University of Southampton

Discussion Papers in
Management

Selecting DEA Specifications and Ranking Units via PCA

C. Serrano Cinca
Department of Accounting and Finance
University of Zaragoza, Spain

and

C. Mar Molinero
Department of Management
University of Southampton, UK

Number M01-3

December 2001

ISSN 1356-3548

SELECTING DEA SPECIFICATIONS AND RANKING UNITS VIA PCA

By

C. Serrano Cinca

Department of Accounting and Finance
University of Zaragoza, Spain.
serrano@posta.unizar.es

C. Mar Molinero

Department of Management
University of Southampton, UK.
camm@socsci.soton.ac.uk

This version: November 2001

Address for correspondence: C. Serrano-Cinca: Department of Accounting and Finance, Fac. CC Económicas y Empresariales, Univ. Zaragoza, Gran Vía 2, Zaragoza (50.005) SPAIN serrano@posta.unizar.es

SELECTING DEA SPECIFICATIONS AND RANKING UNITS VIA PCA

ABSTRACT

DEA model selection is problematic. The estimated efficiency for any DMU depends on the inputs and outputs included in the model. It also depends on the number of outputs plus inputs. It is clearly important to select parsimonious specifications and to avoid as far as possible models that assign full high efficiency ratings to DMUs that operate in unusual ways (mavericks). A new method for model selection is proposed in this paper. Efficiencies are calculated for all possible DEA model specifications. The results are analysed using Principal Components Analysis. It is shown that model equivalence or dissimilarity can be easily assessed using this approach. The reasons why particular DMUs achieve a certain level of efficiency with a given model specification become clear. The methodology has the additional advantage of producing DMU rankings.

KEY WORDS

DEA model selection, Data envelopment analysis, efficiency, principal component analysis, cluster analysis.

SELECTING DEA SPECIFICATIONS AND RANKING UNITS VIA PCA

1. INTRODUCTION

Various model selection methods have been suggested in DEA. Most researchers decide "a priori" what the specification of the model should be, without considering any alternatives. But it is possible that a variable included in the model in this way may contribute little or nothing to the calculation of efficiency values. The converse is also true, it is possible that a variable for which data is available, and has not been included in the model on a priori considerations, may be important in the determination of efficiencies. A methodology aimed at guiding model selection in DEA is clearly desirable. Two interesting model selection approaches are due to Norman and Stocker (1991) and to Ruiz et al (2001). Norman and Stocker (1991) assess the need to include a variable by correlating the values of the variable under consideration with efficiency values obtained from the model that excludes it. Ruiz et al (2001) prove that the contribution of a variable to efficiency can be assessed by estimating efficiencies twice, once with the reduced model -which does not include the variable-, and once with the total model -which includes the variable. However, as any empirical study demonstrates, models that appear to be similar are not exactly equivalent. As variables enter or leave the specification, some DMUs become 100% efficient or loose this characteristic. Both methodologies rely, to a certain extent, on judgement for final model selection. This judgement is made with little reference to the original data set, which becomes obscured in a mass of mathematical details. DEA provides, for each DMU, just a score. It is not very informative as to the way in which inputs and outputs contribute to the efficiency calculation. There are many ways of achieving similar levels of efficiency when various inputs and outputs are involved. It is necessary to look beyond DEA, study the reasons why DMUs achieve a certain degree of efficiency, and the reasons why the various models are, or are not, equivalent. Here we propose a methodology based on multivariate statistical analysis.

DEA and multivariate statistical techniques have been used in combination in various studies. Since the number of efficient DMUs depends on the number of inputs and outputs in the model, it is important to control for the number of inputs plus outputs; Pedraja et al (1999). Adler and Golany (2001) use principal components as inputs and outputs, and thus reduce the data that is fed into the DEA model. A very similar approach is followed by Vargas and Bricker (2000). Multivariate statistical methods and DEA are also used in sequence, to obtain a more complete understanding of the data; examples are Mancebon and Mar Molinero (2000), Bradley et al (2001), and Nath (2001). Zhu (1998) uses PCA as an alternative to the DEA model, although he is aware of the limitations of the procedure. Premachandra (2001) demonstrates the soundness of Zhu's approach and extends it.

Zhu's (1998) approach is based on the realisation that a ratio of a single output to a single input can be a feasible solution in the efficiency frontier and that, by studying such ratios, and their linear combinations, it is possible to rank efficient units. This is certainly a new perspective on the ranking of efficient DMUs which has recently been based on the concept of superefficiency introduced by Andersen and Petersen (1993), although other ranking methods have also been proposed; for example, Doyle and Green (1994) proposed a method based on the cross-efficiency matrix; Sinuany-Stern and Friedman (1998) proposed the use of discriminant analysis; and the same authors also put forward a methodology based on canonical correlation analysis; Friedman and Sinuany-Stern (1997). Raveh (2000a) uses the co-plot, a simplified version of Multidimensional Scaling (MDS) methods to rank Greek banks. A description of the co-plot methodology can be found in Raveh (2000b).

