Which Factors Determine Academic Performance of Economics Freshers ?. Some Spanish Evidence*

Juan J. Dolado* & Eduardo Morales**

(*) Universidad Carlos III & CEPR & IZA
(**) Harvard University

This draft: October 20, 2008

ABSTRACT

This paper analyses the impact of several factors potentially affecting academic performance of Economics freshers (first-course undergraduates) at Universidad Carlos III de Madrid over the period 2002-2005. Outcomes are the grades obtained by these students in three core subjects which differ in their requirements of math skills. Controls include specialization track at high school, type of school, parental education background college attainment, grades obtained at the university entry-exam, gender and nationality. Our main finding is that those students who completed a technical track at high school tend to perform much better in subjects involving math skills than those who followed a social sciences track (supposedly tailor-made for future economics students) and that the latter do not perform significantly better in subjects where prior training in economics helps. Moreover, students from public schools prevail in the lower and upper parts of the grades distributions while females tend to perform on average better than males.

JEL Classification: I21 and I29

Keywords: academic performance, pre-university determinants, selection biases, quantile regressions.

* Corresponding author: Juan J. Dolado (dolado@eco.uc3m.es). We are indebted to a Co-Editor and two anonymous referees for many helpful suggestions which greatly improved the paper, and to Marisol Somolinos for providing us with the students´ administrative records at UC3M. We also thank Antonio Cabrales, Libertad González, Carmelo Nuñez, Tommaso Nannicini and participants at the 2007 PEW conference (Madrid), EEA-ESEM (Budapest), SAE (Granada) and EEEPE workshop (Amsterdam) for insightful comments on preliminary drafts of this paper. The second author is currently a graduate student in the Economics Ph. D. program at Harvard. Financial support from the European Commission under the project *The Economics of Education and Education Policy in Europe* (MRTN-CT-2003-50496), Consolider-Ingenio 2010 (MEC), Excelecon (CM) (Dolado), and Consejería de Educación de la Comunidad de Madrid (Morales) is gratefully acknowledged.

"G.M., an excellent student with straight A´s in the high-school track of social sciences, including Maths, wished to become an economist, like his father was. When he enrolled in an Economics degree at the university, he failed miserably in the first course of Maths… Which is the reason for this failure?. One likely explanation is the inadequacy of the Maths taught in high school and those required for such degree…this failure is not due to a poor performance of high-school teachers, but rather to the lack of information about the level of Maths required in Economics." [Pedro Álvarez Martínez, "La influencia del bachillerato en el fracaso en Economía" EL PAÍS, 09/06/2003]

1. Introduction

It is often claimed that Economics is the discipline with the highest need for formalism in theory-building among social sciences. Thus, undergraduate courses in Economics require a good background in basic mathematics besides prior training in introductory economics and economic history. In this paper we present some empirical evidence about what sort of prior (pre-university) qualifications are related to good academic performance by first-year undergraduates (hereafter labeled as *freshers*) enroled in an Economics degree at one of the Spanish public universities.

In particular, we are interested in examining whether factors correlated with success in math-intensive subjects differ from those which improve performance in other economics subjects with less mathematical content and where prior training is useful. To do so, we focus on the students´ performance in three core subjects ordered in decreasing level of math requirements: Mathematics, Introductory Economics and Economic History (henceforth, *Maths*, *Introecon* and *Econhist*, respectively). [1] More precisely, one of our main goals here is to analyze whether the social sciences track taken at high school - supposedly the one tailor-made for future students of university-level economics- helps to improve performance in these subjects. As highlighted in the newspaper quotation at the epigraph of the paper, the increasing number of freshers who struggle at the early stages of this degree, due to their weak math training at high school, is raising a growing concern among higher-education pundits.

Our evidence relies upon individual-level data collected by us using a sample of almost 400 freshers at Universidad Carlos III de Madrid (UC3M) who took exams in the above-mentioned subjects between the 2002/03 and 2005/06 courses. [2] The students in our sample were enroled at UC3M during the first semester of their four-year BA degree (*Licenciatura*) in Economics (LE) or

---

[1] The syllabus in *Maths* consists of limits, differentiation and integrals. *Introecon* is a basic Microeconomics course (consumer theory and theory of the firm) and *Econhist* deals with the long-run development process in Western Europe.

[2] The choice of freshers is dictated by data availability. However, the fact that the withdrawal rate is rather high (about 30%) during the completion of a degree in the Spanish university system implies that the estimates obtained for this restricted group of students are likely to be less affected by selection biases than the estimates for students in more advanced courses (due to the attrition of those who drop out after the first course).

Business Administration (LADE).[3] "Repeaters", either directly from these degrees or coming from alternative ones (typically from engineering or medicine) are excluded from our sample in order to isolate the treatment effects of variables like the high school track or the type of high school .

Information is available on the following variables : type of high school (public, private and charter [4]) during their upper-secondary education (two years of *Bachillerato* at 16 and 17 years of age); specific training received during this period (i.e., the *Bachillerato* specialization track); and the grade obtained granting access to a university-level degree, which is used as a proxy for unobserved skills. This grade is a weighted average of: (i) the grade obtained in a (nationwide) centralised entry exam taking place at the universities (*Selectividad*) just after students complete high-school at 18 years of age (with a weight of 40%), and (ii) an average of the high school grades obtained during the two courses of *Bachillerato* (with a weight of 60%). These variables, in addition to gender, nationality and some family- background characteristics, are the basic inputs we use to explain outcomes (grades awarded in the final examination of the three university-level subjects) using an achievement production-function approach.

At this stage, it is worth stressing that an important shortcoming of our dataset is that we lack detailed information related to parental background (indeed, the only available *proxy* is an indicator of whether any of the parents has a college degree). However, drawing upon previous evidence on the existence of a strong correlation between type of school and (missing) family´s socio-economic status (see Calero, 2006), we claim that this variable may be a good proxy for family characteristics. The reason is that quality of education (student/teacher ratios, computer facilities, foreign languages, etc.) in non-public schools is generally considered to be higher than in most public schools, in exchange for annual tuition fees of about € 6 to 8 k. in private schools and around € 2 to 3 k. in charter schools. To the extent that families where (at least one of the) parents have a college degree are wealthier, they can afford to send their children to private and charter schools. Indeed, the sample correlation between our parental-background dummy variable and a dummy for non-public schools turns out to be very high (0.89). Therefore, given the difficulties in disentangling separate effects for each of these two indicators in our empirical approach, we will interpret their impacts as indistinctly stemming from *family background.*

Grades in the Spanish education system are numerical, ranging from 0 to 10, leading to five categorical grades. Grades below 5 imply a *Suspenso* (Fail in the anglosaxon system), between 5 and 7 is an *Aprobado* (Third), between 7 and 9 is

---

[3] During the first year , freshers in LADE have the same subjects as those in LE. Thus, for brevity, we will refer in the sequel to all of them as Economics freshers.

[4] A charter school (*concertado*) is a school subsidized by the public sector, typically run by religious orders.

a *Notable* (Lower Second), between 9 and 10 is a *Sobresaliente* (Upper second) and 10 (or very close to that grade) is a *Matrícula de Honor* (First or Distinction). The categorical grades will be used to describe the data in Section 2 where, to simplify notation, we will denote them with the labels SUS, AP, NOT, SOB and MH, respectively. By contrast, the numerical grades (available from the archives of UC3M) are the ones which will be employed in implementing the regressions discussed in Section 3.

Our empirical strategy relies upon two different econometric approaches. First, we run least-squares (OLS and IV) regressions to explain the outcomes (grades achieved in the three subjects at hand), analysing potential biases in each instance. Secondly, we measure the impact of the determinants on the dependent variable at different points in its conditional distribution, by means of quantile regressions (QR). In this fashion, we will be able to provide a sense of how the impact of the explanatory variables may differ throughout the grades distribution. For example, one may find that a particular covariate, while seemingly important at the mean as a determinant of the outcome, may in fact have different impacts across students with high or low grades.

Our paper falls into a large literature that examines the determinants of university students´ academic performance (see, e.g., Dearden et al., 1998, Smith and Naylor, 2001, for the UK; Eide and Showalter, 1998, for the US; Marcenaro and Navarro, 2007, for Spain; and Hanushek, 1986, for a good overview of the literature). However, very few of these studies have focused on factors affecting performance in specific subjects of a college degree, as we do here. To our knowledge, the only exception is a recent literature evaluating the effect of remedial math instruction on the perfomance in *Principles of Economics* in US and UK universities, respectively (see, e.g., Pozo and Stull, 2006 and Lagerlöf and Seltzer, 2008). In this sense, we expect that our findings here can shed some light on the above-mentioned debate on the adequacy of the currently available system of high-school tracks in Spain.

The rest of the paper is structured as follows. Section 2 describes the dataset. In Section 3, we present the alternative econometric approaches and discuss potential biases. Section 4 offers the main results. Section 5 contains a discussion about how representative is the sample and checks for some potential selection biases. Finally, Section 6 concludes.

