A latent class stochastic frontier model for evaluating technical efficiency of Italian Universities: a Generalized Maximum Entropy Approach

Luigi Biggeri¹ Tiziana Laureti² Luca Secondi³

¹ National University Evaluation Committee, Rome and University of Florence- Department of Statistics- email: biggeri@ds.unifi.it
² University of Naples “Parthenope”- Department of Mathematics and Statistics – email: tiziana.laureti@uniparthenope.it
³ National University Evaluation Committee, Rome – email: secondi@ds.unifi.it

One of the most important issues in stochastic frontier models is to take account of the unobserved heterogeneity among production units operating in different production environments (Greene, 2005). Individual production units carry out their production in different environments characterized by various external factors which can influence their technology but are not under their control.

Therefore, production possibilities are expected to differ in a cross-section of firms, and a set of different technologies may simultaneously coexist at any given time. If that is the case, the evaluation of technical efficiency cannot be performed by considering a common technology. In other words, assuming that firms share the same technology when it is inappropriate to do so will result in misleading efficiency measurement and confusion between technological differences (heterogeneity) and individual inefficiency (Tsionas, 2002).

In the university formation process framework this issue is particularly important since the various educational processes are characterized by a stronger heterogeneity in both inputs and outputs than that observed in the “classical” production of goods and services. Bearing in mind the requirements of the neoclassical production theory we can model the university service as a production process where inputs such as capital, labour and organisational factors are used for producing teaching and research outputs. This can be regarded as a process of human capital formation where the university transforms a cultural “raw” material into a cultural “refined” material (output) through the training it delivers by using a number of inputs such as teachers, textbooks, class-rooms, etc. In this context, the student who is submitted to the cultural transformation process can be considered an input however it is important to bear in mind that the student is also a user of the educational service thus playing a more or less active role in carrying out the process. Therefore the students’ individual characteristics, i.e. gender, age, school leaving certificates, individual psychological, social and household characteristics can be considered exogenous variables, which can affect the results of the process and therefore directly influence the error term describing the inefficiency component.

So, we can not neglect the students’ characteristics without introducing a bias in the estimation of the production function of educational production process and in the measurement of technical efficiency (Ferrari and Laureti, 2004, Laureti, 2008).

In order to solve the above mentioned problems, we must model inefficiency and heterogeneity separately in the same model in order to separate the two effects. Recently various methods have been proposed that allow for heterogeneity in stochastic frontier models including random coefficient models (Tsionas, 2002; Greene, 2002, 2005), Markov switching frontier models (Tsionas and Kumbhakar, 2004) and threshold stochastic frontier model (Yélou, Larue, and Tran, 2010).
In this context the latent class models are particularly appealing since they assume that heterogeneity is induced by a set of observable exogenous variables which determines a latent sorting of the production units (Greene, 2002, 2005; Orea and Kumbhakar, 2004).

The aim of this paper is to suggest a latent class approach for evaluating the technical efficiency of the Italian universities from 2003 to 2006, by using the Generalized Maximum Entropy Method (Golan et al. 1996, Golan, 2008). The advantage of these models is that they can consider technological heterogeneity since the production units are classified into a finite number of classes \(Q\) with the same technological parameters in each class without using exogenous sample separation information.

In the case of the Italian Universities it would be possible to classify the faculties (or degree courses) into different groups by using \textit{a priori} sample separation information (e.g. standard cost per student) however this type of classification would not be shared and above all may not be able to show the actual differences among these faculties. On the other hand the latent class model is particularly suitable for obtaining an appropriate classification which is another aim of this paper.

In order to carry out the efficiency analysis we have assumed that there is a latent sorting of the observations in the data set into unobserved \(Q\) latent classes. The probabilities of class membership is a function of a vector of faculty-specific variables, such as the territorial location (Northern, Central, Southern and Islands) of the universities, the size of the university or the size of the faculty itself, measured as the number of students enrolled, and whether it is a state or a private university.

By using the GME approach we can overcome the ill posed situation which can easily arise when the unknown number of classes \(Q\) is high and in this case the total number of parameters to be estimated, concerning each class model and class probability function, will be greater than the number of observations.

Therefore, by using the suggested approach we can distinguish between cross (faculty) heterogeneity unrelated to inefficiency and inefficiency itself.

Moreover, we have assumed that the inefficiency term is a function of a set of exogenous explanatory variables, specified by the students’ characteristics, and of the time variable (year) as well in order to consider a time-varying specification of technical inefficiency by following the approach suggested by Battese and Coelli (1995). In this context we can avoid the specific distributional assumption (i.e the non-negative truncated normal distribution) with the GME approach.

It is worth noting that our analysis of the efficiency of first level degrees is based on data coming from the newly established Italian University Student Archive which includes the students’ elementary data and their academic progress. As the first researchers to use data from this archive in Italy, we have constructed a specific panel data base, considering all the faculties in Italy as production units with reference to four cohorts of students enrolled in first level degree courses (undergraduate three year course programs) in the academic years 2003/04, 2004/05, 2005/06 and 2006/07. For each cohort, we followed the students’ progress up to the academic year 2009/10.

Besides the information regarding the students’ characteristics – coming from the above mentioned Italian University Student Archive – we have included data concerning the university facilities and equipment such as the quality of teaching, classroom availability, laboratories, libraries, location of the offered services, facilities for teaching and research activities, management and administration as well as the working conditions for the technical and administrative personnel collected yearly by the National University Evaluation Committee. Most variables are observed at faculty level while other variables are observed at university level only.
In order to carry out the evaluation of technical efficiency we firstly analysed the entire educational process, where the freshmen are transformed into graduates, and then we focused on a sub-process, that is the first year of the degree courses since the first stage of the process is especially critical for obtaining a degree. For modelling these processes we used the Cobb–Douglas (CD) production function since the assumptions behind its use are plausible in view of the theoretical model, which describes the education formation process in the university system. In particular, the unitary elasticity of substitution employed in the CD production function, which is a borderline case between a high and low degree of substitutability, can be suitable for describing the faculties’ attitude in carrying out the production process (Laureti, 2008).

The output measure in the first process we modelled is represented by the Full Credit Equivalent Students in the first year, measured in terms of students who obtained the required 60 credits; while for the second process we considered the number of graduates who graduated within institutional time (three years). The first output can be considered as a “semi-manufactured” product obtained by the educational process and at the same time as indicators of the growth in human capital while the number of graduates is the “finished product” of the process.

Inputs in the production function include: the number of freshmen; full and/or associate professors and/or researchers, seats in lecture halls; seats in laboratories; books in the library; etc.

As potential exogenous variables influencing the inefficiency term we included the percentage of females; the percentage of freshmen with the best results of secondary school leaving examination (90-100/100); the percentage of freshmen with lyceum diplomas; the percentage of freshmen aged over 22; the percentage of freshmen from other areas (outside the University region)

Finally we considered variables concerning the size and territorial location (Northern, Central, Southern and Islands) of the universities and an economic indicator (GDP per capita) of the province in which the university is located as determinants of the latent class probabilities.

References


