ELSEVIER
CrossMark
Available online at www.sciencedirect.com
ScienceDirect
Procedia Economics and Finance 17 (2014) 79 – 88
Procedia
Economics and Finance
www.elsevier.com/locate/procedia

Innovation and Society 2013 Conference, IES 2013

Estimating the Relative Efficiency of Secondary Schools by Stochastic Frontier Analysis

Sergio Scippacercola[a,1], Luigi D'Ambra[a]

[a] *Economics, Management, Institutions Department, Federico II University of Naples, via Cinthia, Napoli, Italia*

Abstract

This study has the aim of evaluating the efficiency of schools of secondary education. We used the first data (pilot survey) gathered from an official survey, in progress, performed by the school management of the Campania Region. The survey covers attributes regarding the environment, territorial context and economic resources. We adopted the Stochastic Frontier Analysis to estimate the efficiency and a Tobit regression model in order to discuss which factors might affect the efficiency.

Selection and peer-review under responsibility of the Organizing Committee of IES 2013.

Keywords: efficiency; secondary schools; stochastic frontier analysis; Tobit model

1. Introduction

In this period of economic crisis in Europe, the efficiency of the school system is object to attention, weighs quite heavily on the resources of the country. Still today the schools differ dramatically in quality [Hanushek (1986)]. An efficient and quality education is the basis of strategic intervention to strengthen the human capital endowment. Only a good school system can affect the cognitive skills of students, can help to increase productivity, social mobility and the full enjoyment of citizenship rights in the society. Hence the growing interest in the measurement of the efficiency level of learning of the students, of the skills acquired and of the ability to use them practically in daily life and in the workplace [Hanushek and Woessmann (2010)].

The educational process may be considered as the result of a process of production that uses a variety of inputs to determine one or many outputs [Cordero-Ferrera et al. (2008); Dolton et al. (2003); Mizala et al. (2002)] . The main approaches to the problem are the Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA) [Charnes et al. (1978); Ray (2004, 1991); Ruggiero (1996)].

This work aims at assessing the performance of Schools on the judicious management of human resources, structural and economic and educational organization. We investigate the determinants and the degree of efficiency in the utilization of resources in terms of human capital and financial aspects, by employing the SFA [Rao et al. (2005)].

[1] Corresponding author: Tel. +39-081-675010.
E-mail address: sergio.scippacercola@unina.it

Selection and peer-review under responsibility of the Organizing Committee of IES 2013.
doi:10.1016/S2212-5671(14)00881-8

In this study we have taken into consideration part of the data of a broader fact-finding investigation which the School Board of the Campania Region is carrying out.

The paper is organized as follows. The main differences between the Stochastic Frontier Analysis and the Data Envelopment Analysis are described in Section 2; in addition, we discuss here the reasons that lead to the choice of the SFA instead of the DEA. Section 3 introduces the Stochastic production frontier methodology and a *second stage* procedure, based on a Tobit regression model, that will single out the factors that could influence the efficiency. In Section 4, we show the data that are the subject of the proposed methodology. Follow the model specification (Section 5) and the discussion of the main findings (Section 6). Finally, Section 7, presents our conclusions.

2. The DEA and SFA approach

The educational process is very complex and we we can only imperfectly summarize it in a production function. In the field of education, non-parametric techniques have been mainly used, such as the Data Envelopment Analysis (DEA), while the SFA has been used in the School context [Mizala et al. (2002)] and, although more demanding in terms of assumptions, it is less sensitive to the presence of outliers and allows to make inferences on the contribution of inputs.

DEA is a *non-parametric deterministic approach* [Cooper et al. (2007)] that uses the mathematical programming to identify the efficient frontier, and it does not impose functional forms. DEA is based on the chosen inputs and outputs of entities that are named Decision Making Units (DMUs). For example, all the schools (DMUs) are compared relative to the best performing schools.

The main advantage of DEA is that it does not require a priori hypothesis about the analytical form of the production function. Indeed, DEA determines the production function by applying minimization techniques on the data. Differently by the regression analysis, the DEA is based on extremal observations, and this brings to the main disadvantage of DEA, i.e., that the frontier is sensitive to the extreme observations. Further, DEA postulates the absence of random errors and that all deviations from the frontier denote inefficiency of the DMUs.