2. Data and Descriptive Statistics

The data is made up of a questionnaire distributed among four cohorts of freshers enrolled in the bilingual group of the BA degree (*Licenciatura*) in Economics during the academic years 2002/03 to 2005/06. [5] All these freshers

[5] Being a student in the "bilingual" group means that, except for a few subjects (e.g. those related to Law), all teaching takes place in English. Admission to this group is conditional on passing an English exam organized by UC3M. Courses are organized on a semester basis and

were taught *Maths* by the same instructor (one of us) in classrooms with a maximum capacity of about 100 students per group. We solicited information from them about the type of school (public, charter and private) they attended during high school (two years of *Bachillerato* at ages 16 and 17), the kind of training they received during this period (there are four types of *Bachillerato* tracks: technical, natural sciences & health, social sciences and humanities which are chosen by students at age 15 just before starting their first high-school year, i.e., *Primero de Bachillerato*) and the grades they obtained both in the *Selectividad* exam when they finished high school at age 18 and in the two *Bachillerato* courses at ages 16 and 17. To avoid measurement errors, these last two pieces of information were cross-checked with the UC3M administrative records. The response rate to the questionnaire was 96.5%, yielding a sample of 386 individuals (leaving repeaters aside, who represented 7.3% of the population).

A brief description of the relevant variables is provided in Table 1 where we present the conditional distribution of (categorical) grades given students´ characteristics plus the unconditional frequencies of the latter in the last column. For expository purposes, we have grouped the five categorical grades into three broader categories: S (SUS), AN (AP+NOT) and SM (SOB+MH). Overall, 49.2% of the students are male whilst 89.3% are Spanish. By family background, 27.3% had (at least) one of their parents with a college degree. [6] By type of school, 42% come from public schools, 21% from charter schools and 37% from private schools. By type of *Bachillerato* track, 67% have done social sciences, 26% the technical one and the remaining 7% did natural sciences & health (3%) and humanities (4%). It should be emphasized that the high school training in mathematics is more intense in the technical and natural sciences & health *Bachilleratos* than in social sciences. These are the three tracks where students take two compulsory annual math courses, whereas it is only an optional subject for those enroled in humanities. [7] Table 2 summarizes the contents of the different tracks in terms of compulsory and optional subjects.

[Tables 1 and 2 about here]

Table 1 shows that male students are less successful than female students in passing these subjects (except in *Introecon*). Likewise, those coming from public schools with social sciences or humanities *Bachilleratos* tend to do worse. Interestingly, however, students from public schools (mostly with a technical track) do rather well in achieving the highest category (SM=2) in all the three subjects. Thus, students from these high schools seem to have a U-shaped

---

there are ten subjects in the first year (five in each semester), with exams taking place in February and June.

[6] The indicator variables Parent (0) and (1) denote no parent with a university degree and at least one parent with such a degree, respectively.

[7] We checked that all students in our sample coming from the humanities track had taken a math course as one of the optional subjects in this high-school track.

distribution across grades. The lower tail contains those who chose social sciences or humanities tracks whilst in the upper tail there are those who did more scientific-oriented tracks. Given that 70% of high-school students in Spain are enroled into the public education system, the latter fact could be explained by the existence of higher competition among the best students in these high schools, particularly among those completing a technical track. Hence, comparing the best students in this track (equivalent in all observable characteristics but with different school backgrounds), it seems that the ones with a public school background are likely to be drawn from a higher point in the underlying ability distribution. As mentioned before, the conditional distributions of type of school and parental college background are remarkably similar, which will render difficult to identify their separate effects on outcomes. It is also worth noticing that foreign students exhibit a slightly higher variability in grades than natives. Finally, the last row in Table 1 presents the correlations between the numerical marks in each of the subjects and the marks in the *Selectividad* exam. These correlations range between 0.50 and 0.67, being largest in the case of *Maths*.

Figure 1 depicts the (kernel) densities of the (numerical) grades in the three subjects. The distributions in *Econhist* and *Introecon* are unimodal with the former being the one more shifted to the right (i.e., higher probability of a pass grade). Conversely, the density of *Maths* is bimodal and it is the one more shifted to the left (i.e,. lower probability of a pass grade). [8] To achieve comparability across subjects in the estimation of the impacts of the different pre-university determinants, we will use the standarized grades in the empirical analysis. Hence, the estimated effects are measured in terms of the corresponding standard deviations (s.d.´s). To convert these grades into numerical ones, one should multiply the former by the s.d´s of the grades.

[Figure 1 about here]

Figure 2, in turn, displays the densities of the *Selectividad* and *Bachillerato* grades which are similar to a conventional skills distribution. As expected, the latter tend to be uniformly larger than the former illustrating the fact that students tend to do worse in centralized exams than in those taking place at their own schools (the grade gap between *Bachillerato* and *Selectividad* is 0.51 with a s.d. of 0.29). Since it is sometimes argued that non-public schools tend to inflate the *Bachillerato* grades of their students more than public schools, Figure 3 depicts the respective gaps for these two broader school types. We find some supporting evidence for this claim in our sample: the grade gap in non-public schools (0.66, s.d= 0.26) is larger than in public schools (0.31, s.d=0.18). Thus,

[8] The moments (mean and s.d.) of the three distributions are as follows: *Maths* (5.17, 2.50), *Introecon* (5.70, 1.57) and *Econhist* (6.48, 1.44)

being less distorted, the *Selectividad* grades are the ones chosen as a more appropriate proxy for unobserved ability in the empirical section. [9]

[Figures 2 and 3 about here]

3. Econometric approaches

3.1 A brief overview of the production function approach

We rely upon an extensive literature analyzing school outcomes in developing and developed countries using a production function approach; cf. Hanushek (1986, 1995), Case and Deaton (1999), Bjorklund et al. (2003), Todd and Wolpin (2003) and Glewwe and Kremer (2005). Accordingly, outcomes are explained as a function of several inputs in the following manner:

$$y_{i0} = x_{i0}{'}\delta_1 + \delta_2 a_{i0} + u_{i0} \qquad (1)$$
$$y_{i1} = x_{i1}{'}\beta_1 + \beta_2 a_{i1} + u_{i1} \qquad (2)$$

where y_{i0} and y_{i1} $(i = 1,2,...,n)$ represent some metric of academic performance (grades) in two different points in time $(t = 0,\ 1)$: before entering university (e.g., grades in the *Selectividad* exam, denoted by y_{i0}^S, or grades in *Bachillerato,* denoted by y_{i0}^B) and at the university (i.e., grades in each of the three different subjects , y_{i1}), respectively; a_{it} is unobserved ability in each period; x_{it} is a vector of inputs containing the individual and family background characteristics discussed above plus the high-school track; and u_{it} are zero-mean i.i.d. disturbances.

Assuming that the regressors in (1) and (2) are uncorrelated with the disturbances, and that $a_{i1} = a_{i0} + v_i$, where v_i is an error term, a standard solution (see Hanushek, 1986) to control for unobserved ability is to solve for for a_{i0} in (1), so that y_{i0} becomes a proxy for a_{i1}. This implies that (2) can be rewritten as:

$$y_{i1} = x_{i0}{'}\varsigma_1 + x_{i1}{'}\varsigma_2 + \varsigma_3 y_{i0} + w_i\,. \qquad (3)$$

However, to the extent that $w_i = u_{i1} - (\beta_2 / \delta_2)u_{i0} + \beta_2 v_i$, "errors-in-variables" (u_{i0} is correlated with y_{i0}) and other endogeneity problems (potential correlation of v_i with some of the regressors) are bound to invalidate the use of OLS as a consistent estimation approach for the parameters in (3). For example, as will become clear below, one of our inputs of interest, i.e., the high-school

[9] Notice that some of the students have a *Selectividad* grade below 5 (the pass grade) because, as mentioned earlier, the centralized exam grade only accounts for 40% of the overall mark.

track, is present in both (1) and (2) since it is chosen before taking the *Bachillerato/Selectividad* exams and it may not be random. For example, it is often argued that the best students are the ones who choose the technical track in high school. Thus, if we find that this track improves academic performance in an Economics degree, it does not follow necessarily that a randomly allocated high-school student to the technical track would later on exhibit a better performance in such a degree. This would lead to potential biases. [10]

In what follows, we briefly review some of the procedures proposed in the literature to circumvent these shortcomings. The first one is based on the following assumptions: (i) treating ability as a fixed effect, i.e., $a_{i0} = a_{i1} = a_i$, (ii) assuming that $x_{i0} = x_{i1} = x_i$, insofar as the dates of the *Selectividad* exam and the university exams are rather close in time, and (iii) $Cov(x_{it}, u_{i0}) = Cov(x_{it}, u_{i1}) = 0$, *(t=0,1)*. Further, under the additional assumption of $\beta_2 = \delta_2$, subtracting (1) from (2) yields:

$$(y_{i1} - y_{i0}) = x_i{'}(\beta_1 - \delta_1) + (u_{i1} - u_{i0}). \tag{4}$$

Note that (4) mimics the well-known first- differencing or "within" estimation approach to control for fixed effects in panel data. Thus, under the previous assumptions, one can obtain consistent estimates of the relative gains in academic achievement before and after entering university. Nonetheless, there are good reasons to suspect that $\beta_2 \neq \delta_2$ since some of the innate abilities to succeed in high school may be different from those required at university-level studies. In a such a case, estimation of (4) will not be not be a valid procedure.

