Journal of Economics and Econometrics Vol. 55, No.2, 2012 pp. 1-24

ISSN 2032-9652 E-ISSN 2032-9660

Measuring Public Owned University Departments' Efficiency: A Bootstrapped DEA Approach

GEORGE E. HALKOS†, NICKOLAOS G. TZEREMES*, STAVROS A. KOURTZIDIS‡

ABSTRACT

This article uses Data Envelopment Analysis (DEA) in order to determine the performance levels of 16 departments of a public owned university. 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 estimates. The study illustrates how the recent developments in efficiency analysis and statistical inference can be applied when evaluating institutional performance issues. The results reveal the existence of misallocation of resources or/and inefficient application of departments' policy development.

JEL Classification: C60, C67, I20, I23.
Keywords: Departments' efficiency, Data Envelopment Analysis, bootstrap techniques, Kernel density estimation.

† Department of Economics, University of Thessaly. Correspondence address: Korai 43, 38333, Volos, Greece. Tel.: +30 24210 74920. E-mail: halkos@econ.uth.gr.

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

Journal of Economics and Econometrics Vol. 55, No. 2, 2012 pp. 1-24

1 INTRODUCTION

The increasing demand for evaluation of public entities, programs and policies worldwide is a result of governmental desire for accountability. Governments demand the public organizations to operate efficiently and achieve their targets consuming the least possible resources. Universities are complex public organizations which consume multiple inputs to produce multiple outputs. As such entities, universities are in the scope of governments for utilization and allocation of resources and achieving a more efficient operation. In order to monitor and accomplish their goals, governments need proper evaluations. Generally, evaluation of educational and other pubic programs may prove powerful tools for a government who wishes to promote social and public interests. For this purpose, various approaches have been developed in order to evaluate the academic efficiency.

The primary attempts to measure academic efficiency involve performance indicators (PIs), each of which measures the input or the output of a homogeneous set of products. The most commonly used PI in the case of universities is the number of publications (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.

An alternative approach is the econometric approach which led to the development of the stochastic frontier approach (SFA) and has been applied by several researchers in order to evaluate the performance of higher education institutes (Graves, Marchand and Thompson 1982; Hirsch et al., 1984; Johnes 1988, 1997; Cohn, Rhine and Santos 1989; De Groot, McMahon and Volkwein 1991; Johnes 1996; Izadi et al., 2002; Johnes and Johnes 2009). SFA is a parametric approach where statistical inference can be used. On the contrary, as a parametric approach requires the determination of a functional form.

Another way to measure efficiency is the mathematical approach and its basic tool is the Data Envelopment Analysis (DEA). According to Bougnol and Dula (2006), DEA is a suitable tool for assessing the performance in higher education. The attractive aspect of DEA is the incorporation of multiple inputs and outputs measured in different units. Also, DEA does not require specification of a functional form;

however statistical inference cannot be used and the method is sensitive to extreme values (Johnes and Johnes 2009).

Furthermore, 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 no assumptions regarding the exogenous factors or the measurement error. Also, its non-stochastic nature does not allow confidence intervals to be calculated. However the latter has been tackled by Simar and Wilson (1998, 2000) who use a bootstrap methodology in order to approach the distribution and to calculate confidence intervals.

The DEA approach has been used for higher education institutes in many countries around the world such as Australia (Madden, Savage and Kemp 1997; Avrikan 2001; Abbott and Doucouliagos 2003; Carrington, Coelli and Rao 2005), China (Ng and Li 2000; Johnes and Yu 2008), Finland (Raty 2002), Germany (Fandel 2007), India (Tyagi, Yadav and Singh 2009), Israel (Sinuany-Stern, Mehrez and Barboy 1994), Italy (Agasisti and Johnes 2010), the United Kingdom (Athanassopoulos and Shale 1997; Flegg et al., 2004; Johnes 2006a, 2006b), and Taiwan (Kao and Hung 2008).

Our study, by applying the above advances of statistical inference in DEA models, measures the departments' efficiency of a state owned Greek university, the University of Thessaly. Moreover, the article demonstrates how bootstrap techniques can be applied into institution efficiency measurement and thus to obtain bias corrected efficiency estimates and confidence intervals, in contrast with the straightforward applications of DEA techniques.

The article 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 article commenting on the derived results and the implied policy implications.

2 LITERATURE REVIEW

The vast majority of the literature examines the efficiency across universities either on institutional or on departmental level. The most notable studies are Tomkins and Green (1988), Johnes and Johnes

(1993, 2009), Athanassopoulos and Shale (1997), Madden, Savage and Kemp (1997), Colbert, Levary and Shaner (2000), Avrikan (2001), Abbott and Doucouliagos (2003), Flegg et al. (2004), Carrington, Coelli and Rao (2005), Johnes (2006) and Johnes and Yu (2008).

