EERI

Economics and Econometrics Research Institute

# An application of statistical interference in DEA models: An analysis of public owned university departments' efficiency

George E. Halkos, Nickolaos G. Tzeremes, Stavros A. Kourtzidis

EERI Research Paper Series No 17/2010

ISSN: 2031-4892

**EERI**
**Economics and Econometrics Research Institute**
Avenue de Beaulieu
1160 Brussels
Belgium

Tel: +322 299 3523
Fax: +322 299 3523
www.eeri.eu

# An application of statistical interference in DEA models: An analysis of public owned university departments' efficiency

by

**George Emm. Halkos[1], Nickolaos G. Tzeremes, Stavros A. Kourtzidis**

*Department of Economics, University of Thessaly, Korai 43, 38333, Volos, Greece*

**Abstract**

This paper uses Data Envelopment Analysis (DEA) model formulations in order to determine the performance levels of 16 departments of the University of Thessaly. Particularly, the constant returns to scale (CRS) and variable returns to scale (VRS) models have been applied alongside with bootstrap techniques in order to determine accurate performance measurements of the 16 departments. The study illustrates how the recent developments in efficiency analysis and statistical inference can be applied when evaluating institutional performance issues. The paper provides the efficient departments and the target values which need to be adopted from the inefficient departments in order to operate in the most productive scale size (MPSS). Moreover it provides bias corrected estimates alongside with their confidence intervals. The analysis indicates that there are strong inefficiencies among the departments, emphasizing the misallocation of resources or/and inefficient application of departments policy developments.

***Keywords***: *University efficiency; DEA; Bootstrap techniques; Kernel density estimation.*

[1] **Corresponding author:** George Emm. Halkos, Director of Postgraduate Studies, Director of the Operations Research Laboratory, Deputy Head, Department of Economics, University of Thessaly, Korai 43, 38333, Volos, Greece. http://www.halkos.gr. Email: halkos@uth.gr. Tel.: 0030 24210 74920. FAX: 0030 24210 74772

## 1. Introduction

Several studies have tried to measure institutions efficiency facing several problems. According to Johnes and Johnes (1993) the basic problem in measuring the efficiency of higher education institutes, is how to aggregate the heterogeneous inputs and outputs, in the absence of market prices. In order to measure the efficiency, price indicators (PIs) were developed, each of which measures the input or the output of a homogenous set of products. The most commonly used PI in the case of universities is the number of publications (Moed et al. 1984; Harris 1988; Johnes 1990). However, Glass et al. (2006) argue that PIs focus only on one variable, without being capable of including the multiple inputs and outputs that are necessary in higher education institutes. Also, PIs fail to aggregate multiple inputs and outputs because they cannot provide objective weights, which could help to succeed it.

An alternative way of assessing the efficiency is the econometric approach, which defines a production function and assumes that deviations from it are composed of two terms, inefficiency and randomness. Inefficiency follows an asymmetric half-normal distribution and random error term. This represents randomness and includes the exogenous factors as well as the econometric error, which follows the normal distribution. Basic features of econometric approach are the assumption of production technology and the strict parametric nature (Worthington 2001). The econometric approach leads to the development of the stochastic frontier approach (SFA) and has been applied by several researchers to evaluate the performance higher education institutes (Verry and Layard 1975; Graves et al. 1982; Hirsch et al. 1984; Johnes 1988, 1997; Cohn et al. 1989; De Groot et al. 1991; Glass et al. 1995; Johnes 1996; Izadi et al. 2002).

The last approach for measuring the efficiency is the mathematical approach and its basic tool is the Data Envelopment Analysis (DEA). DEA measures the relative efficiency of an institute and objectivity is the big advantage provided, while it calculates the best possible weights for aggregating multiple inputs and outputs. In opposition to the previous approach, DEA does not require determining any functional form, uses the least possible restrictions and only requires the convexity hypothesis (Banker et al. 1986). DEA offers freedom in the choice of the variables, which can be measured in different units. An important advantage is the calculation of shadow prices and slack variables (Stiakakis and Fouliras 2009). Specifically, shadow prices can answer which efficient decision making unit (DMU) is a benchmark for the inefficient under assessment DMU (Johnes and Johnes 1993).

However, DEA assumes that deviations from the efficient frontier are the result of inefficiency. This could lead to overstatement or understatement of the results while there are not any assumptions for the exogenous factors or measurement error. Also, its non-stochastic nature does not allow confidence intervals to be calculated. However this has been tackled by Atkinson and Wilson (1995) and Simar and Wilson (1998, 2000) who use a bootstrap methodology, which applies Monte Carlo techniques in order to approach the distribution and to calculate confidence intervals.

Our study, by applying those advances of statistical inference in DEA models, measures the efficiency of the University of Thessaly departments. Moreover, the paper demonstrates how bootstrap techniques can be applied into institution efficiency measurement and can improve the results obtained by the straightforward application of DEA techniques.

The paper is organized as follows. Section 2 reviews the existing relative literature whereas section 3 presents the various variables used in the formulation of the proposed models. In section 4 the techniques adopted both in theoretical and mathematical formulations are presented. Section 5 discusses the empirical findings of our study. The final section concludes the paper commenting on the derived results and the implied policy implications.

## 2. Literature Review

Lindsay (1976) argues that a public principal does not measure the value of a product by its market price, but from its characteristics. Public authority can evaluate only the most obvious characteristics and this implies that economic resources are directed towards them. On the contrary, private enterprises evaluate all the characteristics of a product. Sisk (1981) applied Lindsay's theory to academic institutions, however he used only one input and one output. Ahn et al. (1988) extended Sisk's research by adding multiple inputs and outputs and they used a DEA model to check the hypothesis that public universities are more efficient than private universities. They used capital and labour as inputs and teaching and research as outputs, measured by the number of full time equivalents separately for undergraduate and postgraduate teaching and the amount of federal grants and contracts respectively.

