How students' exogenous characteristics affect faculties' inefficiency. A heteroscedastic stochastic frontier approach

Zotti, Roberto and Barra, Cristian

University of Salerno

February 2014

How students' exogenous characteristics affect faculties' inefficiency. A heteroscedastic stochastic frontier approach

Abstract

By using a heteroscedastic stochastic frontier model, this paper focuses on how students' exogenous characteristics (such as personal demographic information, pre-enrollment educational background and household economic status) affect faculties' inefficiency. Using individual data on freshmen enrolled at a public owned university in Italy over the 2002-2008 period, we focus both on the direction of this influence on technical inefficiency and on the magnitude of the related partial effects. A measure of R^2 has also been calculated in order to evaluate the overall explanatory power of the exogenous variables used. The empirical evidence reveals the validity of the heteroscedastic assumption, giving credit to the use of some students' individual characteristics according to which the inefficiency is allowed to change. Moreover, the estimates suggest that the university could improve the students' performances by investing in labour inputs.

Keywords: Stochastic frontier analysis; Technical inefficiency estimates; Heteroscedasticity; Higher education.

JEL-Codes: I21, I23; C14; C67

1. Introduction

Two main methods have been extensively applied in the literature for measuring efficiency: non-parametric and parametric. In particular, the non-parametric methods, such as the DEA (Data Envelopment Analysis) and FDH (Free Disposable Hull), proposed by Charnes et al. (1978) and due to the original contribution of Farrell (1957), are based on deterministic frontier models (see also Cazals et al. 2002). DEA model, extended by Banker et al. (1984), is especially adequate to evaluate the efficiency of non-profit entities that operate outside the market, since for them performance indicators, such as income and profitability, do not work satisfactorily (for more theoretical details on DEA see Coelli et al. 1998, Cooper et al. 2004, Ramanathan, 2003 and Thanassoulis, 2001). Instead, the parametric approaches, such as Stochastic Frontier Approach (SFA), Distribution-Free Approach (DFA) and Thick Frontier Approach (TFA) are based on stochastic frontier models (see Aigner et al. 1977). These two main approaches have not only different features, but also advantages and disadvantages (Lewin and Lovell, 1990). On one hand, the non-parametric method does not require the building of a theoretical production frontier, but the imposition of certain, a priori, hypotheses about the technology (free-disposability, convexity, constant or variable returns to scale). However, if these assumptions are too weak, the level of inefficiency could be systematically underestimated in small samples, generating inconsistent estimates. Furthermore, this method is very sensitive to the presence of outliers. On the other hand, the parametric method uses a theoretical analysis to construct the efficient frontier, it's not sensitive to extreme values because imposes some assumptions on the error distribution, but must deal with the problem of decomposing the error term. In particular, SFA, proposed by Aigner et al. (1977), Meeusen and Van den Broeck (1977) and Battese and Corra (1977), assumes that the error term is composed by two components with different distributions (see Kumbhakar and Lovell (2000) for analytical details on stochastic frontier analysis). The first component, regarding the "inefficiency", is asymmetrically distributed (typically as a semi-normal), while the second component, concerning the "error", is distributed as a white noise. In this way, it is necessary to assume that both components are uncorrelated (independent) to avoid distortions in the estimates. Nowadays, the most widely applied SFA technique is the model proposed by Battese and Coelli (1995) to measure technical efficiency across production units[1].

Concerning the topic of performance measurement in tertiary education, only a few studies emphasize the import role of students' individual characteristics on universities' inefficiency (among the exceptions Laureti (2008) and Zoghbi et al. (2013)). Our main aim is, therefore, to contribute to this literature trying to give new insights and to find new empirical evidence. By using individual data taken from several cohorts of students, in this paper SFA is preferred to estimate a stochastic production function for higher education[2]. Relaxing the homoscedasticity assumption, we take into account a vector of exogenous student's characteristics in order to examine whether they have a key role in explaining the inefficiency of the faculties within the University of Salerno. We focus both on the direction of the influence of these exogenous factors on technical inefficiency and on the magnitude of the related partial effects. The validity of the heteroscedastic assumption is tested using a Likelihood Ratio (LR) test which allows us to identify the fit of the model. A measure of R^2 is also calculated in order to evaluate the overall explanatory power of the exogenous variables on technical inefficiency. The rest

---

[1] Instead, DFA, unlike SFA, assumes that the "error" is fixed over time, while the "inefficiency" component is normally distributed. Finally, TFA assumes that the radial distance between the efficient frontier and the lowest and highest quartile is the white noise, while the deviation between the two quartiles indicates inefficiency.