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) divided 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.

Madden, Savage and Kemp (1997) included as inputs the number of auxiliary staff and administrative staff except from academic staff. Flegg et al. (2004) and Johnes and Yu (2008), support that the number of students must be included as an input along with capital and labour. Colbert, Levary and Shaner (2000) conduct a slightly different study from other authors. They measure the relative efficiency of the top 24 MBA programs in USA along with 3 foreign MBA programs, using three output sets, one for student satisfaction, one for student's employer satisfaction and one for both.

All researches mentioned so far measure the efficiency among similar departments of different universities. Sinuany-Stern, Mehrez and Barboy (1994) were the first who measured the efficiency among departments of the same university and specifically at Ben-Gurion University. The same direction is followed by other researches such as King (1997), Arcelus and Coleman (1997), Sarrico and Dyson (2000), Tauer, Fried and Fry (2007), Kao and Hung (2008) and Tyagi, Yadav and Singh (2009). The comparison of departments with different subjects within a university is challenged by a number of authors as it is controversial whether the departments can be considered as homogenous. Although, the benefits of DEA as a supplement tool for the decision maker in resource allocation and utilization within the university, render DEA a necessary tool overcoming the possible drawbacks (Sinuany-Stern, Mehrez and Barboy 1994). Furthermore, in the present article we focus on resource utilization among departments and not on academic performance, which according to Kao and Hung

(2008) surpasses the problem of different subjects among the departments. Additionally, Tyagi, Yadav and Singh (2009) argue that departments inside a university may be considered as homogenous because they operate in similar activities and are willing to achieve similar goals.

As already noted, one of the drawbacks of the basic DEA technique is that it provides no indication whether these simple scores vary statistically significant. Bootstrap techniques have been developed in order to overcome this problem and they are used to estimate 95% confidence intervals for each DMU (Simar and Wilson, 1998, 2000). These techniques are applied to DEA estimators which are biased by construction and eliminate this bias. Bootstrap techniques have been used in higher education by Johnes (2006a, 2006b).

3 DATA

The most widely used inputs across the literature are student, staff and capital inputs (Johnes and Yu 2008). This study uses as inputs the number of academic staff, the number of auxiliary staff (teaching aide, technical and administrative), the number of students (undergraduates, postgraduates, and doctorate) and total income (governmental funding). Ideally, we would consider each category as a separate input, e.g. three inputs for students, as postgraduates are more resource intense than undergraduates and doctorate students are more resource intense than the other two (Carrington, Coelli and Rao 2005). Then we would impose additional constraints at the model in order to ensure that postgraduates are assigned with a greater weight than undergraduates and doctorate students with a greater weight than the other two (Beasley 1990, 1995; Athanassopoulos and Shale 1997).

If we follow this approach then our model consists of eleven inputs and five outputs. Given that the DMUs are only sixteen this would render the DEA model infeasible. Some authors like Kao and Hung (2008) and Tyagi, Yadav and Singh (2009) pre-assign weights in order to compose aggregate measures and make the model more concise. We follow the later approach and we pre-assign weights to three inputs, academic staff, auxiliary staff and number of students and two outputs, graduates and research.

One may argue that given pre-assign arbitrary weights may compromise the final results. In order to avoid this problem, we perform various robustness checks by solving a number of DEA models altering each time the attributed weights and we test if the change in resulting efficiencies is statistically significant by applying a Mann-Whitney test. The results show that the choice of the weights does not alter the results, so our arbitrary choice is not affecting the calculated efficiencies.

The first input is the number of academic staff which is used commonly in the existing literature (Tomkins and Green 1988; Johnes and Johnes 1993; Flegg et al., 2004; Tyagi, Yadav and Singh, 2009) and it is constituted only by faculty members. There are four ranks of faculty members (professors, associate professors, assistant professors and lecturers), so we pre-assigned weights to each rank in order to construct a proper aggregated measure of academic staff (Madden, Savage and Kemp, 1997). Weights are pre-assigned based on the assumption that a professor is expected to produce more research work than a lecturer (Carrington, Coelli and Rao, 2005). Thus, professors are assigned with 1, associate professors with 0.75, assistant professors with 0.5 and lecturers with 0.25. These weights are chosen so the distance between two ranks is 1/4=0.25.

The second input, also used by Arcelus and Coleman (1997), Madden, Savage and Kemp (1997), Flegg et al. (2004) and Tyagi, Yadav and Singh (2009), 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 have an outcome in limiting 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 is 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. In contrast with other studies, total number of students is preferred from full-time equivalents as the required data is unavailable (Agasisti and Johnes, 2010). As we already noted, there are three student ranks according to their resource intensity (undergraduates, postgraduates and doctorate students). Thus, doctorate students are assigned with 1, postgraduates with 0.666 and undergraduates with 0.333.