Vice versa, the SFA, detailed in the next Section, is a *parametric approach* that hypothesizes a functional form and use the data to econometrically estimate the parameters of this function.

SFA requires functional forms on the production frontier, and assumes that units may deviate from the production frontier not only due to technical inefficiency but also measurement errors, statistical noise or to other non-systematic factors. In addition, the SFA requires strong distribution assumptions of both statistical random errors (i.e., normal distribution) and the non-negative technical inefficiency random variables (i.e., half-normal or truncated normal distribution) [Coelli et al. (2005)].

3. The Stochastic Frontier Analysis and the Tobit model

3.1. The SFA

The Stochastic Frontier Analysis searches for the production function, which represents the maximum output attainable given a certain quantity of inputs [Rao et al. (2005)].

The first stage of SFA consists in the specification and in the estimation of the stochastic frontier production function and in the estimation of technical inefficiency effects, under the assumption that these inefficiency effects are identically distributed. The SFA methodology allows functional form and the breakdown of the inefficiency error term. A production function is defined as the schedule of the maximum amount of output that can be produced from a specified set of inputs, given the existing technology.

The problem is to determine empirically the maximum potential of a production unit. The ratio of the observed value to the maximum potential output obtainable from a particular set of inputs is the technical efficiency of a production unit.

The model of the Stochastic Frontier Analysis [Rao et al. (2005)] is:

$$\ln y_i = \mathbf{x}_i'\boldsymbol{\beta} + v_i - u_i \tag{1}$$

where y_i is the output of producer i, $\mathbf{x}_i$ is a vector of inputs, $\boldsymbol{\beta}$ is a vector of $K + 1$ parameters to be estimated.

We assume that:

- $v_i \approx iid\, N(0, \sigma_v^2)$ is the noise or error term or the measure of effects independent by producer; v_i is assumed to have constant variance (homoskedasticity);
- u_i is a non-negative random variable measuring the technical *inefficiency*, iid, with $N^+(0, \sigma_u^2)$ (half-normal or $N^+(\mu, \sigma_u^2)$ normal-truncated model or exponential or gamma);
- v_i and u_i are distributed independently of each other and of the regressors.

3.2. The Technical Efficiency

We can define the Technical Efficiency (TE) as the ratio of realised output to the stochastic frontier output:

$$\ln TE_i = \ln y_i - \ln \widehat{y_i} = \ln(y_i/\widehat{y_i}) = -u_i \qquad (0 \leq TE_i \leq 1) \tag{2}$$

The parameters of stochastic frontier function are estimated by the maximum likelihood method. An estimation of stochastic frontier is facilitated by the use the reparametrization proposed by Battese and Corra (1977):

$$\sigma^2 = \sigma_v^2 + \sigma_u^2 \qquad \gamma = \frac{\sigma_u^2}{\sigma^2} \qquad (0 \leq \gamma \leq 1). \tag{3}$$

The prediction of individual technical efficiencies involves the unobservable technical inefficiency effects u_i. The best predictor of u_i is the conditional expectation of u_i, given the value of $\epsilon_i = v_i - u_i$ [Battese and Coelli (1988)].

If the parameter $\gamma = 0$ then the variance of the technical inefficiency effect is zero and so the model reduces to the traditional mean response function, a specification with parameters that can be consistently estimated using OLS.

If γ is close to one, it indicates that the deviations from the frontier are due mostly to the technical inefficiency.

When $\gamma = 1$, one-sided error component dominates the symmetric error component and the model is the deterministic production function with no noise.

3.3. The Tobit model

The *second stage* consists in the specification of a regression model for the predicted effects of the technical inefficiency. The Tobit model by [Schnedler (2005)] is an appropriate tool, since the efficiency scores are censored, and they cannot exceed 1 nor be lower than 0. The idea at the basis of the Tobit model is that it observes the variable only within bounded limits. If the value of an unobservable dependent variable lies outside the limits, we let it equals to the value at the limit. Inefficiency effects are simultaneously conditioned on several specific factors and estimated using the parameterisation with mean [Battese and Coelli (1995)]:

$$\widehat{TE}_i = \delta_0 + \mathbf{z'}_i \boldsymbol{\delta} \tag{4}$$

where $\mathbf{z}_i$ is the vector of the explanatory variables and δ_0 and $\boldsymbol{\delta}_j$ ($j = 1, 2, \ldots, J$) are respectively a parameter and a vector of J parameters to be estimated. The technical inefficiency effects u_i are frequently estimated in a first step and the determinants of inefficiency are obtained in a second-stage regression. However, this may induce both bias and inefficiency in the estimations. To assess the Technical INefficiency $\widehat{TIN}$ instead of technical efficiency $\widehat{TE}$ in order to directly assess the relationship between inefficiency and other variables, $\widehat{TIN}$ can be calculated using the following formula:

$$\widehat{TIN} = \frac{1 - \widehat{TE}}{\widehat{TE}} \tag{5}$$

and then the Tobit regression method must be applied using $\widehat{TIN}$ instead of $\widehat{TE}$. The technical inefficiency scores ($\widehat{TIN}$) assume values between 0 and infinity [Nakil (2004)]. Although the two-step approach seems reasonable, assuming that any inefficiencies that have been found can be explained by additional factors in a second stage, it contradicts the assumption made in the first stage of identically distributed inefficiency effects on the stochastic frontier.

The main hypotheses of interest of the Stochastic Frontier Analysis are:

$$H_0 : \beta_1 = \ldots = \beta_q = 0 \qquad q \leq K. \tag{6}$$

The omission of u_i is equivalent to impose the constraint specified in the null hypotheses, i.e. :

$$H_0 : \gamma = \delta_0 = \ldots = \delta_J = 0. \tag{7}$$

This indicates that the inefficiency effects in the frontier model are not present (no efficiency).

Null hypotheses of interest are tested using the generalized likelihood ratio. The generalized likelihood-ratio statistic λ is given by:

$$\lambda = -2 \ln [L(H_0)/L(H_1)] = -2 [\ln L(H_0) - \ln L(H_1)] \tag{8}$$

where $L(H_0)$ and $L(H_1)$ are the values of the likelihood function under the specifications of the null and the alternative hypotheses, H_0 and H_1 respectively. Special care must be taken when the likelihood test involves a null hypothesis that includes $\gamma = 0$.

The null hypothesis $H_0 : \gamma = 0$ specifies that the effects of the technical inefficiency are not stochastic. We reject the null hypothesis of no technical inefficiency effects given the specifications of the stochastic frontier and of the inefficiency effect model. In this case that H_0 is true, the generalized likelihood-ratio statistics, LR, has an asymptotic distribution which is a mixture of chi-square distributions [Coelli (1995); Kodde and Palm (1986)].

4. The data

The transposition of SFA methodology into the educational field is relatively simple in the theory, but it has significant difficulties when considering the definition of the production and of the factors that can affect the learning process. For example, there are factors, such as the characteristics of the teachers, the knowledge of the students and the interactions with the school Manager that are difficult to include in an empirical model.

The data available in the our study include for each shool the following groups of variables: environment variables (the demographic, social, and family context), incoming resources (structural, financial and technological endowment), the results produced at end of the school cycle (learning outcomes) and variables of opening and interactions with the territory (projects, networks, training). In this study, we selected from all the variables only the ones that are deemed the most interesting for the analysis. At the end of the survey that is currently performing by the Regional School District, there will partecipate more than a thousand of schools. But, in this work, we examined only thirty-five schools that provided coherent and valid data in the school year 2012-2013.

4.1. The Input and Output variables

The school-leaving data play an important role, because they take part into the selection mechanism for the university access. Thus, we chose as output variable the number of students who have passed the secondary school-leaving examination with a score greater than 80/100, compared with the total number of examined students. This measure has been reputed to be a good indicator of quality of the period spent in the school by the students. Other indicators, that had been used in other studies, such as the average grade achieved or the number of students repeating the last year, were not available.

The independent variables (i.e., the inputs) that we considered for the SFA are structural, financial, technological, human resources, and environmental variables. In particular, the environmental variables were used in the Tobit model.

There is a large set of input variables that could potentially explain the differences of technical efficiency among the schools in the sample.

