A Work Project, presented as part of the requirements for the Award of a

Masters Degree in Economics from the NOVA – School of Business and Economics.

DETERMINANTS OF SUCCESS OF NOVA SBE'S UNDERGRADUATE STUDENTS

DINO MIGUEL ALVES

STUDENT NUMBER 607

A project carried out on the Applied Policy Analysis major, under the supervision of:

Ana Balcão Reis, Maria do Carmo Seabra e Luís Catela Nunes

September 15th, 2014

Determinants of Success of Nova SBE's Undergraduate Students

Abstract

This Work Project analyses the determinants of academic success of undergraduate students enrolled in 2009/10 at Nova School of Business and Economic, using information about students' characteristics and skills to predict their future educational success, measured by four different outcomes. The main finding is that internal high school grade is a better predictor of achievement than the score of mathematics national exam. Additionally, gender, age, economic background and distance between home and Nova SBE seem to be correlated with performance. On the other hand, the estimates suggest that attending a private high school or following the economics track at high school are irrelevant for predicting educational outcomes, after controlling for the internal high school grade and the mathematics national exam score.

JEL classification: I21, 129

Keywords: predictors of success; higher education; national exams; cumulative process of education.

Acknowledgments

I would like to thank Professors Ana Balcão Reis, Maria do Carmo Seabra and Luís Catela Nunes for all the motivation and advice. I am also grateful to Dr. Bertolino Campaniço, who provided the most crucial element for this research: the data.

I dedicate this work to all those who stood by my side during my academic path, with whom I learned through this *cumulative process*.

Introduction

Founded in 1978, the Nova School of Business and Economics (Nova SBE) is a top Portuguese[1] public school in the areas of Economics, Management and Finance recognized by its excellence in teaching, outstanding research and international outlook. Amongst all the exceptional resources, the quality of Nova SBE's students[2] is certainly part of the formula for success; but is it possible to optimize it as an input?

From an administrative perspective, it is critical to know who are these students and why their performances differ after being accepted at Nova SBE. Therefore, this research aims at identifying the determinants of educational success at Nova SBE by understanding how its students' performance can be predicted by prior characteristics, observed at the application stage. Although it is difficult to infer any conclusions on causality, this predictive approach may constitute a key tool in order to improve management efficiency and to define a better "recruitment" strategy[3] at Nova SBE.

Success in education can be measured in terms of multiple outcomes such as students' personal satisfaction, first job's wage, post-graduate achievements, final Grade Point Average (GPA) or even time needed to graduate. Strongly focused on final GPA, this report also analyses the determinants of the number of ECTS completed in the first year,

[1] Portugal has a binary Higher Educational System with 57 Universities and 76 Polytechnics, a total of 133 public or private institutions. The National Agency for Assessment and Accreditation of Higher Education reports 3134 courses (Bachelors, Masters and PhDs) offered by public institutions and 1088 courses provided by private ones, (A3ES, 2012). According to Directorate General for Higher Education of Portugal (DGES), in 2009 there were fourteen institutions teaching Economics (1173 vacancies) and more than seventy offering Management and Administration (3760 vacancies) (DGES, 2010).

[2] Nova SBE admits about 400 bachelors' students each year. In 2009, as happened before, the vacancies were fully occupied and the average application grade of accepted candidates was about 17.2 out of 20, attesting the high standard of the school and its students. Furthermore, it was the institution with highest demand for Management (more than 460 students chose it as first option for only 210 available vacancies) and the second one in Economics (DGES, 2010).

[3] As part of the public system, Nova SBE cannot select or reject undergraduate students. Instead, it only defines the weight of each application criteria, within thresholds defined by the Ministry. In this case, the final application grade of candidates is computed as the simple mean between internal high school grade and the specific national exam score (mathematics).

students' performance in three core subjects and their probability of dropping out without any degree[4].

Assuming education is a result of a cumulative process, the following questions are addressed: First, is maths national exam score a reliable predictor of success, compared to the internal high school grade[5]? Secondly, do students from private schools perform better than those who attended the public system? And third, is there any evidence that the economics track at high school (supposedly the one tailor-made for those who wish to follow related subjects) is related to a better performance at Nova SBE? Answering these questions implies considering the influence of additional predictors such as personal characteristics, family and social background.

This work project is structured as follows. The next section explores the academic literature used as support. Section 3 presents the dataset and some descriptive statistics. The econometric procedures and methodological challenges are explained in Section 4, whereas the most relevant empirical results are discussed in Section 5. Finally, the last section considers possible extrapolations, suggests future extensions and ends with some conclusions.

Literature Review

Supported by econometrics, genetics and psychology, Economics of Education is a multidisciplinary field, which combines different approaches to study human capital and

[4] Besides being one of the EU countries with the lowest number of graduates, Portugal reports critically low survival rates in the first year of undergraduate degrees and also too long average durations for graduation. Consequently, students' drop-out and failure creates a significant inefficiency of public investment in Portuguese higher education. (Education at a Glance, OECD, 2013)

[5] The internal high school grade is an average of all subjects' grades taught during three years of high school, being much more related with a continuous work. However, it is difficult to compare students across different schools only based on this measure because it depends on specific school policies and it is more vulnerable to artificial grade-inflation than national exams.

its achievements. Assuming the link between wages, productivity and education, each individual derives his optimal demand for schooling by maximizing welfare under some personal, familiar, cultural and environmental constraints.

$$Y_{it} = F(Y_{it-1}, S_{it}, E_{it}, P_{it}, A_i, X_{it})^6 \quad [1]$$

Famous in economic literature, the equation above was proposed by Eric Hanushek (1979) who named it *educational production function,* drawing an analogy between the knowledge acquisition process of human beings and the production process of firms.

As Todd and Wolpin (2003) so clearly explained, achievement is the outcome of a cumulative education process in which current and past inputs are combined with individuals' genetic ability. Hence, learnings at university depend not only on school inputs[7] and personal characteristics but also on students' achievements at high school, measured at the point they apply to Nova SBE.

This work does not address the effects of current inputs while students are at Nova SBE. Instead, it is focused on determining the best measure of previous achievement, and on understanding how inputs in previous stages affect the final outcome. Similar questions were studied in a wide variety of literature, yet there is an extraordinary lack of consensus about these predictors of academic achievement.

An inspiring study on determinants of GPA for students of the University of California (San Diego) has been published by Betts and Morell (US, 1999), providing some support to college administrators in U.S. on the selection process of new applicants.