Tomkins and Green (1988) measured the efficiency of twenty accounting departments of English universities by running six DEA models. Particular interest presents the inclusion of research postgraduate students, as well as the number of publications as a measure for research and the number of academic staff as a measure for teaching. Johnes and Johnes (1993) divide publications into categories: papers in academic journals, letters in academic journals, articles in professional journals, articles in popular journals, authored books, edited books, published official reports and contributions to edited works. Moreover, an article was identified if it was published in a journal which was included in Diamond's list (Diamond 1989).

Madden et al. (1997) include as inputs the number of teaching aide staff and administrative staff except from academic staff. Also, they argue that the proper measure of teaching is the number of graduating students because it incorporates the quality into teaching under the assumption that more graduating students implies higher teaching quality. Flegg et al. (2004) and Johnes and Yu (2008) support that the number of students must be included as an input together with capital and labor.

All researches mentioned so far measure the efficiency among similar departments of different universities. Sinuany-Stern et al. (1994) were the first who measure the efficiency among departments of the same university and specifically at Ben-Gurion University. The same direction is followed by some researches like King (1997), Arcelus and Coleman (1997) and Sarrico and Dyson (2000).

Colbert et al. (2000) measured the efficiency of twenty four MBA programs based on the pleasure of students and academic staff. Ng and Li (2000) applied the methodology of Li and Ng (1995) at eighty four Chinese universities. They divided efficiency in to technical, allocative and reallocative efficiency. Avrikan (2001) and Abbott and Doucouliagos (2003) decomposed technical efficiency in pure technical efficiency and scale efficiency.

Wong and Beasley (1990) used proportions for restricting weight flexibility in order to improve DEA model. This technique was used in academic education by Beasley (1990, 1995), Athanassopoulos and Shale (1997) and Carrico et al. (1997).

Finally, there are some other researches that combine DEA with multiple criteria decision making (MCDM). One of the first attempts to combine these methods was made by Golany (1988) who proposed the use of an interactive multiobjective linear-programming procedure, which is supposed to help the decision maker to set the real effective production points for a given set of inputs. Halme et al. (1999) included the preferences of decision maker in the traditional DEA model and Korhonen et al. (2001) applied this method to academic institutes. These two methods were combined by Caballero et al. (2004) in a three-phase procedure.

**3. Data**

As a public institution, university uses multiple inputs to produce multiple outputs. In this study we use as inputs the number of academic staff, the number of auxiliary staff (teaching aide staff, technical and administrative staff), the number of students (undergraduates, postgraduates, doctorate students) and total income (governmental funding).

The number of academic staff is used almost in all bibliography (Tomkins and Green 1988; Johnes and Johnes 1993) and it is constituted only by faculty members. There are four ranks of faculty members (professor, associate professor, assistant professor and lecturer), so we assigned weights to each rank in order to construct a proper aggregated measure of academic staff (Madden et al. 1997). Weights were assigned based on the assumption that a professor is expected to produce more research work than a lecturer. Thus, professors were assigned with 1, associate professors with 0.75, assistant professors with 0.5 and lecturers with 0.25. These weights were chosen so the distance between two ranks to be 1/4=0.25.

The second input, has been also used by Arcelus and Coleman (1997) and Madden et al. (1997), is the auxiliary staff, which is constituted by teaching aide, technical and administrative staff. This input is used under the assumption that teaching, administrative and technical duties have a negative influence on the research of academic staff because they limit their available time for research. Therefore, higher auxiliary staff means higher expected research (Johnes 1988). We assigned weights to each category of auxiliary staff as before. Teaching aide staff was assigned with 1, while technical and administrative staff was assigned with 0.5.

The third input is the number of students, which according to Flegg et al. (2004) and Johnes and Yu (2008) can be included as an input. Like academic staff, there are three student ranks (undergraduates, postgraduates and doctorate students) so we assigned weights to each one. Thus, doctorate students were assigned with 1, postgraduates with 0.666 and undergraduates with 0.333.

The fourth input is the total income which is used by the vast majority of the bibliography in many forms (Tomkins and Green 1988; Beasley 1990; Sinuany-Stern et al. 1994; Athanassopoulos and Shale 1997). Sometimes income can be found as total income or total grants and other times can be found as income from research or from other sources.

As it is widely accepted by whole bibliography, the outputs that are produced by a university are teaching and research. Some researches measure teaching by the hours a professor teaches, which is a convenient approach because it's easy for a researcher to collect this data. However, this measure does not include the quality of teaching. A simple way to include quality into teaching is to measure the number of graduating students. The assumption is that higher number of graduating students means higher quality of teaching (Madden et al. 1997). Again, we assigned weights to each student rank. Thus, postgraduates were assigned with 1 and undergraduates with 0.5.

Academic research is the most controversial output. Although it is widely accepted as an output, it can be measured in various ways. The two main ways to measure research is the income from research (Ahn et al. 1988, Beasley 1990, 1995; Flegg et al. 2004) and the number of publications (Zinuany-Stern et al. 1993; Johnes and Johnes 1993; Johnes and Yu 2008). In the first case, the argument is that more significant research will attract more income. However, this is an indirect measurement, while the number of publications is a direct measurement of academic research and we prefer to use it in our research.

