



University of Pennsylvania
ScholarlyCommons

Publicly Accessible Penn Dissertations

1-1-2014

Two Essays On The Economics Of Education

Nicolas Grau

University of Pennsylvania, ngrau@sas.upenn.edu

Follow this and additional works at: <http://repository.upenn.edu/edissertations>

 Part of the [Education Commons](#), and the [Labor Economics Commons](#)

Recommended Citation

Grau, Nicolas, "Two Essays On The Economics Of Education" (2014). *Publicly Accessible Penn Dissertations*. 1291.
<http://repository.upenn.edu/edissertations/1291>

This paper is posted at ScholarlyCommons. <http://repository.upenn.edu/edissertations/1291>
For more information, please contact libraryrepository@pobox.upenn.edu.

Two Essays On The Economics Of Education

Abstract

In this dissertation I address different topics in education policy, taking advantage of utilizing both micro-data and economic theory. The dissertation consists of two chapters, both using Chilean data. In chapter 1, The Impact of College Admissions Policies on The Performance of High School Students, I empirically evaluate the effects of college admissions policies on high school students' performance. In particular, I empirically demonstrate how increasing equality of opportunity may lead to a boost in average academic effort and shed light on the efficiency of alternative affirmative action policies. The results of this chapter suggest that affirmative action should not be seen only as a way to democratize the access to tertiary education, but also as a way to increase the motivation and performance of high school students. Methodologically speaking, this research contributes to the economic literature by estimating a rank-order tournament with heterogeneous-ability contestants.

In Chapter 2, A Dynamic Model of Elementary School Choice, I study how parents choose a primary school for their child. The approach of this chapter has three main contributions to the previous literature. The empirical strategy allows me to distinguish between first among different sources of observed preferences for private vis-`a-vis public schools, and second among different causes of unequal access to high-quality schools. In the paper I model and empirically estimate how parents may have misperceptions about school quality, because test scores depend on school quality and on the socioeconomic status (SES) of the school's population, parents can confound these two effects, confusing high quality schools with schools that have higher SES students. The paper contributes to the sparse literature on structural estimation with bounded rationality.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Economics

First Advisor

Kenneth I. Wolpin

Keywords

affirmative action, economics of education, school choice, structural estimation

Subject Categories

Education | Labor Economics

TWO ESSAYS ON THE ECONOMICS OF EDUCATION

Nicolás A. Grau Veloso

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania in Partial
Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2014

Supervisor of Dissertation

Kenneth I. Wolpin

Professor, Department of Economics
Chair of the Economics Department

Graduate Group Chairperson

George J. Mailath

Professor, Department of Economics

Dissertation Committee:

Flavio Cunha

Assistant Professor, Department of Economics

Petra E. Todd

Professor, Department of Economics

Kenneth I. Wolpin

Professor, Department of Economic
Chair of the Economics Department

TWO ESSAYS ON THE ECONOMICS OF EDUCATION

COPYRIGHT

Nicolás A. Grau Veloso

2014

Para Catalina, gracias por tu amor y solidaridad

Acknowledgements

I could not have completed this dissertation without the help and support of a number of people. My advisor and chair, Kenneth I. Wolpin, provided me with helpful comments, discussions, and support while guiding me during my years at the University of Pennsylvania. The professors at my committee were also of tremendous help; Flavio Cunha and Petra Todd thanks for your valuable support, comments and help.

This dissertation also benefits greatly from my classmates. From them I not only learned economics, but also how to be a better person and friend.

I would like to thank to my parents, Paulina and Francisco, and my brother Matias, for their unconditional support and for giving me a life full of love and opportunities.

I give my deepest thanks to my wife Catalina, for her love, support, and our life of happiness.

Finally, I would like to thank my sponsor, the Government of Chile. I hope to be able to contribute to the development of Chile and its people.

ABSTRACT
TWO ESSAYS ON THE ECONOMICS OF EDUCATION

Nicolás A. Grau Veloso
Kenneth I. Wolpin

In this dissertation I address different topics in education policy, taking advantage of utilizing both micro-data and economic theory. The dissertation consists of two chapters, both using Chilean data. In chapter 1, *The Impact of College Admissions Policies on The Performance of High School Students*, I empirically evaluate the effects of college admissions policies on high school students' performance. In particular, I empirically demonstrate how increasing equality of opportunity may lead to a boost in average academic effort and shed light on the efficiency of alternative affirmative action policies. The results of this chapter suggest that affirmative action should not be seen only as a way to democratize the access to tertiary education, but also as a way to increase the motivation and performance of high school students. Methodologically speaking, this research contributes to the economic literature by estimating a rank-order tournament with heterogeneous-ability contestants. In Chapter 2, *A Dynamic Model of Elementary School Choice*, I study how parents choose a primary school for their child. The approach of this chapter has three main contributions to the previous literature. The empirical strategy allows me to distinguish between first among different sources of observed preferences for private vis-à-vis public schools, and second among different causes of unequal access to high-quality schools. In the paper I model and empirically estimate how parents may have misperceptions about school quality, because test scores depend on school quality and on the socioeconomic status (SES) of the school's population, parents can confound these two effects, confusing high quality schools with schools that have higher SES students. The paper contributes to the sparse literature on structural estimation with bounded rationality.

Contents

Acknowledgements	iv
1 The Impact of College Admissions Policies on The Performance of High School Students	1
1.1 Introduction	1
1.2 The Model	8
1.3 The Chilean System for College Admissions and Data Description . .	15
1.4 Empirical Specification and Identification	22
1.5 Estimation	25
1.6 Results	27
1.7 Counterfactual Experiments	34
1.8 Conclusion	41
2 A Dynamic Model of Elementary School Choice	43
2.1 Introduction	43
2.2 Literature Review	46
2.3 The Chilean Educational System	49
2.4 The Model	51
2.5 Data and Empirical Implementation	60
2.6 Results	66
2.7 Conclusions	77
A The Impact of College Admissions Policies on The Performance of High School Students	79
A.1 Existence and Uniqueness	79
A.2 Variable Descriptions	87
A.3 Identification	91
A.4 Likelihood	93
A.5 Results	96
A.6 Counterfactual Experiments	106
B A Dynamic Model of Elementary School Choice	114
B.1 Figures and Tables	114
B.2 Estimated Parameters and Model's Inputs	117

B.3 Model Fit and Counterfactual Experiments (Complete Sample) . . .	124
--	-----

List of Tables

1.1	Cardinality of the student types groups	18
1.2	Universities' payoffs and cutoff scores	21
1.3	Correlations and Variance Decomposition for Effort Measures	31
2.1	Fraction of students changing school by grade (with respect to baseline in percentage points)	74
2.2	Student fraction by school type at first grade (with respect to baseline in percentage points)	75
2.3	Gain of knowledge ($KT - K0$) by parents education (with respect to baseline in standard deviations).	76
A.1	Variable Descriptions	87
A.2	Independent Variables	89
A.3	Primary Education Students' Performance	89
A.4	Secondary Education Students' Performance	90
A.5	Measures of Effort in Primary Education	90
A.6	Measures of Effort in Secondary Education	90
A.7	Measures of Learning Skills	90
A.8	Primary School Attending Regression	96
A.9	Repetitions Linear Probability Regression	97
A.10	Secondary School Attending Regression	98
A.11	Two Stage Least Square for Primary Education Students' Performance	99
A.12	Two Stage Least Square for Secondary Education Students' Performance	100
A.13	Second Stage Parameters	101
A.14	Model fit by different groups	102
B.1	Fraction of students changing school	114
B.2	Total school change	115
B.3	Descriptive Statistics of the Two Samples	116
B.4	Knowledge Production Function	118
B.5	GPA Production Function	119
B.6	Parameters of the Utility Function	121
B.7	Parameters of the Admission Probability Function	122
B.8	Parameters that Determine Parents' Quality Perception	123
B.9	Parameters of the Parents Involment Probability Function	123

B.10 Fraction of students changing school by grade (with respect to baseline in percentage points), complete sample.	125
B.11 Student fraction by school type at first grade (with respect to baseline in percentage points), complete sample.	125
B.12 Gain of knowledge ($KT - K0$) by parents education (with respect to baseline in standard deviations), complete sample.	125

List of Figures

1.1	Imputed fifth year wages and locally weighted regression	20
1.2	Final-score cutoffs for 2009 university admissions process	29
1.3	Model fit in tests determining final score	30
1.4	The impact of types on tests	33
1.5	Average effort: SES-Quota versus changing GPA's weight	39
1.6	Expected tests and GPA: SES-Quota versus changing GPA's weight	40
2.1	Fraction of students changing their school by grade	68
2.2	Average total change by parents education	68
2.3	Student fraction by school type	69
2.4	Average quality of the school selected by grade	70
2.5	Gain in Knowledge by parents education ($K_T - K_0$)	71
2.6	Quality misperception by school administration type (2004)	72
2.7	Increase in the fraction of students in schools with higher quality	73
A.1	Tests 2006	102
A.2	Tests 2004	103
A.3	Fraction of the students taking the PSU by groups	104
A.4	Conditional probabilities of learning skill types by groups	105
A.5	Impact of introducing quotas by SES on universities' socioeconomic composition	106
A.6	Impact of changing GPA weight from 0.4 to 0.5 on universities' socioeconomic composition	107
A.7	Impact of changing GPA weight from 0.4 to 0.7 on universities' socioeconomic composition	108
A.8	The impact on effort of quota by SES	109
A.9	The impact on effort of changing GPA weight	109
A.10	Impact of Quota system on tests by SES and universities	110
A.11	Impact of SES-Quota system on who is taking the PSU	111
A.12	Impact of changing GPA weight on who is taking the PSU	111
A.13	The impact on final-score cutoff and college admissions of introducing SES-Quota system, with and without endogenous effort	112
A.14	The impact on final-score cutoff of changing the GPA weight, with and without endogenous effort	112

A.15	The impact of changing the GPA weight on university admissions, with and without endogenous effort	113
B.1	Histogram of the Size of the Choice Sets	117
B.2	Distribution of Estimated Quality by School Type in 2004	118
B.3	GPA Standards	120
B.4	Model fit for the complete sample	124

Chapter 1

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

1.1 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

¹For 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).

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 which students decide their effort and whether or not to take the college admissions test,

²Theoretically 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.

³It 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)).

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

⁴There 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.

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

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 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 counterfactual experiments. (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 Assunção (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

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

⁷To 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.

⁸Among 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.

this is Ferman and Assunção (2011), who used difference-in-difference techniques and quasi-experimental data from Brazilian secondary education, where political forces abruptly imposed an admissions quota for two of Rio De Janeiro's top public universities. They estimate that the quota altered incentives, thus producing a 5.5% decrease in standardized test scores among the favored group, 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 university 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

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

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.

The paper proceeds as follows. Section 1.2 details the features of the model. Section 1.3 describes the Chilean college admissions process, explaining the main features of the data. Section 1.4 discusses the empirical implementation of the model and proves the identification of the model's parameters. Section 1.5 presents the estimation procedure. In Section 1.6, the model fit is discussed along with other aspects of the estimation results. Section 1.7 describes the counterfactual experiments results. Finally, Section 1.8 concludes and discusses future research.

¹⁰Vukina 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.

1.2 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 $i \in \{1, 2, \dots, M\}$ 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 $\{R_1, R_2, \dots, R_N\}$, 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

¹¹Although 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.

¹²This 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.

¹³To 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.

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 ($FC_i \sim N(\bar{FC}, \sigma_{fc}^2)$), 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:

$$FS_i = P_{pm} * PM_i + P_{pv} * PV_i + P_g * GPA_i, \quad (1.1)$$

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:

$$PM_i = \beta_0^{pm} + x_i \beta_1^{pm} + e_i \beta_2^{pm} + \lambda_i \beta_3^{pm} + \varepsilon_i^{pm}, \quad (1.2)$$

$$PV_i = \beta_0^{pv} + x_i \beta_1^{pv} + e_i \beta_2^{pv} + \lambda_i \beta_3^{pv} + \varepsilon_i^{pv}, \quad (1.3)$$

$$GPA_i = \beta_0^g + x_i \beta_1^g + e_i \beta_2^g + \lambda_i \beta_3^g + \varepsilon_i^g. \quad (1.4)$$

$\varepsilon_i^k \sim N(0, \sigma_k^2)$, $\varepsilon_i^k \perp \varepsilon_i^{k'} \forall k \neq k'$ 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 ($\{r_2, r_3, \dots, r_N\}$) 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 and smaller than r_{n+1} . 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:

$$U^0(e) = \theta_1 R_1 + \theta_2 GPA(e) - \frac{e^2}{2}, \quad (1.5)$$

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

$$U^1(e) = \theta_1 \sum_{n=1}^N R_n 1(r_n \leq FS(e) < r_{n+1}) + \theta_2 GPA(e) - FC - \frac{e^2}{2}, \quad (1.6)$$

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 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.

Student's Problem

Given a vector r , the optimization problem for those who do and do not take the national college admissions test can be written as:¹⁶

¹⁴ $r_1 = -\infty$.

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

¹⁶ Φ denotes the standard normal distribution function.

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

$$\max_{e \geq 0} U_i^1(e) = \tag{1.7}$$

$$\max_{e \geq 0} \left\{ \theta_1 \sum_{n=1}^{N-1} (R_n - R_{n+1}) \Phi \left(\frac{r_{n+1} - a_{1i}e - a_{0i}}{\sigma_\eta} \right) + \theta_1 R_N + \theta_2 (b_{0i} + b_{1i}e) - FC - \frac{e^2}{2} \right\}. \tag{1.8}$$

Where:

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

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

$$TCAT_i = \begin{cases} 1 & \text{if } \max_{e \geq 0} U_i^1(e) \geq \max_{e \geq 0} U_i^0(e) \\ 0 & \text{if } \max_{e \geq 0} U_i^1(e) < \max_{e \geq 0} U_i^0(e) \end{cases} \tag{1.9}$$

Lemma 1: Given a vector r , the student's problem (1.7) 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 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 student's problem is characterized by the following first order conditions:

For those who do not take the college admissions test:

$$\hat{e}_i^0 = \theta_2 b_{1i}. \quad (1.10)$$

For those who take the college admissions test:¹⁸

$$\hat{e}_i^1 = \theta_1 \sum_{n=1}^{N-1} (R_{n+1} - R_n) \phi \left(\frac{r_{n+1} - a_{1i} \hat{e}_i^1 - a_{0i}}{\sigma_\eta} \right) \frac{a_{1i}}{\sigma_\eta} + \theta_2 b_{1i}, \quad (1.11)$$

\Rightarrow

$$TCAT_i = \begin{cases} 1 & \text{if } D_i \geq FC_i \\ 0 & \text{if } D_i < FC_i \end{cases} \quad (1.12)$$

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

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

¹⁷In 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.

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 (1.11) 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

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} , $\forall i$:

$$- \hat{e}_i^0 = \theta_2 b_{1i},$$

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

$$- \hat{D}_i = (U_i^1(\hat{e}_i^1, \hat{r}) + FC_i) - U_i^0(\hat{e}_i^0).$$

- $\forall n = 2, \dots, N$:

¹⁹This is shown in Appendix A.1.

²⁰Therefore, \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 $TCAT_i$.

²¹Again, 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.

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

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 $\sum_i m_i \Phi \left(\frac{\theta_1(R_N - R_1) - \bar{F}C}{\sigma_{fc}} \right) > \sum_{\delta=2}^N S_\delta$, 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.1.

Although there is not a proof for $N > 2$, in Appendix A.1 I present a result which limits the potential extent of multiple equilibria. In particular, it narrows the possibility of having *high* and *low* effort equilibria.

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 first order conditions of the student's problem 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.

1.3 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

²²CASEN 2009 (Chilean survey for socioeconomic characterization).

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 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).

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

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

²⁵The 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.

- **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. 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 A.2 there is a description of the variables considered in this paper along with some statistics.

Something worth highlighting is the fact that all independent variables that deter-

²⁶The 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.

mine the effort decision are discrete.²⁷ 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 1.1), the assumption of a continuum of agents within each type is closer to the data specification.

Table 1.1: 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_2).

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).²⁸ Thus, I have one final-score cutoff for each student who took the college admissions test.

²⁷All 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).

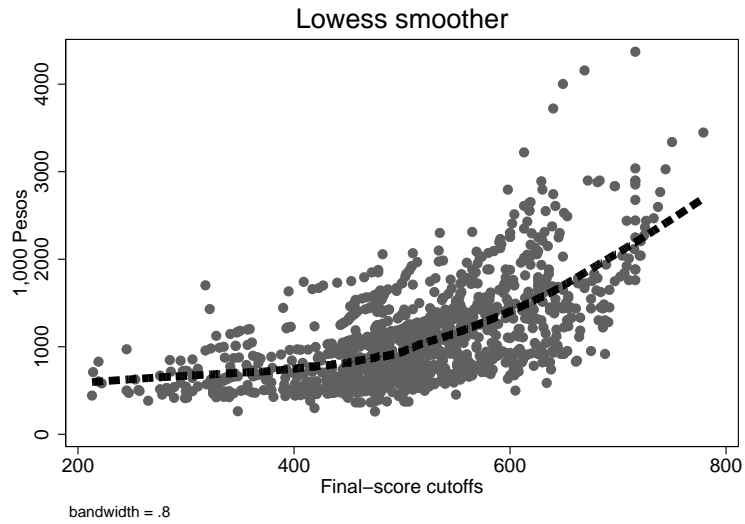
²⁸They are close to the mode in my data base. They can not be exactly the mode in order to have weights that add to one.

- I use the information of the fifth year’s wages for each area of study.²⁹
- 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 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 1.1.
- 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 1.1. 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.³⁰

²⁹The resulting final-score cutoffs are quite similar if I use first year’s wages.

³⁰I use k-means clustering algorithm.

Figure 1.1: Imputed fifth year wages and locally weighted regression



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 and r_{n+1} . This means that all my counterfactual experiments will assume that the share of students who come directly from high school is invariant to policy experiments. In Table 1.2, the resulting final-score cutoff (r), payoffs (R), and seats available (S) for each university are presented.

Table 1.2: 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.

1.4 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.³¹

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:

$$PM_i = \beta_0^{pm} + x_i^h \beta_1^{pm} + e_i^h \beta_2^{pm} + \lambda_i \beta_3^{pm} + \varepsilon_i^{pm}, \quad \forall i \text{ s.t. } TCAT_i = 1, \quad (1.13)$$

$$PV_i = \beta_0^{pv} + x_i^h \beta_1^{pv} + e_i^h \beta_2^{pv} + \lambda_i \beta_3^{pv} + \varepsilon_i^{pv}, \quad \forall i \text{ s.t. } TCAT_i = 1, \quad (1.14)$$

³¹In 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.

$$GPA_i^h = \beta_0^{gh} + x_i^h \beta_1^{gh} + e_i^h \beta_2^{gh} + \lambda_i \beta_3^{gh} + \varepsilon_i^{gh}. \quad (1.15)$$

High school performance and effort measurements:

$$SIMCE_{ji}^h = \beta_0^{sjh} + x_i^h \beta_1^{sjh} + e_i^h \beta_2^{sjh} + \lambda_i \beta_3^{sjh} + \varepsilon_i^{sjh}, \quad j \in \{verbal, math\}, \quad (1.16)$$

$$Me_{ji}^h = x_i^{ejh} \beta_1^{ejh} + e_i^h \alpha^{ejh} + \varepsilon_i^{ejh}, \quad j \in \{1, \dots, J_{eh}\} \quad J_{eh} \geq 2. \quad (1.17)$$

Primary school performance, learning skill and effort measurements:

$$SIMCE_{ji}^p = \beta_0^{sjp} + x_i^p \beta_1^{sjp} + e_i^p \beta_2^{sjp} + \lambda_i \beta_3^{sjp} + \varepsilon_i^{sjp}, \quad (1.18)$$

$$j \in \{verbal, math, natural\ science, social\ science\},$$

$$GPA_i^p = \beta_0^{gp} + x_i^p \beta_1^{gp} + e_i^p \beta_2^{gp} + \lambda_i \beta_3^{gp} + \varepsilon_i^{gp}, \quad (1.19)$$

$$Me_{ji}^p = x_i^{ejp} \beta_1^{ejp} + e_i^p \alpha^{ejp} + \varepsilon_i^{ejp}, \quad j \in \{1, \dots, J_{ep}\} \quad J_{ep} \geq 2, \quad (1.20)$$

$$M\lambda_{ji}^p = x_i^{\lambda jp} \beta_1^{\lambda jp} + \lambda_i \alpha^{\lambda jp} + \varepsilon_i^{\lambda jp}, \quad j \in \{1, \dots, J_\lambda\} \quad J_\lambda \geq 2. \quad (1.21)$$

In this setup, I assume that all the ε_i s are normally and independently distributed.³² Namely, conditional on observable variables, the correlation across equations is only given by the unobserved skill heterogeneity. In Appendix A.2, there is a description of the different dependent and independent variables used in the estimation. The following are relevant for the identification analysis:

³²There 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.

- $M\lambda_{1i}^p$, 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: $\alpha^{e1h} = \alpha^{e1p} = \alpha^{\lambda1p} = 1$. 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.³³ 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.³⁴
- 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:³⁵

$$\theta_{3i} = \theta_3^1 1(\textit{Like math} = 2) + \theta_3^2 1(\textit{Like math} = 3) + \theta_3^3 1(\textit{Like spanish} = 2) + \theta_3^4 1(\textit{Like spanish} = 3).$$

³³I also assume that $x_i^{\lambda jp}$, $x_i^{e jp}$ and x_i^p do not have elements in common, and the same for $x_i^{e jh}$ and x_i^h ; this just for simplicity.

³⁴Using attendance as a measure of effort is a common practice; see for example Hastings, Neilson, and Zimmerman (2012)

³⁵In 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 $\tilde{FC}_i \sim N(\bar{FC} \exp(-\theta_{3i}), \sigma_{fc}^2 \exp(-2\theta_{3i}))$.

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 A.3, has three steps. First, I identify the final score's expectation and variance.³⁶ Second, I non parametrically identify the distribution of learning skills. Third, I identify the utility parameters from different moments of the measures of effort.

1.5 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 ((1.13), (1.14), (1.15), (1.16) and (1.18)) 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

³⁶If $Var(\varepsilon_i^{pm})$, $Var(\varepsilon_i^{pv})$ and $Var(\varepsilon_i^{gh})$ are identified, then σ_η is also identified.

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.³⁷ 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.³⁸ Second, the first order conditions, which lack a closed form solution, should only be solved for the 2,596 student types.
- Calculate the likelihood function.
- Continue with a new guess until finding the Ω_2 that maximizes the likelihood function.³⁹

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

³⁷This 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).

³⁸Because I only need to calculate the first order conditions of the student's problem, the estimation method used is maximum likelihood as opposed to simulated maximum likelihood.

³⁹This is done using the derivative free solver, HOPSPACK.

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 A.4, I describe the contributions of different data to the likelihood.

1.6 Results

The first stage estimation results are presented in Appendix A.5 (Tables A.8, A.9, and A.10). 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.⁴⁰

Second, in the case of the OLS regressions where one of the secondary education performances is the dependent variable (Table A.12), 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

⁴⁰The F statistics are: 16.99 (Primary School Attending Regression), 103.19 (Secondary School Attending Regression), and 58.09 (Repetitions Linear Probability Regression).

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 A.11). 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.⁴¹ Furthermore, the effect of the predicted learning skills is positive and highly relevant in all equations.⁴²

The parameters estimated in the second stage are shown in Appendix A.5 (Table A.13).⁴³ 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.

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 first order conditions of the student's problem are 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.⁴⁴ In this regard, Figure 1.2 shows that the simulated vector r captures the trend and magnitudes of the data fairly well.

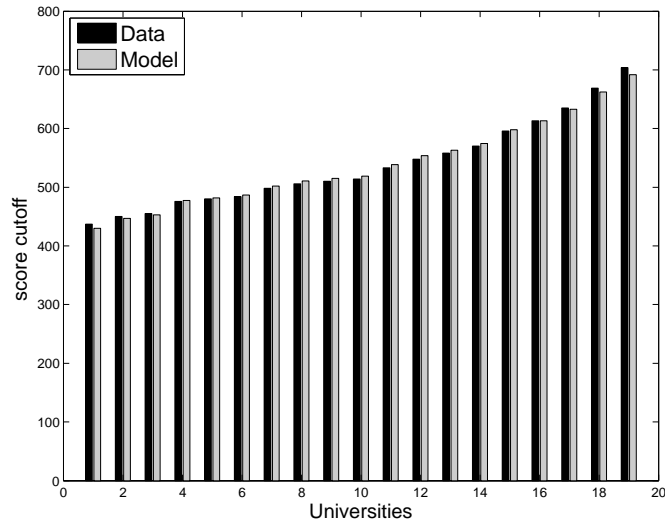
⁴¹In both cases, the effect is statistically significant.