[2] As Ferrari and Laureti (2005) have pointed out, "part of the controversy over the subject of performance measurement in higher education focuses on analysis used".

of the paper is organized as follows. Section 2 describes the methodology, Section 3 illustrates the research design, Section 4 gives information about the data, Section 5 describes the empirical results and finally Section 6 concludes.

2. Empirical Methodology

Considering a cross section framework, a conventional stochastic frontier model can be expressed as follows:

$$y_i = f(x_i, \beta)\exp\{v_i - u_i\} \tag{1}$$

where y_i is the output of faculty i; x_i is a vector of input quantities of faculty i; β is a vector of unknown parameters; $f(x_i, \beta)$ is the production function or conventional regression model (adding the stochastic error, v_i); v_i is a vector of random variables related to the idiosyncratic or stochastic error term of faculty i assumed to be $i.i.d.$ $N(0, \sigma_v^2)$ and independent of the u_i, while u_i is a vector of non-negative random variables measuring the inefficiency term of faculty i. Using the scaling property in equation (1), u_i is assumed to be heteroscedastic and, in particular, distributed as $\sigma_{ui}^2 = exp(z_i; \delta)$ times a half-normal distribution, i.e. $N^+(0,1)$, where z_i is a vector of individual students' characteristics employed to explain faculty-specific inefficiency, and δ denotes a vector of unknown coefficients. In other words, the inefficiency of faculty i is assumed to systematically vary with respect to the different characteristics associated to each student. For instance, if a student of faculty A performs better than a student of faculty B, then the faculty A has a high probability to achieve the best practice frontier. The validity of the heteroscedastic assumption is tested using a LR test which allows us to identify the fit of the model and to confirm the imposition of some exogenous students' characteristic in the variance of the inefficiency term.

The distribution of the inefficiency component, i.e. half-normal, produces a model that satisfies the scaling property. The main advantage of this property is that changes the scale of the inefficiency distribution, but not its shape. Moreover, for the same reasons provided by Laureti (2008), the idiosyncratic error term is assumed to be homoscedastic, i.e. the random variables affect the output of students with the same dispersion and magnitude. All coefficients of parameters in equation (1) are estimated using "maximum likelihood estimation". Specifically, a Cobb-Douglas production function is preferred in this paper especially because it allows us to overcome the multicollinearity problem associated to estimate a few number of parameters with respect to the translog function.

Indexing the exogenous factors with $k = 1, \dots, K$, marginal effects have been calculated, as suggested by Wang (2002) and Wang (2003), through the measures $\partial[E(u_i|x_i, z_i)]/\partial z_{ik}$ and $\partial[V(u_i|x_i, z_i)]/\partial z_{ik}$, which could be interpreted, respectively, as the partial effect of z_{ik} on y_i[3] and the partial effect of z_{ik} on production uncertainty.

Following Liu and Myers (2009), we measure how well the vector of exogenous variables explains faculties' inefficiency through an R^2 type measure, as below reported:

$$R_z^2 = \frac{\sum_{i=1}^{n}\left[\hat{E}(u_i|z_i) - \frac{1}{n}\sum_{i=1}^{n}\hat{E}(u_i|z_i)\right]^2}{\sum_{i=1}^{n}\left[\hat{E}(u_i|z_i) - \frac{1}{n}\sum_{i=1}^{n}\hat{E}(u_i|z_i)\right]^2 + \sum_{i=1}^{n}\hat{V}(u_i|z_i)} \tag{2}$$

where $\hat{E}$ and $\hat{V}$ indicate sample estimates of the mean and variance of u_i conditional on z_i.

---

[3] As y_i is log output, it could be also interpreted as the semi-elasticity of output with respect to the exogenous variables.

3. Research design: The production set

3.1. Inputs

The following variables have been used as inputs: number of professors per student, total expenditure on professor per student, number of lecture halls and student's final high school grades.

The first input is the professor-student ratio (PSR). It is a measure of labour input and it aims to capture the human resources assigned by the faculties on teaching activities. We include the total number of academic staff (professors, associate professors, researchers and assistant professors).

The second input is the total expenditure on professor per student (EPS). It is a measure of the cost of teaching and it aims to measure the financial resources the faculty allocates for teaching activities.