A critical question is how many journals will be used in the research. The inclusion of too few journals might bias the result in favour of departments which produce general research against the departments which produce specialized research. On the contrary, the inclusion of too many journals means that an article in an infamous journal has the same value of with an article in a famous journal (Johnes 1988). Many researches have used only the articles published in the most reputable journals, but these researches refer to British universities in most of the cases, whereas academic staff tends to publish in widely recognized journals (Johnes 1988). According to Harris (1988), Australian academics, with a few exceptions, tend to publish in less recognized journals. This proposition stands for Greek academics too. Thus, we followed Harris' research and we included all articles in refereed journals.

An important element that we took care is the number of authors in an article. So, if the under evaluation author is the first, second or third author of the article, then this article is included in a category that receives higher weight than the category which refers to articles with four or more authors.

Thus, in academic research the following categories with their weights were included. Articles in foreign journals (author 1st - 3rd) were assigned with 1, articles in Greek journals (author 1st - 3rd) with 0.86, books, monographs and chapters in books were considered of the same value and were assigned with 0.71, articles in conferences (author 1st - 3rd) with 0.57, articles in foreign journals (with 4 or more authors) were assigned with 0.43, articles in Greek journals (with 4 or more authors) were assigned with 0.29 and articles in conferences (with 4 or more authors) were assigned with 0.14. Along with articles in conferences we measure discussion papers as Madden et al. (1997) did in their research.

Dyson et al. (2001) raised some issues that must be examined in a DEA model. In the present paper, we will deal with two of the raised issues, the homogeneity of Decision Making Units (DMU) and the number of variables. In order to be homogenous, DMUs must have a similar range of activities and produce similar outputs. The activities of all the departments are teaching and research. Teaching is measured by the number of graduating students and research is measured by the number of publications which are both similar for all the departments. However, it would be useful if we could include other forms of research such as laboratorial research (however it is difficult to be measured). Additionally, DMUs must use a similar range of inputs, as is true in our case. Our inputs are the number of academic staff, the number of auxiliary staff, the number of students and the total income, which are all similar for every department. The last assumption for homogeneity is all DMUs to operate in a similar environment, which is true because all departments operate under the legal framework which is the same for all the Greek universities. Moreover, departments operate under the framework of the same university.

According to Dyson et al. (2001) the number of DMUs must be at least, $2 \times m \times s$ where m is the number of inputs and s the number of outputs. In our case

$2 \times 4 \times 2 = 16$ is equal with the number of DMUs under evaluation indicating a "proper" number of inputs/outputs used.

The data for the number of academic and auxiliary staff, the number of undergraduate and postgraduate students, the number of graduating students and total income were collected from the annual internal report of Evaluation Quality Unit of the University of Thessaly, from the Office of Academic Affairs and from the departments' secretariats and they refer to the period 2009-2010. The data for the publication were provided from the departments' official websites and from annual internal report of the Evaluation Quality Unit.

## 4. Methodology

### 4.1 Efficiency measurement

Efficiency analysis was dated back to the work of Debreu (1951), Koopmans (1951) and Farrell (1957) who were the first to measure empirically the efficiency of production units. Following the notation by Simar and Wilson (2008) we can imply that the process of production is constrained by the production set $\Psi$ which is the set of physically attainable points $(x, y)$ so that:

$$\Psi = \left\{ (x, y) \in \Re_{+}^{N+M} \middle| x \ \ can \ \ produce \ \ y \right\} \qquad (1)$$

where $x \in \Re_{+}^{N}$ is the input vector and $y \in \Re_{+}^{M}$ is the output vector. In that respect the efficient boundary of $\Psi$ is the locus of optimal production plans. This boundary is called the production frontier and can be expressed as:

$$\partial\Psi = \left\{ (x, y) \in \Psi \middle| (\theta x, y) \notin \Psi, \forall 0 < \theta < 1, (x, \lambda y) \notin \Psi, \forall \lambda > 1 \right\} \qquad (2).$$

According to Daraio and Simar (2007) the locus of optimal production plans can be either input or output oriented. In the input oriented framework the input requirement set and its efficient boundary aims to reduce the input amounts keeping the present output levels. In contrast the output oriented framework seeks to maximize the output levels keeping the present input levels. The choice between input and output orientation is based on whether the decision maker controls most the inputs or the outputs. This study uses the assumption of output orientation since public universities have greater control of the research produced and the graduates (outputs). In contrast with the inputs which the amounts of are directly controlled by the Greek

Ministry of Education, Lifelong Learning and Religious Affairs and indirectly by the Universities' departments. Therefore, the production set $\Psi$ is characterized by output feasibility sets defined for all $x \in \Re_{+}^{N}$ as:

$$Y(x) = \left\{ y \in \Re_{+}^{M} \middle| (x, y) \in \Psi \right\} \quad (3),$$

and the output oriented efficiency boundary $\partial Y(x)$ is defined for a given $x \in \Re_{+}^{N}$ as:

$$\partial Y(x) = \left\{ y \middle| y \in Y(x), \lambda y \notin Y(x), \forall \lambda > 1 \right\} \quad (4),$$

and the Debreu-Farrell output measure of efficiency for a production unit located at $(x, y) \in \Re_{+}^{N+M}$ is:

$$\lambda(x, y) = \sup \left\{ \lambda \middle| (x, \lambda y) \in \Psi \right\} \quad (5).$$

