{pm}, \beta^{pv}, \beta^{gh}\}$, $\{Var(\varepsilon_i^{pm}), Var(\varepsilon_i^{pv}), Var(\varepsilon_i^{gh})\}$, $\{\theta, \bar{FC}, \sigma_{fc}, \sigma_\eta\}$ and the distribution of λ. The identification strategy, developed in Appendix C, has three steps. First, I identify the final score's expectation and variance.[37] Second, I non parametrically identify the distribution of learning skills. Third, I identify the utility parameters from different moments of the measures of effort.

5 Estimation

The estimation is carried out in two stages. In the first stage, following the identification analysis presented above and the standard approach to deal 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 second stage, using relevant parameters from the first stage, 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 stage, which only takes a few seconds, leaving just a few parameters to be estimated in the second stage.[38] In terms of numbers, 161 parameters are estimated in the first stage, whereas 84 are estimated in the second stage.

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

- 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 general equilibrium is not required.[39] Second, the first order conditions, which lack a closed form solution, should only be solved for the $2,596$ student types.

[37] If $Var(\varepsilon_i^{pm})$, $Var(\varepsilon_i^{pv})$ and $Var(\varepsilon_i^{gh})$ are identified, then σ_η is also identified.

[38] This is a big gain in time, given that in each iteration the model needs to be solved (which takes around 30 seconds for each set of parameters).

[39] Because I only need to calculate the partial equilibrium of my model, the estimation method used is maximum likelihood as opposed to simulated maximum likelihood.

- Calculate the likelihood function.
- Continue with a new guess until finding the Ω_2 that maximizes the likelihood function.[40]

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 types. This approach has two advantages: first, it is consistent with the model, in which there is a mass of students for each type. Indeed, these discrete unobserved types allows for multiple students for each type (which permits a better approximation to the theoretical equilibrium). Second, it speeds up the estimation, because the student optimization has to be solved just once per student type in each iteration. Meanwhile, some of the parameters that are estimated in the second stage can also be estimated in the first stage (e.g., the factor loadings as shown in the identification argument). I prefer estimating those parameters in the second stage to give to the model a better chance of fitting the data (the model is solved just in the second stage). 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 D, I describe the contributions of different data to the likelihood.

6 Results

The first stage estimation results are presented in Appendix E.1 (Tables 11, 12, and 13). 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 significances are generally all fine. Although in some cases the r squared is fairly small, the instruments are not weak.[41]

Second, in the case of the OLS regressions where one of the secondary education performances is the dependent variable (Table 15), which are the equations whose parameters determine the effort decision, the estimated parameters are as expected in terms of statistical significances, magnitudes, and signs. In particular, the magnitude of the parameters related to effort and learning skills are quite relevant.

Finally, the second stage OLS for the primary education performance presents some problems (Table 14). Indeed, the effect of effort (predicted with instruments) on SIMCEs is in the wrong direction. Nevertheless the effect is in the expected direction for the GPA equation.[42]

[40] This is done using the derivative free solver, HOPSPACK.

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

[42] In both cases, the effect is statistically significant.

Furthermore, the effect of the predicted learning skills is positive and highly relevant in all equations.[43]

The parameters estimated in the second stage are shown in Appendix E.2 (Table 16).[44] As in the first stage, the vast majority of the estimated parameters have the expected sign. The only exceptions are two of the effort cost's parameters θ_3^3 and θ_3^4. Given the non-linear relationship between the parameters and 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 given the estimated parameters. Due to the size of the database, I only draw one vector of shocks per student. Although in the estimation procedure only the partial equilibrium is solved, because the final-score cutoff (r, the general equilibrium object) comes from data, in the simulation I have to calculate the general equilibrium. Thus, the first element to consider in model fit analysis is how close the simulated r_n are in respect to the ones that come from data.[45] In this regard, Figure 2 shows that the simulated vector r captures the trend and magnitudes of the data fairly well.

Though the model shows a good fit in all the aspect 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 3 shows that the model replicates the test distribution observed in the data.[46] Moreover, in Appendix E.3, Table 17 shows that the model is able to replicate student performance across different groups relatively well, although it shows some discrepancies in socioeconomic groups 3 and 4.[47]

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 since one of the two decisions considered in my model is whether to take the national admissions tests. Indeed, Figure 9 (Appendix E.3) shows how the simulation of the model replicates this fraction, particularly the patterns and, with some discrepancies, the magnitude, across gender and high school socioeconomic groups, maternal and paternal education, and high school categories (public, private subsidized, and private non-subsidized).

[43] The parameters are negative because the variables are ordered from more to fewer skills.

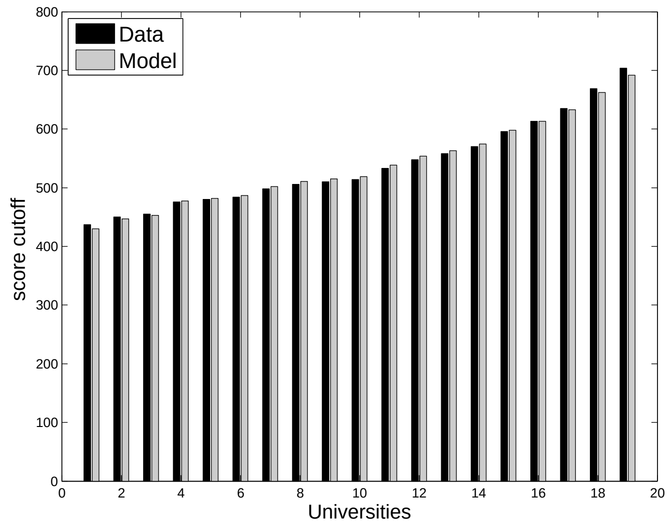
[44] Some parameters are estimated in both stages. In that case, I keep for simulations the ones estimated in the second stage.

[45] The computational algorithm to solve the general 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 final-score cutoff r^0. (3) Given r^0 and the parameters of the model, calculate the optimal effort and optimal decision about taking the PSU, for each student. (4) Given the shocks and effort decisions, calculate the new final-score cutoff (r^1), which solves the general equilibrium condition. (5) Stop if this new r^1 is close enough to r^0 ($\max_{n\in\{1,\dots,N-1\}} |r_n^0 - r_1^1| < \epsilon$), otherwise restart from point (2) with r^1 as the new guess.

[46] The discrepancies in the case of high school GPA are because the data is discrete and there are agglomerations in some grades, something that can not be replicated by the model.

[47] Appendix E.3 contains Figures 7 and 8, which show the model fit of the densities for the remaining tests (2004 and 2006), where all of them show good fit.

Figure 2: Final-score cutoffs for 2009 university admissions process



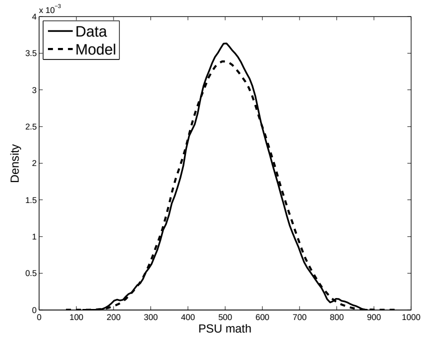
The second student decision modeled is how much effort to exert in high school. In the context of this paper, with many measures of effort, it is not totally clear how to assess the model fit in the effort dimension. However, following the factor models literature, I propose four ways to evaluate such a fit, namely: (1) the correlation between the measures and effort (simulated by the model); (2) the sign and statistical significance of the factor loadings, i.e., parameters that multiply the latent effort decision in each measurement equation; (3) the share of total variance due to estimated effort; and (4) the ratio between the share of total variance due to estimated effort (when effort is modeled) and the ratio of share of total variance due to estimated effort (when effort is not modeled and its distribution, conditional on X, is non-parametrically estimated). Because the latter involves an estimation procedure that needs explanation, I first focus on the former three criteria.

In this respect, Table 3 presents mixed evidence. On one hand, both the correlations and the signs of the factor loadings are in the right direction, positive. On the other hand, in all of the cases, the share of total variance due to estimated effort is quite small, where in the best case it is just above 2%. As it is shown in the third column, the share of total variance due to controls is also small for those measurement equations that include controls.

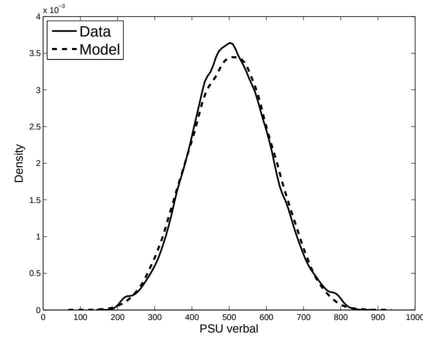
As discussed in a previous section, all the remaining measures of high school effort explain a small fraction of the variance of high school attendance. Thus, part of the reason why the share of total variance due to estimated effort is quite small could be the small correlation among measures of effort. In other words, the problem could be that these measures only share a small part of information (the latent factor). However, this issue can also be explained, since there could be different reasons why students exert effort in high school, and my model captures only one of them. To distinguish between these two possible explanations, measurement error

Figure 3: Model fit in tests determining final score

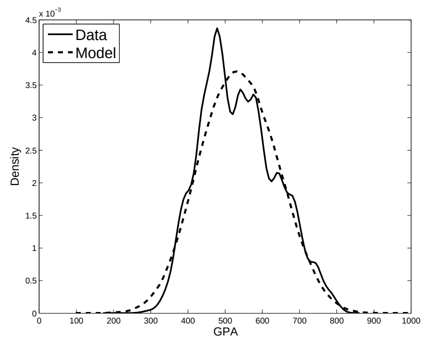
(a) PSU math



(b) PSU verbal



(c) GPA at high school



versus modeling drawbacks, I build criteria four (described below).

From the identification analysis, it is clear that it is possible, in the sense that all the parameters are identified, to estimate the parameters of this set of equations (tests and measures) without using the theory developed in this paper to calculate student effort (conditional on parameters). In particular, as is usual in factor analysis (e.g., Heckman, Stixrud, and Urzua (2006)), I assume that the student effort is drawn from a mix of three normals, which allows for enough flexibility in estimating the density of the factor. Therefore, by doing this estimation I find the density of effort which is consistent with all measures of effort and other tests, since the only thing that I change in respect to the previous estimation is how to calculate high school effort. In this case, simulated maximum likelihood is required.[48] This nonparametric estimation should capture all the information that is not observed and is consistent with the tests and measures of effort. In this context, I conceptualize this information as the density of effort, where such a latent effort decision is not necessary due to considerations of how effort is going to change future chances of being accepted to a better university. In other words, this density establishes a benchmark for my model. The variance that is not explained by this distribution is not captured by any

[48] I also use a reduced form approach for the equation that determines the test taking decision.

Table 3: Correlations and Variance decomposition for effort measures

	Corr(Measure,effort) (Factor Loading)	Share of Total Residual Variance due to estimated effort (Theory)	Share of Total Residual Variance due to controls	Ratio of Share of Total Residual Variance due to estimated effort (Theory/non paramteric)
Attendance	0.136162	0.022099		0.11
Parents perception about student effort	0.075444 (0.109920)	0.003973		0.12
Reading school books at home	0.096300 (0.040786)	0.000546	0.019111	0.11
Using a proper space to study at home	0.122651 (0.046577)	0.000717	0.017227	0.12
Using calculator to study at home	0.101369 (0.038827)	0.000493	0.010817	0.11

theoretical model of effort, since it is due to pure measurement error.[49]

The last column of Table 3 shows that around 11-12% of the variance of the nonparametric distribution of effort is captured by my model. Such a result implies that if the model is correct, only 11-12% of the variance of effort could be explained by modeling how student behavior is determined by future chances of being admitted to a better university. Moreover, this means that, though building a model of effort requires strong assumptions and abstractions from reality, the main problem is the noisiness of the measures of effort.

Is this a relevant issue? I do think that, in general, it could be an issue, but that is not the case in this paper. Under the regular assumption that the errors are iid, having highly noisy measures should affect the precision of the estimated parameters, in particular the standard errors of the factor loadings. However, in this paper all the standard errors are small enough to have statistical significance.[50]

6.2 Unobserved Types

As usual in structural estimations, discrete unobserved types improve the fit of the model. Although in this paper I depart from this tradition by using measures for latent unobserved learning skills, it is still the case that these types have a relevant role in fitting the data. In fact, Table 4 shows that the impact of these types on tests are between 0.5 and 1.5 standard deviations (medium low versus low), 1 and 2.5 standard deviations (medium high versus low) and 2 and 4 standard deviations (high versus low).

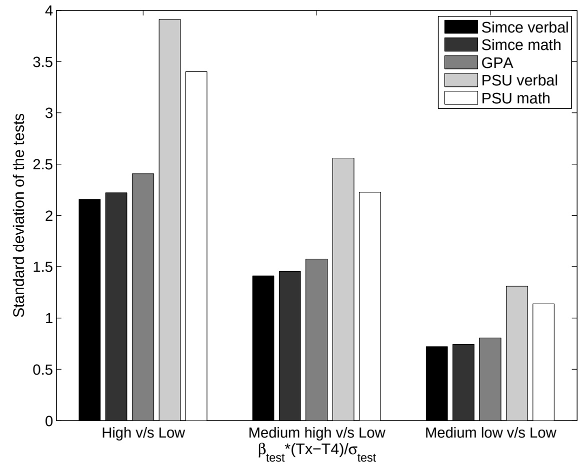
In this approach, it is possible to check the validity of the assumption used in the estimation, that types are independent of X. Indeed, given the estimated π_t (i.e., the unconditional probability of being type t), the conditional probabilities can be recovered by the Bayes rule, such that:[51]

[49] The details of this estimation and the resulting parameters are available upon request.

[50] The only exception is the fixed cost parameter.

[51] These conditional probabilities are used in all the simulations and counterfactual experiments performed in this paper.

Figure 4: The impact of types on tests



$$\pi^i_{t|x} = \frac{\pi_t L_i(\Omega|Type_\lambda = t)}{\sum_\tau \pi_\tau L_i(\Omega|Type_\lambda = \tau)}.$$

Consequently it is possible to see how these probabilities vary across different groups. In fact, Figure 10 (Appendix E.4) shows that the independence assumption does not seem that restrictive: there are not any relevant differences in conditional probabilities across gender, maternal education, paternal education, and high school categories. However, there are some important differences across socioeconomic and urban/rural high school conditions.

7 Counterfactual Experiments

Two policies (counterfactual exercises) are performed in this paper, where both are intended to equalize opportunities. In the first one, a SES-Quota system is established, which imposes that, for each university type, the SES distribution is the same as the population. In other words, if, in the whole system there are $x\%$ of students attending high schools of socioecomic group i, then there should be $x\%$ of students belonging to each high school type in each university type. In practice, the way to get this outcome is by having a tournament within each socioeconomic group (keeping the weights constant for each PSU test and GPA), such that the seats available for students attending high schools socioeconomic group g in university type n is equal to $S_n * \left(\frac{\#students\ SES\ g}{\#students\ in\ the\ system}\right)$, in which case there are five vectors r (one for each socioeconomic group).