3.2 Biases in estimating average treatment effects by OLS

There are, however, other ways of achieving consistency that do not require the assumptions leading to (4). They are often based on the use of information on students´ academic achievements before being subject to a particular treatment (e.g. the choice of a particular track or of a type of high-school). Bonhomme and Sauder (2008) provide a nice illustration of this approach by using pre-treatment outcome variables as instrumental variables (IVs) for post-treatment ones in order to obtain consistent estimates of the *Average Treatment Effect* (ATE) of the choice of type of school on later outcomes in the UK. Unfortunately, we lack this type of information because, besides the *Selectividad* exam, Spanish high-school students do not have any other centralized examination, since the abolition of the *Reválida de Cuarto de Bachillerato.* The latter was a centralized examination taken when students were 14 years old, which was abolished in

[10] We could, however, argue that, since there are no repeaters in our sample, enroling in an Economics degree can be interepreted to a large some extent as a random choice. The fact that there is an important fraction (33%) of freshers in our sample who did not follow a social sciences *Bachillerato* (and that are not repeaters) probably reflects a high degree of uncertainty about what college degree they would subsequently enrol in.

1975. Thus, lacking any genuine pre-treatment variables, our estimated ATE is likely to be biased.

Nothwithstading, under some (admittedly restrictive) alternative assumptions to be laid out below, it can be shown that a slightly different IV approach to estimate (3) can yield estimates of the ATEs which are unambiguously downward biased. Note that this is a rather helpful result since getting a lower bound means that, if our estimated ATE is quantatively and statistically relevant, then the true ATE will be even more important.

To illustrate the intuition behind this approach, let us assume, for concreteness, that we are solely interested in estimating the ATE of choosing a given high-school track at 16 years of age (treatment) whereas the reference category is choice of any other track (control). Let D be a dummy variable for those individuals who are treated, so that the coefficient of D in (2) becomes the true ATE. [11] Then, choice of y_{i0}^S as a proxy for unobserved ability (for the reasons discussed in Section 1), and denoting by c_i the remaining set of regressors in the vector x_i , will lead to the following equivalent expressions to (1) and (2): [12]

$$y_{i0}^S = \delta_0' c_i + \delta_1 D_i + \delta_2 a_i + u_{i0}^S, \qquad (5)$$

$$y_{i1} = \beta_0' c_i + \beta_1 D_i + \beta_2 a_i + u_{i1}, \qquad (6)$$

where β_1 is the true ATE we are looking for.

Then, solving for a_i in (5) and replacing it into (6), yields the following equation:

$$y_{i1} = (\beta_0 - \frac{\beta_2}{\delta_2}\delta_0)' c_i + (\beta_1 - \frac{\beta_2}{\delta_2}\delta_1) D_i + \frac{\beta_2}{\delta_2} y_{i0}^S + (u_{i1} - \frac{\beta_2}{\delta_2} u_{i0}^S), \qquad (7)$$

From (7), it becomes clear that the best we can achieve with this approach is a consistent estimate of the slope on D_i , i.e., $\beta_1 - (\beta_2 / \delta_2)\delta_1$, which differs from the true ATE, β_1, unless $\delta_1 = 0$, a case that we discarded before since the choice of high-school track is previous to the *Selectividad* exam. However, this biased ATE can still be useful to make statements about the true ATE if we impose the following reasonable assumption.

A.1: *(i) sign (*β_1*)=sign*(δ_1)*, and (ii) sign (*β_2*)=sign*(δ_2) *>0.*

---

[11] Notice that we are also assuming that the potential outcomes are a linear function of the different covariates and the treatment effect is common across all individuals.

[12] It is also assumed that $c_{i1} = c_{i0} = c_i$, and that $a_{i1} = a_{i0} = a_i$.

The idea behind A.1.(i) is that if a particular high-school track prepares students better/worse for future university-level exams, it also prepares them better/worse for the *Selectividad* exam. Assumption A.2.(ii), which just ensures that $(\beta_2/\delta_2)>0$, is trivial since ability is always thought to improve performance in both exams. This means that, if we are able to find a consistent estimation procedure for the parameters in (7), the estimated ATE of D_i under A.1 would yield a downward biased estimate (in absolute terms) of the true ATE. In other words, if the the estimated ATE is found to be positive, the true ATE would be even more positive whereas, if it is negative, the true ATE will be even more negative. Such a lower bound will be denoted in the sequel as *LB* $(=\beta_1-(\beta_2/\delta_2)\delta)$.

Notice that, in principle, OLS will not yield a consistent estimate of LB in (7). First, due to the presence of u_{i0}^s in the composed error term in (7) and the fact that, from (5), u_{i0}^s is correlated with y_{i0}^S, we will get an inconsistent (downward biased) estimate of the coefficient on y_{i0}^S due to "errors- in- variables". Second, the fact that, from (5), y_{i0}^S and D_i are also correlated implies that the estimated coefficient on D will also be inconsistent.

To be more precise about the sign of the bias on LB, let us first assume that $E(D_iu_{i0})=E(D_iu_{i1})=E(c_iu_{i0})=E(c_iu_{i1})=E(y_0^Su_{i1})=0$. Next, after applying by the projection matrix $M_c=I-c(cc')^{-1}c'$ to both sides of (7), so that *c* is removed from its RHS, let us denote the projected regressand in the (second- step) regression as $\tilde{y}_1$ ($\tilde{y}_1=M_c\, y_1$, etc.), the matrix of regressors $(\tilde{D},\tilde{y}_0^S)$ as $\tilde{X}$, the composed error term as $\tilde{w}$, and the vectors of parameters and OLS estimators as *b* and $\hat{b}_{ols}$, respectively. Using the vector notation, we can easily compute the sign of the bias on *LB* by selecting the first element (i.e., the one pertaining to the coefficient on D_i) in the (2x1) vector of probability-limits, given by $p\lim(\hat{b}_{ols}-b)=E(\tilde{X}'\tilde{X})^{-1}E(\tilde{X}'\tilde{w})$. This yields:

$$p\lim(LB_{ols}-LB)=(\beta_2/\delta_2)\frac{Cov(\tilde{D},\tilde{y}_0^s)Var(\tilde{u}_0^S)}{\det(E(\tilde{X}'\tilde{X}))}\equiv\xi, \tag{8}$$

whose sign depends on the sign of $Cov(\tilde{D},\tilde{y}_0^s)$ since (β_2/δ_2)>0 from A.1.(ii) and $\det E(\tilde{X}'\tilde{X})>0$. Assuming as before that the regressors in (5) and (6) are uncorrelated with their respective error terms, equation (5) implies that $Cov(\tilde{D},\tilde{y}_0^S)=\delta_1Var(\tilde{D})+\delta_2Cov(\tilde{D},\tilde{a})$. Since the first term is always positive and $\delta_2>0$, it becomes clear that the sign of ξ will depend on the sign and size of $Cov(\tilde{D},\tilde{a})$. If it is negative and sufficiently large (i.e., the least able students choose track *D*), ξ could be zero or negative, in which case LB_{ols} (= $\beta_1-(\beta_2/\delta_2)\delta_1+\xi$) remains a genuine lower bound of the true ATE. By contrast, if it turns out to zero or positive (i.e., if there is no relation or if the most able students choose track *D)*, the bias will be unambiguously positive, making us no longer sure that LB_{ols} is smaller than β_1. However, even in this problematic case, a lower bound could be achieved if the sizes of remaing terms involved in the RHS of (8) lead to a sufficiently small positive bias. The only way to check whether ξ is sufficiently small would be to compare the OLS results with those obtained from an alternative IV estimation procedure that yields consistent estimates of the parameters in (7). This issue is discussed next.

3.3 Instrumental variables

To obtain consistent estimates of the parameters in (7), we adopt an IV approach based on using students´ *Bachillerato* grades, denoted by y_{i0}^{B}, as an instrument for y_{i0}^{S}.[13] Specifically, as in (5) and (6), we assume that the equation determining y_{i0}^{B} is :

$$y_{i0}^{B} = \gamma_0 ´c_i + \gamma_1 D_i + \gamma_2 a_i + u_{i0}^{B}. \tag{9}$$

Let us further make the additional assumption.