The DEA estimator was first operationalized as linear programming estimators by Charnes et al. (1978) assuming the free disposability and the convexity of the production set $\Psi$. It involves measurement of the for a given unit $(x, y)$ relative to the convex hull of $X_n = \left\{ \left( x_i, y_i \right), i = 1, ..., n \right\}$ and it assumes constant returns to scale (CRS):

$$\hat{\lambda}_{CRS}(x, y) = \sup \left\{ \begin{array}{l} \lambda \middle| \lambda y \le \sum_{i=1}^{n} \gamma_i y_i ; x \ge \sum_{i=1}^{n} \gamma_i x_i \quad for \quad (\gamma_1, ..., \gamma_n) \\ such \quad that \quad \gamma_i \ge 0, i = 1, ..., n \end{array} \right\} \quad (6).$$

Later, Banker et al. (1984) developed a DEA estimator allowing for variable returns to scale (VRS) as:

$$\hat{\lambda}_{VRS}(x, y) = \sup \left\{ \begin{array}{l} \lambda \middle| \lambda y \le \sum_{i=1}^{n} \gamma_i y_i ; x \ge \sum_{i=1}^{n} \gamma_i x_i \quad for \quad (\gamma_1, ..., \gamma_n) \\ such \quad that \quad \sum_{i=1}^{n} \gamma_i = 1; \ \gamma_i \ge 0, i = 1, ..., n \end{array} \right\} \quad (7).$$

**4.2 Efficiency bias correction and confidence internals construction**

DEA estimators are biased by construction and thus biased correction techniques need to be adopted for the improvement of the efficiency scores obtained (Halkos and Tzeremes, 2010). Following Simar and Wilson (1998, 2000) we perform the bootstrap procedure for the DEA estimators in order to obtain biased corrected results. The bootstrap procedure is a data-based simulation method for statistical inference (Daraio and Simar 2007, p.52). Some of its main applications[2] are the correction for the bias and construction of confidence intervals of the efficiency estimators (Simar and Wilson, 1998; 2000), applications to Malmquist indices (Simar and Wilson, 1999), statistical procedures for comparing the efficiency means of several groups (Simar and Wilson 2008), test procedures to assess returns to scale (Simar and Wilson, 2002) and criterion for bandwidth selection (Simar and Wilson, 2002; 2008).

Suppose we want to investigate sampling distribution of an estimator $\hat{\theta}$ of an unknown parameter $\theta$, where $\mathrm{P}$ is a statistical model (data generating process, or DGP) and $\hat{\theta} = \hat{\theta}(X)$ is a statistical function of $X$. Therefore by the proposed procedure we try to evaluate the sampling distribution of $\hat{\theta}(X)$, to evaluate the bias, the standard deviation of $\hat{\theta}(X)$ and to create confidence intervals of any parameter $\theta$. By generating data sets from a consistent estimator $\hat{\mathrm{P}}$ of $\mathrm{P}$ from data $X : \hat{\mathrm{P}} = \mathrm{P}\left(\hat{\Psi}, \hat{f}(.,.)\right)$, we denote $X^* = \left\{\left(X_i^*, Y_i^*\right), i = 1,\dots,n\right\}$ the data set generated from $\hat{\mathrm{P}}$.

The estimators of the corresponding quantities of $\hat{\Psi}$ and $\hat{\delta}(x, y)$ (in terms of the output-distance function as in Shephard, 1970), output-distance function) can be defined by the pseudo sample corresponding to the quantities $\hat{\Psi}^*$ and $\hat{\delta}^*(x, y)$. Using the methodology proposed by Simar and Wilson (1998, 2000) the available bootstrap distribution of $\hat{\delta}^*(x, y)$ will be almost the same with the original unknown sampling distribution of the estimator of interest $\hat{\delta}(x, y)$ and therefore it can be expressed as:

$$\left(\hat{\delta}^*(x, y) - \hat{\delta}(x, y)\right)\Big|\hat{\mathrm{P}} \overset{approx.}{\sim} \left(\hat{\delta}(x, y) - \delta(x, y)\right)\Big|\mathrm{P} \tag{8}$$

[2] See Halkos and Tzeremes (2010) for application of bootstrap techniques on SMEs data.

A bias corrected estimator can then by defined as:

$$\tilde{\delta}(x,y)=\hat{\delta}(x,y)-\widehat{bias}\left(\hat{\delta}(x,y)\right)=2\hat{\delta}(x,y)-\frac{1}{B}\sum_{b=1}^{B}\hat{\delta}_b^*(x,y) \tag{9}$$

Finally, the bootstrap confidence interval for $\delta(x,y)$ can be defined as:

$$\left[\hat{\delta}(x,y)-\hat{\alpha}_{1-a/2},\hat{\delta}(x,y)-\hat{\alpha}_{a/2}\right] \tag{10}$$

### 4.3 Testing for returns to scale

In order to choose between the adoption of the results obtained by the CCR (Charnes et al., 1978) and BCC (Banker et al., 1984) models in terms of the consistency of our results obtained we adopt the method introduced by Simar and Wilson (2002). Therefore, we compute the DEA efficiency scores under the CRS and VRS assumption and by using the bootstrap algorithm described previously we test for the CRS results against the VRS results obtained such as:

$$H_o:\Psi^{\vartheta}\ \ is\ \ CRS\ \ against\ \ H_1:\Psi^{\vartheta}\ \ is\ \ VRS \tag{11}$$

The test statistic can be computed as:

$$T(X_n)=\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\theta}crs,n(X_i,Y_i)}{\hat{\theta}vrs,n(X_i,Y_i)} \tag{12}$$

Then the p-value of the null hypotheses can be approximated by the proportion of bootstrap samples as:

$$p-value=\sum_{b=1}^{B}\frac{I\left(T^{*,b}\leq T_{obs}\right)}{B} \tag{13}$$

where B is 2000 bootstrap replications, I is the indicator function and $T^{*,b}$ is the bootstrap samples and original observed values are denoted by $T_{obs}$.