The fourth input is the total income from research which is used by the vast majority of the literature in many forms (Tomkins and Green 1988; Beasley 1990; Sinuany-Stern, Mehrez and Barboy 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 in the literature, the outputs that are produced by a university are teaching, research and services (Avrikan 2001). The drawback is the difficulty to find the appropriate data information. In this article we include only teaching and research as outputs and we exclude services from our model. A simple way to quantify teaching is to measure the number of graduating students. The hypothesis is that higher number of graduating students is related to higher quality of teaching (Madden, Savage and Kemp 1997). Once more, we assign weights to each student rank. Thus, postgraduates are 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 core ways to measure research is the income from research (Ahn, Charnes and Cooper 1988; Beasley 1990, Flegg et al., 2004) and the number of publications (Sinuany-Stern, Mehrez and Barboy 1993; Johnes and Johnes 1993; Johnes and Yu 2008). This study chooses the number of publications as a proxy for the research because the income from research does not reflect academic research (papers, conferences etc.) in Greek Universities but income from other research activities. This leads us to treat "income from research" as any other income and use it as an input, while number of publications is used as an output.

A critical question is how many journals will be used in the research. The inclusion of a very small number of 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 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 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.

According to Carrington, Coelli and Rao (2005), Worthington and Lee (2008) and Tyagi, Yadav and Singh (2009), "weighted publications" is the most suitable measure of research. Thus, articles in foreign journals are assigned with 1; articles in Greek journals with 0.75; books, monographs and chapters in books are considered of the same value and are assigned with 0.50; and articles in conferences with 0.25. Along with articles in conferences we measure discussion papers in the same category (Madden, Savage and Kemp 1997).

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 2008-2009. The data for the publication were provided from the departments' official websites and from annual internal report of the Evaluation Quality Unit.

Descriptive statistics for all variables appear in table 1. It is notable that the standard deviations are high for all variables which imply large inequalities among the departments. This is an indication of how important is for the university to manage its inputs-outputs.

Table 1: Descriptive statistics

Variable	Mean	Median	St.Dev.	Minimum	Maximum
Academic staff	16.19	13.50	11.90	9.00	60.25
Auxiliary staff	30.84	29.00	13.79	8.50	65.50
Number of students	251.40	223.30	143.20	114.60	729.70
Income from research	1112712.00	644173.00	1139139.00	139113.00	4245165.00
Number of graduates	306.00	163.00	520.00	42.00	2209.00
Number of publications	49.63	51.00	22.44	17.50	106.00

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 Ψ which is the set of physically attainable points (x,y) so that:

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

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 Ψ is the locus of optimal production plans. This boundary is called the production frontier and can be expressed as:

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

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 Ψ is characterized by output feasibility sets defined for all $x \in \Re_{+}^{N}$ as:

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

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

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

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

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

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

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

Later, Banker, Charnes and Cooper (1984) developed a DEA estimator allowing for variable returns to scale (VRS) as:

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

4.2 Efficiency bias correction and confidence internals construction

Following Simar and Wilson (1998, 2000) we perform the bootstrap procedure for the DEA estimators in order to obtain biased corrected results (see Appendix for computational details). The bootstrap procedure is a data-based simulation method for statistical inference (Daraio and Simar 2007, p.52). Some of its main applications[1] 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).

---

[1] The essence of bootstrapping efficiency scores has been highlighted by several authors. For further applications of the bootstrap technique see Zelenyuk and Zheka (2006), Simar and Zelenyuk (2007) and Halkos and Tzeremes (2010).

The bootstrap bias estimate for the original DEA estimator $\hat{\lambda}_{DEA}(x,y)$ can be calculated as:

$$\widehat{BIAS}_B\left(\hat{\lambda}_{DEA}(x,y)\right) = B^{-1}\sum_{b=1}^{B}\hat{\lambda}^{*}_{DEA,b}(x,y) - \hat{\lambda}_{DEA}(x,y) \tag{8}$$

Furthermore, $\hat{\lambda}^{*}_{DEA,b}(x,y)$ are the bootstrap values and B is the number of bootstrap replications. Then a biased corrected estimator of $\lambda(x,y)$ can be calculated as:

$$\begin{aligned}\hat{\hat{\lambda}}_{DEA}(x,y) &= \hat{\lambda}_{DEA}(x,y) - \widehat{BIAS}_B\left(\hat{\lambda}_{DEA}(x,y)\right) \\ &= 2\,\hat{\lambda}_{DEA}(x,y) - B^{-1}\sum_{b=1}^{B}\hat{\lambda}^{*}_{DEA,b}(x,y)\end{aligned} \tag{9}$$