A2: *(i) The variables c_i, a_i and D_i are orthogonal to the error terms in (5), (6) and (9), and (ii) the error terms in these equations are uncorrelated among them.*

Then, projecting again both sides of (7) and (9) on *c*, and denoting the matrix of projected IVs as $\tilde{Z} = (\tilde{D}, \tilde{y}_0^B)$ and the vector of IV estimates as $\hat{b}_{IV}$, imply that $p\lim(\hat{b}_{IV} - b) = E(\tilde{Z}´\tilde{X})^{-1} E(\tilde{Z}´\tilde{w})$. Thus, selecting the second element in such a (2x1) vector yields:

$$p\lim(LB_{IV} - LB) = \frac{Var(\tilde{D})Cov(\tilde{y}_0^B, \tilde{w}) - Cov(\tilde{D}, \tilde{y}_0^B)Cov(\tilde{D}, \tilde{w})}{\det(E(\tilde{Z}´\tilde{X}))} = 0, \tag{10}$$

since $Cov(\tilde{y}_0^B, \tilde{w}) = Cov(\tilde{D}, \tilde{w}) = 0$ (from A.2), whereas $\det E(\tilde{Z}´\tilde{X}) \neq 0$. In other words, if we accept A.2, then y_{i0}^{B} will be uncorrelated with u_0^S, u_0^B and u_1, so that this instrument works in eliminating the asymptotic bias in (8). Hence, as mentioned earlier, a comparison of the OLS and IV estimates of (7) will provide us with some indication of how serious the bias term (ξ) is.

How credible is A.2 ?. In principle, two criticisms can be made against this assumption. First, it can be argued that, given our limited information on students´ family background (i.e., parents´ college attainment and school type), omitted variables could lead to non-zero correlation between regressors and disturbances. Second, since y_{i0}^{B} is clearly not a pre-treatment covariate, there

[13] We are grateful to the Co-editor and one of the referees for pointing out this solution to us.

can be a non-zero correlation between the instrument and u_1 (and hence with w), therefore invalidating the IV approach.

Regarding the first criticism, admittedly not much can be done about it with our available dataset except to accept the strong assumption that our two proxies capture "almost perfectly" parental background. As discussed in Section 1, we claim that this is acceptable insofar as parents´ education and choice of school are bound to be good proxies for parental inputs (social status) in children´s education. Moreover, teaching facilities are bound to be strongly related to type of school. Thus, our claim seems plausible. As for the second criticism, one could argue that the *Bachillerato* grades are also "almost perfectly" correlated with students´ academic performance *before* they join *Bachillerato* at age 16. In other words, our implicit assumption is that someone performing well (badly) in earlier stages of primary and secondary education is bound to have a similar performance in *Bachillerato*. By contrast, the novelty of taking the *Selectividad* exam at a centralized level (i.e., in a different environment from the exams taken at high school) may yield a somewhat different performance, as illustrated in Figure 3 by the worse grades obtained by students in this exam.

In case the previous arguments do not convince the reader, we will also report, for robustness, OLS regressions (to be interpreted as partial correlations) of university-subject grades on high- school track dummies and other exogenous regressors, therefore excluding the potentially endogenous post-treatment covariates- i.e., y_{i0}^S and (possibly) type- of school - from the set of regressors.

3.4 Additional caveats

An additional caveat to be considered is the possibility that students enroled in the technical or natural sciences & health tracks may face toughter entry-exams (except in those subjects which are common across all tracks) than those doing social sciences or humanities. In such a case, this may bias upwards their ATEs since high grades in higher education will be correlated with lower entry-exam grades.

To check this issue, Figures 4a and b display the distributions of *Bachillerato* and *Selectividad* grades by high-school track, respectively. We can observe that students in the technical track do better than the rest in both exams. Since there is no control for ability in these distributions, the most plausible interpretations of this fact is that either they have higher ability or that they are better prepared. Hence, the reader should be aware that all the results presented below are likely to be subject to this caveat that cannot be addressed since we lack information on the determinants of choice of tracks. By contrast, another type of selectivity bias with respect to the overall population of freshers in

UC3M, stemming from the fact that the freshers in our sample belong to a bilingual group, for which knowledge of English is required, can be addressed (see Section 5 below).

[Figures 4a and 4b about here]

4. Empirical results

The results obtained with the different econometric approaches are discussed in this section.

4.1 Least-Squares: OLS and IV

Tables 3a-3c show the results of estimating (7) by OLS and IV in pooled regressions. The dependent variable is the standardized (numerical) grade in each of the three subjects taken by fresher i , while the controls are: a gender dummy (female=1), a nationality dummy (foreigner=1), *Selectividad* grades (numerical), two dummies for type of school (charter=1 and private schools=1), three dummies for high-school tracks (natural sciences & health=1, technical=1 and humanities=1), a dummy for parental background (university degree=1), and finally three dummies for cohorts. Thus, the reference group corresponds to Spanish male students from public schools with a high school track in social sciences, whose parents lack a college degrees, and who took the (February) exams in the 2002/03 course.

[Tables 3a to 3c about here]

Before discussing the results, it is worth highlighting that, as mentioned in Section 2, we found high multicollinearity in all specifications between the controls capturing type of school and parental background. [14] In all instances, the estimated coefficients on the corresponding dummies were not significant but, when one of these controls was skipped from the regression, the estimated coefficient of the one left in showed up highly significant. Since the dummies for school type had higher significance level on their own than the ones for parental college attaintment, for sake of brevity we will only report in the sequel the specifications including the former control. Hence, an appropriate interpretation for the estimated ATEs of type of school is that they are capturing, in addition to their own genuine effect, other family background effects that we cannot separately identify. Further, to confirm that the *Selectividad* exam grades are in fact a post-treatment variable with respect to the choice of high-school track, we also ran a regression (not reported here) of those grades on the remaining set of controls in (7), yielding similar qualitatitive results to those obtained when the subject grades are used as the dependent variables. As discussed in Section 3, this implies that endogeneity could be a potential problem. For this reason, we will start by reporting results from

[14] As mentioned in Section 1, the correlation coefficient between the parental background dummy and a single dummy for non-public schools is 0.89.

simple regressions of subject grades on the individual characteristics, excluding the (potentially endogenous) *Selectividad* grades and the dummies for school type. These should be interpreted as partial correlations that could be very informative if we are unable to properly address the endogeneity biases.

Columns (1) in each of the three Tables present the partial correlations excluding both the *Selectividad* grades and the school type dummies. Columns (2) offer the results from estimating specification (7) with all inputs. In columns (3), the linear specification is augmented with interaction terms between the *Selectividad* grades and type of track to check whether the difficulty of this exam differs across tracks. Finally, the last set of columns displays the IV estimates where the *Bachillerato* grade is used as an instrument for the *Selectividad* grade. In all cases, robust standard errors are reported in parentheses.

The most important finding is that the IV estimates of (7) are very similar to the OLS ones. This is not surprising, given that the computed (partial) covariance terms in the RHS of (8), $Cov(\hat{D}, \hat{y}_0^s)$, are fairly small (0.121, 0.004 and -0.025 for the the technical, natural sciences & health, and humanities tracks, respectively). Moreover, the correlation between the *Selectividad* and *Bachillerato* grades is very high (0.96). To the extent that our assumption that the *Bachillerato* grades are a good proxy for academic performance before *Bachillerato* is a plausible one, this means that the *Selectividad* grades are also capturing a significant fraction of the students´ skills at the pre-treatment stage, implying that the OLS bias, ξ, is bound to be small. Therefore, this result yields favourable support to using OLS in estimating (7) to obtain a lower bound for the true ATE. This seems to be further confirmed by the results obtained from the partial-correlation regressions in columns (1) which are are qualitatively similar to the OLS and IV results. [15] Finally, the interaction terms do not seem to matter. In all instances, having followed a technical track seems to lead to a better performance in *Maths* and *Introecon* (and not significatively worse in *Econhist*) than having completed a social sciences track (reference category). By contrast, students with a humanities background tend to do uniformly worse. Interestingly, female students tend to perform better in *Maths* than male students whereas there are no statistically significant differences between domestic and foreign students.

[15] As regards the ATE of *Bachillerato* track, we find larger (in absolute value) estimates for the technical (positive) and the humanities (negative) tracks in columns (1) than in columns (2), indicating that these variables are the ones which are more correlated with the omitted covariates (*Selectividad* grades and type of school). Likewise, the estimated coefficients on the female gender dummy seem to be larger in columns (1) than in columns (2) for *Maths* and *Econhist*.

Given the similarity between the OLS and IV results, we will focus next on highlighting the main results in column (2) of each subject. The largest effects are found for the *Selectividad* grade and the technical track dummy. For example, an extra point in the university entry-exam leads to about 0.60 extra s.d.´s (1.54, 0.95 and 0.83 points, respectively) relative to the reference group in each subject. Likewise, having completed the technical track leads to 0.68 extra s.d.´s (1.7 points) in Maths and 0.45 s.d.´s (0.7 points) in *Introecon*, without any significant gain in *Econhist*, whereas 0.5 extra s.d.´s (1.25 points) in *Maths* are achieved by those who followed the natural sciences & health track. By contrast, the humanities track has a penalty of almost 0.6 s.d´s (1.5 points) in that subject. Regarding the effect of type of school (subject to the key interpretational caveat mentioned earlier), having attended a non-public school (or coming from a higher-educated family) is related to be a better grade on average. For example, coming from a private school leads to 0.3 and 0.25 extra s.d´s. in *Maths* (0.73 points) and *Introecon* (0.38 points), respectively, relative to coming from a public school. As regards gender, female students get 0.14 s.d.´s (0.38 points) more than their male classmates in *Maths*, without significant differences in the remaining subjects. Lastly, with the exception of the 2004/05 course, the cohort dummies are significantly negative. Despite the short sample period, this gives some support to the extended opinion among several pundits that training in high schools has been deteriorating over time due to expanding participation in secondary education. Notwithstanding, this effect might be contaminated by the presence of different instructors in two of the three subjects.

