

The Effective Use of Student Time: A Stochastic Frontier Production Function Case Study

Peter Dolton, Oscar D. Marcenaro and Lucia Navarro



June 2001

Published by
Centre for the Economics of Education
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

© Peter Dolton, Oscar D. Marcenaro and Lucia Navarro, submitted February 2001

ISBN 0 7530 1441 6

Individual copy price: £5

The Centre for the Economics of Education is an independent research centre funded by the Department of Education and Employment. The view expressed in this work are those of the authors and do not necessarily reflect the views of the Department of Education and Employment. All errors and omissions remain the authors.

Executive Summary

In the economics of education there are relatively few studies which have focused on the mechanism of human capital acquisition. That is, exactly how do people acquire knowledge and what is the relationship between the learning environment and the educational achievement of those receiving the education? The relationship between student study time allocation and examination performance is little understood and is the subject of this research paper.

Education can be regarded as a production process in which a variety of individual study inputs are used to determine a multidimensional output. From the standpoint of the educational institution the way in which resources are used to transform students into well-qualified graduates is of importance. Should individual universities and the taxpayer fund longer and more time intensive courses or are the gains from the extra expenditure worthwhile? From the perspective of the individual student – how best should they allocate their time between formal study in lecture attendance, self-study and leisure and other activities?

The accepted technique for modelling the process of exam performance is the educational production function. This study models the existence of a university production function based on individual student data relating to examination performance. We model the allocation of student time into formal study (lectures and classes) and self study and its relationship to university examination scores using a stochastic frontier production function. The estimation of potential rather than an average educational production function provides an opportunity for estimating the extent of higher education inefficiency. This case study uses unique time budget data and detailed personal records from one university in Spain.

Our econometric results would suggest important policy implications for the university authorities and educational planners. In addition the results may be suggestive for the individual student in their choice of study time and potentially for parents seeking to support their sons and daughters in higher education. Our results suggest:

- Within the formal system of teaching in Spain, both formal study and self study are significant determinants of exam scores but that the former may be up to four times more important than the latter. Hence a student who wishes to maximise their examination score should attend all lectures and classes and minimise their absence from any formal tuition provided by the

university. A logical corollary to this result is that the student should not overindulge in leisure time.

- There is a clear payoff to minimizing the amount of time spent on travel and domestic activities. These results could also have implications for parents who wish to support their student sons and daughters.
- Most obviously for universities the significance of formal study time on performance suggests that they should do all that is in their power to encourage student attendance at lectures and classes or even to make them compulsory university authorities may need to review how many formal contact hours are necessary in each subject. Indeed, if universities operate in a quasi-competitive environment where student performance by university is compared and subsequent employment outcomes are used as performance indicators of universities, they may need to devote more resources to teaching their students.
- University authorities should review the amount of time taken to study for a degree as our results suggest that the academic year could be lengthened and the duration of a degree course shortened if more hours of lectures and classes were presented. Indeed this issue has been on the policy agenda in Spain with the possible shortcoming of degree studies from 5 to 4 years.
- A higher level of student state financial support is most conducive to more favourable exam performance.
- Family income does not seem to matter in the provision of advantage. What does offer an advantage is having provided an educationally privileged background with fewer brothers and sisters (to have to share parental attention).
- Finally, our results provide some support for the view that this ability is not symmetrically distributed among university students and that possibly each student may be constrained by what is possible for someone with their ability. Indeed, controlling for unobserved ability we obtain the result that input of self study time may not matter at all. We also find that self study time may be insignificant if ability bias is corrected for.

The Effective Use of Student Time: A Stochastic Frontier Production Function Case Study

Peter Dolton, Oscar D. Marcenaro and Lucia Navarro

1.	Introduction	1
2.	The Student Time Allocation Problem	2
3.	Stochastic Frontier Model	6
4.	The Malaga University Student Time Survey	9
5.	Econometric Results	13
6.	Identification, Endogeneity, Instrumental Variables and Ability Bias	22
7.	Conclusions and Policy Implications	31
	Appendices	33
	References	49

Acknowledgements

The authors acknowledge comments on a previous version of this paper presented in Newcastle University.

Peter Dolton is a Professor of Economics at the University of Newcastle. Oscar D. Marcenaro and Lucia Navarro are both at the University of Malaga.

The Centre for the Economics of Education is an independent research centre funded by the Department of Education and Employment. The view expressed in this work are those of the

authors and do not necessarily reflect the views of the Department of Education and Employment. All errors and omissions remain the authors.

1. Introduction

Education is a fundamental contributory factor in the social and economic development of a country. There are a growing number of studies that examine the role which human capital acquisition plays in the economy. Relatively few of them have focused their attention on the mechanism of human capital acquisition. That is, exactly how do people acquire knowledge and what is the relationship between the learning environment and the educational achievement of those receiving the education? Such questions are particularly important given the rapid expansion of all higher education sector in all OECD countries.

From an economic point of view, education can be regarded as a production process in which a variety of individual study inputs are used to determine a multidimensional output, in the form of present and future satisfaction¹. From the standpoint of the educational institution the way in which resources are used to transform students into well-qualified graduates is of importance. Should individual universities and the taxpayer fund longer and more time intensive courses or are the gains from the extra expenditure worthwhile? From the perspective of the individual student – how best should they allocate their time between formal study in lecture attendance, self-study and leisure and other activities? Most research which estimates how students achieve their examination grades simply examines the relation between pre-university and university exam scores controlling for personal characteristics and fails to consider how students spend their time in the study process. Indeed there has been a general lack of research on how student time (and its balance) transforms into examination performance. We address this issue in this paper.

Although there have been many studies of educational production the evidence would suggest that we are still a long way from understanding how education is produced in terms of how hours studying is transformed into knowledge. Therefore, there is a rationale for new empirical studies which attempt to shed further light on the process by which these different inputs are transformed into educational outputs.

¹ By present satisfaction, we mean the amount of free time which the subject derives from his status as a student. In contrast, future satisfactions is a result of the possibility of access to the job market under advantageous circumstances and the social recognition which a high level of attainment affords the student.

The accepted technique for modelling the educational process of exam performance is the educational production function. This study models the existence of a university production function based on individual student data on examination performance. We adopt a production frontier approach, the deviations within which, could be due to errors in specification or measurement or the inefficiency in the process of production. The estimation of potential rather than average educational production functions provides an opportunity for estimating the extent of this higher education inefficiency.

In particular we will investigate the level of inefficiency produced in the transformation of the use students make of their time into educational performance, using the stochastic frontier model. To do this we will use case study data from the higher education system in Spain. This approach is of particular interest if we bear in mind the virtual absence of studies which have followed this line of inquiry.²

The remainder of this paper is organised as follows: in section two we present a simple theoretical model of the students time allocation problem, in section three we set out the stochastic frontier model and the benefits in relation to other possible specifications. In section four we provide a description of the survey of students in one Spanish university. The simple production function econometric results are presented in section five. Section six examines issues of: identification, the endogeneity of pre-university exam scores and unobserved ability bias. Finally, in section seven we discuss possible policy implications and summarise the conclusions.

2. The Student Time Allocation Problem

The seminal paper on the allocation of time by Becker (1965) appeals to the problem of student time allocation in the motivation of his treatment. He then goes on to model time allocation in conjunction with income and the demand for goods which takes us away from the main topic of study in this paper. We are primarily concerned with student time allocation

² We have found reference to a dated study by Harris (1940) but virtually nothing since then. The exception is Lassibille and Navarro (1986) who present a deterministic model to explain the use which university students make of their time.

between study and leisure (and sleep) and do not consider student earnings and demands for goods³.

Students Preferences over Study Time and Leisure

Assume the student preferences over study time/leisure and exam performance can be represented by the utility function:

$$U=U(P,L) \tag{1}$$

where P=Performance in Exams

L=Leisure

Assume that exam performance is a sufficient statistic for all future earnings and prospects and hence consumption of goods. Assume $u_P, u_L > 0$, and that the utility function is convex⁴.

For convenience we will consider leisure to be a sum of two components. The first is the 16 hours of the day over which the individual student is “free” to choose between self study, leisure and sleep. The second is the notional 8 hours per weekday which can be apportioned to formal study in lectures and classes or “stolen” additional leisure. This framework is not a necessary condition for formal analysis but merely an analytical convenience to facilitate diagrammatic analysis.

Time Constraint and Exam Performance

Assume each student can convert time spent on self study, S , and time spent on formal education, F , into examination performance, P , but that this relation is conditional on their, individual specific, innate ability (or intelligence) A .

$$P=P(F,S,A) \tag{2}$$

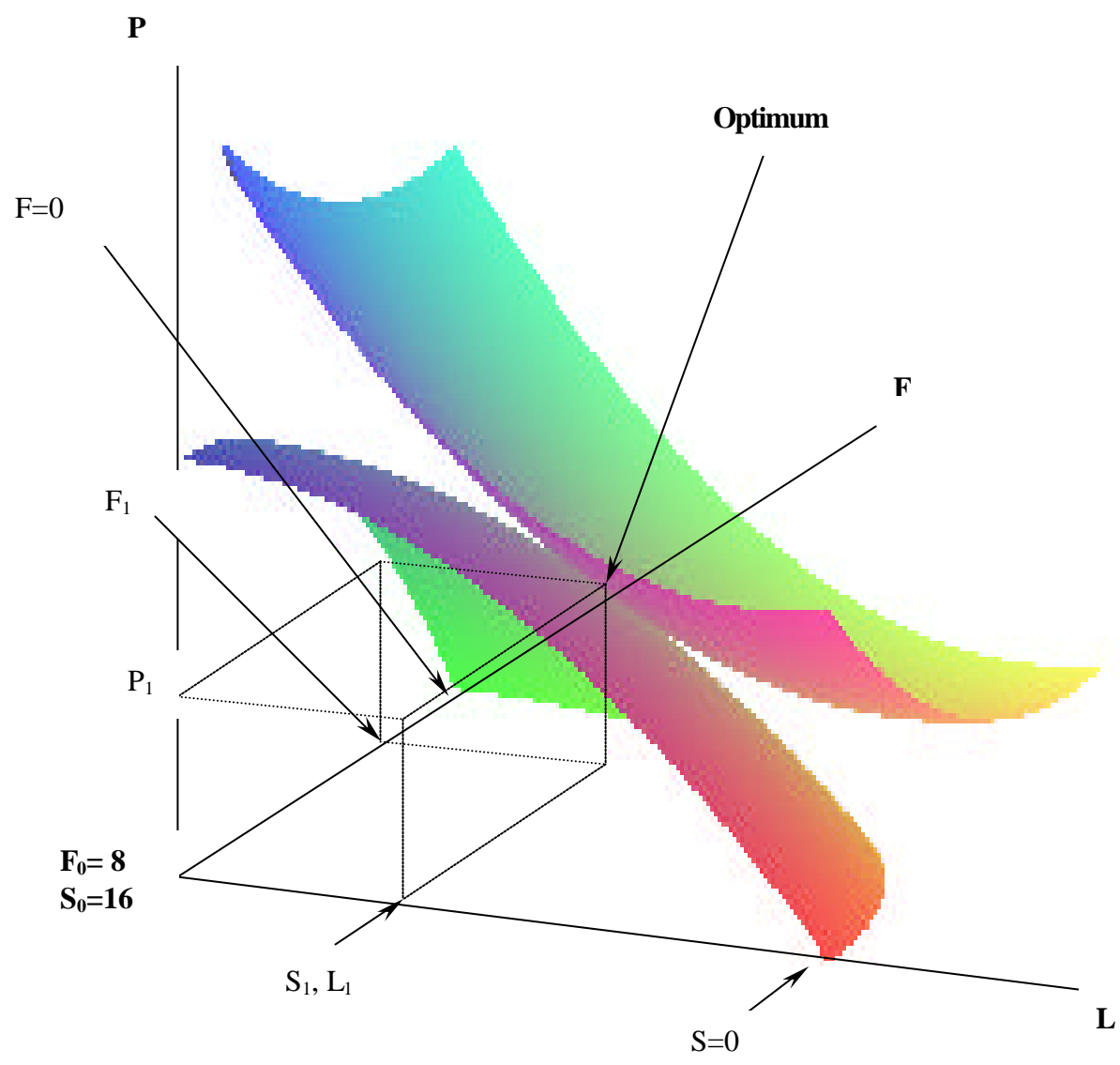
where $P_F > 0$, $P_S > 0$ and $P_A > 0$. However we may wish to assume that there is diminishing returns to study time after some amount of self study and formal education (i.e. $P_{SS}<0$, $P_{FF}<0$)

³ Whilst this is an inevitable simplification, in our data most students do not work in the labour market and hence Becker’s analysis is less directly relevant. In a context where students also have to decide how much time to spend working in the labour market this would not be true.

⁴ Notice that by assumption in this model students derive disutility from extra time spent studying. Our model would therefore be inappropriate for those exceptional students who derive positive utility from extra study hours.

which may also be individual specific. The position may be illustrated for an individual of fixed specific ability \bar{A} by the Figure 1 below⁵.

Figure 1

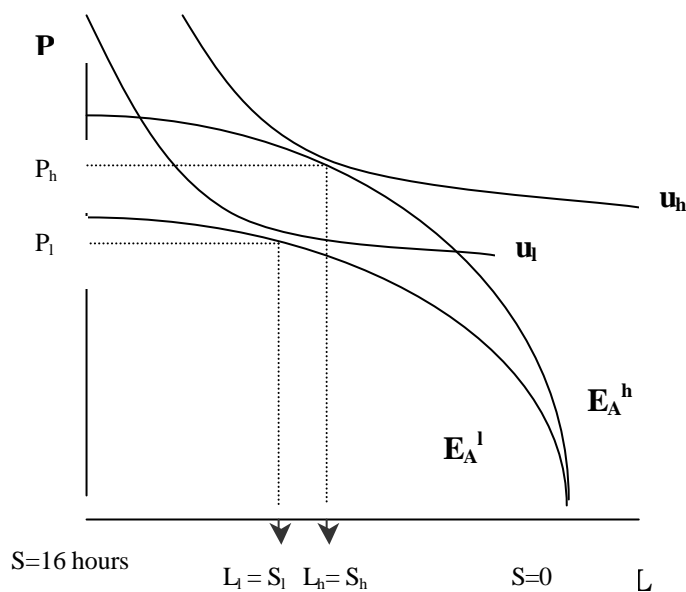


⁵ We assume for simplicity that the time spent sleeping is constant and equal to eight hours per day.

