2020 Vol 2, Issue 1, February 2020

2020-02

Efficiency ranking using Principal Component Analysis

Florou, Giannoula

Research Institute for Entrepreneurship Development (RIED): Neapolis University, Pafos

Efficiency ranking using Principal Component Analysis

*Florou Giannoula, **Anastasiadou Sofia

*International Hellenic University, gflorou@teiemt.gr
**University of Western Macedonia, sanastasiadou@uowm.gr

Abstract
Many studies analyze the efficiency ranking of various organizations, using mainly the Data Envelope Analysis (DEA) with input and output variables. DEA is a non-parametric method, which estimates efficiency of one organization in relation to the best organization in the specific field. The method of DEA is based on linear programming and on measuring the efficiency of a production process, which represents the highest amount of output produced by given amount of input, in a specific time frame.
This paper presents an alternative method of efficiency ranking using the Principal Component Analysis. Organizations are presented in the first principal component plane, using the ratios of output variables per input variables. We rank organizations using the first components coefficients. We analyze the advantages and disadvantages of the alternative method and we present an application of this method in ranking the efficiency of the Greek Technological Institutes.

Keywords: DEA, Principal Components Analysis, Efficiency, Greek, Technological Institutes

1. Introduction

Data Envelope Analysis (DEA) is a methodology often used to measure efficiency – productivity of various organisations. It uses linear programming methodology and compares units – organisations with common input and output variables aiming at discovering the most efficient ones. When the first classic models of DEA appeared (Charnes et al., 1978; Banker et al., 1984), there was no evaluating ranking of organisations, but only a classification in relation to the best one or ones.

Later, modified models of Data Envelope Analysis were used, in order for the ranking of the units-organizations to occur (Andersen & Petersen, 1993). The method was further developed and was connected to the use of other methodologies in different scientific fields.

Principal Components Analysis (PCA) is a method of the multidimensional Data Analysis, which is used for the main components of a dataset to be found, extracting new, independent variables as linear combinations of the initial variables. These new variables are much fewer than the initial variables, are called main components and describe data in the best possible way. Various studies were published (Premachandra, 2001; Kardiyen & Orkcu, 2006), comparing Principal Components Analysis and Data Envelope Analysis in organization ranking as far as their effectiveness is concerned, not only using real data but also simulation data.

After that, in the second chapter of the present paper we present the two methods briefly. In the third chapter, we display the data of this paper which deal with real data of Greek Technological Educational Institutions. We describe their modification, as we propose, into data of input-output form, so that they can be used in effectiveness ranking via Principal Component Analysis. In the fourth section we analyze the effectiveness ranking and in the fifth chapter the conclusions of our paper are presented.

2. Data Envelope Analysis and Principal Components Analysis

Data Envelope Analysis is a non-parametric method, proposed by Charnes, Cooper, Rhodes for measuring relative efficiency of units, similar in terms of services or production. Various ways are used to estimate efficiency. They derive mainly from the output per input ratio (Kardiyien & Orkcu, 2006, Mavris, et al., 2019).

The comparison of each unit- organization is achieved by maximizing the fraction resulting from the sum of the weighted output to the sum of the weighted output. Provided that the respective ration of each unit is less or equal to 1.

max $h_{j0}=\Sigma u_r y_{rj0}/\Sigma v_i x_{ij0}$ r=1….s, i=… m

With the limitations $\Sigma u_r y_{rj} - \Sigma v_i x_{ij} <= 0$ j=1,2,…n

$u_r, v_i > \varepsilon$ for every i, r (Derpanis, 2009).

The method consists in our effort to maximize the gravity coefficients of the fraction, so that efficiency has maximum 1. If the ration of an organization is 1, it is characterized as effective, while if it is lower than 1, it is considered non effective. We cannot interfere a great deal with the maximization process of the algorithm, nor with the way of calculating gravity coefficients.

Principal Component Analysis is a Data Analysis Method, which is used for data table of n objects (in lines) and p variables (in columns). As a result of the method k uncorrelated new variables derive (k<p), which are a linear combination of the initial variables, and which explain a large part of the variability of the initial data.