⁴²The parameters are negative because the variables are ordered from more to fewer skills.

⁴³Some parameters are estimated in both stages. In that case, I keep for simulations the ones estimated in the second stage.

⁴⁴The 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 ($\max_{n \in \{1, \dots, N-1\}} |r_n^0 - r_n^1| < \epsilon$), otherwise restart from point (2) with r^1 as the new guess.

Figure 1.2: Final-score cutoffs for 2009 university admissions process

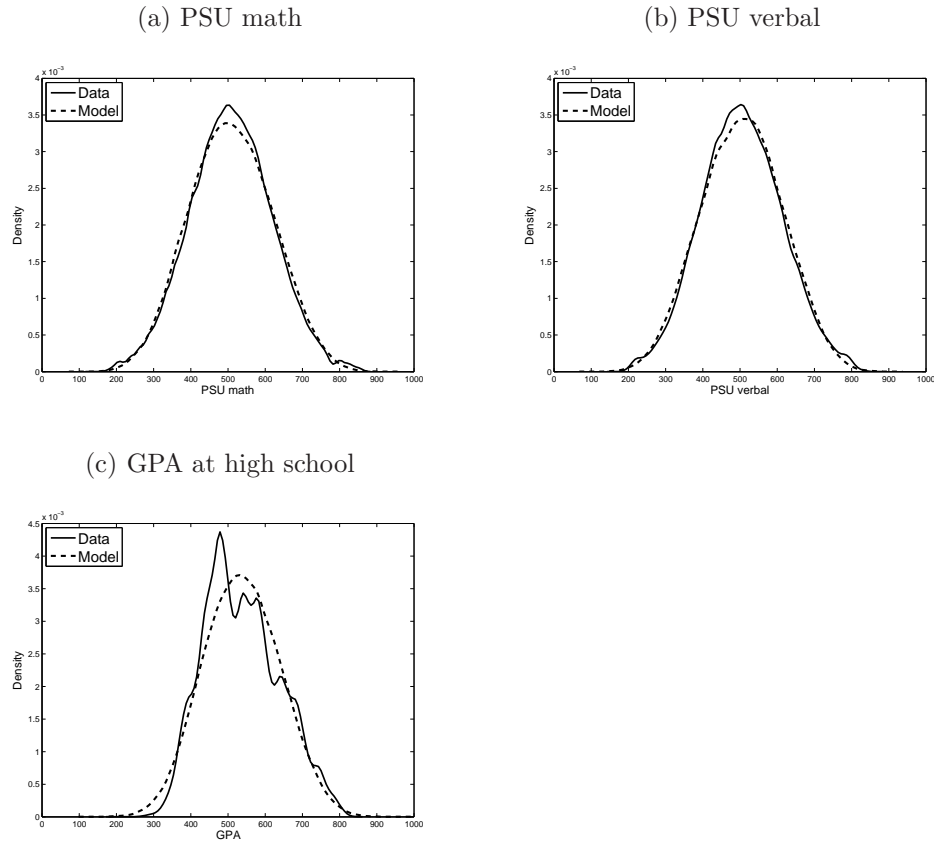


Though the model shows a good fit in all the aspect of the data, given that the goal of this paper is to study how different college admissions policies may affect high school students' behavior, I focus my attention on the model fit for those tests that are relevant in the admissions process, along with the student test decision. Figure 1.3 shows that the model replicates the test distribution observed in the data.⁴⁵ Moreover, in Appendix A.5, Table A.14 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 4.⁴⁶

⁴⁵The 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.

⁴⁶Appendix A.5 contains Figures A.1 and B.4, which show the model fit of the densities for the remaining tests (2004 and 2006), where all of them show good fit.

Figure 1.3: Model fit in tests determining final score



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 A.3 (Appendix A.5) 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).

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 1.3 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.

Table 1.3: 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 parametric)
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

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 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.⁴⁷ 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 theoretical model of effort, since it is due to pure measurement error.⁴⁸

The last column of Table 1.3 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

⁴⁷I also use a reduced form approach for the equation that determines the test taking decision.

⁴⁸The details of this estimation and the resulting parameters are available upon request.

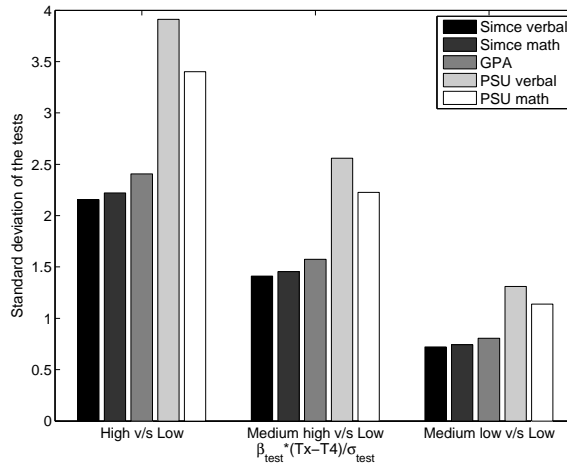
noisiness of the measures of effort.

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.⁴⁹

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 1.4 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).

Figure 1.4: The impact of types on tests



⁴⁹The only exception is the fixed cost parameter.

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:⁵⁰

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

Consequently it is possible to see how these probabilities vary across different groups. In fact, Figure A.4 (Appendix A.5) 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.

1.7 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 socioeconomic 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

⁵⁰These conditional probabilities are used in all the simulations and counterfactual experiments performed in this paper.

weight was increased, which in practice implies that the probability of attending better universities for students from low income high schools is increased.⁵¹ 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.⁵²

The first aspect to review from these experiments is how do they change the universities' socioeconomic composition, which is presented in Figures A.5, A.6 and A.7 (Appendix A.6). 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

⁵¹It is also checked for what happens when it is decreased.

⁵²For 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.

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.⁵³

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 A.8 shows that the SES-quota implementation increases the average effort of high school students by 0.3 standard deviations. Similarly, Figure A.9 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 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.⁵⁴

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 A.10 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

⁵³I 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).

⁵⁴There 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.

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 A.11 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 A.12), 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 A.11 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

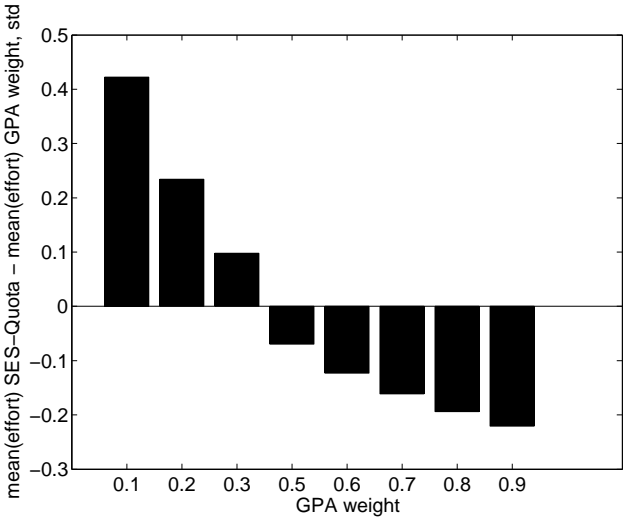
college admissions rules in order to have substantial variations in high school student performance. In this context it is pertinent to ask how relevant this is to model effort.

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 A.13 show that there is an important difference between the optimal effort's final-score cutoffs and the fixed effort's final-score cutoffs,⁵⁵ 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 A.14 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 A.15). 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.

⁵⁵It should be kept in mind that such a counterfactual experiment implies 5 final-score cutoffs per university.

Figure 1.5: Average effort: SES-Quota versus changing GPA's weight

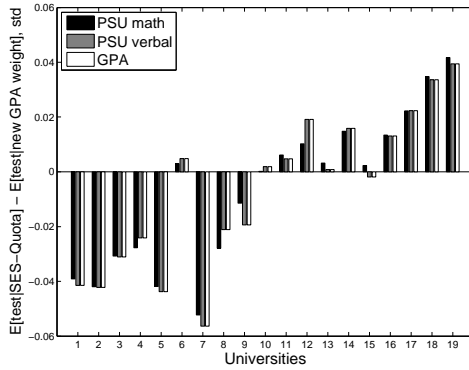


Finally, I discuss which college admissions rule leads to the most efficient student allocation. I first simulate the estimated model for different GPA weights and calculate the resulting socioeconomic composition among universities from each of these exercises. Then, I impose these quotas in the SES-quota system. As a result, I can compare outcomes of the two policy experiments while having the same socioeconomic composition in both cases.

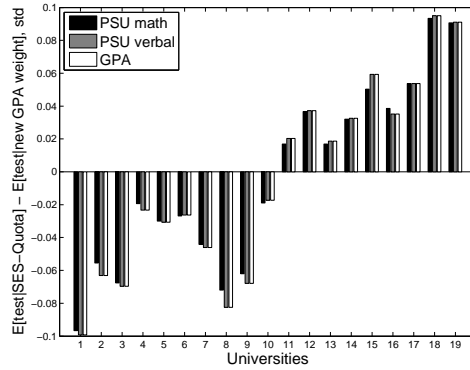
As Figure 1.5 shows, the first point is that changes in the GPA weight imply a higher increase in average effort than for the SES-Quota system. This is mainly because the estimated effort marginal productivity is much higher in the GPA production function than in the production functions of the two PSU tests.

Figure 1.6: Expected tests and GPA: SES-Quota versus changing GPA's weight

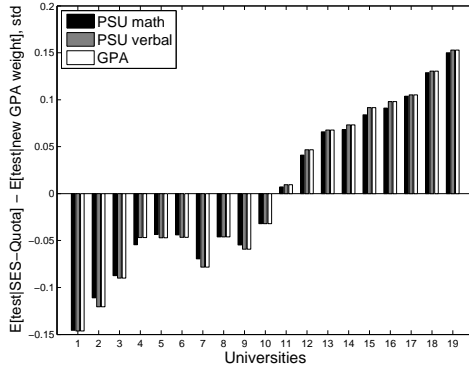
(a) GPA weight = 0.5



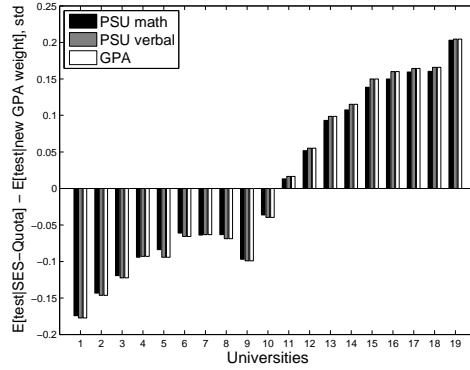
(b) GPA weight = 0.6



(c) GPA weight = 0.7



(d) GPA weight = 0.8



However, this does not mean that changing the GPA weight is the preferred system to achieve equal opportunities. Instead, Figure 1.6 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.

1.8 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.

Chapter 2

A Dynamic Model of Elementary School Choice

2.1 Introduction

A frequent topic in policy debates is what should be the role – if any – of market incentives in education provision. Given that parents' choice is the critical mechanism to increase school quality in a market-oriented educational system, the literature has focused on the extent to which parents consider school quality when they make their decisions, and how this consideration is heterogeneous across parents. To understand parents' school choice, and the potential heterogeneity in their preferences, one must separate the effects of differences in their preferences, in perceptions about quality, and in choice sets. Distinguishing these three elements is a complex task given that, in general, these determinants of parents' choice are not observable.

In this paper, I build and estimate a dynamic model of elementary school choice. To this end, I use detailed Chilean administrative data for the students who entered 1st grade in 2004. As many authors have emphasized (e.g., Gallego and Hernando (2008) and Hsieh and Urquiola (2006)), the Chilean system is probably the most massive school choice program in the world, hence the importance of studying the determinants of school choice in this context.

I model elementary school choice at the end of each academic year, allowing for parental heterogeneity along several dimensions: their ability to understand public information about quality (standardized tests), how much they care about school quality (measured as the school's contribution to standardized test scores), their involvement in the school attended by their child, and their choice set.⁵⁶ By estimating the structural parameters of the model, I am able to assess the empirical relevance of these components in explaining both the observed preference for private over public schools and the unequal access to high quality schools.

In the model, parents care about different characteristics of primary schools, such as the school's socioeconomic composition, quality, religious affiliation, location, type of administration (i.e., public, subsidized private and non-subsidized private), tuition fee, and GPA standard. Parents do not perfectly observe school quality. To estimate the quality of each school, they can access two different sources of information. First, every year they observe the performance of each school on a standardized test, which is made public with a one year lag. Parents can have different levels of misperception in processing this information; because test scores depend on school quality and on the socioeconomic status (SES) of the school, parents can confound these two effects, confusing high quality schools with schools that have higher SES students. Second, parents also differ in their exogenous level of involvement in the schooling process of their child, which implies that those who are involved in their child's school observe the quality of that particular school without misperception.

I estimate the parameters of the model by simulated maximum likelihood, using the Monte Carlo integration and interpolation method (Keane and Wolpin (1994)). To build the database of students, I use the administrative panel data from 2004 to 2011, which includes the school attended by each student in each year, their average grade, the municipality where they live and where the school is located, and some basic demographic information. Because the sample of students entering 1st grade

⁵⁶In Chile, at least in my sample period, schools were allowed to select students based on academic and non-academic characteristics (e.g., parents' marital status).

in 2004 took the SIMCE test in 4th grade (2007) and in 8th grade (2011), I merge this panel with information from parent surveys associated with those rounds of SIMCE administration, including mother's and father's information. To build the database of schools, I use the test scores from the SIMCE test and the information collected from SIMCE parents' surveys for the years 2002, 2005-2011.⁵⁷ Test scores are used to estimate school quality for every year. The surveys include questions about school tuition fees, and information about the elements considered by the school in the admission process. Furthermore, from administrative data of the Ministry of Education, I collect information about schools' religious affiliation, if any.

The results show that parents do care –but in a moderate way– about school quality, that more involved parents care marginally more about school quality, and that parents' decisions are not sensitive to quality after the first decision (1st grade). Moreover, the results also suggest that parents have an important misperception about school quality, which results in a less favorable opinion about the quality of public schools, relative to private schools. This result supports the idea that parents may have difficulties in isolating a school's quality from its socioeconomic composition when they observe test scores. However, given that quality is not very relevant for their decision, such a misperception only partially affects parents' choices.

Regarding the debate about why parents choose private schools over public schools, the results show that, if parents were only concerned about quality, they would choose public schools more often. The same would be true if they did not have a misperception about quality. However, the results suggest that admission rules are binding restrictions and that relaxing them would increase the demand for private schools. The simulations also show that schools' admission rules and household location are both important in explaining the rise in the achievement gap between students from different SES.

⁵⁷4th grade SIMCE for years 2002, 2005-2009 and 8th grade SIMCE for years 2004, 2007, 2009 and 2011.

The paper has three main contributions. First, to the best of my knowledge, this is the first paper that structurally estimates a dynamic model of elementary school choice. The dynamic nature of school choice has particular relevance in the Chilean context, where around 30% of the students switch schools at least once between 1st and 8th grade (excluding those who moved to other municipalities and those who fail at least one year). Second, the structural approach followed in this paper allows me to quantify different causes of unequal access to high quality schools and of the higher demand for private schools than for public schools. Finally, the model considers the difficulties that parents may have in processing and understanding information about school quality, which contributes to the scarce literature on structural estimation with bounded rationality, as well as to the literature that uses observed choices to infer agents' information.

The structure of the paper is as follows: Section 2.2 presents a review of the related literature; Section 2.3 briefly describes the Chilean educational system; Section 2.4 introduces the model; Section 2.5 discusses the data and the procedure to estimate the model; and Section 2.6 presents the results and the analysis of the counterfactual experiments. Section 2.7 concludes.

2.2 Literature Review

This paper is related to several strands in the literature. First of all, it is related to the papers that evaluate the role of competitive market incentives in education provision. On one side, there are theoretical papers, such as Epple and Romano (1998), and McMillan (2004), which debate the potential for those incentives, specifically tuition vouchers, to increase schools' quality and to make significant improvements for poor families.⁵⁸ On the other side, there are empirical studies that show mixed evidence

⁵⁸In a survey of this literature, Epple and Romano (2012) conclude: *Research taking account of distinctive features of the education "market" has shown that early arguments touting the virtues of laissez-faire flat-rate vouchers were overly optimistic. However, the research does not vindicate voucher opponents who use shortcomings of the laissez-faire voucher to justify the wholesale dismissal of vouchers.*

regarding this debate; see, for example, Hsieh and Urquiola (2006), and Angrist, Bettinger, and Kremer (2006).⁵⁹

Authors have studied the determinants of parents' school choice because this is one of the important mechanisms that could explain the shortcomings in the implementation of market-oriented policies in education. In particular, they have studied whether and to what extent parents consider school quality when they make their choice.⁶⁰ For instance, in an interesting paper, Hastings and Weinstein (2008) used a natural experiment and a field experiment that provided direct information on school test scores to lower-income families in a public school choice plan, finding a significant increase in the probability that those families would choose higher-performing schools. In an alternative strategy, several studies have focused on estimating the value that parents place on school quality by calculating how much more people pay for houses located in areas with better schools (e.g., Black (1999) and Kane, Riegg, and Staiger (2006)).

One caveat about this literature is that, in general, it measures school quality using average school test scores. The problem with this approach is that, from the point of view of the parents, this average is not relevant; what is relevant is what their child's performance would be if she were to attend a particular school. Given that sorting is a common feature in education, these two conditional expectations should not coincide. Exceptions to this general problem are shown in Mizala and Urquiola (2013), Neilson (2013), and Rothstein (2006). For instance, Mizala and Urquiola (2013) use a sharp regression discontinuity to estimate the effect that being identified as a SNED winner (a program which seeks to identify effective schools, controlling for schools' SES) has on schools' enrollment, finding no consistent evidence that winning a SNED award affects this outcome.⁶¹

⁵⁹Bettinger (2011) reviews the cases of Chile, Colombia, and Sweden, emphasizing the context-specific nature of the results.

⁶⁰See for example Alderman, Orazem, and Paterno (2001), and Bast and Walberg (2004).

⁶¹Mizala, Romaguera, and Urquiola (2007) present evidence indicating that, in the case of Chile, once we control for the students' socioeconomic status, the remaining part of the test scores are very volatile from year to year. Hence, they argue that producing a meaningful ranking of schools

Several authors study the determinants of school choice in the Chilean context.⁶² For instance, Gallego and Hernando (2008), using a semi-structural approach, find results that suggest that the school choice implemented in Chile increased overall student welfare, but they also find that there is a lot of heterogeneity in the size and even the sign of the welfare change. Along the same lines, Chumacero, Gómez, and Paredes (2011), using a database that accurately estimates the distance between the household and school, find that both quality and distance are highly valued by households. In a recent and novel paper, Neilson (2013) study the effects of targeted school vouchers on the outcomes of poor children in Chile; his findings suggest that this program effectively raised competition in poor neighborhoods, pushing schools to improve their academic quality. Finally, Mark, Elacqua, and Buckley (2006) study how parents construct their school choice sets and comparing this to what they say they are seeking in choosing schools. Their results indicate that parental decisions are influenced by demographics.⁶³

This paper is also related to the literature that models individuals' economic decisions incorporating bounded rationality. In general, this literature follows the idea that, as Simon (1986) points out, cognitive effort is a scarce resource, and the knowledge and computational power of the decision-maker are always limited. In an interesting paper, which is one of the few papers that perform a structural estimation with bounded rationality, Houser, Keane, and McCabe (2004) develop a Bayesian procedure for classification of subjects into decision rule types in choice experiments, finding that, in a very difficult dynamic problem, more than a third of the experimental subjects followed a rule very close to the optimal (expected wealth maximizing) rule.

that may inform parents and policymakers may be harder than is commonly assumed.

⁶²Chile's school choice policies will be described in the next section.

⁶³There are several studies that try to study the effect of the voucher system implementation in Chile. Although the evidence is mixed regarding its effect on school quality, there is more agreement on the negative effect of this policy on student socioeconomic segregation (Auguste and Valenzuela (2006); Gauri (1999); and Hsieh and Urquiola (2006)). In a different approach, Bravo, Mukhopadhyay, and Todd (2010) find that educational vouchers increased educational attainment, high school graduation, college attendance and graduation, and wages.

Because school choice is a complex task, which involves gathering and processing information, different authors have studied the presence of bounded rationality in that context. For instance, Schneider, Teske, Marshall, and Roch (1998) find that, on average, low-income parents have very little accurate information about objective conditions in the schools.⁶⁴ However, even though levels of objective information held by parents are low, their actual choice of schools reflects their preferences in education. Along the same lines, Azmat and Garcia-Montalvo (2012) conclude that, as well as parents' education, information gathering and information processing are important determinants for the quality of school choice.

Finally, this paper is also related to the literature that attempts to infer agents' information using observed choices, such as: Carneiro, Hansen, and Heckman (2003); Cunha, Heckman, and Navarro (2005); and Navarro (2011).

2.3 The Chilean Educational System

In 1981, the Chilean military government created a voucher market in the educational system, which was part of a broader reform that also included the decentralization of public schools (which were transferred to municipalities) and the introduction of flexibility in teachers' contracts.⁶⁵ This reform transformed the way schools were funded by the government, establishing a system where private and public schools were paid per student, with a flat voucher, on the basis of attendance.

Since then, the allocation of public resources has been mainly determined by parents' decisions. However, in practice, this decision has had several restrictions: schools can select students based on their previous performance, tests, and the characteristics of their parents (e.g., marital status and religion). On top of that, since 1994, when a co-payment law was passed, schools that are eligible for public funding can also charge a tuition fee; in that case, depending on the amount charged, there is a

⁶⁴The same is found by Henig (1996).

⁶⁵For a summary of these reforms, see Gauri (1999) and Mizala and Romaguera (2000).

discount to the school's subsidy.⁶⁶

This reform consolidated a system of mixed provision of education, with three types of schools; municipal (public), private subsidized (voucher-private), and entirely private (non voucher-private). The first two receive most of their funds from state vouchers, and, since 1994, privately subsidized schools may additionally charge a tuition fee. In 2013, over 90% of the Chilean students attending public and private schools received funding via vouchers.⁶⁷

In order to guide parents' decisions and to measure the student learning process, a new testing system, SIMCE, came into existence in 1988. The SIMCE is an annual nationwide standardized test. Its results have been public information for more than two decades, publicized in part by listings in major newspapers of individual schools' performance. The government also uses SIMCE scores to allocate resources.⁶⁸

More than 30 years after the reform, there are several clear stylized facts. First, there has been a massive migration from public to private schools. Indeed, the student fraction in the public system went from 78%, in 1981, to 38% in 2012.⁶⁹ Secondly, enrollment in voucher-private schools was accelerated after passage of the co-payment law.⁷⁰ Thirdly, the magnitude of socioeconomic school segregation is very high (and higher than the geographical segregation), and has increased slightly over the last decade (Valenzuela, Bellei, and Ríos (2014)). Finally, despite important increases in the public budget allocated to education, Chile's performance is relatively poor when compared with similar countries (Chumacero, Gómez, and Paredes (2011)).

Another salient feature of the Chilean system, which is consistent with its "free

⁶⁶Epple and Romano (2008) emphasize the consequences of this selection mechanism for the outcomes of an educational voucher system.

⁶⁷Source: Ministry of Education, Chile.

⁶⁸Meckes and Carrasco (2010) describe SIMCE's main features, purposes, institutional framework, and strategies for communicating results.

⁶⁹There is a debate, and mixed evidence, about whether voucher private schools have higher quality than public schools. In a meta-analysis, Drago and Paredes (2011) find that voucher-private schools have a small advantage over public schools. On the contrary, Bellei (2009) finds that voucher-private schools are no more effective than public schools, and that they may be less effective.

⁷⁰See Larrañaga (2004).

choice” design, is that a fairly large number of parents switch schools at some point during primary school. In this regard, Table B.2 of the Appendix B.1 shows that, in any grade, around 4-7% of the parents change their child’s school, and that more educated parents are more likely to do so.⁷¹ Moreover, Table B.1 of the Appendix B.1 shows that more than 30% of parents changed their child’s school at least one time during primary school.⁷²

2.4 The Model

I consider a model in which each family $i \in \{1, 2, \dots, I\}$ decides among their possible elementary school alternatives in each of T (finite) discrete periods of time, where T is the end of the elementary cycle. The educational market is composed of J schools. The parents’ decision is restricted in two ways. First, each parent i , has a specific choice set $\Lambda_i \subseteq \{1, 2, \dots, J\}$. The cardinality of Λ_i is denoted by $S(\Lambda_i)$. Second, each school $j \in \Lambda_i$ may or may not admit the student i based on a rule that will be described below.