In the second counterfactual experiment, I simulate what would happen if the GPA weight was increased, which in practice implies that the probability of attending better universities for students from low income high schools is increased.[52] 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 they capture the differences in high school quality, which is highly correlated with income.

From these exercises, I study the impact on effort, tests, and probability of taking the college admissions test. Moreover, I compare both systems in terms of efficiency. By having the same socioeconomic composition by university, I study which system implies the most efficient student allocation, where efficiency means allocating students with respect to their expected GPA and PSU test.[53]

The first aspect to review from these experiments is how do they change the universities' socioeconomic composition, which is presented in Figures 11, 12 and 13 (Appendix F). On one hand, the first set of plots confirms the outcome of SES-Quota system, namely, that each socioeconomic group is proportionally represented in each university. On the other hand, increasing GPA weights implies more low-income students attending top universities. For example, 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). 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 five universities who belong to SES 1 is doubled, the fraction of the students admitted to the top three universities who belong to SES 2 is also doubled, and the same is true for the top university for SES 3. All these increments are at the expense of higher socioeconomic groups (SES 4 and 5).

From these results, there are two features worth highlighting, which are relevant to keep in mind for the next paragraphs. First, because this is a tournament, where the seats and "prizes" are fixed, there are winners and losers. Second, the effect of the SES-Quota system (the one presented in this paper) is much more aggressive in how the college selection system distributes opportunities than changing GPA weights.[54]

The main goal in this paper is to see how changes in students' opportunities may affect their behavior in high school. In this respect, Figure 14 shows that the SES-quota implementation increases the average effort of high school students by 0.3 standard deviations. Similarly, Figure 15 shows that the changes in GPA weight imply increases in students' average effort from 0.2 to 0.8 standard deviations, depending on the magnitude of the weight's change.

Furthermore, these plots show the importance of the interaction between the two student decisions (i.e., exerted effort and taking the PSU), in the sense that the highest reactions in exerted effort come from those students who also change their decision on taking the college admissions test. For instance, for those students who were not taking the national tests in the baseline simulation, who become takers once the GPA weight is changed, the increase in average exerted effort is from 0.5 to 0.9 standard deviations. The opposite occurs for those

[52] It is also checked for what happens when it is decreased.

[53] For all the simulations and counterfactual experiments, I use the same shocks for each student. In this way, the changes in behavior are only due to changes in colleges admissions rules.

[54] I don't include more plots with different weights but the reader can request the results for a broader set of weights (0.1, 0.2, 0.3, 0.5, 0.6, 0.7, 0.8, 0.9).

who pass from taking to not taking the tests. However, even for those students who do take the college admissions test in both scenarios, there is an important increment in average effort, both in the SES-quota system and when the GPA weight is changed.[55]

Given the linear form of the tests' production function, the effects of these changes in admissions rules on tests is a linear function of the effect on effort. In particular, Figure 16 presents the numbers for the SES-quota experiment. In this case, for those students who attend SES 1 or 2 high schools, the average PSU (math and verbal) increases by around 0.05 standard deviations and by around 0.1 in high school GPA. 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. Finally, even though the magnitudes of these changes are small, there is an important effect on the average final score at each university, which brings attention to the relevance of the change that this experiment produces in the admission system.

As pointed out above, admissions rules also affect the test-taking decision, which is natural since in my model, due to test cost, students take the national test when they have fair chances of being admitted to a good university. Indeed, Figure 17 shows that the implementation of SES-Quota system increases (decreases) the PSU participation by about $5 - 20$ percentage points for socioeconomic groups 1 and 2 (3, 4, and 5). Interestingly, for the entire population these effects cancel each other out, which is consistent with this being a tournament, where the new admissions policy does not change the number of seats per university. In the case of changing the GPA weight (Figure 18), the effect across socioeconomic groups is more moderate, in the range of $1 - 8$ percentage points.

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 performed in this paper allows for such an analysis. In particular, the second plot of Figure 17 shows that, when introducing the SES-Quota system, the new PSU-takers are noticeably more skilled (i.e., higher learning skill type) than those who decide to abandon the admissions process, i.e., not taking the PSU. In the case of changing GPA weights, this result depends on the variation extent, namely, it is the same as the SES-Quota system for new weights equal to 0.5 and 0.6 and goes in the opposite direction for higher weights.

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 by 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 student performance. In this context it is pertinent to ask how relevant this is to model effort.

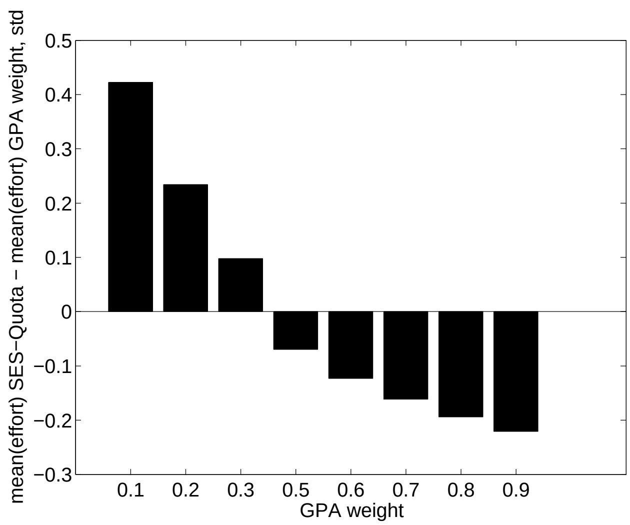
In this regard, I compare how the final-score cutoff and the admission of each university would change, given the described counterfactual experiments, in two scenarios: (1) with optimal effort (i.e., simulating the model) and (2) with fixed effort (i.e., the effort exerted in the baseline scenario). The results plotted in Figure 19 show that there is an important difference between

[55] There are no changes for those who do not take the college admissions test in both scenarios. This is by construction, given that the same shocks are used in all the simulations and counterfactual experiments.

the optimal effort's final-score cutoffs and the fixed effort's final-score cutoffs,[56] 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 goes from 0.2 to more than 1.5 standard deviations. Moreover, only 55% of the students are admitted to the same university in both scenarios.

Figure 20 shows that when these two scenarios are compared given a change in GPA weight from 0.4 to 0.5 (from 0.4 to 0.7), the differences in final-score cutoffs change from 0.01 to 0.025 (from 0.01 to 0.025). However, even in the cases where the effects are moderate, only 70% (50%) of the students are admitted to the same university in both scenarios (Figure 21). Thus, this evidence supports the idea that modeling efforts and the decision to take the PSU is important in order to anticipate what would happen to the main outcomes of the college admissions system.

Figure 5: Average effort: SES-Quota versus changing GPA's weight



Finally, I discuss which college admissions rule leads to the most efficient student allocation. 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.

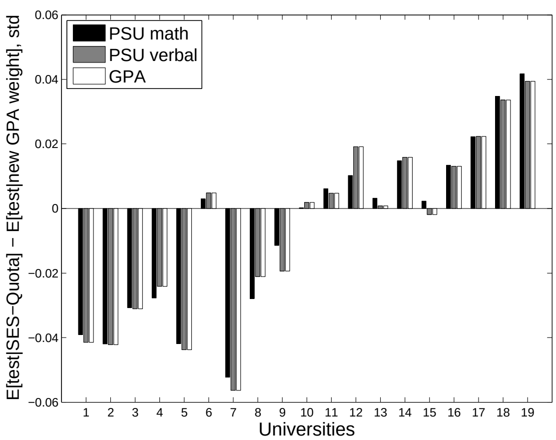
As Figure 5 shows, the first point is that changes in the GPA weight imply a higher increase in average effort than for the SES-Quota system. This is mainly because the estimated effort marginal productivity is much higher in the GPA production function than in the production functions of the two PSU tests.

[56] It should be kept in mind that such a counterfactual experiment implies 5 final-score cutoffs per university.

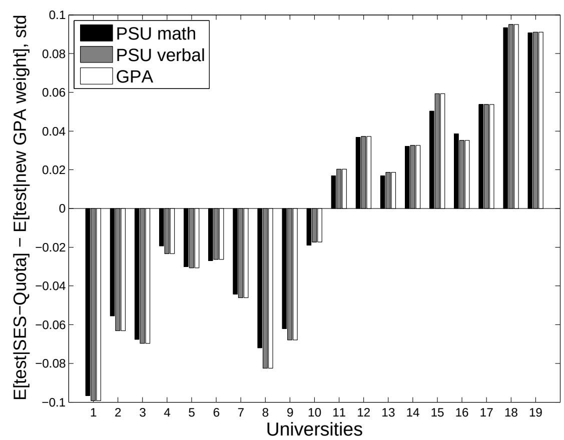
However, this does not mean that changing the GPA weight is the preferred system to achieve equal opportunities. Instead, Figure 6 shows that the higher the GPA weight, the larger the advantage of SES-Quota system, in terms of expected PSU test scores and GPA of the students admitted at top universities. This result is because, 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. Therefore the latter keeps the weights of each shock constant, which attenuates the risk of admitting a bad student due to one extremely positive shock (the three shocks are independent). In sum, the SES-Quota system implies, in expectation, a better student allocation, keeping the level of equal opportunities constant, because it is able to achieve this goal using the existing information more efficiently.

Figure 6: Expected tests and GPA: SES-Quota versus changing GPA's weight

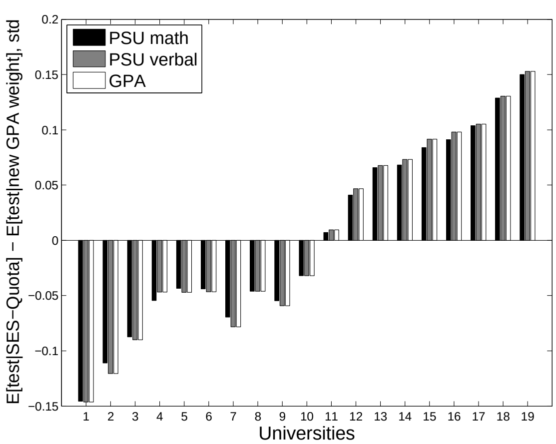
(a) GPA weight = 0.5



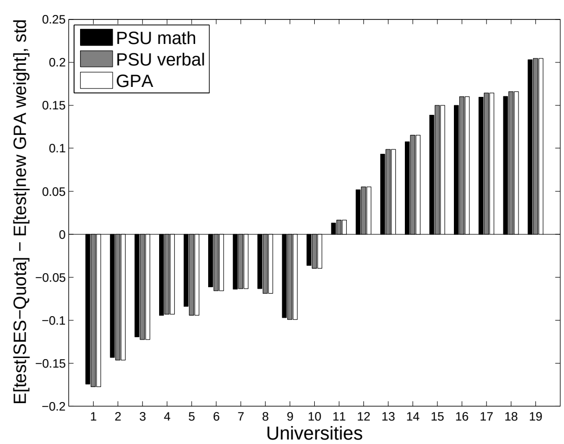
(b) GPA weight = 0.6



(c) GPA weight = 0.7



(d) GPA weight = 0.8



8 Conclusion

To answer the question of this paper, it would be best to have data before and after some admissions policy changes. This ideal data would make it easier to capture the effects of admissions rules on high school student performance. In the absence of such data, structural estimations allow for *ex ante* policy evaluation. Yet, even with such data, the structural approach will be needed in order to study the effect of several policies, as in this paper. The current paper is one of the first steps in studying the structural relationship between high school student effort and their probabilities of being admitted to a good university.

Given the well known difficulty in measuring effort and the level of abstraction that the model needs to be tractable, it is valid to question the reliability of the paper's results. In my opinion, even though the model makes relevant abstractions from reality in order to be tractable and estimable, the current paper can be seen as a reasonable model of the college admissions system of Chile, with reasonable parameters, estimated as rigorously as possible. Yet, this exercise is only capable of giving a rough idea about what could happen if college admissions rules change.

In terms of results, the main lesson from this paper is that it is qualitatively and quantitatively important to consider how a college admissions system may impact high school student behavior. In particular, there are good theoretical and empirical reasons why increasing the level of equal opportunities in college access may boost the effort exerted by high school students. The results of this paper support that claim. Moreover, this paper sheds some light on which admissions system could be optimal in the sense of having an efficient student allocation conditional on delivering the desired change in universities' socioeconomic composition.

There are two interesting avenues for future work. In terms of the model, it would be an interesting, but difficult, extension to consider more than one major per university. I can see in the data where non-mandatory PSUs (e.g., history, biology) were taken by each student (if any). Thus it would be possible to have a better idea of what type of major she was considering when making the test decision. This new multi-major model will imply a specific tournament for each of these majors (with specific vector of final-score cutoff). Given that the effort decision is a non-linear function of the final-score cutoffs, having a better approximation to the real vector of cutoffs may lead to a relevant improvement in the matching between the model and the data.

In terms of method, the paper exploits the interaction between theoretical and factor analysis models. It is left to future research to formalize this analysis with some tests to establish whether the endogenous modeled variable (high school student effort in this case) effectively represents the latent variable.

References

Arcidiacono, P. (2005): "Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?," *Econometrica*, 73(5), 1477–1524.

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

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(1), 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(3), 416–434.

Chan, J., and E. Eyster (2003): "Does Banning Affirmative Action Lower College Student Quality?," *American Economic Review*, 93(3), 858–872.

Cunha, F., and J. J. Heckman (2008): "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43(4).

Cunha, F., J. J. Heckman, and S. M. Schennach (2010): "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3), 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(4), pp. 1491–1527.

Domina, T. (2007): "Higher Education Policy as Secondary School Reform: Texas Public High Schools after Hopwood," *Educational Evaluation and Policy Analysis*, 29(3), 200–217.

Eckstein, Z., and K. I. Wolpin (1999): "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities," *Econometrica*, 67(6), 1295–1340.

Epple, D., R. Romano, and H. Sieg (2006): "Admission, Tuition, and Financial Aid Policies in the Market for Higher Education," *Econometrica*, 74(4), 885–928.

Eren, O., and D. J. Henderson (2008): "The Impact of Homework on Student Achievement," *Econometrics Journal*, 11(2), 326–348.

Ferman, B., and J. Assunção (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(2), 319–355.

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(3), 411–482.

Hickman, B. R. (2010): "Effort, Race Gaps and Affirmative Action: A Structural Policy Analysis of US College Admissions," Working paper, Department of Economics, University of Chicago.

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

Holzer, H., and D. Neumark (2000): "Assessing Affirmative Action," *Journal of Economic Literature*, 38(3), 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(2), pp. 247–252.

Lazear, E. P., and S. Rosen (1981): "Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89(5), 841–64.

Long, M. C. (2004): "College Applications and the Effect of Affirmative Action," *Journal of Econometrics*, 121(1-2), 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(1), 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(4), pp. 945–966.

Stinebrickner, R., and T. R. Stinebrickner (2004): "Time-use and College Outcomes," *Journal of Econometrics*, 121(1-2), 243–269.

Stinebrickner, R., and T. R. Stinebrickner (2008): "The Causal Effect of Studying on Academic Performance," *The B.E. Journal of Economic Analysis & Policy*, 8(1), 14.

Todd, P. E., and K. I. Wolpin (2003): "On The Specification and Estimation of The Production Function for Cognitive Achievement," *Economic Journal*, 113(485), F3–F33.