However, according to Simar and Wilson (2008) this bias correction can create an additional noise and the sample variance of the bootstrap values $\hat{\lambda}^{*}_{DEA,b}(x,y)$ need to be calculated. The calculation of the variance of the bootstrap values is illustrated below:

$$\hat{\sigma}^2 = B^{-1}\sum_{b=1}^{B}\left[\hat{\lambda}^{*}_{DEA,b}(x,y) - B^{-1}\sum_{b=1}^{B}\hat{\lambda}^{*}_{DEA,b}(x,y)\right]^2 \tag{10}$$

In addition we need to avoid the bias correction illustrated in (9) unless:

$$\frac{\left|\widehat{BIAS}_B(\hat{\lambda}_{DEA}(x,y))\right|}{\hat{\sigma}} > \frac{1}{\sqrt{3}} \tag{11}$$

By expressing the output oriented efficiency in terms of the Shephard (1970) output distance function we can construct bootstrap confidence intervals for $\hat{\delta}_{DEA}(x,y)$ as:

$$\left[\hat{\delta}_{DEA}(x,y) - \hat{\alpha}_{1-a/2}, \hat{\delta}_{DEA}(x,y) - \hat{\alpha}_{a/2}\right] \tag{12}$$

4.3 A bootstrap test for choosing CCR or BCC model

In order to choose between the adoption of the results obtained by the CCR (Charnes, Cooper and Rhodes 1978) and BCC (Banker, Charnes and Cooper, 1984) models in terms of the consistency of our results obtained we adopt the test 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 (see Appendix for details) we test for the CRS results against the VRS results as follows:

$$H_o : \Psi^{\vartheta} \quad is \quad CRS \quad against \quad H_1 : \Psi^{\vartheta} \quad is \quad VRS \tag{13}$$

The test statistic can be computed as:

$$T\left(X_n\right) = \frac{1}{n}\sum_{i=1}^{n} \frac{\overset{\wedge}{\lambda}\, crs, n\left(X_i, Y_i\right)}{\overset{\wedge}{\lambda}\, vrs, n\left(X_i, Y_i\right)} \tag{14}$$

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

$$p - value = \sum_{b=1}^{B} \frac{I\left(T^{*,b} \leq T_{obs}\right)}{B} \tag{15}$$

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

5 EMPIRICAL RESULTS

First we test for the existence of constant or variables returns to scale (equations 13-15) and approximate the p-value by using the bootstrap algorithm described previously. For this test we obtained 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[2]. Table 2 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

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

the departments of Primary Education, Medical School, Veterinary Science, Physical Education & Sport Science and Economics are reported to be efficient (efficiency score =1). Similarly the lowest performances are reported for the Departments of Special Education (0.558) and of Computer & Communication Engineering (0.637). In addition the departments of Biochemistry & Biotechnology (0.939) and of Ichthyology & Aquatic Environment (0.925) 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) alongside with the estimated bias (Bias) and its standard deviation (std). As can be realized under the bias correction the efficiency scores have changed significantly although the departments with lowest performance are reported to be the same. These are the departments of Special Education (0.49) and Computer & Communication Engineering (0.549).

The biased corrected results indicate that the departments of Primary Education, Medical School, Veterinary Science, Physical Education and Sport Science and Economics 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 departments of Economics and Medical School have winder bounds compared to the other departments indicating that the biased efficiency scores may have higher values compared to the other university departments. Similarly the departments of Primary Education, Veterinary Science, Physical Education and Sport Science, Biochemistry and Biotechnology and Ichthyology and Aquatic Environment 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 efficiency scores.