For the individual represented in this figure his/her utility would be maximised by taking L_1 leisure which results in S_1 time spent on self study, F_1 is the time spent on formal education, and exam performance P_1 . Notice in this framework $(S_0 - S_1)$ is the amount of “free” leisure and sleep time taken and $(F_0 - F_1)$ is the amount of additional “stolen” leisure taken which is non-attendance at lectures and classes.

Notice that this simple theory is rich enough to explain the possibility that some individuals who allocate less time to study may end up with higher exam performance, simply due to their higher ability and their more efficient conversion of study time to exam performance. This position is illustrated in Figure 2. For convenience assume formal study time to be fixed and consider only the choice of self study time. This diagram illustrates two individuals, a high ability person, h , and a low ability person, l .

Figure 2



Even with identical study/leisure preferences it is possible with a different self study time to exam performance frontiers indexed by ability, to generate the situation in which the high ability student may study for less time $S_h < S_1$ and still achieve higher exam performance

$P_h > P_l$. Of course the position in Figure 2 is only one possibility as, in general, the optimal choice of L and P depends on the shape of the E transformation and preferences⁶.

3. Stochastic Frontier Model

As outlined previously, we can compare the behaviour of a student to that of a firm which attempts to obtain an output by the transformation of a set of inputs. In general terms, this process can be represented by the following equation,

$$y_i = \mathbf{a} + X_i' \mathbf{b} + \mathbf{e}_i \quad i = 1, \dots, n \quad (3)$$

y_i being a measure of educational performance of individual i , x_i is a vector of their explanatory variables, \mathbf{e}_i a random disturbance, $\mathbf{\beta}$ a vector of slope coefficients and α a fixed but unknown population intercept. The size of the sample is represented by the value n .

The idea of this model is that each student's examination performance is affected by random factors, which are inherently unobservable and distributed normally. These may be associated with assignment to an inspiring teacher, being a member of a good mutual or self help study group, finding the ideal textbook to study from and a whole array of other stochastic factors. The second element which is unobservable in a student's potential performance is that their achievement potential is constrained by their inherent (unobservable) ability. This means that each student is limited by how effectively they can "convert" study hours into favourable exam results. The frontier in this context is notionally provided by the students who are most efficient at this conversion. We can effectively measure all other students "inefficiency" or degree of lower ability (as measured relative to the most able students in their cohort). It may be appropriate that the distribution of this unobservable term is asymmetric and we should explicitly model this in a way that allows this to be tested. The stochastic production function facilitates this. The likelihood is that the asymmetry could result from the sorting process of higher education, as it will only admit the top 35% or so, of students from the pre-university exam results distribution (see Appendix A for details of the participation rate in Spain). This means that the selected population who enter university are

⁶ I.e. there are the analogue of income and substitution effects when washing out how P and L change as

the selected right tail of the ability distribution⁷. Figure B1 shows how those who enter university are a selected subsample of the whole potential applicant population. This is the main rationale for the use of the stochastic frontier production function.

If we define y_i as the maximum potential performance which students can obtain for any given combination of inputs, the equation (3) can function as an educational frontier production model. This representation requires some assumption concerning the disturbance term. The two hypothesis which appear to satisfy the greatest level of acceptability, lead us to differentiate between the deterministic frontier model and the stochastic frontier model. Both models have in common the parametric nature of their specifications. Doubtless, the first result of considering that any deviation of an observation from the theoretical maximum potential is to be attributed solely to some kind of inefficiency in the educational process of production. From the analytical point of view, this assumes,

$$y_i = \mathbf{a} + X_i' \mathbf{b} - u_i \quad i = 1, \dots, n \quad (4)$$

$$u_i \geq 0$$

where u_i represents the inefficiency term⁸. In contrast, the stochastic frontier production, as outlined by Aigner, Lovell and Schmidt (1977), Meeusen and Van den Broeck (1977) and Battese and Corra (1977) rely on the premise that the deviations from the production function are due to statistical noise. Such a stochastic factor cannot be attributed to the process of production, and hence should not be embedded in the inefficiency term. The equation (3) representing this final hypothesis is expressed thus⁹,

$$y_i = \mathbf{a} + X_i' \mathbf{b} + v_i - u_i \quad i = 1, \dots, n \quad (5)$$

$$u_i \geq 0$$

A changes.

⁷ Notice from Figure B1 that this selection process is not a strict truncation as some of these students with low pre-university scores (4.0-6.0) decide to go university and some do not. Formally the stochastic frontier requires a strict truncation, hence the model is only an approximation of the data.

⁸ Aigner and Chu (1968) suggested two methods for estimating the parameters, assuming that the residuals u_i are positive. These methods are linear programming and quadratic programming.

⁹ As a result this model can be regarded as a generalisation of the standard regression model, the distinguishing feature of which is the presence of a one sided error (u_i).

where v_i is usually assumed to be a normally random variable (distributed independently of u_i) with mean zero and variance \mathbf{s}_v^2 , and u_i a non negative error¹⁰ typically assumed to be a half-normal distributed variable, with $\mathbf{s}_u^2 > 0$. Furthermore, we assume both components of the compound disturbance to be independent and identically distributed (i.i.d) across observations. In this model $\lambda = \mathbf{s}_u^2 / \mathbf{s}_v^2$, which is a measure of the degree of asymmetry of the ($v_i - u_i$) disturbance term. The larger is λ the more pronounced will be the asymmetry and the correspondingly the OLS estimation is less justified.

Several other specifications can be assumed for the u_i inefficiency term, apart from the half normal distribution Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977) presented a model which introduced an exponentially distributed disturbance. Later, Stevenson (1980) and Greene (1980) development an alternative specification which used a gamma distribution¹¹. The difficulties of interpreting the latter have led to a greater number of models which use a half normal or exponential specification. It appears that there is no objective criteria for choosing between the two specifications apart from the judgement of the individual researcher. Nevertheless, Battese and Coelli (1988) suggested that the half-normal is the most useful formulation which we could use.

It should be pointed out that there are other methods of analyzing production data, for example Data Envelopment Analysis (DEA)¹². Here, for reasons of space, we are not going to examine this method of analysis¹³. However the motivation for DEA techniques is the same as that which leads us to propose a stochastic frontier model. In contrast to the deterministic approaches of the deterministic frontier and DEA models, the stochastic frontier allows that the variance observed in student performance to be attributed not only to inefficiencies on the educational system but also to incomplete model specification or student heterogeneity. This comparative advantage which the stochastic model has proven to be important when the educational system is analysed, given that the complexity of the factors making up the process of production is such that the factors which can be observed in practice, only make up a small

¹⁰ If we were estimating a cost function u_i would be a non positive error.

¹¹ See Fried, Lovell and Schmidt (1993) for a broader discussion of this issue.

¹² For applications of this technique in the field of education, see Johnes and Johnes (1993), Chalos and Cherian (1995) and Kirjainen and Loikkanen (1998).

¹³ Norman and Stocker (1991) present a comprehensive description of this type of analysis.

proportion of the whole. Consequently, whatever deviation there is from the maximum attained performance will contain a strong stochastic component, the identification of which will prove to be crucial when drawing conclusions referred to possible inefficiency sources. There are two additional reasons for not using DEA in our analysis. Firstly, the parametric approach is easier to interpret for firms as institutions. Secondly, we do not need to identify individual observations as inefficient or measure the degree of inefficiency associated with any particular student.

4. The Malaga University Student Time Survey

The data we use to study the student time allocation process is taken from a survey conducted in April 1999 on first and final year students from the different qualifications offered at the University of Malaga. In total, the sample comprises 3722 observations taken from students from forty different subject areas¹⁴. In the survey information was collected about personal characteristics, family and school background and academic attributes. Detailed information was requested relating to the amount of time they dedicated to their normal activities. A clear distinction was made between time use on an average weekday and at weekends (details of the questions on time use are reported in Appendix B). The data were collected in the classroom by using a self-completed questionnaire which used individual student identification numbers. This procedure ensured confidentiality and anonymity. The names of each student were not recorded and it was made clear that the survey was not for official university purposes. Hence students were not left with any misunderstanding about the data collection. It was emphasized that the survey was for research purposes only and the data would not be retained or used by the university administration for academic, teaching or assessment purposes. Hence students had no incentive to lie or misreport their responses. Student identification numbers were later used to merge pre-university examination records from central university administrative data.

There is the potential for some bias since the respondents were those who had attended university classes when the survey was carried out, as a result absent students were

¹⁴ This sample represents 9.5% all students who matriculated at the University of Malaga during the academic year 1998-1999.

not followed up. Since attendance is around 60% there are a significant minority who do not appear in our sample. This may lead to a higher response rate amongst the more successful students (who usually attend classes more regularly) and the reader should be aware of this when generalizing from the results in this paper. There is also the possibility that sampling only from those attending lectures that we have a sample which is biased in relation to the relative importance of the formal study versus self study balance. If there are a significant number of successful students, not in our sample, who utilise self-study more predominantly in their study schedule then we may understate the importance of self study relative to formal study in the production of exam results. However if absentees are predominantly the worst performing students (which is most likely), then our results may slightly understate the importance of formal study time. All tables show t-statistics alongside the coefficient estimates, to facilitate robustness interpretations of the results.

A major difficulty in any study of time use is to get respondents to accurately remember their time allocation. Juster and Stafford (1991) report that there are many potential biases in asking people to record time use. They suggest that asking respondents to keep a diary is a preferred survey method¹⁵.

Unfortunately this was not possible in this study. Juster and Stafford (1991) do however offer some reassurance to this study in an important respect. Namely they suggest that reporting error is minimized when responses involve recording “daily work patterns” with “regular schedules”¹⁶. This finding is of most importance if we consider recording information about student study time. All students know how many hours of contact time are involved in their weekly time table, hence to calculate actual contact time they only have to make some adjustment for non-attendance. Likewise the remainder of their weekly schedule will have a regular pattern which may facilitate a reasonable estimate of self study time.

Further support for the validity of our data comes from the construction of the questionnaire which prompts them to be logically consistent in terms of their total hours adding

¹⁵ However recent evidence, see Mulligan, Schneider and Wolfe (2000) suggests that time budget studies of this type have biased samples since participating in the survey interferes too much with the lives of the subjects. Hence this result would support our data collection method.

¹⁶ See Juster and Stafford (1991), p. 482.

up to these that are available. Around 81% of our respondents record a time budget which adds up to a consistent 24 hours day.

Looking at our average student our data suggests they allocate their weekly time according to the Table 1 below. This allocation is a plausible one. From Table 1 we see that the average person’s time is not quite fully exhausted both on weekdays and weekends. This is a common finding of time budget studies and may be accounted for with other miscellaneous time consuming activities not listed in our questionnaire.

Table 1: Students weekly time allocation (hours)

	Weekdays	Weekend
<i>Formal Education</i>	28.4	0
<i>Self Study</i>	15.12	4.80
<i>Private Tuition</i>	0.95	0
<i>IT/Language</i>	2.25	0.3
<i>Travel/Domestic</i>	8.87	2.0
<i>Leisure</i>	20.46	14.74
<i>Paid Work</i>	1.63	0.5
<i>Sleep</i>	38.8	20.84
Total	116.48	43.18

Further biases are possible in the recording of time use. Any measurement error which is systematically related to observed characteristics, e.g. gender, is not a problem as we can condition for this in our estimation. In addition any “pure” measurement error in time recording will also not be a problem since, provided it is random, its influence is captured in the stochastic error term. Of more concern is the possibility of bias generated by a systematic error based on unobservable characteristics. One important example might be that those students who performed badly on their exam might seek some self-justification by underreporting their study time -hence allowing themselves to find an excuse for their poor performance- which did not involve recognising that they may have low ability. It has to be acknowledged that there is very little which can be done about this type of measurement error.

In the Appendix B, two tables are presented with the statistics describing the variables used in our estimations. In the first of these the means and standard deviations for the total number of subjects is presented, as well as differences by gender. After deleting observations from the sample which had missing values of one or more of the variables our sample is

reduced to 1976 students. Definitions of the variables are given in the Appendix B. The tables in this appendix indicate that women constitute slightly more than 50% of the sample.

This pattern is specially marked in the areas of Health, Arts and Non Technical University School, as oppose to the Pure Sciences and Engineering (Higher and Technical University Schools) where the proportion is still low¹⁷.

When examining the use students make of their time, the first factor to take into account is the time spent attending university classes. The table shows that men, on average, spend the same amount of time attending classes as women each day, but around two hours less in self study. This factor could lead one to think that women put greater effort into their studies, which could be an explanatory factor for their higher performance. This, in quantitative terms, translates into an average grade of almost 0.4 points higher. Something similar happens to the time spent attending IT and language classes. Both, women and men, spend the same amount of time receiving supplementary private tuition, but the latter spend, on average, twenty minutes more acquiring IT and language skills.

The time spent on travel and domestic tasks is higher for women who spend, on average, four hours and forty minutes whereas men spend only two hours and forty. From this we can infer, on the one hand, that women usually participate more in domestic tasks and on the other, that according to our information, 52 % of male students reported to have their own means of transport. In contrast, only 29 % of women responded affirmatively to this question. The latter factor proves to be an important saving in time spent on travel. Finally, it could be interesting to remark that a 43 % of men declared that the principal reason why they decided to study at University was to earn more money by doing an university degree and/or to have more chance of finding a job, but only a 27 % of women declared this as the principal reason.

