ON EFFICIENCY OF HIGHER EDUCATION WITH DATA ENVELOPMENT ANALYSIS AND REGRESSION

Marilena-Aura DIN [a*], Cristina COCULESCU [a]

[a] *Romanian American University of Bucharest, Romania*

ABSTRACT
The efficiency of higher education systems is a constant concern at the level of each country. At the national policy level, governments are interested to know the determinant factors to enhance the higher education efficiency and whether increasing the state's contribution to the financing of the sector are a significant factor in increasing the efficiency in Higher Education. Our paper addresses so-called output-oriented Data Envelopment Analysis (DEA) requiring mathematical programming as a first stage and the regression of the obtained efficiency scores on explanatory factors as a second step. These allow us a cross-country analysis of the efficiency of higher education sector in different states with different GDP but as homogeneous as possible in the sample, on models that combine as input the level of state contribution to the financing of the sector from the total expenditure, and as output the percentage of graduates with diploma from the total population of the same age and employability rate. The analysis in the two steps, respectively the DEA followed by the regression, examines which of the non-discretionary inputs is statistically significant, respectively to what extent the state's contribution to the financing of systems or other factors such as GDP per capita influences the efficiency of the higher education systems chosen for comparison.

KEYWORDS: *data envelopment analysis, efficiency, higher education, regression.*

1. INTRODUCTION

The efficiency of higher education systems is a constant concern at the level of each country. At the national policy level, governments are interested to know the determinant factors to enhance the higher education efficiency and whether increasing the state's contribution to the financing of the sector are a significant factor in increasing the efficiency in Higher Education. It is expected that the higher the state allocations for the sector, the more efficient the sector to reach its targets, both as outputs of the system and as outcomes.

In recent years, more and more research works have begun to approach tertiary education efficiency and effectiveness across countries using the DEA approach as a research methodology. DEA is a non-parametric optimization method, from the class of mathematical programming problems that are different from the econometric approach – (Stochastic Frontier Analysis).

An efficiency assessment in public tertiary education systems across EU countries, Japan and the US with semi-parametric methods and stochastic frontier analysis have proved that a good quality secondary system, output-based funding rules, institutions' independent evaluation and staff policy autonomy are positively related to efficiency (Aubyn et al., 2009). Moreover, their study provides evidence that public spending on tertiary education is more effective in what concerns labour productivity growth and employability when it is coupled with efficiency.

[*] Corresponding author. E-mail address: *marilena.din@gmail.com*

In their research paper, Nadoveza Jelić and Gardijan Kedžo (2018) deals with tertiary education efficiency computed using Data Envelopment Analysis (DEA), and with effectiveness across 24 European Union countries in four sub-periods between 2004 and 2015.

Solihin et al. (2017) analyzed by DEA the efficiency of local government expenditure on education sector. Measurement of the effectiveness of government spending was done by using panel data regression. Furthermore, their study showed that government expenditure in educational sector is relatively inefficient; government expenditure for education has no significant impact on educational index, while household expenditure for education and GDP per capita has positive significant impact on the education index. Thus, they proved that government expenditure for educational sector is not effective improving educational index.

Agasisti (2011) has analyzed by DEA input oriented, under variable returns to scale (VRS) in two stages, higher education sectors in 18 OECD countries, and find that in some specification funding for tertiary education is negatively correlated with its efficiency. For the second stage his analysis used Tobit regression.

Wolszczak-Derlacz and Parteka (2011) have analyzed institutions in seven European countries and conclude that more efficient universities have a larger number of different departments, a larger proportion of females among the academic staff, a higher percentage of funds from external sources, and are older.

These results created the premises for our approach. The purpose of our paper is to assess the relative technical efficiency of the higher education sector in a cross-country analysis of European states with different GDP. For the most part, the proposed analysis should consists of two stages, respectively so-called output-oriented Data Envelopment Analysis (DEA) requiring mathematical programming as a first stage and the regression of the obtained efficiency scores on explanatory factors as a second step, for assessing the effectiveness of Government spending on efficiency of higher education sector.

Our paper is organized as follows. After introduction, the second section covers the methodology of research, including general overview of the topic, such as the description of the concepts used, the rationale for the selection of the inputs and outputs, questions under research and inputs and outputs measurement from the perspective of the expected results, general consideration about how it works DEA, and the proposed DEA models results. The analysis is followed by conclusions and further research.