[6] In this conceptual model, Y_{it-1} measures the previous attainment (the student's situation at entrance), S_{it} indicates years of schooling, E_{it} stands for school resources (teachers' quality, class size, facilities, equipments, …), P_{it} reflects the quality of colleagues, A_i measures individual (unobservable) ability and X_{it} reflects any other dimensions such as parents' education or social background.

[7] Note that all individuals are Nova SBE students and there is no class-data; hence, they are assumed to have the same level of attendance, and to benefit from the same educational resources, school environment and peer effects. Yet, it is reasonable to assume that Bachelor partially contains this information, given that students tend to be assigned to different classes according to whether they are studying Economics or Management.

They confirm the relevance of standardized tests (SAT) and high school grades as predictors of final GPA (explaining, on average, about 28.25% of the predicted outcome). However, the inclusion of information on the student's personal background, the socioeconomic environment of high schools, and the quality of school resources significantly improves predictions of student's university GPA (R^2=0,3129). Moreover, they found that boys, racial minorities and students with poor high school environments perform relatively worse at university.

Smith and Naylor (UK, 2001) and more recently Marcenaro and Navarro (Spain, 2007) also investigated the role of high school results, family and social background on the success of British and Spanish university students. The common conclusions are the enormous importance of prior achievements (measured by A-level results for UK and *Selectividad* for Spain), the better performance of women and the high correlation between parents' academic or professional background and undergraduates' success[8]. Additionally, these papers support evidence that students who live with their parents and those who receive social support perform better.

As far as it was possible to assess, the literature does not contemplate any study specifically focused on predicting the success of economics and management bachelors, considering the whole curriculum.

On the other hand, much has been written about whether proficiency in mathematics determines achievement and success on introductory economic courses, a debate which is far from over. Becker and Watts (US, 1996) refer to an old survey indicating that instructors on the Principles of Economics "view students' ability to make numerical

[8] Hassink and Kiiver (Germany, 2007) estimates indicate that household income has a more important effect on the educational success of children at a more advanced stage in their education, while the effect of parental education (mainly mother's education) seems to be stable over time. However, the second effect tends to disappear if other measures of prior achievement are considered because they already include this effect, as a cumulative process.

calculations as important", believed that "algebra is somewhat important" and "graphs are extremely important". Surprisingly, those instructors did not consider calculus as an important skill for their course. A review by Cohn et al. (1998) lists a large number of studies which claim or reject a significant effect of having a maths background on the learning of economics[9]. A different problem concerns the reliability of test scores, such as the maths national exam, as a predictor of success[10]. Also investigated but equally controversial is whether following the economics track at high school affects performance in Principles of Economics courses[11].

Gender effects and type of schools are two additional dimensions that deserve to be highlighted. In contrast with the economics of education's benchmark, Anderson, Benjamin and Fuss (Canada, 1994), Lopus (US, 1997), Ballard and Johnson (US, 2004), Borg and Stranahan[12] (US, 2002), Parri and Aas (Estonia, 2006) all found a gender pattern: when controlling for academic and social background, male students perform better than female students in economics and business. Regarding the type of school, it is only included as a predictor of success among economics and management students

[9] Myatt and Waddell (US, 1990), Raimondo, Esposito and Gershenberg (US, 1990), Anderson, Benjamin and Fuss (Canada, 1994), Douglas and Sulock (US, 1995), Ballard and Johnson (US, 2004), claimed that students with stronger maths background perform better, while Brasfield, McCoy and Milkman (US, 1992), and Cohn et al. (US, 1999) did not find significant effects.

[10] Several critics point out the limitations of standardized test scores arguing that teachers may "teach to the test", and that these measures only evaluate students' ability to deal with a particular type of questions, ignoring creativity and deeper problem-solving skills. For instance, in Estonia, Parri and Aas (Estonia, 2006) found no statistical significance of national examination scores in predicting the performance of university students in economics and management.

[11] For e.g., Moyer and Paden (US, 1970), Myatt and Waddell (US, 1990) and Lopus (US, 1997) identified positive effects from prior studying of economics. In contrast, Palmer, Carliner and Romer (US, 1979), Reid (US, 1983) and Dolado and Morales (Spain, 2009) argued that those students who did not follow economics in high school are more likely to succeed in university introductory courses. These second results are counter intuitive but were explained by the lower attendance rate of students with economics background, the high level of work to which students from scientific fields are exposed and the inadequate program of economics taught at high school.

[12] Borg and Stranahan give a biological reasoning for this effect: "Females mature more rapidly than males causing them to develop higher verbal skills but lower spatial and numerical skills relative to males. Since the latter skills are more useful in the types of multiple choice tests that are used to measure economic knowledge, males appear to have an advantage in acquiring economic knowledge. This explanation may be the reason that the gender differential is maller in the studies that used essay and other types of questions to measure economic knowledge (p.590)"

by Dolado and Morales (Spain, 2009), who argue that having attended a private school improved the performance of first year students in Universidad de Madrid[13]. Horowitz and Spector (US, 2005) and Evans and Schwab (US, 1995) reached similar results with a special emphasis on religious schools, controlling for selection bias.

Finally, this research follows a previous duration analysis of the time needed to graduate among Nova SBE's students by Alves (2014), which suggests that students who enrol at a regular age and those who receive social support have a higher chance of graduating in the regular 3 year period. Moreover, high school grades and maths scores are also positively correlated with success, even though the coefficient of maths exams is much smaller. In contrast, displaced students tend to take more time to graduate.

Data and descriptive statistics

The dataset used in this research, gathered from three different sources (SIGES/CSE, CNA/DGES and SAS/UNL), refers to 363 undergraduate students of Economics and Management Bachelors, who enrolled through *Contingente Geral*[14] of the National Admission Process in 2009/10 at the Nova School of Business and Economics, Universidade Nova de Lisboa, Portugal.

For each individual, the relevant information[15] can be classified as (i) *Personal information*, which includes: gender, age, distance of residence from Nova SBE; (ii)

[13] In 2009, Universidade do Porto also published a report about the determinants of success of all its students. The most controversial result was the negative effect of attending a private high school, justified by a greater ability of public school's students to find solutions on their own; nevertheless there are concerns that these results are undermined by some econometric problems.

[14] The dataset includes 42 more students who enrol through different application procedures and 15 who were transferred from other schools or bachelor courses, but none of them are considered.

[15] The dataset includes also other variables such as nationality, scores obtained in national exams of portuguese and economics, school rankings, order of preference in applications, parents' job situation, purchasing power index of student's municipality, duration of Bachelors and Principles of Management grades. However, these variables are not included or analysed in this report because they do not present significance for predicting success.