After significance tests, the following variables have been kept on the list of the potential determinants of technical efficiency, that represent characteristics of the school and of the management/production:

- the number of teachers per 100 students (x_1), it measures the relationship between the teaching staff and all students; it is used to assess the presence of teachers with respect to all of the students;
- the number of teachers per classroom (x_2), it assesses the distribution and allocation of the teaching staff, it can be considered a measure of the intensity of the teaching provision;
- the number of students per classroom (x_3), it evaluates the level of overcrowding of the classes;
- teachers who have more than ten years of service (x_4), it assesses the presence in the school of teachers with many years of teaching experience;
- extra funds added to ordinary income received by the school (x_5). This indicator evaluates the existence of additional activities that involve the teaching staff and the student and shows the presence of initiatives and willingness to develop projects;
- the total area of the classrooms (x_6), it is a structural indicator that provides a measure of the comfortability and livability of the environment;
- the total area of the library (x_7), it is an indicator of structural and technological equipment and provides a measure of the cultural richness of the school.

4.2. The environmental variables

It is well known that the local context influences the development of the educational process in schools. Therefore, it is important to identify the main context variables that can determine the evolution of the educational process. Since all the schools in our data are in the same province, we do not take into account variables such as the unemployment rate, the presence of foreign population and the crime rate. Thus, we consider the following indicators of context:

- the number of foreign students enrolled in the school with respect to the total number of students, it provides us with an indicator of possible integration difficulties and slowdowns in the educational process (x_8);
- the number of students enrolled in the first classes in the current school year compared to the number of students enrolled at the upper classes in the previous school year. This indicator provides an evaluation of the attraction of the institution (x_9);
- the number of students who drop out of school compared to the total number of students. It allows us to assess the lack of interest in the study or the unfitness to follow the educational process for other scholastic interests (x_{10});
- the number of parents who participated in collegiate bodies with respect to the number of eligible parents. It provides the level of involvement in the educational process and the socio-cultural level of the population where the school is situated (x_{11});
- the number of students enrolled in the first classes with equal or higher rating compared to the total number of subscribers at the lower classes. It allows us to evaluate the cultural characteristics of the students who belongs to the school (x_{12}).

The main statistics of the input and output variables are shown in Table 1.

5. The model specification

The choice of the model is based on the Box-Cox transformation [Box and Cox (1964)]. We started including all variables and interactions in the model. We also included three dummy variables (Type of school: Technical Institutes or not; School Size: by number of students less than 800 or more than 799; School territorial distribution: City or Region), but they were removed after they have been determined not significant.

The choice of the functional form has been carried out under the hypothesis of a parsimonious model by likelihood ratio test and AIC criteria [Akaike (1987)]. The final model is:

$$\ln y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 \ln x_{i4} + \beta_5 \ln x_{i5} + \beta_6 \ln x_{i6} + \beta_7 \ln x_{i7} + v_i - u_i \tag{9}$$

Table 1. Statistics of the variables

Variable	Mean	Variance	Skewness	Kurtosis
y (number of students who passed the secondary school-leaving examination with a score greater than 80/100 compared to the total number of examined students)	31.74	376.82	0.28	2.55
x_1 (number of teachers per 100 students)	13.44	78.28	3.81	18.48
x_2 (number of teachers per classroom)	2.50	0.40	1.23	5.97
x_3 (number of students per classroom)	21.46	43.48	-0.15	4.86
x_4 (teachers who have more than ten years of service)	40.48	958.61	0.92	3.45
x_5 (extra funds added to ordinary income received by the school)	28487.3	1.85e+10	5.22	29.36
x_6 (total area of the classrooms)	1586.06	542094.63	-0.08	2.50
x_7 (total area of the library)	76.31	3949.63	1.62	4.82
x_8 (the number of foreign students with respect to the total students)	1.92	3.24	1.09	3.53
x_9 (number of students enrolled in the first classes compared to the number of students enrolled at the upper classes in the previous school year)	102.34	642.22	-0.48	5.27
x_{10} (number of students who drop out of school compared to the total number of students)	4.70	71.08	4.01	20.60
x_{11} (number of parents who participated in collegiate bodies	179.00	40402.00	1.33	4.30
with respect to the number of eligible parents)	179.00	40402.00	1.33	4.30
x_{12} (number of students enrolled in the first classes with equal or higher compared to eight the total number of subscribers at the lower classes)	28.17	456.25	1.05	3.56

where i refers to the i-th school, y is the number of students who passed the secondary school-leaving examination with a score greater than 80/100, x_1 is the number of teachers per 100 students, x_2 is the number of teachers per classroom, x_3 is the number of students per classroom, x_4 is the teachers who have more than ten years of service, x_5 are the extra funds added to the ordinary income received by the school, x_6 is the total area of the classrooms and x_7 is the total area of the library. Variables V and U are defined as described in the Section 3.