The third input is the number of lecture halls (LH). It is a measure of capital input and it aims to capture part of the equipment which is used by the faculty for teaching purposes.

The forth input is represented by the student's final high school grades (HGS). It is a measure which aims to capture the quality of the students on arrival at university[4]. It can be considered as a proxy of the knowledge and skills of students when entering tertiary education.

3.2. Outputs

The output choice is firstly related to the standard way of measuring student's performances. Moreover, it is also related to the institutional background we have to deal with. To be more precise, in order to reach European standards, the Italian higher education system has been reformed mainly in the 1990s and at the beginning of the 2000s[5]. The quality of education was meant to increase by reducing the number of people who leave the universities and by improving the academic performances in a way that students would acquire a number of credits and exams as close as possible to those theoretically obtainable in a given year and would get the degree in a time as close as possible to the one legally established by the degree course regulations.

Following these guidelines, as well as the previous literature, we consider the grade point average (GPA) of freshmen as outcome, representing the grades, weighted by the corresponding credits, obtained by the students after the 1st year, calculated in the following way (for each student i and exam j):

$$Grade\ point\ average\ (GPA)_i = \frac{\sum_{j=1}^{n} grades_{ij} * credits_{ij}}{\sum_{j=1}^{n} credits_i} \tag{3}$$

It has been considered the GPA at the end of the first year for two reasons. Firstly, because we want to measure the performances in a specific moment of time for all freshmen. Secondly, and more importantly, the choice follows the Italian Ministry of Education, Universities and Research guidelines, according to which, higher education institutions are evaluated also on the base of indicators such as the number of students leaving the university after the first year or the number of

[4] There is a strong evidence that pre-university academic achievement is an important determinant of the students' performances (Boero et al. 2001; Smith and Naylor, 2001; Arulampalam et al. 2004; Lassibille, 2011). The underlying theory is that ability of students lowers their educational costs and increases their motivation (DesJardins et al. 2002).

[5] Universities have started to be funded according to their level of virtuosity, in order to achieve higher research performances and to promote academic excellence. Both quantitative and qualitative indicators were developed to accurately evaluate their productivity in research and teaching.

students who enroll in the second year having acquired a certain amount of credits. In other words, the transition between the first and the second year has been considered as the main checkpoint to evaluate the regularity of the educational path.

3.3. Exogenous variables

As Laureti (2008) underlined, "it seems inadequate to assume that the variability of the student efficiency behaviour in carrying out the formation process is the same for each student". Therefore, given that several exogenous variables are available, the use of a heteroscedastic stochastic frontier model is particularly suitable for our analysis, to adequately measure the effects of students' exogenous characteristics on faculties' inefficiency. In order to do that, we include in the variance of the inefficiency component (Caudill et al. (1995), Hadri (1999) and Wang (2002)) the following explanatory variables, which vary from student to student: type of high school (LYC) is a dummy variable with value 1 if the student has attended a lyceum and value 0 otherwise, gender (MAL) is a dummy which has a value of one if the student is a male and zero otherwise, age (AGE) is a continuous variable, the age of the freshmen, residence[6] (RES) is a continuous variable measuring the distance between the university campus and the student's place of residence, income[7] (INC) is a dummy taking the value of one is the student has a high income (income greater than 35,000.00 €) and value 0 otherwise, dropout[8] (DOUT) is a dummy which takes the value of one if the student has dropped out by the end of the 1st year and zero otherwise; type of faculty (FAC), is a set of dummies assigning each student to a faculty, and finally time trend (YEAR), a set of dummies related to the academic year of enrollment, captures the influence of technical change leading a shift in the output over time. These factors will be modelled as variables which directly influence the variability of the inefficiency term. As Zoghbi et al. (2013) pointed out, "these variables are considered exogenous in the sense that they influence the production process but are not themselves either inputs or outputs". They, in fact, affect the efficiency with which inputs are converted into outputs.

3.4. Specification of the models

The diversification in the quality of the inputs allows us to explore whether the results obtained are sensitive to the specification adopted, as summarized by the Table 1 below.

---

[6] About the residence variable, the main literature usually refers to the student's residence in the city, province or region where the university is located. Given the geographical peculiarity of the province of Salerno respect to the other provinces of Campania and given the geographical position of the University of Salerno (the campus is located approximately 15 kilometers from the city of Salerno), the residence variable measures, for each student, the distance, in kilometers, of the student's residence from the university location. In order to calculate that distance, Google map has been used. Specifically the distance is considered as the best and fastest way, suggested by Google map, to reach the university campus.