Academic history: type of high school (public or private), internal high school grades, mathematics national exam scores and specialization track followed at high school (economics, sciences or other); (iii) *Nova SBE information*: bachelor's degree (management or economics), and round of enrolment[16]; (iv) *Social and economic background*: parents' educational level and value of scholarship attributed in case of social support; (v) *Academic output*: academic situation in 2013 (graduated, drop out or still studying), final GPA, number of ECTSs completed in the first year and grades obtained in core subjects' (Calculus I, Principles of Microeconomics and Principles of Macroeconomics).

Table 1 Descriptive statistics on students' characteristics

	Economics	Management	Total	%
Male	89	83	172	47.4%
Female	76	115	191	52.6%
Public *(high school)*	107	148	255	70.3%
Private *(high school)*	49	42	91	25.1%
Displaced Students *(more than 40 km from Nova)*	41	40	81	22.3%
Regional Students *(less than 40 km from Nova)*	124	158	282	77.7%
Regular Age *(18 years old or younger)*	149	165	314	86.5%
Older Age *(more than 18 years old)*	16	33	49	13.5%
Sciences Track *(at high school)*	46	69	115	31.7%
Economics Track *(at high school)*	108	107	215	59.2%
Other Track *(at high school)*	2	11	13	3.6%
International *(high school)*	1	11	12	3.3%
Parents with HE *(at least one has higher education)*	103	128	231	63.6%
Parents without HE *(no one has higher education)*	62	70	132	36.4%
1st Round *(application to Nova SBE)*	140	154	294	81.0%
2nd or 3r Round *(application to Nova SBE)*	27	42	69	19.0%
With Social Support	16	26	42	11.6%
No Social Support	149	172	321	88.4%
Total	165	198	363	100.0%

The majority of this population are students who don't receive social support, who finished high school within the regular time and enrolled at Nova SBE in the regular

[16] The Portuguese application procedure is national, based on *numerus clausus* filled by candidates in order of application grades. The procedure is divided into three rounds allowing students who weren't accepted for their first choices (or decided to change) to apply again with a new selection process to fill the empty vacancies.

round of applications. A significant share of this population followed the economics track at public high schools and lives close to Nova SBE with their parents, who have a higher education degree. Although both groups are fairly similar, there are some differences: girls seem to be more likely to choose Management, students with economic background from high school are more likely to choose Economics and so are students from private high schools.

Table 2 Descriptive statistics on educational outcomes

	All sample			Economics			Management		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Application Grade	17.2	15.1	19.5	17.4	15.1	19.5	17.1	15.6	19.5
High School Grade	16.8	13.5	20.0	17.2	13.5	19.3	16.5	13.9	20.0
Maths Exam Score	17.6	11.0	20.0	17.7	11.0	20.0	17.6	15.0	20.0
Final GPA (N = 270)[17]	14.2	11.4	18.6	14.3	12.1	18.6	14.1	11.4	16.7
Completed ECTS in 2009	43.4	0.0	67.5	46.3	0.0	67.5	40.9	0.0	60.0
Calculus (N= 327)[17]	13.4	10.0	20.0	13.5	10.0	20.0	13.2	10.0	19.0
Microeconomics (N=330) [17]	13.2	10.0	19.0	13.3	10.0	19.0	13.1	10.0	19.0
Macroeconomics (N=321) [17]	12.5	10.0	20.0	12.6	10.0	20.0	12.4	10.0	18.0

Compared with Economics, the last accepted student in Management had a higher application grade (15.05 and 15.60, respectively). However, by looking to the simple means (Table 2), it is clear that economics students who applied with better high school grades achieved a better performance at Nova SBE: they obtained higher average GPAs, completed more ECTSs in the first year and had faster academic paths. Moreover, Table A. 1 (Appendix) shows that the share of economics students who finished the first degree in 4 years period (83.6%) contrasts with the bad results of Management, with only two in each three students being able to graduate. The percentage of students who were still trying to finish Management in 2013/14 was twice as high as in Economics and the same is true of the dropout rate. The reason behind this is unclear, but it is not

[17] The number of individuals refers to those who graduate (270) or completed each core subject.

believable to assume Economics is easier since the first two years are very similar in both bachelors.

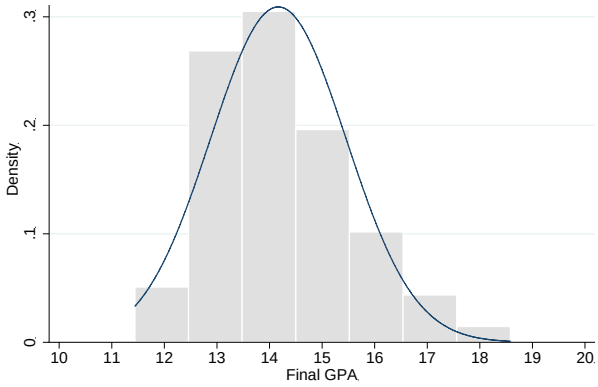
Some preliminary results are particularly suggestive. The first unexpected finding is the lack of correlation between internal high school grades and maths exam scores ($corr_{hsg,mes}$ = 0.077) [18]. Secondly, the high school grades seem to be more highly correlated with final GPA ($corr_{hsg,gpa}$ = 0.498) than the maths exam scores ($corr_{mes,gpa}$ = 0.301). Then, the third interesting feature is that the correlations of high school grade, maths exam score and final GPA are much higher among students from private schools[19], who also perform better in every dimension.

Table 3 and Figure 1 describe the distribution of students' final GPA, with a special emphasis on the quantile used later in the econometric models.

Table 3

Percentile	GPA
Min	11.44
10%	12.67
25%	13.14
50%	13.99
75%	15.02
90%	15.86
Max	18.58

Figure 1



Additionally, some interesting conclusions will be highlighted considering the density

[18] At 5 percent level, the correlation test does not reject the null hypothesis of no correlation. However, when only students who graduate are considered, the correlation slightly increases (0.117). Since they do not express the same information, high school grade and maths exam score should be tested for each outcome in order to obtain their individual and joint significance.
[One should keep in mind that these correlations are only small for this selected sample. It is not reasonable to believe that the same happens for the national population.]

[19]

Average Grades	Public	Private
High School Grade	16.71	17.07
Maths Exam Score	17.63	17.59
Final GPA	14.04	14.36

For private schools: $corr_{hsg,mes}$ = 0.125; $corr_{hsg,gpa}$ = 0.526; $corr_{mes,gpa}$ = 0.422.
For public schools: $corr_{hsg,mes}$ = 0.065; $corr_{hsg,gpa}$ = 0.494; $corr_{mes,gpa}$ = 0.302.

of Final GPA (Figure A. 1 displays these distributions considering nine different dimensions). Besides confirming that economics students performed much better than management ones, it seems that boys achieved better GPA than girls (at least in the distribution tails). Moreover, students older than 18 at the start showed worse results, as well as those who attended public high school, followed the economics track at high school, enrolled after the first round or whose parents don't have degrees. Assuming displaced students face more costs and adaptation problems, a sharper effect would be expected, however, the advantage of local students was only clear in the distribution tails. Likewise, students who receive social support don't showed significant differences, which probably means scholarships are promoting equality.