In Table B2 in the Appendix B, a descriptive analysis of the variables is presented, by subject. The main conclusions are presented here. A large difference can be observed in the average performance of students. Thus students of Pure Sciences and Engineering (Higher and Technical University Schools) demonstrate a low average grade, perhaps due to the difficulty of their studies and the high proportion of male students (which could explain that Engineering students are those who spend more time attending private classes). At the other extreme,

Health students are found to have around two points more than the average score, something which could be explained by the rigorous selection process they endure prior to university entrance (as showed by the average university entrance marks) and, equally, to their greater capacity to study. In addition, the variable “number of hours of self study” shows that students of Health, Pure Science and Engineering are the ones who dedicate the most time to their studies. This fact is specially relevant to the first of these subjects with approximately five more hours of study than Non Technical University Schools students (those who spend the least time of their studies). As is logical to assume differences of opposite sign are reflected in the amount of time spent on leisure for both groups of students.

Finally, those who come from “Non Technical University Schools” have the smallest percentage of fathers and mothers who have undertaken university studies, this could be an important explanatory variable in describing the university performance of this type of student.

5. Econometric Results

In this section we discuss the results obtained in the estimation of the stochastic frontier specified in section 2 (equation 5). As we pointed out in the introduction, it is not quite obvious what the outputs of educational process are; i.e. it has a multidimensional output. Bowles (1970) suggests the educational system performs two primary economic functions: socialisation and selection. The first function, socialisation, involves the indoctrination of values and beliefs. The second function, selection, involves the direct effects of schooling on the productivity of the workers¹⁸. Unfortunately neither the social nor the economic dimension are directly quantifiable from our data. Our measurement of educational output is based on the average scores obtained by the students during the first semester (academic year 1998-99). These achievement scores must be considered as proxies for productive performance because of the previously outlined unobservable stochastic elements¹⁹.

Columns one and two of Table 2 contain the Ordinary Least Squares (OLS) and maximum likelihood (ML) estimates of the educational production function. Column one of that

¹⁷ The qualifications which have been included in each area are detailed in diagram 1 of the appendix A.

¹⁸ The relationship between education and productivity has been broadly developed by the “Human Capital Theory” and the “Screening Hypothesis”.

table records the results obtained by OLS and column two presents the ML estimates of the first specification, both use raw (non normalised)²⁰ examination scores (which take no explicit account of subject differences in scores) as dependent variable.

The overall goodness of fit of this model estimated by OLS (as indicated by $\bar{R}^2 = 0.22$) may be considered satisfactory given the difficulty of observing the heterogeneous factors which impact on the educational production process. On the other hand, the result of the likelihood ratio test indicates that the model estimated by ML is significant at standard tolerance levels (for all the specifications).

Estimation by OLS gives unbiased and consistent estimates of all parameters of the frontier function with the exception of the constant term. Hence, we get essentially all the information we would like except the position of the frontier. As a result, the slope coefficients generated by OLS are similar to those obtained by ML. The major difference will be found in the estimation of the intercept term (α), due to the inconsistency of the OLS estimator. The empirical results confirm this theoretical discussion. In fact, the intercept term shows the greatest divergence between both estimates. Hence we will only discuss the coefficients obtained by means of ML estimations in all the specifications.

Examining the variables relating to personal characteristics, age has a positive impact on educational achievement, this may result from the maturity acquired by doing other things before studying or as a consequence of a increased capacity to organise their studies and a better knowledge of the university framework. Alternatively delayed entry to university could have made the student more determined and focused hence more efficient with their time. The result relating to the gender variable is striking. According to the descriptive statistics, women (reference group) attain the highest scores, however the gender coefficient of this variable is insignificant. This could be due (as highlighted in section 3 of this paper) to the fact that this group probably perform better because of the greater time spent studying. In this case it would be the variable “self study” which would explain the higher female students’ output.

¹⁹ Some studies of the educational system’s productivity have used achievement test scores as an output measure. Hanushek (1986) has discussed the shortcomings of this measure.

²⁰ Later in this section we normalise university exam scores to control for subject differences in selection and assessment. Here, in Table 2, we simply use raw scores.

Table 2: Stochastic Educational Production Function (output non normalised)

Variables	Specification I				Specification II		Specification III	
	OLS		ML		ML		ML	
	Coefficient	t	Coefficient	t	Coefficient	t	Coefficient	t
Constant	4.43***	6.852	5.80***	9.358	5.739***	8.905	7.811***	16.199
Personal Characteristics								
Age	0.060***	3.496	0.066***	3.535	0.071***	3.661		
Gender	0.046	0.493	0.030	0.330	0.018	0.198	0.046	0.526
Married	0.410	1.065	0.237	0.440	0.208	0.384	0.389	0.737
Nationality	-0.110	-0.252	-0.210	-0.571	-0.232	-0.645	-0.314	-0.851
Geographic Area	0.116	1.200	0.090	0.968	0.075	0.799	0.066	0.729
Own transport	0.078	0.922	0.086	1.027	0.081	0.967	0.002	0.031
Time Use^a								
Formal Education	2.513***	4.532	2.723***	5.311	2.672***	5.214	2.052***	3.953
Self Study	0.543***	2.704	0.626***	3.335	0.640***	3.428	0.623***	3.417
Private Tuition	-2.238**	-1.972	-2.175**	-2.055	-2.021*	-1.888	-1.517	-1.382
IT/Language	-0.313	-0.586	-0.138	-0.252	-0.102	-0.187	-0.379	-0.720
Travel/Domestic	-1.146***	-3.124	-1.181***	-3.389	-1.159***	-3.384	-1.117***	-3.361
Leisure	-0.175*	-1.703	-0.206**	-2.042	-0.210**	-2.099	-1.169*	-1.695
Paid Work	-0.232	-0.615	-0.257	-0.699	-0.232	-0.626	-0.379	-1.025
Subjects								
Arts	0.299**	2.241	0.394***	2.755	0.409***	2.892	0.433***	3.183
Health	0.366*	1.646	0.339	1.323	0.395	1.562	0.681***	2.832
Engineering	-1.320***	-6.896	-1.303***	-6.450	-1.324***	-6.581	-1.395***	-7.034
Pure Sciences	-0.902***	-5.163	-0.792***	-4.655	-0.807***	-4.776	-0.825***	-4.947
Non Technical University	0.362***	2.674	0.515***	3.583	0.577***	4.046	0.522***	3.797
Technical University	-0.971***	-6.299	-0.931***	-5.724	-0.863***	-5.263	-1.000***	-6.512
Schools	0.427***	3.655	0.460***	3.915	0.447***	3.827	0.481***	4.308
Parents' Characteristics								
Mother university studies	-0.004	-0.023	-0.050	-0.289	-0.043	-0.252	-0.037	-0.219
Parents divorced	-0.219	-1.212	-0.244	-1.316	-0.192	-1.042	-0.184	-1.038
Orphan	-0.108***	-3.096	-0.100***	-3.076	-0.978***	-2.980	-0.101***	-3.105
Family size	-0.0002	-0.271	-0.0003	-0.036	-0.0002	-0.208	-0.0003	0.320
Family income	-0.120	-0.571	-0.150	-0.689	-0.183	-0.843	-0.036	-0.175
Residence								
University Residence	-0.120	-1.129	-0.111	-1.071	-0.102	-0.976	-0.114	-1.123
Rent flat	0.095	1.141	0.122	1.468	0.151*	1.819	0.148*	1.829
Motivation of the students								
Satisfaction	-0.272***	-3.354	-0.269***	-3.348	-0.267***	-3.327	-0.278***	-3.556
Ambition								
Other characteristics related to the students' background								
Grant	0.0008***	2.564	0.0009***	2.916	0.0009***	3.198	0.0009***	3.049
Bachillerato L.O.G.S.E.					-0.227**	-2.044		
Via F. P.					-0.517***	-3.486		
Via R.E.M.					0.840	1.494		
First year student							-0.943***	-11.94
Parameters for compound error								
λ			1.623***	7.338	1.623***	7.464	1.504***	7.602
$S_{\epsilon} = (S_v^2 + S_u^2)^{1/2}$			2.258***	22.74	2.246***	23.05	2.138***	24.029
				1		6		

S_v^2		1.402	1.389	1.401
S_u^2		3.695	3.657	3.170
Number of observations	1976	1976	1976	1976
\bar{R}^2	0.22			
$F_{(27,1949)}$	17.09***			
$-2(\log_R - \log_U)$		318.86***	354.90***	235.89***

Note ^a: The time use variables are expressed in minutes per hour.

Note: * coefficient significantly different from zero at 10% confidence level; ** coefficient significantly different from zero at 5% level; *** coefficient significantly different from zero at 1% level.

Marital status, nationality²¹ and geographic area are not significant at the standard confidence levels. The indicator variable representing whether a student has their own means of transport has a positive but insignificant coefficient, due possibly to the fact that these students will make important savings in their time spent on travel, and therefore its effect is showed by “travel/domestic” variable.

The second group of explanatory variables measures the use students make of their time. The principle result which deserves attention is that the time spent in formal university study in lectures, seminars, classes and laboratory sessions is positive and highly significant in the determination of student performance. This suggests that there is a direct effect of increased hours spent at the university in formal study. What is less clear in this result is the extent to which this result reflects two different effects: firstly, the higher number of hours provided by the university for the study of a particular subject or secondly the higher rate of attendance by the student at the formal sessions which have been provided for him or her. Unfortunately with our data it is not possible to distinguish between these two separate effects. All we know is the number of hours spent in formal contact time. We do not know how many hours were scheduled for that student.

An equally important result is that self study time is positive and significant as a determinant of performance, but has a much smaller coefficient than the time spent in formal university study. In other words, a student who spends an extra hour at the university in formal study (*ceteris paribus*) will get better results than those who increase their self study time by one hour.

The most straightforward interpretation of this result is that in terms of producing exam performance, lectures and formal study are up to four times more efficient than self study.

However care should be exercised in the interpretation of this result and its generalizability to other educational institutions, and in particular, to other countries.

Specifically, caution should be expressed since the Spanish higher education system is very structured and the work required for any course is carefully prescribed during lectures and classes. Course structures are very regulated and little is expected of the students in terms of original research, creative writing or investigate study. Most courses have set textbooks and a prescribed curriculum. A lot of time is spent in lectures and classes, in instruction and practise for the examinations by working through of past examination papers. In such a system we may expect the return to formal study time to be higher than a more flexible system such as that which operates in the UK.

An interpretation of our main result is that each person has a finite capacity to take in subject matter. Hence after a period of intensive self study a person's capacity for learning new concepts by further time input may be strictly constrained. Hence the efficient allocation of effort may be to study for relatively short periods of time.

A second possible explanation of this result is that opportunity for self study hours is strictly constrained (once one has allowed for formal study time, leisure, travel and sleep). In this interpretation most students only have a limited range of hours to choose to study. In particular in many subjects of study the formal contact hours may be quite high.

On the other hand, it can be seen from Table 2 that time spent in private tuition has a negative effect on students performance. This clearly shows that students who need to attend private tuition are those who are less capable, i.e. with low ability or low motivation, or both. This negative result disappear (in all specifications) when the dummy for first year students is included. The result implies that it is mainly for new, inexperienced students that the private tuition effect is negative. In contrast there is not significant evidence of the influence of time spent learning or improving languages or computer knowledge on students results.

Time spent on travel and domestic tasks, and leisure both have a negative (and significant) influence on scores. The intuition of this result is clear if we think about the student's available time constraint and that time spent travelling or in domestic activities is not available

²¹ It seems logical as far less than 1% of the sample are overseas students.

for study. Therefore he will have to choose among university work (i.e. formal education and study) and other activities.

Time spent working for money has no statistically significant effect. This result is possibly because of the low proportion of students working in the labour market. A possible explanation for this is the low level of university fees in Spain and the level of state grants which obviate the need for students to supplement their income.

Seven dummy variables are used to measure the differential impact on output due to subject differences. The reference group is Social Sciences, so the coefficients measure the effect of each subject relative to this group. From this table one can see that the students who study Health, Arts, and those from Non Technical University Schools attain higher scores than the reference group. In contrast students of Pure Sciences and Engineering (Higher and Technical University Schools) underperform relative to those in Social Sciences. In general these subject variables are highly significant, therefore we can think about the possible existence of large differences across students from different subjects.

Since university selection and entry standards are very different by subject (e.g. Medicine and Engineering -Higher University Schools- are the subjects in most demand and hence entry requirements are highest) it is important to control for this heterogeneity in the determination of performance outcome. One way of doing this is to simply add dummy variables by subject as we have done in Table 2. Alternatively it could be argued that not only are the type of students entering each subject different in ability but also in terms of the marking and assessment schemes. This means that a score of 5.00 in Medicine may mean something totally different to the same score in Arts. To allow for this possibility (and test the robustness of our results on study time for across subject heterogeneity) we normalise our scores within subject. Hence we measure each person's performance relative to the mean score of their subject peers. In this way we aim to control for ability differences by subject of the students, and the possible heterogeneity of assessment methods by subject.

In the normalised²² results we would expect that the variance of the dependent variable would be substantially reduced and hence the R² of OLS on the scope for the regressors to

²² *Normalised Scores* = $\frac{Scores_i - Average\ Scores_{subject}}{Standard\ Deviation_{subject}}; i = 1, \dots, n$

explain this dependent variable would be considerable reduced. This is confirmed in Table 3. All the coefficients in this table are of the same sign as those tabulated in Table 2 but are smaller (in absolute terms) as a result of the origin and scale change.