[7] About the family income variable, a good measure of family income was lacking. A measurement of the student's household economic situation (ISEE), which takes into account the household income, personal estate and number of members, has been used as a proxy for income. Specifically, those students who declared an income higher than 35,000.00 € are classified as high income students.

[8] A broader dropout definition rather than the formal one used by the Universities administration offices has been considered. A student drops out, in line with some previous research, both when she officially withdraws from the university (the so called "rinunciatari") presenting a formal request to the student office, and when she does not renew the registration leaving the degree program in which had been enrolled. Students who do not renew their registration but asked to move to another university are not considered as dropout. Moreover, students who do not renew their registration but are found to be enrolled in another Faculty of University of Salerno are not considered dropout either. This is to avoid to put together forms of leaving behavior different in their characteristics. According to Tinto (1975), failure to separate permanent dropout from temporary and transfer behaviors has often led institutional and state planners to overestimate substantially the extent of dropout from higher education.

Table n. 1 – Specification of inputs, outputs and exogenous factors in SFA models [1]

	Model A	***Model B***
Inputs	*PSR; LH; HSG*	*EPS; LH; HSG*
Outputs	*GPA*	*GPA*
Explaining the inefficiency	*LYC; MAL; AGE; RES; INC; DOUT*	*LYC; MAL; AGE; RES; INC; DOUT*

Notes:
PSR: professor-student ratio
LH: number of lecture halls
HSG: high school grades
EPS: expenditure on professor per student
GPA: grade point average at the end of the 1st year
LYC: students who attended a lyceum
MAL: males
AGE: age of the students
RES: distance between the campus and the student's residence
INC: students with high income
DOUT: students who drop out at the end of the 1st year

1 Faculty dummies (FAC) are included in the variance of the inefficiency component while the time trend(YEAR) is included both in the production function and in the variance of the inefficiency component.

Following the equation (1) in section 2, the benchmark model (Model A in Table 1) is specified as follows:

$$y_i = ln(GPA_i) \tag{4}$$

$$x_i = \{ln(PSR_i)\,; ln(LH_i)\,; ln(HSG_i); YEAR_t\} \tag{5}$$

$$z_i = \{LYC_i;\ MAL_i;\ AGE_i;\ RES_i;\ INC_i;\ DOUT_i; FAC_i;\ YEAR_t\} \tag{6}$$

where GPA is the value of output as a continuous variable measuring the grades, weighted by the corresponding credits, obtained by the students at the end of the first year. PSR (the number of professors per enrolled student), LH (the number of lecture halls) and HSG (a continuous variable measuring the students' high school grades) are used as inputs. Keeping constant the output side, we explore whether the amount of expenditure on professor per student (EPS) might represent an alternative way of measuring the labour input (Model B in Table 1). In both specifications, the type of high school attended (LYC), gender (MAL), age (AGE), the distance between the campus and the student's residence (RES), student economic status (INC) and the dropout at the end of the first year (DOUT) are used as exogenous variables (z-vector) in order to capture their impact on the variance of the inefficiency term. Faculty dummies (FAC) are also included in the variance of the inefficiency component. Finally, the time trend (YEAR) captures two different effects: (i) technical change over time when it's included in the production function and (ii) inefficiency change over time when it's included in the variance of inefficiency term[9].

4. Data

The production frontier model is estimated by using individual data taken from an unique administrative dataset on post-reform students[10] enrolled at University of Salerno in the academic years 2002/2003 (Cohort 2002), 2003/2004 (Cohort 2003), 2004/2005 (Cohort 2004), 2005/2006, (Cohort 2005), 2006/2007 (Cohort 2006), 2007/2008, (Cohort 2007) and 2007/2008, (Cohort 2008). The dataset gathers information about individuals' characteristics (gender, age), educational background and pre-enrollment characteristics (type of high school attended, high school diploma score), households'

[9] We also estimated the models without including the time trend both in the production function and variance of the inefficiency term and the results (available on request) do not change.

[10] By post-reform students we mean students enrolled in a three year degree program after that a reform process of the Italian higher education system has been implemented, in order to meet the objectives of the Bologna process, by the Decree n. 509/99. The reform changed the length of the degree programs introducing a first Degree which lasts three years, followed by a two-year Specialized Degree.

financial conditions (family declared income) and general information about the university careers and performances (exams passed and credits acquired). See more detailed descriptive statistics in Table 2 below[11].