In this paper, we propose the use of Principal Component Analysis, to estimate effectiveness – efficiency of the organizations. This method yields equally good results with Data Envelope Analysis, while at the same time it classifies organizations according to their effectiveness (Kardiyien & Orkcu, 2006). The advantages of the proposed method in relation with DEA are simplicity, brevity, better interpretation of results and the combination of main principles so as to have a comprehensive picture of effectiveness.

3. Conversion of input-output data in ratios

In Data Envelope Analysis a total fraction is used for output/input ratios. In using Principal Components Analysis, as we suggest, we calculate every possible fraction of output/input, which may result from the data variables. Thus, we create a new table of data with r organizations (in lines) and p new variables – ratios (in columns).

Afterwards, by applying Principal Components Analysis to the new table, we try to find principal components describing in the best possible way the organizations, minimizing thus, the number of variables. In most cases, ranking in the plane of the 1^{st} or 2^{nd} component is enough to reveal a good picture of the comparative ranking of the organizations of interest to us.

We choose the k main components, the sum of eigenvalues of which (interpreted variability and respective component gravity) is higher than 0.80 (80%). We recalculate the coordinates of r organizations and initial variables, using now only the k main components. The new coordinates are negative or positive. If for one main component the coordinates of all the units- organizations are negative, the gravity of this main component is considered negative. If the coordinates of all the units- organizations are positive, the gravity of this main component is considered positive. If most coordinates of units-organizations are negative, the gravity of this main component is considered negative, while in the opposite case it is regarded positive.

4. Implementation in Greek Technological Institutions

We use Greek Technological Educational Institutions (TEI) as an example of implementation of the proposed methodology. The following input and output variables have been suggested by other writers, in their related articles, to measure the effectiveness of TEI (Katharaki-Kathatakis, 2010).

Input Variables

- Number of Teaching Staff
- Number of other staff
- Number of active students
- Total expenses (except payroll costs of permanent staff)

Output Variables

- Number of graduates

- Number of publications and references
- Level of revenues of Special Research Account of every TEI

We collected real data for Greek TEI, as they have been published by the Greek Statistic Authority for the year 2011, concerning the number of students, graduates and staff (www.statistics.gr).

The data for operating costs occurred on the basis of the grants distribution of the Ministry of Finance for 2011.

The number of publications resulted from the research of the National Documentation Center and deals with the five-year period 2006-2010. It is based on the publications of the Web of Sciences database (EKT, 2012).

Unfortunately, we were not able to find reliable data for the level of revenues of every Special Research Account of TEI and for this reason we did not use this output variable. The data collected, are presented on table 1.

Table 1: Initial data

DATA ENTRY OF ACADEMIC YEAR 2011/2012	2011 OPERATING COSTS (In ths)	PermanentTS	Lecturers on CONTRACT	TOTAL of students of WINTER SEMESTERS	TOTAL of students BEYOND REGULAR SEMESTERS	Publications	TOTAL of graduates 2010/11
T.E.I OF ATHENS	9920	498	875	14936	12955	482	3483
T.E.I OF CRETE	5160	196	492	10706	10787	373	1447
ALEXANDER T.E.I OF THESSALONIKI	8640	293	545	12615	11813	331	2249
T.E.I OF KAVALA	3680	133	228	5903	5416	108	838
T.E.I OF WESTERN MACEDONIA	5120	119	303	9658	12680	131	1266
T.E.I OF LARISA	5160	210	507	10608	9478	170	1637
T.E.I OF PATRAS	4920	94	462	11168	9890	82	1846
T.E.I OF PIRAEUS	6120	147	480	6530	6488	142	1336
T.E.I OF SERRES	2760	74	156	6126	7529	71	670
T.E.I OF KALAMATA	2680	48	117	2872	2519	85	658
T.E.I OF MESOLOGGI	2840	63	121	4498	3244	64	629
T.E.I OF CHALKIDA	2520	66	149	7340	7085	98	818
T.E.I OF LAMIA	2240	63	99	3827	3877	94	576
T.E.I OF EPIRUS	4280	88	203	7580	6010	99	1049
T.E.I OF IONIAN ISLANDS	2304	25	131	3078	1314	28	253

On table 2 the ratios of output variables to input variables of every TEI are presented, as they were calculated with respective divisions.