Table 3: Stochastic Educational Production Function (output normalised)

Variables	Specification I				Specification II		Specification III	
	OLS		ML		ML		ML	
	Coefficient	T	Coefficient	t	Coefficient	t	Coefficient	t

Constant	-0.684*	-1.857	0.122	0.338	0.096	0.254	1.341***	4.699
Personal Characteristics	0.039***	3.809	0.041***	3.772	0.044***	3.891		
Age	-0.0004	-0.009	-0.009	-0.198	-0.017	-0.354	-0.026	-0.529
Gender	0.251	1.116	0.176	0.551	0.169	0.525	0.287	0.953
Married	-0.082	-0.320	-0.122	-0.563	-0.131	-0.611	-0.185	-0.842
Nationality	0.068	1.203	0.056	1.021	0.048	0.867	0.048	0.887
Geographic Area	0.042	0.857	0.045	0.919	0.043	0.875	-0.004	-0.080
Own transport								
Time Use^a	1.196***	3.765	1.275***	4.303	1.247***	4.195	0.919***	3.082
Formal Education	0.286***	2.498	0.329***	3.075	0.328***	3.074	0.333***	3.206
Self Study	-1.312**	-1.983	-1.319**	-2.051	-1.245**	-1.920	-1.015	-1.533
Private Tuition	-0.181	-0.588	-0.107	-0.340	-0.071	-0.226	-0.262	-0.877
IT/Language	-0.668***	-3.119	-0.697***	-3.391	-0.682***	-3.362	-0.675***	-3.436
Travel/Domestic	-0.098*	-1.644	-0.112**	-1.919	-0.115**	-1.971	-0.086	-1.472
Leisure	-0.216	-0.980	-0.204	-0.939	-0.184	-0.839	-0.256	-1.173
Paid Work								
Parents' Characteristics	0.241***	3.530	0.261***	3.893	0.252***	3.765	0.270***	4.211
Mother university stud.	-0.249	-	-0.045	-0.441	-0.044	-0.440	-0.042	-0.426
Parents divorced	-0.132	0.253	-0.145	-1.314	-0.120	-1.107	-0.112	-1.073
Orphan	-0.067***	-1.248	-0.065***	-3.370	-0.064***	-3.284	-0.066***	-3.436
Family size	-0.0001	-3.288	-0.0001	-0.177	-0.0002	-0.364	0.0001	0.241
Family income		-0.281						
Residence								
University Residence	-0.117		-0.128	-1.043	-0.145	-1.185	-0.058	-0.503
Rent flat	-0.078	-0.961	-0.074	-1.216	-0.070	-1.151	-0.082	-1.385
Motivation of the student		-1.261						
Satisfaction	0.058		0.069	1.410	0.082*	1.689	0.073	1.545
Ambition	-0.140***	1.196	-0.143***	-3.016	-0.141***	-2.988	-0.151***	-3.279
Other characteristics of the students		-2.960						
Grant	0.0005***		0.0005***	2.970	0.0006***	3.226	0.0005***	3.247
Bachillerato		2.638			-0.119*	-1.818		
L.O.G.S.E.					-0.262***	-3.044		
Via F. P.					0.548*	1.800		
Via R.E.M.							-0.536***	-11.66
First year student								
Parameters for compound error			1.403***	6.277	1.397***	6.325	1.283***	6.278
λ			1.276***	19.913	1.269***	20.040	1.207***	20.60
$S_{\epsilon} = (S_v^2 + S_u^2)^{1/2}$								6
Number of observations								
S_v^2				0.549		0.546		0.550
S_u^2				1.080		1.065		0.906
\bar{R}^2	1976		1976		1976		1976	
$F_{(27, 1949)}$	0.04							
$-2(\log_R - \log_U)$	4.51***		111.92***		128.83***		240.38***	

Note ^a: The time use variables are expressed in minutes per hour.

Note: * coefficient significantly different from zero at 10% confidence level; ** coefficient significantly different from zero at 5% level; *** coefficient significantly different from zero at 1% level.

Unsurprisingly, the family income variable is not significant, since family income may be expected to affect the demand of higher education but not necessarily the students performance at this educational level. We also include dummy variables related to the type of accommodation, but these have no influence on scores.

Two variables which merit attention are those relating to the motivations of students²³. According to the first such variable, “satisfaction”, those students who did not get into the course they wanted to at university do not perform significantly worse than those who did. In contrast, if the principal reason why they decided to study at University was to earn more money and/or to have a better chance of finding a job, then their academic performance falls.

The remaining variables considered in our estimation pick up some other characteristics related to the students’ educational background. The coefficients of these variables indicate: firstly, there is a clear positive correlation between the amount of state financial support received by the grant holders and their academic results²⁴. This suggests that funds devoted to grants may be used as an important tool from the educational policy point of view. Secondly, the students who continued to higher education from *Bachillerato L.O.G.S.E.* seem to perform worse as compared to those coming from B.U.P. (Secondary School in the old educational system). This means that the changes introduced by the reform of the educational system (at the Secondary School level) do not contribute to improved student performance. Nevertheless, this result must be treated with caution, because the reform of the educational system (L.O.G.S.E, 1990) had only just been introduced and only affected the first year students in our survey²⁵. As pointed out in Appendix A the vocational track is for the less academic students, thus the negative sign found in the estimations for the variable “Via F.P.” is not unexpected.

The last variable included in Tables 2 and 3 is a dummy variable, which enables us to distinguish between the differential impact on educational achievement of first year students as

²³ As can be seen from Table C1 (Appendix C), the results of our estimations do not change in a significant way when motivation variables are dropped.

²⁴ The level of the state support grant is “means-tested” on family income but is in addition payable only to those students with a specified minimum level of exam performance. In our data this variable is correlated with family income but uncorrelated with pre-university exam performance.

²⁵ This is the reason why the variables “Via *Bachillerato LOGSE*” and “First year students” are included in different specifications.

compared to final year students²⁶. The negative sign of this variable may be a consequence of the lower capacity of first year students to organise their studies and an inferior knowledge of the university framework compared to final year students. Alternatively, in the first year of higher education students may be deliberately taking more leisure time in order to make friends and appreciate the university experience since they are aware that the burden of getting a job and the importance of exam performance will become relatively much more important in their later years of university study.

The robustness of our results has been tested in several different ways. As reported in Table C2 (Appendix C), when a Cobb-Douglas functional form²⁷ is used (instead of a linear functional form) to examine the sign and significance level of the variables considered in the different specifications. Basically these remain the same as in Tables 2, 3. Hence, Table C2 provides additional evidence about the stability of the coefficients found.

Finally, it is interesting to examine the variance decomposition provided by estimates of the stochastic frontier, since it allows for both noise and inefficiency. The variance of the composite error (ϵ) is not $s_\epsilon^2 = s_v^2 + s_u^2$. As Greene (1993) points out $Var(u) = \left(\frac{p-2}{2}\right) s_u^2$, due to the asymmetry of the disturbance term. Hence the contribution of the variance of u to the total variance is,

$$\frac{Var(u)}{Var(\epsilon)} = \frac{\left(\frac{p-2}{2}\right) s_u^2}{\left(\frac{p-2}{2}\right) s_u^2 + s_v^2}$$

as a consequence, approximately 50 percent²⁸ of the variance of the composite error is caused by educational process inefficiency, while the remaining 50 percent represents unexplained variability. A possible interpretation of this result is that a portion of the unexplained variance in estimated educational production function may be due to time use inefficiencies by students.

²⁶ The variable “age” is not included in Specification III because of the high correlation with the variable “first year students”.

²⁷ Additional estimation was undertaken using a transcendental logarithmic functional form (translog). However our model includes too many independent variables to find a stable set of coefficients for the interaction effects of the variables.

²⁸ This figure does not vary much among the specifications reported (53 % in specifications I and II, and 49% in specification III).

Another possible instrument to measure the relative weight of the inefficiency in our estimations is the parameter λ (inefficiency component of the model)²⁹. This parameter is defined as:

$$I = \frac{s_u^2}{s_v^2}$$

Since s_u^2 represents about twice as much as s_v^2 (as is showed in Table 3), the value of λ reinforces the argument above. In all our estimations the λ parameter is highly significant which indicates that the use of the frontier production function is appropriate.

6. Identification, Endogeneity, Instrumental Variables, and Ability Bias

Until now we have adopted a very structural approach to an important possible bias in our estimations. This bias results from the unobservable nature of ability. The stochastic frontier approach takes a mechanistic approach to the problem of unobserved ability by effectively modelling the selection to university as a partitioning of the ability distribution as described earlier. Ideally we would wish to include the pre-university performance endogenously into our regression model as it may be of importance in the empirical problem of study time allocation. One possible solution to this problem is to take pre-university examinations scores as indicators of ability. However this has the problem that the unobservable factors which play an important role in the determination of pre-university exam results (like for example motivation, and determination) may also be determinants of university exam results. This would lead us to suggest that pre-university exam scores were endogenous to university exam scores. An instrumental variable procedure offers one solution to potentially correct for expected bias which may affect the input coefficients³⁰. An alternative solution is to use the residuals from the pre-university regression as a proxy for unobserved ability. We explore each of these approaches below after setting out the econometric models.

So far in our estimation we have used a simple production function framework to assess the relationship between the university examination score achieved and the time inputs put into this process by the student. In doing so we (in accordance with the literature) made a number of simplifying econometric modelling assumptions. Most specifically there are two key

²⁹ In the simple regression model with symmetrical disturbances $\lambda=0$.

³⁰ For a detailed explanation of this econometric procedure see, e.g., Green (2000).

assumptions which need to be tested in the context of our estimation: firstly, what role should the modelling of pre-university exam scores play in the process of modelling university exam performance, and secondly how might the results of our econometric modelling be affected by the treatment of unobserved ability. We will show in this section that these two questions are inter-linked in the sense that the variable relating to pre-university performance may be endogenous to university performance and the lack of a measure of ability directly affects our interpretation of the stochastic error terms in the model.

This section closely follows the survey article on econometric methodology in this area written by Todd and Wolpin (2000) and the econometric testing methodology detailed in the literature on Hausman-Wu tests and the those for IV estimation suggested by Bound et al (1995). Hence we cannot lay claim to any new methodological approaches but hope to be rigorous about a familiar problem.

Structural Form Econometric Model

We start by setting out the different models which can be estimated by adapting slightly the framework set out in Todd and Wolpin (2000). We specify four possible additional estimates of the production function for university achievement which have different econometric assumptions which limit their interpretation. The structural form of this model is specified in equations (6) and (7). Where y_{i0} and y_{i1} are respectively the pre-university and university performance scores, X_{i0} and X_{i1} are the family and other socio-economic characteristics (assumed to be non-stochastic) which influence exam performance at time period 0, before university and time period 1, at university respectively. The variable T_{i1} represents the time allocated to study which is focus of our research³¹ and m_{i0} and m_{i1} represent the ability of individual i at time 0 and time period 1 respectively. The stochastic error terms in the two equations are u_{i0} and u_{i1} .

$$y_{i0} = \mathbf{a}_0 + X_{i0}' \mathbf{b}_0 + \mathbf{d}_0 \mathbf{m}_{i0} + u_{i0} \quad (6)$$

$$y_{i1} = \mathbf{a}_1 + X_{i1}' \mathbf{b}_1 + \mathbf{g} y_{i0} + \mathbf{d}_1 \mathbf{m}_{i1} + \mathbf{p}_1 T_{i1} + u_{i1} \quad (7)$$

Simple Production Function Estimation

The simple production function estimator can be described by making some assumptions about the structural form model in equations (6) and (7):

$$E(u_{i1} / X_{i1}) = 0$$

$$E(u_{i1} / T_{i1}) = 0$$

$$d_0 m_{i0} = d_1 m_{i1} = 0$$

$$g = 0$$

Estimates:

$$y_{i1} = \mathbf{a}_1 + X_{i1}' \mathbf{b}_1 + p_1 T_{i1} + u_{i1} \quad (8)$$

The simplest model of production which we have been estimating assumes that pre-university performance plays no role in university exam performance and that ability either does not matter or cannot be measured. We also need to assume that both X_{i1} and T_{i1} are exogenous. In previous sections we have estimated a specific form of (8) which structurally allows the selection on ability by modelling the production function with a stochastic frontier. We now consider estimation methods which attempt to treat equations (6) and (7) jointly.

“Gain” or Fixed Effects Estimation

The fixed effects estimator which Todd and Wolpin (2000) call the ‘gains estimator’, has been popular in the labour economics literature as it would appear to solve the problem of unobservables. However this ‘fix’ to the problem of unobservables is not really a ‘solution’ since it rests on the often unjustifiable assumption that the unobservable effect (like unobservable ability) is fixed across time periods. Typically ability may develop as the student progresses. (See Todd and Wolpin (2000) for a detailed discussion of this point.) The value added model assumes:

³¹ We do not need to split it into formal and self study in this notation as this split poses no extra conceptual issues.

$$E(u_{i1} / X_{i1}) = 0$$

$$E(u_{i1} / T_{i1}) = 0$$

$$d_{i0} \mathbf{m}_0 = d_{i1} \mathbf{m}_1 = 0$$

$$\mathbf{g} = 1$$

Estimates using (7) – (6):

$$(y_{i1} - y_{i0}) = (\mathbf{a}_1 - \mathbf{a}_0) + X_{i1}' \mathbf{b}_1 - X_{i0}' \mathbf{b}_0 + \mathbf{p}_1 T_{i1} + \mathbf{e}_i \quad (9)$$

where $\mathbf{e}_i = (u_{i1} - u_{i0})$.

We do not report the estimation of this model as it is a restricted version of the “value added estimator” which we now consider.

Value Added Estimator

A generalised version of the fixed effects estimator is described by Wolpin and Todd (2000). They call this the value added estimator which has the following structure:

$$E(u_{i1} / X_{i1}) = 0$$

$$E(u_{i1} / y_{i1}) = 0$$

$$E(u_{i1} / T_{i1}) = 0$$

$$d_{i0} \mathbf{m}_0 = d_{i1} \mathbf{m}_1 = 0$$

$$\mathbf{g} \neq 0, \mathbf{g} \neq 1$$

The model suggests the estimation of:

$$y_{i1} = \mathbf{a}_1 + X_{i1}' \mathbf{b}_1 + \mathbf{g} y_{i0} + \mathbf{p}_1 T_{i1} + u_{i1} \quad (10)$$

i.e. it is an attempt to estimate equation (7) of the structural form on the assumption that unobserved ability is unimportant or that the y_{i0} variable is an adequate proxy for unobserved ability.

An alternative way of considering this estimation is to estimate equation (6) (without unobserved ability) and then compute [(7)– \mathbf{g} (6)]. i.e.:

$$(y_{i1} - \mathbf{g}y_{i0}) = (\mathbf{a}_1 - \mathbf{g}\mathbf{a}_0) + X'_{i1}\mathbf{b}_1 - X'_{i0}\mathbf{b}_0\mathbf{g} + \mathbf{p}'_1T_{i1} + \mathbf{e}_{i1} \quad (11)$$

where $\mathbf{e}_{i1} = (u_{i1} - \mathbf{g}u_{i0})$.