After analyzing the statistics, one must recognize that these students are not broadly representative of Portugal higher education's population. The majority of them are urban, born in Lisbon and come from a strong social-economic background with college-educated parents.

Methodology and econometric approaches

Following Todd and Wolpin (2003), this research uses a Value-Added specification to portray the cumulative process of knowledge acquisition in the following model:

$$y_{i0} = x_{i0}'\delta_1 + \delta_2 a_i + u_{i0} \quad [2]$$

$$y_{i1} = x_{i1}'\beta_1 + \beta_2 a_i + u_{i1} \quad [3]$$

,where y_{i0} $(i = 1, 2, \dots, n)$ represents student's attainment before entering university (i.e., high school grade and mathematics exam score) and y_{i1} is the outcome in analysis (e.g., final GPA). The vector of personal characteristics, family background and school inputs is given by x_{it} and a_i measures the unobserved characteristics (including innate

ability), assumed as constant over time. Residuals u_{it} are zero-mean i.i.d. and uncorrelated with regressors.

By solving for a_i in [2] and substituting in [3], one gets the university's outcome as a function of previous achievement, current and historical inputs:

$$y_{i1} = y_{i0}'\alpha_1 + x_{i1}'\alpha_2 + x_{i0}'\alpha_3 + w_{i1} \quad [4]$$

This model has limitations regarding endogeneity, which will be addressed later, but is completely compatible with the predictive intention of this research.

For estimation purposes, different regressors, controls and econometric approaches shall be used according to the characteristics of each outcome.

i. Ordinary Least Squares

The standard method used to predict final GPA and number of completed ECTS in the first year is the Ordinary Least Squares. The OLS estimation is a useful tool for summarizing the average relationship between an outcome variable and a set of regressors, based on the conditional mean $E(\mathrm{y}|\mathrm{x})$.

ii. Quantile Regression

In order to provide a more complete picture of final GPA and completed ECTS, the Quantile Regression (QR) is the suggested methodology. This approach gives information about the relationship between the outcome and the regressors at different quantiles or percentiles of the conditional distribution of the dependent variable. Compared with OLS, the QR estimator is less limited, much more robust to outliers, and seems to be especially suitable for heteroskedastic data in the sense that it avoids assumptions about parametric distribution of residuals.

According to Cameron and Trivedi (2005) , the purpose of this approach is to find the β_{q} that better fits in each quantile ($0 < q < 1$). While the OLS estimator minimizes the

sum of the square of residuals, the QR estimator minimizes the weighted sum of the absolute-deviation of residuals. In other words, considering $\varepsilon_{iq} = (y_i - x_i'\beta_q)$ as the residual, the q[th] QR estimator minimizes the objective function

$$\mathrm{Q}(\beta_\mathrm{q}) = \sum_{i:\varepsilon_{iq}\geq 0}^{N} q\left|\varepsilon_{iq}\right| + \sum_{i:\varepsilon_{iq}<0}^{N}(1-q)\left|\varepsilon_{iq}\right| \qquad [5]$$

If q = 0.1, for example, then much less weight is placed on prediction for observations with $y_i \geq x_i'\beta_{0.1}$ than for observations with $y_i < x_i'\beta_{0.1}$. The conditional median estimator, also used in this research, is a very particular case of this method (q=0.5)[20], obviously distinct from the conditional mean estimator (OLS).

iii. Tobit Model

When success is measured based on grades in core subjects, it includes only grades for the first year. So, if a student failed in his first attempt, a zero is reported. In such a case, although the data for regressors is observed, one only knows that the outcome is lower or equal to 9 out of 20, but the value itself is unobserved.

In this situation, with censored data, the Ordinary Least Squares (OLS) regression does not yield consistent parameter estimates because the sample has an artificial excess of zeros, which are not representative of the population. The parametric Tobit model, instead, assumes a standard normal distribution of residuals, $u_i \sim \mathcal{N}(0, \sigma^2)$, and computes the estimators considering the probability of each observation being censored.

The Tobit model considers the following structure:

$$y_{i1}^* = \alpha + x_i'\beta_1 + u_i \qquad y_{i1} = \begin{cases} y_{i1}^* & , y_{i1}^* > 9 \\ 0 & , y_{i1}^* \leq 9 \end{cases} \qquad [6]$$

[20] Minimization of the function $\mathrm{Q}(\beta_{0.5}) = \sum_i^N |y_i - x_i'\beta_{0.5}|$.

iv. Binary Outcome Models

Considering binary outcomes, just like the decision of dropping out (or not), requires an econometric approach that correlates the probability of leaving college with regressors.

The linear probability model (LPM), which follows a standard OLS, could be an appropriate estimation method, but it does not reflect some specific features of probabilities such as being always between 0 and 1. To overcome the nonsense of predicting probabilities outside this interval, the econometric theory suggests logit and probit models, which are computed by maximum likelihood and assume respectively the logistic and the standard normal distributions. With such procedures, the effect of any regressor on the probability of dropping out is not directly given by the coefficient but by the corresponding marginal effect.

v. Other Limitations

An important challenge in Economics of Education is how to address a major limitation for a researcher: missing data. With a very complete dataset and the right controls properly measured, the econometric procedures would produce unbiased and consistent estimators, which would express the effect of individual characteristics and previous attainment on the educational outcome. Ideally, it would be possible to infer conclusions on causality and even predict the effect of policy changes.

Unfortunately, the available dataset does not include enough information: besides of representing a very specific group of high-performing students (without any control group), it is clear that unobserved characteristics play an important role on both educational choices (inputs) and results (outcomes).

The first problem is the fact that those students who were not accepted at Nova SBE were not observed. Without any information about this group it is not possible to

evaluate the effect of changes to the recruitment policy. Moreover, these "untreated" students could perform differently if they had been accepted at Nova SBE; without a counterfactual it is difficult to evaluate a new policy. Heckman's selection model is a useful procedure to control for this selection bias, yet, it only works when the determinants of selection are observed in treated and untreated individuals, which is not the case.