## 5. Empirical Results

Firstly we test for the existence of constant or variables returns to scale (equations 11-13) and by approximating the p-value by using the bootstrap algorithm described previously we obtained for this test a p-value of 0.98> 0.05 (with B=2000) hence, we cannot reject the null hypothesis of constant returns to scales and thus the CCR model need to be adopted in our analysis[3]. Table 1 reports the results obtained under the hypothesis of constant returns to scale (however, the VRS estimators are very similar to the CRS estimators). As can be realised the departments of primary education, medical school, veterinary science, physical education & sport science and the department of economics are reported to be efficient (efficiency score =1). Whereas, the lowest performances are reported for the departments of special education (0.5574) and the department of computer & communication engineering (0.646). In addition the department of biochemistry & biotechnology (0.9587) and the department of ichthyology & aquatic environment (0.9378) are reported to have high efficiency scores. When we apply the bootstrap algorithm on the efficiency scores obtained we calculate the biased corrected efficiency scores (CRS BC) along side with the estimated bias ($\widehat{Bias}$) and its standard deviation ($\hat{\sigma}$). As can be realized under the bias correction the efficiency scores have changed significantly however the departments with lowest performance are reported to be the same, these are the departments of special education (0.5574) and computer &communication engineering (0.646).

However the departments with highest performances are reported to have small changes on their efficiency scores (0.7 to 0.79). The department of agriculture crop, production & rural environment, the department of physical education & sport science, the department of primary education, the faculty of veterinary science and the medical school are reported to have the highest efficiency scores. But a closer look is needed on the lower (LB) and upper (UB) bounds before any conclusions can be made. Indeed the department of Economics has winder bounds (0.7 to 0.99) indicating that the biased efficiency scores can have higher values compared to the other university departments. Similarly the departments of biochemistry & biotechnology, the medical school, primary education, ichthyology & aquatic environment, veterinary science and the department of physical education and sport science have greater ranges of biased corrected efficiency scores. This variation indicates the different resource allocation and research policies among the universities departments implying greater variability in their estimated efficiencies scores.

---

[3] The results under the VRS assumption are available upon request.

**Table 1 about here**

Figure 1 presents the density estimates of the original and the biased corrected efficiency estimates (CRS) alongside with the lower and upper bounds of the efficiency scores. For the calculation of the density estimates we have used the “normal reference rule-of-thumb” approach bandwidth selection (Silverman 1986) and a second order Gaussian kernel. It appears that the original CRS are leptokurtic and almost identical with the upper bound of the biased corrected efficiency scores whereas the bias corrected efficiency scores appear to be leptokurtic and quite similar with lower bounds estimates. The leptokurtic distributions indicate that there is a rapid fall-off in the density as we move away from the mean. Furthermore, the pickedness of the distribution suggests a clustering around the mean with rapid fall around it. The density estimates appear to support graphically the previous findings which indicate that among the departments in the University of Thessaly there are different resource allocation policies and inefficiencies in the application of University’s general development policy. In addition it appears that the outputs used (research and graduates) are being part of different policy perspectives among the university’s departments.

**Figure 1 about here**

Following Banker (1984) we use the optimal values of $\sum_{i=1}^{n} \gamma_i^*$ which are given by the efficient departments in order to calculate the most productive scale size (MPSS) of the inefficient departments. Table 2 provides the scale sizes that departments should operate in order to be efficient. For instance, the department of Agriculture Crop, Production and Rural Environment in order to operate at its MPSS needs to increase the research and graduates’ levels by 42%. The benchmarks (or the reference set) for the department of Agriculture Crop, Production and Rural Environment are given by the department of Primary Education and the department of Physical Education and Sport Science. It seems difficult to compare these three departments to its thematic and scientific nature however the two reference sets are more closely in terms of the amounts of inputs/outputs to the department of Agriculture Crop, Production and Rural Environment than the other departments and therefore they show (by providing coefficients $\gamma_i^*$) how outputs can be increased in order to make the department under evaluation efficient.

Furthermore, Table 2 provides the relation between the proportional change of inputs and the resulting proportional change in outputs (returns to scales- RTS). As such constant returns to scale arise when a department produces $n$ per cent increase in output by an $n$ per cent rise in all inputs. However if output rises by a larger percentage than inputs, there are increasing return to scales (IRS). Whereas, if output rises by a smaller percentage than inputs, there are decreasing returns to scale (DRS). As can be realized only the department of Urban Planning and Regional Development and the department of Computer & Communication Engineering report DRS.

**Table 2 about here**

## 5. Conclusions

This paper applies an efficiency analysis in all the departments of University of Thessaly. By applying inferential approach on DEA efficiency scores the paper measures the efficiency of 16 university departments. The majority of the existing studies similar to ours (Sinuany-Stern et al. 1994; King 1997; Arcelus and Coleman 1997; Sarrico and Dyson 2000) evaluate the performance of university departments however it is the first time (to our knowledge) that bootstrap techniques are used in DEA formulation measuring university departments' performance. Furthermore, the bootstrap techniques have provided consistency to the original biased CRS results.

Moreover, by applying the inferential approach and bootstrapped procedures we derived the general conclusion that there are strong inefficiencies among the departments, indicating misallocation of resources or/and inefficient application of departments policy developments. Additionally, our paper provides input and output target values for policy implications and evaluation among the departments of the University of Thessaly. Finally, this study provides evidence of how the advances and recent developments in efficiency analysis can be applied for an effective evaluation of performance issues in public owned universities overcoming traditional DEA related problems.