2. THE METHODOLOGY OF RESEARCH

2.1 General overview

The DEA method is a non-parametric one, which evaluates the operating efficiency of some comparison units - DMUs, ranking them so that each DMU will know how to improve their efficiency compared to other, more efficient DMUs that are on the frontier of efficiency.

The efficiency score, empirical efficient frontier, performance evaluation, benchmarking, direction of the increase in efficiency are some of the results that recommend this method, besides the fact that the DEA technique does not imply the specification of a function that transforms inputs into outputs; the principle being similar to a black box in which a set of inputs goes in and a set of outputs goes out, given the observed data.

Standard efficiency definition is given by the general formula:

$$Eff = \frac{\sum O}{\sum I} \qquad (1)$$

In our analysis of efficiency we preferred that the outputs be more outcomes oriented, by introducing variables such as the employment rate, which tends to the effectiveness approach as a concept. Also, the number of graduates as output is complemented by the graduation of the

specialties in science, technology, computing, engineering, for which it is known that the quality of the human resource from teaching and research, as well as the investments in equipment, matter.
Also, we tried to refine the analysis of strict efficiency by approaching as the input of quality criteria such as the ratio of the number of students to the academic staff. Nadoveza Jelić and Gardijan Kedžo (2018) demonstrated that quality considerations affect the relative positions of countries regarding their efficiency scores.

2.2 Inputs and outputs DEA specification

We used here a DEA output - oriented model, under variable returns to scale (VRS) where the quantity and quality of the inputs has to be fixed.
A DEA output - oriented model examines whether a system produces enough output for the available resource inputs. Here, different DEA models combine as inputs the level of state contribution to the financing of the sector from the total expenditure, and as outputs different combinations among which the percentage of graduates with diploma from the total population of the same age and employability rate, as in the Table 1.
VRS refers to Variable Return to Scale, respectively, increasing and decreasing return to scale (Banker et al., 1984).
There are considerable problems of defining and measuring the inputs and outputs of higher education production process (Johnes, 2004). One of them is the statistical significance of the input and output variables included in the multiple models.

Table 1. Measures for inputs and outputs

	Inputs		**Outputs:**
1.	PEGDP-Public expenditure on tertiary education - as % of gross domestic product (GDP)	1.	GHE - Graduates in tertiary education per 1000 of population aged 20-29
2.	RSAS - Ratio of students to academic staff at tertiary education level	2.	ERHE Employment rates for tertiary education graduates (%)
		3.	PHE30_34 Population aged 30-34 with tertiary education attainment
		4.	STEM20_29 Tertiary education graduates, in science, math., computing, engineering, manufacturing, construction per 1000 of population aged 20-29

Source: adapted from Eurostat
(tertiary education correspond to ISCED levels 5-8)

2.3 Questions under research

Under DEA analysis the main question is how much efficiently the higher education systems convert their inputs (i.e. public funds and human resources) into the outcome oriented outputs (i.e. employment rates of graduates and EU target indicator given by tertiary educational attainment, in addition to the percent of graduates of total population aged 20-29, or STEM graduates)?
Which are the best practice systems on the frontier of efficiency?
Which efficient systems can be utilized as benchmarks for improvement. In other words, in order to adopt these best practices to improve quality, each inefficient system of higher education under

evaluation has to be projected on the efficiency frontier to find which is the virtual target given by the convex combination of peer countries performing better.

2.4 How it works DEA

For the efficiency of one DMU_0, the DEA method determine if there is *a set of weights* (lambda's) so that the *convex combination of n DMUs perform better than DMU_0*, which is under evaluation.
In case there is such set of weights, this DMU_0 is *inefficient*, and otherwise, the DMU_0 is *efficient*, and it is on the *efficiency frontier*.
Thus, a linear programming problem (L.P.P.) is solved for each DMU_0 that is under evaluation in terms of efficiency. If we have DMU_j, with j = 1, ... n, s outputs Yrj from m inputs Xij, where r = 1, ..., s and i = 1, ..., m, Φ the objective function, i.e. the efficiency function, with the constrains that all the convex linear combinations of these inputs be upper bounded by the input of the DMU_0 unit under evaluation, and the convex linear combinations of all the outputs be lower than ΦY_{r0}. Then, for the DEA output oriented model, the mathematical model of the efficiency is the linear programming problem which run for each DMU_0 as follows:

$$\Phi^* = \max \Phi$$
Subject to:
$$\begin{cases} \sum_{j=1}^{n} x_{i,j} \cdot \lambda_j \le x_{i_0}, \ i = 1,\ldots,m \\ \sum_{j=1}^{n} y_{r,j} \cdot \lambda_j \ge \Phi \cdot y_{r_0}, \ r = 1,\ldots,s \end{cases} \quad (2)$$
$$\sum_{j=1}^{n} \lambda_j = 1, \ \lambda_j \ge 0, \ \text{j=1,\ldots,n}$$