Table 2: Ratios of output variables to input variables

DATA ENTRY OF ACADEMIC YEAR 2011/2012	PUBLICATIONS OPERATING COSTS	PUBLICATIONS of TS	PUBLICATIONS of lec. on CONTRACT	PUBLICATIONS of STUDENTS	GRADUATES-OPERATING COSTS	GRADUATES_TS	GRADUATES_CONTRACT
T.E.I OF ATHENS	0.049	0.968	0.551	0.032	0.351	6.994	3.981
T.E.I OF CRETE	0.072	1.903	0.758	0.035	0.280	7.383	2.941
ALEXANDER T.E.I OF THESSALONIKI	0.038	1.130	0.607	0.026	0.260	7.676	4.127
T.E.I OF KAVALA	0.029	0.812	0.474	0.018	0.228	6.301	3.675
T.E.I OF WESTERN MACEDONIA	0.026	1.101	0.432	0.014	0.247	10.639	4.178
T.E.I OF LARISA	0.033	0.810	0.335	0.016	0.317	7.795	3.229
T.E.I OF PATRAS	0.017	0.872	0.177	0.007	0.375	19.638	3.996
T.E.I OF PIRAEUS	0.023	0.966	0.296	0.022	0.218	9.088	2.783
T.E.I OF SERRES	0.026	0.959	0.455	0.012	0.243	9.054	4.295
T.E.I OF KALAMATA	0.032	1.771	0.726	0.030	0.246	13.708	5.624
T.E.I OF MESOLOGGI	0.023	1.016	0.529	0.014	0.221	9.984	5.198
T.E.I OF CHALKIDA	0.039	1.485	0.658	0.013	0.325	12.394	5.490
T.E.I OF LAMIA	0.042	1.492	0.949	0.025	0.257	9.143	5.818
T.E.I OF EPIRUS	0.023	1.125	0.488	0.013	0.245	11.920	5.167
T.E.I OF IONIAN ISLANDS	0.012	1.120	0.214	0.009	0.110	10.120	1.931

The means and standard deviations of the variables – ratios we calculated, are displayed on table 3.

Table 3: Means and standard deviations

	Mean	Standard deviation
PUBLICATIONS_OPERATING_COSTS	.0323	.014694
PUBLICATIONS_TS	1.1686	.339062
PUBLICATIONS_CONTRACT	.5099	.210450
PUBLICATIONS_STUDENTS	.0191	.008722
GRADUATES_OPERATING_COSTS	.2615	.063831
GRADUATES_TS	**10.122**	3.365236
GRADUATES_CONTRACT	4.162	1.142793
GRADUATES_STUDENTS	**.1535**	.042857

The highest rate is of the ratio of graduates per number of TS, with a mean of 10 graduates and standard deviation of 3.5 graduates per member of TS.

The lowest rates regard the ratio of publications per number of students and per amount of operating costs.

Following that, by means of Principal Components Analysis, we detect those, out of the ratios we calculated, which affect the effectiveness of TEI. Instead of a total output/input fraction, used by Data Envelope Analysis, we propose the use of every possible output/input fraction for the variables of output and input, we mentioned in the previous chapter.

The results of Principal Components Analysis, which arose by means of SPSS software, are demonstrated in the following tables 4, 5, 6.

Table 4: Component Matrix

	Component		
	1	2	3
PUBLICATIONS_OPERATING_COSTS	.883	-.207	.093
PUBLICATIONS_TS	**.731**	.066	-.490
PUBLICATIONS_CONTRACT	**.888**	.074	-.382
PUBLICATIONS_STUDENTS	**.898**	-.258	.270
GRADUATES_OPERATING_COSTS	.314	.597	.526
GRADUATES_TS	-.326	**.824**	-.095
GRADUATES_CONTRACT	.386	**.705**	-.360
GRADUATES_STUDENTS	.453	.240	**.697**

Extraction Method: Principal Component Analysis.

As we observe on the table 4 of correlations, the first principal component is mainly related to the number of publications. The second principal component is mostly related to the number of graduates. The third principal component is related to the ratio of graduates to the number of students.