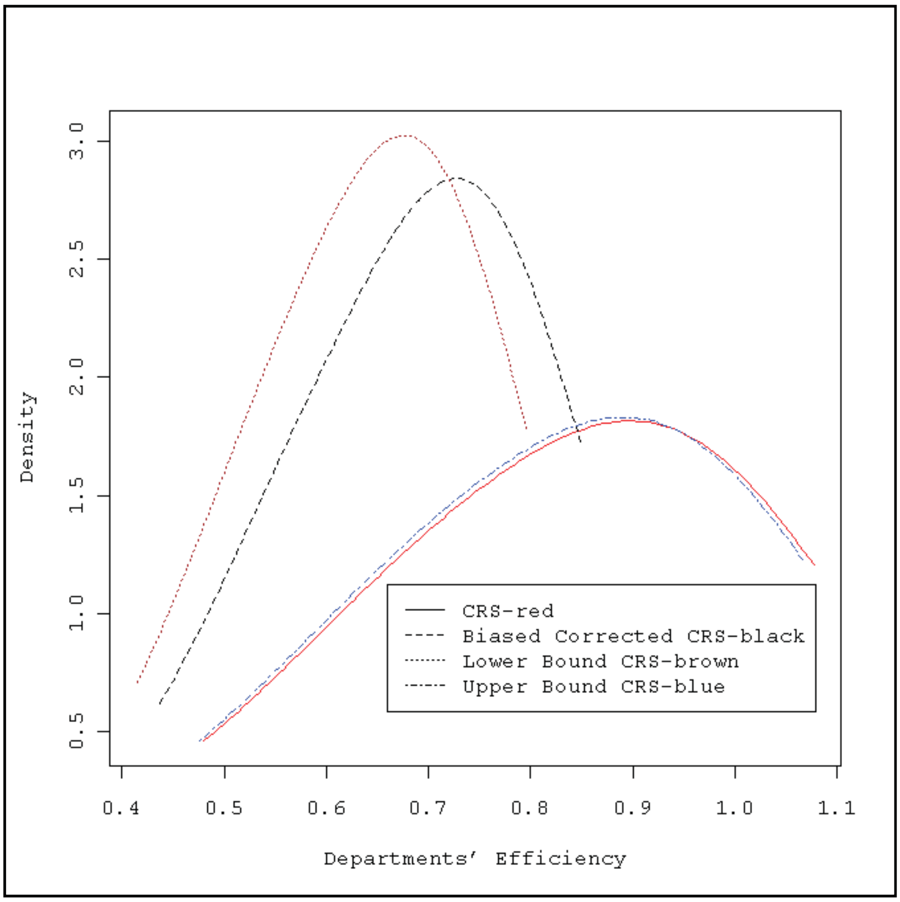
Figure 1 presents the density estimates of the original and the biased corrected efficiency estimates (CRS) alongside with the lower and upper bounds of the efficiency scores. For the calculation of the density estimates we have used the "normal reference rule-of-thumb" approach bandwidth selection (Silverman 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 peakedness 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.

Table 2: Estimated efficiency scores, estimated bias and estimated bias' standard deviations

a/a	Departments	CRS	CRS (BC)	Bias	std	LB	UB
1	Mechanical Engineering	0.701	0.631	0.071	0.002	0.587	0.696
2	Urban Planning & Regional Developm.	0.870	0.750	0.119	0.006	0.696	0.861
3	Civil Engineering	0.730	0.608	0.122	0.005	0.581	0.724
4	Architecture	0.739	0.574	0.165	0.015	0.536	0.732
5	Computer & Communication Engineer.	0.637	0.549	0.088	0.003	0.508	0.632
6	Primary Education	1.000	0.770	0.230	0.025	0.737	0.990
7	Preschool Education	0.692	0.598	0.094	0.003	0.563	0.684
8	Special Education	0.558	0.490	0.067	0.001	0.466	0.552
9	History, Archaeology & Social Anthrop.	0.861	0.745	0.115	0.004	0.701	0.854
10	Agriculture, Crop & Rural Environment	0.899	0.804	0.095	0.003	0.755	0.892
11	Ichthyology and Aquatic Environment	0.925	0.692	0.233	0.036	0.645	0.916
12	Medical School	1.000	0.748	0.252	0.042	0.697	0.992
13	Veterinary Science	1.000	0.752	0.248	0.039	0.706	0.991
14	Biochemistry and Biotechnology	0.939	0.698	0.241	0.040	0.652	0.931
15	Physical Education and Sport Science	1.000	0.794	0.206	0.017	0.763	0.992
16	Economics	1.000	0.749	0.251	0.042	0.700	0.993

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 3 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 the graduates' levels by 68.7%.

Figure 1: Kernel density functions of CRS efficiency estimates using Gaussian Kernel and the approach bandwidth



The benchmarks (or the reference set) for this department are given by the departments of Primary Education and 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 other departments within the university and therefore they show (by providing coefficients γ_i) how inputs can be decreased and outputs increased in order to make the department under evaluation efficient.

Table 3: Scale efficient targets and MPSS of the departments

		Efficient Target (%)	Output			
a/a	Departments	Research	Graduates	Benchmarks	$\sum \gamma_i$	RTS
1	Mechanical Engineering	98.38	98.38	6,15	0.72	IRS
2	Urban Planning & Regional Developm.	8.92	8.92	6,15	1.06	DRS
3	Civil Engineering	90.10	90.10	6,15	0.72	IRS
4	Architecture	259.32	259.32	6,15	0.38	IRS
5	Computer & Communication Engineer.	43.81	43.81	6,15,16	1.09	DRS
6	Primary Education	0.00	0.00		1.00	CRS
7	Preschool Education	236.28	76.34	6,16	0.82	IRS
8	Special Education	110.98	110.98	6,15,16	0.85	IRS
9	History, Archaeology & Social Anthrop.	151.79	57.53	6,16	0.74	IRS
10	Agriculture, Crop & Rural Environment	68.71	68.71	6,15	0.66	IRS
11	Ichthyology and Aquatic Environment	112.17	112.17	6,15	0.51	IRS
12	Medical School	0.00	0.00		1.00	CRS
13	Veterinary Science	0.00	0.00		1.00	CRS
14	Biochemistry and Biotechnology	215.48	215.48	12,15,16	0.34	IRS
15	Physical Education and Sport Science	0.00	0.00		1.00	CRS
16	Economics	0.00	0.00		1.00	CRS

Furthermore, table 3 provides the relation between the proportional change in 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 a n per cent rise in all inputs. However if outputs rise by a larger percentage than inputs, there are increasing return to scales (IRS) while if outputs increase by a smaller percentage than inputs then there are decreasing returns to scale (DRS). As can be realized only the departments of Urban Planning and Regional Development and of Computer & Communication Engineering report DRS.