Parents’ Utility

Let $D_{it} \in \Lambda_i$ be the school chosen by parent i at time t . The flow utility of parents i when their child is attending school j at time t is given by:

$$u_{ijt} = \beta_{ki}K_{ijt} + \beta_y\mathbf{Y}_{jt} + \beta_z\mathbf{Z}_{ij} + \beta_gG_{ijt} + C 1(D_{it-1} \neq j) + \beta_{eg}[G_{ijt-1} - \widehat{G}_{ijt-1}] \\ + 1(D_{it-1} = j) + \epsilon_{ijt}^u$$

where K_{ijt} is the knowledge achieved by student i in school j at time t (j can be different across years), \mathbf{Y}_{jt} is a vector of characteristics of the school j (e.g.

⁷¹These figures do not include parents who change the municipality where they live, or students who repeat a grade. If one considers those cases, this fraction rises to around 11% (Zamora (2011)).

⁷²These levels of student mobility are similar to what is observed in other countries. For instance, Hanushek, Kain, and Rivkin (2004) show that, in Texas’ public schools, one-third of all children switch schools at least once between grades 4 and 7, excluding changes due to the transition from elementary to middle school.

socioeconomic composition and price), \mathbf{Z}_{ij} is a measure of the distance between the attributes of school j and the tastes of individual i (e.g., religion and location), G_{ijt} is the GPA obtained by the student, C is the direct cost of changing school, $1(A)$ is a function that takes 1 when A is true, \widehat{G}_{ijt-1} is the expected GPA given the information at $t - 2$, and ϵ_{ijt}^u is an iid shock.

In the final period, there is a utility that also captures all future payoffs, such that

$$u_{ijT} = \beta_{Ki}^T K_{ijT} + \beta_y \mathbf{Y}_{jT} + \beta_z \mathbf{Z}_{ij} + C \mathbf{1}(D_{iT-1} \neq j) + \beta_{eg} [G_{ijT-1} - \widehat{G}_{ijT-1}] \\ \mathbf{1}(D_{iT-1} = j) + \beta_s^T SEC_j + \beta_{ag}^T \overline{G}_{ijT} + \beta_{ta}^T TA_{iT} + \epsilon_{ijT}^u,$$

where TA_{it} represents the time, in years, that student i has been attending the current school and SEC_j takes one when school j also offers secondary level grades (from 9th to 12th) and zero otherwise. Including the latter in the terminal utility captures the changing costs that parents are forced to incur in $T + 1$ when their child attends a school that does not offer the secondary class level. Furthermore, \overline{G}_{ijT} and TA_{iT} are included because, as will be noted, they determine the future chances of being admitted in the desired high school.

Student knowledge

I model student knowledge as a cumulative process. In particular, let q_{jt} be the quality of school j at time t and q_{ijt} the quality of the school attended by student i at time t , such that $q_{ijt} = \sum_{j=1}^J q_{jt} \mathbf{1}(D_{it} = j)$; thus, the learning process is given by:

$$K_{i0} = \alpha_0 X_i, \\ K_{ijt} = K_{ijt-1} + \alpha_1 q_{ijt}.$$

Therefore, the knowledge achieved by student i in school j at time t , K_{ijt} , is a function of student i 's previous knowledge and the quality of the school she attends that

year, q_{ijt} . In addition, the initial knowledge only depends on student i 's characteristics X_i (i.e., parents' education).

GPA function

Grades in elementary school are determined by the following production function:

$$G_{ijt} = \lambda_{0jt}^t + \lambda_{1jt}^t K_{ijt} + \lambda_2^t TA_{it} + \varepsilon_{ijt}^g,$$

This specification captures the idea that each school may have a particular way to map knowledge onto grades. In particular, the higher the value of λ_{0jt} , the more likely it is that students perform well in school j . Moreover, even conditioning on student knowledge, TA_{it} has an effect on grades. This accounts for the fact that it may take time for new students to learn the characteristics of the evaluation system of each school.

Probability of admittance

Parents are restricted in their choices to the extent that schools have the right of admittance. Let AD_{ijt} be a binary variable, which is unobservable for the econometrician, that equals one if student i can enter school j at time t and zero otherwise, such that:

$$AD_{ijt} = \begin{cases} 1 & \text{if } \varrho_{ijt} - \varepsilon_{ijt}^{ad} \geq 0 \\ 0 & \text{if } \varrho_{ijt} - \varepsilon_{ijt}^{ad} < 0 \end{cases}$$

where

$$\begin{aligned} \varrho_{ijt} = & \varphi_0 + \varphi_q q_{jt} Sel_{jt} + \varphi_{0k} Sel_{jt}^k + \varphi_{1k} K_{it-1} Sel_{jt}^k + \varphi_{0g} Sel_{jt}^g + \varphi_{1g} \bar{G}_{it-1} Sel_{jt}^g \\ & + \varphi_{0mr} Sel_{jt}^m + \varphi_{1mr} MR_{ijt-1} Sel_{jt}^m + \varphi_{0r} Sel_{jt}^r + \varphi_{1r} REL_i Sel_{jt}^r + \varphi_s Sel_{jt}^o \\ & + \varphi_{0ns} New_{jt} + \varphi_{1ns} New_{jt} * Size_{jt} + \varphi_{fe} X_i * fee_{jt} + \varphi_{sx} X_i * Sel_{jt}. \end{aligned}$$

and Sel_{jt}^k takes one when school j selects students based on academic tests and zero otherwise; Sel_{jt}^g takes one when school j selects students based on previous grades and zero otherwise; Sel_{jt}^r takes one when school j selects students based on students' religion and zero otherwise; takes one when school j selects students based on parents' marital status and zero otherwise Sel_{jt}^m ; Sel_{jt}^o takes one when school j selects students based on other reasons and zero otherwise.⁷³ These variables may all equal one at the same time.⁷⁴ New_{jt} takes one when the school is new (or doesn't offer the previous grades), $Size_{jt}$ is the size of this new school, and fee_{jt} denotes the tuition fee. Finally, MR_{ijt} takes one if parents are married and zero otherwise, and REL_i takes one if parents are religious and zero otherwise.⁷⁵

Then, assuming that ε_{ijt}^{ad} are iid, following a logistic distribution, the probability of admission is described by:

$$Pr(AD_{ijt} = 1) = \frac{\exp(\varrho_{ijt})}{1 + \exp(\varrho_{ijt})}$$

To have a tractable likelihood calculation, I assume that ε_{ijt}^{ad} is realized before parents make the D_{it} decision.⁷⁶ Moreover, to simplify the solution of the model, I assume that, for any student, there is always at least one school willing to admit her. In particular,

$$h \in \operatorname{argmax}_{j \in \Lambda_i}(\varrho_{ijt}) \Rightarrow AD_{iht} = 1.$$

I chose this specification for the random process of AD_{ijt} for two reasons: given the rich information that I have about the admission rules of each school it seems

⁷³In the empirical implementation, all these variables are proportions, instead of binary variables. This is because I construct these variables from parents' surveys, and in each school their answers are not always the same. Thus, for instance, in the empirical implementation, Sel_{jt}^k is the fraction of parents in school j who affirm that school j selects students based on an academic test.

⁷⁴ Sel_{jt} is an index to measure how selective is school j at time t . In the empirical implementation of this model, $Sel_{jt} = (Sel_{jt}^k + Sel_{jt}^g + Sel_{jt}^o) * \frac{1}{3}$.

⁷⁵Contreras, Sepulveda, and Bustos (2010) present evidence indicating that student selection is a widespread practice among private subsidized schools.

⁷⁶This means parents do not apply to schools. Instead, at the end of each period they know their feasible set for the next period and they pick the feasible school that maximizes their expected utility.

reasonable to, and considering the challenge in separately identifying the parameters of this process and the parameters of the utility function when AD_{ijt} is latent. In this regard, it should be noted that there is no variable that enters in the same way in the admission probability function and in the utility function.⁷⁷

Moreover, this specification takes advantage of the interaction between the features of the model and data availability. For instance, if the school selects students based on an academic test (with $\varphi_{1k} > 0$, as expected), then the higher K_{it-1} , the higher the probability of i being admitted at j .

Parents' information and perception about quality

Parents have two sources of information about school quality. Firstly, they observe the results of the standardized tests for all the schools, which is public information. Secondly, they may observe the quality of the school which their child is attending, which is private information.

Regarding public information, it is assumed that standardized tests are measures of school quality, whose values also depend on the characteristics of the student. Thus, in this model, school quality is defined as a school's contribution to learning (i.e., value added), such that:

$$ST_{ijt}^m = q_{jt} + \theta_2^m X_i + \varepsilon_{ijt}^m,$$

$$ST_{ijt}^s = \theta_0^s + \theta_1^s q_{jt} + \theta_2^s X_i + \varepsilon_{ijt}^s,$$

$$ST_{ijt}^n = \theta_0^n + \theta_1^n q_{jt} + \theta_2^n X_i + \varepsilon_{ijt}^n,$$

$$ST_{ijt}^{sc} = \theta_0^{sc} + \theta_1^{sc} q_{jt} + \theta_2^{sc} X_i + \varepsilon_{ijt}^{sc},$$

where ST_{ijt}^m denotes the math test score, where s is for Spanish, n for natural science, and sc for social science. I define \tilde{q}_{jt} as the estimation of the expected quality of school j at time t , given the public information \mathbf{ST}_{jt} , such that: $\tilde{q}_{jt} = E[q_{jt} | \widehat{\mathbf{ST}}_{jt}, X]$.

⁷⁷The few variables that are in both functions are interacting with other variables in the function that determine the probability of admission.

The model is flexible in terms of how parents access and understand the information about schools' quality. In the first place, there are different types of parents with regard to their ability to distinguish the school's contribution from the students' contribution to test scores. In the second place, there are different types of parents with regard to their involvement in the schooling process of their child, which determines whether they observe the quality of that school. Namely, only involved parents have access to private information. The first is denoted *parent's cognitive skill*, whereas the second is denoted *parent's school involvement*. The school involvement type of parent i is given by $\psi_i \in \{0, 1\}$, where 1 means involved.⁷⁸

To present how parents access and understand the information about the quality of schools, I divide the analysis into three cases: (1) their perception about the quality of the schools not attended by their child; (2) the involved parents' perception about the quality of the school attended by their child; and (3) the non-involved parent's perception about the quality of the school attended by their child. In all three cases, what matters is parents' perception at the end of $t-1$ (when they make the choice of school for period t), about school quality at time t , given their information at time $t-1$, i.e. $E_{t-1}[q_{jt}|D_{it-1}, \psi_i]$.⁷⁹

Case 1: Schools not attended by their child.

$$E_{t-1}[q_{jt}|D_{it-1} \neq j] = \tilde{q}_{jt-2} + \sum_{\chi \in A} \eta_{\chi}^i (\overline{ST}_{jt-2}^{\chi} - \theta_{0,t-2}^{\chi} - \theta_{1,t-2}^{\chi} q_{jt-2}),$$

where, $\eta_{\chi}^i = \eta_{1,\chi}^1 X_i + \eta_{2,\chi} S(\Lambda_i)$ and $A = \{m, s, n, sc\}$. Thus, η depends on parents' education and the size of the choice set ($S(\Lambda_i)$), where the latter is motivated by the bounded rationality literature.⁸⁰

In this case, if – as expected – $\eta_{\chi}^i \geq 0$, then parents will overestimate the quality for

⁷⁸This is an exogenous, time invariant, parents' characteristic and therefore it does not depend on the school's characteristic.

⁷⁹In all these cases, they use their current estimation of quality as their prediction for future values.

⁸⁰A survey in Conlisk (1996).

schools whose students have, on average, highly educated parents.⁸¹ Moreover, given the fact that in the Chilean educational system standardized tests are published one year after taken, even if parents did not have a misperception about quality (i.e., $\eta = 0$), when they choose the school for time t (at the end of $t - 1$), they would estimate school qualities using the public information at $t - 2$.

In sum, the use of public information presents two potential drawbacks: it is published after a delay, and parents may have difficulties in interpreting it, namely, when they observe the tests, they can have problems in isolating school quality from the socioeconomic composition of its students.

Case 2: School attended by their child, when parents are involved in that school ($\psi_i = 1$).⁸²

$$E_{t-1}[q_{jt}|D_{it-1} = j, \psi_i = 1] = q_{jt-1}$$

Thus, parents who are involved in their child's school observe the quality of that school without distortion and without lag.⁸³

Case 3: School attended by their child, when parents are not involved in that school ($\psi_i = 0$).

$$E_{t-1}[q_{jt}|D_{it-1} = j, \psi_i = 0] = \tilde{q}_{jt-2} + \sum_{\chi \in A} \eta_{\chi}^i (\overline{ST}_{jt-2}^{\chi} - \theta_{0,t-2}^{\chi} - \theta_{1,t-2}^{\chi} \tilde{q}_{jt-2}),$$

$$A = \{m, s, n, sc\}.$$

Thus, parents who are not involved in their child's school have, for that school, the same information that they have for all the other schools (public information).

⁸¹Given the functional form of the standardized tests, $\overline{ST}_{jt-2}^{\chi} - \theta_{0,t-2}^{\chi} - \theta_{1,t-2}^{\chi} \tilde{q}_{jt-2}$ is the part of the average test, of subject χ , that is not explained by school quality. Hence, this is the part explained by the socioeconomic composition of the school.

⁸²I allow for β_{ki} being different for this type of parent.

⁸³This assumption is supported by the evidence presented in Azmat and Garcia-Montalvo (2012), who find that knowing about and/or visiting more schools is related to more accurately assessing local schools.

In this context, parents' perception about their child's knowledge is given by the following expressions:

- $E_0[K_{i0}|\psi_i] = K_{i0}$.
- $E_t[K_{it}|D_{it}, \psi_i] = E_t[K_{it-1}|D_{it}, \psi_i] + \alpha_1 \sum_{j=1}^J E_t[q_{jt}|D_{it}, \psi_i]1(D_{it} = j)$.
- $E_{t-1}[K_{it}|D_{it-1}, \psi_i] = E_{t-1}[K_{it-1}|D_{it-1}, \psi_i] + \alpha_1 \sum_{j=1}^J E_{t-1}[q_{jt}|D_{it-1}, \psi_i]1(D_{it} = j)$.

I denote $\tilde{K}_{it}^a = E_a[K_{it}|D_{ia}, \psi_i]$, $a = \{t-1, t\}$.

Decision Timing and Solution of the Model

At the end of period $t-1$, the following random variables are realized: (1) Utility idiosyncratic shocks: $\epsilon_{ijt}^u \forall i, j$; (2) the right of admittance shocks: ϵ_{ijt}^{ad} (hence, AD_{ijt}) $\forall i, j$; and (3) test scores, published with lag: $\{\mathbf{ST}_{t-2}^m, \mathbf{ST}_{t-2}^l, \mathbf{ST}_{t-2}^n, \mathbf{ST}_{t-2}^{sc}\}$. Given this information, parents decide D_{it} , taking into consideration the expected flow utility at t and the expected future payoff associated with each school.⁸⁴

The model is solved by backward recursion, where the dynamic decision is driven by the state variables (Ω_{it}) .⁸⁵

$$\Omega_{it} = \begin{cases} \{D_{it}, TA_{it}, \tilde{K}_{ijt}^t, \bar{G}_{ijt}, q_{ijt}, \mathbf{ST}_{t-1}, \epsilon_{it}^{ad}, \epsilon_{it}^u\} & \text{if } \psi_i = 1 \\ \{D_{it}, TA_{it}, \tilde{K}_{ijt}^t, \bar{G}_{ijt}, \mathbf{ST}_{t-1}, \epsilon_{it}^{ad}, \epsilon_{it}^u\} & \text{if } \psi_i = 0 \end{cases}$$

I define Ω_{it}^- as the state variables which are observed by the econometrician, such that:⁸⁶

$$\Omega_{it}^- = \begin{cases} \{D_{it}, TA_{it}, \tilde{K}_{ijt}^t, \bar{G}_{it}, q_{ijt}, \mathbf{ST}_{t-1}\} & \text{if } \psi_i = 1 \\ \{D_{it}, TA_{it}, \tilde{K}_{ijt}^t, \bar{G}_{it}, \mathbf{ST}_{t-1}\} & \text{if } \psi_i = 0 \end{cases}$$

⁸⁴ ϵ_{ijt}^g is realized after the decision of D_{it} is made.

⁸⁵ $TA_{it} = 1 + 1(D_{it} = D_{it-1})TA_{it-1}$ and $\bar{G}_{ijt} = \frac{\bar{G}_{ijt-1}*(t-1) + G_{ijt}}{t}$.

⁸⁶ Observed conditional on types.

To consider the school's right of admittance, I redefine the flow utility as:

$$\bar{u}_{ijt} = \tilde{u}_{ijt}(\epsilon_{ijt}^{ad}) + \epsilon_{ijt}^u$$

where,

$$\tilde{u}_{ijt}(\epsilon_{ijt}^{ad}) = \begin{cases} \beta_k K_{ijt} + \beta_x \mathbf{Y}_{jt} + \beta_z \mathbf{Z}_{ij} \\ + \beta_g G_{ijt} + C1(D_{it-1} \neq j) \\ + \beta_{eg}[G_{ijt-1} - \hat{G}_{ijt-1}]1(D_{it-1} = j) & \text{if } AD_{ijt}(\Omega_{it-1}^-, \epsilon_{ijt}^{ad}) = 1 \\ -\infty & \text{if } AD_{ijt}(\Omega_{it-1}^-, \epsilon_{ijt}^{ad}) = 0 \end{cases}$$

The solution to this dynamic problem is fully characterized by the integrated value function, $\bar{V}(\Omega_{it-1}^-)$, such that:⁸⁷

$$\bar{V}(\Omega_{it-1}^-) = \int \max_{j \in \Lambda_i} \{E_{t-1} \tilde{u}_{ijt}(\epsilon_{ijt}^{ad}) + \epsilon_{ijt}^u + \delta E_{t-1} [\bar{V}(\Omega_{it}^-) | \Omega_{it-1}^-, D_{it} = j]\} dG_\varepsilon(\varepsilon_{it}),$$

$$\varepsilon_{it} = [\epsilon_{it}^u \ \epsilon_{it}^{ad}]'$$

Then, defining the auxiliary function $v(\Omega_{it-1}, D_{it} = h)$ as:

$$v(\Omega_{it-1}, D_{it} = h) = E_{t-1} \tilde{u}_{iht}(\epsilon_{it}^{ad}) + \epsilon_{iht}^u + \delta E_{t-1} [\bar{V}(\Omega_{it}^-) | \Omega_{it-1}^-, D_{it} = h]$$

$$\Rightarrow D_{it} \in \operatorname{argmax}_{j \in \Lambda_i} \{v(\Omega_{it-1}, D_{it} = j)\} \quad \forall t \in \{1, 2, \dots, T-1\}$$

At the end of $T-1$:

⁸⁷ $\epsilon_{it}^u = \{\epsilon_{ijt}^u\}_{j \in \Lambda_i}$ and $\epsilon_{it}^{ad} = \{\epsilon_{ijt}^{ad}\}_{j \in \Lambda_i}$.

$$\tilde{u}_{ijT}(\epsilon_{ijT}^{ad}) = \begin{cases} \beta_K K_{ijT} + \beta_x \mathbf{Y}_{jT} + \beta_z \mathbf{Z}_{ij} + \beta_{ta}^T T A_{iT} \\ + \beta_{eg} [G_{ijT-1} - \bar{G}_{ijT-1}] 1(D_{iT-1} = j) \\ + C 1(D_{iT-1} \neq j) + \beta_s^T SEC_j + \beta_{ag}^T \bar{G}_{ijT} & \text{if } AD_{ijt}(\Omega_{iT-1}^-, \epsilon_{ijT}^{ad}) = 1 \\ -\infty & \text{if } AD_{ijt}(\Omega_{iT-1}^-, \epsilon_{ijT}^{ad}) = 0 \end{cases}$$

$$v(\Omega_{iT-1}, D_{iT} = j) = E_{T-1} \tilde{u}_{ijT}(\epsilon_{iT}^{ad}) + \epsilon_{ijT}^u,$$

$$\Rightarrow D_{iT} \in \operatorname{argmax}_{j \in \Lambda_i} \{v(\Omega_{iT-1}, D_{iT} = j)\}.$$

2.5 Data and Empirical Implementation

Data Description

The main source of information in this paper is the administrative panel data from 2004 to 2011 on all students in the country from the Ministry of Education of the government of Chile. This panel includes the school attended every year, the average grade, the municipality where the student lives and where the school is located, and some basic demographic information. As mentioned, the sample of students entering 1st grade in 2004 took the SIMCE test in 4th grade (2007) and 8th grade (2011), and I merge this panel with information from parent surveys that are carried out during the SIMCE process. These contain mother's and father's education, whether they care about school religion, and their marital status.

In order to characterize schools, I use SIMCE test scores and the information collected from SIMCE parents' surveys for the years 2002, 2005-2011.⁸⁸ Test scores are used to estimate schools' quality for every year. The surveys include questions about

⁸⁸4th grade SIMCE for years 2002, 2005-2009 and 8th grade SIMCE for years 2004, 2007, 2009 and 2011.

school tuition fees, and whether the school considered some of the following elements in the admission process: a student test, previous GPA, parents' marital status (and whether they had a religious wedding), and a general category to account for any other information considered in the admission. Furthermore, from administrative data of the Ministry of Education, I collect information about each school's religious affiliation.

Finally, from the SIMCE of 2011, 8th grade for my cohort, I use the answers to two types of questions as determinants of parent involvement. First, I use the questions to parents:

1. *How often do you attend the periodic parents' meeting of your child's class?*
2. *Name the first three reasons why you chose your child's current school.*

Second, I use the questions to students, *How often does one of your parents do each of the following activities?:*

1. *She or he explains to me the class material that I don't understand.*
2. *She or he helps me to study.*

Empirical Implementation

Two inputs are needed to estimate the model, namely, the measures of school quality and parents' choice set. Moreover, to gain in speed, and given the detailed information that I have, I estimate the parameters of the knowledge production function and the parameters of the grade production function outside of the model.

Estimating Measures of Quality

As presented above, the observable test scores have the following functional form:

$$ST_{ijt}^{\chi} = \theta_0^{\chi} + \theta_1^s q_{jt} + \theta_2^{\chi} X_i + \varepsilon_{ijt}^{\chi}, \quad \chi \in \{m, s, n, sc\}.$$

I estimate the factor loadings (θ) and the distribution of q_{jt} by EM algorithm, assuming that the latter follows a mixture of normal distributions:⁸⁹

$$p(q_{jt}) = \sum_{l=1}^3 \pi_l N(\mu_{jt}^l, \sigma_{jt}^l).$$

Then, $E[q_{jt}|\mathbf{ST}_{jt}, X]$ is estimated as $\sum_{l=1}^3 \hat{\pi}_l \hat{\mu}_{jt}^l$.⁹⁰ Figure B.2 (Appendix B.2), shows the distribution of estimated school quality by school type in 2003, which is consistent with Bellei (2009), in the sense that, when one does not control for peer effects, voucher-private schools have higher quality than public schools. Because I want to understand parents' decisions and to what extent they base such decisions on school quality, it makes sense to consider peer effects as part of the definition of school quality.⁹¹

To estimate the parameters of the knowledge production function, I run the following OLS regression:⁹²

$$\tilde{K}_{ijT} = \alpha_0 X_i + \alpha_1 \sum_{t=1}^T \tilde{q}_{ijt} + \vartheta_{ijT}$$

Parents' Choice Set

Given that, in principle, parents may choose any school in the country, it is a hard empirical problem to define the choice set Λ_{it} . To do so, I classify families in G groups, grouped by their home location (municipality) and their level of education, then:

⁸⁹In this estimation I only use students who have attended the same school during the first four years of primary school. X_i is a vector of parents education. In the estimation, I allow that μ_{jt}^l and σ_{jt}^l depend on school characteristic.

⁹⁰Test scores are available in 4th grade at the elementary school level, and in 8th and 10th grade in alternating years. This precludes including student fixed effects to estimate the school quality in every year of my sample.

⁹¹However, this should be kept in mind in the analysis of the counterfactual experiments.