**Acknowledgements**

We would like to thank the Vice Rector (Economic and Student Affairs, Public and International Relations) of the University of Thessaly Prof. Michel Zouboulakis for the provision of data which are part of the annual quality evaluation report given to the Ministry of Education, Lifelong Learning and Religious Affairs. In addition we would like to thank all the departments' head administrative staff for the provision of vital information. Any remaining errors are solely the authors' responsibility.

**References**

Abbott M, Doucouliagos C (2003) The efficiency of Australian universities: a data envelopment analysis. Econ Edu Rev 22:89-97.

Ahn T, Charnes A, Cooper WW (1988) Some statistical and DEA evaluations of relative efficiencies of public and private institutions of higher learning. Socio Econ Plan Sci 22:259-269.

Arcelus FJ, Coleman DF (1997) An efficiency review of university departments. Int J Syst Sci 28: 721-729.

Athanassopoulos A, Shale E (1997) Assessing the comparative efficiency of higher education institutions in the UK by means of data envelopment analysis. Edu Econ 5: 117-133.

Atkinson SE, Wilson PW (1995) Comparing mean efficiency and productivity scores from small samples: A bootstrap methodology. J Prod Anal 6: 137-152.

Avrikan N (2001) Investigating technical and scale efficiencies of Australian Universities through data envelopment analysis. Socio Econ Plan Sci 35: 57-80.

Banker RD, Conrad RF, Strauss RP (1986) A comparative application of data envelopment analysis and translog methods: An illustrative study of hospital production. Manage Sci 32: 30-44.

Banker RD (1984) Estimating Most Productive Scale Size Using Data Envelopment Analysis," Eur J Oper Res 17:35-44.

Banker RD, Charnes A, Cooper WW (1984) Some Models for Estimating Technical and scale inefficiencies in DEA. Manage Sci 32: 1613-1627.

Beasley JE (1990) Comparing university departments. OMEGA-Int J Manage S 18:171-183.

Beasley JE (1995) Determining teaching and research efficiencies. J Oper Res Soc 46:441-452.

Caballero R, Galache T, Gomez T, Molina J, Torrico A (2004) Budgetary allocations and efficiency in the human resources policy of a university following multiple criteria. Econ Edu Rev 23:67-74.

Carrico CS, Hogan SM, Dyson RG, Athanassopoulos AD (1997) Data envelopment analysis and university selection. J Oper Res Soc 48:1163-1177.

Charnes A, Cooper WW, Rhodes E (1978) Measuring the Efficiency of Decision Making Units. Eur J Oper Res 2: 429-444.

Cohn E, Rhine S, Santos M (1989) Institutions of higher education as multi-product firms: Economies of scale and scope. Rev Econ Stat 71:284-290.

Colbert A, Levary RR, Shaner MC (2000) Determining the relative efficiency of MBA programs using DEA. Eur J Oper Res 125: 656-669.

Daraio C, Simar L (2007) Advanced robust and nonparametric methods in efficiency analysis. Springer Science, New York.

De Groot H, McMahon W, Volkwein F (1991) The cost structure of American research universities. Rev Econ Stat 73: 424-431.
Debreu G (1951) The Coefficient of Resource Utilization. Econometrica 19: 273-292.
Diamond A (1989) The core journal of economics. Curr Contents 21: 4-11.
Farrell MJ (1957) The measurement of the Productive Efficiency. J R Stat Soc A Stat CXX 3: 253-290.
Flegg AT, Allen DO, Field K, Thurlow TW (2004) Measuring the efficiency of British universities: A multi-period data envelopment analysis. Edu Econ 12:231-249.
Glass JC, McCallion G, McKillop DG, Rasaratnam S, Stringer KS (2006) Implications of variant efficiency measures for policy evaluations in UK higher education. Socio Econ Plan Sci 40:119-142.
Glass JC, McKillop DG, Hyndman N (1995) Efficiency in the provision of university teaching and research: An empirical analysis of UK universities. J Appl Econom 10:61-72.
Golany B (1988) An interactive MOLP procedure for the extension of DEA to effectiveness analysis. J Oper Res Soc 39:725-734.
Graves P, Marchand R, Thompson R (1982) Economics departmental rankings: Research incentives, constraints and efficiency. Am Econ Rev 72:1131-1141.
Halkos GE, Tzeremes NG. (2010) The effect of foreign ownership on SMEs performance: An efficiency analysis perspective. J Prod Anal. doi: 10.1007/s11123-010-0174-2.
Halme M, Joro T, Korhonen P, Salo S, Wallenius J (1999) A value efficiency approach to incorporating preference information in data envelopment analysis. Manage Sci 45:103-115.
Harris GT (1988) Research output in Australian university economics departments, 1974-83, Aust Econ Pap 27:102-110.
Hirsch B, Austin R, Brooks J, Moore B (1984) Economics departmental rankings: comment. Am Econ Rev 74:822-826.
Izadi H, Johnes G, Oskrochi R, Crouchley R (2002) Stahastic frontier estimation of a CES cost function: the case of higher education in Britain. Econ Edu Rev 21:63-71.
Johnes G (1988) Determinants of research output in economics departments in British universities. Res Pol 17:171-178.
Johnes G (1990) Measures of research output: University departments of economics in the UK, 1984-8. Econ J 100:556-560.
Johnes G. (1997). Costs and industrial structure in contemporary British higher education. Econ J 107:727-737.
Johnes G, Johnes J (1993) Measuring the research performance of UK economics departments: An application of data envelopment analysis. Oxford Econ Pap 45:332-347.
Johnes J (1996) Performance assessment in higher education in Britain. Eur J Oper Res 89:18-33.
Johnes J, Yu L (2008) Measuring the research performance of Chinese higher education institutions using data envelopment analysis. China Econ Rev 19:679-696.
King WD (1997) Input and output substitution in higher education. Econ Lett 57:107-111.
Koopmans TC (1951) An Analysis of Production as an Efficient Combination of Activities. In Koopmans TC (ed) Activity Analysis of Production and Allocation, Cowles Commission for Research in Economics, Monograph No. 13, Wiley, New York.