6 CONCLUSIONS

The majority of the relative existing studies (like Sinuany-Stern, Mehrez and Barboy 1994; King 1997; Arcelus and Coleman 1997; Sarrico and Dyson 2000; Tauer, Fried and Fry 2007; Kao and Hung 2008; Tyagi, Yadav and Singh 2009) evaluate the performance of departments of the same field for different university departments. However our paper applies for the first time to our knowledge bootstrap techniques in order to perform an efficiency analysis among different departments of the same university. As an illustrative example our paper evaluates the performance of sixteen different departments of the University of Thessaly. Nevertheless the proposed approach can be extended to different departments in different universities. Our DEA model contributes to the existing literature of efficiency analysis of departments of different fields within the same university (King 1997; Arcelus and Coleman 1997; Sarrico and Dyson 2000; Tauer, Fried and Fry 2007; Kao and Hung 2008; Tyagi, Yadav and Singh 2009) by applying the inferential approach and the latest developments on bias correction on the obtained efficiency scores as has been introduced by Simar and Wilson (1998, 2000, 2002, 2008). Similar to the pre mentioned studies examining the departments' efficiency among the same institutions, our paper focuses on resource utilization among departments and not on academic performance, which according to Kao and Hung (2008) surpasses the problem of different subjects among the departments. In addition our study assumes that departments inside a university may be considered as homogenous because they operate in similar activities and are willing to achieve similar goals (Tyagi, Yadav and Singh 2009). Furthermore, since the university under examination is a stated own university, the funding from the government goes to the university's central administration and then it is allocated to the different departments based on different resource criteria. Therefore, the inclusion of bootstrapping procedures producing bias corrected results in our analysis and the adoption of the resource allocation view of the departments' efficiency evaluation minimizes the heterogeneity related problems regarding the comparability of departments from different fields and thus it provides consistent results.

Finally, the empirical results obtained suggest that there are strong inefficiencies among the departments, indicating misallocation of resources or/and inefficient application of departments policy developments. Additionally, the article provides output target values for policy implications and evaluation among the departments. These targets provide benchmarks for the long term sustainability of the departments. Finally, our 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

The authors would like to thank two anonymous reviewers for their helpful and constructive comments on an earlier version of our paper. Any remaining errors are solely the authors' responsibility. We would also like to thank Prof. M. Zouboulakis, Vice Rector of Economic and Student Affairs, Public and International Relations of the University of Thessaly 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. Finally, we would like to thank all the departments' head administrative staff for the provision of vital information.

REFERENCES

Abbot, M. and C. Doucouliagos. 2003. "The efficiency of Australian universities: a data envelopment analysis." Economics of Education Review 22: 89-97.

Agasisti, T. and G. Johnes. 2010. "Heterogeneity and the evaluation of efficiency: the case of Italian universities." Applied Economics 42: 1365-1375.

Ahn, T., A. Charnes, and W.W. Cooper. 1988. "Some statistical and DEA evaluations of relative efficiencies of public and private institutions of higher learning." Socio-Economic Planning Sciences 22: 259-269.

Arcelus, F.J. and D.F. Coleman. 1997. "An efficiency review of university departments." International Journal of System Science 28: 721-729.

Athanassopoulos, A. and E. Shale. 1997. "Assessing the comparative efficiency of higher education institutions in the UK by means of data envelopment analysis." Education Economics 5: 117-133.

Avrikan, N. 2001. "Investigating technical and scale efficiencies of Australian Universities through data envelopment analysis." Socio-Economic Planning Science 35: 57-80.

Banker, R.D. 1984. "Estimating Most Productive Scale Size Using Data Envelopment Analysis." European Journal of Operational Research 17: 35-44.

Banker, R.D., A. Charnes, and W.W. Cooper. 1984. "Some Models for Estimating Technical and scale inefficiencies in DEA." Management Science 32: 1613-1627.

Beasley, J.E. 1990. "Comparing university departments." OMEGA 18: 171-183.

Bougnol, M.L. and J.H. Dula. 2006. "Validating DEA as a ranking tool: An application of DEA to assess performance in higher education." Annals of Operations Research 145: 339-365.

Carrington, R., T. Coelli, and D.S.P. Rao. 2005. "The performance of Australian universities: Conceptual issues and preliminary results." Economic Papers 24: 145-163.