The estimation results of this model are presented in the first column of Table 4. The results are dramatic in their comparison to the stochastic frontier estimation in Table 3. The most important finding is that self study time is now insignificant and the included y_{i0} variable gives a \mathbf{g} estimate of .362. This suggests that when ability is included as a proxy by y_{i0} the impact of more self study time is negligible. The continued importance of formal study time implies that the only (decision variable) input which matters in terms of student performance is the impact of time in lectures and classes. This would suggest that ability directly constrains student performance and that this is largely unaffected by extra time in self study.

Instrumental Variables Estimation

Another estimation solution which is often adopted in the case of measurement error or endogenous variables is the technique of instrumental variables. This technique requires us to find correlates (X_{i0}), of y_{i0} which are not correlated with u_{i0} or y_{i1} . Formally we can write the model as:

$$E(u_{i1} / X_{i1}) = 0$$

$$E(u_{i1} / T_{i1}) = 0$$

$$\mathbf{d}_{0\ i0} \mathbf{m} = \mathbf{d}_{1\ i1} \mathbf{m} = 0$$

X_{i0} significantly correlated with y_i

X_{i0} are valid exclusion restrictions from (7)

u_{i0} and u_{i1} uncorrelated

The procedure consists of estimating (6):

$$y_{i0} = \mathbf{a}_0 + X'_{i0}\mathbf{b}_0 + u_{i0}$$

compute predicted values of y_{i0} :

$$\hat{y}_{i0} = \hat{\mathbf{a}}_0 + X'_{i0}\hat{\mathbf{b}}_0$$

using these predicted values in equation (7) gives:

$$y_{i1} = \mathbf{a}_1 + X_{i1}' \mathbf{b}_1 + \hat{g}_{i0} + \mathbf{d}_0 \mathbf{m}_1 + \mathbf{p}_1 T_{i1} + u_{i1}$$

The results obtained from the first step of the instrumental variable procedure can be seen in Appendix C (Table C3). We report different specifications in order to control the problems stemming from the correlation among different personal and parents' characteristics (specification IV seems to be the most satisfactory). The major conclusion from this table are that those students who attend private Secondary Schools show higher pre-university academic achievements, and those attending *Bachillerato L.O.G.S.E.* get worse pre-university exam results than reported by students who attended the old *Bachillerato (B.U.P.)*.

The estimated values for the instrumental variable (pre-university exam results) are incorporated into the second stage of the procedure (i.e. in the base specifications showed in Tables 2 and 3). The coefficient estimates from the second stage of the instrumental variable procedure are reported in Table 4.

The major conclusion from this table is that the results for all the variables are virtually identical to those shown in Tables 2 and 3. Therefore there is little evidence of possible biases in previous estimations.

Our first step estimation results used in computing the IV of y_{i0} in Table 4 are included in Specification V in Table C3 in the Appendix C. We performed the Bound et al. (1995) tests which suggested that the instruments we used are valid with a F statistic of 14.15 which is significant at the 1% level. The partial r squared is 0.0259. None of the regressors used in this specification are significant in explaining university exam results. Superficially these results suggest that the IV approach may be a suitable technique for handling the problem of the endogeneity of y_{i0} in the y_{i1} equation. We can also see this as y_{i0} being an imperfect proxy of \mathbf{m}_0 with measurement error. In either case the technique of IV estimation would be justified³². The results of this estimation are presented in the second column section of Table 4. They show us that the IV variable is not significant in the determination of y_{i1} . This conclusion is supported in the Hausman-Wu mis-specification test which shows that H_0 , the hypothesis that there is no systematic difference between the regression coefficients in the original model

³² Note that the estimation of (12) is by OLS since the IV procedure is only strictly valid for this estimation technique.

and the IV estimated model, is accepted with a $c^2=2.60$. The reason for this result becomes clear if we write out the IV model and perform some simple algebraic manipulation.

$$\begin{aligned} y_{i1} &= \mathbf{a}_1 + X_{i1}' \mathbf{b}_1 + \mathbf{g}(\hat{\mathbf{a}}_0 + X_{i0}' \hat{\mathbf{b}}_0) + \mathbf{p}_1 T_{i1} + u_{i1} \\ y_{i1} &= (\mathbf{a}_1 + \mathbf{g}\mathbf{a}_0) + X_{i1}' \mathbf{b}_1 + X_{i0}' \mathbf{b}_0 \mathbf{g} + \mathbf{p}_1 T_{i1} + u_{i1} \end{aligned} \quad (12)$$

It is clear for the form of (6) that the IV estimation will not be significantly different from the Simple Production Function estimator of equation (8) because although there is a term in \hat{y}_{i0} in the estimation the form of (12) makes it clear that in reality this is only changing the composition of the X regressors relating to family and personal background in a marginal manner.

Residual Unobserved Ability Estimation

The final estimation procedure we examine explicitly attempts to control for unobserved ability. If we assume that a major component of the residuals on the y_{i0} equation consist of the omitted variable associated with unmeasured ability then we can at least partially control for this in the y_{i1} equation by using the residuals as an extra regressor. Setting out this model more formally:

In this model we assume:

$$E(u_{i1} / X_{i1}) = 0$$

$$E(u_{i1} / T_{i1}) = 0$$

$$\mathbf{m}_{i0} = \mathbf{m}_{i1}$$

Estimating the residuals from equation (6):

$$\hat{r}_{i0} = (\mathbf{d}_0 \mathbf{m}_{i0} + u_{i0}) = (y_{i0} - \hat{\mathbf{a}}_0 + X_{i0}' \mathbf{b}_0) \quad (13)$$

Substituting \hat{r}_{i0} into (7) as a proxy for y_{i0} :

$$y_{i1} = \mathbf{a}_1 + X_{i0}' \mathbf{b}_0 + \mathbf{p}_1 T_{i1} + \mathbf{q} \hat{r}_{i0} + \mathbf{d}_1 \mathbf{m}_{i1} + u_{i1}$$

using (13) we get:

$$y_{i1} = \mathbf{a}_1 + X_{i0}' \mathbf{b}_0 + \mathbf{p}_1 T_{i1} + \mathbf{q} \mathbf{d}_0 \mathbf{m}_{i0} + \mathbf{q} u_{i0} + \mathbf{d}_1 \mathbf{m}_{i1} + u_{i1}$$

rearranging:

$$y_{i1} = \mathbf{a}_1 + X_{i0}' \mathbf{b}_0 + \mathbf{p}_1 T_{i1} + (\mathbf{q} \mathbf{d}_0 + \mathbf{d}_1) \mathbf{m}_{i0} + \mathbf{e}_i \quad (14)$$

where $\mathbf{e}_i = (\mathbf{q} u_{i1} + u_{i0})$.

This shows us that the coefficient on the residuals term can be interpreted as an estimate of the importance of unobserved ability.

The results of estimating this model are reported in the final columns of Table 4. The results strongly support the value added model estimates as they are very similar. They suggest again that self study time is unimportant when the proxy for ability is included. As such, these results are further support for the finding that an individual's ability will constrain what is feasible in terms of examination performance and there is limited scope for influencing this university outcome by further self study time.

Table 4: Stochastic Educational Production Function-IV (Pre-University results)

	Specification II (Output Normalised)					
	Value Added Estimation ^b		Instrumental Variable Estimation ^c		Residuals Unobserved Ability Estimation ^c	
	Coefficient	t	Coefficient	t	Coefficient	t
Constant	-2.786	-7.003	-0.491	-0.287	-0.460	-1.327
Personal Characteristics						
Age	0.060***	5.741	0.040***	2.530	0.038***	3.794
Gender	0.042	0.896	-0.012	-0.204	-0.016	-0.339
Married	0.098	0.335	0.235	1.053	0.155	0.745
Nationality	-0.072	-0.310	-0.080	-0.315	-0.088	-0.369
Geographic Area	0.058	1.117	0.061	1.078	0.063	1.189
Own transport	0.028	0.604	0.034	0.692	0.035	0.761
Time Use^a						
Formal Education	0.823***	2.932	1.196***	3.774	0.760***	2.558
Self Study	0.089	0.867	0.282***	2.468	0.062	0.575
Private Tuition	-1.267**	-2.111	-1.243**	-1.883	-1.242**	-2.019
IT/Language	-0.395	-1.407	-0.139	-0.451	-0.439	-1.526
Travel/Domestic	-0.469**	-2.405	-0.651***	-3.047	-0.476***	-2.384
Leisure	-0.067	-1.182	-0.099*	-1.670	-0.054	-0.976
Paid Work	-0.199	-0.981	-0.178		-0.222	-1.081
Parents' Characteristics						
Mother University studies	0.103*	1.756	0.233**	2.189	0.237***	3.914
Parents divorced	-0.023	-0.237	-0.026	-0.262	-0.012	-0.137
Orphan	-0.014	-1.469	-0.105	-0.998	-0.149	-1.515
Family size	-0.052***	-2.906	-0.065***	-3.228	-0.062***	-3.380
Residence						
University Residence	-0.204*	-1.784	-0.131	-1.076	-0.190*	-1.675
Rent flat	-0.103*	-1.781	-0.076	-1.224	-0.104*	-1.799
Motivation of the students						
Satisfaction	-0.047	-1.017	0.068	1.397	-0.049	-1.070
Ambition	-0.071*	-1.645	-0.143***	-3.027	-0.073*	-1.664
Other characteristics related to the students' background						
Grant	0.0005***	2.814	0.0005***	3.013	0.0004***	2.606
Bachillerato L.O.G.S.E.	0.096	1.514	-0.143	-1.041	-0.114*	-1.817
Via F. P.	-0.468***	5.926	-0.241	-1.514	-0.232***	-2.901
Via R.E.M.	0.787***	3.355	0.575**	1.938**	0.616***	2.340
Instrumental Variable						
Pre-University exam results	0.362***	17.622	-0.029	-0.140		
Residuals from y_{io}					0.372***	16.932
Parameters for compound error						
λ	1.082***	5.283				
$S_{\epsilon} = (S_v^2 + S_u^2)^{1/2}$	1.111***	18.920				
S_v^2						
S_u^2	0.569					
Number of observations	0.666					
\bar{R}^2	1976		1976		1976	
$F_{(27, 1949)}$			0.05		0.17	
-2 (log_R - log_U)			4.67***		16.39***	
	195.99***		119.47***		390.63***	

Note ^a: The time use variables are expressed in minutes per hour.

Note ^b: ML estimations.

Note ^c: OLS estimations.

Note: * coefficient significantly different from zero at 10% confidence level; ** coefficient significantly different from zero at 5% level; *** coefficient significantly different from zero at 1% level.

7. Conclusions and Policy Implications

Careful consideration of our econometric results would suggest important policy implications for the university authorities and educational planners. In addition the results may be suggestive for the individual student in their choice of study time and potentially for parents seeking to support their sons and daughters in higher education.

Most obviously for universities the significance of formal study time on performance suggests that they should do all that is in their power to encourage student attendance at lectures and classes or even to make them compulsory. More difficult is the recognition that subject differences are important. This may mean that university authorities may need to review how many formal contact hours are necessary in each subject. Indeed, if universities operate in a quasi-competitive environment where student performance by university is compared and subsequent employment outcomes are used as performance indicators of universities, they may need to devote more resources to teaching their students. Further implications of our results, which are more difficult to predict (given our unobservables), are whether more effective teaching units could be delivered by operating smaller class sizes or even devoting more formal contact hours to students with lower ability (or lower pre-university test scores)³³. In addition university authorities should review the amount of time taken to study for a degree as our results suggest that the academic year could be lengthened and the duration of a degree course shortened if more hours of lectures and classes were presented. Indeed this issue has been on the policy agenda in Spain with the possible shortcoming of degree studies from 5 to 4 years.

A further implication of our results for government involvement in education is suggested by the importance of financial support for students, since it would appear that a higher level of this support is most conducive to more favourable exam performance. The issue here is what is the optimal level of student support and whether one targets this support at the students from the least well-off families. Our results suggest that “means-tested” support does have a very important impact on students from low income families.

³³ In other countries, for example the United Kingdom, students with deficiencies in subjects like mathematics may have to attend to ancillary classes.

Our results also have some clear conclusions for the individual student. Most clearly the student who wishes to maximise their examination score should attend all lectures and classes and minimise their absence from any formal tuition provided by the university. A logical corollary to this result is that the student should not overindulge in leisure time. In addition the student would be very wise to consider the replacement of time spent attending private tutorials with time attending formal education, as our results suggest that this could have positive effects. In addition our results suggest that there is a clear payoff to minimizing the amount of time spent on Travel and Domestic activities. These results could also have implications for parents who wish to support their student sons and daughters. Most specifically family income does not seem to matter in the provision of advantage. What does offer an advantage is having provided an educationally privileged background with fewer brothers and sisters (to have to share parental attention). When a child is at university a parent who seeks to confer on advantage may consider providing help to minimize time spent commuting and in domestic chores.

A final more difficult area to be clear about is the extent to which any student's unobserved natural ability constrains their potential performance. Our results provide some support for the view that this ability is not symmetrically distributed among university students and that possibly each student may be constrained by what is possible for someone with their ability. Indeed, controlling for unobserved ability we obtain the result that input of self study time may not matter at all. Hence the plausible message maybe to individual students (and their over-ambitious parents) that each one of them has an "ability" limit on what is efficiently achievable in terms of an examination score and that it may be very hard to improve on this by additional input of private study time. An additional warning note to those students with high expectations about earnings on job prospects is that such ambitions may be a misdirection of effort which could detract from the achievement of better final examination marks.

Appendix A:

The Spanish education system

Spanish education has changed considerably over the last decade particularly since the Organic Act on General Management of the Education System (L.O.G.S.E., 1990). There are a variety of different qualifications that students can take and the educational system is divided into two different stages. First, Compulsory Education, which comprises Primary School (*Educación Primaria*) and the first level of Secondary School (*Educación Secundaria Obligatoria*). Second, Non Compulsory Education, consisting of the second level of Secondary School (*Formación Profesional or Bachillerato*), and Higher Education.