Finally, for completeness, we report in Table 4 the OLS results obtained from the "within" specification in (4). The estimates are of course quantitatively different from those in colums (2) of Tables 3a-3c since we are impossing the restriction $(\beta_2 / \delta_2) = 1$. Therefore, they should be interpreted as the net effects of the vector of covariates x_i on the gap between the subject and *Selectividad* grades. Yet, the previous qualitative findings remain similar in this restricted specification.

[Table 4 about here]

4.2 Quantile Regressions

The fact discussed earlier that we may not have well-behaved distributions in the outcome and in some of the other variables implies that least-squares coefficients may yield partial information. Accordingly, in line with a growing literature on the application of this technique to achievement production functions, we use quantile regressions (QR).[16] Following the well-known methodology first proposed in Koenker and Bassett (1978), the model of QR in

[16] Illustrations of the use of QRs in the literature on schooling outcomes can be found, e.g., in Eide and Showalter (1998), Levin (2001), and Marcenaro and Navarro (2007). Given the similarity of the OLS and IV estimates, the applied QRs are solely based on OLS.

this setup can be described as follows. Using numerical grades, let (y_{i1}, g_i) be a random sample, where g_i =(D_i, y^S_{i0}, c_i) and $Q_\theta(y_{i1}|g_i)$ is the conditional θ^{th} quantile of the distribution of y_{i1} given g_i. Then, under the assumption of a linear specification as in (3), the model can be defined as

$$y_{i1} = g_i'\beta_\theta + u_{\theta i1}\,, \qquad Q_\theta(y_{i1}|g_i) = g_i'\beta_\theta \tag{11}$$

where the distribution of the error term $u_{\theta i}$, $F_{u\theta}(\cdot)$, is left unspecified, just assuming that $u_{\theta i1}$ satisfies $Q_\theta(u_{\theta i}|g_i) = 0$. The estimated vector of QR coefficients, $\hat{\beta}_\theta$, is interpreted as the marginal change in the conditional quantile θ due to a marginal change in the corresponding element of the vector of coefficients on g, and can be obtained using the optimization techniques described in Koenker and Bassett (1982).

In order to facilitate the comparison of results across subjects, we choose different quantiles for each subject so that the percentiles become similar in terms of both numerical and categorical grades. These are: θ=0.25 (grade: 2.8, SUS), 0.75 (7.0 NOT) and 0.95 (9.5, SOB) for *Maths*; θ=0.10 (3.8, SUS), 0.80 (7, NOT) and 0.98 (10, MH) for *Introecon*; and, finally, θ=0.10 (4.5, SUS), 0.70 (7.2, NOT) and 0.98 (9.3, SOB) for *Econhist*. Tables 5a-5c report the estimated coefficients at the relevant quantiles (together with the regression at the median, i.e., at θ= 0.50) using the specification in the columss (2) of Tables 3a-3c. For convenience, we reproduce the OLS estimates in the first column (*average*) in order to compare the coefficients at the mean as opposed to the coefficients at the chosen quantiles of the conditional distribution of (numerical) grades. [17]

[Tables 5a to 5c about here]

The QR results offer valuable additional information to the one discussed above.[18] The key result in *Maths* is that the impact of private and charter schools (in the range of 0.2 to 0.4 extra s.d.´s or 0.5 to 1 points relative to public schools) is much weaker at the top quantile, in line with the prevalence of students coming from public schools at the higher part of the grade distribution. A similar effect is observed for the *Selectividad* grades (the most significant variable, together with the technical track), whose effect decreases throughout the distribution. The opposite effect holds for the humanities track. As regards the other subjects, the results are similar, with the only exception that completion of more science-based tracks does not seem to help in *Econhist.* The

[17] For the sake of brevity, we do not report the estimated coefficients on the cohort dummies. However, the pattern of negative coefficients for the 2003/04 and 2005/06 cohorts remains the same across quantiles.

[18] An F test on the joint equality of all the coefficients across the chosen quantiles yields p-values very close to zero, therefore rejecting the null.

natural sciences & health track even has a negative effect at the top quantile. [19] Finally, foreign students seem to perform better than native students at the higher quantiles.

5. Other selection biases

While the biases discussed in Section 3 affect the internal validity of our results, our sample of students has two characteristics that could lead to (favourable) sample selection biases and affect therefore the external validity of our conclusions.

The first one is that UC3M is considered to be one of the Spanish universities with the highest reputation in Economics.[20] That, in principle, could lead to attracting better students than other universities with a lower ranking in this field. Unfortunately, we do not have any control group in order to test for this selection bias. However, there is ample evidence that the mobility of students across regions is very low and the entry-exam grade requested by UC3M to get admission in the Economics degree is a low pass (5.0), despite being somewhat larger in LADE (6.0). These acceptance grades are similar to those requested by most universities. Thus, we conjecture that biases are bound to be minor in this respect.

The second potential selection bias stems from the fact that students in our sample belong to a group is taught in English. Given that Spain is one of the European countries with the lowest share of the population speaking foreign languages (44%), it could be the case that the freshers in this group are not representative of the population of freshers taking lectures in Spanish, which is an ample majority. An indication that this bias could be present is that the proportion of students coming from public schools in the bilingual group (42%) is significantly lower than the corresponding share in the total population of students completing higher-secondary education (66%).

In order to check whether our students in the sample are somewhat different from those enroled in non-bilingual groups, we have used another dataset regarding two groups (taught in Spanish) of freshers in Economics at UC3M during 2002-2006. The aggregate sample size for these control groups is 572 students. Information on these freshers was again obtained from the university archives and relates to gender, nationality, grades at the *Selectividad* exam and on whether students completed high school in the region of Madrid (CM) or in other Spanish regions. Unfortunately, we lack the remaining individual

[19] These results remain qualitatively similar when a multinomial logit setup is used to explain the probabilities of falling into each of the categorical grades, as we did in a previous version of this paper (see Dolado and Morales, 2007).

[20] According to the rankings published in the newspaper EL MUNDO (CAMPUS magazine) since 2007, UC3M is one of the two best universities in Spain to complete *licenciaturas* in Economics or LADE, together with UPF.

information which was used before in analyzing the determinants of outcomes for the bilingual group.

To control for this sort of sample selection bias we estimate a participation equation in the bilingual group as the first step in the conventional two-stage Heckman approach for selection correction. We use the pooled sample of all students (both from the Spanish and bilingual groups) which includes 958 students (=572+386). Given the scarce information available, we use the residence in CM (which is also available for the students in the bilingual group, but has not been used as a covariate in the previous sections) as the identifying variable. The insight for this choice is as follows: if the bilingual group is a (favourably) selected group from the population of students enrolled in Economics degrees at UC3M, then it is likely that a larger share of students from other Spanish regions will enrol in this group, given that there are very few universities in Spain offering bilingual courses.[21]

[Table 6 about here]

The first column in Table 6 presents the results from a first-stage probit model where the dependent variable equals 1 if a student belongs to the bilingual group and 0 otherwise. The covariates are gender, nationality, a dummy variable on residence (CM=1), (numerical) grade at the *Selectividad* exam and the cohort dummies (not reported). Results point out that being a foreigner and living outside CM increase the probability of belonging to the bilingual group whilst the other covariates do not have significant effects. Thus, our identifying strategy seems to work appropriately. The next three columns in Table 6 report the results the OLS estimation of the linear model in columns (2) of Tables 3a-3c but this time augmented with the inverse Mills ratio (*lambda*) from the participation equation. This last term turns out to be always insignificant and, despite some minor quantitative changes in the estimated coefficients, none of the qualitative results stressed above change with the selection correction. Hence, although we cannot completely discard selection biases with respect to the overall population of Spanish freshers in Economics, our results seem to provide valid inference in the context of UC3M undergraduates and, possibly, in relation to the overall population of similar students completing an Economics degree in other universities in Madrid.

6. Conclusions

Our main finding in this paper is that, conditional on our proxy for skills and all the interpretational caveats discussed above, the most important covariate related to academic success in *Maths* (for Economists) courses during the first year of an Economics degree is to have previously followed a technical track in the last two years of *Bachillerato*. Interestingly, we also find that having

[21] The fractions of students living outside the region of Madrid are 18.3% and 12.2% in the bilingual and Spanish groups, respectively. The averages of the entry-exam grades are 6.8 and 6.2 respectively, though a test for equal means does not reject the null with a p-value of 0.13.

completed a social science track, instead of a technical track, does not significantly improve performance in other two subjects with less (*Introecon*) o very little (*Econhist)* mathematical content but which require more prior economics training. Given that the social sciences track was designed by the Spanish education authorities to provide the appropriate training for high school students willing to become economists, it is fairly striking that a background in, e.g., mathematics, physics or chemistry leads to a better academic performance. Another interesting finding is that, among the best students, there is a majority of those coming from public schools. One possible explanation for this finding is that public schools seem to exert higher competition among the best students than non-public schools. This previous higher exposure to competition may help them to adapt better to the competitive environment of public universities. Finally, on average, females tend to do better than males.