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

Urzua, S., and T. Rau (2012): "Higher Education Dropouts, Access to Credit, and Labor Market outcomes: Evidence from Chile," Discussion paper.

Vukina, T., and X. Zheng (2007): "Structural Estimation of Rank-Order Tournament Games with Private Information," *American Journal of Agricultural Economics*, 89(3), 651–664.

A Existence and uniqueness

A.1 Existence

Lemma 2: If $\forall\ i:\ \theta_1(R_N - R_1)a_{1i}^2\phi(1) < \sigma_\eta^2$ and $\sum_i m_i \Phi\left(\frac{\theta_1(R_N-R_1)-\bar{FC}}{\sigma_{fc}}\right) > \sum_{\delta=2}^N S_\delta$ there exists at least one equilibrium.

Proof: To prove the lemma, I show that the conditions for the Brouwer fixed point theorem are satisfied. Let $G_n(r) = r_n - \left(\sum_{\delta=n+1}^N S_\delta - \sum_i m_i \Phi\left(\frac{D_i(r)-\bar{FC}}{\sigma_{fc}}\right)\left[1-\Phi\left(\frac{r_n - e_i^1(r)a_{1i}-a_{0i}}{\sigma_\eta}\right)\right]\right)$, where $r \in \mathbb{R}^{N-1}$, then I define the vector-value function $G(r)$ as:[57]

$$G(r) = \begin{bmatrix} G_1(r) \\ G_2(r) \\ . \\ . \\ . \\ G_{N-1}(r) \end{bmatrix} = \begin{bmatrix} r_1 - \left(\sum_{\delta=2}^N S_\delta - \sum_i m_i \Phi\left(\frac{D_i(r)-\bar{FC}}{\sigma_{fc}}\right)\left[1-\Phi\left(\frac{r_1-e_i^1(r)a_{1i}-a_{0i}}{\sigma_\eta}\right)\right]\right) \\ r_2 - \left(\sum_{\delta=3}^N S_\delta - \sum_i m_i \Phi\left(\frac{D_i(r)-\bar{FC}}{\sigma_{fc}}\right)\left[1-\Phi\left(\frac{r_2-e_i^1(r)a_{1i}-a_{0i}}{\sigma_\eta}\right)\right]\right) \\ . \\ . \\ . \\ r_n - \left(\sum_{\delta=n+1}^N S_\delta - \sum_i m_i \Phi\left(\frac{D_i(r)-\bar{FC}}{\sigma_{fc}}\right)\left[1-\Phi\left(\frac{r_n-e_i^1(r)a_{1i}-a_{0i}}{\sigma_\eta}\right)\right]\right) \end{bmatrix}$$

Hence, proving existence for the general equilibrium is equivalent to showing the existence of a fixed point for $G(r)$. In order to fulfil the Brouwer fixed point theorem's conditions, the vector-valued function $G : M \to M$ should be continuous and M non-empty, compact and convex subset of some Euclidean space $\mathbb{R}^{N-1}$.

Given that the effort decision of any student is bounded by $[\min_i\{\underline{e}_i\}, \max_i\{\bar{e}_i\}]$ it is clear that:[58]

$$r \to \infty \Rightarrow \sum_i m_i \Phi\left(\frac{D_i(r) - \bar{FC}}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r_n - e_i^1(r)a_{1i} - a_{0i}}{\sigma_\eta}\right)\right] \to 0,$$

$$r \to -\infty \Rightarrow \sum_i m_i \Phi\left(\frac{D_i(r) - \bar{FC}}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r_n - e_i^1(r)a_{1i} - a_{0i}}{\sigma_\eta}\right)\right] \to \sum_i m_i \Phi\left(\frac{\theta_1(R_N - R_1) - \bar{FC}}{\sigma_{fc}}\right) > \sum_{\delta=2}^N S_\delta.$$

---

[57] $e_i^1(r)$ stands for the optimal effort decision for those who decide to take the college admissions test given the vector of cutoff scores r.

[58] As $r \to -\infty$

$$D_i = \theta_1\left(\sum_{n=1}^{N-1}(R_n - R_{n+1})\,\Phi\left(\frac{r_n - a_{1i}\hat{e}_i^1 - a_{0i}}{\sigma_\eta}\right)\right) + \theta_1(R_N - R_1) + \theta_2 b_{1i}(\hat{e}_i^1 - \hat{e}_i^0) - \frac{(\hat{e}_i^1)^2 - (\hat{e}_i^0)^2}{2} \to \theta_1(R_N - R_1),$$

because as $r \to -\infty$, $|\hat{e}_i^1 - \hat{e}_i^0| \to 0,\ \forall i$.

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

$$\forall\ n:\ r \to \infty \Rightarrow G_n(r + \varepsilon * \vec{1}) - G_n(r) \to \varepsilon > 0$$
$$\forall\ n:\ r \to -\infty \Rightarrow G_n(r - \varepsilon * \vec{1}) - G_n(r) \to -\varepsilon < 0$$

Therefore, there exist two vectors $\underline{r}$ and $\bar{r}$ such that $\forall\ r < \bar{r} \Rightarrow G(r) < G(\bar{r}) < \bar{r}$ and $\forall\ r > \underline{r} \Rightarrow G(r) > G(\underline{r}) > \underline{r}$.[59] Hence, I can define the set $M = \{r \in \mathbb{R}^{N-1},\ \underline{r} \leq r \leq \bar{r}\}$. This set is not empty, compact and convex.[60]

To show that $G(r)$ is continuous it is sufficient to prove that $\forall i\ e_i(r)$ is continuous.[61] Moreover, applying the Berge's maximum theorem and considering the fact that the effort decision of any student is bounded by $[\min_i\{\underline{e}_i\}, \max_i\{\bar{e}_i\}]$ (compact set), a sufficient condition for the continuity of $e_i^1(r)$ 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:

$$\frac{\partial^2 U_i^1(e)}{\partial e^2} = \theta_1 \sum_{n=1}^{N-1} (R_{n+1} - R_n) \left(\frac{r_n - a_{1i}e - a_{0i}}{\sigma_\eta}\right) \phi \left(\frac{r_n - a_{1i}e - a_{0i}}{\sigma_\eta}\right) \left(\frac{a_{1i}}{\sigma_\eta}\right)^2 - 1$$

But because the first term can not be bigger than $\theta_1(R_N - R_1)\left(\frac{a_{1i}}{\sigma_\eta}\right)^2 \phi(1)$, then[62]

$$\theta_1(R_N - R_1)a_{1i}^2\phi(1) < \sigma_\eta^2 \Rightarrow \frac{\partial^2 U_i^1(e)}{\partial e^2} < 0$$

Moreover, $G(r)$ is well defined for any r because, as it was shown above, for any r there exist optimal efforts for those who take the college admissions test ($e_i^1(r)$) and for those who do not take the test ($e_i^0(r)$). ■

A.2 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 $\{r, e\}$ and $\{r', e'\}$, where without loss of generality $r' > r$,[63] from the general equilibrium definition it is directly shown that:

[59] Because there exist $\bar{r}$ such that $\forall r > \bar{r}$: $\sum_i m_i \Phi\left(\frac{D_i(r) - \bar{FC}}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r_n - e_i^1(r)a_{1i} - a_{0i}}{\sigma_\eta}\right)\right] < \sum_{\delta=n+1}^N S_\delta$, and $\underline{r}$ such that $\forall r < \underline{r}$: $\sum_i m_i \Phi\left(\frac{D_i(r) - \bar{FC}}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r_n - e_i^1(r)a_{1i} - a_{0i}}{\sigma_\eta}\right)\right] > \sum_{\delta=n+1}^N S_\delta$.

[60] To be sure about non-emptiness, it is possible to pick $\underline{r} < 0$ and $\bar{r} > 0$.

[61] If $e_i^1(r)$ is continuous then $D_i(r)$ is also continuous.

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

[63] Notice because $N = 2$, r and r' are scalars. S is the amount of seats offered by the only university.

$$S = \sum_i m_i \Phi\left(\frac{D_i - \bar{FC}}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r - e_i a_{1i} - a_{0i}}{\sigma_\eta}\right)\right]$$
$$S = \sum_i m_i \Phi\left(\frac{D_i^{'} - \bar{FC}}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r' - e_i^{'} a_{1i} - a_{0i}}{\sigma_\eta}\right)\right] \tag{21}$$

To get the contradiction I proceed in two steps. First, I show that the statement: $\forall r' > r, i :$ $\Phi\left(\frac{r' - e_i^{'} a_{1i} - a_{0i}}{\sigma_\varepsilon}\right) - \Phi\left(\frac{r - e_i a_{1i} - a_{0i}}{\sigma_\varepsilon}\right) > 0$, 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:

In fact, let $\Pi_0 = \max_e U_i^0(e)$ and $\Pi_1(r) = \max_e U_i^1(e)$, then $D_i = \Pi_1(r) + FC_i - \Pi_0$.[64] Taking the derivative to D_i with respect to r,[65]

$$\frac{\partial D_i}{\partial r} = \frac{\partial \Pi_1(r)}{\partial r} = (R_1 - R_2)\frac{\theta_1}{a_1}\phi\left(\frac{r - ea_1 - a_0}{\sigma_\eta}\right) < 0$$
$$\Rightarrow \left(\frac{D_i - \bar{FC}}{\sigma_{fc}}\right) > \left(\frac{D_i^{'} - \bar{FC}}{\sigma_{fc}}\right)$$

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

$$\sum_i m_i \left(\Phi\left(\frac{D_i - \bar{FC}}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r - e_i a_{1i} - a_{0i}}{\sigma_\eta}\right)\right] - \Phi\left(\frac{D_i^{'} - \bar{FC}}{\sigma_{fc}}\right)\left[1 - \Phi\left(\frac{r' - e_i^{'} a_{1i} - a_{0i}}{\sigma_\eta}\right)\right]\right) = 0$$

$$\Rightarrow \sum_i m_i \left(\left[1 - \Phi\left(\frac{r - e_i a_{1i} - a_{0i}}{\sigma_\eta}\right)\right] - \left[1 - \Phi\left(\frac{r' - e_i^{'} a_{1i} - a_{0i}}{\sigma_\eta}\right)\right]\right) < 0$$

$$\Rightarrow \sum_i m_i \left(\Phi\left(\frac{r' - e_i^{'} a_{1i} - a_{0i}}{\sigma_\eta}\right) - \Phi\left(\frac{r - e_i a_{1i} - a_{0i}}{\sigma_\eta}\right)\right) < 0$$

where this last inequality contradicts that $\forall r' > r, i :$ $\Phi\left(\frac{r' - e_i^{'} a_{1i} - a_{0i}}{\sigma_\varepsilon}\right) - \Phi\left(\frac{r - e_i a_{1i} - a_{0i}}{\sigma_\varepsilon}\right) > 0$ ■

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.

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

[65] Here, I am assuming that effort is continuous in r (if that is the case, the value function is differentiable), but in the step 2 I also show that $\Pi_1(r) > \Pi_1(r')$ when effort is not continuous in r.

Case 1: effort decision is continuous in r:

Taking a derivative of the first order condition (10), when $N = 2$ implies:[66]

$$\theta_1(R_2 - R_1)\frac{\partial\phi(r)}{\partial r}\left[\frac{1-\frac{\partial e}{\partial r}a_1}{\sigma_\eta}\right]\frac{a_1}{\sigma_\eta} = \frac{\partial e}{\partial r} \;\Rightarrow\; \frac{\partial e}{\partial r} = \frac{\theta_1(R_2-R_1)\frac{\partial\phi(r)}{\partial r}\frac{a_1}{\sigma_\eta^2}}{1+\theta_1(R_2-R_1)\frac{\partial\phi(r)}{\partial r}\left(\frac{a_1}{\sigma_\eta}\right)^2}$$

$$\Rightarrow 1 - \frac{\partial e}{\partial r}a_1 = \frac{1}{1+\theta_1(R_2-R_1)\frac{\partial\phi(r)}{\partial r}\left(\frac{a_1}{\sigma_\eta}\right)^2} \tag{22}$$

Therefore,

$$\frac{\partial\Phi\left(\frac{r-ea_1-a_0}{\sigma_\eta}\right)}{\partial r} = \phi\left(\frac{r-ea_1-a_0}{\sigma_\eta}\right)\left(1-\frac{\partial e}{\partial r}a_1\right)\frac{1}{\sigma_\eta} = \frac{\phi\left(\frac{r-ea_1-a_0}{\sigma_\eta}\right)\frac{1}{\sigma_\eta}}{1+\theta_1(R_2-R_1)\frac{\partial\phi(r)}{\partial r}\left(\frac{a_1}{\sigma_\eta}\right)^2}$$

So, to get the desired result, it is enough showing that $1+\theta_1(R_2-R_1)\frac{\partial\phi(r)}{\partial r}\left(\frac{a_1}{\sigma_\eta}\right)^2 > 0$. In fact, this inequality is ensured by the second order condition:[67]

$$-\theta_1(R_2-R_1)\frac{\partial\phi(r)}{\partial r}\left(\frac{a_1}{\sigma_\eta}\right)^2 - 1 < 0$$

Therefore, it follows that $\frac{\partial\Phi\left(\frac{r-ea_1-a_0}{\sigma_\eta}\right)}{\partial r} > 0$. ■

Case 2: effort decision is discontinuous in r:[68]

Without loss of generality, assume there are two different effort decisions which are optimal at r ($e_h > e_l$). Defining $\Pi_x = \theta_1(R_1-R_2)\Phi\left(\frac{r-e_xa_1-a_0}{\sigma_\eta}\right)+\theta_1R_2+\theta_2(b_0+b_1e_x)-\frac{e_x^2}{2},\ x = l, h$ (the value function for each local equilibrium) and applying the envelope theorem imply:

$$\frac{\partial\Pi_l}{\partial r} - \frac{\partial\Pi_h}{\partial r} = \frac{\theta_1(R_2-R_1)}{\sigma_\eta}\left[\phi\left(\frac{r-e_la_1-a_0}{\sigma_\eta}\right) - \phi\left(\frac{r-e_ha_1-a_0}{\sigma_\eta}\right)\right] \tag{23}$$

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

[66] For simplicity, I suppress the individual sub-index and denote $\phi\left(\frac{r-ea_1-a_0}{\sigma_\eta}\right)$ as $\phi(r)$.

[67] I am assuming away $-\theta_1(R_2-R_1)\frac{\partial\phi(r)}{\partial r}\left(\frac{a_1}{\sigma_\eta}\right)^2 - 1 = 0$.