6. Main findings

Before using the procedure of maximum likelihood the negative skewness of the residual OLS was verified. The negative sign implies that the residuals of the sample have the correct characteristic for the implementation of the procedure of maximum likelihood.

We have analysed three models half-normal, truncated normal and exponential. Half-normal distribution for the efficiency term proved more significant than truncated normal and exponential tested models.

Thus, by using the log likelihood values and the test on $\mu = 0$, we have choose the half-normal model. In Table 2 are summarized the main results.

Table 2. Estimation Results of Frontier Production Functions with dependent variable being the number of students who passed the secondary school-leaving examination with a score greater than 80/100 compared with the total of examined students.

Input Variables/Parameters	Coefficient	Standard Error	z	$P > \|z\|$	95% Confidence Interval	
Constant	2.774614	.3040215	9.13	0.000	2.178743	3.370486
x_1 (Teachers per 100 Students)	.0511661	.007135	7.17	0.000	.0371818	.0651505
x_2 (Teachers per Class)	-.8886777	.0852517	-10.42	0.000	-1.055768	-.7215874
x_3(Students per Class)	.1153029	.0175287	6.58	0.000	.0809473	.1496585
x_4(Teacher with ten years' teaching experience	.1094727	.0643067	1.70	0.089	-.0165661	.2355115
x_5 (Extra funds)	.0489883	.036231	1.35	0.176	-.0220231	.1199998
x_6 (Surface area of the classrooms)	-.0147324	.0094023	-1.57	0.117	-.0331607	.0036959
x_7(Surface area of the library)	.0117525	.0233249	0.50	0.614	-.0339634	.0574684
σ_u	0.7091463	.0847591				
σ_v	0.2152360	.0015276				
γ	0.9154430					
Log likelihood -24.723196		$Prob > \chi^2 = 0.0000$				
Likelihood-ratio test of $\sigma_u = 0$	$\overline{\chi}^2(01) = 26.26$	$Prob \geq \overline{\chi}^2 = 0.0000$				

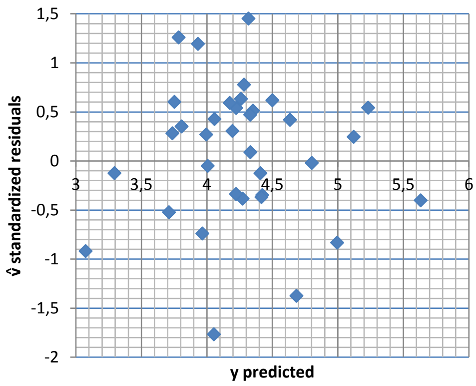

Fig. 1. Scatter plot of $\hat{v}$ standardized residuals

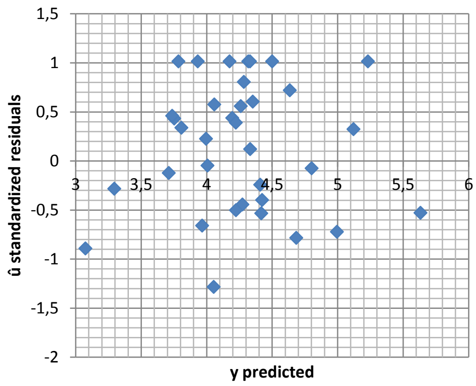

Fig. 2. Scatter plot of $\hat{u}$ standardized residuals

The test for statistical significance of the deterministic inefficiency portion of total error involves computation of a statistic γ. The test is: $H_0 : \gamma = 0; H_1 : \gamma \neq 0$. Using the most basic production function form along with a half-normal deterministic inefficiency error, γ is calculated to be 0.91. The likelihood ratio test statistic for σ_u based on a mixed χ^2 distribution was 26.26 ($p \leq \overline{\chi}^2 = 0.000$) [Kodde and Palm (1986)]. This supports rejection of the null hypothesis. Inefficiency is a significant portion of the total error, and SFA is appropriate for the analysis. The model is significant ($Prob > \chi^2 = 0.000$).