The analysis is carried out on a total number of 32648 freshmen, after deleting outliers, distributed on 9 faculties[12].

Table n. 2 – Definition of the variables and descriptive statistics – Mean values, 2002-2008 period

			Mean values
Inputs			
PSR	*Professor-student ratio*	*Number of professors per enrolled student*	*0.1333 (0.1090)*
EPS	*Expenditure on professor per student*	*Financial expenditure on professor per enrolled student*	*4555. 779 (3688.27)*
LH	*Number of lecture halls*	*Number of lecture halls used by the faculty for teaching purposes*	*17.7742 (8.1004)*
HSG	*High School grades*	*High school final grades*	*79.0193 (12.7763)*
Outputs			
GPA	*Grade Point Average*	*Grade Point Average at the end of the 1st year*	*24.7490 (2.8744)*
Explaining the inefficiency			
LYC	*Type of High School*	*Students who attended a lyceum*	*0.4210 (0.4937)*
MAL	*Gender*	*Males*	*0.4598 (0.4983)*
AGE	*Age*	*Age of the students*	*21.1935 (5.3156)*
RES	*Residence*	*Distance between the university campus and the place of residence*	*42.6076 (61.1251)*
INC	*Income*	*Students with high income*	*0.1735 (0.3787)*
DOUT	*Dropout at the end of the 1st*	*Students who drop out at the end of the 1st year*	*0.3503 (0.4770)*

Source: own calculations using data on students enrolled at the University of Salerno. Standard deviations are shown in parentheses.
Note: Expenditures on professors per student in thousands of deflated 2005 euros.

5. The Empirical Evidence

A Cobb-Douglas production function approach has been applied to Models A and B (see Table 1 for details) in order to estimate the teaching efficiency of faculties within the University of Salerno over the 2002-2008 period, pooling all years together. Table 3 below, describes the results from individual data both for the variables included in the production function and for the variables used to explain the inefficiency term and also shows the average efficiency scores from stochastic frontier estimations. All the estimates have been obtained using the STATA 12 software.

First of all, the null hypothesis that there is no heteroscedasticity in the error term has been tested and rejected, at 1% significance level, using a Likelihood Ratio Test (LR), giving credit to the use of some exogenous variables, according to which the inefficiency term is allowed to change. In other words, the validity of heteroscedastic assumption has been confirmed[13]. The goodness of fit statistics for the inefficiency component (R^2) is 0.50 and 0.49, respectively for models A and B, indicating that almost 50% of the sample variation in inefficiency could be explained by the exogenous variables

[11] Descriptive statistics on each faculty are not presented in the paper and are available on request.

[12] Faculties gather together different departments according to their similarities; they are organized into different subject areas, each offering a number of degree courses, aiming at coordinate the teaching activities. We claim to treat the faculties within the same university as homogenous according to the fact that they operate in similar activities, using both academic and non-academic staff for teaching purposes (see Tyagi et al. 2009).

[13] The test confirms that the imposition of some exogenous students' characteristics in the variance of the inefficiency component improves the fit of our model.

used. Starting from the variables included in the production function, the estimates show that the larger is the number of professors per student as well as the amount of expenditures on professors per student, the higher is the student GPA at the end of the first year. The number of lecture halls is not a statistically significant determinant of the students' performances. The evidence suggests that the university could improve the students' performances by investing in labour inputs. Moreover, according to what we consider a proxy of the student's quality at the entrance, the greater is the percentage of students with a higher secondary school final grade, the higher are the students' performances at the end of the first year.

Table n. 3 - Estimates for the production frontier, for the inefficiency components and average efficiency scores from stochastic frontier estimations - Mean values, 2002-2008 period