[68] Given that the discontinuity is possible only for those who take the college admissions test, for this proof I assume away the possibility of not taking the college admissions test.

$$e_h - e_l = \frac{a_1\theta_1(R_2 - R_1)}{\sigma_\eta}\left[\phi\left(\frac{r - e_l a_1 - a_0}{\sigma_\eta}\right) - \phi\left(\frac{r - e_h a_1 - a_0}{\sigma_\eta}\right)\right]$$

$$\Rightarrow \frac{\partial \Pi_l}{\partial r} - \frac{\partial \Pi_h}{\partial r} = \frac{e_h - e_l}{\sigma_\eta} > 0 \tag{24}$$

Therefore, by (24) 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 $\forall r' > r$ such that the effort decision is not continuous at r for students type i, then $\Phi\left(\frac{r' - e_i' a_{1i} - a_{0i}}{\sigma_\eta}\right) - \Phi\left(\frac{r - e_i a_{1i} - a_{0i}}{\sigma_\eta}\right) > 0$. ■

In the case where $N > 2$, as in this paper, it can be established that $\sum_{n=1}^{N-1} \frac{\partial \overline{G_m}}{\partial r_n} < 0 \ \forall m$, where $\overline{G_m} = G_m - r_m$. This result implies that if $\overline{G(r)} = 0$ (i.e., r is an equilibrium), then $r' = r(a+1)$ where $a \neq 0$, can not be an equilibrium.[69] Loosely speaking, this means that if there is an equilibrium denoted by r, the farther r' departs from r the harder it is to have r' as another equilibrium.

To prove the statement, I proceed in two steps.[70] First, it is proved that $\sum_{n=1}^{N-1} \frac{\partial \overline{G_m}}{\partial r_n} < -\sum_i m_i \Phi\left(\frac{D_i - \overline{FC}}{\sigma_{fc}}\right) \frac{\phi_i(r_m)}{\sigma_\eta}\left[1 - a_{1i}\sum_{n=1}^{N-1} \frac{\partial e_i^1}{\partial r_n}\right]$.[71] Second, I show that $1 - a_{1i}\sum_{n=1}^{N-1} \frac{\partial e_i^1}{\partial r_n} > 0 \ \forall i$.

To get the first result, notice that:

$$\forall n \neq m: \ \frac{\partial \overline{G_m}}{\partial r_n} = \sum_i m_i \Phi\left(\frac{D_i - \overline{FC}}{\sigma_{fc}}\right)[1 - \phi_i(r_m)]\frac{\partial D_i}{\partial r_n}\frac{1}{\sigma_{fc}} + \sum_i m_i \Phi\left(\frac{D_i - \overline{FC}}{\sigma_{fc}}\right)\phi_i(r_m)\frac{\partial e_i^1}{\partial r_n}\frac{a_{1i}}{\sigma_{fc}}$$
$$< \sum_i m_i \Phi\left(\frac{D_i - \overline{FC}}{\sigma_{fc}}\right)\phi_i(r_m)\frac{\partial e_i^1}{\partial r_n}\frac{a_{1i}}{\sigma_{fc}}$$

$$\frac{\partial \overline{G_m}}{\partial r_m} = \sum_i m_i \Phi\left(\frac{D_i - \overline{FC}}{\sigma_{fc}}\right)[1 - \phi_i(r_m)]\frac{\partial D_i}{\partial r_m}\frac{1}{\sigma_{fc}} - \sum_i m_i \Phi\left(\frac{D_i - \overline{FC}}{\sigma_{fc}}\right)\frac{\phi_i(r_m)}{\sigma_{fc}}\left[1 - \frac{\partial e_i^1}{\partial r_m}a_{1i}\right]$$
$$< -\sum_i m_i \Phi\left(\frac{D_i - \overline{FC}}{\sigma_{fc}}\right)\frac{\phi_i(r_m)}{\sigma_{fc}}\left[1 - \frac{\partial e_i^1}{\partial r_m}a_{1i}\right]$$

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

[69] It would be better to show that this is true even when the increase (or decrease) is not proportional across score cutoffs. Such a result is not established in this paper. Moreover, I am not sure about the veracity of the statement.

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

[71] $\phi_i(r_m) = \phi\left(\frac{r_m - e_i^1 a_{1i} - a_{0i}}{\sigma_\eta}\right)$.

$$\sum_{n=1}^{N-1}\frac{\partial\overline{G_m}}{\partial r_n} < -\sum_i m_i\Phi\left(\frac{D_i-\overline{FC}}{\sigma_{fc}}\right)\frac{\phi_i(r_m)}{\sigma_{fc}} + \sum_{n=1}^{N-1}\sum_i m_i\Phi\left(\frac{D_i-\overline{FC}}{\sigma_{fc}}\right)\phi_i(r_m)\frac{\partial e_i^1}{\partial r_n}\frac{a_{1i}}{\sigma_{fc}}$$
$$= -\sum_i m_i\Phi\left(\frac{D_i-\overline{FC}}{\sigma_{fc}}\right)\frac{\phi_i(r_m)}{\sigma_{fc}}\left[1-a_{1i}\sum_{n=1}^{N-1}\frac{\partial e_i^1}{\partial r_n}\right]$$

To establish the second result, I begin taking the derivative to the first order condition for those who decide taking the college admissions test. When that is done, I get:

$$\frac{\partial e_i^1}{\partial r_m} = \theta_1\sum_{n=1}^{N-1}(R_{n+1}-R_n)\left(\frac{r_n-e_i^1a_{1i}-a_{0i}}{\sigma_\eta}\right)\phi_i(r_n)\left(\frac{a_{1i}}{\sigma_\eta}\right)^2\frac{\partial e_i^1}{\partial r_m}-$$
$$\theta_1(R_{m+1}-R_m)\left(\frac{r_m-e_i^1a_{1i}-a_{0i}}{\sigma_\eta}\right)\phi_i(r_m)\frac{a_{1i}}{\sigma_\eta^2}$$
$$= \frac{-\theta_1(R_{m+1}-R_m)\left(\frac{r_m-e_i^1a_{1i}-a_{0i}}{\sigma_\eta}\right)\phi_i(r_m)\frac{a_{1i}}{\sigma_\eta^2}}{1-\theta_1\sum_{n=1}^{N-1}(R_{n+1}-R_n)\left(\frac{r_n-e_i^1a_{1i}-a_{0i}}{\sigma_\eta}\right)\phi_i(r_n)\left(\frac{a_{1i}}{\sigma_\eta}\right)^2}$$

$$\Rightarrow 1-a_{1i}\sum_{n=1}^{N-1}\frac{\partial e_i^1}{\partial r_n} = \frac{1}{1-\theta_1\sum_{n=1}^{N-1}(R_{n+1}-R_n)\left(\frac{r_n-e_i^1a_{1i}-a_{0i}}{\sigma_\eta}\right)\phi_i(r_n)\left(\frac{a_{1i}}{\sigma_\eta}\right)^2} > 0$$

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

B Variable Descriptions

Table 4: Variable Descriptions

Variable	Description
	Independent Variables
SEX	Takes 1 if the students is male and 0 if is female.
EDU_MO1	Takes 1 if there is no information about mother's education (0 otherwise).
EDU_MO2	Takes 1 if student's mother has some courses at the primary education level or she does not have formal education (0 otherwise).
EDU_MO3	Takes 1 if student's mother finished primary education or she has some courses of secondary education (0 otherwise).
EDU_MO4	Takes 1 if student's mother finished secondary education (0 otherwise).
EDU_MO5	Takes 1 if student's mother had or finished technical post secondary education (0 otherwise).
EDU_MO6	Takes 1 if student's mother had some years or finished college education (0 otherwise).
EDU_FAC	Takes 1 if student's father had some years or finished college education (0 otherwise).
DEP_P1	Takes 1 if student's primary school is public (0 otherwise).
DEP_P2	Takes 1 if student's primary school is private and subsidized by the government (0 otherwise).
DEP_P3	Takes 1 if student's primary school is private and not subsidized by the government (0 otherwise).
DEP_S1	Takes 1 if student's high school is public (0 otherwise).
DEP_S2	Takes 1 if student's high school is private and subsidized by the government (0 otherwise).
DEP_S3	Takes 1 if student's high school is private and not subsidized by the government (0 otherwise).
SES_P1	Takes 1 if student's primary school belongs to the first socio-economic group type (0 otherwise).
SES_P2	Takes 1 if student's primary school belongs to the second socio-economic group type (0 otherwise).
SES_P3	Takes 1 if student's primary school belongs to the third socio-economic group type (0 otherwise).
SES_P4	Takes 1 if student's primary school belongs to the forth socio-economic group type (0 otherwise).
SES_P5	Takes 1 if student's primary school belongs to the fifth socio-economic group type (0 otherwise).
SES_P1	Takes 1 if student's high school belongs to the first socio-economic group type (0 otherwise).
SES_P2	Takes 1 if student's high school belongs to the second socio-economic group type (0 otherwise).
SES_P3	Takes 1 if student's high school belongs to the third socio-economic group type (0 otherwise).
SES_P4	Takes 1 if student's high school belongs to the forth socio-economic group type (0 otherwise).
SES_P5	Takes 1 if student's high school belongs to the fifth socio-economic group type (0 otherwise).
RURAL_P	Takes 1 if student's primary school is located in a rural area (0 otherwise).
RURAL_S	Takes 1 if student's high school is located in a rural area (0 otherwise).
LENG_CONT	Is the proportion (reported by the students) of the 8th year verbal test's contents that was covered in classes.
MATH_CONT	Is the proportion (reported by the students) of the 8th year math test's contents that was covered in classes.
NAT_CONT	Is the proportion (reported by the students) of the 8th year natural science test's contents that was covered in classes.
SOC_CONT	Is the proportion (reported by the students) of the 8th year social science test's contents that was covered in classes.
Like_math	*I like to study math*: 1 (strongly agree), 2 (agree), 3 (disagree and strongly disagree).
Like_spanish	*I like to study Spanish*: 1 (strongly agree), 2 (agree), 3 (disagree and strongly disagree).
	Primary Education Students' Performance
SIMCE_V_P	Verbal SIMCE at 8th primary grade.
SIMCE_M_P	Math SIMCE at 8th primary grade.
SIMCE_S_P	Social Science SIMCE at 8th primary grade.
SIMCE_N_P	Natural Science SIMCE at 8th primary grade.
GPA_P	Grade point average at 8th primary grade.

Variable	**Description**
	Secondary Education Students' Performance
SIMCE_V_S	Verbal SIMCE at 2nd secondary grade.
SIMCE_M_S	Math SIMCE at 2nd secondary grade.
PSU_M	Math national test for college admissions.
PSU_V	Verbal national test for college admissions.
GPA_S	Grade point average at 2nd secondary grade.
TAKE_PSU	Takes 1 if the student takes the PSU test (0 otherwise).
	Measures of Effort in Primary Education
ME1	*When I study, I exert effort even if it is a difficult subject*: 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
ME2	*When I study, I try hard to learn*: 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
ME3	*When I study and I am not getting something, I look for additional information*: 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
ATTEN_P	Percentage of attendance in 8th primary grade.
STUDY_LENG	How often the student studies Spanish: 1 (every or almost every day), 2 (some days a week), 3 (Just for exams), and 4 (Never besides the study at class time).
STUDY_MATH	How often the student studies Math: 1 (every or almost every day), 2 (some days a week), 3 (Just for exams), and 4 (Never besides the study at class time).
STUDY_NAT	How often the student studies Natural Science: 1 (every or almost every day), 2 (some days a week), 3 (Just for exams), and 4 (Never besides the study at class time).
STUDY_SOC	How often the student studies Social Science: 1 (every or almost every day), 2 (some days a week), 3 (Just for exams), and 4 (Never besides the study at class time).
	Measures of Effort in Secondary Education
ATTEN_S	Percentage of attendance in 2nd secondary grade.
EFFORT_P	The student exerts effort and she (or he) is persistent.
use_space	How often does the student do homework in the space conditioned to study at home: 1 (never), 2 (rarely), 3 (frequently) and 4 (almost always).
use_sb	How often does the student read textbooks at home: 1 (never), 2 (rarely), 3 (frequently) and 4 (almost always).
use_calc	How often does the student use calculator to study at home: 1 (never), 2 (rarely), 3 (frequently) and 4 (almost always).
	Measures of Learning Skills
MS1	*I feel able to understand the harder subjects covered by the teachers*: 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
MS2	*I trust that I can do excellent homework and exams*: 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
MS2	*If I set a goal about learning well something, I can do it*: 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
MS4	*If I decide not to have poor marks, I really can avoid them*: 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
MS5	*When I study I lose the focus, because I am not good at studying*: 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
REP_BI	Takes 1 if the student has repeated at least one grade, 0 otherwise.

Table 5: Independent Variables

Variable	Mean	Std. Dev.	Min	Max	N
SEX	0.47	0.50	0	1	146319
EDU_MO1	0.16	0.37	0	1	146319
EDU_MO2	0.12	0.32	0	1	146319
EDU_MO3	0.24	0.42	0	1	146319
EDU_MO4	0.26	0.44	0	1	146319
EDU_MO5	0.11	0.32	0	1	146319
EDU_MO6	0.10	0.30	0	1	146319
EDU_FAC	0.14	0.35	0	1	146319
DEP_P1	0.49	0.50	0	1	146319
DEP_P2	0.41	0.49	0	1	146319
DEP_P3	0.10	0.29	0	1	146319
DEP_S1	0.39	0.49	0	1	146319
DEP_S2	0.52	0.50	0	1	146319
DEP_S3	0.09	0.29	0	1	146319
SES_P1	0.08	0.27	0	1	146319
SES_P2	0.29	0.45	0	1	146319
SES_P3	0.36	0.48	0	1	146319
SES_P4	0.18	0.39	0	1	146319
SES_P5	0.09	0.29	0	1	146319
SES_S1	0.16	0.37	0	1	146319
SES_S2	0.37	0.48	0	1	146319
SES_S3	0.26	0.44	0	1	146319
SES_S4	0.12	0.33	0	1	146319
SES_S5	0.09	0.28	0	1	146319
RURAL_P	0.11	0.31	0	1	146319
RURAL_S	0.04	0.19	0	1	146319
LENG_CONT	0.91	0.06	0.31	1	146304
MATH_CONT	0.95	0.05	0.12	1	146318
NAT_CONT	0.84	0.09	0.30	1	146318
SOC_CONT	0.90	0.09	0.07	1	146318

Table 6: Primary Education Students' Performance

Variable	Mean	Std. Dev.	Min	Max	N
SIMCE_M_P	268.64	48.50	116	406	145236
SIMCE_V_P	267.56	48.34	96	392	145944
SIMCE_N_P	271.28	48.84	120	411	146177
SIMCE_S_P	265.98	47.60	113	387	145011
GPA_P	5.87	0.52	4	7	146319

Table 7: Secondary Education Students' Performance

Variable	Mean	Std. Dev.	Min	Max	N
SIMCE_M_S	268.59	62.68	93	427	146041
SIMCE_V_S	266.25	50.49	120	398	146083
PSU_M	508.20	110.70	150	850	113946
PSU_V	505.05	108.85	177	850	113946
TAKE_PSU	0.78	0.42	0	1	113946
GPA_S	537.39	100.93	208	826	146319

Table 8: Measures of Effort in Primary Education

Variable	Mean	Std. Dev.	Min	Max	N
ME1	1.72	0.81	1	4	146319
ME2	2.08	1.00	1	4	146319
ME3	1.52	0.72	1	4	146319
ATTEN_P	95.71	3.87	60	100	146319
STUDY_LENG	2.60	0.72	1	4	146319
STUDY_MATH	2.53	0.80	1	4	146319

Table 9: Measures of Effort in Secondary Education

Variable	Mean	Std. Dev.	Min	Max	N
ATTEN_S	93.78	3.86	71.5	100	146319
EFFORT_P	0.27	0.44	0	1	137532
use_calc	3.71	0.83	2	5	111366
use_sb	4.02	0.79	2	5	114742
use_space	4.18	0.82	2	5	92329

Table 10: Measures of Learning Skills

Variable	Mean	Std. Dev.	Min	Max	N
MS1	1.89	0.81	1	4	145938
MS2	1.59	0.74	1	4	145768
MS3	1.33	0.59	1	4	145581
MS4	1.56	0.80	1	4	145261
REP	0.07	0.25	0	1	144128

The definition of SES (socio-economics 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 vulnerability of the school.