⁹²Where, in a similar fashion as in the estimation of $E[q_{jt}|\mathbf{ST}_{jt}]$, K_{ijT} is estimated by EM algorithm, assuming that K_{ijT} is a latent variable measured by the SIMCE tests at time T. Further, $\tilde{q}_{ijt} = \sum_{l=1}^3 \hat{\pi}_l \hat{\mu}_{ijt}^l$.

$$G_g = \{i, \text{ s.t. } (ed_i, loc_i) = (ed_g, loc_g)\},$$

$$g(i) \Leftrightarrow i \in G_g.$$

$$\Lambda_{it} = \{j, \text{ s.t. } \exists i' \in G_{g(i)} \mid \sum_{t=0}^T 1(D_{i't} = j) > 0\}$$

This means that, by definition, for each pair of parents, the chosen school belongs to their choice set.

Having a large number of families belonging to each group implies that, if no family belonging to group G_g has chosen a particular school, it is because that school is not feasible for that group of families. Figure B.1 (Appendix B.2) shows the distribution of the size of parents' choice set, which indicates that, if anything, this approach is overestimating that size.

Estimating the Parameters of the Grade Production Function

To estimate the parameters in the grade production function, λ_{0jt} and $\lambda_{1j} \forall j, t$, I use the math test to replace K_{ijt} by $ST_{ijt}^m - \varepsilon_{ijt}^m$ in the grade production function, such that:

$$G_{ijt} = \lambda_{0jt} + \lambda_{1jt}ST_{ijt}^m + \lambda_{2t}TA_{ijt} - \lambda_{1j}\varepsilon_{ijt}^m + \varepsilon_{ijt}^g$$

Then, I estimate the parameters of interest by Two Stage Least Squares, using ST_{ijt}^l , ST_{ijt}^n and ST_{ijt}^c as instruments of ST_{ijt}^m .

Estimating the Parameters of the Utility Function

The specification of the utility function includes (in vector \mathbf{Y}_j) school socioeconomic composition dummies, tuition fee, school type dummies (public, voucher-private,

and non voucher-private);⁹³ (in vector \mathbf{Z}_j) a dummy variable that takes one if both the school and parents are religious and zero otherwise, and a dummy variable that takes one if parents live in the municipality where the school is located and zero otherwise.

I estimate the parameters of the utility function by simulated maximum likelihood, using the Monte Carlo integration and interpolation method (Keane and Wolpin (1994)).⁹⁴ Given that this –and any method that solves the dynamic problem in each parameter iteration– is time consuming, I select a sample in the following way: I sort the municipalities belonging to Santiago City in descending order, in terms of their total student population, and I use the students living in the municipalities ranked 1, 3, 5 (odd numbers) ..., 19.⁹⁵ As a result, the final sample for the estimation has 9,752 families and 856 schools.⁹⁶

Given the solution to the dynamic problem, which is fully characterized by the integrated value function $\bar{V}(\Omega_{it-1}^-)$, and assuming that ϵ_{ijt}^u is iid, following a standard type-1 extreme value distribution, then:

$$P(D_{it} = h | \epsilon_{it}^{ad}, \psi_i) = \frac{\exp(E_{t-1}[\tilde{u}_{iht}(\epsilon_{iht}^{ad}) + \delta \bar{V}(\Omega_{it}^-) | \Omega_{it-1}^-, D_{it} = h, \psi_i]) AD_{iht}(\Omega_{it-1}^-, \epsilon_{iht}^{ad})}{\sum_{j \in \Lambda_i} \exp(E_{t-1}[\tilde{u}_{ijt}(\epsilon_{ijt}^{ad}) + \delta \bar{V}(\Omega_{it}^-) | \Omega_{it-1}^-, D_{it} = j, \psi_i]) AD_{ijt}(\Omega_{it-1}^-, \epsilon_{ijt}^{ad})}.$$

Therefore, the probability of a sequence of schools chosen by parents i , \mathbf{D}_i , is given by:⁹⁷

⁹³Because the public system has two types of schools, one from 1st to 6th and the other one from 7th to 12th, in the case of public schools, I allow for a different dummy for each type.

⁹⁴The discount parameter δ is not estimated, but it is assumed equal to 0.95.

⁹⁵The considered municipalities are Estación Central, Huechuraba, La Granja, La Reina, Macul, Melipilla, Pedro Aguirre Cerda, Recoleta, San Miguel, and Ñuñoa. I used 10 of the 33 municipalities of Santiago city.

⁹⁶I also drop the students who fail a year and those who change the municipalities where they live. The former is because I use the student information collected in the 2011 SIMCE (8th grade), information that is obviously missing for those who enter 1st grade in 2004 and fail at least one year between 2004 and 2010. The latter is because the dynamic problem is solved for each student type, where a type is defined by the student location, among other things. Therefore, if I considered people who change their location, I would have to solve the dynamic problem for all the combinations of locations observed in the data, which would dramatically increase the estimation time.

⁹⁷ $\pi_n^i = P(\psi_i = n | X_i)$.

$$P(\mathbf{D}_i) = \prod_{t=1}^T P(D_{it}) = \sum_{n \in \{0,1\}} \pi_n^i \int \sum_{t=1}^T P(D_{it} | \epsilon_{it}^{ad}, \psi_i = n) dG_\epsilon \epsilon_{it}^{ad}.$$

where the log-likelihood function L , is given by:⁹⁸

$$L = \sum_{i=1}^I \log(P(\mathbf{D}_i)).$$

Given that AD_{ijt} is a latent variable, I approximate $P(D_{it})$ by:

$$P(D_{it} = h) \approx \sum_{n \in \{0,1\}} \pi_n^i \frac{1}{N_s} \sum_{\kappa=1}^{N_s} \frac{\exp(E_{t-1}[\tilde{u}_{iht}(\epsilon_{iht}^{ad}) + \delta \bar{V}(\Omega_{it}^-) | \Omega_{it-1}^-, D_{it} = h, \psi_i = n]) AD_{iht}^\kappa}{\sum_{j \in \Lambda_i} \exp(E_{t-1}[\tilde{u}_{ijt}(\epsilon_{ijt}^{ad}) + \delta \bar{V}(\Omega_{it}^-) | \Omega_{it-1}^-, D_{it} = j, \psi_i = n]) AD_{ijt}^\kappa}.$$

where N_s is the number of simulations and the values of AD_{ijt}^κ are drawn from G_ϵ .⁹⁹

Identification

The identification of this model faces two challenges not commonly present in any standard discrete choice dynamic programming models of individual behavior.¹⁰⁰ First, there is a challenge in separately identifying the parameters of the utility function and the parameters of the admission probabilities, without observing parents' applications. Second, there is a challenge in identifying the parameters that determine parents' perception about quality (i.e., η).

The former challenge is overcome through exclusion restrictions, which are naturally developed given the available data and the features of the model. Specifically, there is no variable (nor interaction of variables) that is simultaneously present in the utility function and in the admission probability function. For instance, parents care about

⁹⁸Because each likelihood calculation takes around 5 minutes, I use HOPSPACK (Hybrid Optimization Parallel Search PACKage) to optimize the likelihood function. This program is a derivative-free optimization solver.

⁹⁹In the estimation, I consider 50 simulations for each individual-time data point.

¹⁰⁰For a survey, see Aguirregabiria and Mira (2010), Eckstein and Wolpin (1989) or Rust (1994).

school quality through its effect on student knowledge, a variable that also enters in the admission probability function. However, the admission probability is affected by student knowledge only for schools that select students based on academic tests. Furthermore, given the limitation that the model imposes on the heterogeneity of parents' preferences for quality, the fact that parents do not choose some schools that would give them higher utilities is rationalized by the model as if those schools did not admit such a student.¹⁰¹

The intuition behind the solution for the latter challenge is the following: if those parents who choose schools, not for the estimated quality, but for their average test scores (which is also determined by the socioeconomic composition of the school) have a higher probability of belonging to a particular education group or live in higher proportions in municipalities with a particular pattern in terms of the choice set size, then one can use those correlations to identify the parameters that determine η . Technically speaking, given the fact that the socioeconomic composition of schools enters directly in the utility function, the parameters of η_i are identified given the variation that comes from the interaction – in the utility function – of the socioeconomic composition of the school and the socioeconomic composition of parents i .¹⁰²

2.6 Results

In the Appendix B.2, I show the estimated parameters of the knowledge production function and the production function of grades. In short, almost all the signs are as expected and the magnitudes are, in around two third of the cases, statistically significant. An interesting result, presented in Table B.5, is that, even controlling for student knowledge, the number of years a student stayed in a particular school

¹⁰¹This is what identified the constant of the admission probability function. Another possible approach could be the one developed by Geyer and Sieg (2013).

¹⁰²The difference between test scores and quality is, on average, equal to the contribution of the school's SES to test scores.

positively impacts grades which, in practice, constitutes a heterogeneous switching cost. Moreover, Figure B.3 shows how schools have different standards by which they evaluate their students.¹⁰³

The estimation of the parameters of the utility function are shown in Appendix B.2 (Tables B.6-B.9), where most of them have the expected sign. A noteworthy outcome of this estimation is that parents who are involved in their child's school care more about school quality, i.e., they have a higher parameter β_k . Yet, given the non-linear relationship between the parameters and parents' decision, the best way to assess the relevance of parameter magnitudes is through model fit analysis and counterfactual experiments.

Model Fit

I present the fit of the model under two scenarios. In the first case, I consider the sample used for the estimation. In the second case, whose figures are presented in Appendix B.3, I use the complete sample (all the students in Santiago City).¹⁰⁴ I discuss the fit of the model under these two scenarios, since these are also the two samples that I consider in the counterfactual experiments, and because showing that the fit is similar in the two cases reinforces the point that the model is capturing the main mechanisms that determine parent decision, without overfitting the data.

As Figure 2.1 shows, the simulation of the model overall fits the pattern of the students who switch school by grade. However, the model has difficulty in generating the increase in school switching that occurs at the end of 6th grade. This increase is mainly driven by the entry of new public schools in 7th grade (for the new cycle), something that the model can only partially generate. Moreover, as Figure 2.2 shows, the model does a good job of predicting the 8-year total school changes, by parents

¹⁰³In this context, an easy school is one where the constant is big and the slope is small, hence all the students have good grades and their achieved knowledge has an irrelevant impact on their performance.

¹⁰⁴In Appendix B.1, Table B.3 shows descriptive statistics for these two samples.

education.¹⁰⁵

Figure 2.1: Fraction of students changing their school by grade

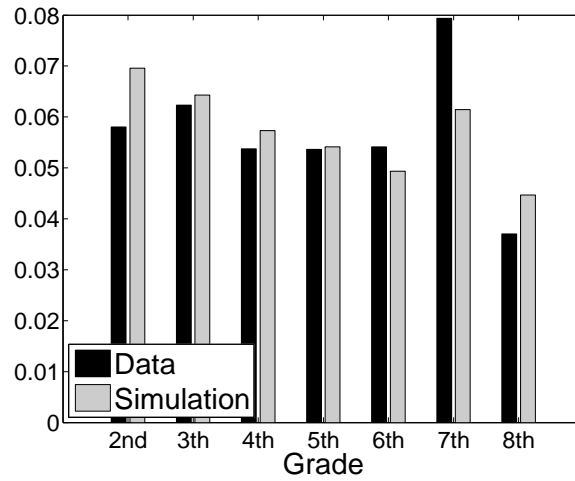
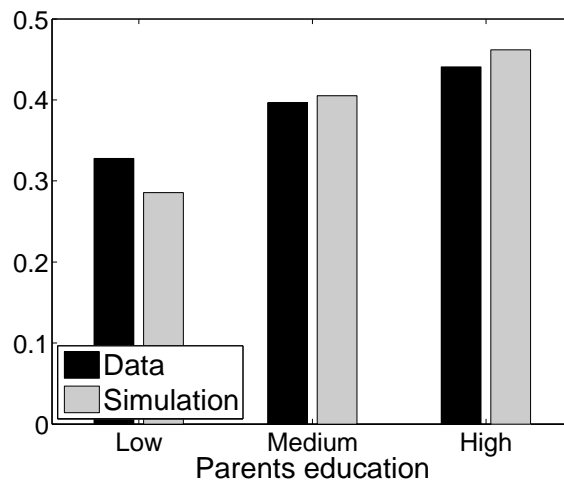


Figure 2.2: Average total change by parents education



Given the kind of counterfactual experiments that I perform, it is relevant to assess how the model fits the data with regard to some patterns of the decision of parents.

¹⁰⁵In the estimation and in the simulation, I collapse the information of parents' education into three categories: (1) both parents did not complete secondary education (low education); (2) One of the parents completed secondary education, but both parents did not attend higher education (medium education); and (3) at least one of the parents attended higher education (high education).

For instance, Figure 2.3 shows how the model fits parents' choice in terms of school type, namely, their decision about attending public, voucher-private or non voucher-private schools. Furthermore, the model is also able to generate the average quality of the schools selected by parents (Figure 2.4), which in the model is determined by how much parents care about quality, the correlation between quality and other features that parents value, and the admissions restrictions that parents face when they make their decision.

Figure 2.3: Student fraction by school type

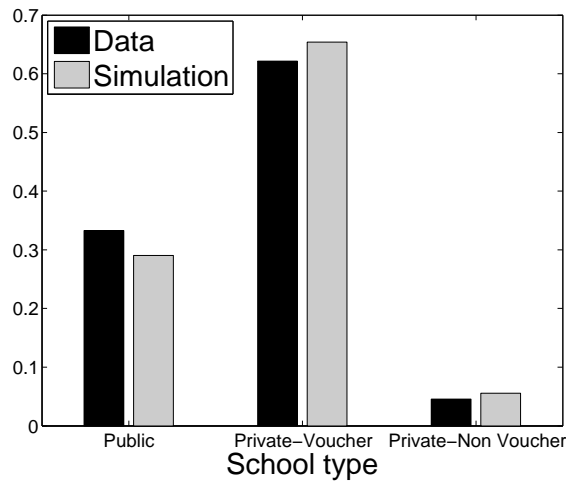
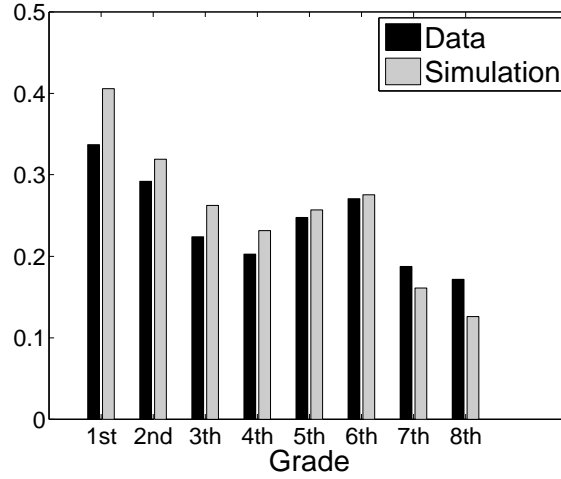


Figure 2.4: Average quality of the school selected by grade

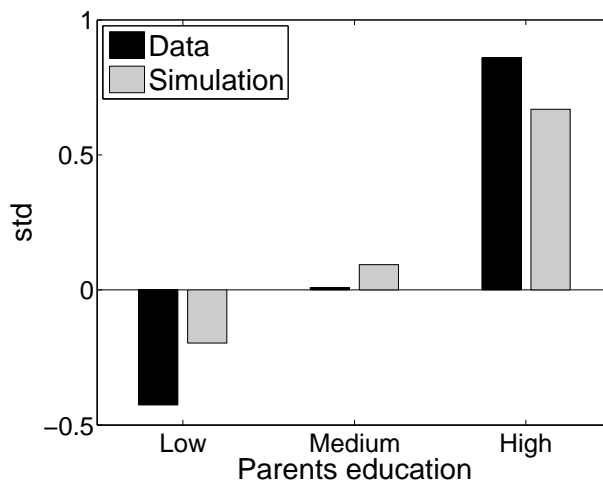


One common feature of many educational systems, which is extremely problematic in the Chilean case,¹⁰⁶ is the fact that students' access to different schools in terms of quality depends on their income. In the context of the model, this means that the initial knowledge gap $K_{0i} - K_{0i'}$, is increased by $K_{Ti} - K_{Ti'} - (K_{0i} - K_{0i'}) = (K_{Ti} - K_{0i}) - (K_{Ti'} - K_{0i'})$.

The model has several channels that can generate this correlation: parents can have differences in preferences about quality, differences in cognitive skills to understand information about schools, differences in involvement in the child's school, and differences in their choice restrictions. Figure 2.5 shows how, in the data and in the model, the knowledge gain is positively correlated with parents' education. This figure also says that the model overpredicts the gain for students with parents of low or medium education, and underpredicts this gain for students with highly educated parents.

¹⁰⁶Valenzuela, Bellei, and Ríos (2014).

Figure 2.5: Gain in Knowledge by parents education ($K_T - K_0$)



Overall, the fit of the model when the sample considers all the students in Santiago City is similar to the model fit when the estimation sample is used (Appendix B.3). The most important difference is that the former underestimates the frequency of school switches.

Parents Perception about Quality

Given the estimated parameters, it is possible to calculate the differences between parents' perception about school quality (which is determined by $\hat{\eta}_i$) and the effective quality of each school ($\tilde{q}_{jt} = E[\widehat{q_{jt}} | \mathbf{ST}_{jt}]$). Moreover, it is interesting to see how the distance between perception and reality affects the three school types differently.

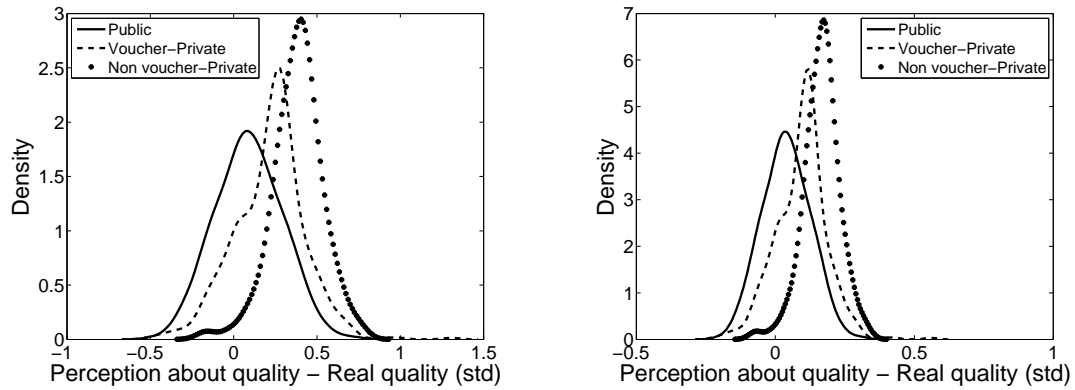
To this end, I take two prototypical parents, one with low education and with a choice set of 50 schools, the other highly educated with the same size choice set, and then calculate which would be their quality perception for each school of the sample. To conclude, I calculate the distance between perception and reality for each school.¹⁰⁷ Figure 2.6 shows the results of this exercise. In the first place, there is an important distance between perception and reality. In the second place, this

¹⁰⁷In practice, what I calculate is $E_{t-1}[q_{jt} | D_{it-1} \neq j] - \tilde{q}_{jt}$.

misperception is less severe for more educated parents. Finally, this misperception biases parents' preferences toward private schools. This bias is driven by the fact that voucher-private schools have more educated parents than public schools, whereas the same is true between non voucher-private and voucher-private schools. As discussed in the model section, because $\eta_\chi^i \geq 0$, parents overestimate the quality for schools whose students have, on average, highly educated parents.

Figure 2.6: Quality misperception by school administration type (2004)

(a) Parents with Low Education ($S(\Lambda) = 50$) (b) Parents with High Education ($S(\Lambda) = 50$)



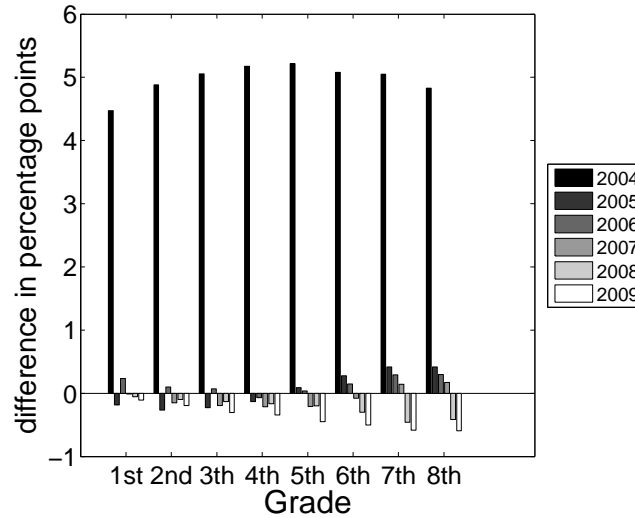
Counterfactual Experiments

To assess how important school quality is in parents' decisions, I simulate the model, randomly picking half of the schools and increasing their quality by 0.5 std, while decreasing the quality of the rest by the same amount. Then, I calculate the increase in the fraction of parents sending their children to the former schools. To see how relevant quality is in the first decision (first grade), vis-a-vis later decisions, I do this exercise by increasing schools' quality in different periods. For instance, I do not affect school quality until t , and I perform these quality changes from $t + 1$ to T .

Figure 2.7 shows the results of these exercises. On one hand, there is a moderate increase, of 4-5 percentage points, in the demand for schools that increase their quality since the first period. On the other hand, the effect is irrelevant when schools

change their quality after the first decision is made (1st grade).

Figure 2.7: Increase in the fraction of students in schools with higher quality



To study the mechanisms in parents' demand that explain the frequency of switching schools, the allocation of students across school types, and the correlation between the gain in knowledge and parents' educational level, I simulate the model under the following scenarios:¹⁰⁸

- No misperception ($\eta = 0$): parents correctly estimate school quality from standardized tests.
- Only quality matters: $U = \beta * K + C * (D_{it} \neq D_{it-1}) + \epsilon$.
- All admitted: $P(AD_{itj} = 1 | X_i) = 1 \forall i, j, t$.
- Random admissions: $AD_{ij} \perp\!\!\!\perp X_i$.
- $C * 0.9$: cost of changing school reduced by 10%.

¹⁰⁸It should be noticed that many of these policies may affect the choice set definition. Thus, given that a choice set is fixed in all these simulations, the effects of these policies are underestimated in this analysis.

- New locations: Parents with the lowest education are relocated to the municipality with highest average quality. Parents with the highest education are relocated to the municipality with lowest average quality.
- All $ED = 3$: All the students have the same knowledge endowment (K_0).

Table 2.1 shows the fractions of parents who switch schools by grade in the baseline simulation (first column), and the differences in percentage points – compared to the baseline – under each of the counterfactual experiments. From this table, it follows that, if parents were just concerned about quality, they would switch more often, which is explained by the fact that the other schools’ characteristics are more stable across the years (SES, price in std, and type). Admissions restrictions play a relevant role in attenuating the frequency of switches. Finally, and more obviously, this frequency is also attenuated by the switch cost.

Table 2.1: Fraction of students changing school by grade (with respect to baseline in percentage points)

	Baseline	No misperception	Only Q maters	All admitted	Random admission	$C * 0.9$	New locations	All $ED = 3$
2nd	6.9%	-0.3	0.6	12.2	0.0	4.9	-1.8	-0.3
3th	6.2%	-0.1	1.4	12.4	0.3	4.7	-1.4	0.0
4th	5.7%	-0.1	1.8	12.2	0.3	4.5	-1.4	-0.1
5th	5.2%	0.1	2.3	12.0	0.4	4.5	-1.1	0.0
6th	4.8%	-0.1	2.4	11.4	0.3	4.1	-1.0	-0.0
7th	6.1%	-0.3	1.6	10.4	-0.1	4.9	-1.8	-0.2
8th	4.5%	0.0	3.0	10.2	-0.2	3.5	-1.1	-0.2

Table 2.2 shows the fraction of parents by school type in the baseline simulation (first column), and the differences in percentage points – compared to the baseline

– under each of the counterfactual experiments. Even though parents’ perception is importantly biased in favor of private schools (with and without vouchers), when they decide based on the real quality, the fraction of parents attending public schools increases by a moderate 1.2 percentage points. This small effect, relative to the size of the misperception, is explained by the fact that parents do not care too much about quality. In fact, if parents were only concerned with school quality, there would be an increase of 1.9 percentage points in the fraction of parents choosing public schools, while this figure would decrease by 2.7 percentage points for voucher-private schools. This basically reflects the fact that the other elements of the utility function (SES of the school, the preference for its type, etc.) lead parents to apply to private schools.¹⁰⁹ Finally, these simulations allow us to see what would happen if less educated parents had a more relaxed choice set constraint. Columns 3, 4, and 7, all tell the same story: less choice set restrictions would lead (less educated) parents to choose private schools more often, though not necessary because of their higher quality.