Table 5: Total Variance Explained

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.471	43.382	43.382	3.471	43.382	43.382
2	1.709	21.368	64.750	1.709	21.368	64.750
3	1.367	17.088	81.837	1.367	17.088	81.837
4	.744	9.294	91.132			
5	.668	8.352	99.484			
6	.027	.335	99.819			
7	.013	.161	99.980			
8	.002	.020	100.000			

Extraction Method: Principal Component Analysis.

As we can observe on table 5 of percentage of explained variance, the first principal component interprets 43.4% of total variance. The second principal component explains 21.4% of total variance. The third principal component interprets 17% of total variance. The three first principal components explain a large percentage, almost 82%, of total variance.

Table 6: The coordinates of TEIs in first, second and third component

	1st Component	2nd Component	3rd Component	Combination coordinate using axis gravity coefficient
T.E.I OF ATHENS	1.03778	-0.13829	2.11755	6.26
T.E.I OF CRETE	1.84673	-1.2602	-0.33234	3.80
ALEXANDER TEI OF THESSALONIKI	0.53642	-0.44250	0.43569	1.70
T.E.I OF KAVALA	-0.37687	-0.94721	0.23745	-2.60
T.E.I OF WESTERN MACEDONIA	-0.49792	0.04150	-0.33840	-2.12
T.E.I OF LARISA	-0.45955	-0.39480	1.14803	-0.70
T.E.I OF PATRAS	-1.29807	2.19294	1.07322	0.71
T.E.I OF PIRAEUS	-0.46978	-0.75721	1.21038	-1.27
T.E.I OF SERRES	-0.63456	-0.21413	-0.54397	-3.31
T.E.I OF KALAMATA	1.20677	1.12998	-0.28434	5.73
T.E.I OF MESOLOGGI	-0.37669	0.23811	-0.66230	-1.81
T.E.I OF CHALKIDA	0.33935	1.12862	-1.11302	1.59
T.E.I OF LAMIA	1.25385	0.36910	-1.16377	3.39
T.E.I OF EPIRUS	-0.41766	0.65020	-0.65775	-1.24
T.E.I OF IONIAN ISLANDS	-1.68980	-1.59615	-1.12643	-10.13

On table 6 the coordinates of TEI on the 3 components appear. On the first principal component the TEI of Athens, Thessaloniki, Crete, Kalamata, Lamia are counterpoised on the diagram with the rest , as it is schematically demonstrated on the diagram 1.

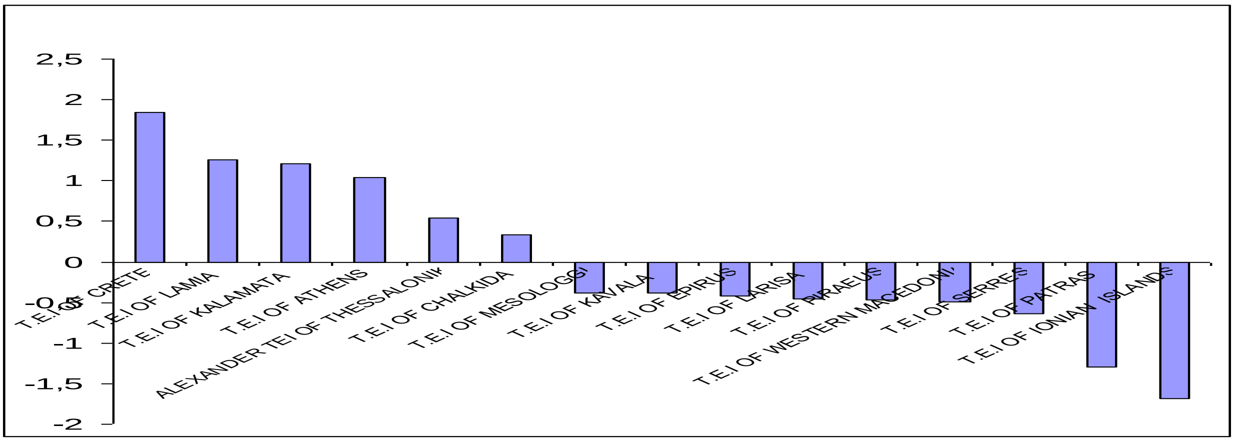

Diagram 1: Ratios of publication (first axis 43.4%)

On the second principal component the TEI of Athens, Thessaloniki, Crete, Kavala, Ionian Islands, Piraeus are placed opposite the others as it is clearly displayed on the diagram 2.