As stressed throughout the paper, some of these findings have to be taken with caution when drawing education-policy implications, since we have not been able to address the issue of non-random selection, particularly in the choice of *Bachillerato* tracks. Moreover, one further qualification is that our outcomes are very short-term (grades at the end of the first semester of freshers) rather than longer-term indicators like final grades or term of completion. Notwithstanding, one possible preliminary lesson to be drawn is that, as suggested in the quotation at the epigraph of the paper, high-school students in the social sciences track, intending to later enrol in an Economics degree, do not seem to get enough math training and therefore struggle in the more mathematically intensive subjects. This may explain the high dropout rate (close to 30%) at the early stages of this degree. A possible solution to this problem could be to adopt the math courses in the technical track as compulsory subjects for those willing to enrol in Economics within the social siences track. Alternatively, extensive remedial math courses could be offered, as UC3M and other Spanish universties currently do. Whether these remedial courses are effective in helping less technically able students is in our future research agenda.

References

1. Bjorklund, A. Edin, P-A., Fredriksson, P. and Kruger, A. (2003), "Education, Equality and Efficiency: An Evaluation of Swedish School Reforms" SNS, Stockholm, Sweden.

2. Bonhomme, S. and U. Sauder (2007), “Accounting for Unobservables in Comparing Selective and Comprehensive Schooling”, CEMFI (mimeo).

3. Calero, J. (2006) “Desigualdades tras la Educación Obligatoria: Nuevas Evidencias”, DT 23/2006. Fundación Alternativas, Madrid.

4. Case, A. and A. Deaton (1999), "School Inputs and Educational Outcomes in South Africa ", *Quarterly Journal of Economics*, 114, 1047-1084.

5. Dearden, L., Ferri, J., and C. Meghir (1998), “ The Effect of School Quality on Educational Attainment and Wages” Institute of Fiscal Studies W.P. 98/3.

6. Dolado, J and Morales, E. (2007), ““Which Factors Determine Academic Performance of Undergraduate Students in Economics? : Some Spanish Evidence”, CEPR DP. 6237.

7. Eide, E. and M.H. Showalter (1998), "The Effect of School Quality on Student Performance ", *Economics Letters*, 58, 345-350.

8. Glewwe, P. and M. Kremer (2005), "Schools, Teachers and Education Outcomes in Developing Countries", forthcoming in *Handbook on the Economics of Education*. Elsevier.

9. Hanushek, E. (1986), "The Economics of Schooling Production and Efficiency in Public Schools”, *Journal of Economic Literature*, 24, 1141-1177.

10. Hanushek, E. (1995), "Interpreting Recent Research on Schooling in Developing Countries", World Bank Research Observer, 10, 227-246.

11. Koenker, R. and G. Bassett (1978), “Regression Quantiles”, Econometrica, 46, 33-50.

12. Koenker, R. and G. Bassett (1982), “Robust Tests for Heteroskedasticity on Regression Quantiles”, *Econometrica*, 50, 43-61.

13. Lagerlöf, J. and A. Seltzer (2008), The Effects of Remedial Mathematics on the Learning of Economics: Evidence from a Natural Experiment”, forthcoming in *Journal of Economic Education*.

14. Levin, J. (2001), " For Whom the Reductions Count: A Quantile Regression Analysis of Class Size and Peer Effects on Scholastic Achievement", *Empirical Economics*, 26, 221-246.

15. Marcenaro, O. and L. Navarro (2007), “ El Éxito en la Universidad: Una Aproximación Cuantílica”, *Revista de Economía Aplicada*, 15, 5-40.

16. Pozo, S. and C. Stull (2006), “ Requiring a Math Skills Unit: Result of a Randommized Experiment”, *American Economic Review P&P*, 92, 437-441.

17. Smith, J. and R. Naylor (2001), “Determinants of Degree Performance in UK Universities: A Statistical Analysis of the 1993 Student Cohort”, *Oxford Bulletin of Economics & Statistics*, 63, 29-60.

18. Todd, P. and K. Wolpin (2003), “ On the specification and Estimation of the Production Function for Cognitive Achievement”, *The Economic Journal*, 113, 3-33.

TABLES AND FIGURES

Table 1: Distributions of Grades by Students´ Characteristics

Covariates/Grades	*Maths* S=0	*Maths* AN=1	*Maths* SM=2	*Introecon* S=0	*Introecon* AN=1	*Introecon* SM=2	*Econhist* S=0	*Econhist* AN=1	*Econhist* SM=2	*U freq* *
Frequency	37.05	53.63	9.32	26.62	68.92	4.66	10.36	85.24	4.40	
Male	42.63	49.47	7.89	24.74	71.05	4.21	11.58	85.79	2.63	49.2
Female	31.63	57.65	10.71	28.06	66.84	5.10	9.18	84.69	6.12	50.8
Public	55.21	35.58	9.21	41.10	52.76	6.14	13.50	82.10	4.40	42.1
Charter	25.93	65.43	8.64	16.05	81.48	2.47	7.41	90.12	2.47	21.1
Private	22.54	67.61	9.85	15.49	80.28	4.23	8.45	86.62	4.93	36.8
Parent. (0)	58.76	36.22	5.02	44.78	53.02	3.20	16.56	78.71	4.73	72.7
Parent. (1)	19.23	71.56	9.17	12.68	77.34	9.98	7.75	86.63	5.62	27.3
Social Sc.	45.53	50.97	3.50	31.52	66.54	1.95	12.06	84.82	3.12	66.8
Tech.	9.09	63.64	27.27	9.10	77.78	13.13	5.05	85.86	9.09	26.2
NSc &Health.	7.69	92.31	0.00	23.08	76.92	0.00	0.00	100.0	0.00	3.1
Hum.	94.12	5.88	0.00	52.94	47.06	0.00	23.53	76.47	0.00	3.9
Spanish	36.23	54.78	8.99	26.38	69.57	4.06	9.86	86.38	3.77	89.3
Foreigner	43.90	43.90	12.20	26.83	63.41	9.76	14.63	75.61	9.76	10.7
Select. grades.**		0.668			0.610			0.498		

Note: (*) The figures in the last column represent the unconditional frequencies of each covariate. (**)The figures in the last row correspond to the correlations between the (numerical) grades in each of the subjects and the university entry-exam (*Selectividad*) grades

Table 2 . Types of *Bachillerato* tracks

Track	Common subjects	Specific subjects
Technical	Language & Literature Philosophy English History	Maths (A) Physics Tech. Drawing + 2
NSc & Health	-----	Biology Chemistry Maths (A) + 2
Social Sciences	-----	Maths (SS) Economics Geography + 2
Humanities	-----	History Latin Geography +2

Note: +2 means that students can take any other two optional subjects that they wish to. Maths (A) and Maths (SS) mean Advanced Maths. and Maths. for social sciences, respectively

Table 3a: Grades Production Function Estimates
MATHS

Dependent variable: Grades (standarised)

Variable	**OLS (1)**	**OLS (2)**	**OLS (3)**	**IV (4)**
Female	0.313^{***} (0.084)	0.137^{**} (0.066)	0.126^{*} (0.067)	0.137^{**} (0.068)
Foreigner	-0.263^{*} (0.162)	-0.068 (0.121)	-0.060 (0.112)	-0.068 (0.1208)
Charter	----	0.219^{***} (0.089)	0.321^{***} (0.114)	0.220^{***} (0.091)
Private	---	0.290^{***} (0.078)	0.336^{***} (0.092)	0.292^{***} (0.078)
Nat. Sc & Health	0.572^{***} (0.166)	0.523^{***} (0.165)	0.537^{***} (0.1734)	0.533^{***} (0.181)
Technical	1.084^{***} (0.101)	0.676^{***} (0.089)	0.715^{***} (0.058)	0.673^{***} (0.091)
Humanities	-0.894^{***} (0.107)	-0.569^{***} (0.095)	-0.612^{***} (0.220)	-0.567^{***} (0.100)
Select. grade	------	0.616^{***} (0.047)	0.619^{***} (0.052)	0.608^{***} (0.049)
Course_0304	-0.433^{***} (0.125)	-0.469^{***} (0.093)	-0.497^{***} (0.103)	-0.377^{***} (0.101)
Course_0405	0.078 (0.114)	0.052 (0.089)	0.040 (0.096)	0.046 (0.087)
Course_0506	-0.057 (0.112)	-0.375^{***} (0.101)	-0.412^{***} (0.106)	-0.378^{***} (0.119)
Select* Nat..Sc. & H.			0.469 (0.432)	
Select*Tech.			-0.021 (0.091)	
Select*Hum.			-0.084 (0.446)	
Nº Obs.	386	386	386	386
R^2	0.334	0.608	0.615	0.607

Note: ***, **, * represent significance at 99, 95 and 90% respectively.
A constant term is included. Omitted group: males, Spanish, public school, social sciences, cohort 2002/03.