If Φ*=1, then the current output level cannot be increased, meaning that the DMU_0 is on the efficiency frontier, i.e. is a *best practice*.
If Φ*>1, then the same input levels would be enough for higher levels of outputs, which means that DMU_0 is *inefficient*.
In the second stage, with the optimal solution of the previous L.P.P., respectively Φ* given by the efficiency model (2), the efficiency score Φ* in the slack model is considered fixed:
Mathematical model – DEA output oriented (2) –model

$$\max \left(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \right)$$
Subject to:
$$\begin{cases} \sum_{j=1}^{n} x_{i,j} \cdot \lambda_j + s_i^- = x_{i_0}, \ i = 1,\ldots,m \\ \sum_{j=1}^{n} y_{r,j} \cdot \lambda_j - s_r^+ = \Phi^* \cdot y_{r_0}, \ r = 1,\ldots,s \end{cases} \quad (3)$$
$$\sum_{j=1}^{n} \lambda_j = 1, \ \lambda_j \ge 0, \ \text{j=1,\ldots,n}$$

A DMU is considered *DEA Efficient* if and only if the efficiency score is equal 1 and zero slack values. If the efficiency score is equal 1, but there are some non-zero slack values, than the DMU is *DEA weakly efficient*.

The target of DMU_0 found as inefficient is given by the next system:

$$\begin{cases} \hat{x}_{i_0} = x_{i_0} - s_i^- \ , \ i = 1, \dots, m \\ \hat{y}_{r_0} = \Phi^* \cdot y_{r_0} + s_r^+ , \ r = 1, \dots, s \end{cases} \quad (4)$$

In a graphical representation of points (x_{ij}, y_{rj}), for the DEA output - oriented model, the efficiency frontier is obtained as a concave polygonal line as in the figure 1., where an inefficient point like DMU_0 is projected from the center of the reference system on the border to find the target of DMU_0. This target represents a virtual unit given by the convex combination of others peer DMUs DMU_A and DMU_B, as suggested in the figure 1.

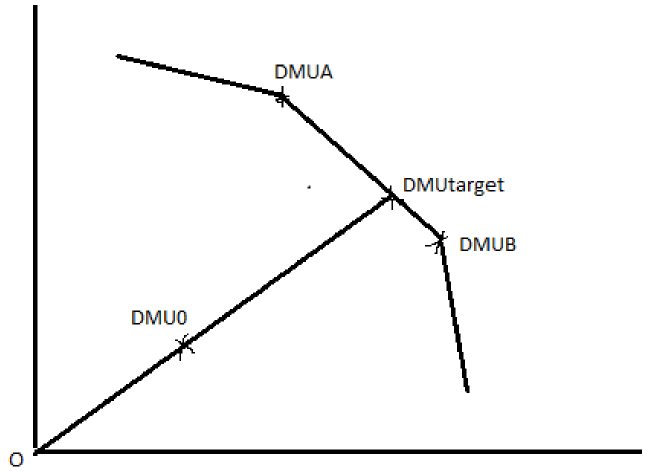

Figure 1. Efficiency target of an inefficient DMU_0

Source: Authors

In other words, DEA method compares a DMU_0 to an efficient target, which is its projection of on the efficiency frontier. DEA method gives also the value of the relative efficiency, as the value less than one, given by the ratio between the distances from DMU0 and from the target to the origin of the reference system. The target of the inefficient DMU_0 appears as a virtual best practice unit. In this way DEA method can also help to find improvements in order to achieve the target, comparing with the other peer DMUs given by the convex combination.

2.5 DEA models results

The data for inputs and outputs were collected from Eurostat data base on a sample of 25 countries, for the year 2016. However, the availability of the collected comparable data influenced the results. Five models have been built, differing in the number of inputs and outputs as in the Table 2, their choice being guided by states of the art, i.e. the previously mentioned studies at the levels of country systems. We run the DEA slacks-based methods (SBM) developed by Tone (2001), oriented under VRS.