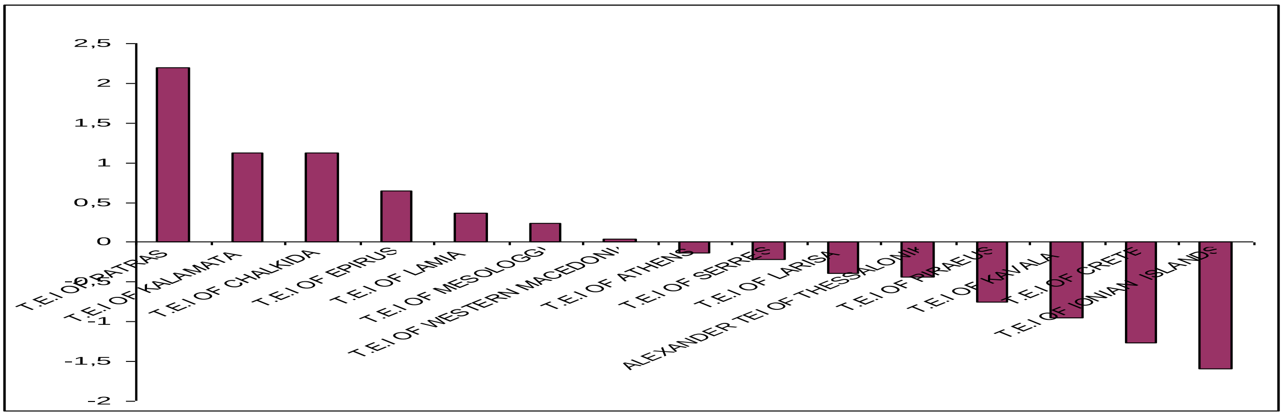

Diagram 2: Ratios of graduates (second axis 21.4%)

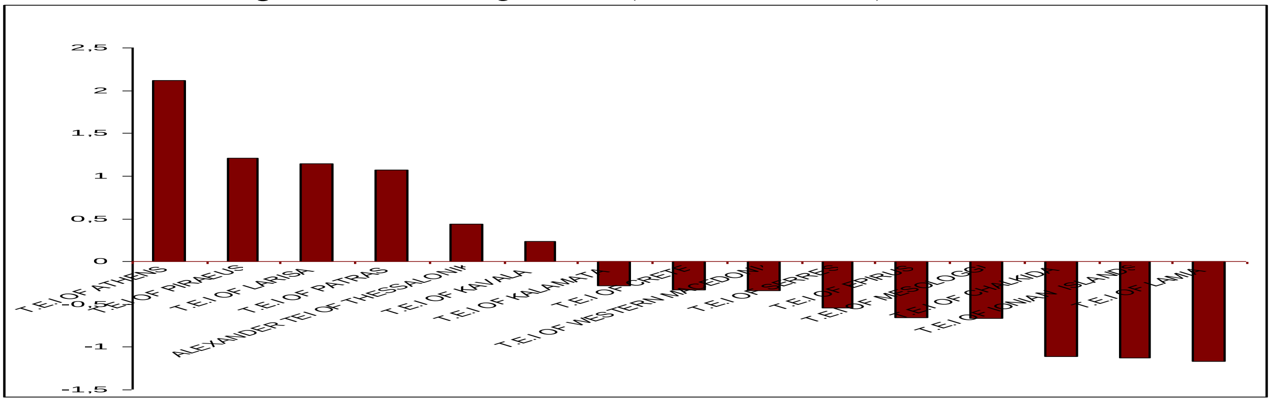

Diagram 3: Ratios of graduates per students (third axis 17%)

On the third principal component the TEI of Athens, Piraeus, Larisa, Patra, Thessaloniki, Kavala are juxtaposed with the rest, as we see schematically on the diagram 3.