Table 3b: Grades Production Function Estimates
INTROECON
Dependent variable: Grades (standarised)

Variable	**OLS (1)**	**OLS (2)**	**OLS (3)**	**IV (4)**
Female	0.082 (0.078)	0.137** (0.066)	0.078 (0.080)	0.077 (0.068)
Foreigner	0.051 (0.190)	0.204 (0.161)	0.205* (0.132)	0.203 (0.163)
Charter	----	0.108 (0.104)	0.250* (0.136)	0.104* (0.105)
Private	---	0.240*** (0.097)	0.329*** (0.111)	0.242*** (0.098)
Nat. Sc & Health	-0.037 (0.164)	-0.044 (0.112)	-0.011 (0.173)	-0.043 (0.120)
Technical	0.851*** (0.121)	0.449*** (0.113)	0.375*** (0.108)	0.453*** (0.116)
Humanities	-0.584*** (0.142)	-0.262* (0.151)	-0.212* (0.145)	-0.265* (0.151)
Select. grade	------	0.606*** (0.063)	0.558*** (0.062)	0.598*** (0.069)
Course_0304	-0.160* (0.096)	-0.233*** (0.079)	-0.222* (0.123)	-0.232*** (0.079)
Course_0405	-0.082 (0.111)	-0.150* (0.082)	-0.161 (0.116)	-0.154* (0.0827)
Course_0506	-0.002 (0.157)	-0.301** (0.137)	-0.288** (0.127)	-0.298** (0.139)
Select* Nat..Sc. & H.			0.098 (0.108)	
Select*Tech.			-0.470 (0535)	
Select*Hum.			-0.084 (0.446)	
Nº Obs.	386	386	386	386
R^2	0.171	0.436	0.446	0.435

Note: As in Table 3a

Table 3c: Grades Production Function Estimates: ECONHIST

Dependent variable: Grades (standarised)

Variable	OLS (1)	OLS (2)	OLS (3)	IV (4)
Female	0.240*** (0.100)	0.102 (0.088)	0.112 (0.091)	0.106 (0.089)
Foreigner	-0.066 (0.183)	0.071 (0.143)	0.081 (0.1149)	0.069 (0.144)
Charter	----	0.183* (0.106)	0.352** (0.154)	0.185* (0.106)
Private	----	0.118 (0.105)	0.274** (0.1125)	0.122 (0106)
Nat. Sc & Health	0.081 (0.177)	0.030 (0.169)	-0.089 (0.173)	0.034 (0.170)
Technical	0.342*** (0.123)	-0.050 (0.121)	0.147 (0.132)	-0.039 (0.123)
Humanities	-0.419** (0.188)	-0.121 (0.187)	-0.073* (0.164)	-0.130 (0.188)
Selectividad grade	-----	0.576*** (0.067)	0.523*** (0.070)	0.560** (0.073)
Course_0304	-0.164* (0.123)	-0.251*** (0.095)	-0.263* (0.138)	-0.247*** (0.096)
Course_0405	-0.144 (0.114)	-0.225*** (0.089)	-0.243* (0.132)	-0.224*** (0.089)
Course_0506	-0.099 (0.161)	-0.384*** (0.147)	-0.381*** (0.143)	-0.376** (0.144)
Select* Nat..Sc. & H.			0.068 (0.077)	
Select*Tech.			0.087 (0.122)	
Select*Hum.			-0.067 (0.060)	
Nº Obs.	386	386	386	386
R^2	0.047	0.296	0.446	0.293

Note: As in Table 3a

Table 4. Grades Production Function Estimates

Dependent variable: Subject grade – Selectividad grade

Variable	***Maths***	***Intrecon***	***Ecohist***
Female	0.075*	0.177**	-0.002
	(0.040)	(0.089)	(0.097)
Foreigner	-0.016	0.255	0.125
	(0.126)	(0.172)	(0.144)
Charter	0.169*	0.050	0.127
	(0.096)	(0.117)	(0.121)
Private	0.188**	0.116	0.068
	(0.087)	(0.102)	(0.083)
Nat. Sc & Health	0.508***	-0.044	0.081
	(0.184)	(0.220)	(0.188)
Technical	0.456***	0.224**	-0.039
	(0.094)	(0.113)	(0.118)
Humanities	-0.335***	-0.192	-0.135
	(0.105)	(0.165)	(0.194)
Course_0304	-0.549***	-0.338**	-0.361***
	(0.098)	(0.093)	(0.1115)
Course_0405	-0.387***	-0.250***	-0.326***
	(0.095)	(0.090)	(0.099)
Course_0506	-0.615***	-0.543***	-0.639***
	(0.121)	(0.151)	(0.136)
Nº Obs.	386	386	386
R^2	0.195	0.135	0.129

Table 5a. QR (and OLS). Maths

Dependent variable: Grades (standarised)

Covariates	Average	θ=25	θ=50	θ=75	θ=95
Female	0.137***	0.176*	0.166**	0.079	0.066
	(0.066)	(0.106)	(0.087)	(0.071)	(0.133)
Foreigner	-0.068	-0.305**	-0.042	0.007	0.487**
	(0.121)	(0.145)	(0.163)	(0.175)	(0.234)
Charter	0.219***	0.170	0.286***	0.320***	-0.234***
	(0.089)	(0.106)	(0.116)	(0.128)	(0.084)
Private	0.290***	0.260***	0.342***	0.337***	-0.061*
	(0.078)	(0.082)	(0.106)	(0.116)	(0.036)
NSc. &Health	0.523***	0.597***	0.534	0.480	0.457*
	(0.165)	(0.129)	(0.217)	(0.345)	(0.273)
Technical	0.676***	0.680***	0.614***	0.520***	0.638***
	(0.089)	(0.111)	(0.1233)	(0.097)	(0.148)
Humanities	-0.569***	-0.366***	-0.634***	-0.746***	-0.796***
	(0.095)	(0.080)	(0.182)	(0.194)	(0.228)
Select. grade	0.616***	0.704***	0.637***	0.593***	0.444***
	(0.047)	(0.056)	(0.065)	(0.051)	(0.060)
Nº Obs.	386	386	386	386	386
Pseudo-R^2	0.607	0.393	0.425	0.428	0.416

Note: As in Table 3a. Cohort dummies have also been included.

Table 5b. QR (and OLS). Introecon

Dependent variable: Grades (standarized))

Covariates	Average	θ=10	θ=50	θ=80	θ=98
Female	0.079	0.027	0.166**	0.020	-0.040
	(0.066)	(0.105)	(0.083)	(0.081)	(0.111)
Foreigner	0.204	0.039	0.175	0.188	0.418**
	(0.161)	(0.125)	(0.163)	(0.213)	(0.214)
Charter	0.108	0.178	0.213***	0.176*	-0.183**
	(0.104)	(0.206)	(0.287)	(0.103)	(0.092)
Private	0.240***	0.357***	0.387***	0.181*	-0.152**
	(0.097)	(0.143)	(0.156)	(0.106)	(0.074))
N Sc. &Health	-0.044	0.277	0.534	-0.268	-0.757*
	(0.113)	(0.229)	(0.678)	(0.245)	(0.143)
Technical	0.449***	0.231	1.413***	0.483***	0.538*
	(0.113)	(0.211)	(0.243)	(0.158)	(0.283)
Humanities	-0.262***	-0.536**	-1.684***	-0.17	-0.096
	(0.151)	(0.278)	(0.282)	(0.164)	(0.148)
Select. grade	0.606***	0.578***	1.637***	0.718***	0.753***
	(0.063)	(0.099)	(0.108)	(0.081)	(0.096)
Nº Obs.	386	386	386	386	386
Pseudo-R^2	0.436	0.264	0.425	0.375	0.506

Note: As in Table 3a. Cohort dummies have also been included.

Table 5c. QR (and OLS). Econhist

Dependent variable: Grades (standarised)

Covariates	**Average**	**θ=10**	**θ=50**	**θ=70**	**θ=98**
Female	0.102	0.089	0.126	0.061	0.084
	(0.088)	(0.165)	(0.163)	(0.0841	(0.211)
Foreigner	0.071	-0.035	0.102	0.217*	0.338*
	(0.143)	(0.225)	(0.163)	(0.125)	(0. 201)
Charter	0.183*	0.211	0.213	0.186	-0.169**
	(0.106)	(0.306)	(0.187)	(0.133)	(0.074)
Private	0.118	0.363*	0.387***	0.111	-0.052*
	(0.105)	(0.193)	(0.126)	(0.085)	(0.030)
NSc. &Health	0.030	0.257	0.193	0.097	-0.785***
	(0.169)	(0.442)	(0.378)	(0.214)	(0.243)
Technical	-0.050	0.133	0.014	0.093*	0.048
	(0.121)	(0.232)	(0.163)	(0.056)	(0.253)
Humanities	-0.121	-0.324**	-0.164*	-0.090	0.365
	(0.187)	(0.314)	(0.282)	(0.154)	(0.286)
Select. grade	0.576***	0.575***	0.637***	0.608***	0.581***
	(0.067)	(0.133)	(0.088)	(0.051)	(0.109)
Nº Obs.	386	386	386	386	386
Pseudo-R^2	0.276	0.204	0.425	0.241	0.367

Note: As in Table 3a. Cohort dummies have also been included.