Table 2.2: Student fraction by school type at first grade (with respect to baseline in percentage points)

	Baseline	No misperception	Only Q maters	All admitted	Random admission	$C * 0.9$	New locations	All $ED = 3$
Public	29.2%	1.2	1.9	-3.8	-3.2	0.6	-0.5	-4.1
Voucher Private	65.1%	-0.6	-2.7	0.4	0.6	-0.1	3.7	2.4
Non voucher Private	5.7%	-0.7	0.8	3.4	2.7	-0.5	-3.2	1.6

Table 2.3 shows the knowledge that students gained between 2004 and 2011, by parents’ education. While the numbers of the baseline simulation are presented in the first column, the numbers in the other columns are the differences in standard

¹⁰⁹It should be noticed that, in this model, peer effects are part of the school quality, which is constant in all the policy experiments. Therefore, in this model it is not possible to study a potential self-fulfilling prophecy, in which parents think that private schools are better, and therefore apply to those schools; those schools select the best students (those who have more educated parents); and, because of that pattern of admissions decisions, and given the peer effect, private schools end up being better than the public ones.

deviations – compared to the baseline – under each of the counterfactual experiments. The first result to notice is that, while an exclusive focus on quality would increase the knowledge gained by students whose parents have medium or high education, this shift in preferences would not have a relevant effect for students whose parents have a low level of education. This confirms the relevance of choice restrictions: for some parents, even if they put more weight on quality, they cannot find a better school for their child. A second element to notice is that both prohibiting schools from making admission decisions based on student characteristics and reallocating the poor families to better municipalities are effective measures to reduce the gap between students with parents with different levels of education.¹¹⁰

Table 2.3: Gain of knowledge ($KT - K0$) by parents education (with respect to baseline in standard deviations).

	Baseline	No misperception	Only Q maters	All admitted	Random admission	$C * 0.9$	New locations	All $ED = 3$
Incompleted High school	-0.240	-0.025	0.009	0.174	0.119	0.002	0.143	0.135
Completed High school	0.096	-0.045	0.112	0.258	0.201	0.002	0.021	0.202
With college studies	0.807	-0.095	0.204	0.153	0.057	-0.031	-0.008	0.024

Appendix B.3 contains the tables that show the results of the counterfactual experiments when using the complete sample, which includes all the students of Santiago City who entered first grade in 2003. Although the complete sample incorporates all the small municipalities that were not part of the estimation sample, the main conclusions (elaborated from Tables 2.1, 2.2, and 2.3) are not affected.

¹¹⁰This can be concluded by looking at columns 5, 8, and 7 of Table 2.3. Notice that in all these counterfactual experiments, the choice set (Λ) is fixed, in the sense that the set is invariant conditional on the municipality where parents live and their educational level. Thus, when a family of parents with low education is relocated from municipality A to municipality B, their new choice set (Λ) is going to be the choice set of a family with low-educated parents who live in municipality B.

2.7 Conclusions

This paper estimates a dynamic model of elementary school choice. To this end, I use detailed Chilean administrative data for the students who entered 1st grade of the elementary cycle in 2003, following them until 8th grade (2011), which in Chile is the end of the elementary cycle. The estimated model considers several elements that are relevant to explain parents' decisions, namely, how much do they care about school quality (and other school characteristics), parents' skill in understanding information about quality (national standardized tests), parents' involvement in the school attended by their children, and their choice set.

Assessing the relevance of these different components contributes to a better understanding of the demand for schools and the role that markets with competitive incentives can have in education. In particular, the structural approach followed in this paper allows me to quantify different sources of unequal access to high quality schools and of the higher demand for private schools than for public schools. In doing so, this paper also contributes to the scarce literature that estimates structural models with bounded rationality, as well as to the literature which uses observed choices to infer agents' information.

Regarding the debate about the extent to which parents base their decisions on school quality, I find that parents do care about school quality, but only to a moderate degree. Moreover, the simulations show that parents' decisions are not sensitive to changes in quality after the first decision (1st grade). I also find that more involved parents care marginally more about school quality.

The results show that parents have an important misperception about school quality, which causes them to have a less favorable opinion about public schools, relative to private schools. This result supports the idea that parents may have difficulty in isolating a school's quality from its socioeconomic composition when they observe test scores. However, given that quality is not very relevant for their decision, such a misperception has only a limited effect on parents' decisions.

Concerning the question of why parents choose private schools over public schools, the results show that, if parents were only concerned about quality, they would choose public schools more often. The result would be the same if they did not have a misperception about quality. However, if parents had more freedom in terms of the schools their children could attend, they would choose private schools more often. This last result suggests that admission rules are binding restrictions and that relaxing them would increase the demand for private schools.

Regarding the causes of the increase in the knowledge gap between students from different socioeconomic backgrounds, simulations show that schools' admission rules and household location are relevant in explaining the rise in this gap. This result supports the papers which argue that Chilean SES school segregation cannot be explained only by geographical segregation.¹¹¹

Finally, it should be noticed that, even though these counterfactual exercises are very useful to compare the effects of different policies on relevant outcomes (e.g., inequality), these are in general small effects. The latter can be partially explained by the fact that, in all the simulations, the choice set is fixed conditional on parents' education and home location. This limitation is something that should be addressed in future research.

¹¹¹See, for example, Valenzuela, Bellei, and Ríos (2014), Elacqua (2012) and Hsieh and Urquiola (2006).

Appendix A

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

A.1 Existence and Uniqueness

Existence

Lemma 2: If $\forall i : \theta_1(R_N - R_1)a_{1i}^2\phi(1) < \sigma_\eta^2$ and $\sum_i m_i \Phi\left(\frac{\theta_1(R_N - R_1) - \bar{F}C}{\sigma_{fc}}\right) > \sum_{\delta=2}^N S_\delta$ 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 $G_n(r) = r_n - \sum_{\delta=n}^N S_\delta + \sum_i m_i \Phi\left(\frac{D_i(r) - \bar{F}C}{\sigma_{fc}}\right) \left[1 - \Phi\left(\frac{r_n - e_i^1(r)a_{1i} - a_{0i}}{\sigma_\eta}\right)\right]$, where $r \in \mathbb{R}^{N-1}$, then I define the vector-value function $G(r)$ as:¹¹²

¹¹² $e_i^1(r)$ stands for the optimal effort decision for those who decide to take the college admissions test given the vector of cutoff scores r .

$$G(r) = \begin{bmatrix} G_2(r) \\ G_3(r) \\ \cdot \\ \cdot \\ \cdot \\ G_N(r) \end{bmatrix}$$

$$G(r) = \begin{bmatrix} r_2 - \left(\sum_{\delta=2}^N S_\delta - \sum_i m_i \Phi \left(\frac{D_i(r) - \bar{F}C}{\sigma_{fc}} \right) \left[1 - \Phi \left(\frac{r_2 - e_i^1(r) a_{1i} - a_{0i}}{\sigma_\eta} \right) \right] \right) \\ r_3 - \left(\sum_{\delta=3}^N S_\delta - \sum_i m_i \Phi \left(\frac{D_i(r) - \bar{F}C}{\sigma_{fc}} \right) \left[1 - \Phi \left(\frac{r_3 - e_i^1(r) a_{1i} - a_{0i}}{\sigma_\eta} \right) \right] \right) \\ \cdot \\ \cdot \\ \cdot \\ r_N - \left(\sum_{\delta=n+1}^N S_\delta - \sum_i m_i \Phi \left(\frac{D_i(r) - \bar{F}C}{\sigma_{fc}} \right) \left[1 - \Phi \left(\frac{r_N - e_i^1(r) a_{1i} - a_{0i}}{\sigma_\eta} \right) \right] \right) \end{bmatrix}$$

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 $G : M \rightarrow M$ 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:¹¹³

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

¹¹³As $r \rightarrow -\infty$

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

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

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

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

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

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 $\forall r > \underline{r} \Rightarrow G(r) > G(\underline{r}) > \underline{r}$.¹¹⁴ 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.¹¹⁵

To show that $G(r)$ is continuous it is sufficient to prove that $\forall i e_i(r)$ is continuous.¹¹⁶ 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 (1.11), it follows that:

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

But because the first term can not be bigger than $\theta_1(R_N - R_1) \left(\frac{a_{1i}}{\sigma_\eta} \right)^2 \phi(1)$, then¹¹⁷

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

¹¹⁴Because there exist \bar{r} such that $\forall r > \bar{r}: \sum_i m_i \Phi \left(\frac{D_i(r) - \bar{F}C}{\sigma_{fc}} \right) \left[1 - \Phi \left(\frac{r_n - e_i^1(r) a_{1i} - a_{0i}}{\sigma_\eta} \right) \right] < \sum_{\delta=n}^N S_\delta$, and \underline{r} such that $\forall r < \underline{r}: \sum_i m_i \Phi \left(\frac{D_i(r) - \bar{F}C}{\sigma_{fc}} \right) \left[1 - \Phi \left(\frac{r_n - e_i^1(r) a_{1i} - a_{0i}}{\sigma_\eta} \right) \right] > \sum_{\delta=n}^N S_\delta$.

¹¹⁵To be sure about non-emptiness, it is possible to pick $\underline{r} < 0$ and $\bar{r} > 0$.

¹¹⁶If $e_i^1(r)$ is continuous then $D_i(r)$ is also continuous.

¹¹⁷The function $x\phi(x)$ is maximized at $x = 1$.

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 ($e_i^0(r)$). ■

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 $r' > r$,¹¹⁸ from the general equilibrium definition it is directly shown that:

$$\begin{aligned} S &= \sum_i m_i \Phi \left(\frac{D_i - \bar{F}C}{\sigma_{fc}} \right) \left[1 - \Phi \left(\frac{r - e_i a_{1i} - a_{0i}}{\sigma_\eta} \right) \right] \\ S &= \sum_i m_i \Phi \left(\frac{D'_i - \bar{F}C}{\sigma_{fc}} \right) \left[1 - \Phi \left(\frac{r' - e'_i a_{1i} - a_{0i}}{\sigma_\eta} \right) \right] \end{aligned} \quad (\text{A.1})$$

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 $D_i = \Pi_1(r) + FC_i - \Pi_0$.¹¹⁹ Taking the derivative to D_i with respect to r ,¹²⁰

$$\begin{aligned} \frac{\partial D_i}{\partial r} &= \frac{\partial \Pi_1(r)}{\partial r} = (R_1 - R_2) \frac{\theta_1}{a_1} \phi \left(\frac{r - e a_1 - a_0}{\sigma_\eta} \right) < 0 \\ &\Rightarrow \left(\frac{D_i - \bar{F}C}{\sigma_{fc}} \right) > \left(\frac{D'_i - \bar{F}C}{\sigma_{fc}} \right) \end{aligned}$$

Therefore, from the later inequality and equations (A.1) it is directly shown that:

¹¹⁸Notice because $N = 2$, r and r' are scalars. S is the amount of seats offered by the only university.

¹¹⁹The value function for those who do not take the college admissions test does not depend on r .

¹²⁰Here, 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 .

$$\begin{aligned}
& \sum_i m_i \left(\Phi \left(\frac{D_i - \bar{F}C}{\sigma_{fc}} \right) \left[1 - \Phi \left(\frac{r - e_i a_{1i} - a_{0i}}{\sigma_\eta} \right) \right] - \Phi \left(\frac{D'_i - \bar{F}C}{\sigma_{fc}} \right) \right. \\
& \quad \left. \left[1 - \Phi \left(\frac{r' - e'_i a_{1i} - a_{0i}}{\sigma_\eta} \right) \right] \right) = 0 \\
& \Rightarrow \sum_i m_i \left(\left[1 - \Phi \left(\frac{r - e_i a_{1i} - a_{0i}}{\sigma_\eta} \right) \right] - \left[1 - \Phi \left(\frac{r' - e'_i a_{1i} - a_{0i}}{\sigma_\eta} \right) \right] \right) < 0 \\
& \Rightarrow \sum_i m_i \left(\Phi \left(\frac{r' - e'_i a_{1i} - a_{0i}}{\sigma_\eta} \right) - \Phi \left(\frac{r - e_i a_{1i} - a_{0i}}{\sigma_\eta} \right) \right) < 0
\end{aligned}$$

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

■

Step 2:

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

Case 1: effort decision is continuous in r :

Taking a derivative of the first order condition (1.11), when $N = 2$ implies:¹²¹

$$\begin{aligned}
\theta_1(R_2 - R_1) \frac{\partial \phi(r)}{\partial r} \left[\frac{1 - \frac{\partial e}{\partial r} a_1}{\sigma_\eta} \right] \frac{a_1}{\sigma_\eta} = \frac{\partial e}{\partial r} & \Rightarrow \frac{\partial e}{\partial r} = \frac{\theta_1(R_2 - R_1) \frac{\partial \phi(r)}{\partial r} \frac{a_1}{\sigma_\eta}}{1 + \theta_1(R_2 - R_1) \frac{\partial \phi(r)}{\partial r} \left(\frac{a_1}{\sigma_\eta} \right)^2} \\
\Rightarrow 1 - \frac{\partial e}{\partial r} a_1 = \frac{1}{1 + \theta_1(R_2 - R_1) \frac{\partial \phi(r)}{\partial r} \left(\frac{a_1}{\sigma_\eta} \right)^2} & \tag{A.2}
\end{aligned}$$

Therefore,

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

¹²¹For simplicity, I suppress the individual sub-index and denote $\phi \left(\frac{r - e a_1 - a_0}{\sigma_\eta} \right)$ as $\phi(r)$.

Thus, 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:¹²²

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

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 :¹²³

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:

$$\frac{\partial\Pi_l}{\partial r} - \frac{\partial\Pi_h}{\partial r} = \frac{\theta_1(R_2 - R_1)}{\sigma_\eta} \left[\phi\left(\frac{r - e_l a_1 - a_0}{\sigma_\eta}\right) - \phi\left(\frac{r - e_h a_1 - a_0}{\sigma_\eta}\right) \right] \quad (\text{A.3})$$

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

$$\begin{aligned} e_h - e_l &= \frac{a_1\theta_1(R_2 - R_1)}{\sigma_\eta} \left[\phi\left(\frac{r - e_l a_1 - a_0}{\sigma_\eta}\right) - \phi\left(\frac{r - e_h a_1 - a_0}{\sigma_\eta}\right) \right] \\ &\Rightarrow \frac{\partial\Pi_l}{\partial r} - \frac{\partial\Pi_h}{\partial r} = \frac{e_h - e_l}{\sigma_\eta} > 0 \end{aligned} \quad (\text{A.4})$$

Therefore, by (A.4) 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.¹²⁴ Loosely speaking, this means

¹²²I am assuming away $-\theta_1(R_2 - R_1)\frac{\partial\phi(r)}{\partial r}\left(\frac{a_1}{\sigma_\eta}\right)^2 - 1 = 0$.

¹²³Given 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.

¹²⁴It 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.

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.¹²⁵ First, it is proved that $\sum_{n=2}^N \frac{\partial \overline{G}_m}{\partial r_n} < -\sum_i m_i \Phi \left(\frac{D_i - \overline{FC}}{\sigma_{fc}} \right) \frac{\phi_i(r_m)}{\sigma_\eta} \left[1 - a_{1i} \sum_{n=2}^N \frac{\partial e_i^1}{\partial r_n} \right]$.¹²⁶ Second, I show that $1 - a_{1i} \sum_{n=2}^N \frac{\partial e_i^1}{\partial r_n} > 0 \forall i$.

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:

$$\begin{aligned} \sum_{n=2}^N \frac{\partial \overline{G}_m}{\partial r_n} &< -\sum_i m_i \Phi \left(\frac{D_i - \overline{FC}}{\sigma_{fc}} \right) \frac{\phi_i(r_m)}{\sigma_{fc}} + \sum_{n=2}^N \sum_i m_i \Phi \left(\frac{D_i - \overline{FC}}{\sigma_{fc}} \right) \phi_i(r_m) \frac{\partial e_i^1}{\partial r_n} \frac{a_{1i}}{\sigma_{fc}} \\ &= -\sum_i m_i \Phi \left(\frac{D_i - \overline{FC}}{\sigma_{fc}} \right) \frac{\phi_i(r_m)}{\sigma_{fc}} \left[1 - a_{1i} \sum_{n=2}^N \frac{\partial e_i^1}{\partial r_n} \right] \end{aligned}$$

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:

¹²⁵For simplicity the result is shown for the case where G is continuous.

¹²⁶ $\phi_i(r_m) = \phi \left(\frac{r_m - e_i^1 a_{1i} - a_{0i}}{\sigma_\eta} \right)$.

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

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

A.2 Variable Descriptions

Table A.1: 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>I like to study math</i> : 1 (strongly agree), 2 (agree), 3 (disagree and strongly disagree).
Like_spanish	<i>I like to study Spanish</i> : 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	<i>When I study, I exert effort even if it is a difficult subject:</i> 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
ME2	<i>When I study, I try hard to learn:</i> 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
ME3	<i>When I study and I am not getting something, I look for additional information:</i> 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>I feel able to understand the harder subjects covered by the teachers:</i> 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
MS2	<i>I trust that I can do excellent homework and exams:</i> 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
MS2	<i>If I set a goal about learning well something, I can do it:</i> 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
MS4	<i>If I decide not to have poor marks, I really can avoid them:</i> 1 (always or almost always), 2 (often), 3 (occasionally) and 4 (never or almost never).
MS5	<i>When I study I lose the focus, because I am not good at studying:</i> 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 A.2: 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 A.3: 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 A.4: 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 A.5: 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 A.6: 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 A.7: 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.

A.3 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:¹²⁷ $E[\delta_i^T | x_i^h] = 0$, $E[\delta_i^T | x_i^{e1h}] = 0$, $E[\delta_i^T | x_i^{\lambda 1p}] = 0$, $E[\delta_i^T | Me_{2i}^h] = 0$ and $E[\delta_i^T | M\lambda_{2i}^p] = 0$ from which β^T , β^{e1h} and $\beta^{\lambda 1p}$ are identified.¹²⁸ 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:

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

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

¹²⁷Because 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.

¹²⁸This implies that all the parameters involved in a_{0i} , a_{1i} and b_{1i} are identified.

Which means that $var(\varepsilon_i^{e1h})$ and $var(\varepsilon_i^{\lambda1h})$ are identified, and consequently the vector $\{var(\varepsilon_i^{pm}), var(\varepsilon_i^{pv}), var(\varepsilon_i^{gh})\}$ is 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:

$$\begin{aligned} \widehat{SIMCE}_{ji}^p &= \lambda_i + \widehat{\varepsilon}_i^{sjp} \\ M\lambda_{ji}^p - x_i^{\lambda1p} \beta_1^{\lambda1p} &= \lambda_i + \varepsilon_i^{\lambda1p} \end{aligned}$$

Therefore, because $\widehat{\varepsilon}_i^{sjp}$ and $\varepsilon_i^{\lambda1p}$ are independent of each other and with respect to λ_i , the distribution of λ is identified (Cunha and Heckman (2008)).¹²⁹

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.¹³⁰

First, notice that when $TCAT_i = 0$, then $Me_{1i}^h - \varepsilon_i^{e1h} = b_{i1}\theta_2$

$$\Rightarrow E[Me_{1i}^h - \varepsilon_i^{e1h} | TCAT_i = 0] = \theta_2 E[b_{i1} | TCAT_i = 0]$$

$$\Rightarrow \theta_2 = \frac{E[Me_{1i}^h | TCAT_i = 0]}{E[b_{i1} | TCAT_i = 0]}$$

which ensures the identification of θ_2 .

¹²⁹The identification can be achieved under much weaker conditions regarding measurement errors. Indeed, independence is not necessary; see Cunha, Heckman, and Schennach (2010).

¹³⁰Here I show the identification when the utility parameters are not individual specific, but the extension is trivial.

Similarly, because $TCAT_i = 1$ implies that $Me_{1i}^h = g(x_i, a_{1i}(\lambda_i), b_{1i}, \theta_1, \theta_2, \sigma_\eta) + \varepsilon_i^{e1h}$, then¹³¹:

$$\begin{aligned} \Rightarrow E[Me_{1i}^h | \lambda_i, x_i, TCAT_i = 1] &= E[g(x_i, a_{1i}(\lambda_i), b_{1i}, \theta_1, \theta_2, \sigma_\eta) | \lambda_i, x_i, TCAT_i = 1] \\ &\Rightarrow \int_{\lambda} E[Me_{1i}^h | \lambda, x_i, TCAT_i = 1] f(\lambda) d\lambda \\ &= \int_{\lambda} E[g(x_i, a_{1i}(\lambda_i), b_{1i}, \theta_1, \theta_2, \sigma_\eta) | \lambda, x_i, TCAT_i = 1] f(\lambda) d\lambda \end{aligned}$$

which allows for the identification of θ_1 .

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

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

A.4 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 $T_i \in \{PSUM_i, PSUV_i\}$:

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

¹³¹ $g(x_i, a_{1i}(\lambda_i), b_{1i}, \theta_1, \theta_2, \sigma_\eta)$ is the implicit function associated with the first order condition (1.11).

$$Pr(TCAT|x_i^h, e_i^h, \lambda_t, \Omega) = \Phi\left(\frac{D_i - \bar{FC}}{\sigma_{fc}}\right)^{TCAT_i} \left(1 - \Phi\left(\frac{D_i - \bar{FC}}{\sigma_{fc}}\right)\right)^{1 - TCAT_i}. \quad (\text{A.5})$$

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

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

$$F_i(\text{high school tests} | Type_\lambda = t) = \prod_{T_i} f(T_i|x_i^h, e_i^h, \lambda_t, \Omega). \quad (\text{A.6})$$

Similarly, the contributions to the likelihood of high school effort measures are described by:¹³²

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

$$F_i(\text{high school effort measures}) = \prod_j f(Me_{ji}^h|x_i^{ejh}, e_i^h, TCAT_i, \Omega). \quad (\text{A.7})$$

Along the same lines, the contributions to the likelihood of the unobserved learning skill measures are described by:¹³³

$$f(M\lambda_{ji}^p|x_i^{\lambda jp}, \lambda_t, \Omega) = \phi\left(\frac{M\lambda_{ji}^p - x_i^{\lambda jp} \beta_1^{\lambda jp} - \lambda_t \alpha^{\lambda jp}}{\varepsilon^{ejh}}\right) \frac{1}{\sigma_{\varepsilon^{ejh}}}, \quad j \in \{1, \dots, J_\lambda\}$$

¹³²Here for simplicity the effort measurements are assumed to be continuous, but in the estimation I use ordered probit specifications.

¹³³Again, these measures are assumed to be continuous, but in the estimation I use ordered probit specifications.

$$F_i(\text{learning skill measures} \mid \text{Type}_\lambda = t) = \prod_j f(M\lambda_{ji}^p \mid x_i^{\lambda jp}, \lambda_t, \Omega) \quad (\text{A.8})$$

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

$$T_i \in \{GPA_i^p, SIMCE_{math,i}^p, SIMCE_{verbal,i}^p, SIMCE_{socialscience,i}^p, SIMCE_{naturalscience,i}^p\}.$$

Given that, conditional on λ_i , x_i^h and e_i^h , the ε_i are independent across tests, the contribution to the likelihood is given by¹³⁴:

$$f(T_i \mid x_i^p, e_i^p, \lambda_t, \Omega) = \phi \left(\frac{T_i - \beta_0^T - x_i^h \beta_1^T - (\widehat{M}e_{1i}^p - x_i^{e1p} \beta_1^{e1p}) \beta_2^T + \lambda_t \beta_3^T}{\sigma_{\varepsilon^T}} \right) \frac{1}{\sigma_{\omega^T}},$$

$$F_i(\text{primary school tests} \mid \text{Type}_\lambda = t) = \prod_{T_i} f(T_i \mid x_i^p, e_i^p, \lambda_t, \Omega). \quad (\text{A.9})$$

Therefore, the likelihood contribution for the i th individual is thus:

$$L_i(\Omega) = \quad (\text{A.10})$$

$$\log \left(\sum_t F_i(\text{high school tests} \mid \text{Type}_\lambda = t) F_i(\text{high school effort measures} \mid \text{Type}_\lambda = t) F_i(\text{learning skill measures} \mid \text{Type}_\lambda = t) F_i(\text{primary school tests} \mid \text{Type}_\lambda = t) \pi_t \right) \quad (\text{A.11})$$

¹³⁴ $\widehat{M}e_{1i}^p = \widehat{\delta}_1 + \sum_{m=2}^{J_{ep}} M e_{mi}^p \widehat{\delta}_m$ and $\omega_i^T = \varepsilon_i^T - \varepsilon_i^{e1p} \beta_2^T$, where the $\widehat{\delta}$ s are the OLS coefficients.