Table 2. Types of DEA models used for analysis

		Model I (1I-2O)	**Model II (1I-3O)**	**Model III (2I-3O)**	**Model IV (2I-4O)**	**Model V (2I-3O)**
Input 1	PEGDP	x	x	x	x	x
Input 2	RSAS			x	x	x
Output 1	GHE20_29	x	x	x	x	x
Output 2	ERHE	x	x	x	x	x
Output 3	PHE30_34		x	x	x	
Output 4	STEM20_29				x	x

Source: Own elaboration

The first stage DEA results for all models are presented in Table 3 and 4. As can be seen from Table 4, efficiency scores are senzitive to the choice of inputs and outputs. As well as the number of efficient DMUs increases as the number of inputs and outputs increases.

Table 3. DEA Efficiency scores

		M-I	**M-II**	**M-III**	**M-IV**	**M-V**
No.	DMU	Score	Score	Score	Score	Score
1	Belgium	0.8973	0.8928	0.8928	0.7918	0.7393
2	Bulgaria	1	1	1	1	1
3	Czechia	0.9039	0.8574	0.8574	0.8901	0.8972
4	Denmark	1	1	1	0.9999	1
5	Germany	0.7229	0.6892	0.7825	0.8037	0.779
6	Estonia	0.6798	0.7481	0.7814	0.737	0.647
7	Greece	0.7324	0.8188	0.8188	1	0.7727
8	Spain	0.8163	0.832	1	1	1
9	France	1	0.9995	0.9996	1	1
10	Italy	0.6597	0.6319	0.6319	0.6633	0.6542
11	Cyprus	0.6797	0.821	0.821	0.7631	0.5615
12	Latvia	0.8967	0.8738	0.8738	1	1
13	Lithuania	1	1	1	1	1
14	Hungary	0.7821	0.7501	0.8079	0.8186	0.8321
15	Malta	0.9998	0.9986	1	1	1
16	Netherlands	0.9062	0.8731	0.9027	0.9999	0.9999
17	Poland	0.982	0.9293	0.9649	0.9781	0.9886
18	Romania	0.8002	0.6971	0.6971	0.766	0.8621
19	Slovenia	1	1	1	1	1
20	Slovakia	0.7635	0.7285	0.7347	0.6749	0.6638
21	Finland	0.8549	0.8778	0.8858	0.903	0.8699
22	Sweden	0.7276	0.7981	1	1	0.8296
23	United Kingdom	0.989	0.9444	0.9508	1	1

		M-I	M-II	M-III	M-IV	M-V
24	Norway	0.8158	0.8602	1	1	0.8825
25	Turkey	0.6382	0.6248	0.6248	0.5295	0.5097
No. of inputs/outputs		1I/2O	1I/3O	2I/3O	2I/4O	2I/3O
No. of Eff. DMUs		5	4	8	11	9

Source: Own elaboration

Table 4. Efficiency ranks

		M-I	M-II	M-III	M-IV	M-V
No.	DMU	Rank	Rank	Rank	Rank	Rank
1	Belgium	11	9	13	19	20
2	Bulgaria	1	1	1	1	1
3	Czech Republic	10	14	16	16	12
4	Denmark	1	1	1	12	1
5	Germany	21	23	20	18	18
6	Estonia	22	20	21	22	23
7	Greece	19	17	18	1	19
8	Spain	14	15	1	1	1
9	France	1	5	9	1	1
10	Italy	24	24	24	24	22
11	Cyprus	23	16	17	21	24
12	Latvia	12	11	15	1	1
13	Lithuania	1	1	1	1	1
14	Hungary	17	19	19	17	16
15	Malta	6	6	1	1	1
16	Netherlands	9	12	12	12	10
17	Poland	8	8	10	14	11
18	Romania	16	22	23	20	15
19	Slovenia	1	1	1	1	1
20	Slovakia	18	21	22	23	21
21	Finland	13	10	14	15	14
22	Sweden	20	18	1	1	17
23	United Kingdom	7	7	11	1	1
24	Norway	15	13	1	1	13
25	Turkey	25	25	25	25	25

Source: Authors

Table 5. Efficiency scores - Descriptive statistics

	M-I	M-II	M-III	M-IV	M-V
Average	0.8499	0.8499	0.8811	0.8928	0.8596
Max	1	1	1	1	1
Min	0.6382	0.6248	0.6248	0.5295	0.5097
St Dev	0.1263	0.1192	0.1218	0.1399	0.1559

Source: Authors

According with the model M-V, the most efficient systems are in the countries Bulgaria, Denmark, Spain, France, Latvia, Lithuania, Malta, Slovenia, and United Kingdom.
Romania is ranked to the 15 place. If our country wants to become efficient, it will have to aim at the position of Bulgaria, Latvia and Slovenia – as peers, with the improvements given by the projection values from Table 6a and 6b.