Problems with efficiency of estimation can arise when the variance of the dependent variable varies across the data. c affects standard errors, and thus determinations of significance of a given variable. Standard tests for heteroscedasticity following a linear regression are not available for frontier maximum likelihood estimation.

In cross-section models, it can appear in either of the error terms, and affect interferences concerning the production technology parameters and the error components. Thus, it can also have effects on the inefficiency estimates. If heteroskedasticity appears in the inefficiency term, the problem is more severe since both the estimates of production technology and inefficiency are biased. If v_i is heteroskedastic, only the inefficiency estimates are affected. If both error terms are heteroscedastic, the effect is not clear since the unmodeled heteroskedasticity causes biases in opposite directions. In such case, the overall bias can be small, however, one can visually inspect scatter plots of the residuals and fitted values (y) for patterns in the data. Because the distribution of u_i has been modeled to be dependent on the explanatory variables, heteroskedasticity of u_i would be expected. However, it is not expected for v_i.

Therefore, before to analyse the results, we have verified the homoskedasticity hypothesis by the visualization of the predicted values with respect to the estimated residual component of v_i (Fig. 1), we remember that the mean and variance for half normal u_i [Aigner et al. (1977)] are:

$$E(u) = \sigma_u \sqrt{\frac{2}{\pi}} \qquad Var(u) = \frac{(\pi - 2)}{\pi}\sigma_u^2 \tag{10}$$

so we compute the estimated residual component of u_i (Fig. 2). No some outliers appear to affect variance in the mid-range of predicted values, no strong pattern of heteroskedasticity is apparent.

The coefficients of the variables are positive except in the case of the x_2 variable (Teacher per Class) and x_6 (Surface area of the Classrooms). The impact of the number of teachers per 100 students on output is positive and significant.

The results (Table 1) show that the production inputs such as teachers for 100 students and the number of students per class has a positive and significant impact on the determination of the production frontier. Positive, but the presence is not significant of teachers with more than ten years teaching experience; this is important for educational outcomes in all models considered: in fact, this variable is useful for the measurement of the quality of teachers, as opposed to those measuring the quantity.

For the determination of the border any extra revenue funds, the total area of the classes and the presence of school libraries are not significant. It would seem from this scenario that exposed human resources are most crucial for the definition of the production frontier.

The teacher - class ratio has a negative impact on results, which indicates that it is a sensitive variable for improving efficiency: an increase in the number of teachers, however, the average ratio between teachers and classes are kept almost constant (if the number of classes increases proportionally to the number of students per class decreasing).

Overall, the efficiency estimates derived under the three distributional assumptions utilized in this analysis appear to be quite highly correlated. As one might expect, there is a very high degree of correlation between the estimates derived from the half-normal and truncated normal distributions, with a correlation coefficient of 0.995. Between the half-normal and exponential, there is a correlation of 0.85 and between the truncated normal and the exponential the correlation coefficient is 0.84.

All these seem to suggest that the efficiency estimates derived from the application of the stochastic frontier model are relatively robust to the distributional assumptions made, even though the estimations from the exponential distribution appear to be quite low. The robustness of the results under different distributional assumptions is even more evident when considering the rank order of efficiency estimates under the three distributions.

The Spearman rank order correlation coefficient for the ranked estimates produced under the half-normal distribution compared to those derived under the exponential assumption is 0.89. The same calculated statistic for the half-normal distribution compared to the truncated normal distribution is 0.995 and the same statistic for the truncated normal distribution versus the exponential distribution is 0.91 .

6.1. One stage model - two stage model

SFA can be performed using one-stage and two-stage variations. In a two-stage approach, SFA is used to produce the inefficiency estimates for each observation. These estimates are then regressed against proposed explanatory variables of the inefficiency, generally using the Tobit model [Kumbhakar and Lovell (2003)]. We point out that a two-stage model requires assuming that the explanatory variables of the inefficiency and the independent variables (i.e. inputs) in the production function are uncorrelated. If they are correlated, then the estimates of β, σ_u, and σ_v are biased because of the omission of the theorized inefficiency explanatory variables from the production function.