The second concern is endogeneity, created by the omission of relevant variables (e.g., income or ability) which are not only directly correlated with current educational outcomes but also with previous achievement (included as a regressor). Analytically, this problem is observable in equation [4]: given that w_{i1} depends on u_{io}, which is correlated with y_{i1}, if $\alpha_1 \neq 1$ [21], the residuals are correlated with y_{i1} and it is not possible to disregard endogeneity. Consequently, the OLS estimates become biased and only an IV approach with proper instruments (which are not available) could solve the problem.

The final limitation to this analysis is related to students' choices and self-selection. In theory, if the structural model was changed by the implementation of a new recruitment policy, students would prepare themselves differently for final exams, and then the variables that measure previous achievement would also change.

Combined, these limitations create significant biases and undermine any conclusions about causality or policy evaluations. However, when it comes to forecasting it is possible to disregard endogeneity and selection biases. Besides, without changing the structural model, the analysis of each parameter clarifies its corresponding relevance and impact on the educational outcome.

[21] This inequality assumes that the effect of omitted variables on outcome varies over time ($\beta_2 \neq \delta_2$), which is a reasonable argument given that some inputs are much more likely to influence success in previews stages of educational process than others (Hassink and Kiiver, 2007)

Empirical results

This chapter presents the econometric models used to predict each educational outcome and discusses its results. Different combinations of inputs are used according the specific characteristics of the various measures of success, admitting different approaches and considering multiple questions.

For each measure of success, the baseline predictive model is the following:

$$success_i = \alpha + \beta_1 maths_i + \beta_2 hsgrade_i + X_i'\beta_3 + u_i\,, \qquad [7]$$

where $maths$ represents the score obtained by the student in the national maths exam, $hsgrade$ corresponds to the internal high school grade, and X is a vector of other controls (inputs and personal characteristics). As discussed before[17], $maths$ and $hsgrade$ are analysed jointly and separately for every outcome in order to understand which one best predicts success.

Still, the inclusion of additional covariates increases not only the predictive power of the model but also it helps to identify other determinants of success. Therefore, beyond the baseline model, each outcome will involve further analysis considering different combinations of regressors in order to define a benchmark model.

i. Outcome: Final GPA

Considering final GPA as the dependent variable, the first set of regressions contains the baseline split into regressions (1) and (2) and complete in regression (3). Regression (4) considers an interaction of the baseline with a dummy variable measuring the type of high school (pub assumes 1 for students who attended public high school and 0 otherwise). Finally, regression (5) is taken as benchmark and includes a larger range of characteristics, as follows: BSc identifies the student's bachelor's degree being 1 for

economics students and 0 for management; age equals to 1 for students enrolled with the regular age (18 years old) and 0 for older students; $male$ is the gender dummy and assumes the value 0 for women; $displ$(aced) identifies students whose original address is more than 40km away from Nova SBE; eco(nomics), int(ernational) and $other$ refer to high school tracks, the science track being used as the reference group. Thus, the reference group for this analysis are female management students born before 1991, who live close to Nova SBE and have a high school background in science.

Table 4. Determinants of Final GPA (OLS) [22]

	(1)	(2)	(3)	(4)	(5)
Maths	0.310***	-	0.253***	0.292***	0.280***
Hsgrade	-	0.561***	0.529***	0.528***	0.545***
Pub	-	-	-	-0.193	-
BSc	-	-	-	-	-0.164
Age	-	-	-	-	0.389*
Male	-	-	-	-	0.244*
Disp	-	-	-	-	-0.350**
Eco	-	-	-	-	0.002
Int	-	-	-	-	1.000***
Other	-	-	-	-	-0.126
Constant	8.696***	4.634***	0.706	0.156	-0.398
N	270	270	270	259	270
R^2	0.091	0.248	0.308[23]	0.338	0.364
Adj. R^2	0.087	0.245	0.303	0.330	0.342

Legend: * p<.1; ** p<.05; *** p<.01

The first clear result from Table 4 is that maths exam score and high school grade are both positively correlated with strong significance. Yet, the internal grade's coefficient is larger, which confirms stronger partial correlations (even when considering other controls) and higher predicting power over final GPA (measured by R^2=26.7%)[23].

The fourth regression analyses the effect of public versus private high schools on achievement but the extra dummy variable has no statistical significance, meaning that the type of school attended does not predict final GPA.

[22] In order to correct for heteroskedasticity (that is rarely not present on cross sectional data), this chapter's estimations use robust standard errors (except the logit model).

[23] Assuming that the correlation between Maths and Hsgrade is almost zero, it is possible to decompose the coefficient of determination (R^2) of regression (3). The result is that high school grade predicts 26.7% of changes on final GPA, while maths exam score only influences 5.3%.

Finally, the benchmark regression stresses the relevance of high school grades and maths exam scores but also considers other significant covariates such as age, gender and attendance of an international high school (positively correlated) and distance from Nova (negatively correlated).

According to these results, and considering everything else constant, studying at an international high school increases final GPA by one value, being the largest coefficient. The negative effect of being a displaced student was already expected, since it probably means living with less parental supervision and spending more time on domestic chores instead of studying. Due to the similarity between Economics and Management Bachelors, the insignificance of such coefficient is comprehensible. Although it is only significant at the 10 percent level, the age coefficient has the expected sign, assuming that students who followed a regular academic path, without taking breaks or failing years, are more likely to succeed.

Additionally, there is a counter-intuitive result regarding the lack of relevance of following the economics track or any other in high school, but this is also coherent with previous findings[11]. Also in line with the literature[12], male students tend to perform better than females, *ceteris paribus*.

The OLS regressions of Table 4 only consider average effects; nevertheless, it is important to analyse if the conclusions are consistent throughout the overall distribution. The quantile regressions of Table 5 evaluates the link between final GPA and the regressors at 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} quantiles in the conditional distribution of the dependent variable. From Table 5 it is possible to conclude that some results are consistent throughout the distribution while others depend on the quantile being analysed.