Pupils attend Primary School from 6 to 12 years old. Students attend first level of Secondary School from 13 to 16 (which is the statutory leaving age). At age 16, pupils who satisfactorily achieve the stipulated academic target are awarded the *Graduado de Educación Secundaria Obligatoria*. After age 16 students may choose to leave the education system completely (around 15.2 % in the academic year 1999-2000) or stay on at school. Those who stay on at school follow one of the two distinct tracks: the vocational (*Ciclos Formativos de Formación Profesional*) track or the academic track (*Bachillerato LOGSE*)³⁴.

The vocational track is for the less academic students who can choose from a variety of vocational qualifications based upon practical subjects such as computing, hairdressing, office skills, etc. Students who succeed in the first two years of vocational education obtain a Certificate called *Ciclo Formativo de Formación Profesional (medium level)*. For those continuing beyond the *medium level* there is a wide range of higher vocational qualifications *Ciclo Formativo de Formación Profesional (higher level)*, with more than sixty specialities.

The academic track is for the more able students who study at school for a further two years (*Bachillerato*). After completing this stage, they have the option to continue to higher education. Students can opt for either a 3 years (first-cycle) degree, which can be technical (*Escuelas Universitarias Técnicas*) or non-technical (*Escuelas Universitarias no Técnicas*), or a 4-5-6 years (first and second cycle) degree (*Facultades and Escuelas Superiores*). For both sorts of education, entrance is competitive, as places are limited. This

necessitates the use of a rationing device. In the case of Faculties (*Facultades*) and Higher Technical Schools (*Escuelas Técnicas Superiores*), the students will take an university entrance exam (*Selectividad*) on several subjects; the weighted average of the marks obtained in the entrance exam and during the Baccalaureate (*Bachillerato*) years (or *higher level of Ciclo Formativo de Formación Profesional*) will be used as the rationing device. However, for those who prefer a shorter degree the entrance exam is not necessary, because they are filtered simply by means of the average marks obtained during the *Bachillerato* years (or *higher level of Ciclo Formativo de Formación Profesional*).

Students who have taken higher vocational qualifications are more likely to attend a technical short degree. The choice between a short degree and a lengthier one will also be largely based on the candidate's academic ability.

Before this system was introduced (in 1992), the schooling was only compulsory up until the age of fourteen. The second major difference between the old (General Education Act of 1970) and the new system is that the students coming from the previous one had to take the decision about dropout or, in the opposite case, following one of the two distinct academic (*Bachillerato Unificado Polivalente*) or vocational (*Formación Profesional*) tracks from the age of about fourteen.

The Spanish Higher Education System

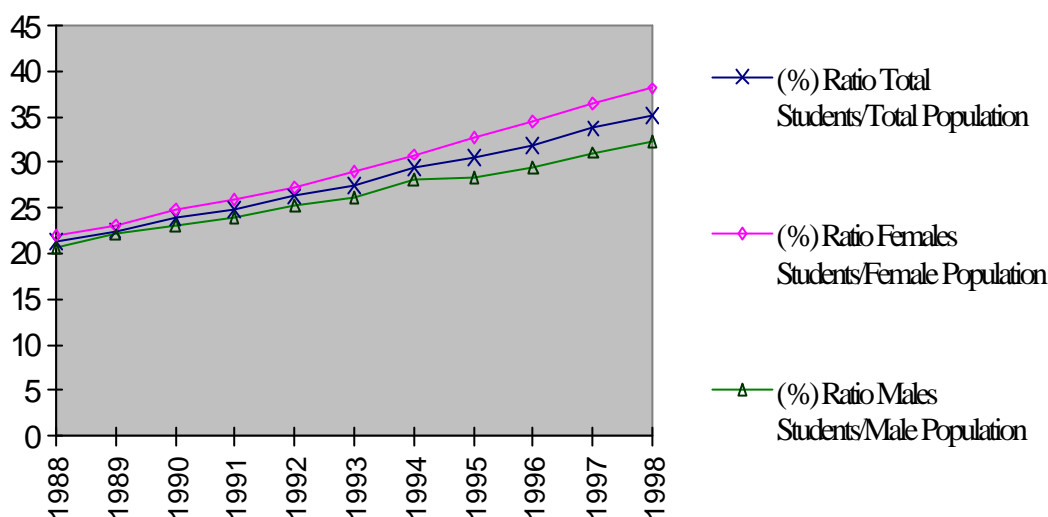
In the last three decades, Spain has undergone a “democratisation” of secondary education, which has resulted in a growth in the demand for higher education. This is exemplified by the fact that in the last ten years (1988 to 1998) the number of students entering higher education³⁵ grew by 52.3 % whilst the total size of the population of university entry age fell by 11 %. In the same period, the total number of registered students rose by 62 % (this trend is more marked in the case of women where the figure

³⁴ There is a third possible track called R.E.M., which is consequence of an experimental plan implemented by the government in a few centres of Secondary Education. But only around 0.5% of the students follow this track.

³⁵ In Spain the concepts higher education and university education can be considered synonymous due to the lack of a significant non-university sector of higher education. This situation is due to change gradually in the coming years, provided the most recent reforms (L.O.G.S.E, 1990) of the Spanish education system continue to be consolidated.

is 70.4 %) which implies, in absolute terms, a total of almost 1.6 million university students, 58.9 % of whom are women. Figure A1 shows the relationship between students registered in

Figure A1: Participation Ratio: The Ratio Students Registered/Population of University age (19-24); (1988-1998)



Source: Statistics of the Higher Education in Spain (INE).

higher education and the total population at university entry age.

This trend has been accentuated by high levels of unemployment (the highest in the European Union)³⁶, which have reduced the relative real cost of studying at university, and by the growing role of the public sector in financing education, which has facilitated, from an economic perspective, the participation of students from socially and economically less privileged groups in the higher education system. An additional important factor which has a bearing on our data are the fundamental changes taking place in the role of women in Spanish society, which have led to their increasing participation in higher education. The main motivation for this increased participation by women is the clear incentives to overcome discrimination and enter the job market³⁷.

In this context of expansion of demand it is especially significant to develop techniques of measurement for the internal efficiency of the higher education system.

³⁶ The unemployment rate in Spain in 1998 was 20.8 %, in contrast to the average percentage of 10.8% in the European Union.

³⁷ See Garcia (1999) for a detailed analysis of this problem in Spain.

Appendix B

The variables used in our models are defined below:

Personal Characteristics Variables

University scores: Average scores obtained during the first semester of the 1998-1999 academic year (scaled from 0 to 10 points).

Students distribution by scores

Scores	Population ^a (%)	Sample (%)
0-5	29.91	35.02
5-6.5	39.37	33.25
6.5-8.5	21.45	26.82
8.5-9.5	7.88	4.35
9.5-10	1.39	0.6
N	39130	1976

Note a: Students registered at the University of Malaga.

Pre-university entrance results: In the case of Faculty or Higher Technical Schools students, weighted average of the marks obtained in the university entrance exam (50%) and during the *Bachillerato* or *Higher level of Formación Profesional* years (50%); in the case of Technical and Non Technical University Schools students, average of the marks obtained during the *Bachillerato* or *Higher level of Formación Profesional*. This scores are scaled from 0 to 10 points.

Age: Age of the student measured in years.

Gender: Equals 1 if respondent's gender is male, 0 if female.

Married: Equals 1 if married, 0 otherwise.

Nationality: Equals 1 if respondent's nationality is Spanish and 0 otherwise.

Geographic Area: Equals 1 if respondent lives in Malaga Capital during the academic year and 0 otherwise.

Own transport: Equals 1 if respondent uses own transport and 0 otherwise.

Oldest: Equals 1 if respondent is the eldest sibling and 0 otherwise.

Youngest: Equals 1 if respondent is the youngest sibling and 0 otherwise.

Time Use Variables

The questions on time use asked in the questionnaire are:

“What is the average allocation of your time on ‘a normal weekday’ in the following activities”

“What is the average allocation of your time on ‘a normal weekend’ in the following activities”

Formal Education: Average time spent daily attending university classes³⁸.

Self Study: Average time spent daily and on weekends studying.

Private Tuition: Average time spent, daily and on weekends, attending private tuition.

IT/Language: Average time spent, daily and on weekends, learning or improving languages or computer knowledge.

Travel/Domestic: Average time spent, daily and on weekends, on housework and journeys.

Leisure: Average time spent, daily and on weekends, on leisure activities.

Paid Work: Average time spent, daily and on weekends, on paid work.

Subjects Effects Variables

Arts: Equals 1 if respondent is registered at Arts (University Faculties) and 0 otherwise.

Health: Equals 1 if respondent is registered at Health (University Faculties) and 0 otherwise.

Engineering: Equals 1 if respondent is registered at Engineering (Higher Technical Schools) and 0 otherwise.

Pure Sciences: Equals 1 if respondent is registered at Pure Sciences (University Faculties) and 0 otherwise.

Social Sciences: Equals 1 if respondent is registered at Social Sciences (University Faculties) and 0 otherwise (base case).

Non Technical University Schools: Equals 1 if respondent is registered at a Non Technical University School and 0 otherwise.

³⁸ Note that around 3% of students claimed they spent no time in private self study. When questioned further these students confirmed that they did not need to spend any time in self study as they could prepare for their examinations by intensive study just prior to the date of the exam.

Technical University Schools: Equals 1 if respondent is registered at a Technical University School and 0 otherwise.

Parents' Characteristics Variables

Father University Studies: Equals 1 if respondent's father has an university degree and 0 otherwise.

Father non University Studies: Equals 1 if respondent's father has less than university degree and 0 otherwise (*base case*).

Mother University Studies: Equals 1 if respondent's mother has an university degree and 0 otherwise.

Mother non University Studies: Equals 1 if respondent's mother has less than university degree and 0 otherwise (*base case*).

Parents divorced: Equals 1 if respondent's parents are divorced and 0 otherwise.

Orphan: Equals 1 if respondent's mother and/or father is dead and 0 otherwise.

Family size: Number of family members.

Family income: Family income *per capita* (i.e. Household income divided by the number of family members) measured in thousands of pesetas.

Residence Variables

University Residence: Equals 1 if respondent lives in an university residence and 0 otherwise.

Rent Flat: Equals 1 if respondent lives in a rented flat and 0 otherwise.

Parents' House: Equals 1 if respondent lives in parents' house and 0 otherwise (*base case*).

Other characteristics related to the students' background

Grant: Sum of money measured in thousands of pesetas obtained by the individual student from the state.

Bachillerato L.O.G.S.E.: Equals 1 if respondent went on to higher education after the *Bachillerato L.O.G.S.E.* and 0 otherwise.

Via Vocational Training (F.P.): Equals 1 if respondent went on to higher education after the *Vocational Training* and 0 otherwise.

Via *Reforma de las Enseñanzas Medias (R.E.M.)*: Equals 1 if respondent went on to higher education after the *R.E.M.* and 0 otherwise.

Via *Bachillerato Unificado Polivalente (B.U.P.)*: Equals 1 if respondent went on to higher education after the *Bachillerato B.U.P.* and 0 otherwise (base case).

First year student: Equals 1 if respondent is first year student and 0 otherwise.

Private: Equals 1 if respondent attended an private school during the Secondary School and 0 otherwise.

Motivation Variables

Satisfaction: Equals 1 if respondent is studying their chosen subject and 0 otherwise.

Ambition: Equals 1 if the principal reason why the respondent decided to study at University was to earn more money and/or to have more chance of finding a job, and 0 otherwise.

The qualifications includes in each subject are detailed below:

Arts: English Language and Linguistics, Spanish Language and Literature, Philosophy, History, History of Art, Geography, Translation, Pedagogy, Audio-visual Communication, Public Relations.

Health: Medicine, Speech Therapy.

Engineering: Computer Science, Telecommunications, Industrial, Civil.

Pure Sciences: Biology, Mathematics, Chemistry.

Social Sciences: Law, Management and Business Administration, Economics, Psychology.

Non Technical University Schools: Education, Nursing, Physiotherapy, Tourism Sciences, Industrial Relations, Public Administration, Business Administration (Diploma).

Technical University Schools: Technical Engineering (Industrial, Civil, Computer Sciences, Telecommunications).

Table B1: Descriptive Statistics by gender

Variables	Total		Female		Male	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Average Marks (Pre-University)	6.71	1.02	6.77	1.01	6.64	1.02
Average Marks (University)	5.60	1.79	5.77	1.83	5.38	1.71
Personal Characteristics						
Age	20.63	2.35	20.32	2.57	20.99	2.31
Gender	0.46	-	-	-	-	-
Married	0.01	-	0.01	-	0.007	-
Nationality	0.99	-	0.99	-	0.99	-
Geographic Area	0.79	-	0.80	-	0.78	-
Own transport	0.40	-	0.29	-	0.52	-
Time Use^a						
Formal Education	5.68	1.68	5.66	1.63	5.69	1.74
Self Study	7.80	4.80	8.63	4.95	6.83	4.46
Private Tuition	0.23	0.81	0.23	0.78	0.23	0.84
IT/Language	0.73	1.75	0.59	1.46	0.90	2.02
Travel/Domestic	3.75	2.99	4.64	3.03	2.70	0.26
Leisure	18.83	9.19	17.87	8.86	19.98	0.94
Paid Work	0.80	2.50	0.72	2.43	0.90	0.26
Subjects						
Arts	0.28	-	0.36	-	0.20	-
Health	0.04	-	0.05	-	0.03	-
Engineering	0.07	-	0.02	-	0.12	-
Pure Sciences	0.08	-	0.08	-	0.08	-
Non Technical University Schools	0.26	-	0.34	-	0.17	-
Technical University Schools	0.15	-	0.04	-	0.28	-
Parents' Characteristics						
Father university studies	0.21	-	0.19	-	0.22	-
Mother university studies	0.14	-	0.13	-	0.16	-
Parents divorced	0.06	-	0.06	-	0.05	-
Orphan	0.04	-	0.05	-	0.04	-
Family size	4.55	1.18	4.53	1.19	4.58	1.15
Family income	64.86	47.78	62.93	46.64	67.15	49.02
Residence						
University Residence	0.04	-	0.05	-	0.03	0.16
Rent flat	0.24	-	0.24	-	0.24	0.42
Motivation of the students						
Satisfaction	0.69	-	0.68	-	0.71	-
Ambition	0.34	-	0.27	-	0.43	-
Other characteristics related to the students' background						
Grant	73.03	136.01	76.13	139.75	69.35	131.43
Bachillerato L.O.G.S.E.	0.14	-	0.17	-	0.10	-
Via F. P.	0.08	-	0.07	-	0.08	-
Via R.E.M.	0.006	-	0.006	-	0.007	-
First year student	0.57	-	0.63	-	0.51	-
Private Centre	0.22	-	0.22	-	0.22	-

Note ^a: The time use variables are expressed in hours.