Table 6a. Efficiency target/projections for inputs

PEGDP			RSAS		
Data	Projection	Diff.(%)	Data	Projection	Diff.(%)
0.71	0.71	0	16.8	16.5253	-1.635

Source: Authors

Table 6b. Efficiency target/projections for outputs

GHE20_29			ERHE			STEM20_29		
Data	Projection	Diff.(%)	Data	Projection	Diff.(%)	Data	Projection	Diff.(%)
50	65.2676	30.535	84.2	84.2	0	14.4	16.9142	17.46

Source: Authors

Table 7. Results for HESs efficiency

DMU	DEA Eff	Rank	Peers			
Belgium	0.7393	20	France	Slovenia		
Bulgaria	1	1	Bulgaria			
Czechia	0.8972	12	Bulgaria	Latvia	Slovenia	
Denmark	1	1	Denmark			
Germany	0.779	18	Malta	Slovenia	United Kingdom	
Estonia	0.647	23	Malta	Slovenia	United Kingdom	
Greece	0.7727	19	Bulgaria	Slovenia		
Spain	1	1	Spain			
France	1	1	France			
Italy	0.6542	22	Bulgaria	Slovenia		
Cyprus	0.5615	24	France	Slovenia		
Latvia	1	1	Latvia			
Lithuania	1	1	Lithuania			
Hungary	0.8321	16	Bulgaria	Lithuania	Malta	Slovenia
Malta	1	1	Malta			
Netherlands	0.9999	10	Netherlands	United Kingdom		
Poland	0.9886	11	France	Lithuania	Malta	Slovenia
Romania	0.8621	15	Bulgaria	Latvia	Slovenia	
Slovenia	1	1	Slovenia			
Slovakia	0.6638	21	Malta	Slovenia	United Kingdom	
Finland	0.8699	14	Malta	Slovenia	United Kingdom	
Sweden	0.8296	17	Malta	Slovenia		
United Kingdom	1	1	United Kingdom			
Norway	0.8825	13	Malta	Slovenia		
Turkey	0.5097	25	Slovenia			

Source: Authors

3. FURTHER RESEARCH AND CONCLUSIONS

Some HE systems are for sure oriented more on scientific research or business orientation, than on teaching, but the present analysis makes no distinction between them, considering the efficiency of all types of activities. Another one problem that needs attention is that of homogeneity in the regressed sample, which leads to the idea of grouping countries on clusters for GDP per capita. The findings suggest that results are highly sensitive to methodology and models.
A serious problem that needs attention is to study the sensitivity of the results. Given the possibility of running DEA on a wide variety of specifications, it is necessary to apply significance tests of the input and output variables included in the model. Pastor et al. (2002) developed a test to evaluate the significance of nested models in a radial DEA.
Additional problems with the two stage procedure is that DEA efficiency estimates are serially correlated making standard methods of inference invalid (Simar & Wilson, 2004).
In our DEA model, only discretionary inputs are included, which can be modified by the decision making system. But educational outcomes may also be influenced by environmental factors, i.e. non-discretionary inputs, which no longer depend on the decision making system.
The question arises that if non-discretionary or environmental inputs are taken into account alongside the input of interest, namely the state's contribution to the financing of the tertiary education sector (GF), which of these environmental inputs has significant effects on the efficiency of higher education systems? The socio-economic factors such as the level of education of the parents - PED or the material state of the families, measured by GDP per capita – GDPC, which may influence the performance of the students and ultimately the educational outcomes.
Therefore, in the second stage we will consider the efficiency score obtained with the DEA method, as a dependent variable in a multiple regression with the exogenous factors given by the non-discretionary variables GDPC and PED, and the discretionary variables like GF, or PF, as in the Table 8. Johnes (2006) asserts that, according to the issue addressed by Ruggiero in 1996, the generally approach of including in the DEA efficiency analysis of all inputs, whether controllable or not, can produce results which do not make adequate allowance for higher education institutions (HEIs), and their inefficiency may be overestimated as a consequence.