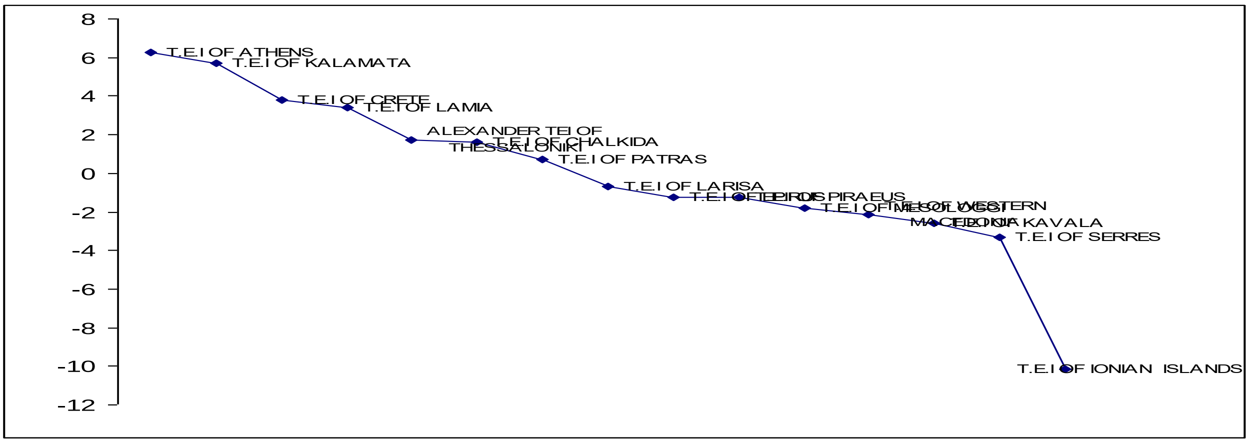

Diagram 4: Total effectiveness ranking of TEI

For the ranking of TEI as far as total effectiveness is concerned, we propose the use of their coordinates in all three principal components. We calculate a combinatory coordinate, taking the three coordinates, multiplying each one with a gravity coefficient. This gravity coefficient results from the variance percentage explained by each principal component. In the last column of the table 6 we have the rates and on the diagram 4 we see TEI ranking on the basis of their total effectiveness.

5. Conclusions

The use of Principal Components Analysis was presented as an alternative for Data Envelope Analysis. The main advantage is easiness, interpretation of each main axis and the related ranking of the points. We compared Greek TEI in terms of their effectiveness using input and output variables.

References

Charnes A., Cooper W.W. & Rhodes E. (1978). The efficiency of decision making units. *European Journal of Operation Research*, Vol. 2, pp. 429-444.

Banker R.D., Charnes A., Cooper W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, Vol. 30, no 19, pp. 1078-1092.

Andersen, P. & Petersen N. (1993). A procedure for ranking efficient units in Data Envelopment Analysis. *Management Science*, Vol. 39, no 10, pp. 1261-1264..

Premachandra, I.M. (2001). A note on DEA vs principal component analysis. An improvement to Joe Zhu's approach. *European Journal of Operation Research*, Vol. 132, pp. 553-560. .

Kardiyen, F. & Orkcu H.H. (2006). The Comparison of Principal Component Analysis and Data Envelopment Analysis in Ranking of Decision Making Units. *Journal of Science*, Vol 19 (2), pp. 127-133. 2006.

EKT National Documentation Centre (2012). Greek Academic Institutes Analysis, http://metrics.ekt.gr/el/report02 (in Greek).

Derpanis, D. (2009). Data Envelope Analysis, Phd Thesis, University of Pireas. (in Greek).

Katsamanis G. (2009). Data Envelope Analysis, Phd Thesis, University, of Crete, Chania, (in Greek).

Katharaki, M. & Katharakis G. (2010). A comparative assessment of Greek universities' efficiency using quantitative analysis, *International Journal of Education Research* 49, pp. 115-128.

Mavris, E., Anastasiadou, S., Florou, G, Efthymios Valkanos, E., Anastasiadis, L. (2019).
Data Evelopment Analysis (DEA): An application of dearegarding Greek tertiary e ducation. *Journal of Business Paradigms, vol 4. No 1*. pp. 78-99.