To characterize student families I only use information of SIMCE 2006. This is because if I had also used 2004 information I would have lost more data, since some parents do not answer the questionnaire.

C Identification

Step 1, final score's expectation and variance:

Let $T_i \in \{PM_i, PV_i, GPA_i^h\}$, it is direct that

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

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:[72] $E[\delta_i^T | x_i^h] = 0$, $E[\delta_i^T | x_i^{e1h}] = 0$, $E[\delta_i^T | x_i^{\lambda 1p}] = 0$, $E[\delta_i^T | Me_{2i}^h] = 0$ and $E[\delta_i^T | M\lambda_{2i}^p] = 0$ from which β^T, β^{e1h} and $\beta^{\lambda 1p}$ are identified.[73] Therefore, $\{\beta^{pm}, \beta^{pv}, \beta^{gh}\}$ are identified.

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

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

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

Which means that $var(\varepsilon_i^{e1h})$ and $var(\varepsilon_i^{\lambda 1h})$ are identified, and consequently $\{var(\varepsilon_i^{pm}), var(\varepsilon_i^{pv}), var(\varepsilon_i^{gh})\}$ are also identified.

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

The nonparametric identification of $f(\lambda)$ and $f(e^h|x)$ can be proved following an analysis similar to Cunha and Heckman (2008). First, proceeding in a similar fashion as before, with two measures for each latent variable, 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:

$$\begin{aligned} \widehat{SIMCE}_{ji}^p &= \lambda_i + \widehat{\varepsilon}_i^{sjp} \\ M\lambda_{ji}^p - x_i^{\lambda 1p}\beta_1^{\lambda 1p} &= \lambda_i + \varepsilon_i^{\lambda 1p} \end{aligned}$$

[72] Because the effort decision is taken before the shocks' realization, such a decision is independent of the measurement errors, 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 the selection is not an issue for identification.

[73] This implies that all the parameters involved in a_{0i}, a_{1i} and b_{1i} are identified.

Therefore, because $\widehat{\varepsilon}_i^{sjp}$ and $\varepsilon_i^{\lambda 1p}$ are independent of each other and with respect to λ_i, the distribution of λ is identified (Cunha and Heckman (2008)).[74]

It is worth noting that, along the same lines, it is possible to prove the nonparametric identification of $f(e^h|x)$. This would allow another way to identify the utility parameters.

Step 3, parameters of the utility function:

Once the distribution of λ is identified, it is possible to identify the utility parameters.[75] First, notice that when $TCAT_i = 0$, then $Me^h_{1i} - \varepsilon_i^{e1h} = b_{i1}\theta_2$

$$\Rightarrow E[Me^h_{1i} - \varepsilon_i^{e1h}|TCAT_i = 0] = \theta_2 E[b_{1i}|TCAT_i = 0]$$

$$\Rightarrow \theta_2 = \frac{E[Me^h_{1i}|TCAT_i = 0]}{E[b_{1i}|TCAT_i = 0]}$$

which ensures the identification of θ_2.

Similarly, because $TCAT_i = 1$ implies that $Me^h_{1i} = g(x_i, a_{1i}(\lambda_i), b_{1i}, \theta_1, \theta_2, \sigma_\eta) + \varepsilon_i^{e1h}$, then[76]:

$$\Rightarrow E[Me^h_{1i}|\lambda_i, x_i, TCAT_i = 1] = E[g(x_i, a_{1i}(\lambda_i), b_{1i}, \theta_1, \theta_2, \sigma_\eta)|\lambda_i, x_i, TCAT_i = 1]$$

$$\Rightarrow \int_\lambda E[Me^h_{1i}|\lambda, x_i, TCAT_i = 1]f(\lambda)d\lambda = \int_\lambda E[g(x_i, a_{1i}(\lambda_i), b_{1i}, \theta_1, \theta_2, \sigma_\eta)|\lambda, x_i, TCAT_i = 1]f(\lambda)d\lambda$$

which allows for the identification of θ_1.

Finally, the identification of $\bar{FC}$ and σ_{fc} is trivial since

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

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

[74] The identification can be achieved under much weaker conditions regarding measurement errors. Indeed, independence is not necessary; see Cunha, Heckman, and Schennach (2010).

[75] Here I show the identification when the utility parameters are not individual specific, but the extension is trivial.

[76] $g(x_i, a_{1i}(\lambda_i), b_{1i}, \theta_1, \theta_2, \sigma_\eta)$ is the implicit function associated with the first order condition (10).

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

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

$$Pr(TCAT|x_i^h, e_i^h, \lambda_t, \Omega) = \Phi\left(\frac{D_i - \bar{FC}}{\sigma_{fc}}\right)^{TCAT_i}\left(1 - \Phi\left(\frac{D_i - \bar{FC}}{\sigma_{fc}}\right)\right)^{1-TCAT_i} \tag{25}$$

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

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

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

$$F_i(high\ school\ tests \mid Type_\lambda = t) = \prod_{T_i} f(T_i|x_i^h, e_i^h, \lambda_t, \Omega) \tag{26}$$

Similarly, the contributions to the likelihood of high school effort measures are described by:[77]

$$f(Me_{ji}^h|x_i^{ejh}, e_i^h, TCAT_i, \Omega) = \left[\phi\left(\frac{Me_{ji}^h - x_i^{ejh}\beta_1^{ejh} - e_i^{1h}\alpha^{ejh}}{\sigma_{\varepsilon^{ejh}}}\right)\frac{1}{\sigma_{\varepsilon^{ejh}}}\right]^{TCAT_i}$$

$$\left[\phi\left(\frac{Me_{ji}^h - x_i^{ejh}\beta_1^{ejh} - e_i^{0h}\alpha^{ejp}}{\sigma_{\varepsilon^{ejh}}}\right)\frac{1}{\sigma_{\varepsilon^{ejh}}}\right]^{1-TCAT_i}, \ j \in \{1, ..., J_{eh}\}$$

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

$$F_i(high\ school\ effort\ measures) = \prod_j f(Me^h_{ji}|x^{ejh}_i, e^h_i, TCAT_i, \Omega) \qquad (27)$$

Along the same lines, the contributions to the likelihood of the unobserved learning skill measures are described by:[78]

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

$$F_i(learning\ skill\ measures \mid Type_\lambda = t) = \prod_j f(M\lambda^p_{ji}|x^{\lambda jp}_i, \lambda_t, \Omega) \qquad (28)$$

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[79]:

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

$$F_i(primary\ school\ tests \mid Type_\lambda = t) = \prod_{T_i} f(T_i|x^p_i, e^p_i, \lambda_t, \Omega) \qquad (29)$$

Therefore, the likelihood contribution for the ith individual is thus:

$$L_i(\Omega) = log\Bigg(\sum_t F_i(high\ school\ tests \mid Type_\lambda = t)\ F_i(high\ school\ effort\ measures \mid Type_\lambda = t)$$
$$F_i(learning\ skill\ measures \mid Type_\lambda = t)\ F_i(primary\ school\ tests \mid Type_\lambda = t)\ \pi_t\Bigg) \qquad (30)$$

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

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

E Results

E.1 First satge parameters

Table 11: Primary School Attending Regression (output: ATTEN_P_hat)

VARIABLES	(1) ATTEN_P
ME1==2	-0.133***
	(0.0256)
ME1==3	-0.214***
	(0.0360)
ME1==4	-0.218***
	(0.0787)
ME2==2	0.0586**
	(0.0258)
ME2==3	0.115***
	(0.0284)
ME2==4	0.124***
	(0.0393)
ME3==2	-0.0500*
	(0.0263)
ME3==3	-0.123***
	(0.0422)
ME3==4	-0.280***
	(0.107)
STUDY_LENG==2	0.0140
	(0.0473)
STUDY_LENG==3	-0.0481
	(0.0497)
STUDY_LENG==4	-0.380***
	(0.0672)
STUDY_MATH==2	0.0277
	(0.0405)
STUDY_MATH==3	-0.0286
	(0.0432)
STUDY_MATH==4	-0.0208
	(0.0573)
Constant	95.82***
	(0.0435)
Observations	146,319
R-squared	0.002
F statistic	16.99

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Some variables are omited due to perfect multicolinearity.

Table 12: Repetitions Linear Probability Regression (output: REP_hat)

VARIABLES	(1) REP_BI
MS1==2	0.00174
	(0.00154)
MS1==3	0.0290***
	(0.00224)
MS1==4	0.0307***
	(0.00581)
MS2==2	0.00425**
	(0.00166)
MS2==3	0.0212***
	(0.00283)
MS2==4	0.0287***
	(0.00947)
MS3==2	0.00502**
	(0.00196)
MS3==3	0.0172***
	(0.00404)
MS3==4	0.0183
	(0.0132)
MS4==2	0.000835
	(0.00166)
MS4==3	0.0142***
	(0.00300)
MS4==4	0.00883**
	(0.00416)
Constant	0.0532***
	(0.00107)
Observations	141,916
R-squared	0.006
F statistic	58.09

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Some variables are omited due to perfect multicolinearity.

Table 13: Secondary School Attending Regression (output: ATTEN_S_hat)

VARIABLES	(1) ATTEN_S
EFFORT_P	0.152***
	(0.0278)
use_space==3	0.111
	(0.0819)
use_space==4	0.371***
	(0.0791)
use_space==5	0.498***
	(0.0798)
use_sb==2	-0.635***
	(0.102)
use_sb==3	-0.353***
	(0.0397)
use_sb==4	-0.0721**
	(0.0312)
use_calc==3	0.376***
	(0.0771)
use_calc==4	0.837***
	(0.0778)
use_calc==5	0.873***
	(0.0803)
Constant	93.14***
	(0.105)
Observations	83,366
R-squared	0.013
F statistic	103.2

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Some variables are omited due to perfect multicolinearity.

Table 14: Two Stage Least Square for Primary Education Students' Performance

VARIABLES	(1) simcev SIMCE_V_P	(2) simcem SIMCE_M_P	(3) simcen SIMCE_N_P	(4) simces SIMCE_S_P	(5) gpap GPA_P
SEX	-8.655***	9.499***	9.414***	9.565***	-0.164***
	(0.232)	(0.222)	(0.228)	(0.230)	(0.00258)
RURAL_P	1.562***	1.439***	3.548***	2.933***	0.0627***
	(0.463)	(0.449)	(0.434)	(0.459)	(0.00529)
SES_P2	1.055**	-0.880*	0.0666	0.626	-0.0681***
	(0.529)	(0.513)	(0.488)	(0.526)	(0.00605)
SES_P3	10.99***	8.463***	9.793***	11.44***	-0.0697***
	(0.581)	(0.559)	(0.536)	(0.573)	(0.00652)
SES_P4	28.05***	28.06***	29.29***	30.13***	-0.0390***
	(0.653)	(0.634)	(0.615)	(0.643)	(0.00724)
SES_P5	42.33***	48.86***	46.75***	42.92***	0.0632***
	(1.040)	(1.016)	(1.044)	(1.007)	(0.0114)
EDU_MO2	-6.537***	-5.747***	-6.131***	-5.259***	-0.0186***
	(0.454)	(0.430)	(0.431)	(0.447)	(0.00512)
EDU_MO3	-1.687***	-1.345***	-2.424***	-1.734***	0.0456***
	(0.380)	(0.359)	(0.367)	(0.374)	(0.00424)
EDU_MO4	6.452***	5.097***	5.362***	6.826***	0.138***
	(0.374)	(0.355)	(0.366)	(0.368)	(0.00414)
EDU_MO5	9.035***	7.438***	8.870***	10.36***	0.146***
	(0.467)	(0.452)	(0.469)	(0.462)	(0.00517)
EDU_MO6	14.68***	14.04***	16.16***	16.51***	0.209***
	(0.526)	(0.508)	(0.529)	(0.519)	(0.00585)
EDU_FAC	6.646***	7.574***	8.040***	6.743***	0.0583***
	(0.410)	(0.401)	(0.420)	(0.407)	(0.00455)
DEP_P2	1.392***	0.733***	2.320***	2.267***	-0.103***
	(0.289)	(0.276)	(0.283)	(0.287)	(0.00318)
DEP_P3	-0.515	-1.043	1.070	-2.413***	-0.136***
	(0.851)	(0.830)	(0.882)	(0.822)	(0.00923)
ATTEN_P_hat	-6.594***	-13.26***	-8.170***	-7.535***	0.262***
	(0.751)	(0.711)	(0.748)	(0.732)	(0.00813)
REP_hat	-347.8***	-430.5***	-360.9***	-379.0***	-6.906***
	(5.985)	(5.640)	(5.874)	(5.837)	(0.0652)
LENG_CONT	77.31***				
	(2.009)				
MATH_CONT		168.1***			
		(2.775)			
NAT_CONT			69.81***		
			(1.258)		
SOC_CONT				37.98***	
				(1.248)	
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

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

Table 15: Two Stage Least Square for Secondary Education Students' Performance

VARIABLES	(1) simcev SIMCE_V_S	(2) simcem SIMCE_M_S	(3) psum PSU_M	(4) psuv PSU_V	(5) gpas GPA_S
SEX	-7.136***	11.66***	23.28***	3.641***	-28.31***
	(0.271)	(0.322)	(0.579)	(0.594)	(0.572)
RURAL_S	-3.703***	-5.968***	-12.84***	-12.15***	-0.624
	(0.763)	(0.927)	(1.768)	(1.869)	(1.650)
SES_S2	10.02***	13.94***	22.80***	24.38***	-4.249***
	(0.443)	(0.530)	(1.058)	(1.091)	(0.935)
SES_S3	29.05***	39.20***	82.08***	77.91***	11.86***
	(0.506)	(0.603)	(1.142)	(1.178)	(1.078)
SES_S4	44.82***	65.23***	134.3***	122.1***	30.27***
	(0.635)	(0.750)	(1.357)	(1.420)	(1.339)
SES_S5	53.15***	79.01***	166.9***	151.2***	57.56***
	(1.289)	(1.443)	(2.538)	(2.630)	(2.818)
EDU_MO2	-4.213***	0.235	-1.655	-9.321***	16.25***
	(1.456)	(1.737)	(3.323)	(3.430)	(2.888)
EDU_MO3	-0.757	4.018**	2.099	-4.819	14.84***
	(1.430)	(1.704)	(3.238)	(3.348)	(2.830)
EDU_MO4	6.147***	11.08***	15.17***	10.40***	24.08***
	(1.427)	(1.699)	(3.221)	(3.329)	(2.824)
EDU_MO5	9.046***	13.78***	20.05***	18.37***	26.28***
	(1.459)	(1.734)	(3.271)	(3.381)	(2.899)
EDU_MO6	16.54***	21.63***	37.82***	37.63***	43.86***
	(1.482)	(1.760)	(3.310)	(3.421)	(2.959)
EDU_FAC	6.973***	9.051***	20.59***	20.08***	16.39***
	(0.451)	(0.528)	(0.880)	(0.913)	(0.975)
DEP_S2	-2.134***	-2.407***	-14.71***	-11.13***	-13.03***
	(0.315)	(0.378)	(0.693)	(0.710)	(0.657)
DEP_S3	-1.673	-0.909	-1.440	-1.454	-9.691***
	(1.167)	(1.298)	(2.256)	(2.316)	(2.555)
ATTEN_S_hat	3.817***	4.724***	11.52***	10.54***	26.94***
	(0.296)	(0.356)	(0.641)	(0.657)	(0.614)
REP_hat	-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