A.5 Results

First satge parameters

Table A.8: Primary School Attending Regression

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_LEN==2	0.0140 (0.0473)
STUDY_LEN==3	-0.0481 (0.0497)
STUDY_LEN==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 omitted due to perfect multicollinearity.

Table A.9: Repetitions Linear Probability Regression

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 omitted due to perfect multicollinearity.

Table A.10: Secondary School Attending Regression

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 omitted due to perfect multicollinearity.

Table A.11: Two Stage Least Square for Primary Education Students' Performance

	(1)	(2)	(3)	(4)	(5)
	simcev	simcem	simcen	simces	gpap
VARIABLES	SIMCE_V_P	SIMCE_M_P	SIMCE_N_P	SIMCE_S_P	GPA_P
SEX	-8.655*** (0.232)	9.499*** (0.222)	9.414*** (0.228)	9.565*** (0.230)	-0.164*** (0.00258)
RURAL_P	1.562*** (0.463)	1.439*** (0.449)	3.548*** (0.434)	2.933*** (0.459)	0.0627*** (0.00529)
SES_P2	1.055** (0.529)	-0.880* (0.513)	0.0666 (0.488)	0.626 (0.526)	-0.0681*** (0.00605)
SES_P3	10.99*** (0.581)	8.463*** (0.559)	9.793*** (0.536)	11.44*** (0.573)	-0.0697*** (0.00652)
SES_P4	28.05*** (0.653)	28.06*** (0.634)	29.29*** (0.615)	30.13*** (0.643)	-0.0390*** (0.00724)
SES_P5	42.33*** (1.040)	48.86*** (1.016)	46.75*** (1.044)	42.92*** (1.007)	0.0632*** (0.0114)
EDU_MO2	-6.537*** (0.454)	-5.747*** (0.430)	-6.131*** (0.431)	-5.259*** (0.447)	-0.0186*** (0.00512)
EDU_MO3	-1.687*** (0.380)	-1.345*** (0.359)	-2.424*** (0.367)	-1.734*** (0.374)	0.0456*** (0.00424)
EDU_MO4	6.452*** (0.374)	5.097*** (0.355)	5.362*** (0.366)	6.826*** (0.368)	0.138*** (0.00414)
EDU_MO5	9.035*** (0.467)	7.438*** (0.452)	8.870*** (0.469)	10.36*** (0.462)	0.146*** (0.00517)
EDU_MO6	14.68*** (0.526)	14.04*** (0.508)	16.16*** (0.529)	16.51*** (0.519)	0.209*** (0.00585)
EDU_FAC	6.646*** (0.410)	7.574*** (0.401)	8.040*** (0.420)	6.743*** (0.407)	0.0583*** (0.00455)
DEP_P2	1.392*** (0.289)	0.733*** (0.276)	2.320*** (0.283)	2.267*** (0.287)	-0.103*** (0.00318)
DEP_P3	-0.515 (0.851)	-1.043 (0.830)	1.070 (0.882)	-2.413*** (0.822)	-0.136*** (0.00923)
ATTEN_P_hat	-6.594*** (0.751)	-13.26*** (0.711)	-8.170*** (0.748)	-7.535*** (0.732)	0.262*** (0.00813)
REP_hat	-347.8*** (5.985)	-430.5*** (5.640)	-360.9*** (5.874)	-379.0*** (5.837)	-6.906*** (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*** (71.99)	1,387*** (68.18)	997.0*** (71.72)	955.6*** (70.25)	-18.65*** (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 A.12: Two Stage Least Square for Secondary Education Students' Performance

	(1)	(2)	(3)	(4)	(5)
	simcev	simcem	psum	psuv	gpas
VARIABLES	SIMCE_V_S	SIMCE_M_S	PSU_M	PSU_V	GPA_S
SEX	-7.136*** (0.271)	11.66*** (0.322)	23.28*** (0.579)	3.641*** (0.594)	-28.31*** (0.572)
RURALS	-3.703*** (0.763)	-5.968*** (0.927)	-12.84*** (1.768)	-12.15*** (1.869)	-0.624 (1.650)
SES_S2	10.02*** (0.443)	13.94*** (0.530)	22.80*** (1.058)	24.38*** (1.091)	-4.249*** (0.935)
SES_S3	29.05*** (0.506)	39.20*** (0.603)	82.08*** (1.142)	77.91*** (1.178)	11.86*** (1.078)
SES_S4	44.82*** (0.635)	65.23*** (0.750)	134.3*** (1.357)	122.1*** (1.420)	30.27*** (1.339)
SES_S5	53.15*** (1.289)	79.01*** (1.443)	166.9*** (2.538)	151.2*** (2.630)	57.56*** (2.818)
EDU_MO2	-4.213*** (1.456)	0.235 (1.737)	-1.655 (3.323)	-9.321*** (3.430)	16.25*** (2.888)
EDU_MO3	-0.757 (1.430)	4.018** (1.704)	2.099 (3.238)	-4.819 (3.348)	14.84*** (2.830)
EDU_MO4	6.147*** (1.427)	11.08*** (1.699)	15.17*** (3.221)	10.40*** (3.329)	24.08*** (2.824)
EDU_MO5	9.046*** (1.459)	13.78*** (1.734)	20.05*** (3.271)	18.37*** (3.381)	26.28*** (2.899)
EDU_MO6	16.54*** (1.482)	21.63*** (1.760)	37.82*** (3.310)	37.63*** (3.421)	43.86*** (2.959)
EDU_FAC	6.973*** (0.451)	9.051*** (0.528)	20.59*** (0.880)	20.08*** (0.913)	16.39*** (0.975)
DEP_S2	-2.134*** (0.315)	-2.407*** (0.378)	-14.71*** (0.693)	-11.13*** (0.710)	-13.03*** (0.657)
DEP_S3	-1.673 (1.167)	-0.909 (1.298)	-1.440 (2.256)	-1.454 (2.316)	-9.691*** (2.555)
ATTEN_S_hat	3.817*** (0.296)	4.724*** (0.356)	11.52*** (0.641)	10.54*** (0.657)	26.94*** (0.614)
REP_hat	-316.3*** (6.754)	-462.2*** (8.013)	-736.7*** (14.32)	-675.2*** (14.95)	-971.5*** (13.58)
Constant	-90.85*** (27.87)	-185.4*** (33.50)	-611.9*** (60.30)	-512.2*** (61.84)	-1,936*** (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

Second stage parameters

Table A.13: Second Stage Parameters

Utility								
θ_1	0.0000139	(0.0000008)	θ_3^2	0.00306	(0.00018)	F^C	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								
$\alpha_e(\text{effort_p})$	0.1099	(0.0076)	Cut_{sb}^2	3.1656	(0.8147)	Cut_{sp}^3	4.7551	(0.9201)
$\alpha_{const}(\text{effort_p})$	-10.9229	(0.7171)	Cut_{sb}^3	4.3483	(0.8149)	$\beta_{ca}^{eh}(\text{effort})$	0.0388	(0.0084)
σ_{atten}	3.8072	(0.0073)	$\beta_{sp}^{eh}(\text{effort})$	0.0466	(0.0098)	$\beta_{ca}^{eh}(\text{ses_s2})$	0.0002	(0.0103)
$\beta_{sb}^{eh}(\text{effort})$	0.0408	(0.0087)	$\beta_{sp}^{eh}(\text{ses_s2})$	0.0520	(0.0128)	$\beta_{ca}^{eh}(\text{ses_s3})$	0.0602	(0.0117)
$\beta_{sb}^{eh}(\text{ses_s2})$	-0.0650	(0.0108)	$\beta_{sp}^{eh}(\text{ses_s3})$	0.1168	(0.0142)	$\beta_{ca}^{eh}(\text{ses_s4})$	0.2092	(0.0150)
$\beta_{sb}^{eh}(\text{ses_s3})$	-0.0994	(0.0123)	$\beta_{sp}^{eh}(\text{ses_s4})$	0.2183	(0.0173)	$\beta_{ca}^{eh}(\text{ses_s5})$	0.2941	(0.0185)
$\beta_{sb}^{eh}(\text{ses_s4})$	-0.0287	(0.0155)	$\beta_{sp}^{eh}(\text{ses_s5})$	0.3943	(0.0205)	$\beta_{ca}^{eh}(\text{edu_fac})$	0.0447	(0.0115)
$\beta_{sb}^{eh}(\text{ses_s5})$	0.3718	(0.0190)	$\beta_{sp}^{eh}(\text{edu_fac})$	0.1128	(0.0123)	Cut_{ca}^1	1.9459	(0.7840)
$\beta_{sb}^{eh}(\text{edu_fac})$	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)	$\lambda(\text{Type1})$	0.0001	(0.000017)
Cut_{ms1}^1	0.0402	(0.0063)	Cut_{ms3}^1	0.9588	(0.0077)	$\lambda(\text{Type2})$	0.0561	(0.0006)
Cut_{ms1}^2	1.1359	(0.0070)	Cut_{ms3}^2	1.9305	(0.0094)	$\lambda(\text{Type3})$	0.1078	(0.0011)
Cut_{ms1}^3	2.4898	(0.0098)	Cut_{ms3}^3	0.9588	(0.0077)	$\lambda(\text{Type4})$	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 .

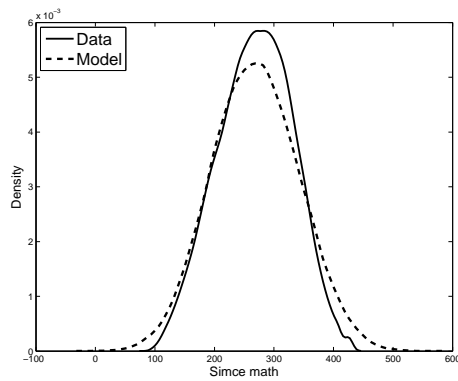
Model fit

Table A.14: 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 A.1: Tests 2006

(a) Simce math 2006



(b) Simce verbal 2006

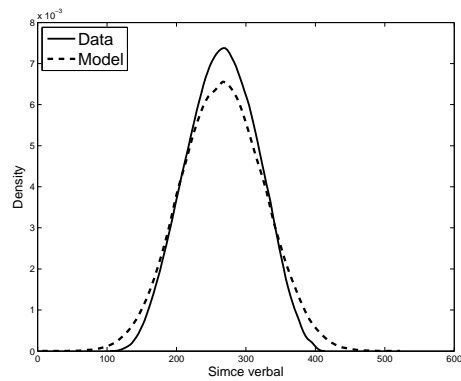
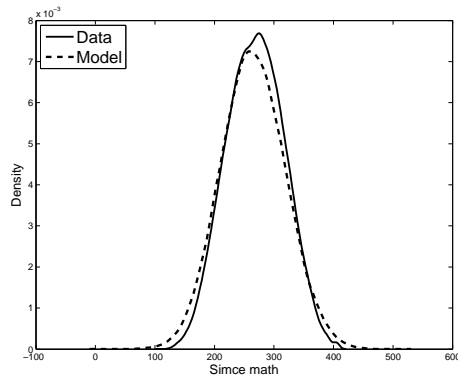
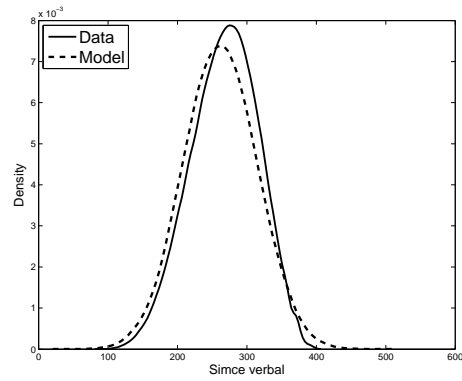


Figure A.2: Tests 2004

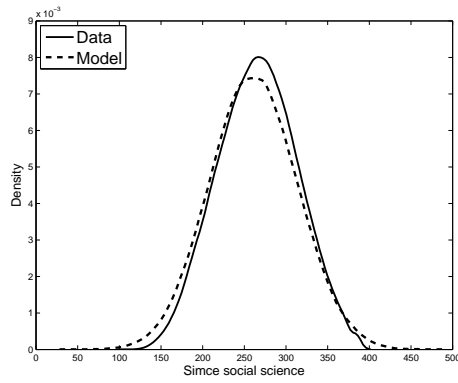
(a) Simce math 2004



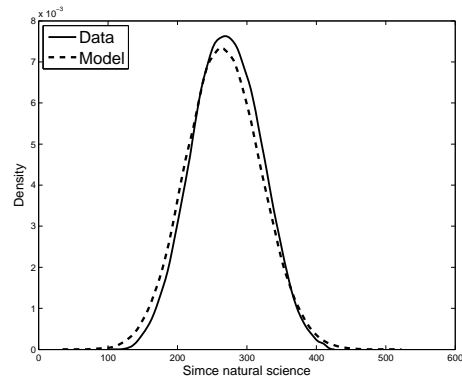
(b) Simce verbal 2004



(c) Simce natural science 2004



(d) Simce social science 2004



(e) GPA 2004

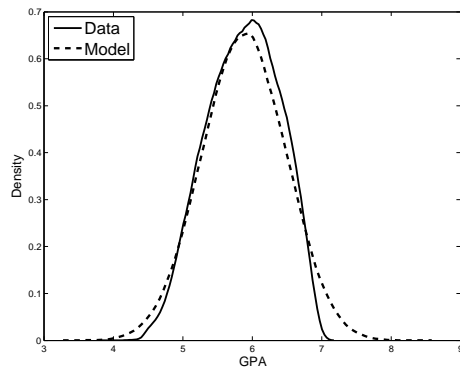
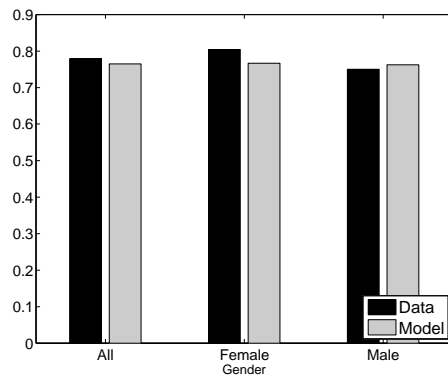
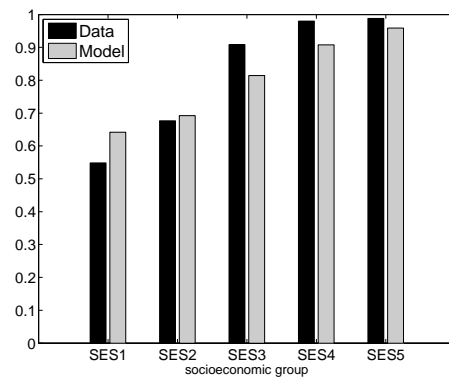


Figure A.3: Fraction of the students taking the PSU by groups

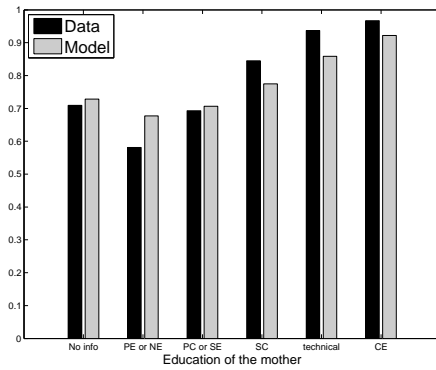
(a) By gender



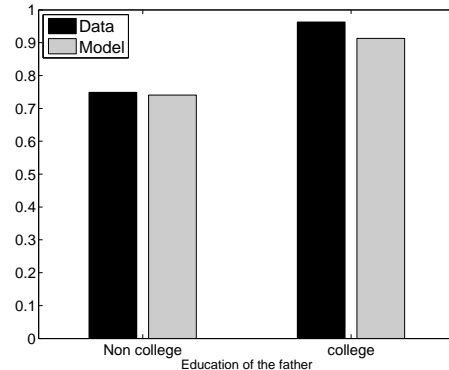
(b) By high school SES



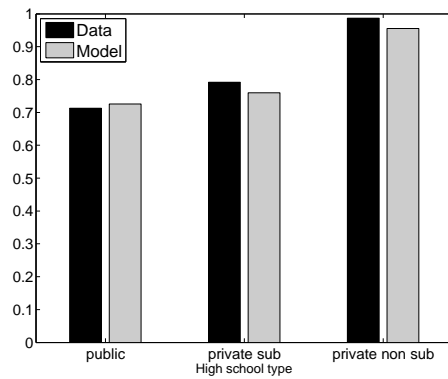
(c) By mother's education



(d) By father's education



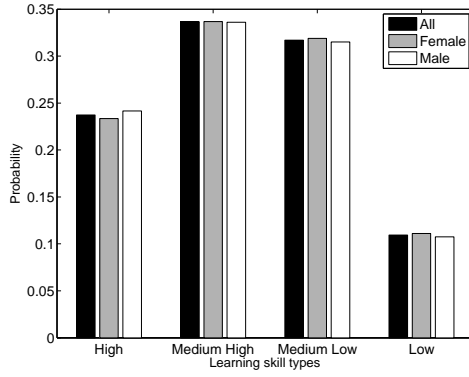
(e) By high school categories



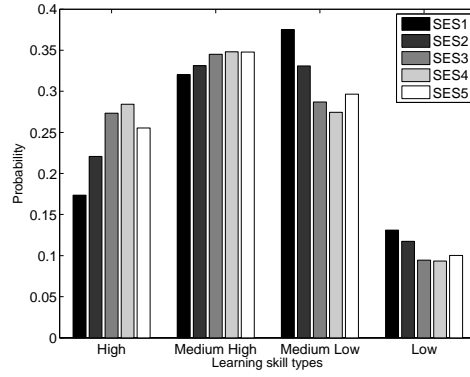
Unobserved types

Figure A.4: Conditional probabilities of learning skill types by groups

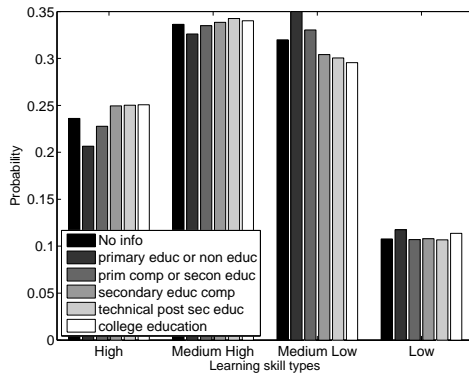
(a) All and by gender



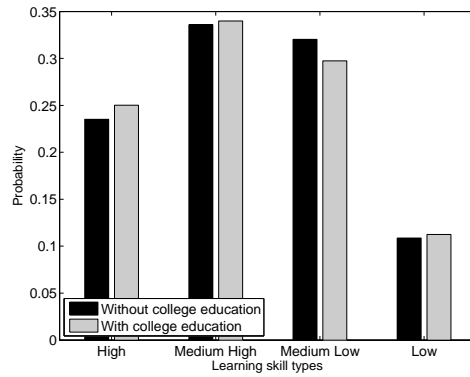
(b) By high school SES



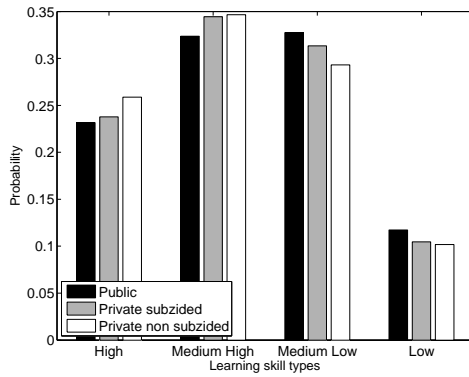
(c) By mother's education



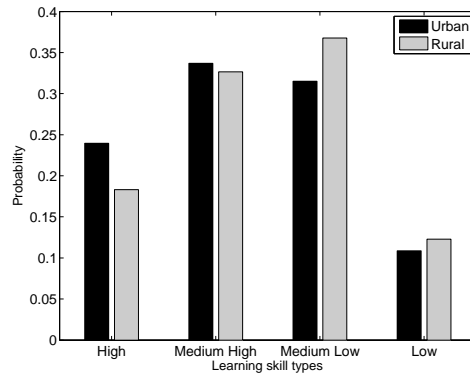
(d) By father's education



(e) By high school categories



(f) By rurality



A.6 Counterfactual Experiments

Figure A.5: Impact of introducing quotas by SES on universities' socioeconomic composition

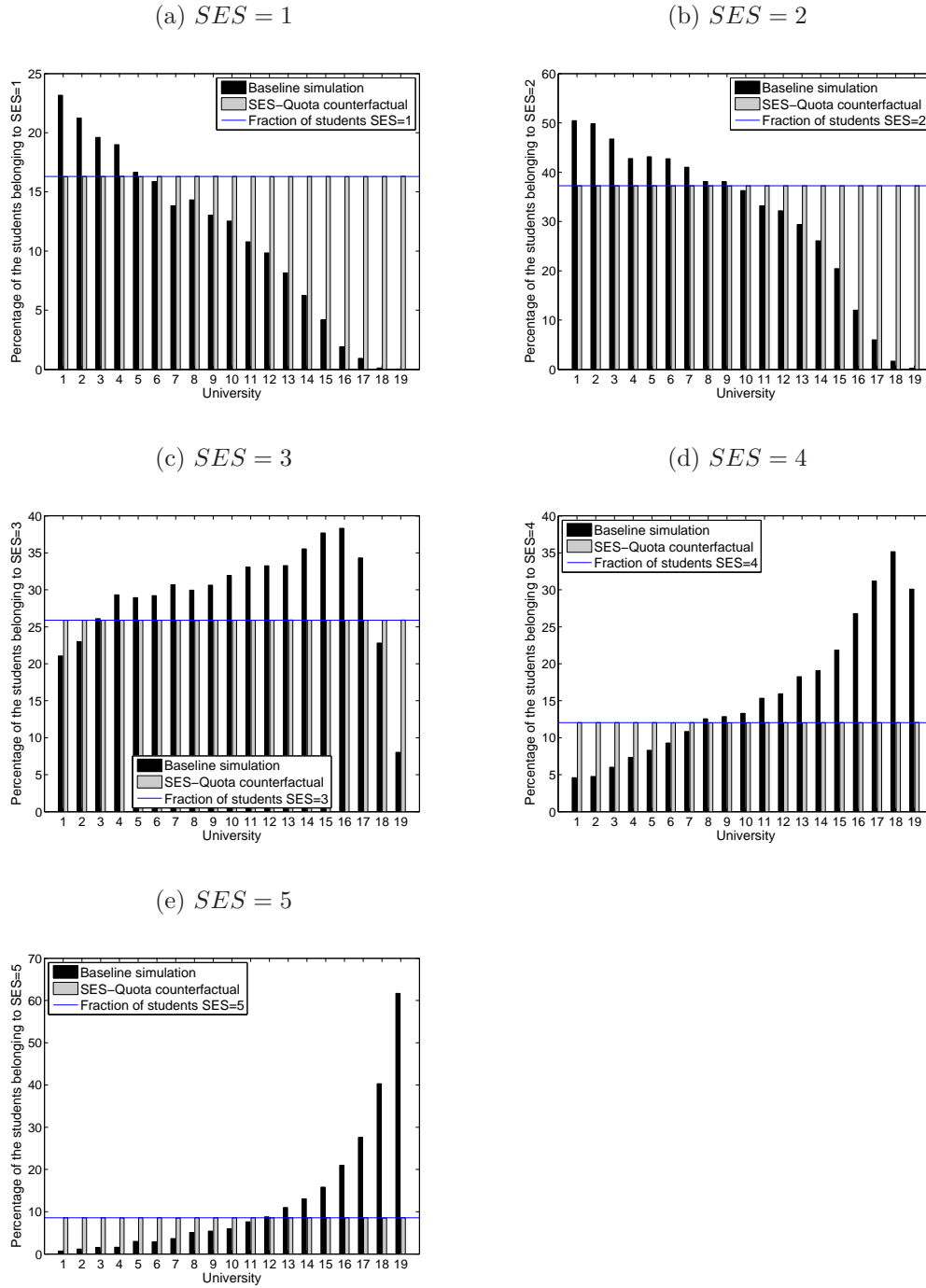
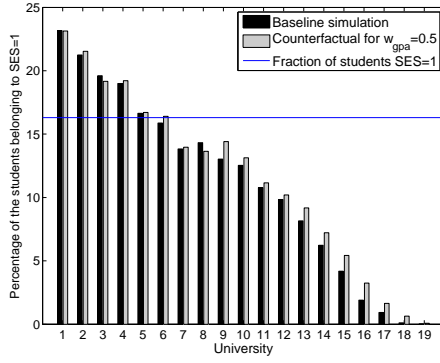
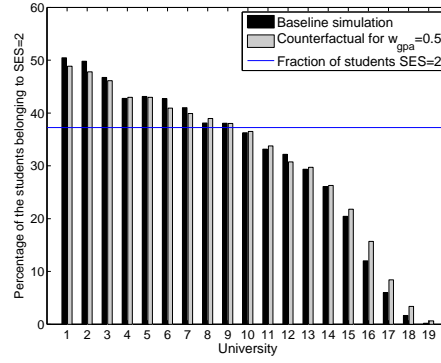