Korhonen P, Tainio R, Wallenius J (2001) Value efficiency analysis of academic research. Eur J Oper Res 130:121-132.
Li SK, Ng YC (1995) Measuring the productive efficiency of a group of firms. Int Adv Econ Res 1: 377-390.
Lindsay C (1976) A theory of government enterprise. J Polit Econ 84:1061-1077.
Madden G, Savage S, Kemp S (1997) Measuring public sector efficiency: A study of economics departments at Australian Universities. Edu Econ 5:153-168.
Moed HF, Burger WJM, Frankfort JG, Van Raan AFJ (1985) The use of bibliometric data for the measurment of university research performance. Res Pol 14:131-149.
Ng YC, Li SK (2000) Measuring the research performance of Chinese higher education institutions: An application of data envelopment analysis. Edu Econ 8:139-156.
Sarrico CS, Dyson RG (2000) Using DEA for planning in UK universities-An institutional perspective. J Oper Res Soc 51:789-800.
Shephard RW (1970) Theory of Cost and Production Function. Princeton University Press, Princeton, NJ.
Simar L, Wilson P (2008) Statistical interference in nonparametric frontier models: recent developments and perspectives. In: Fried H, Lovell CAK, Schmidt S (eds) The measurement of productive efficiency and productivity change. Oxford University Press, New York.
Simar L, Wilson PW (1998) Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. Manage Sci 44:49–61.
Simar L, Wilson PW (1999) Estimating and Bootstrapping Malmquist Indices. Eur J Oper Res 115:459-471.
Simar L, Wilson PW (2000) A general methodology for bootstrapping in non-parametric frontier models. J Appl Stat 27:779-802.
Simar L, Wilson PW (2002) Nonparametric tests of returns to scale. Eur J Oper Res 139:115-132.
Sinuany-Stern Z, Mehrez A, Barboy A (1994) Academic departments efficiency via dea. Comput Oper Res 21:543-556.
Sisk D (1981) A theory of government enterprise: University Ph. D. Production. Public Choice 37:357-363.
Stiakakis Emm, Fouliras P (2009) The impact of environmental practices on firms' efficiency: the case of ICT-producing sectors. Oper Res Int J 9:311–328.
Tomkins C, Green R (1988) An experiment in the use of data envelopment analysis for evaluating the efficiency of UK university departments of accounting, Financ Accountability Manage 4:147-164.
Verry DW, Layard PRG (1975) Cost functions for university teaching and research. Ecn J 85:55-74.
Wong Y-HB, Beasley JE (1990) Restricting weight flexibility in data envelopment analysis. J Oper Res Soc 41:829-835.
Worthington A (2001) An empirical survey of frontier efficiency measurement techniques in education. Educ Econ 9:245-268.

Table 1: Estimated efficiency scores, estimated bias and estimated bias' standard deviations.

| a/a | **Departments** | **CRS** | **CRS (BC)** | $\hat{Bias}$ | $\hat{\sigma}$ | **LB** | **UB** |
|---|---|---|---|---|---|---|---|
| 1 | Mechanical Engineering | 0.7012 | 0.6273 | 0.0739 | 0.0019 | 0.5842 | 0.6952 |
| 2 | Urban Planning and Regional Development | 0.8898 | 0.7531 | 0.1367 | 0.0068 | 0.7049 | 0.8815 |
| 3 | Civil Engineering | 0.7303 | 0.6103 | 0.1200 | 0.0048 | 0.5795 | 0.7233 |
| 4 | Architecture | 0.7423 | 0.5823 | 0.1600 | 0.0136 | 0.5443 | 0.7366 |
| 5 | Computer & Communication Engineering | 0.6460 | 0.5555 | 0.0905 | 0.0032 | 0.5156 | 0.6415 |
| 6 | Primary Education | 1.0000 | 0.7761 | 0.2239 | 0.0266 | 0.7325 | 0.9923 |
| 7 | Preschool Education | 0.6921 | 0.5984 | 0.0937 | 0.0030 | 0.5594 | 0.6863 |
| 8 | Special Education | 0.5574 | 0.4880 | 0.0694 | 0.0013 | 0.4619 | 0.5525 |
| 9 | History, Archaeology and Social Anthropology | 0.8605 | 0.7441 | 0.1164 | 0.0043 | 0.6998 | 0.8537 |
| 10 | Agriculture Crop, Production and Rural Environment | 0.8937 | 0.7985 | 0.0952 | 0.0032 | 0.7458 | 0.8857 |
| 11 | Ichthyology and Aquatic Environment | 0.9378 | 0.7166 | 0.2212 | 0.0347 | 0.6569 | 0.9303 |
| 12 | Medical School | 1.0000 | 0.7612 | 0.2388 | 0.0408 | 0.7017 | 0.9915 |
| 13 | Veterinary Science | 1.0000 | 0.7752 | 0.2248 | 0.0293 | 0.7242 | 0.9918 |
| 14 | Biochemistry and Biotechnology | 0.9587 | 0.7336 | 0.2251 | 0.0369 | 0.6705 | 0.9525 |
| 15 | Physical Education and Sport Science | 1.0000 | 0.7850 | 0.2150 | 0.0227 | 0.7487 | 0.9913 |
| 16 | Economics | 1.0000 | 0.7504 | 0.2496 | 0.0400 | 0.7004 | 0.9900 |

Figure 1: Kernel density functions of CRS efficiency estimates using Gaussian Kernel and the appropriate bandwidth (normal reference rule-of-thumb).