Another major problem is that the first stage requires the assumption that u_i be distributed identically and independently across all observations, yet the second stage assumes that a functional relationship exists between the inefficiency represented by u_i in the first stage and the explanatory variables. An early approach, seemingly more straightforward, included the efficiency variables and their parameters in the model as (1).

A more recent strategy, as specified by Kumbhakar et al. (1991), significantly advanced the one-stage approach so that the estimation of the efficiency and of the parameters of proposed inefficiency explanatory variables is accomplished simultaneously, thereby avoiding problems associated with the two-stage approach. Its form is as follows:

$$\ln(y_i) = f(x_i;\beta) + v_i - u_i \tag{11}$$

$$u_i \sim N^+(z_i\delta, \sigma_u^2) \tag{12}$$

Several researchers have empirically shown that this one-stage procedure leads to less biased and more efficient results. Given its econometric advantages and the goals of this analysis, the final model of the efficiency used in this analysis employs this one-stage approach. Software Stata allows to use the one-stage when a truncated-normal distribution is specified (which also covers the more specific half-normal distribution).

In this paper we have considered the two stage approach because of the low correlations among the input variables and the environmental variables. The choice of a parsimonious model is based on lower AIC .

As regards the environmental variables, by Tobit regression (Table 3) with inefficiency as dependent variable, we show that only the cultural level of the students in input is negative and slightly significant, while all the other variables are not significant.

6.2. Efficiency Analysis for different dummy variables

In the sample there are schools of various types and of different sizes and spatial distribution. In order to identify whether the technical efficiency calculated using the stochastic frontier is dependent on the above-mentioned context variables, we proceeded to perform some non-parametric tests on the technical efficiency according with the types:

Table 3. Results of Tobit Regression Model with Inefficiency as Dependent Variable.

Variables	Coefficient	Standard Error	t	$P > t$	95% Confidence Interval	
Constant	.9614183	.7791851	-1.84	0.076	-.3537952	.0188992
x_8 (foreign presence)	-.167448	.091245	-1.84	0.076	-.3537952	.0188992
x_9 (attraction for that specific institution)	.0092977	.006333	1.47	0.152	-.003636	.0222315
x_{10} (dropouts)	-.0250788	.0195147	-1.29	0.209	-.0649332	.0147756
x_{11} (parents participation)	-.000091	.0008493	-0.11	0.915	-.0018255	.0016434
x_{12} (cultural level input)	-.019536	.0081774	-2.39	0.023	-.0362365	-.0028356
Log likelihood	-46.635494	$Prob > \chi^2 = 0.0350$				

- Type of school: Technical Institutes or not;
- School Size: by number of students less than 800 or more than 799;
- School territorial distribution: City or Region.

To investigate the effects of the factors of interest, Wilcoxon tests were conducted. Table 4 presents the results of efficiency difference by three factors. All factors reveal no significant differences on the efficiency score. We conclude that the three dummy variables do not have different impacts on the efficiency.

Table 4. Two-sample Wilcoxon rank-sum (Mann-Whitney) test.

Ranksum efficiency by school type: H_0: efficiency (Technical Institute) = efficiency (Non Technical Institute) z = 1.375 P(efficiency (Technical Institute) > efficiency (Non Technical Institute)) = 0.643	$Prob > \|z\| = 0.1692$
Ranksum efficiency by School Size: H_0: efficiency (Schools with fewer than 800 students) = efficiency (Schools with more than 799 students) z = -1.078 P(efficiency (Schools with fewer than 800 students) > efficiency (Schools with more than 799 students)) = 0.643	$Prob > \|z\| = 0.2811$
Ranksum efficiency by territorial distribution: H_0: efficiency (City) = efficiency (Region) z = 0.872 P(efficiency (City) > efficiency (Region)) = 0.589	$Prob > \|z\| = 0.3833$

7. Conclusions

This study introduced a model to estimate the school efficiency using Stochastic Frontier Analysis. The estimation of the efficiency of an institution of secondary education is beneficial to improve the educational process, because it provides feedback to the school manager and points out the lacks into the educational process.

By the results of the analysis, we state that the production inputs such as, the number of teachers per 100 students and the number of students per class, have significant impact on the determination of the production frontier. Moreover, the number of teachers with more than ten years teaching experience is useful variable to measure the quality of the teachers, as opposed to those which measure the quantity.