Figure A.6: Impact of changing GPA weight from 0.4 to 0.5 on universities' socio-economic composition

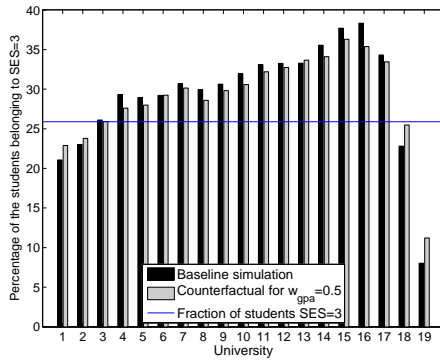
(a) $SES = 1$



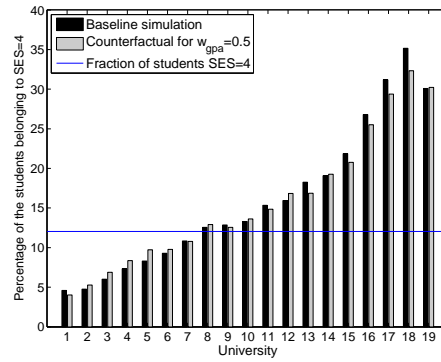
(b) $SES = 2$



(c) $SES = 3$



(d) $SES = 4$



(e) $SES = 5$

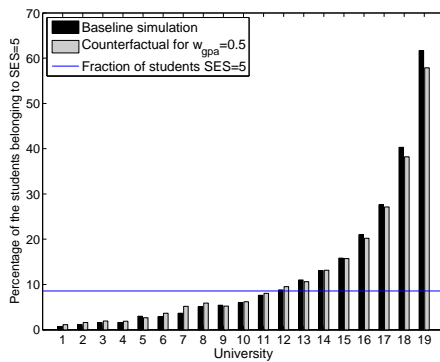
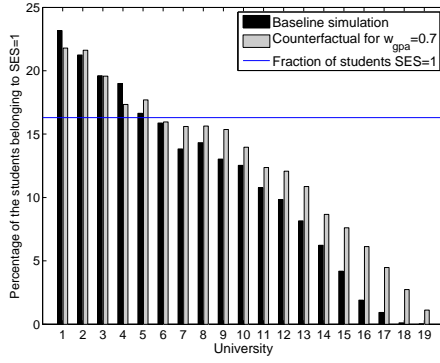
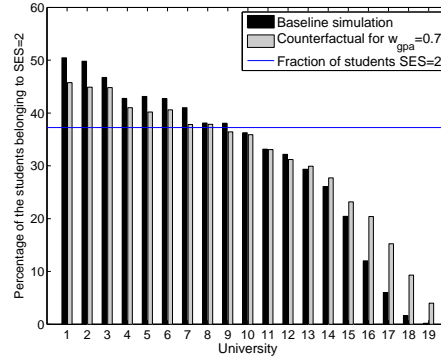


Figure A.7: Impact of changing GPA weight from 0.4 to 0.7 on universities' socio-economic composition

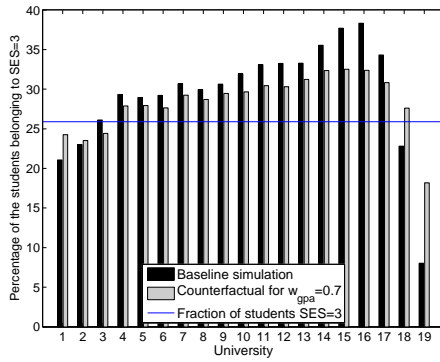
(a) $SES = 1$



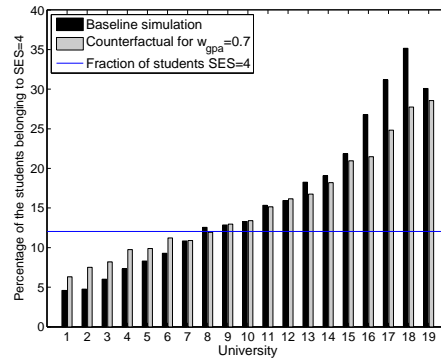
(b) $SES = 2$



(c) $SES = 3$



(d) $SES = 4$



(e) $SES = 5$

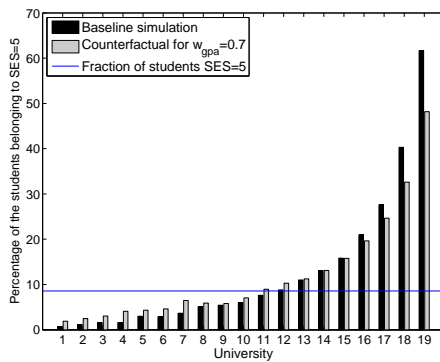
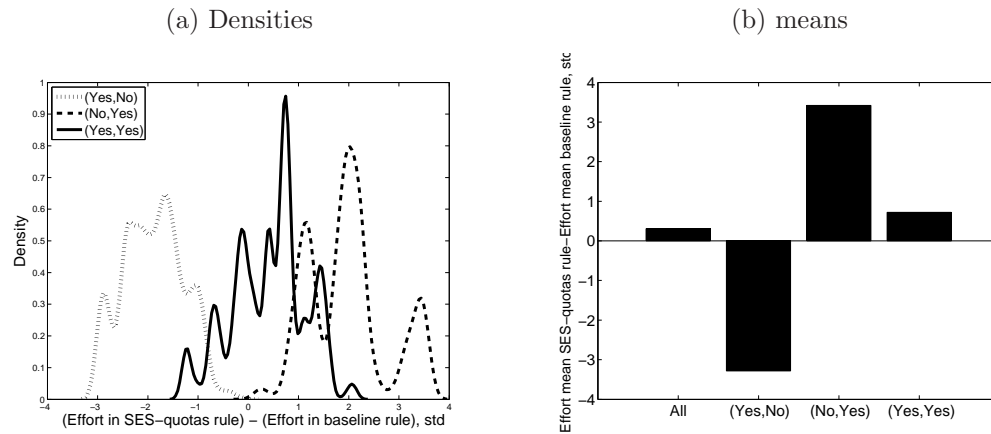


Figure A.8: The impact on effort of quota by SES



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 A.9: The impact on effort of changing GPA weight

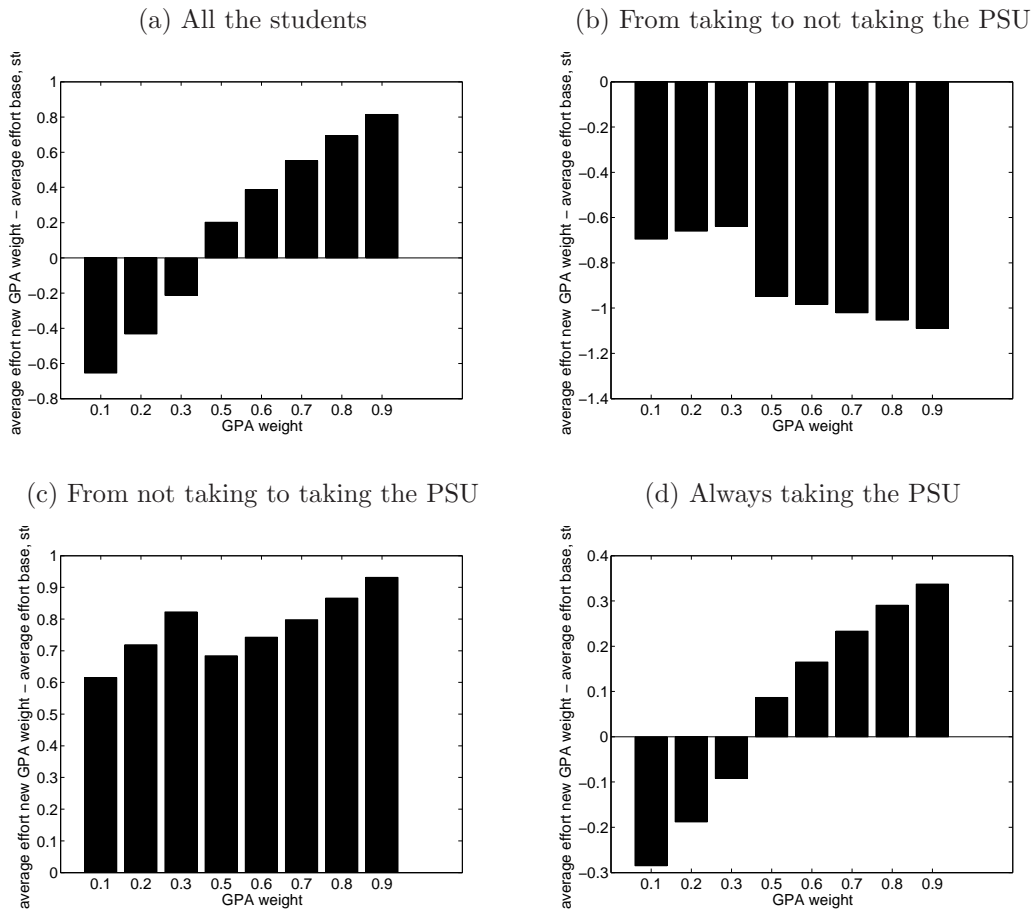
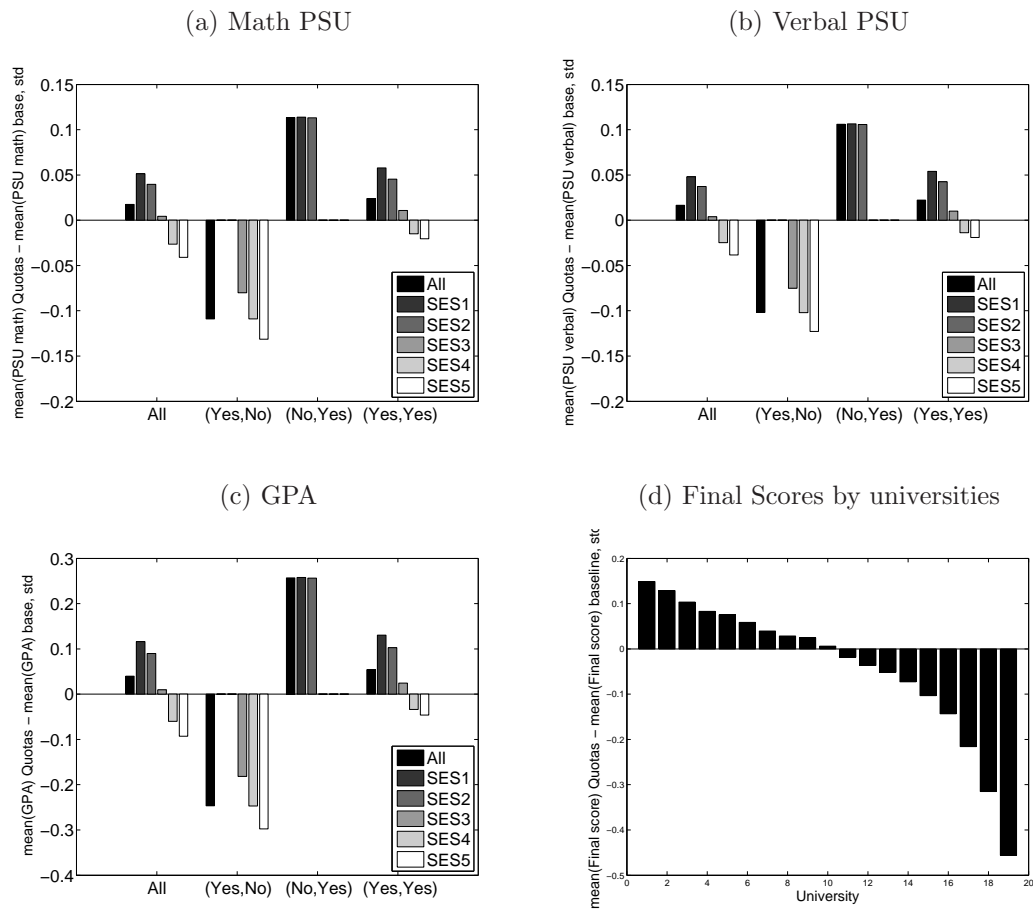


Figure A.10: Impact of Quota system on tests by SES and universities



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 A.11: 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

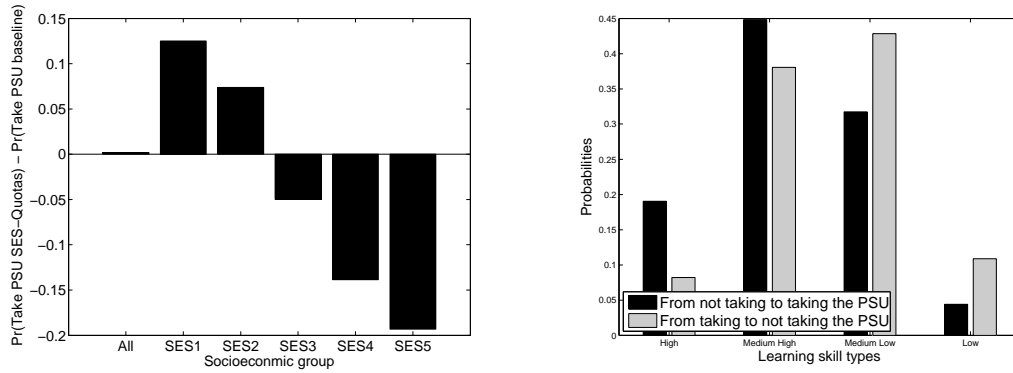


Figure A.12: Impact of changing GPA weight on who is taking the PSU

- (a) Change in the fraction of student taking the PSU by SES
 (b) Impact on the PSU-takers' learning skill distribution

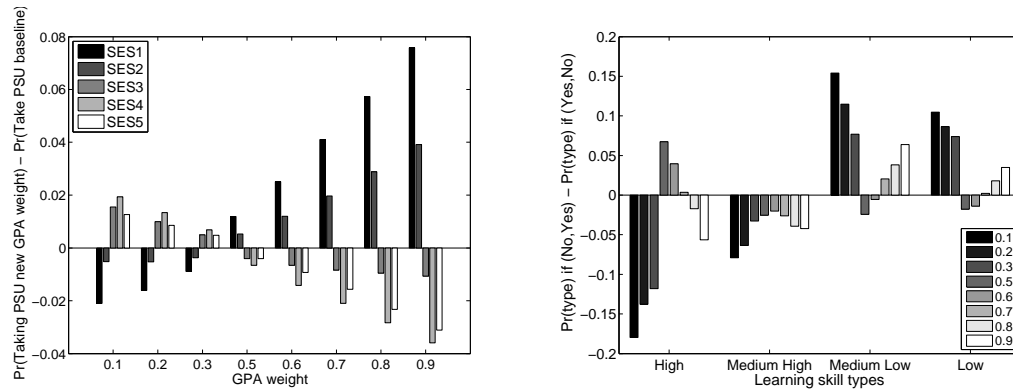


Figure A.13: The impact on final-score cutoff and college admissions of introducing SES-Quota system, with and without endogenous effort

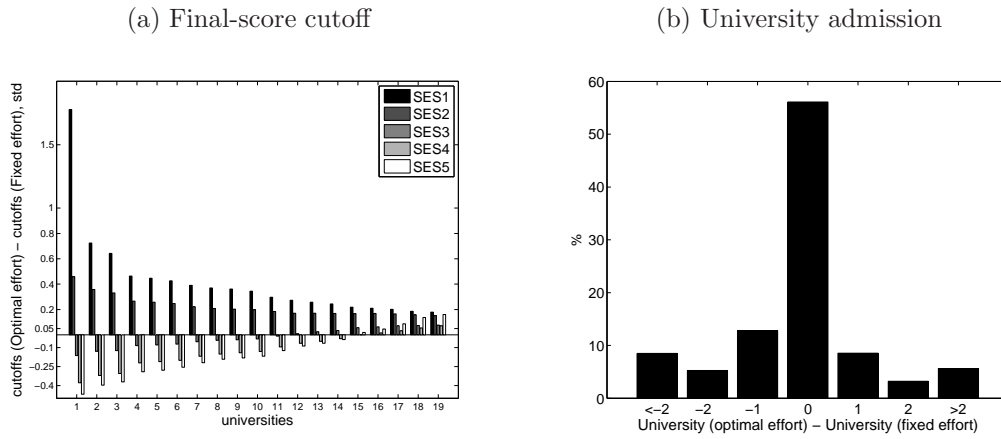


Figure A.14: The impact on final-score cutoff of changing the GPA weight, with and without endogenous effort

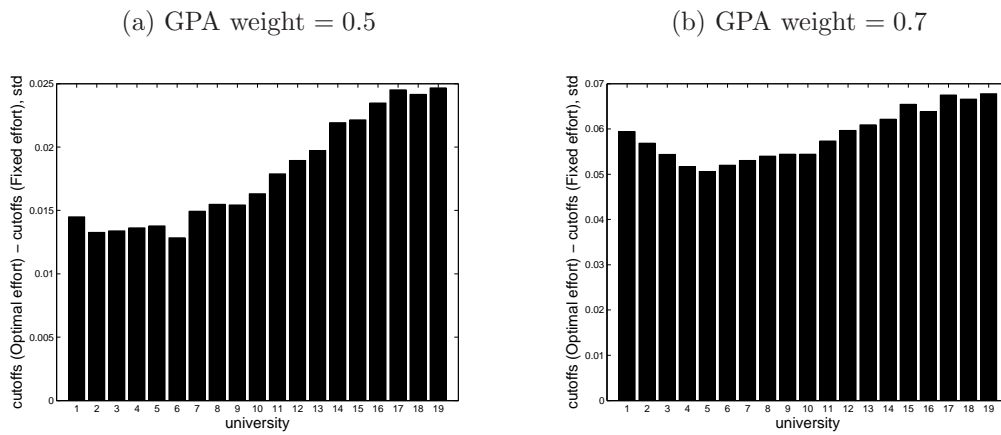
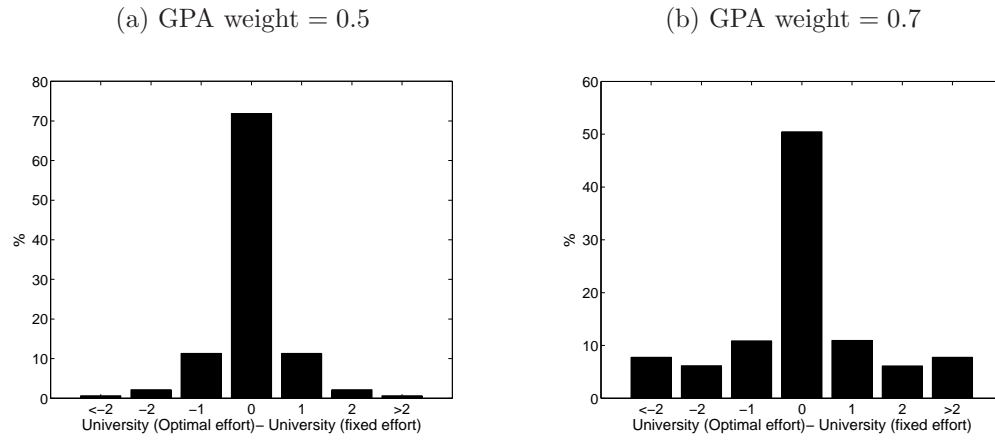


Figure A.15: The impact of changing the GPA weight on university admissions, with and without endogenous effort



Appendix B

A Dynamic Model of Elementary School Choice

B.1 Figures and Tables

Table B.1: Fraction of students changing school

	2nd	3th	4th	5th	6th	7th	8th	N
All	0.06	0.06	0.05	0.05	0.05	0.07	0.04	77432
Mother's education								
Incomplete elementary school	0.04	0.04	0.04	0.04	0.03	0.04	0.03	6962
Complete elementary school	0.06	0.05	0.04	0.05	0.03	0.05	0.03	8182
Incomplete secondary school	0.06	0.05	0.05	0.05	0.05	0.06	0.04	11085
Complete secondary school	0.06	0.06	0.05	0.05	0.05	0.08	0.04	29667
Complete or incomplete college	0.06	0.07	0.06	0.06	0.06	0.09	0.06	21536

Note: I drop the students who fail at least one class between 1st and 8th grade, and I also drop the students whose families switch the municipality where they live.

Table B.2: Total school change

	0	1	2	>2	N
All	0.69	0.24	0.05	0.01	77432
Mother's education					
Incomplete elementary school	0.78	0.19	0.03	0.01	6962
Complete elementary school	0.75	0.20	0.04	0.01	8182
Incomplete secondary school	0.72	0.22	0.05	0.01	11085
Complete secondary school	0.68	0.26	0.05	0.01	29667
Complete or incomplete college	0.65	0.27	0.07	0.02	21536

Note: I drop the students who fail at least one class between 1st and 8th grade, and I also drop the students whose families switch the municipality where they live.

Table B.3: Descriptive Statistics of the Two Samples

	Estimation Sample	Complete Sample
Parents with Low Education	0.209	0.187
Parents with Medium Education	0.428	0.384
Parents with High Education	0.363	0.429
Religious Family	0.212	0.260
Parents Attend School Meeting*	0.946	0.948
Parents Explain Class Material**	0.469	0.474
Parents Help Student Study***	0.402	0.391
Parents Care about Quality****	0.377	0.395
N	9752	19819

All the figures are proportions.

(*) From the question: How often do you attend the periodic parents' meeting of your child's class? I construct a dummy variable that takes one if parentst always or almost always attend, and zero otherwise.

(**) From the question: How often does one of your parents explain you the class material that you don't understand? I construct a dummy variable that takes one if the answer is always or almost always attend, and zero otherwise.

(***) From the question: How often does one of your parents help you to study? I construct a dummy variable that takes one if the answer is always or almost always attend, and zero otherwise.

(****) From the question: Name the first three reasons why you chose your child's current school.; I construct a dummy variable that takes one if parentst name school quality as one of the reasons, and zero otherwise.