Table 6: Probit and Grades Production Function Estimates (with selection correction)

Dependent variable: Grades (standarised)

Variable	**Participation Probit (Bil=1)**	**Variable**	***Maths***	***Intrecon***	***Econhist***
Female	0.024 (0.033)	Female	0.126** (0.062)	0.081** (0.041)	0.098 (0.092)
Foreigner	0.094** (0.045)	Foreigner	-0.072 (0.060)	0.195 (0.146)	0.076 (0.156)
Select. grade	0.125 (0.247)	Charter	0.189*** (0.089)	0.112 (0.107)	0.212* (0.115)
Residence (CM)	-0.168** (0.083)	Private	0.312*** (0.084)	0.274*** (0.102)	0.126 (0.112)
		N Sc & Health	0.496*** (0.195)	-0.041 (0.220)	0.027 (0.261)
		Technical	0.694*** (0.086)	0.473*** (0.106)	0.005 (0.131)
		Humanities	-0.554*** (0.181)	-0.293* (0.172)	-0.093 (0.242)
		Select. grade	0.592*** (0.046)	0.572*** (0.052)	0.556*** (0.059)
		Lambda	0.051 (0.047)	0.023 (0.062)	0.026 (0.075)
Nº Obs.	958	Nº Obs.	386	386	386
Pseudo- R^2	0.178	R^2	0.607	0.435	0.275

Note: As in Table 3a. Cohort dummies have also been included.

Figure 1: Distributions of Subject Grades

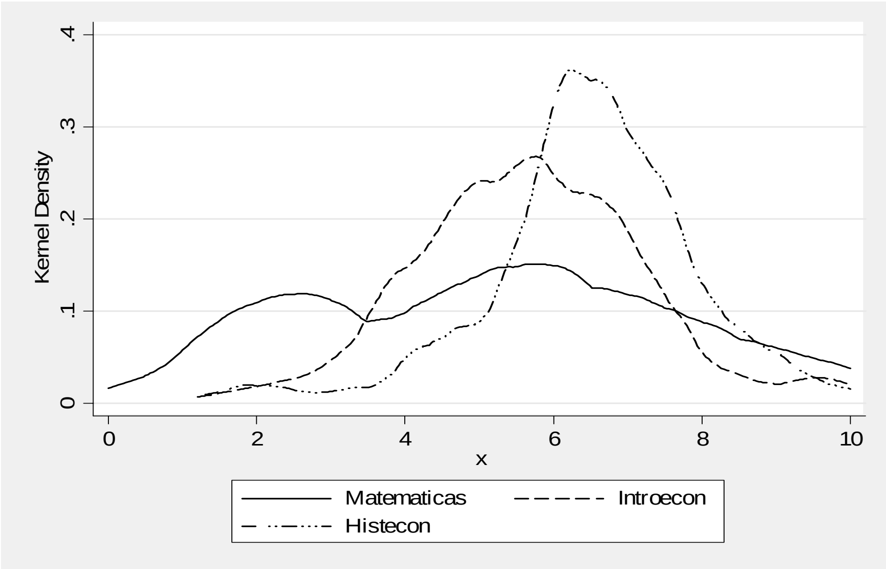

Figure 2: Distributions of *Selectividad* and *Bachillerato* grades

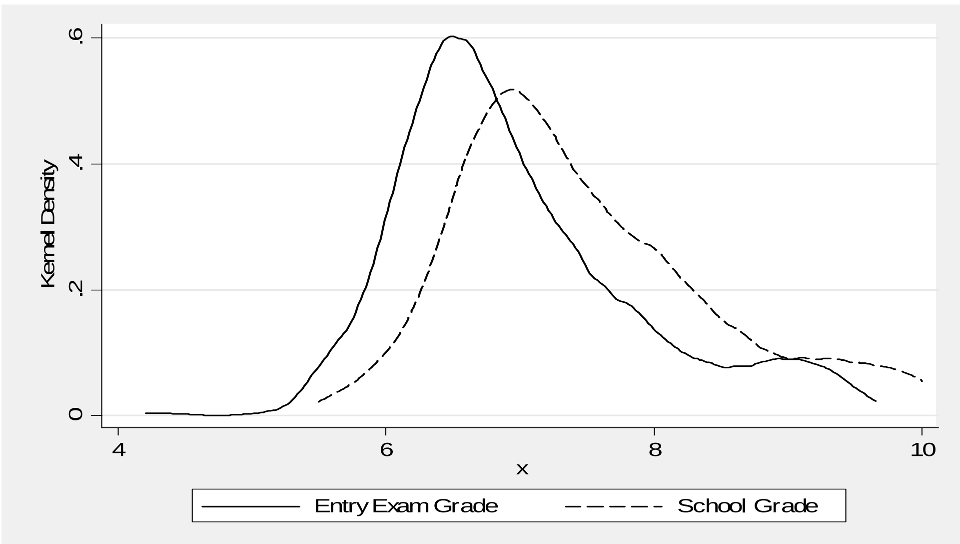

Figure 3: Gaps between *Bachillerato* and *Selectividad* grades by type of school

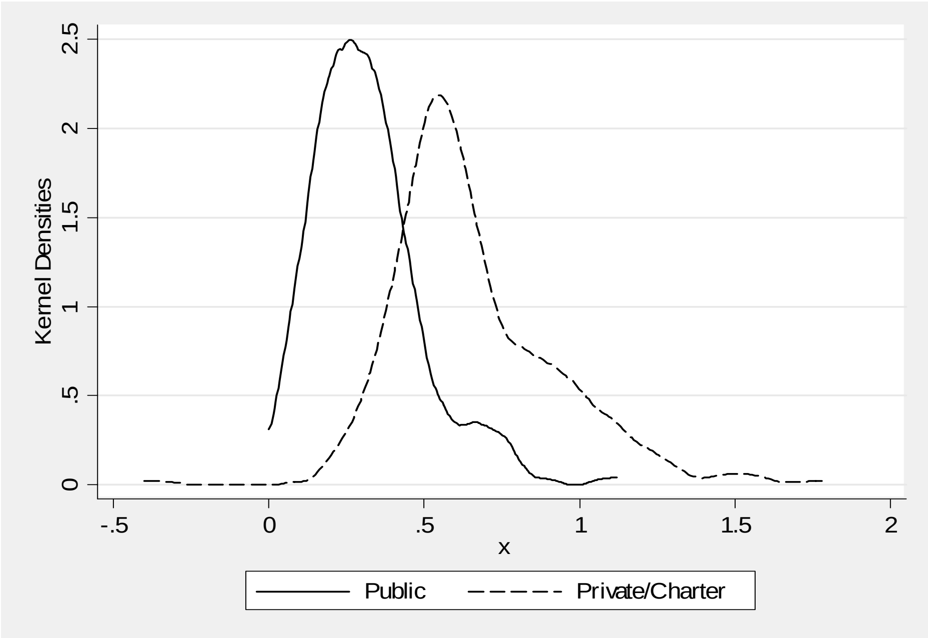

Figure 4a: Distributions of *Bachillerato* grades by track

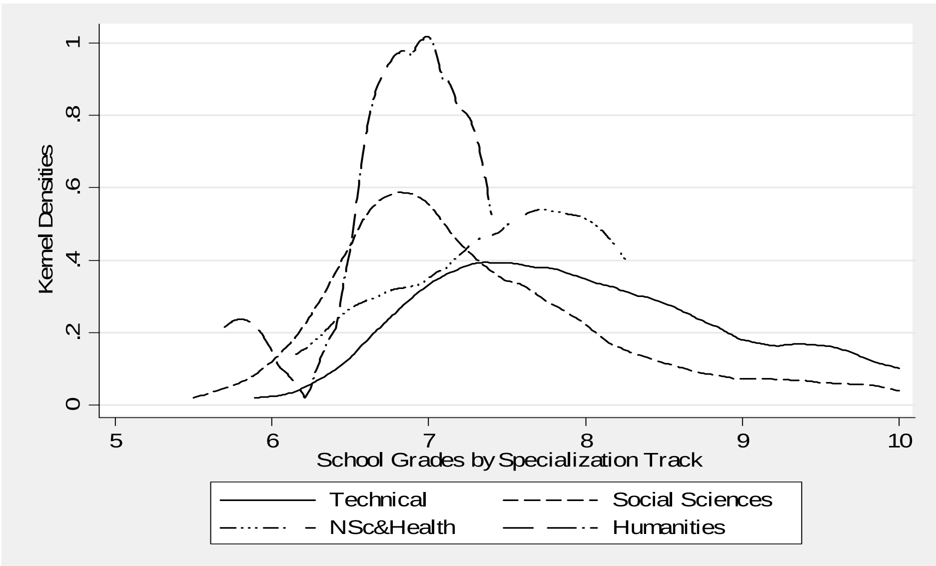

Figure 4b: Distributions of *Selectividad* grades by track