Robust standard errors in parentheses

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

E.2 Second stage parameters

Table 16: Second Stage Parameters

Utility								
θ_1	0.0000139	(0.0000008)	θ_3^2	0.00306	(0.00018)	FC	0.0000001	(0.0050620)
θ_2	3.6468447	(0.1390378)	θ_3^3	-0.00075	(0.00011)	σ_{fc}	4.1806149	(0.2548031)
θ_3^1	0.0008679	(0.0001100)	θ_3^4	-0.00099	(0.00016)			
Production function of tests								
β_λ^{smp}	-474.1999	(6.0361)	β_e^{svh}	3.9462	(0.6749)	β_{const}^{pv}	-475.8645	(123.3146)
β_λ^{svp}	-423.4805	(5.9159)	β_λ^{svh}	-584.0101	(8.1491)	β_e^{gh}	25.5289	(0.9754)
β_λ^{snp}	-8.1700	(0.0000)	β_{const}^{svh}	-83.4840	(63.3909)	β_λ^{gh}	-1049.5653	(12.1455)
β_λ^{ssp}	-7.5346	(0.0000)	β_e^{pm}	11.2758	(1.3678)	β_{const}^{gh}	-1792.2654	(91.5724)
β_λ^{gp}	-6.5496	(0.0767)	β_λ^{pm}	-1183.4871	(14.9213)	σ_{pm}	56.3147	(0.1301)
β_e^{smh}	4.8272	(0.8289)	β_{const}^{pm}	-563.7606	(128.6623)	σ_{pv}	53.5401	(0.1403)
β_λ^{smh}	-714.3026	(9.9824)	β_e^{pv}	10.5432	(1.3109)	σ_{gh}	70.5981	(0.1610)
β_{const}^{smh}	-176.2544	(77.8534)	β_λ^{pv}	-1293.8195	(15.5825)			
Measures of student effort at high school								
α_e(effort_p)	0.1099	(0.0076)	Cut_{sb}^2	3.1656	(0.8147)	Cut_{sp}^3	4.7551	(0.9201)
α_{const}(effort_p)	-10.9229	(0.7171)	Cut_{sb}^3	4.3483	(0.8149)	β_{ca}^{eh} (effort)	0.0388	(0.0084)
σ_{atten}	3.8072	(0.0073)	β_{sp}^{eh} (effort)	0.0466	(0.0098)	β_{ca}^{eh} (ses_s2)	0.0002	(0.0103)
β_{sb}^{eh} (effort)	0.0408	(0.0087)	β_{sp}^{eh} (ses_s2)	0.0520	(0.0128)	β_{ca}^{eh} (ses_s3)	0.0602	(0.0117)
β_{sb}^{eh} (ses_s2)	-0.0650	(0.0108)	β_{sp}^{eh} (ses_s3)	0.1168	(0.0142)	β_{ca}^{eh} (ses_s4)	0.2092	(0.0150)
β_{sb}^{eh} (ses_s3)	-0.0994	(0.0123)	β_{sp}^{eh} (ses_s4)	0.2183	(0.0173)	β_{ca}^{eh} (ses_s5)	0.2941	(0.0185)
β_{sb}^{eh} (ses_s4)	-0.0287	(0.0155)	β_{sp}^{eh} (ses_s5)	0.3943	(0.0205)	β_{ca}^{eh} (edu_fac)	0.0447	(0.0115)
β_{sb}^{eh} (ses_s5)	0.3718	(0.0190)	β_{sp}^{eh} (edu_fac)	0.1128	(0.0123)	Cut_{ca}^1	1.9459	(0.7840)
β_{sb}^{eh} (edu_fac)	0.0837	(0.0116)	Cut_{sp}^1	2.6673	(0.9198)	Cut_{ca}^2	3.5939	(0.7843)
Cut_{sb}^1	1.8099	(0.8143)	Cut_{sp}^2	3.6418	(0.9199)	Cut_{ca}^3	4.5660	(0.7843)
Measures and distribution of the learning skill								
α_{ms1}^λ	5.6102	(0.0883)	α_{ms3}^λ	4.4054	(0.0953)	λ(Type1)	0.0001	(0.000017)
Cut_{ms1}^1	0.0402	(0.0063)	Cut_{ms3}^1	0.9588	(0.0077)	λ(Type2)	0.0561	(0.0006)
Cut_{ms1}^2	1.1359	(0.0070)	Cut_{ms3}^2	1.9305	(0.0094)	λ(Type3)	0.1078	(0.0011)
Cut_{ms1}^3	2.4898	(0.0098)	Cut_{ms3}^3	0.9588	(0.0077)	λ(Type4)	0.1620	(0.0016)
α_{ms2}^λ	4.2743	(0.0888)	α_{ms4}^λ	3.8976	(0.0827)	π_1	0.2375	(0.0158)
Cut_{ms2}^1	0.4338	(0.0070)	Cut_{ms4}^1	0.5333	(0.0066)	π_2	0.3364	(0.0160)
Cut_{ms2}^2	1.4490	(0.0081)	Cut_{ms4}^2	1.4438	(0.0074)	π_3	0.3169	(0.0156)
Cut_{ms2}^3	2.6893	(0.0128)	Cut_{ms4}^3	2.1435	(0.0084)	π_4	0.1092	

The estimated parameters for the unobserved types probabilities are: p_1, p_2 and p_3, where $\pi_1 = \frac{p_1}{p_1+p_2+p_3+1}$. The reported (SE_i) refers to p_i.

E.3 Model fit

Table 17: Model fit by different groups

	PSU math		PSU verbal		GPA	
	Model	Data	Model	Data	Model	Data
All	508	508	505	505	537	538
Female	494	496	500	502	550	550
Male	525	522	511	509	523	523
SES 1	423	418	422	417	516	506
SES 2	452	453	453	454	517	511
SES 3	517	528	515	526	542	548
SES 4	581	590	573	581	570	582
SES 5	640	638	626	623	611	619
F wo college	490	490	488	488	529	528
F w college	597	602	589	592	591	599
Public	475	476	472	474	530	528
Private Sub	503	503	502	502	530	531
Private non Sub	637	635	622	621	607	616

Figure 7: Tests 2006

(a) Simce math 2006 (b) Simce verbal 2006

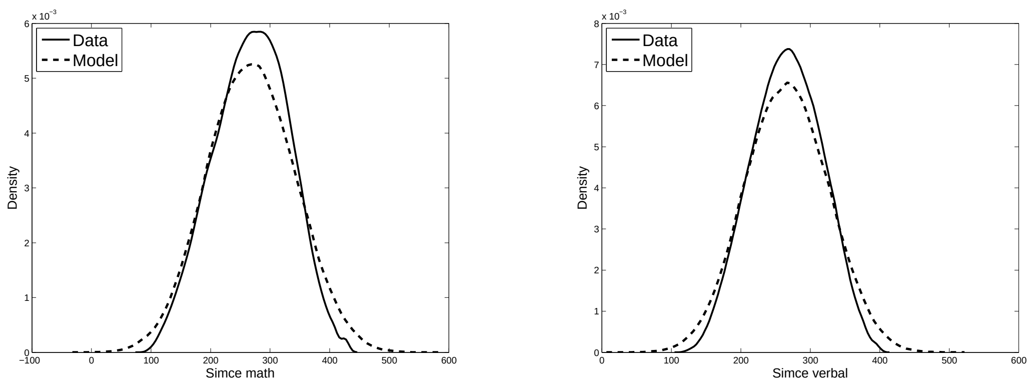
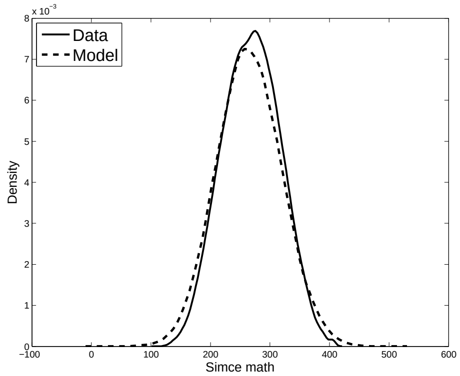
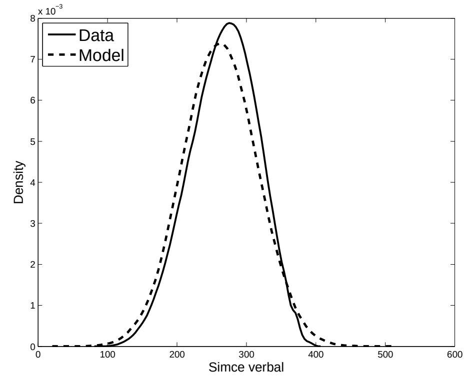


Figure 8: Tests 2004

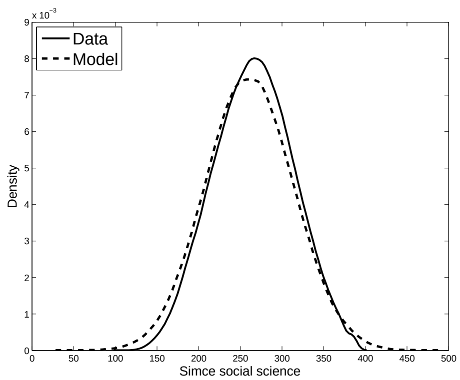
(a) Simce math 2004



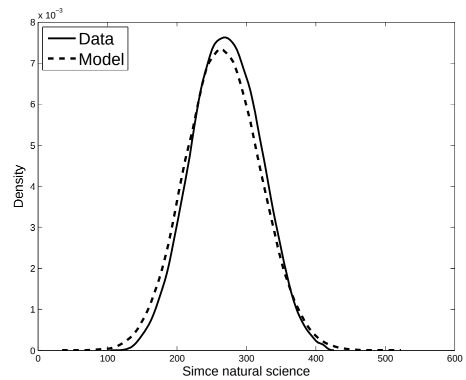
(b) Simce verbal 2004



(c) Simce natural science 2004



(d) Simce social science 2004



(e) GPA 2004

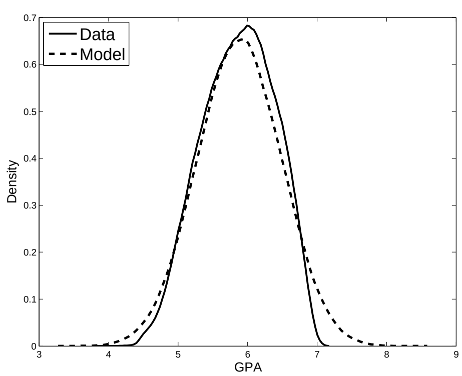
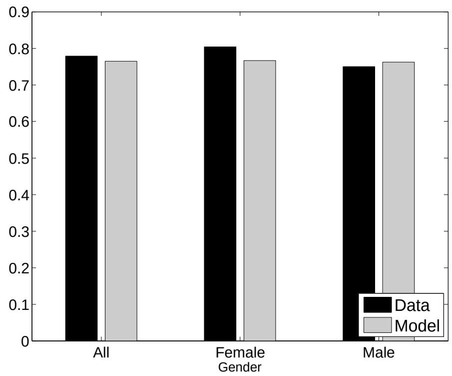
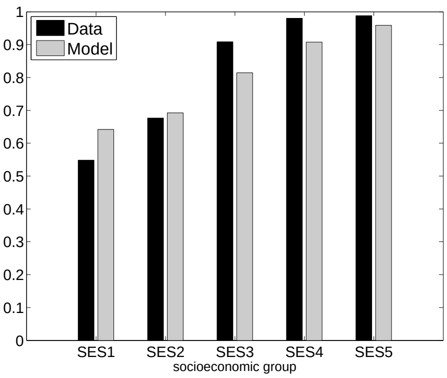