Variables	***Model A***	***Model B***
	Coefficients	***Coefficients***
Output: *Grade point average at the end of the 1st year (GPA)*		
Inputs		
Professor-student ratio (PSR)	*0.0093****	
	(0.0021)	
Expenditure on professor per student (EPS)		*0.0086****
		(0.0023)
Number of lecture halls (LH)	*0.0008*	*0.0010*
	(0.0024)	*(0.0025)*
High School grades (HSG)	*0.2434****	*0.2437****
	(0.0035)	*(0.0035)*
Explaining the inefficiency		
Students who attended a lyceum (LYC)	*-0.6962****	*-0.6988****
	(0.0307)	*(0.0307)*
Males (MAL)	*0.2277****	*0.2280****
	(0.0315)	*(0.0315)*
Age (AGE)	*-0.0285****	*-0.0286****
	(0.0043)	*(0.0044)*
Residence (RES)	*0.0004*	*0.0004*
	(0.0002)	*(0.0002)*
Students with high income (INC)	*-0.0943***	*-0.0944***
	(0.0384)	*(0.0385)*
Students who drop out at the end of the 1st year (DOUT)	*0.0061****	*0.0068****
	(0.0023)	*(0.0023)*
Log-likelihood	*28466.537*	*28464.252*
LR statistic	*5697.89*	*5884.00*
Wald statistic	*4776.68*	*4771.84*
R_z^2	*0.5000*	*0.4981*
Faculty dummies (FAC)	*Yes*	*Yes*
Time trend (YEAR)	*Yes*	*Yes*
Average efficiency scores from stochastic frontier estimations		
Faculties	***Average efficiency scores***	***Average efficiency scores***
Faculty n. 7	*1.0000*	*1.0000*
Faculty n. 5	*0.9806*	*0.9803*
Faculty n. 6	*0.9575*	*0.9578*
Faculty n. 2	*0.9358*	*0.9371*
Faculty n. 9	*0.9213*	*0.9217*
Faculty n. 3	*0.9211*	*0.9215*
Faculty n. 8	*0.9044*	*0.9045*
Faculty n. 4	*0.8973*	*0.8977*
Faculty n. 1	*0.8907*	*0.8910*

Notes: Standard errors are shown in parentheses.
* p<0.10; ** p<0.05; *** p<0.01
Average efficiency scores from stochastic frontier estimations are presented in descending order of average efficiencies.
Faculty dummies are included in the variance of the inefficiency component while the time trend is included both in the production function and in the variance of the inefficiency component.

Considering the exogenous factors included in the analysis, our evidence shows that the students' characteristics have an important role in describing the variance of the inefficiency term. A more academic oriented high school attendance decreases faculties' inefficiency. This confirms that the secondary school track chosen represents a channel through which the family environment (consolidating the intergenerational correlation in the educational attainment) influences the level of education completed (Checchi et al. 2013, Carneiro and Heckman, 2005). A positive and statistically significant coefficient has been found on the variable gender, meaning that, ceteris paribus, males are related to a higher level of faculties' inefficiency. What we consider a proxy of the student economic status appears to well explain the inefficiency term, too. The negative and statistically significant estimated coefficient on the variable income indicates that the variance between students with high family income is lower than between those with low family income, ceteribus paribus, showing a lower inefficiency for high income students. Particularly interesting is the positive and statistically significant estimated coefficient of the average dropout rate in the faculty at the end of the 1st year. This indicates that, all other variables being constant, inefficiency rises with the presence of students that dropout from the university. This is in line with the main reforms the Italian higher education system went through, according to which the quality of education was meant to increase also by reducing the number of students who leave the universities[14]. The negative and significant estimated coefficient associated with age indicates that, ceteris paribus, inefficiency decreases with age. The estimated coefficient relative to the student place of residence is positive but it is not statistically significant.

In order to quantify the effects of the exogenous factors, we compute the partial effects, as Table 4 summarizes, along with their standard errors.

Table n. 4 - Partial effects of exogenous factors

	Model A	***Model B***
Partial effects on $E(u_i\|x_i, z_i)$		
Students who attended a lyceum (LYC)	-0.02427	-0.02428
	(0.00048)	(0.00048)
Males (MAL)	0.00919	0.00916
	(0.0005)	(0.0005)
Age (AGE)	-0.00077	-0.00076
	(0.00005)	(0.00005)
Residence (RES)	0.000014	0.000013
	(0.000004)	(0.000004)
Students with high income (INC)	-0.00388	-0.00388
	(0.0006)	(0.0006)
Students who drop out at the end of the 1st year (DOUT)	0.00023	0.00025
	(0.00003)	(0.00003)
Partial effects on $V(u_i\|x_i, z_i)$		
Students who attended a lyceum (LYC)	-0.00264	-0.00265
	(0.00008)	(0.00008)
Males (MAL)	0.00114	0.00114
	(0.00008)	(0.00008)
Age (AGE)	-0.00004	-0.00004
	(0.000009)	(0.000009)
Residence (RES)	0.000001	0.000001
	(0.000006)	(0.000006)
Students with high income (INC)	-0.00037	-0.00037
	(0.0001)	(0.0001)
Students who drop out at the end of the 1st year (DOUT)	0.00003	0.00003
	(0.0001)	(0.000006)

Notes: Standard errors are shown in parentheses.