Table 5. Determinants of Final GPA (Quantile Regression)

Variable	OLS (5)	Q=10th	Q=25th	Q=50th	Q=75th	Q=90th
Maths	0.280***	0.166**	0.260**	0.294***	0.334***	0.355***
Hsgrade	0.545***	0.384***	0.469***	0.553***	0.631***	0.658***
BSc	-0.164	-0.428**	-0.219	-0.278	-0.118	-0.282
Age	0.389*	0.363	0.768**	0.624	0.430	0.404
Male	0.244*	-0.100	-0.093	0.132	0.386**	0.908***
Disp	-0.350**	-0.321	-0.470	-0.154	-0.455**	-0.505*
Eco	0.002	-0.162	-0.248	0.072	-0.022	0.227
Int	1.000***	0.682	0.769	0.905*	1.721***	1.577***
Other	-0.126	-0.495	0.965	0.193	-0.142	-0.631
Constant	-0.398	3.505	0.513	-0.976	-2.333	-2.574
N	270	270	270	270	270	270
Pseudo R^2	(0.364)	0.158	0.160	0.223	0.266	0.278

The coefficient for internal high school grade is almost twice the maths national exam for every quantile, being both always significant as well as growing at similar rates along the conditional distribution. In contrast, bachelor's degree is only relevant for the 1st quartile while gender, being displaced and having an international high school background seem to gain relevance among top students. The differences between OLS and QR coefficients (along the distribution) are described in Figure A. 2 (Appendix).

ii. Outcome: Grades of Core Subjects

As first year courses, Calculus, Microeconomics and Macroeconomics are core subjects common to both economics and management. Understanding which regressors better predict the performance in each course requires the use of Tobit Model (censored for values lower than 10).

Table 6 presents the baseline and the benchmark structures for each specific subject. With the exception of the baseline for Microeconomics, these models confirm that maths exam scores and internal high school grades are good predictors of success, the first coefficient being always lower than the second one, even for Calculus. These results are in line with the conclusions from Table 4. The benchmark model for the Macroeconomics grade deserves some attention, since it is the only case which shows some evidence of the relevance of studying economics at high school.

Table 6. Determinants by Core Course (Tobit)

Variables	Calculus		Microeconomics		Macroeconomics	
	(3)	(5)	(3)	(5)	(3)	(5)
Maths	0.592***	0.653***	0.223	0.309**	0.361***	0.440***
Hsgrade	0.956***	0.890***	0.937***	0.868***	1.201***	1.140***
BSc	-	-0.152	-	0.031	-	-0.338
Age	-	0.903*	-	0.474	-	0.303
Male	-	-0.068	-	0.703**	-	0.163
Disp	-	-1.051***	-	-0.166	-	-0.914**
Eco	-	0.058	-	0.078	-	0.838**
Int	-	-0.106	-	1.017	-	1.411**
Other	-	-1.408	-	-2.927***	-	-2.975**
Constant	-13.854***	-14.280***	-7.120**	-8.225***	-16.377***	-17.269***
Sigma	2.580***	2.498***	2.768***	2.687***	2.629***	2.527***
N	327	327	330	330	321	321
Pseudo R^2	0.052	0.066	0.032	0.045	0.065	0.085

iii. Outcome: Number of ECTS completed in the first year

Predictors of success may be different for the first and the final year, as students' performance in their freshman year is much more affected by environment and their ability to adapt to new circumstances.

The equation considers performance in the first year in terms of completed ECTS as a function of maths national exam score, internal high school grade, bachelor's degree, displacement and two new dummy variables[24]: $round$ refers to the enrolment stage and assumes 1 if the student was selected only in the second or third rounds, meaning that they missed a significant number of classes in the beginning of the first semester; and sas represents the allowance of public social support and is equal to one for students who received a scholarship in their first year (regardless of its amount).

As far as can be concluded from Table 7, maths national exam score does not predict performance in the first year at Nova. Meanwhile, the coefficient for internal grade is significant but loses influence among those students who completed more ECTS.

[24] Age, gender and track were not significant, while these new variables seem to be strongly correlated with the number of completed ECTS in the first year.

Table 7. Determinants of Completed ECTS in 2009 (OLS and Quantile Regression)

Variable	OLS	Q=10th	Q=25th	Q=50th	Q=75th	Q=90th
Maths	0.453	1.241	0.000	0.611	0.071	0.000
Hsgrade	4.622***	7.858***	6.282***	4.620***	2.262*	0.000
BSc	2.443	8.686*	3.025	-0.584	-0.580	-0.000
Disp	-7.616***	-14.228**	-11.308***	-8.010***	-5.521***	-0.000
Round	-9.126***	-6.425	-11.898***	-10.014***	-8.589**	-7.500*
SAS	2.142*	4.692	3.508**	1.791	1.011	0.000
Constant	-40.537**	-129.196**	-67.374***	-39.415*	18.372	60.000***
N	344	344	344	344	344	344
Pseudo R^2	(0.272)	0.170	0.214	0.183	0.112	0.018

With a significant negative effect, being a displaced student and enrolling after the first round reduces, on average, respectively 7.6 and 9.1 ECTS in the first year, assuming everything else is constant. Being displaced seems to particularly affect students who failed more courses (in the first year, each course corresponds to 7.5 ECTS and students are allowed to enrol in up to 70 ECTS per year).

On the other hand, there is evidence that receiving social support is positively correlated with a better performance in beginners. Although these students are the most needy in economic terms, Table 7 suggests that the incentive mechanism[25] behind the scholarship is effectively reducing the negative income effects.

iv. Outcome: Dropping out

The prediction of dropping out requires the use of a binary outcome model suitable for this type of dependent variable, which equals 1 for students who left Nova SBE without completing their bachelors and 0 for those who graduate within 4 years of study[26] . Table 8 presents the outputs of average marginal effects and marginal effects at means

[25] Students who receive social support must complete at least 60% of their annual ECTS to ensure the renewal of the scholarship.

[26] Students who were still finishing their degrees in 2013/14 were excluded from this analysis.

for two logit models[27]: the baseline model, and a more complex one with two new explanatory variables[28]: $saseur$ is the value of social support received in thousands of euros, and $feduc$ is a dummy variable identifying students whose fathers are college-educated.

Table 8. Probability of Dropping out (Logit Model)

Variables	AME		MEM	
	(1)	(2)	(1)	(2)
Maths	0.008	0.013	0.007	0.010
Hsgrade	-0.054***	-0.029*	-0.050***	-0.022*
BSc	-	-0.089**	-	-0.070**
Disp	-	0.122***	-	0.095***
Round	-	0.085**	-	0.066*
SASeur	-	-0.133*	-	-0.104**
Feduc	-	-0.060*	-	-0.047*
N	307	307	307	307
Pseudo R^2	0.058	0.156	0.058	0.156

Again, according to Table 8, the maths national exam score has no power in predicting the probability of dropping out, a result consistent with other measures of success. On the other hand, having an additional mark in internal grades seems to reduce the predicted probability of dropping out (by 5.4 percentage points for the baseline model). Surprisingly, students of economics are less likely to drop out, while those who enrolled after the first round face higher probabilities of leaving university without a degree, as do students who are displaced. Conversely, it appears that social support decreases the probability of dropping out by 13.3 percentage points for every thousand euros received. Finally, the academic background of the students' family, in this case the father's level of education, can also be considered a positive predictor of success which reduces the likelihood of quitting before graduation.