Table B2: Descriptive Statistics by subject

Variables	Social Sciences		Arts		Health	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Average Marks (Pre-University)	6.64	0.94	6.64	1.01	7.38	1.10
Average Marks (University)	5.69	1.46	5.95	1.71	6.31	1.36
Personal Characteristics						
Age	20.40	2.04	20.33	2.29	19.81	3.13
Gender	0.47	-	0.32	-	0.29	-
Married	0	-	0.01	-	0.01	-
Nationality	0.99	-	0.99	-	0.99	-
Geographic Area	0.77	-	0.80	-	0.87	-
Own transport	0.50	-	0.31	-	0.31	-
Time Use^a						
Formal Education	5.23	1.22	5.51	1.45	6.66	1.50
Self Study	8.59	4.70	7.48	4.73	11.73	6.63
Private Tuition	0.23	0.72	0.19	0.66	0.14	0.47
IT/Language	0.65	1.44	0.67	1.77	0.44	1.40
Travel/Domestic	3.50	3.17	4.06	3.18	4.11	3.00
Leisure	19.70	8.69	19.83	9.54	17.26	10.76
Paid Work	0.43	1.52	0.85	2.55	0.65	2.36
Parents' Characteristics						
Father university studies	0.27	-	0.18	-	0.29	-
Mother university studies	0.20	-	0.13	-	0.24	-
Parents divorced	0.05	-	0.08	-	0.05	-
Orphan	0.03	-	0.05	-	0.02	-
Family size	4.54	1.11	4.44	1.15	4.50	1.20
Family income	74.50	54.06	62.23	43.37	80.49	63.86
Residence						
University Residence	0.02	-	0.04	-	0.15	-
Rent flat	0.22	-	0.22	-	0.21	-
Motivation of the students						
Satisfaction	0.73	-	0.62	-	0.69	-
Ambition	0.37	-	0.28	-	0.19	-
Other characteristics related to the students' background						
Grant	55.25	122.26	80.32	80.32	95.35	150.77
Bachillerato L.O.G.S.E.	0.13	-	0.18	-	0.18	-
Via F. P.	0	-	0.03	-	0.07	-
Via R.E.M.	0.004	-	0.007	-	0	-
First year student	0.57	-	0.63	-	0.75	-
Private Centre	0.31	-	0.17	-	0.37	-

Note ^a: The time use variables are expressed in hours.

Table B2 (continued)

Variables	Engineering		Pure Sciences		Non Technical University Schools		Technical University Schools	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Average Marks (Pre-Univ.)	7.52	0.99	6.88	1.01	6.58	0.94	6.52	0.96
Average Marks (University)	4.51	1.55	4.86	1.79	6.00	1.92	4.79	1.53
Personal Characteristics								
Age	20.24	1.94	19.97	1.61	20.73	2.45	21.89	2.42
Gender	0.84	-	0.47	-	0.29	-	0.85	-
Married	0.007	-	0	-	0.01	-	0.02	-
Nationality	0.99	-	0.98	-	0.99	-	0.99	-
Geographic Area	0.84	-	0.76	-	0.79	-	0.76	-
Own transport	0.46	-	0.41	-	0.38	-	0.49	-
Time Use^a								
Formal Education	6.47	1.52	5.95	1.77	5.49	1.81	5.90	1.93
Self Study	8.72	4.16	9.12	5.17	6.98	4.54	7.11	4.20
Private Tuition	0.51	1.25	0.21	0.65	0.23	0.92	0.22	0.79
IT/Language	0.63	1.35	0.28	0.89	0.58	1.39	1.51	2.60
Travel/Domestic	3.42	3.06	3.14	2.56	3.88	2.74	3.47	2.96
Leisure	16.96	7.82	17.91	9.64	18.41	8.89	18.68	9.01
Paid Work	0.31	1.38	0.68	2.20	0.82	2.70	1.26	3.07
Parents' Characteristics								
Father university studies	0.39	-	0.26	-	0.14	-	0.18	-
Mother university studies	0.28	-	0.19	-	0.09	-	0.11	-
Parents divorced	0.06	-	0.07	-	0.04	-	0.05	-
Orphan	0.07	-	0.03	-	0.06	-	0.05	-
Family size	4.67	1.21	4.60	1.30	4.60	1.19	4.64	1.15
Family income	78.35	47.17	69.85	52.01	58.10	39.39	61.48	53.44
Residence								
University Residence	0.09	-	0.03	-	0.02	-	0.03	-
Rent flat	0.32	-	0.24	-	0.22	-	0.29	-
Motivation of the students								
Satisfaction	0.91	-	0.70	-	0.67	-	0.75	-
Ambition	0.36	-	0.37	-	0.35	-	0.42	-
Other characteristics related to the students' background								
Grant								
Bachillerato L.O.G.S.E.	43.55	107.46	63.17	134.79	85.20	139.77	63.82	131.53
Via F. P.	0.06	-	0.12	-	0.15	-	0.07	-
Via R.E.M.	0	-	0	-	0.14	-	0.19	-
First year student	0.007	-	0.13	-	0.004	-	0.007	-
Private Centre	0.50	-	0.60	-	0.58	-	0.37	-
	0.32	-	0.19	-	0.21	-	0.19	-

Note ^a: The time use variables are expressed in hours.

Figure B1

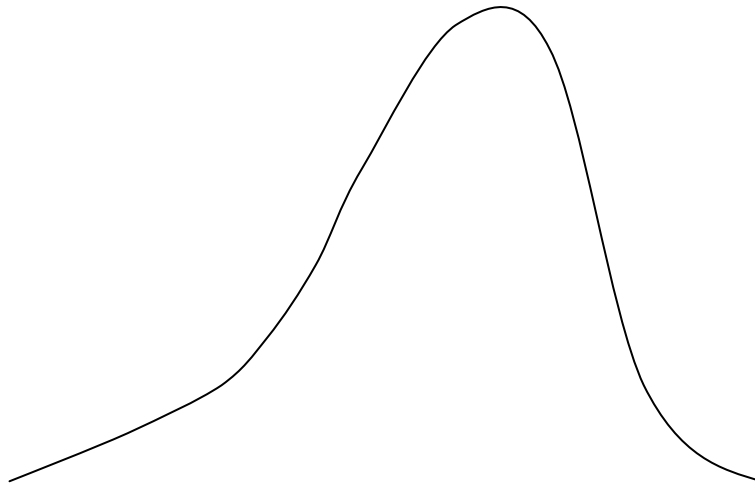


Figure B2

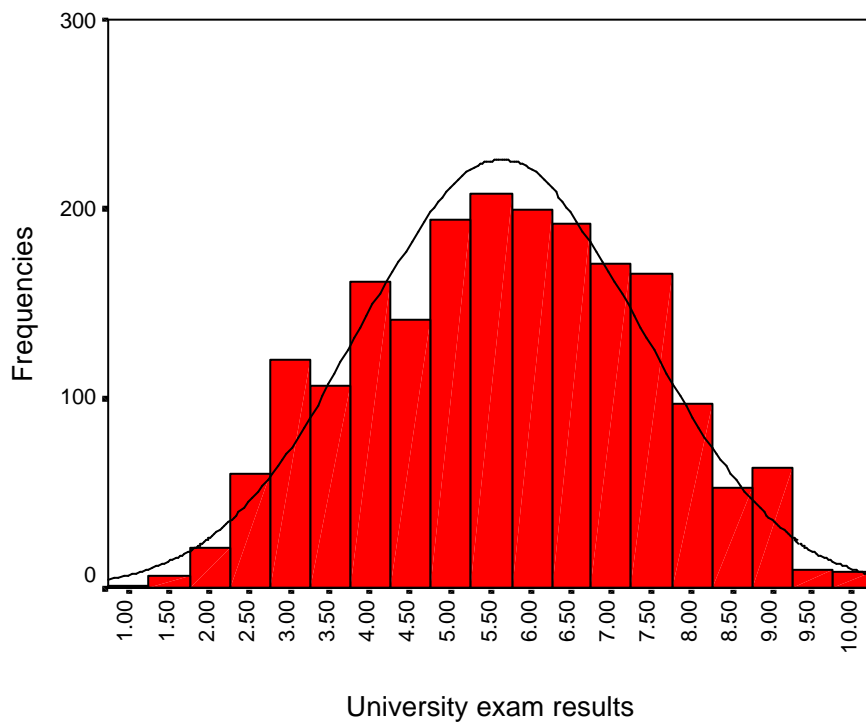


Figure B3

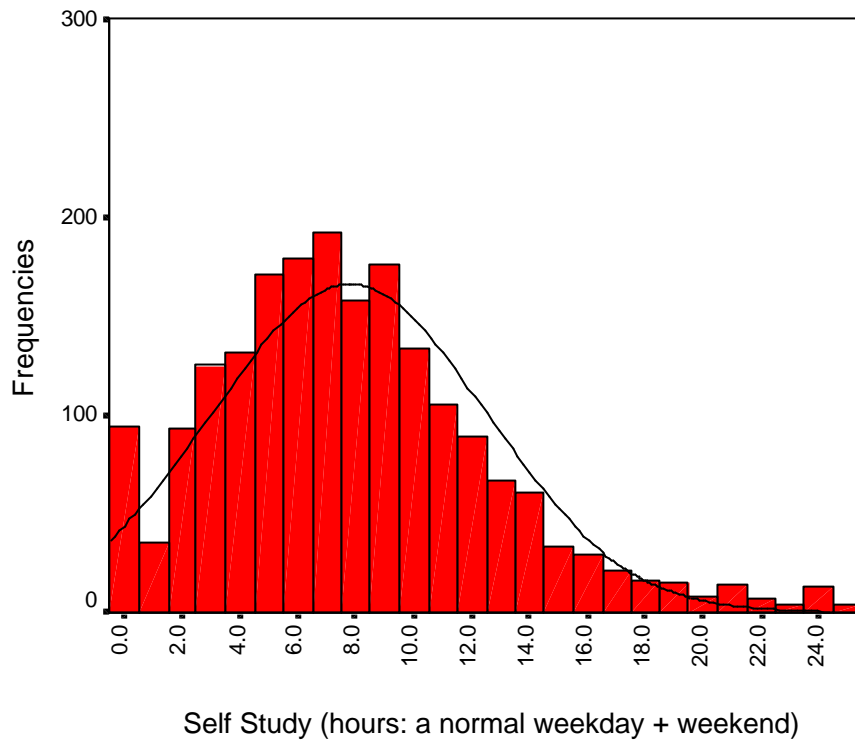
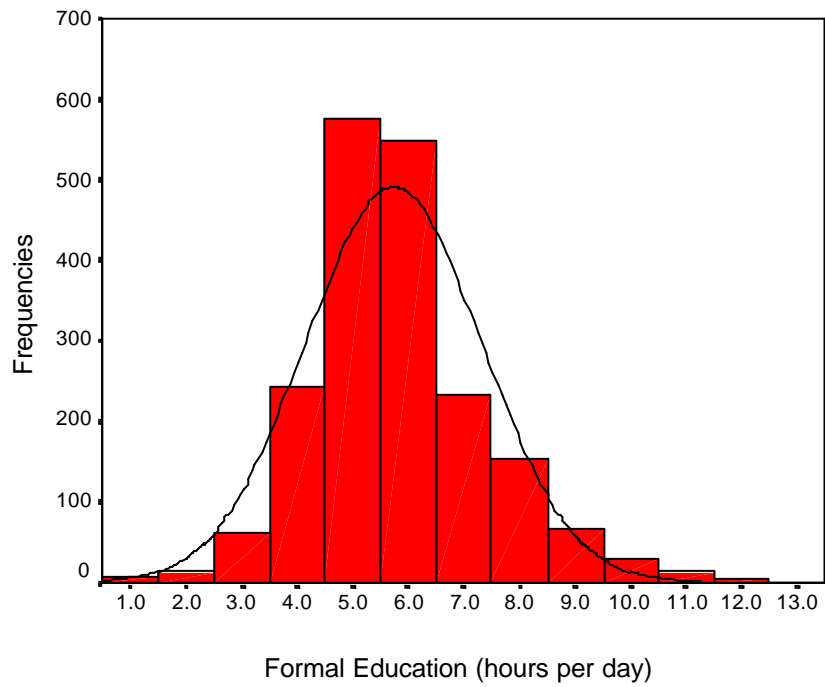


Figure B4



Appendix C

**Table C1: Stochastic Educational Production Function
(dropping motivation effects)**

Variables	Specification II			
	Output normalised		Output non normalised	
	Coefficient	t	Coefficient	t
Constant	5.817***	9.253	0.130	0.355
Personal Characteristics				
Age	0.069***	3.589	0.043***	3.803
Gender	-0.026	-0.287	-0.038	-0.763
Married	0.256	0.479	0.194	0.611
Nationality	-0.257	-0.720	-0.141	-0.670
Geographic Area	0.048	0.515	0.034	0.617
Own transport	0.069	0.824	0.037	0.749
Time Use^a				
Formal Education	2.734***	5.306	1.290***	4.320
Self Study	0.668***	3.559	0.343***	3.212
Private Tuition	-1.918*	-1.816	-1.179*	-1.840
IT/Language	0.015	0.028	-0.009	-0.030
Travel/Domestic	-1.177***	-3.432	-0.689***	-3.396
Leisure	-0.216**	-2.143	-0.119**	-2.023
Paid Work	-0.170	-0.452	-0.153	-0.686
Subjects				
Arts	0.412***	2.898		
Health	0.418*	1.645		
Pure Sciences	-0.820***	-4.800		
Engineering	-1.277***	-6.362		
Non Technical University Schools	0.569***	3.921		
Technical University Schools	-0.866	-0.306		
Parents' Characteristics				
Mother University studies	0.459	-1.053	0.259***	3.869
Parents divorced	-0.052	-0.306	-0.049	-0.490
Orphan	-0.194	-1.053	-0.120	-1.103
Family size	0.100***	-3.056	-0.065***	-3.349
Family income	0.00005	0.053	-0.00007	-0.119
Residence				
University Residence	-0.173	-0.792	-0.136	-1.104
Rent flat	-0.088	-0.848	-0.062	-1.022
Other characteristics related to the students' background				
Grant	0.009***	3.320	0.0006***	3.317
Bachillerato L.O.G.S.E.	-0.217**	-1.951	-0.114*	-1.749
Via F. P.	-0.501***	-3.369	-0.256***	-2.964
Via R.E.M.	0.832	1.506	0.542*	1.801
Parameters for compound error				
λ	1.588***	7.319	1.356***	1.356
$\sigma_{\epsilon} = (\sigma_v^2 + \sigma_u^2)^{1/2}$	2.242***	22.678	1.263***	1.263
σ_v^2		1.428		0.562
σ_u^2		3.601		1.034
Number of observations		1976		1976
-2 (log_R - log_U)		339.75***		116.45***

Note ^a: The time use variables are expressed in minutes per hour.