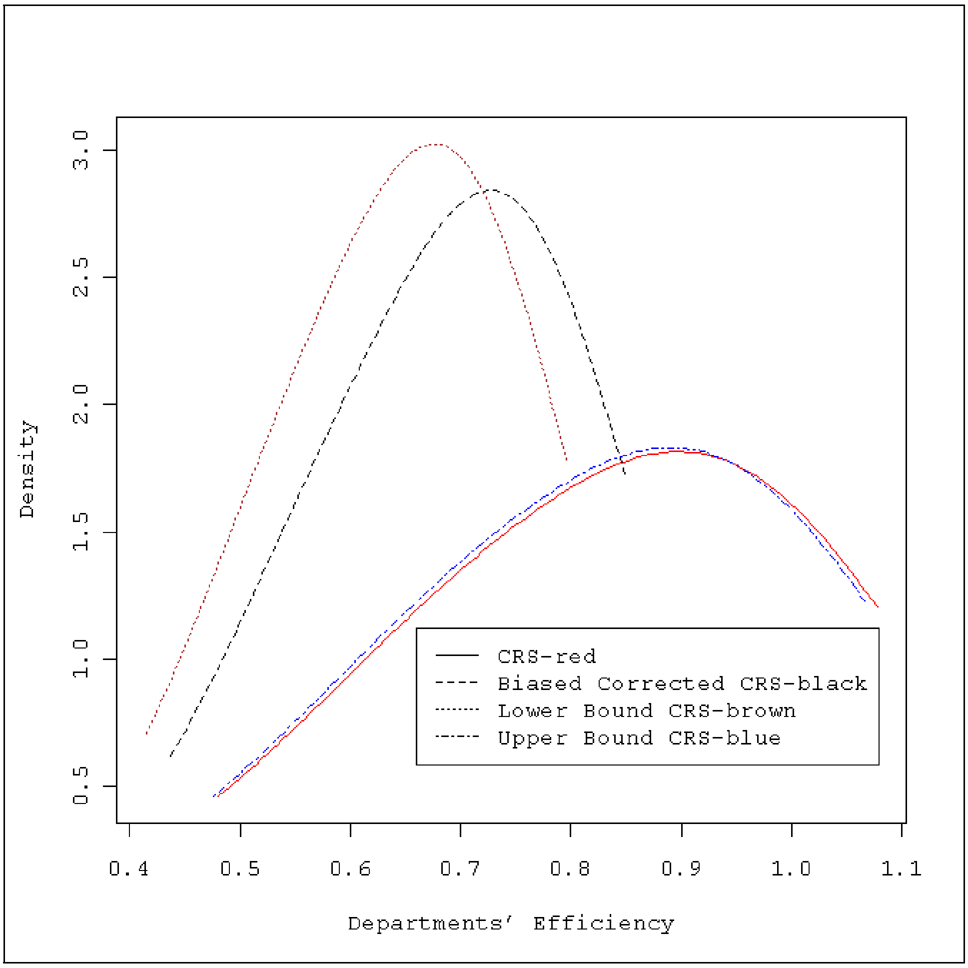

Table 2: Scale efficient targets of the departments

| a/a | DEPARTMENTS | Efficient Output Target (%) | | | |
|---|---|---|---|---|---|
| | | Research | Graduates | Benchmarks | RTS |
| 1 | Mechanical Engineering | 49.5735 | 49.5735 | 6,15 | Increasing |
| 2 | Urban Planning and Regional Development | 1.5983 | 1.5983 | 6,15 | Decreasing |
| 3 | Civil Engineering | 47.2307 | 47.2307 | 6,15 | Increasing |
| 4 | Architecture | 71.7512 | 71.7512 | 6,15 | Increasing |
| 5 | Computer & Communication Engineering | 28.8698 | 28.8698 | 6,15,16 | Decreasing |
| 6 | Primary Education | 0.0000 | 0.0000 | | Constant |
| 7 | Preschool Education | 75.2959 | 43.2909 | 6,16 | Increasing |
| 8 | Special Education | 52.6766 | 52.6766 | 6,15,16 | Increasing |
| 9 | History, Archaeology and Social Anthropology | 61.8666 | 36.5209 | 6,16 | Increasing |
| 10 | Agriculture Crop, Production and Rural Environment | 41.9540 | 41.9540 | 6,15 | Increasing |
| 11 | Ichthyology and Aquatic Environment | 51.3282 | 51.3282 | 6,15 | Increasing |
| 12 | Medical School | 0.0000 | 0.0000 | | Constant |
| 13 | Veterinary Science | 0.0000 | 0.0000 | | Constant |
| 14 | Biochemistry and Biotechnology | 68.3485 | 68.3485 | 12,15,16 | Increasing |
| 15 | Physical Education and Sport Science | 0.0000 | 0.0000 | | Constant |
| 16 | Economics | 0.0000 | 0.0000 | | Constant |