[27] Probit Models were also computed but the conclusions were very similar and presented a lower likelihood value.

[28] It also excludes irrelevant variables and includes these two because they are highly correlated with drop-out rate. One should note that $feduc$ refers only to the father, while the statistics presented on Table 1 refer to at least one of the parents. This is the egression in which father's education is statistically significant.

Conclusions

Considering a cumulative approach to education, this research analyses the determinants of success among Nova School of Business and Economics' undergraduates and uses academic and personal information about 363 students just before they entered Nova SBE, in 2009, to predict different measures of performance in their first year and at the time of graduation. The objective of such a work is to understand the most relevant prior characteristics for economics and management bachelors, and represents a useful tool to improve Nova SBE's selection of candidates with a focus on academic success. The recognition of endogeneity and sample-selection bias prevented conclusions about causality and suggested a predictive approach to the problem.

The main finding of this research is that the internal high school grade is a much better predictor of students' achievement than the score of the maths national exam, showing larger coefficients and much higher predictive power. Regarding ECTS in the first year or the dropout probability, the exam score does not even have statistical significance.

Does it mean that Nova SBE should disregard the maths national exam as an entry requirement[29]? Probably not. Firstly, this study is focused only on academic measures of success; it is possible that if other measures were considered, like for instance first wage, it could present different results. Secondly, since there is no information about those students who have not been accepted at Nova SBE, it is difficult to evaluate the impact of a change on the "recruitment" policy of Nova SBE. Third, the dataset only considers students who enrolled in 2009/10 and, since then, the structure of national exams has changed, which may undermine the present conclusions in more recent years. Finally, changing the

[29] The relevance of high school exams is not in question: As national standardized tests, they play a very important role in terms of accountability of the whole schooling system, allowing the comparison of students across different school, committing schools and teachers to their work and assuring all contents are covered at high school.

application rules would also change students' prior decisions about how to prepare for applying to Nova SBE. This specific problem of endogeneity is very difficult to overcome, given that a new structural model would change students' choices about type of secondary school, effort or time of study for each exam.

Apart from these conclusions, other relevant findings should be mentioned. Results suggest that following the economics track at high school and attending a private school seems to be irrelevant in predicting success at Nova SBE, after controlling for the internal high school grade and the mathematics national exam score . On the other hand, some controls, such as gender, age, and family background were revealed to have some influence on undergraduates' achievement. In line with the literature, male students perform better, as well as those who are not displaced and who enrolled at Nova at 18 years of age or less. Also of interest is the positive partial correlation between first-year success and being accepted in the first application round at Nova or receiving social support, as well as the positive influence of fathers' educational background in reducing the dropout probability.

In the future, it would be useful to use data for different school years to check the robustness of these results and also to include other controls such as family income and class-level information. To overcome the sample selection problem it would be pertinent to get information about those students who scored just below the threshold. Additionally, it would be interesting to predict first wage as an outcome; this data is not available yet, but it could become an important instrument at Nova SBE.

Considering the specific characteristics of Nova SBE and its students, it would be dangerous to generalize these conclusions to the national education system as a whole. However, such an evaluation should be implemented by other courses and institutions in order to improve the efficiency of the Portuguese Higher Education System.

References

A3ES, Agência de Acreditação e Avaliação do Ensino Superior. 2012. *O Sistema de Ensino Superior em Portugal*. Lisbon.

Alves, Dino. 2014. "Undergraduate Students' Success at Nova SBE: A censored duration analysis in discrete time." (Microeconomics Research Note, Monograph)

Anderson, Gordon, Dwayne Benjamin, and Melvyn Fuss. 1994. "The Determinants of Success in University Introductory Economics Courses." *Journal of Economic Education*, 25(2): 99-119.

Ballard, Charles, and Marianne Johnson. 2004. "Basic Math Skills and Performance in an Introductory Economics Class." *Journal of Economic Education*, 35(1): 3-23.

Becker, William, and Michael Watts. 1996. "Chalk and talk: A national survey of teaching undergraduate economics". *American Economic Review*, 86 (May): 448–54.

Betts, Julian and Darlene Morell. 1999. "The determinants of undergraduate grade point average: the relative importance of family background, high school resources, and peer group effects." *Journal of Human Resources*, 34(2): 268-293.

Borg, Mary, and Harriet Stranahan. 2002. "The effect of gender and race on students performance in principles of economics: the importance of personality type." *Applied Economics*, 34: 589–598.

Cameron, Colin, and Pravin Trivedi. 2005. *Microeconometrics: Methods and Applications*. New York: Cambridge University Press

Cohn, Elchanan, Sharon Cohn, Donald Balch and James Bradley Jr. 1998. "The Effects of Mathematics Background on Student Learning in Principles of Economics." *Journal of Education for Business*, 70(1): 18-22.

DGES. 2010. *Acesso ao Ensino Superior: dez anos de concurso nacional (2000-2009)*. Lisbon.

Dolado, Juan, and Eduardo Morales. 2009. "Which factors determine academic performance of economics freshers? Some Spanish evidence." *Investigaciones Económicas*, 33(2):179-210

Evans, Williams, and Robert Schwab. 1995. "Finishing High School and Starting College: Do Catholic Schools Make a Difference?." *Quarterly Journal of Economics*, 110: 74 - 94.

Fernandes, Graça, and Margarida Lopes. 2008. "ISEG undergraduate students: determinants of academic performance." *ECER Conference "From teaching to learning"*.

Hanushek, Eric. 1979. "Conceptual and empirical issues in the estimation of educational production functions." *Journal of Human Resources*, 14(3):351-388.

Hanushek, Eric. 1986. "The economics of schooling: production and efficiency in public schools." *Journal of Economic Literature*, 24: 1141–1177.

Hanushek, Eric. 1996. "School resources and students performance." In: *Does Money Matter? The Effect of School Resources on Student Achievement and Adult Succes*, ed. Burtless, G., 43–73. Washington, D.C.: Brookings Institution.

Hassink, Wolter, and Hanna Kiiver. 2007. "Age-dependent Effects of Socio-economic Background on Educational Attainment – Evidence from Germany." *Discussion Papers*, 26.

Heckman, James. 1979. "Sample Selection Bias as a Specification Error." *Econometrica*, 47: 153–161.

Horowitz, John, and Lee Spector. 2005. "Is There a Difference Between Private and Public Education on College Performance?." *Economics of Education Review*, 24: 189-195.

Koenker, Roger, and Kevin Hallock. 2001. "Quantile Regression." *Journal of Economic Perspectives*, 15(4): 143-156.