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

Cohn, E., S. Rhine, and M. Santos. 1989. "Institutions of higher education as multi-product firms: Economies of scale and scope." The Review of Economics and Statistics 71: 284-290.

Colbert, A., R.R. Levary, and M.C. Shaner. 2000. "Determining the relative efficiency of MBA programs using DEA." European Journal of Operational Research 125: 656-669.

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

De Groot, H., W. McMahon, and F. Volkwein. 1991. "The cost structure of American research universities." The Review of Economics and Statistics I 73: 424-431.

Debreu, G. 1951. "The Coefficient of Resource Utilization." Econometrica 19: 273-292.

Dyson, R.G., R. Allen, A.S. Camanho, V.V. Podinovski, C.S. Sarrico, and E.A. Shale. 2001. "Pitfalls and protocols in DEA." European Journal of Operational Research 132: 245-259.

Fandel, G. 2007. "On the performance of universities in North Rhine-Westphalia, Germany: Government's redistribution of funds judged using DEA efficiency measures." European Journal of Operational Research 176: 521-533.

Farrel, M.J. 1957. "The measurement of productive efficiency." Journal of Royal Statistical Society A 120: 253-290.

Flegg, A.T., D.O. Allen, K. Field, and T.W. Thurlow. 2004. "Measuring the efficiency of British universities: A multi-period data envelopment analysis." Education Economics 12: 231-249.

Glass, J.C., G. McCallion, D.G. McKillop, S. Rasaratnam, and K.S. Stringer. 2006. "Implications of variant efficiency measures for policy evaluations in UK higher education." Socio-Economic Planning Sciences 40: 119-142.

Graves, P., R. Marchand, and R. Thompson. 1982. "Economics departmental rankings: Research incentives, constraints and efficiency." The American Economic Review 72: 1131-1141.

Halkos, G.E. and N.G. Tzeremes. 2010. "The effect of foreign ownership on SMEs performance: An efficiency analysis perspective." The Journal of Productivity Analysis 34: 167-180.

Harris, G.T. 1988. "Research output in Australian university economics departments, 1974-83." Australian Economic Papers 27: 102-110.

Hirsch, B., R. Austin, J. Brooks, and B. Moore. 1984. "Economics departmental rankings: comment." The American Economic Review 74: 822-826.

Izadi, H., G. Johnes, R. Oskrochi, and R. Crouchley. 2002. "Stochastic frontier estimation of a CES cost function: The case of higher education in Britain." Economics of Education Review 21: 63-71.

Johnes J. and L. Yu. 2008. "Measuring the research performance of Chinese higher education institutions using data envelopment analysis." China Economic Review 19: 679-696.

Johnes, G. 1988. "Determinants of research output in economics departments in British universities." Research Policy 17: 171-178.

Johnes, G. 1990. "Measures of research output: University departments of economics in the UK, 1984-8." The Economic Journal 100: 556-560.

Johnes, G. 1997. "Costs and industrial structure in contemporary British higher education." The Economic Journal 107: 727-737.

Johnes, G. and J. Johnes. 1993. "Measuring the research performance of UK economics departments: An application of data envelopment analysis." Oxford Economic Papers 45: 32-347.

Johnes, G. and J. Johnes. 2009. "Higher education institutions' costs and efficiency: Taking the decomposition a further step." Economics of Education Review 28: 107-113.

Johnes, J. 1996. "Performance assessment in higher education in Britain." European Journal of Operational Research 89: 18-33.

Johnes, J. 2006a. "Data envelopment analysis and its application to the measuring of efficiency in higher education." Economics of Education Review 25: 273-288.

Johnes, J. 2006b. "Measuring efficiency: A comparison of multilevel modelling and data envelopment analysis in the context of higher education." Bulletin of Economic Research 58: 75-104.

Kao, C. and H.T. Hung. 2008. "Efficiency analysis of university departments: An empirical study." Omega 36: 653-664.

King, W.D. 1997. "Input and output substitution in higher education." Economic Letters 57: 107-111.

Koopmans, T.C. 1951. "Analysis of production as an efficient combination of activities." In T.C. Koopmans, eds. Activity Analysis of Production and Allocation. New York: Wiley.

Madden, G., S. Savage, and S. Kemp. 1997. "Measuring public sector efficiency: A study of economics departments at Australian Universities." Education Economics 5: 153-168.

Ng, Y.C. and S.K. Li. 2000. "Measuring the research performance of Chinese higher education institutions: An application of data envelopment analysis." Education Economics 8: 139-156.

Raty, T. 2002. “Efficient facet based efficiency index: A variable returns to scale specification.” Journal of Productivity Analysis 17: 65-82.

Sarrico, C.S. and R.G. Dyson. 2000. “Using DEA for planning in UK universities-An institutional perspective.” Journal of the Operational Research Society 51: 789-800.