B.2 Estimated Parameters and Model's Inputs

Model's Inputs and Parameters Estimated Outside the Model

Figure B.1: Histogram of the Size of the Choice Sets

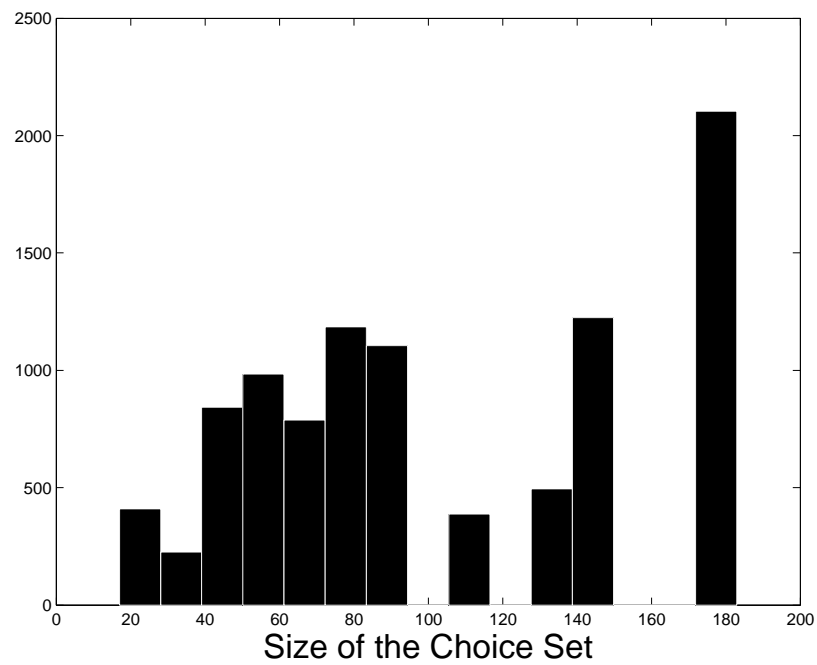


Figure B.2: Distribution of Estimated Quality by School Type in 2004

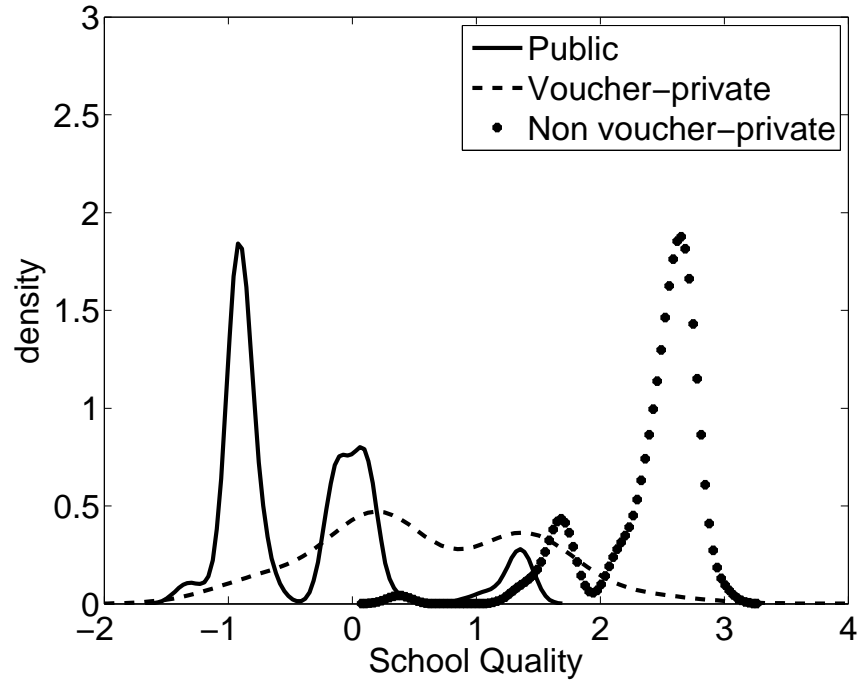


Table B.4: Knowledge Production Function

	Coeff	Std Err.
Accumulated School Quality at 8th Grade	1.07	0.0124
Parents with Medium Level of Education	5.52	0.2523
Parents with High Level of Education	8.72	0.3007
Constant	255.49	0.2047

$N = 55,421$

$R^2 = 0.2113$

Note: This is the result of estimating the Knowledge Production Function

$$\tilde{K}_{ijT} = \alpha_0 X_i + \alpha_1 \sum_{t=1}^T \tilde{q}_{ijt} + \vartheta_{ijT}.$$

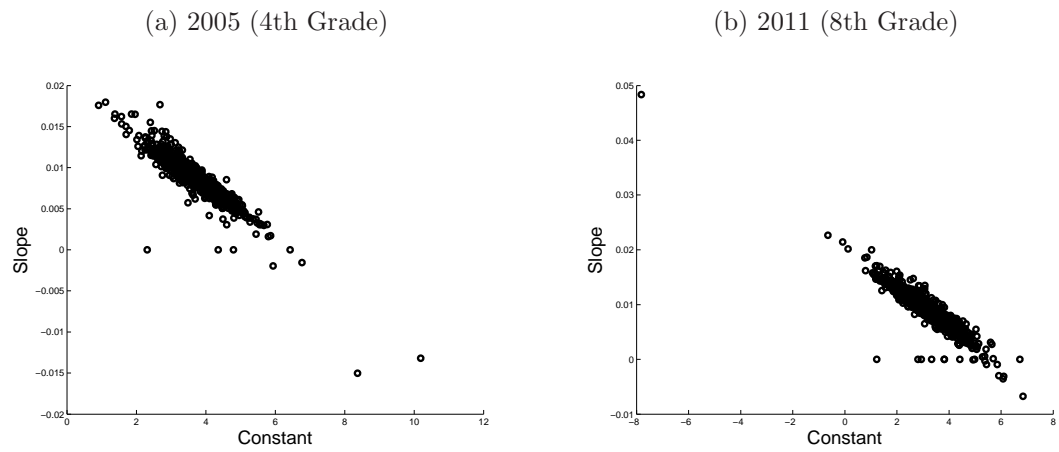
The omitted dummy variable is Parents with Low Level of Education.

Table B.5: GPA Production Function

	4th 2005		4th 2006		4th 2007		4th 2008		4th 2009		8th 2011	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
Entered in 2nd	-0.07	0.0034	-0.06	0.0034	-0.06	0.0038	-0.06	0.0039	-0.07	0.0034	-0.05	0.0054
Entered in 3th	-0.08	0.0033	-0.08	0.0033	-0.08	0.0036	-0.07	0.0038	-0.07	0.0033	-0.04	0.0050
Entered in 4th	-0.11	0.0030	-0.11	0.0030	-0.12	0.0035	-0.10	0.0037	-0.11	0.0032	-0.06	0.0048
Entered in 5th											-0.06	0.0045
Entered in 6th											-0.08	0.0045
Entered in 7th											-0.10	0.0040
Entered in 8th											-0.17	0.0046
Constant	3.63	0.0063	3.72	0.0062	3.74	0.0066	3.75	0.0066	3.72	0.0065	3.02	0.0097
N	214,661		210,177		162,552		169,500		178,685		161,920	
Adjusted R-squared	.549		.545		.559		.537		.558		.465	

Note: This is the result of estimating the GPA production function $G_{ijt} = \lambda_{0j}^t + \lambda_{1j}^t K_{ijt} + \lambda_2^t TA_{it} + \varepsilon_{ijt}^g$, where I only present the values for λ_2^t . The omitted dummy variable is Entered in 1st grade.

Figure B.3: GPA Standards



Note: Given the GPA production function $G_{ijt} = \lambda_{0j}^t + \lambda_{1j}^t K_{ijt} + \lambda_2^t TA_{it} + \varepsilon_{ijt}^g$; λ_{0j} is the constant and λ_{0j} the slope.

Parameters estimated by Simulated Maximum Likelihood

Table B.6: Parameters of the Utility Function

	Coeff	Std Err.		Coeff	Std Err.
Student Knowledge in $t < T$ (β_{kt})	0.0000	0.0045	Student Knowledge in T (β_{kT})	0.0546	0.0184
$\Delta\beta_{kt}$ of Involved Parents	0.0000	0.0119	$\Delta\beta_{kT}$ of Involved Parents	0.0024	0.0453
Medium Low SES School	0.0376	0.0224	Medium SES School	0.0960	0.0232
Medium High SES School	0.0972	0.0268	High SES School	0.1200	0.0379
School Tuition Fee	-0.0140	0.0068	Voucher-Private School	-0.0006	0.0053
Non Voucher-Private School	0.0611	0.0218	Religious School * Religious Parents	0.2525	0.0173
Student GPA	0.0311	0.0500	GPA Correction $G_{ijt-1} - \hat{G}_{ijt-1}$	0.0478	0.0081
Home and School in Different Municipalities	-0.3454	0.0056	School Offers Secondary Education	0.0402	0.0354
GPA at time T	0.0600	0.4542	Years Enrolled in Current School (at T)	0.0838	0.0091
Switch Cost	-5.9096	0.0286	Public School (2nd cycle)	1.4814	0.0527

Table B.7: Parameters of the Admission Probability Function

	Coeff	Std Err.		Coeff	Std Err.
Constant (1st Grade)	-0.3150	0.0133	School Quality * Sel^k (1st Gr.)	-0.1000	0.0125
School Quality * Sel^g (1st Gr.)	-0.2107	0.0183	School Quality * Sel^o (1st Gr.)	-0.0964	0.0199
Sel^k (1st Gr.)	-0.0614	0.0103	$K * Sel^k$ (1st Gr.)	0.0033	0.0001
Sel^m (1st Gr.)	-0.0050	0.0111	$MR * Sel^m$ (1st Gr.)	0.0006	0.0055
Sel^r (1st Gr.)	-0.0280	0.0121	Religious* Sel^r (1st Gr.)	0.0009	0.0213
Sel^o (1st Gr.)	-0.0221	0.0082	Tuition Fee * Low Education (1st Gr.)	-1.9624	0.2115
Tuition Fee * Medium Education (1st Gr.)	-1.0137	0.0263	Tuition Fee * High Education (1st Gr.)	-0.1745	0.0203
School selectivity(+) * Low Ed. (1st Gr.)	-2.2012	0.2312	School selectivity * Medium Ed. (1st Gr.)	-0.9976	0.0842
School selectivity * High Ed. (1st Gr.)	0.0017	0.0073	Constant (2nd-8th Grade)	-0.7455	0.0172
School Quality * Sel^k (2nd-8th)	-0.1000	0.0123	School Quality * Sel^g (2nd-8th)	-0.1888	0.0141
School Quality * Sel^o (2nd-8th)	-0.0999	0.0160	Sel^k (2nd-8th)	-0.0525	0.0158
$K * Sel^k$ (2nd-8th)	0.0002	0.0001	Sel^g (2nd-8th)	-0.0200	0.0067
$GPA * Sel^g$ (2nd-8th)	0.0010	0.0007	Sel^m (2nd-8th)	-0.0103	0.0006
$MR * Sel^m$ (2nd-8th)	0.0020	0.0027	Sel^r (2nd-8th)	-0.0200	0.0102
Religious* Sel^r (2nd-8th)	0.0019	0.0655	Sel^o (2nd-8th)	-0.0100	0.0092
New School (2nd-8th)	1.0087	0.3220	New School * School Size (2nd-8th)	0.0025	0.0029
Tuition Fee * Low Education (2nd-8th)	-0.7037	0.1729	Tuition Fee * Medium Education (2nd-8th)	-0.5036	0.0460
Tuition Fee * High Education (2nd-8th)	-0.1967	0.0160	School selectivity(++) * Low Ed. (2nd-8th)	0.4190	0.1552
School selectivity * Medium Ed. (2nd-8th)	0.1221	0.0052	School selectivity * High Ed. (2nd-8th)	-0.0058	0.0087

(+) School Selectivity at first grade is defined as $(Sel^k + Sel^o)/2$.

(++) School Selectivity between 2nd and 8th grade is defined as $(Sel^k + Sel^o + Sel^g)/3$.

Table B.8: Parameters that Determine Parents' Quality Perception

	Coeff	Std Err.		Coeff	Std Err.
Low Education (Math SIMCE)	0.0003	0.6083	Medium Education (Math SIMCE)	0.0094	0.0563
High Education (Math SIMCE)	0.0003	0.3222	Size of the Choice Set (Math SIMCE)	0.0074	0.0006
Low Education (Spanish SIMCE)	0.0000	0.5754	Medium Education (Spanish SIMCE)	0.0002	0.1638
High Education (Spanish SIMCE)	0.0002	0.1724	Size of the Choice Set (Spanish SIMCE)	0.0000	0.0000

Table B.9: Parameters of the Parents Involment Probability Function

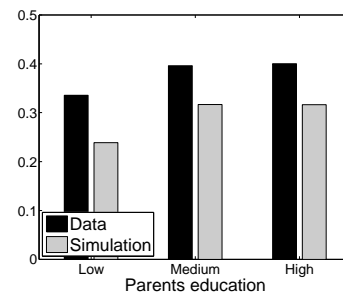
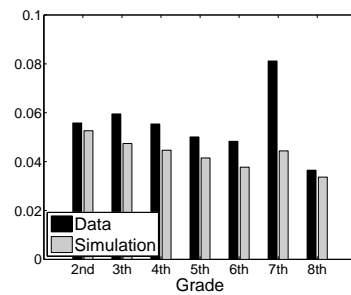
	Coeff	Std Err.		Coeff	Std Err.
Constant	-0.8873	0.1223	Parents Attend School Meetings	0.2544	1.4731
Parents Explain Class Material	0.0248	0.2497	Parents Help to Study	-0.2164	0.2939
Parents Care about Quality	0.2335	0.2577			

B.3 Model Fit and Counterfactual Experiments (Complete Sample)

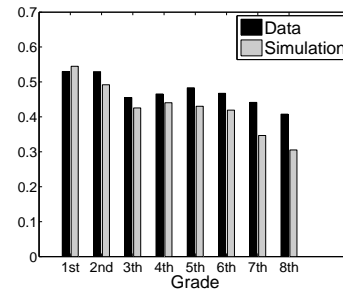
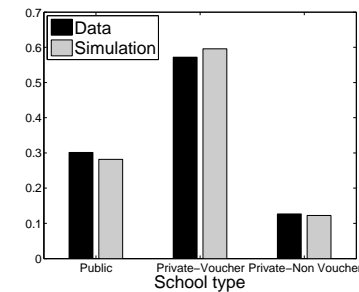
Model Fit

Figure B.4: Model fit for the complete sample

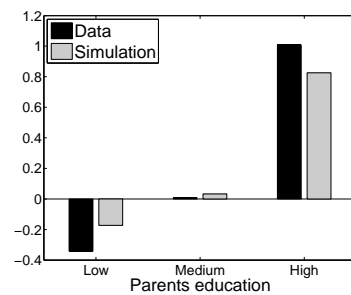
- (a) Fraction of student changing their school by grade
- (b) Average total change by parents education



- (c) Student fraction by school type
- (d) Average quality of the school selected by grade



- (e) Gain in Knowledge by parents education



Counterfactual Experiments

Table B.10: Fraction of students changing school by grade (with respect to baseline in percentage points), complete sample.

	Baseline	No misperception	Only Q maters	All admitted	Random admission	$C * 0.9$	New locations	All $ED = 3$
2nd	5.3%	-0.2	0.4	10.5	0.2	3.7	-1.0	-0.1
3th	4.8%	-0.1	1.0	10.1	0.2	3.7	-0.8	-0.1
4th	4.4%	-0.1	1.3	9.9	0.2	3.4	-0.7	-0.1
5th	4.1%	-0.1	1.8	9.7	0.2	3.3	-0.6	-0.1
6th	3.7%	-0.1	1.9	9.2	0.2	3.1	-0.7	-0.1
7th	4.4%	-0.2	1.6	8.5	0.0	3.3	-1.0	-0.2
8th	3.4%	-0.1	2.3	8.2	-0.1	2.7	-0.6	-0.2

Table B.11: Student fraction by school type at first grade (with respect to baseline in percentage points), complete sample.

	Baseline	No misperception	Only Q maters	All admitted	Random admission	$C * 0.9$	New locations	All $ED = 3$
Public	27.9%	0.9	2.3	-3.3	-2.3	0.7	0.1	-2.8
Voucher Private	59.4%	-0.3	-2.9	-0.6	-0.7	-0.2	9.9	1.9
Non voucher Private	12.6%	-0.6	0.7	3.9	3.0	-0.5	-9.9	0.9

Table B.12: Gain of knowledge ($KT - K0$) by parents education (with respect to baseline in standard deviations), complete sample.

	Baseline	No misperception	Only Q maters	All admitted	Random admission	$C * 0.9$	New locations	All $ED = 3$
Low educated	-0.206	-0.016	0.021	0.099	0.053	0.001	0.130	0.071
Medium educated	0.042	-0.029	0.103	0.205	0.127	-0.003	-0.003	0.133
High educated	1.007	-0.060	0.168	0.152	0.069	-0.027	-0.327	-0.014

Bibliography

- AGUIRREGABIRIA, V., AND P. MIRA (2010): “Dynamic Discrete Choice Structural Models: A Survey,” *Journal of Econometrics*, 156(1), 38 – 67, Structural Models of Optimization Behavior in Labor, Aging, and Health.
- ALDERMAN, H., P. F. ORAZEM, AND E. M. PATERNO (2001): “School Quality, School Cost, and the Public/Private School Choices of Low-Income Households in Pakistan,” *The Journal of Human Resources*, 36(2), pp. 304–326.
- ANGRIST, J., E. BETTINGER, AND M. KREMER (2006): “Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia,” *American Economic Review*, 96(3), 847–862.
- ARCIDIACONO, P. (2005): “Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?,” *Econometrica*, 73(5), 1477–1524.
- AUGUSTE, S., AND J. P. VALENZUELA (2006): “Is It Just Cream Skimming? School Vouchers in Chile,” Discussion paper.
- AZMAT, G., AND J. GARCIA-MONTALVO (2012): “The Role of Awareness, Information Gathering and Processing in School Choice,” Working Papers 639, Barcelona Graduate School of Economics.
- BAST, J. L., AND H. J. WALBERG (2004): “Can parents Choose the Best Schools for their Children?,” *Economics of Education Review*, 23(4), 431 – 440, Special Issue In Honor of Lewis C. Solman.

- BELLEI, C. (2009): “The Private-Public School Controversy: The Case of Chile,” in *School Choice International*, ed. by P. Peterson, and R. Chakrabarti, pp. 165–192. Cambridge, MA: MIT Press.
- BETTINGER, E. (2011): “Chapter 7 - Educational Vouchers in International Contexts,” in *Handbook of The Economics of Education*, ed. by S. M. Eric A. Hanushek, and L. Woessmann, vol. 4 of *Handbook of the Economics of Education*, pp. 551 – 572. Elsevier.
- BLACK, S. E. (1999): “Do Better Schools Matter? Parental Valuation of Elementary Education,” *The Quarterly Journal of Economics*, 114(2), 577–599.
- 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.
- BRAVO, D., S. MUKHOPADHYAY, AND P. E. TODD (2010): “Effects of School Reform on Education and Labor Market Performance: Evidence from Chile’s Universal Voucher System,” *Quantitative Economics*, 1(1), 47–95.
- 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.
- CARNEIRO, P., K. T. HANSEN, AND J. J. HECKMAN (2003): “2001 Lawrence R. Klein Lecture Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice,” *International Economic Review*, 44(2), 361–422.

- CHAN, J., AND E. EYSTER (2003): “Does Banning Affirmative Action Lower College Student Quality?,” *American Economic Review*, 93(3), 858–872.
- CHUMACERO, R. A., D. GÓMEZ, AND R. D. PAREDES (2011): “I Would Walk 500 Miles (if it Paid): Vouchers and School Choice in Chile,” *Economics of Education Review*, 30(5), 1103–1114.
- CONLISK, J. (1996): “Why Bounded Rationality?,” *Journal of Economic Literature*, 34(2), 669–700.
- CONTRERAS, D., P. SEPULVEDA, AND S. BUSTOS (2010): “When Schools Are the Ones that Choose: The Effects of Screening in Chile,” *Social Science Quarterly*, 91(s1), 1349–1368.
- CUNHA, F., J. HECKMAN, AND S. NAVARRO (2005): “Separating Uncertainty from Heterogeneity in Life Cycle Earnings,” *Oxford Economic Papers*, 57(2), 191–261.
- 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.
- DRAGO, J. L., AND R. PAREDES (2011): “La Brecha de Calidad en la Educación Chilena,” *Cepal Review*, 104, 167–180.

- ECKSTEIN, Z., AND K. I. WOLPIN (1989): “The Specification and Estimation of Dynamic Stochastic Discrete Choice Models: A Survey,” *The Journal of Human Resources*, 24(4), pp. 562–598.
- (1999): “Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities,” *Econometrica*, 67(6), 1295–1340.
- ELACQUA, G. (2012): “The Impact of School Choice and Public Policy on Segregation: Evidence from Chile,” *International Journal of Educational Development*, 32(3), 444 – 453.
- EPPLE, D., AND R. ROMANO (2008): “Educational Vouchers and Cream Skimming,” *International Economic Review*, 49(4), 1395–1435.
- (2012): “Economic Modeling and Analysis of Educational Vouchers,” *Annual Review of Economics*, 4(1), 159–183.
- 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.
- EPPLE, D., AND R. E. ROMANO (1998): “Competition between Private and Public Schools, Vouchers, and Peer-Group Effects,” *The American Economic Review*, 88(1), pp. 33–62.
- 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.

- GALLEGO, F. A., AND A. E. HERNANDO (2008): “On the Determinants and Implications of School Choice: Semi-Structural Simulations for Chile [with Comments],” *Economía*, 9(1), pp. 197–244.
- GAURI, V. (1999): *School Choice In Chile: Two Decades of Educational Refor*, Pitt Latin American Studies. University of Pittsburgh Press.
- GEYER, J., AND H. SIEG (2013): “Estimating a Model of Excess Demand for Public Housing,” *Quantitative Economics*, 4(3), 483–513.
- HANUSHEK, E. A., J. F. KAIN, AND S. G. RIVKIN (2004): “Disruption versus Tiebout Improvement: the Costs and Benefits of Switching Schools,” *Journal of Public Economics*, 88(9?10), 1721 – 1746.
- 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.
- HASTINGS, J. S., AND J. M. WEINSTEIN (2008): “Information, School Choice, and Academic Achievement: Evidence from Two Experiments,” *The Quarterly Journal of Economics*, 123(4), 1373–1414.
- 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.
- HENIG, J. (1996): “The Local Dynamics of Choice: Ethnic Preferences and Institutional Responses,” in *Who chooses? Who loses? Culture, institutions, and the unequal effects of school choice*, ed. by R. E. B. Fuller, and G. Orfield, pp. 93–117. New York: Teachers College Press.
- 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.
- HOUSER, D., M. KEANE, AND K. MCCABE (2004): “Behavior in a Dynamic Decision Problem: An Analysis of Experimental Evidence Using a Bayesian Type Classification Algorithm,” *Econometrica*, 72(3), 781–822.
- HSIEH, C.-T., AND M. URQUIOLA (2006): “The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile’s Voucher Program,” *Journal of Public Economics*, 90(8-9), 1477–1503.
- 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.
- KANE, T. J., S. K. RIEGG, AND D. O. STAIGER (2006): “School Quality, Neighborhoods, and Housing Prices,” *American Law and Economics Review*, 8(2), 183–212.
- KEANE, M. P., AND K. I. WOLPIN (1994): “The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence,” *The Review of Economics and Statistics*, 76(4), 648–72.
- LARRAÑAGA, O. (2004): “Competencia y Participación Privada: La Experiencia Chilena en Educación,” Working Papers wp208, University of Chile, Department of Economics.
- 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.
- MARK, G. ELACQUA, AND J. BUCKLEY (2006): “School Choice in Chile: Is it Class or the Classroom?,” *Journal of Policy Analysis and Management*, 25(3), 577–601.
- McMILLAN, R. (2004): “Competition, Incentives, and Public School Productivity,” *Journal of Public Economics*, 88(9?10), 1871 – 1892.
- MECKES, L., AND R. CARRASCO (2010): “Two decades of SIMCE: an Overview of the National Assessment System in Chile,” *Assessment in Education: Principles, Policy and Practice*, 17(2), 233–248.
- MIZALA, A., AND P. ROMAGUERA (2000): “School Performance and Choice: The Chilean Experience,” *The Journal of Human Resources*, 35(2), pp. 392–417.
- MIZALA, A., P. ROMAGUERA, AND M. URQUIOLA (2007): “Socioeconomic Status or Noise? Tradeoffs in the Generation of School Quality Information,” *Journal of Development Economics*, 84(1), 61–75.
- MIZALA, A., AND M. URQUIOLA (2013): “School markets: The impact of Information Approximating Schools’ Effectiveness,” *Journal of Development Economics*, 103(0), 313 – 335.
- NAVARRO, S. (2011): “Using Observed Choices to Infer Agent’s Information: Reconsidering the Importance of Borrowing Constraints, Uncertainty and Preferences in College Attendance,” Discussion paper.
- NEILSON, C. (2013): “Targeted Vouchers, Competition among Schools, and the Academic Achievement of Poor Students,” Discussion paper.
- 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.
- ROTHSTEIN, J. M. (2006): “Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions,” *American Economic Review*, 96(4), 1333–1350.
- RUST, J. (1994): “Structural Estimation of Markov Decision Processes,” in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4 of *Handbook of Econometrics*, chap. 51, pp. 3081–3143. Elsevier.
- SCHNEIDER, M., P. TESKE, M. MARSHALL, AND C. ROCH (1998): “Shopping for Schools: In the Land of the Blind, The One-Eyed Parent May be Enough,” *American Journal of Political Science*, 42(3), pp. 769–793.
- 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.
- SIMON, H. A. (1986): “Rationality in Psychology and Economics,” in *In Rational Choice: The Contrast Between Economics and Psychology*, ed. by R. M. Hogarth, and M. W. Reader, pp. 25–40. Chicago: University of Chicago Press.
- 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.
- VALENZUELA, J. P., C. BELLEI, AND D. D. L. RÍOS (2014): “Socioeconomic School Segregation in a Market-Oriented Educational System. The Case of Chile,” *Journal of Education Policy*, 29(2), 217–241.
- 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.
- ZAMORA, G. (2011): “Movilidad escolar en Chile: Análisis de las Implicancias para la Calidad y Equidad de la Educación.” *Estudios pedagógicos (Valdivia)*, 37, 53 – 69.