Krueger, Alan. 1999. "Experimental estimates of education production functions." *Quarterly Journal of Economics*, 114 (2):497-532.

Lagerlöf, Johan, and Andrew Seltzer. 2009. "The effects of remedial mathematics on the learning of economics: Evidence from a natural experiment." *Journal of Economic Education*, 40 (2): 115-137 .

Lopus, Jane. 1997. "Effects of the high school economics curriculum on the learning in the college principles class". *Journal of Economic Education*, 28 (2): 143–153.

Marcenaro, Oscar, and Lucía Navarro. 2007. "El éxito en la universidad: una aproximación cuantílica," *Revista de Economía Aplicada*, 15:5-40.

Moyer, Eugene, and Donald Paden. 1970. "Economics Achievement and Mathematics. Training." *Journal of Economic Education*, 2: 104-106.

Myatt, Anthony, and Charles Waddell. 1990. "An Approach to Testing the Effectiveness of the Teaching and Learning of Economics in High School." *Journal of Economic Education*, 21(3): 355-363.

Naylor, Robin and Jeremy Smith. 2004. "Determinants of educational success in higher education". In *International Handbook of the Economics of Education*, 415-461. ed. Geraint Johnes and Jill Johnes. Cheltenham: Edward Elgar Publishing.

OECD .2013. *Education at a Glance 2013: OECD Indicators*. Paris: OECD Publishing.

Palmer, John, Geoffrey Carliner, and Thomas Romer. 1979. "Does high school economics help?." *Journal of Economic Education*, 10 (Spring): 58-61.

Parri, Janne, and Kadric Aas. 2006. "National examination scores as predictors of university students' performance in Estonia." *Trames,* 10 (3), 255-267.

Reid, Roger. 1983. "A note on the environment as a factor affecting student performance in principles of economics". *Journal of Economic Education*, 14 (4): 18–22.

Sampaio, Gustavo. 2013. "College Dropouts and Entrance Test Scores: A Censored Quantile Regression Approach." *2013 Annual Meeting, Agricultural and Applied Economics Associatio*n, 150324.

Smith, Jeremy and Robin Naylor. 2001. "Determinants of degree performance in UK universities: A statistical analysis of the 1993 student cohort." *Oxford Bulletin of Economics & Statistic*s, 63: 29-60.

Tobin, James. 1958. "Estimation of relationships for limited dependent variables." *Econometrica*, 26 (1): 24–36.

Todd, Petra, and Kenneth Wolpin. 2003. "On the specification and estimation of the production function for cognitive achievement." *The Economic Journal*, 113: 3-33.

Universidade do Porto, Reitoria. 2012. *Sumário analítico do Percurso dos estudantes admitidos na Universidade do Porto em 2008/2009, pelo regime geral*. Porto.

Appendix

Table A. 1 Additional descriptive statistics on educational outcomes

	All sample		Economics		Management	
	N	%	N	%	N	%
Did not Enrol	17	4,7%	5	3,0%	12	6,1%
Dropouts	37	10,2%	11	6,7%	26	13,1%
Dropouts (1st year)	22	6,1%	6	3,6%	16	8,1%
Dropouts (2nd year)	12	3,3%	5	3,1%	7	3,5%
6 Semesters	158	43,5%	78	47,3%	80	40,4%
7 Semesters	73	20,1%	37	22,4%	36	18,2%
8 Semesters	39	10,7%	23	13,9%	16	8,1%
9 Semesters or more	39	10,7%	11	6,7%	28	14,1%

Figure A. 1 Final GPA distributions according to 9 different dimensions

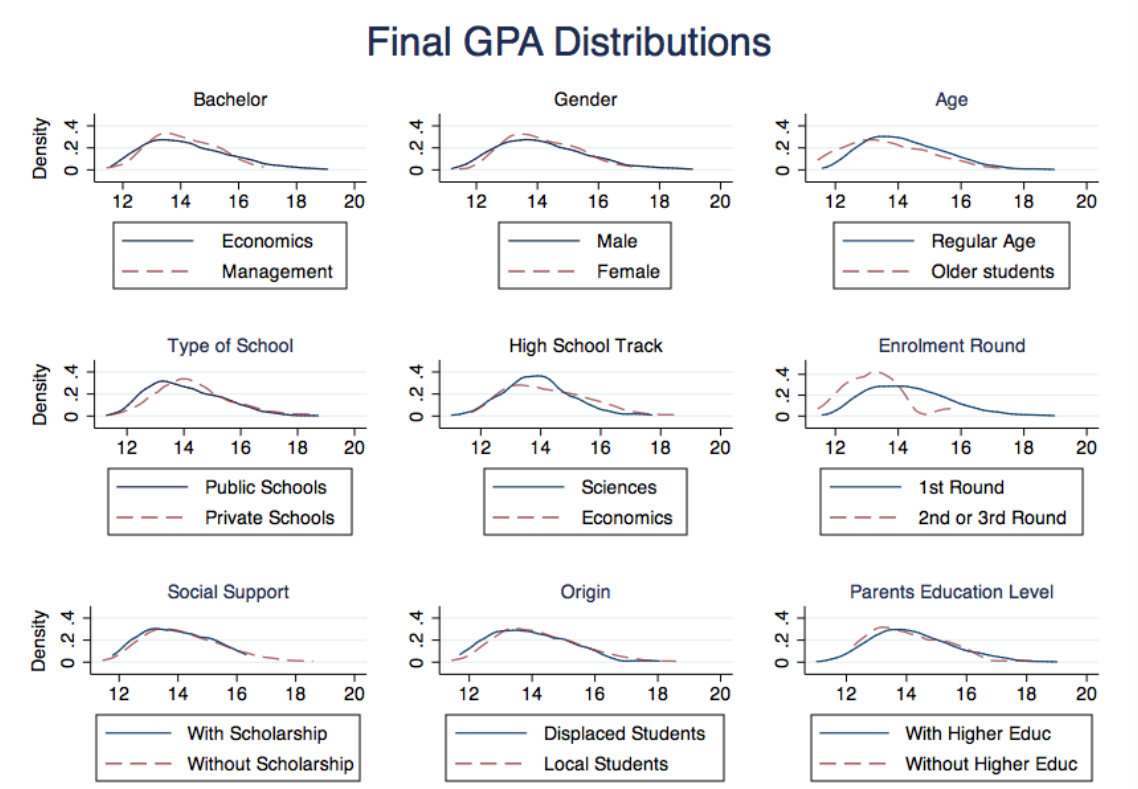


Figure A. 2 Comparison on OLS and QR coefficients for each regressor