[14] The Italian Ministry of Education, Universities and Research clearly considers a high dropout rate as a signal of a system that does not work perfectly. As a consequence, especially the transition between the 1st and the 2nd year has been strongly incentivized.

The marginal effects on $E(u_i|x_i, z_i)$ and $V(u_i|x_i, z_i)$ measure how an increase or decrease in the exogenous variables changes the expected inefficiency and the production uncertainty, respectively (Bera and Sharma, 1999). To obtain the semi-elasticity of the dummy variables, we take the anti-log of the dummy coefficient, subtract 1 from it, and multiply the difference by 100%.

Looking at the partial effects on $E(u_i|x_i, z_i)$, results show that students who attended an academic high school tend to be more efficient. Since $\partial E(\ln y)/\partial Lyc = -\partial E(u)/\partial Lyc$, the effects turns into an increase in output by almost 2.5%, while males tend to be less efficient than females with a decrease in output by 1%. Moreover, considering the variable income, there is evidence that students with high income decrease technical inefficiency. The effect translates into an increase in output by 0.4%. Turning to the variable related to the dropout, the positive average marginal effect means that an increase in the percentage of student who drop out at the end of the first year increases technical inefficiency, being associated with a decrease in output by 0.02%. Although the effect is quantitatively small, it is statistically significant. The average marginal effect of the variable age indicates that an increase in age decreases technical inefficiency. If age increases by a year, the output increases by about 0.08%.

Considering the partial effects on $V(u_i|x_i, z_i)$, production uncertainty decreases with age, students who attended a lyceum and students with high income. On the other hand, males and the presence of students who dropped out increase uncertainty.

6. Concluding remarks

This paper analyses how students' exogenous characteristics affect the inefficiency of the faculties within the University of Salerno over the 2002-2008 period, using individual data. We use a heteroscedastic stochastic frontier model. This approach has two components: one being a stochastic production frontier according to which the faculties' inefficiency is measured, and the other one being a one-sided error term which captures technical inefficiency. Specifically, we let the inefficiency term to depend on students' characteristics (such as personal demographic information, pre-enrollment educational background and household economic status), focusing both on the direction of the influence and on the magnitude of the partial effects.

The LR test confirms that the homoscedastic hypothesis is not suitable in our analysis, suggesting that the student's exogenous characteristics have an important role in explaining the inefficiency of the faculties within the University of Salerno. A measure of R^2 has been calculated, revealing that for both the models considered about 50% of the sample variation in the inefficiency component could be explained by the exogenous variables used. A positive and statistically significant relationship between two measures of labour input (such as the number of professors per student and the amount of expenditures on professors per student) and the students' performances has been found. This evidence suggests that the university could improve the students' performances by investing in labour inputs. Moreover, considering a proxy of the student's quality at the entrance, the greater is the percentage of students with a higher secondary school final grade, the higher are the students' performances. Exogenous variables do have a statistically significant effect on technical inefficiency. We found evidence that the presence of students from a more academic oriented high school has a positive effect on faculties' efficiency, confirming the importance of the secondary school track chosen. On the other hand, a negative relationship between the average dropout rate and the faculties' efficiency has been found. This is in line with the main reform the Italian higher education system went through, according to which the quality of education was meant to increase also by reducing the number of students who leave the universities.

The aim of our analysis is to give a contribution to the literature which attempts to estimate a production function for tertiary education, paying particular attention on the fundamental role that students' characteristics play on faculties' inefficiency. Hopefully, universities' regulators might take advantage of these studies through appropriate policy decisions in order to make the higher education system more efficient. Even though more empirical evidence has to be provided, the use of this approach could be useful in order to get more information on how efficiency analyses of higher education institutions might be improved when exogenous student's characteristics are taken into account.

References

Aigner, D., Lovell, K. and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6, 21–37.

Arulampalam, W., Naylor, R. A. and Smith, J. P. (2004). A hazard model of the probability of medical school drop-out in the UK. Journal of the Royal Statistical Society: Series A (Statistics in Society), 167 (1), 157-178.

Banker, R.D., Charnes, A. and Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in DEA. Management Science, 32, 1613-1627.