Table 8. Regression determinants

	Exogenous Factors	
Non-discretionary variables	GDP per capita	GDPC
	The level of education of the parents	PED
Discretionary variables	Funding of Tertiary education from central government (GF) in %	GF

Source: Authors

The regression analysis will examine which of the non-discretionary inputs is statistically significant, namely GDP per capita and parents' education level, or the discretionary inputs namely the state's financial contribution, and respectively household expenditure for education influences the efficiency of the higher education systems chosen for comparison. In other words, to what extent can the state intervene through the financial contribution to the efficiency of higher education systems? Is the efficiency of the sector less sensitive to the state's financial contribution compared to the influence of the non-discretionary factors, or is it possible that spending from private sources (i.e. funds attracted from other private sources, local business collaborations, or tuition fees) may have greater effects on the efficiency of the sector. This would mean differences between the responsibilities of spending public vs. private money.

For the estimation of the regression (5), the dependent variable is the DEA efficiencies, that takes values only between 0 and 1. This restriction is natural and does not arise from censorship

$$DEA = \beta_0 + \beta_1 \cdot GDPC + \beta_2 \cdot PED + \beta_3 \cdot GF + u \qquad (5)$$

If it is not the case of censored at 0 and 1, means that is not the situation of Tobit Regression. Linear regression models can be problematic for variables that are restricted in range because linear models will typically predict values outside that range for sufficiently extreme values of the predictors. For outcomes restricted to the 0 to 1 range, it is often used approaches like Fractional Regression or Beta Regression. For the further research will take in account how fractional regression can be applied, also to longitudinal data, and how to handle endogeneity in those models. Also continue with findings if is true the suspicion that higher efficiency is inversely correlated with increased levels of state funding, i.e. the degree of state contribution could be negatively correlated with efficiency.

For the moment, just from DEA results we have inefficiency of higher education systems even for the countries with high percentage of state revenues. But has the GDP per capita the greatest influence on the efficiency of the systems relative to their inputs? Also, how much matters the efficient spending for employability?

In any case, it is clear so far that financing mechanisms have the potential to alter the efficiency of higher education providers, and systems, but still not clear in what directions. The answers could have broad policy implications, even optimizing the efficiency of the systems by increasing the private contribution on expenses of higher education systems.

REFERENCES

Agasisti, T. (2011). Performances and spending efficiency in higher education: a European comparison through non-parametric approaches. *Education Economics*, *19*(2), 199-224.

Aubyn, M. St., Pina, A., Garcia, F., & Pais, J. (2009). Study on the efficiency and effectiveness of public spending on higher education. *Economic Papers No. 390.* Brussels: European Commission, European Comission, Directorate-General for Economic Affairs Publications. Retrieved from https://ec.europa.eu/economy_finance/publications/pages/publication16267

Banker, R. D., Charnes, A., & Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30*(9), pp1078-1092.

Johnes, J. (2004). Efficiency measurement in Johnes, G & Johnes, J (eds). *The International Handbook on the Economics of Education*, Edward Elgar, Cheltenham.

Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review, 25*(3), 273-288.

Nadoveza Jelić, O. & Gardijan Kedžo, M. (2018). Efficiency vs effectiveness: an analysis of tertiary education across Europe. *Public Sector Economics, 42*(4), 381 – 414.

Pastor, J.T., Ruiz, J.L., & Sirvent, I. (2002). A statistical test for nested radial DEA models. *Operations Research, 50*(4), 728-735.

Simar, L., & Wilson, P.W. (2004). Estimation and inference in two-stage, semiparametric modelsof production process. *Discusion Paper*, Institut de Statistique, Universite Catholique de Louvain, Belgium.

Solihin, A., Mursinto, D., & Sugiharti, L. (2017). Efficiency and Effectiveness of Government Expenditure on Education at Districts/Cities Level in East Java Indonesia. *Asian Social Science, 13*(8), 91-102.

Tone, K. (2001). A slacks based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, *130*, 498-509.

Wolszczak-Derlacz, J. & Parteka, A. (2011). Efficiency of European public higher education institutions: a two-stage multi-country approach, *Scientometrics, 89*(3), 887-917.