Note: * coefficient significantly different from zero at 10% confidence level; ** coefficient significantly different from zero at 5% level; *** coefficient significantly different from zero at 1% level.

**Table C2: Stochastic Educational Production Function
(Cobb-Douglas functional form)**

Variables	Specification I		Specification II		Specification III	
	Coefficient t	T	Coefficient	t	Coefficient t	t
Constant	1.732***	6.943	1.740***	6.390	2.479***	21.400
Personal Characteristics						
Log. Age	0.209***	2.990	0.226***	3.045		
Gender	0.005	0.343	0.0001	0.005	0.007	0.484
Married	-0.046	-0.441	-0.051	-0.474	-0.014	-0.172
Nationality	-0.050	-0.999	-0.062	-1.200	-0.061	-1.113
Geographic Area	0.009	0.609	0.009	0.599	0.004	0.254
Own transport	0.014	1.094	0.014	1.002	0.004	0.286
Time Use^a						
Log. Formal Education	0.021***	2.555	0.021***	2.485	0.019*	1.710
Log. Self Study	0.009**	2.027	0.009**	2.085	0.011***	2.509
Log. Private Tuition	-0.008**	-1.950	-0.007**	-1.951	-0.005	-1.340
Log. IT/Language	0.027**	1.931	0.030**	2.108	0.021	1.494
Log. Travel/Domestic	-0.016**	-2.197	-0.016**	-2.240	-0.016**	-2.149
Log. Leisure	-0.072***	-4.192	-0.077***	-4.555	-0.064***	-3.693
Log. Paid Work	-0.0004	-0.137	0.0001	0.028	-0.002	-0.524
Subjects						
Arts	0.069***	2.985	0.068***	2.925	0.074***	3.269
Health	0.049	1.179	0.054	1.274	0.099***	2.419
Engineering	-0.186***	-6.405	-0.190***	-6.442	-0.219***	-7.617
Pure Sciences	-0.078***	-3.024	-0.082***	-3.190	-0.099***	-3.898
Technical University Schools	-0.119***	-4.907	-0.113***	-4.519	-0.219***	-5.633
Non Technical University Schools	0.102***	4.474	0.109***	4.710	0.093***	4.110
Parents' Characteristics						
Mother university studies	0.073***	3.835	0.074***	3.845	0.085***	4.510
Parents divorced	-0.020	-0.777	-0.020	-0.771	-0.021	-0.761
Orphan	-0.036	-1.334	-0.031	-1.109	-0.029	-1.050
Log. Family size	-0.054**	-2.324	-0.059***	-2.427	-0.066***	-2.754
Log. Family income	0.008	0.650	0.004	0.316	-0.006	0.509
Residence						
University Residence	-0.038	-1.185	-0.041	-1.287	-0.026	-0.808
Rent flat	-0.017	-1.094	-0.015	-0.960	-0.018	-1.134
Motivation of the students						
Satisfaction	0.029**	2.159	0.033***	2.376	0.034***	2.481
Ambition	-0.036***	-2.788	-0.034***	-2.636	-0.038***	-2.974
Other characteristics of the students						
Log. Grant	0.008***	2.800	0.008***	2.827	0.007***	2.508
Bachillerato L.O.G.S.E.			-0.026	-1.399		
Via F. P.			-0.079***	-3.508		
Via R.E.M.			0.136	1.331		
First year student					-0.128***	-9.326
Parameters for compound error						
λ		9.224	5.494***	9.301	5.036***	10.710
$S_{\varepsilon} = (S_v^2 + S_u^2)^{1/2}$	5.512***	49.097	0.540***	49.148	0.522***	49.626
S_v^2	0.009		0.009		0.010	
S_u^2	0.284		0.281		0.262	
Number of observations	1976		1976		1976	
-2 (log_R - log_U)	302.78***		320.77***		399.86***	

Note ^a: The time use variables are expressed in minutes per hour.

Note: * coefficient significantly different from zero at 10% confidence level; ** coefficient significantly different from zero at 5% level; *** coefficient significantly different from zero at 1% level.

Table C3: Pre-University exam results estimations

	Specification I		Specification II		Specification III		Specification IV		Specification V	
	Coefficient	T	Coefficient	t	Coefficient	t	Coefficient	t	Coefficient	t
Constant	7.647***	36.346	7.775***	36.715	7.694***	36.579	7.672***	36.325	6.484***	155.218
Personal Characteristics										
Age	-0.051***	-5.156	-0.057***	-5.736	-0.054***	-5.407	-0.052***	-5.225		
Gender	-0.160***	-3.646	-0.148***	-3.333	-0.157***	3.562	-0.157***	-3.551		
Oldest	0.160***	3.135	0.155***	3.002	0.157***	3.074	0.160***	3.125	0.144***	2.711
Youngest	0.138**	2.388	0.117**	2.000	0.123**	2.124	0.143**	2.466	0.129**	2.143
Parents' Characteristics										
Father University studies	0.269***	4.165			0.353***	6.028				
Mother University studies	0.227***	3.078					0.358***	5.328		
Family income	0.0009*	1.726	0.002***	4.336	0.001**	2.427	0.001***	2.452	0.002***	3.601
Other characteristics of the students										
Private	0.061	1.106	0.135**	2.454	0.068	1.223	0.099*	1.806	0.267***	4.798
Bachillerato L.O.G.S.E.	-0.545***	-8.236	-0.560***	-8.391	-0.549***	-8.288	-0.549***	-8.274		
Via F.P.	0.677***	7.971	0.618***	7.239	0.671***	7.895	0.646***	7.607		
Via R.E.M.	-0.465*	-1.668	-0.475*	-1.684	-0.474*	-1.696	-0.460*	-1.646		
\bar{R}^2	0.11		0.09		0.10		0.10		0.03	
F _(11, 1964)	23.11***		22.64***		24.37***		23.49***		14.15***	

Table C4: Formal Education vs. Self Study (different specifications)

	Base model (Specification I)		Base model extended with pre-university via (Specification II)		Base model extended with Mature/non mature students (Specification III)	
	Non Normalised	Normalised	Non Normalised	Normalised	Non Normalised	Normalised
Formal Education	2.723	1.275	2.672	1.247	2.052	0.919
Self Study	0.626	0.329	0.640	0.328	0.623	0.333
	4.3	3.9	4.2	3.8	3.3	2.8

References

- Aigner, D. J. and Chu, S-F. (1968), "On Estimating the Industry Production Function", American Economic Review, 58 (4) (September), pp. 826-839.
- Aigner, D. J., Lovell, C. A. and Schmidt, P. (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models", Journal of Econometrics, 6 (1) (July), pp. 21-37.
- Battese, G. and Coelli, T. (1988), "Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data", Journal of Econometrics, 38, pp. 387-399.
- Battese, G. and Corra, G. (1977), "Estimation of a Production Frontier Model: With Application to the Pastoral Zone of Eastern Australia", Australian Journal of Agricultural Economics, 21, pp. 167-179.
- Becker, G. S. (1964), *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, New York: National Bureau of Economic Research, Columbia University Press, (2^a edition 1975).
- Becker, G. S. (1965), "A Theory of the Allocation of Time", The Economic Journal, vol. 75, Issue 299, pp. 493-517.
- Becker, G. S. (1981), *Treatise on the Family*, Harvard University Press, Cambridge, Mass.
- Bound, J., Jaeger, D., and Baker, R. (1995), "Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak", Journal of the American Statistical Association, vol. 90, No. 430.

- Bowles, S. (1970), "Towards an Educational Production Frontier", in W. L. Hansen (Ed.), *Education, Income and Human Capital*, New York, Columbia University Press, 11-60.
- Chalos, P. and Cherian, J. (1995), "An Application of Data Envelopment Analysis to Public Sector Performance Measurement and Accountability", Journal of Accounting and Public Policy.
- Coleman, J. S., et al. (1966), *Equality of Educational Opportunity* ("Coleman Report"), 2 vols, Washington, D. C.
- Fried, H. O., Lovell, C. A. K. and Schmidt, S. (1993), *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford University Press, 1993. – XII.
- Garcia-Crespo, D. (1999), "Promotion in the Spanish Labour Market: Differences by Gender", W.P. 11-99. Targeted Socio-Economic Research: Schooling, Training and Transition. LEO-CRESEP.
- Greene, W. (1990), "A Gamma Distributed Stochastic Frontier Model", Journal of Econometrics, 46, pp. 141-163.
- Greene, W. (1999), "Frontier Production Functions", from the Handbook of applied econometrics, Vol. II: Microeconomics. Edited by Hashem M. Pesaran and Peter Schmidt (Blackwell Publishers).
- Greene, W. (2000), *Econometric Analysis*, Fourth edition. Edited by Prentice Hall, Inc.
- Hanushek, E. A. (1986), "The Economics of Schooling Production and Efficiency in Public Schools", Journal of Economic Literature, XXIV, pp. 1147-1177.

- Harris, D. (1940), "Factors Affecting College Grades: A Review of the Literature, 1930-1937", Psychological Bulletin, vol. 37, Issue 3.
- Johnes, G. and Johnes, J. (1993), "Measuring the Research Performance of UK Economics Departments: An Application of Data Envelopment Analysis", Oxford Economic Papers, 45, pp. 332-347.
- Juster, F. and Stafford, F. (1991), "The Allocation of Time: Empirical Findings, Behavioral Models, and Problems of Measurement", Journal of Economic Literature, vol. XXIX (June 1991), pp. 471-522.
- Juster, F. and Stafford, F. (1986), "Response Errors in the Measurement of Time Use", Journal of the American Statistical Association, 81 (394), pp. 390-402.
- Kirjavainen, T. and Loikkanen, A. (1998), "Efficiency Differences of Finnish Senior Secondary School: An Application of DEA and Tobit Analysis", Economics of Education Review, vol. 17, n° 4, pp. 377-394.
- Lassibille, G.; Navarro, L. (1990), *El Valor del Tiempo en la Universidad*. Textos mínimos. Universidad de Málaga.
- Meeusen, W. and van den Broeck, J. (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error", International Economic Review, 18, pp. 435-444.
- Meyer, R. (1997), "Value-Added Indicators of School Performance: A Primer". Economics of Education Review, vol. 16, Issue 3.
- Mincer, J. (1958), "Investment in Human Capital and Personal Income Distribution", Journal of Political Economy, 66 (August), pp. 281-302.

Mulligan, C., Schneider, B. and Wolfe, R. (2000), "Time Use and Population Representation in the Sloan Study of Adolescents", NBER Technical Working Paper 265.

Norman, M. and Stocker, B. (1991), *Data Envelopment Analysis. The Assessment of Performance*, John Wiley & Sons.

Schultz, T. W. (1960), "Capital Formation by Education", Journal of Political Economy, 68, pp. 571-583.

Stevenson, R. (1980), "Likelihood Functions for Generalised Stochastic Frontier Estimation", Journal of Econometrics, 13, pp. 58-66.

Todd, P. and Wolpin, K. (2000), "On the Specification and Estimation of the Production Function for Cognitive Achievement", University of Pennsylvania, mimeo.

CENTRE FOR THE ECONOMICS OF EDUCATION
Recent Discussion Papers

- | | | |
|---|---|--|
| 9 | Peter Dolton
Mary Silles | Over-Education in the Graduate Labour Market: Some Evidence from Alumni Data |
| 8 | Arnaud Chevalier
Gauthier Lanot | The Relative Effect of Family and Financial Characteristics on Educational Achievement |
| 7 | Arnaud Chevalier | Graduate Over-Education in the UK |
| 6 | Barbara Sianesi and
John Van Reenen | The Returns to Education: A Review of the Macro-Economic Literature |
| 5 | C. Harmon
H. Osterbeek
I. Walker | The Returns to Education
A Review of Evidence, Issues and Deficiencies in the Literature |
| 4 | L. Dearden
S. McIntosh
M. Myck
A. Vignoles | The Returns to Academic and Vocational Qualifications in Britain |
| 3 | S. McIntosh
A. Vignoles | Measuring and Assessing the Impact of Basic Skills on Labour Market Outcomes |
| 2 | A. Vignoles
R. Leva..if
J. Walker
S. Machin
D. Reynolds | The Relationship Between Resource Allocation and Pupil Attainment: A Review |
| 1 | A. Vignoles
T. Desai
E. Montado | An Audit of the Data Needs of the DfEE Centres for the Economics of Education and the Wider Benefits of Learning |

To order a discussion paper, please contact the Publications Unit
Tel 020 7955 7673 Fax 020 7955 7595