Battese, G.E. 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, 169-179.

Battese, G.E. and Coelli, T.J. (1995). A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics, 20, 325-332.

Bera, A.K. and Sharma, S.C. (1999), Estimating production uncertainty in stochastic frontier production function models. Journal of Productivity Analysis, 12, 187-210.

Boero G., McNight, A., Naylor, R. and Smith, J. (2001). Graduates and graduate labour markets in the UK and Italy. Lavoro e Relazioni Industriali, 2, 131-172.

Carneiro, P. and Heckman, J.J. (2005). The evidence on credit constraints in post-secondary schooling. Economic Journal, 112, 705-734.

Caudill, S., Ford, J. and Gropper, D. (1995). Frontier estimation and firm-specific inefficiency measures in the presence of heteroscedasticity. Journal of Business and Economic Statistics, 13 (1), 105-111.

Cazals, C., Florens, J.P. and Simar, L. (2002). Nonparametric frontier estimation: a robust approach. Journal of Econometrics, 106, 1–25.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2, 429-444.

Checchi, D., Fiorio, V. and Leonardi, M. (2013). Intergenerational persistence in educational attainment in Italy. Economic Letters, 118 (1), 229-232.

Coelli T, Rao, D.S.P. and Battese, G.E. (1998). An introduction to efficiency and productivity analysis. Boston: Kluwer Academic Publishers.

Cooper, W.W., Seiford, L.M. and Zhu, J. (2004). Handbook on data envelopment analysis. Springer: Kluwer Academic Publishers.

DesJardins, S. L., Ahlburg, D. A. and McCall, B. (2002). A temporal investigation of factors related to timely degree completion. The Journal of Higher Education, 73 (5), 555-581.

Farrel, M.J. (1957). The measurement of productive efficiency. Journal of Royal Statistical Society, 120, 253-290.

Ferrari, G. and Laureti, T. (2005). Evaluating the technical efficiency of human capital formation in the Italian University: Evidence from Florence. Statistical Methods and Applications, 14(2), 243–270.

Hadri, K. (1999). Estimation of a doubly heteroscedastic stochastic frontier cost function. Journal of Business and Economic Statistics, 17 (3), 359-363.

Kumbhakar, S.C. and Lovell, C.A.K. (2000). Stochastic Frontier Analysis, New York, Cambridge University Press.

Lassibille, G. (2011). Student progress in higher education: what we have learned from large-scale studies. The Open Education Journal, 4, 1-8.

Laureti, T. (2008). Modelling exogenous variables in human capital formation through a heteroscedastic stochastic frontier. International Advances in Economic Research, 14 (1), 76–89

Lewin, A.Y. and Lovell, C.A.K. (1990). Frontier analysis: parametric and nonparametric approaches. Journal of Econometrics, 46, 1-245.

Liu, Y. and Myers, R. (2009). Model selection in stochastic frontier analysis with an application to maize production in Kenya. Journal of Productivity Analysis, 31 (1), 33-46.

Meeusen, W. and Van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review, 18, 435-444.

Ramanathan, R. (2003). An Introduction to data envelopment analysis: a tool for performance measurement. New Delhi: Sage Publications India Private Ltd.

Smith, J. and Naylor, R. (2001). Determinants of degree performance in UK universities: A statistical analysis of the 1993 student cohort. Oxford Bulletin of Economics and Statistics, 63, 29–60.

Thanassoulis, E. (2001). Introduction to the theory and application of data envelopment analysis. Norwell, Massachusetts: Kluwer Academic Publisher.

Tinto, V. (1975). Dropout from higher education: a theoretical synthesis of recent research. Review of Educational Research, 45 (1), 89-125.

Tyagi, P., Yadav, S. P. and Singh, S. P. (2009). Relative performance of academic departments using DEA with sensitivity analysis. Evaluation and Program Planning 32, 168-177.

Wang, H. J. (2002). Heteroscedasticity and non-monotonic efficiency effects of a Stochastic Frontier Model. Journal of Productivity Analysis, 18 (3), 241-253.

Wang, H. J. (2003). A stochastic frontier analysis of financing constraints on investment: the case of financial liberalization in Taiwan. Journal of Business & Economic Statistics, 21, 406–419.

Zoghbi, A.C., Rocha, F. and Mattos, E. (2013). Education production efficiency: Evidence from Brazilian universities. Economic Modelling, 31, 94–103.