Figure 9: Fraction of the students taking the PSU by groups

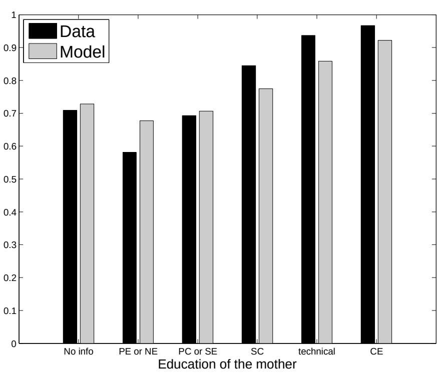
(a) By gender



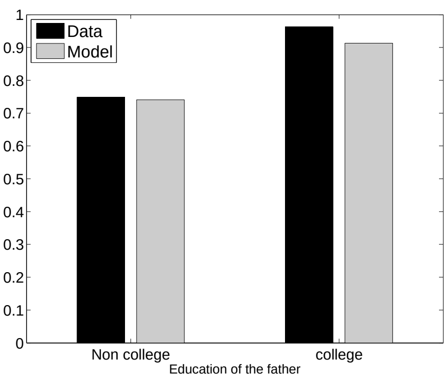
(b) By high school SES



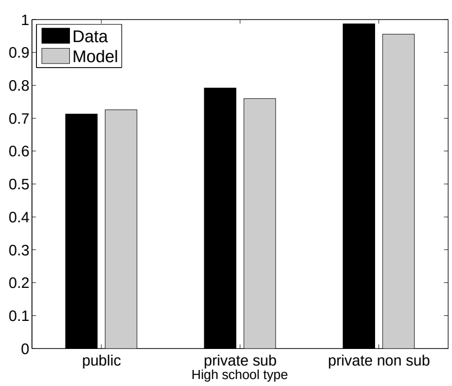
(c) By mother's education



(d) By father's education



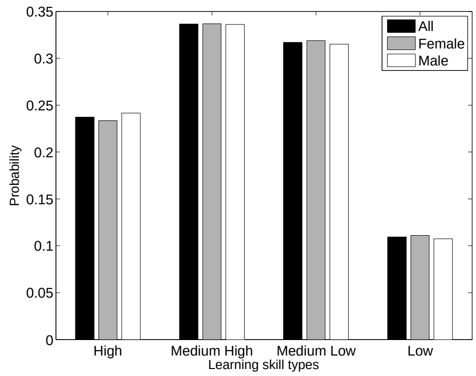
(e) By high school categories



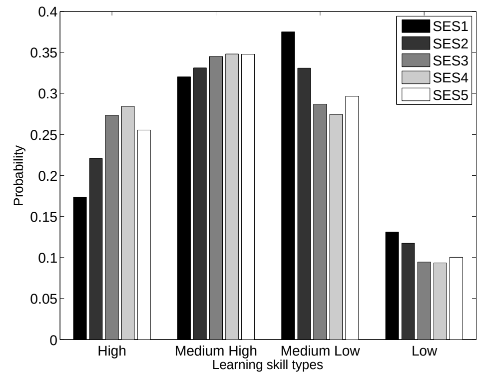
E.4 Unobserved types

Figure 10: Conditional probabilities of learning skill types by groups

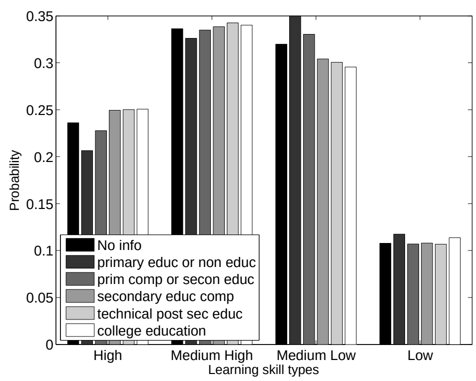
(a) All and by gender



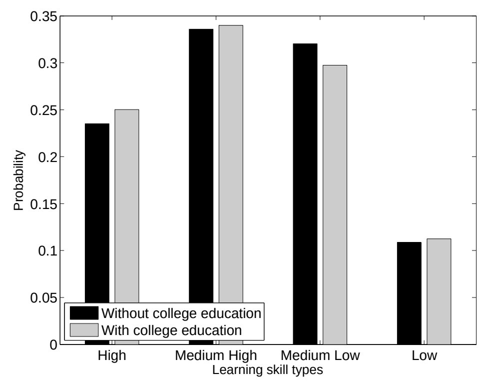
(b) By high school SES



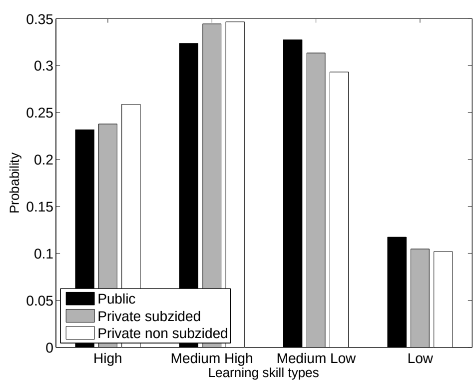
(c) By mother's education



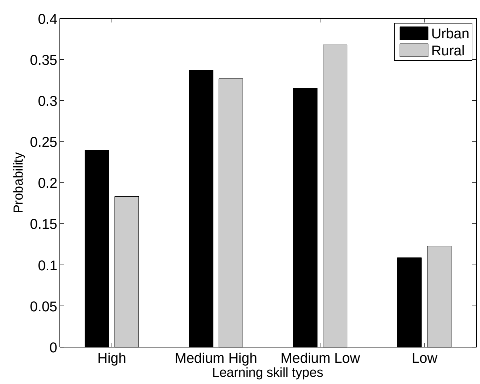
(d) By father's education



(e) By high school categories



(f) By rurality



F Counterfactual experiments

Figure 11: Impact of introducing quotas by SES on universities' socioeconomic composition

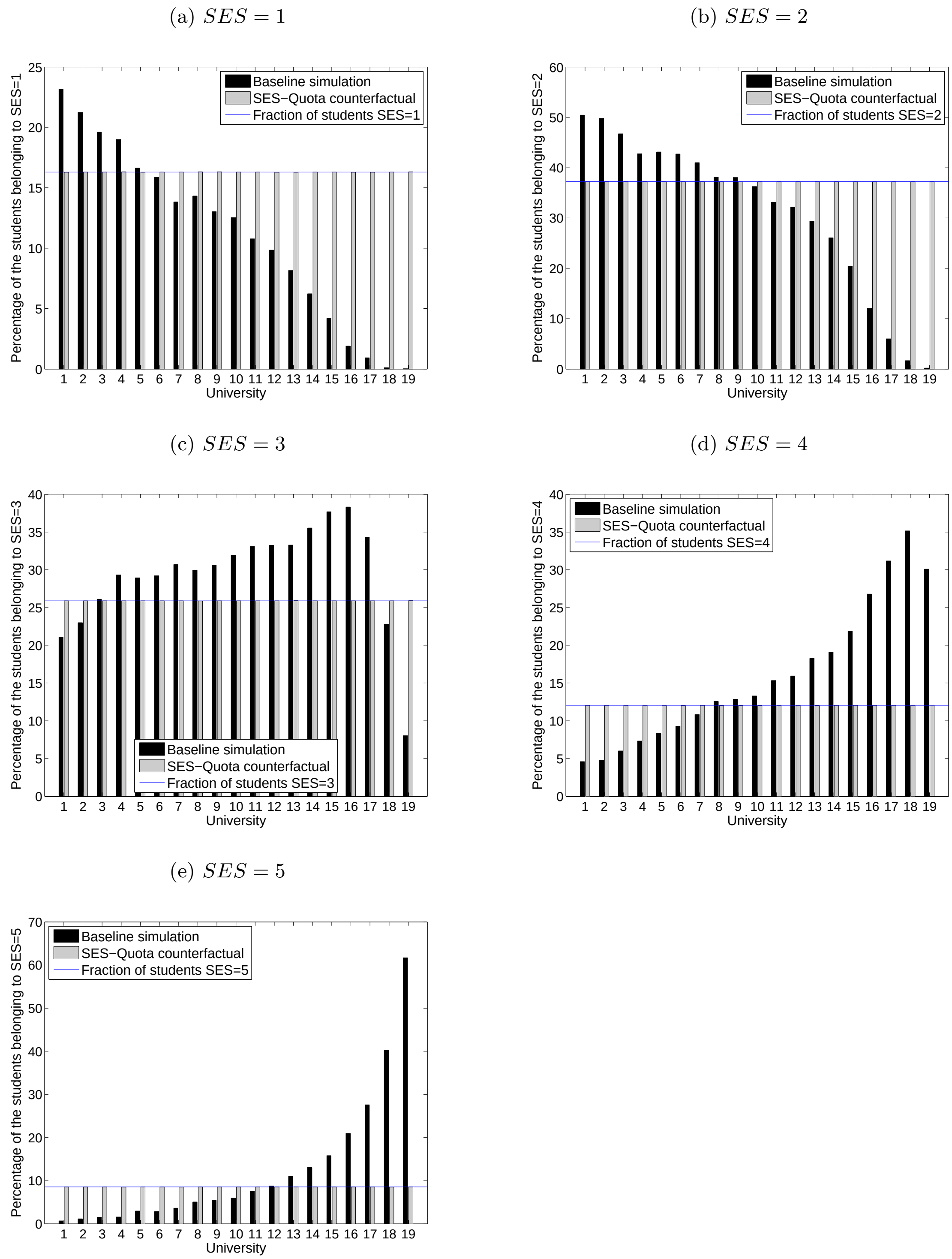
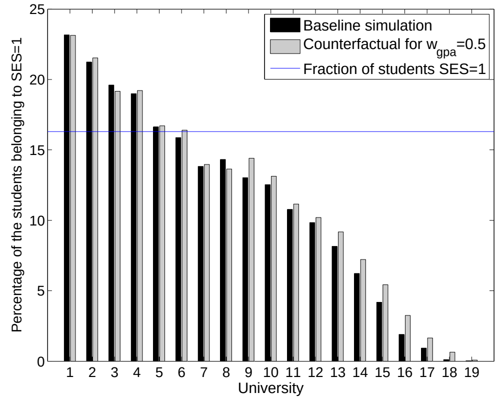
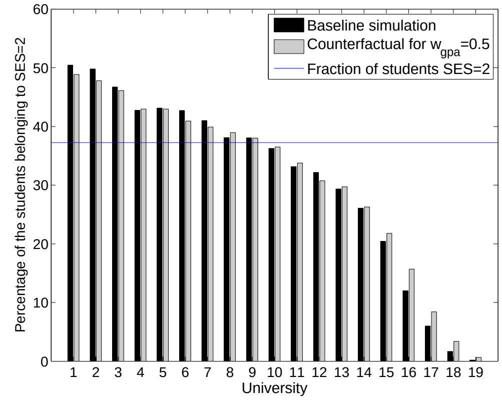


Figure 12: Impact of changing GPA weight from 0.4 to 0.5 on universities' socioeconomic composition

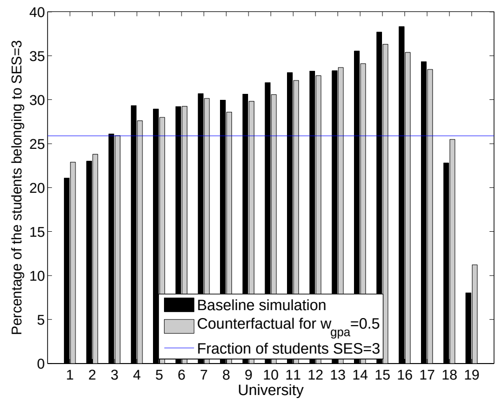
(a) $SES = 1$



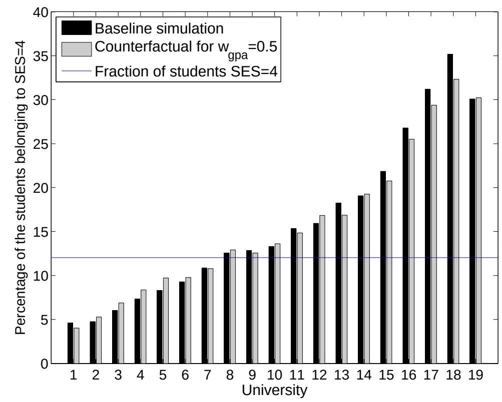
(b) $SES = 2$



(c) $SES = 3$



(d) $SES = 4$



(e) $SES = 5$

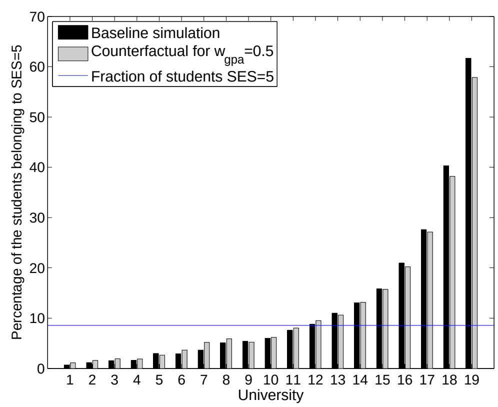
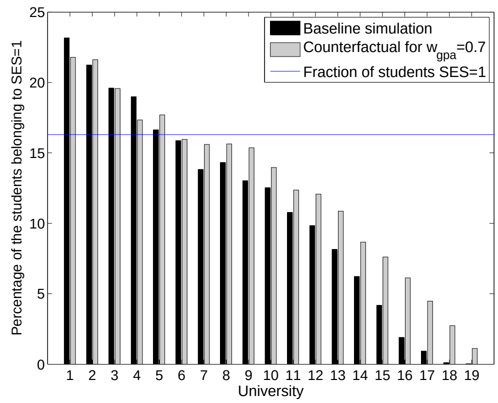
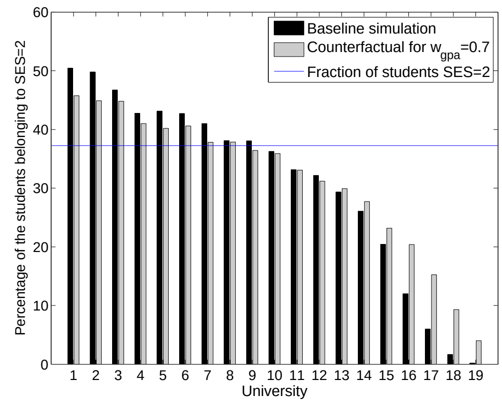


Figure 13: Impact of changing GPA weight from 0.4 to 0.7 on universities' socioeconomic composition

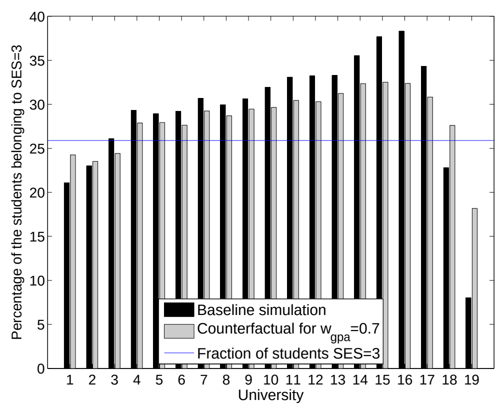
(a) $SES = 1$



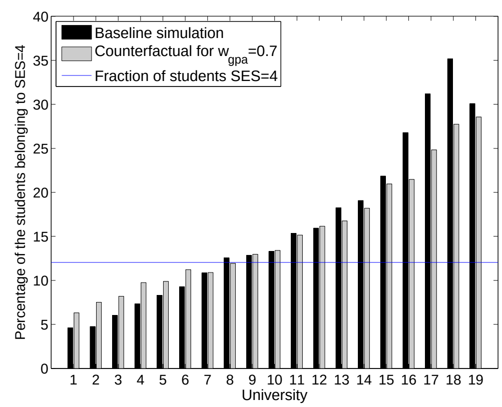
(b) $SES = 2$



(c) $SES = 3$



(d) $SES = 4$



(e) $SES = 5$

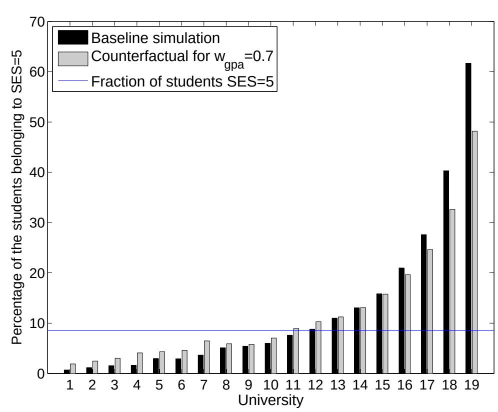
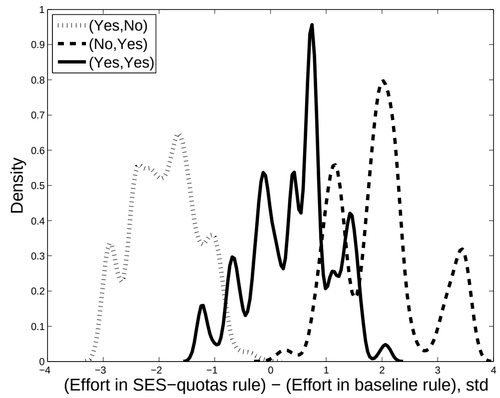
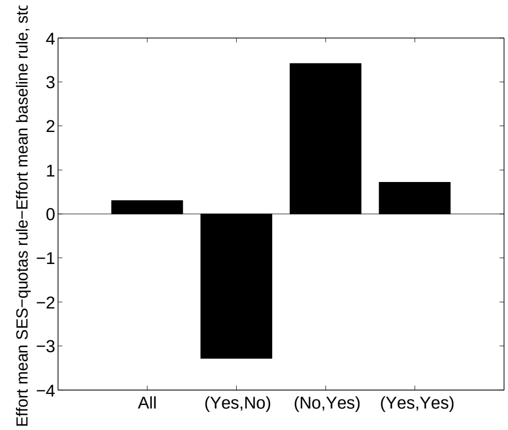


Figure 14: The impact on effort of quota by SES

(a) Densities



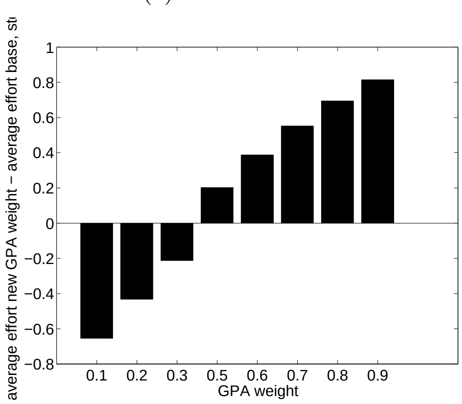
(b) means



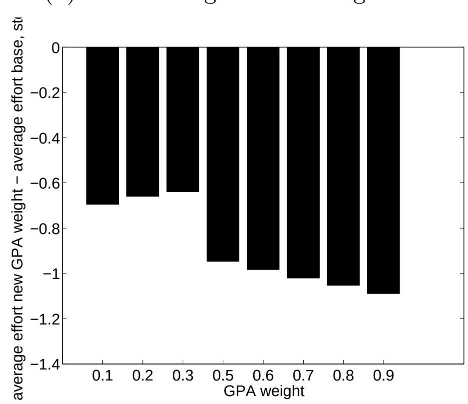
Note: ({Yes,No},{Yes,No}) stands for (Whether the students were taking the PSU in baseline scenario,Whether the students are taking the PSU in counterfactual scenario).