The financial variables such as the extra revenue funds and the structural variables, as the total area of the classes and the presence of school libraries are not significant. It would seem that the human resources are the most important variables for the production frontier.

The methodology described in this work is suitable for the evaluation of the efficiency. Moreover, even with our partial data, the method and the results achieved already provide a useful interpretation of the efficiency frontier for the evaluation of schools. Indeed, the obtained efficiency estimates have been utilized to rank the schools according with the efficiency index and possible sources of inefficiency have been examined with different techniques.

The study is different from the previous ones because it adopts a two-stage model to investigate the effects of the school efficiency.

The Stochastic Frontier Analysis techniques were applied to calculate technical efficiency scores, while the Tobit regression was used to investigate the possible causes of technical efficiency for each context variable.

Acknowledgements

The research has been partially funded in accordance by the agreement between the University of Naples "Federico II" and Nidge University (Prof. Basaran). We thank the School Superintendent of Campania, and in particular Angela Orabona of the *Polo Qualità di Napoli* for providing us with the data for this study.

References

Aigner, D., Lovell, C.A., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of econometrics 6, 21–37.

Akaike, H., 1987. Factor analysis and aic. Psychometrika 52, 317–332.

Battese, G., Coelli, T., 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20, 325–332.

Battese, G.E., Coelli, T.J., 1988. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. Journal of Econometrics 38, 387 – 399.

Battese, G.E., Corra, G.S., 1977. Estimation of a production frontier model: with application to the pastoral zone of eastern australia. Australian Journal of Agricultural and Resource Economics 21, 169–179.

Box, G.E., Cox, D.R., 1964. An analysis of transformations. Journal of the Royal Statistical Society. Series B (Methodological) , 211–252.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency of decision making units. European journal of operational research 2, 429–444.

Coelli, T., 1995. Estimators and hypothesis tests for a stochastic frontier function: A monte carlo analysis. Journal of productivity analysis 6, 247–268.

Coelli, T.J., Rao, D.S.P., O'Donnell, C.J., Battese, G.E., 2005. An introduction to efficiency and productivity analysis. Springer.

Cooper, W.W., Seiford, L.M., Tone, K., 2007. Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software. Springer.

Cordero-Ferrera, J.M., Pedraja-Chaparro, F., Salinas-Jimnez, J., 2008. Measuring efficiency in education: an analysis of different approaches for incorporating non-discretionary inputs. Applied Economics 40, 1323–1339.

Dolton, P., Marcenaro, O.D., Navarro, L., 2003. The effective use of student time: a stochastic frontier production function case study. Economics of Education Review 22, 547 – 560.

Hanushek, E.A., 1986. The economics of schooling: Production and e. Journal of economic literature 24, 1141–1177.

Hanushek, E.A., Woessmann, L., 2010. The economics of international differences in educational achievement. Technical Report. National Bureau of Economic Research.

Kodde, D.A., Palm, F.C., 1986. Wald criteria for jointly testing equality and inequality restriction s. Econometrica 54, 1243–48.

Kumbhakar, S.C., Ghosh, S., McGuckin, J.T., 1991. A generalized production frontier approach for estimating determinants of inefficiency in us dairy farms. Journal of Business & Economic Statistics 9, 279–286.

Kumbhakar, S.C., Lovell, C.K., 2003. Stochastic frontier analysis. Cambridge University Press.

Mizala, A., Romaguera, P., Farren, D., 2002. The technical efficiency of schools in chile. Applied Economics 34, 1533–1552.

Nakil, S., 2004. Information technology, efficiency and productivity: Evidence from korean local governments .

Rao, D.P., O'Donnell, C.J., Battese, G.E., Coelli, T.J., 2005. An introduction to efficiency and productivity analysis. Springer.

Ray, S.C., 1991. Resource-use efficiency in public schools: A study of connecticut data. Management Science 37, 1620–1628.

Ray, S.C., 2004. Data envelopment analysis: theory and techniques for economics and operations research. Cambridge University Press.

Ruggiero, J., 1996. Efficiency of educational production: An analysis of new york school districts. The Review of Economics and Statistics , 499–509.

Schnedler, W., 2005. Likelihood estimation for censored random vectors. Econometric Reviews 24, 195–217.