Shephard, R.W. 1970. Theory of Cost and Production Function. Princeton, NJ: Princeton University Press.

Silverman, B.W. 1986. Density estimation for statistics and data analysis. London: Chapman and Hall.

Simar, L. and P.W. Wilson. 1998. “Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models.” Management Science 44: 49–61.

Simar, L. and P.W. Wilson. 1999. “Estimating and Bootstrapping Malmquist Indices.” European Journal of Operational Research 115: 459-471.

Simar, L. and P.W. Wilson. 2000. “A general methodology for bootstrapping in non-parametric frontier models.” Journal of Applied Statistics 27: 779-802.

Simar, L. and P.W. Wilson. 2002. “Nonparametric tests of returns to scale.” European Journal of Operational Research 139: 115-132.

Simar, L. and P.W. Wilson. 2008. “Statistical interference in nonparametric frontier models: recent developments and perspectives.” In: H. Fried, C.A.K. Lovell, and S. Schmidt, eds. The measurement of productive efficiency and productivity change. New York: Oxford University Press.

Simar, L. and V. Zelenyuk. 2007. “Statistical inference for aggregates of farrell-type efficiencies.” Journal of Applied Econometrics 22: 1367-1394.

Sinuany-Stern, Z., A. Mehrez, and A. Barboy. 1994. “Academic departments efficiency via dea.” Computers & Operations Research 21: 543-556.

Tauer, L.W., H.O. Fried, and W.E. Fry. 2007. “Measuring of academic departments within a college.” Education Economics 15: 473-489.

Tomkins, C. and R. Green. 1988. "An experiment in the use of data envelopment analysis for evaluating the efficiency of UK university departments of accounting." Financial Accountability & Management 4: 147-164.

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

Worthington, A. and B.L. Lee. 2008. "Efficiency, technology and productivity change in Australian universities, 1998-2003." Economics of Education Review 27: 285-298.

Zelenyuk, V. and V. Zheka. 2006. "Corporate governance and firm's efficiency: the case of a transitional country, Ukraine." Journal of Productivity Analysis 25: 143-157.

APPENDIX

A synoptic illustration of Simar and Wilson (1998, 2000) bootstrap algorithm

In order to implement the homogenous bootstrap algorithm for a set of bootstrap estimates $\left\{ \hat{\lambda}_b^* \left(x, y\right) \middle| b = 1, ..., B \right\}$ for a given fixed point $\left(x, y\right)$ the following eight steps must be carried out:

From the original data set we compute $\hat{\lambda}_{CRS}$.

Then we apply the "rule of thump" (Silverman 1986, p.47-48) to obtain the bandwidth parameter h.

We generate $\beta_1^*, ..., \beta_n^*$ by drawing with replacement from the set $\left\{ \hat{\lambda}_1, ..., \hat{\lambda}_n, \left(2 - \hat{\lambda}_1\right), ..., \left(2 - \hat{\lambda}_n\right) \right\}$.

Then we draw $\varepsilon_i^*, i = 1, ..., n$ independently from the kernel function $K\left(.\right)$ and compute $\beta_i^{**} = \beta_i^* + h\varepsilon_i^*$ for each $i = 1, ..., n$.

For each $i = 1,...,n$ we compute β_i^{***} as: $\beta_i^{***} = \bar{\beta}^* + \frac{\beta_i^{**} - \bar{\beta}^*}{\left(1 + h^2 \sigma_k^2 \sigma_\beta^2\right)^{1/2}}$, where $\bar{\beta}^* = \sum_{i=1}^{n} \beta_i^* / n$, $\sigma_\beta^2 = \sum_{i=1}^{n} \left(\beta_i^* - \bar{\beta}^*\right) / n$ and σ_k^2 is the variance of the probability density function used for the kernel function. In addition λ_i^* can then be computed as: $\lambda_i^* = \begin{cases} 2 - \beta_i^{***} \forall \beta_i^{***} < 1 \\ \beta_i^{***} \text{ otherwise} \end{cases}$.

The bootstrap sample is created as: $\mathrm{X}_n^* = \left\{\left(x_i^*, y_i\right) \middle| i = 1,...,n\right\}$ where $x_i^* = \lambda_i^* \hat{x}^{\partial}\left(y_i\right) = \lambda_i^* \hat{\lambda}_i^{-1} x_i$.

We compute the DEA efficiency estimates $\hat{\lambda}_i^*\left(x_i, y_i\right)$ for each of the original sample observations using the reference set X_n^* in order to obtain a set of bootstrap estimates.

Finally, we repeat steps 3 to 7 B times (at least 2000 times) to obtain a set of bootstrap estimates $\left\{\hat{\lambda}_b^*\left(x, y\right) \middle| b = 1,...,B\right\}$.