Figure 15: The impact on effort of changing GPA weight

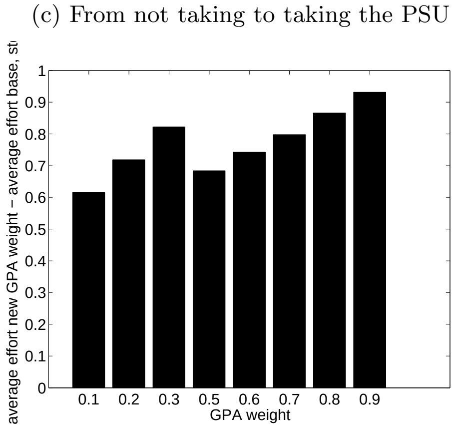
(a) All the students



(b) From taking to not taking the PSU



(c) From not taking to taking the PSU



(d) Always taking the PSU

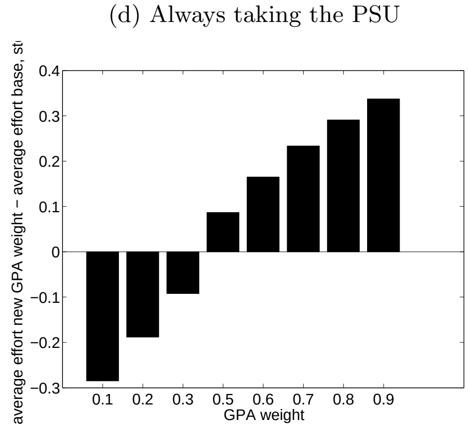
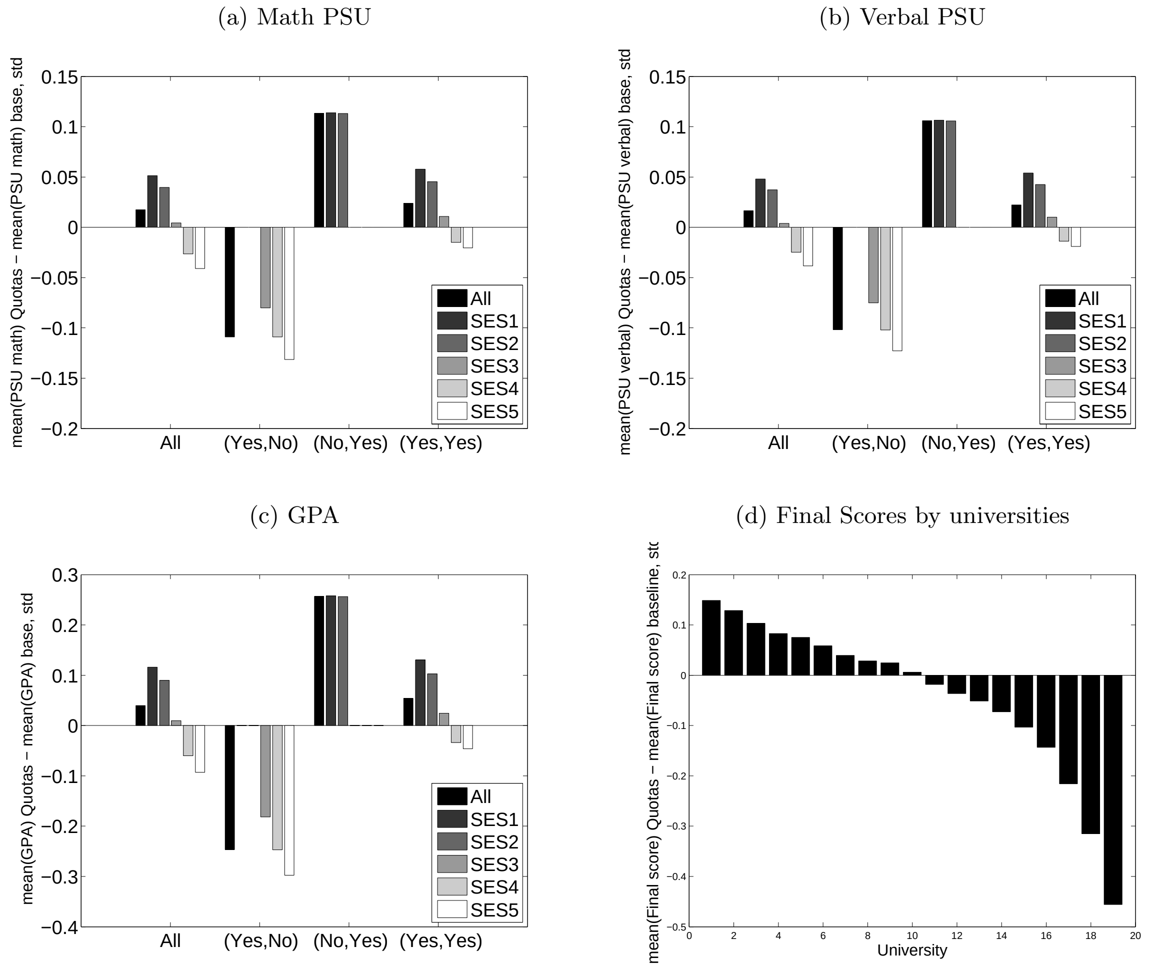


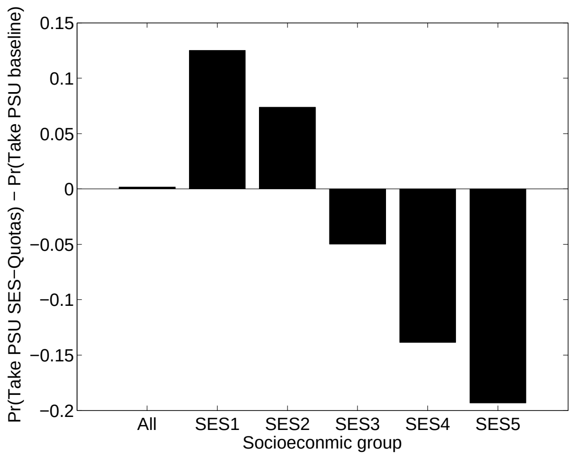
Figure 16: Impact of Quota system on tests by SES and universities



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).

Figure 17: Impact of SES-Quota system on who is taking the PSU

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



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

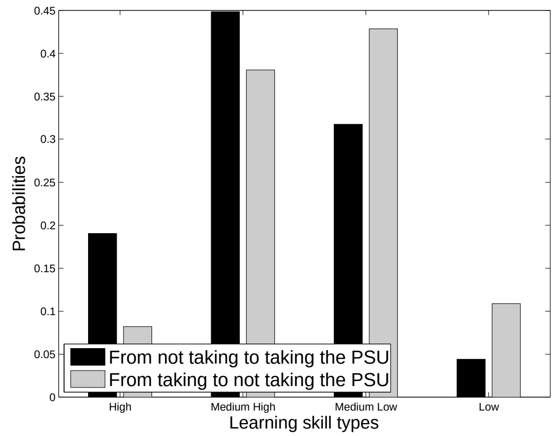
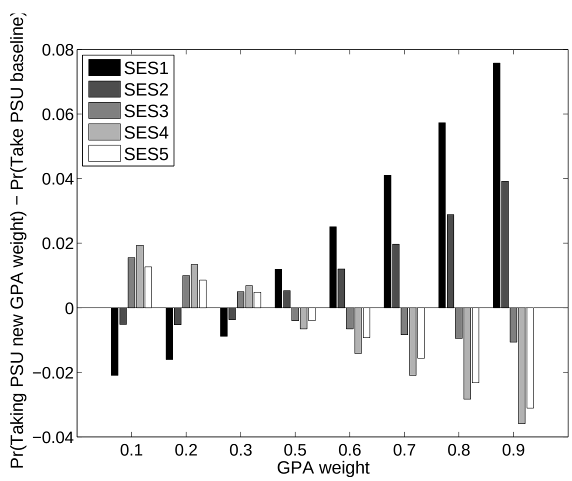


Figure 18: Impact of changing GPA weight on who is taking the PSU

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



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

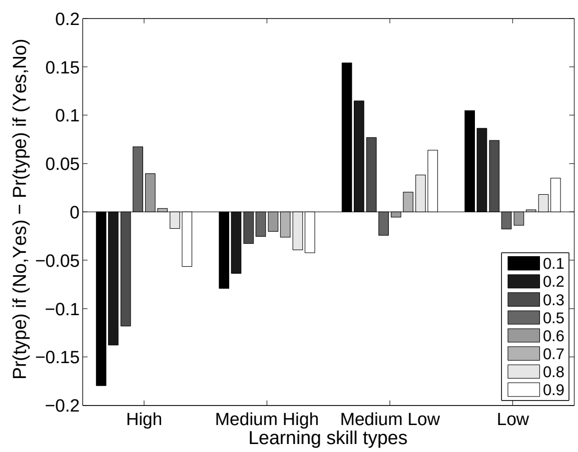


Figure 19: The impact on final-score cutoff and college admissions of introducing SES-Quota system, with and without endogenous effort

(a) Final-score cutoff

(b) University admission

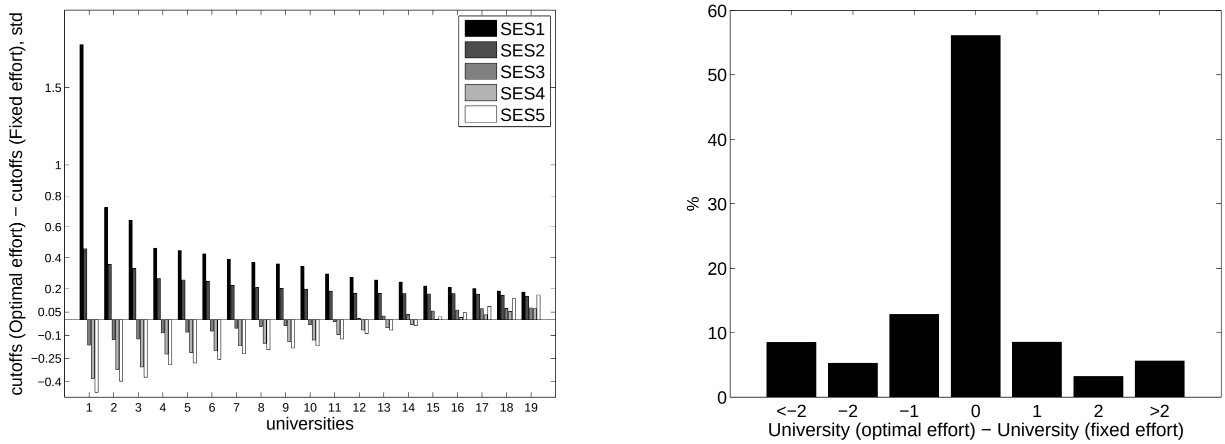


Figure 20: The impact on final-score cutoff of changing the GPA weight, with and without endogenous effort

(a) GPA weight = 0.5

(b) GPA weight = 0.7

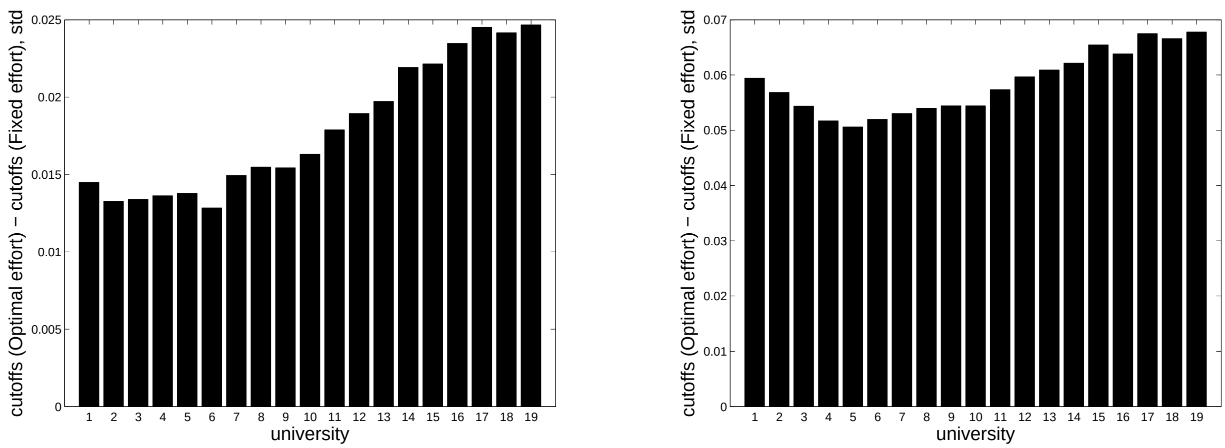
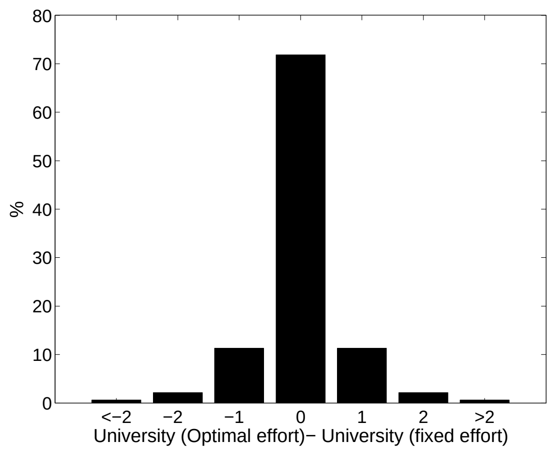


Figure 21: The impact of changing the GPA weight on university admissions, with and without endogenous effort

(a) GPA weight = 0.5



(b) GPA weight = 0.7