Multivariate statistical techniques and DEA can also be used simultaneously as was done by Serrano Cinca et al. (2001). These authors combined MDS and DEA in the context of efficiency in dot.com companies. Data on a set of ratios that combined web metrics –output- and financial information –input- was represented in the form of MDS configurations, on which efficiency ratings were superimposed and used to explore the various strategic objectives of the dot.com companies. Various strategic groups of companies were identified. It was shown that the various strategic groups had different objectives, and that different DEA models were appropriate for each group.

This paper proposes a new approach to guide model selection in DEA and to the ranking of units. The method has the advantage that the ranking extends to inefficient units, something that the method based on superefficiency does not permit. The DEA modelling procedure is embedded in a multivariate statistical framework. The methodology shares with Zhu (1998) the idea that much can be learned from exploring simplifications of the full DEA model within a multivariate statistical framework, although the similarities end here. The procedure proposed attempts to visualise differences and similarities between the efficiencies generated by the various models, by treating efficiencies as variables in a statistical framework.

The model is developed within the context of the data set on Chinese cities published by Zhu (1998) and Premachandra (2000).

The paper is organised as follows. Section 2 describes the data, introduces the notation, and gives the efficiency matrix that will be the object of the modelling procedure. Modelling is described in section 3, which is divided into several subsections to account for the various steps to be followed. The last section contains the conclusions.

2. CASE STUDY: CHINESE CITIES AND DEA.

Table 1 shows the data set on 18 Chinese cities analysed and published by Zhu (1998) and Premachandra (2000). There are two inputs and three outputs defined as follows:

Input 1, (X1): Investment in fixed assets by state-owned enterprises.

Input 2, (X2): Foreign funds actually used

Output 1, (Y1): Total industrial output value

Output 2, (Y2): Total value of retail sales

Output 3, (Y3): Handling capacity of coastal ports

Table 1 about here

The first step in the procedure requires the listing of all possible DEA models that can be derived from possible inputs and outputs. In the Chinese cities example there are two inputs and three outputs, resulting in 21 possible DEA models, each model containing a combination of one or more inputs with one or more outputs. These are shown in Table 2. To make it easy for identification purposes, notation is written in such a way that the inputs and outputs that enter it can be easily identified. For this purposes, the first input, X1, is associated with the letter A in the name; the second input, X2, is associated with the letter B; outputs are associated with numbers in an obvious way. Thus, model A1 in Table 2 contains one input, X1, and one output, Y1. Model A12 contains input X1 and outputs Y1 and Y2. The complete model, the only one selected by Premachandra (2000) is AB123. In fact, both Zhu (1998) and Premachandra (2000) only estimate a DEA model, which is our AB123.

Table 2 about here

Efficiencies from each model were obtained using the constant returns to scale, input oriented version of the algorithm. Table 3 shows the efficiencies obtained.

Table 3 about here

The influence of the model on efficiency can be clearly observed in Table 3. For example, DMU 2 is 100% efficient in twelve models that include output Y3 in their specification (A123, A13, A23, A3, B123, B13, B23, B3, AB123, AB13, AB23 and AB3). But if Y3 is removed from the specification, the efficiency of DMU 2 drops to very low values ranging from 0.11 to 0.33. This was observed by Zhu (1998), who commented: "DMU2 only had a good performance with respect to the ratios that measure the input utilization by the third output".

Something similar could be said about DMU 6 and DMU 10. Sometimes they are 100% efficient and other times they appear to be inefficient. In this case, as in every other, it is possible to scan through Table 3 in search of clues that may explain which inputs or outputs are responsible for the changes. It is, however, desirable to analyse Table 3 in a multivariate analysis context. Models can be treated as variables and efficiencies as observations. The aim is to explore the structure of the data and to visualise its most important features.

It is clear that DMUs 2, 6, 10 are very different even if they all appear to be 100% efficient under the complete model, AB123. Another interesting example is provided by DMUs 1 and 16. A cursory examination of Table 3 suggests that they are not very different, and under the complete model AB123 they both achieve 47% efficiency. Are they similar? If they are not, where are the differences? The method suggested here makes it possible to answer these questions.

It is clear that Table 3 contains much information, but that it also contains redundancy, since some DEA models may be equivalent, and some may contain independent information. Multivariate techniques aimed at data reduction and representation such as PCA and Hierarchical Cluster Analysis (HCA) may be used in this context. This is done in the next section.

3. DEA AND PCA. EFFICIENCIES AS VARIABLES IN A MULTIVARIATE STATISTICAL ANALYSIS.

This section will be concerned with the analysis of Table 3 within a multivariate statistical context. First, PCA will be used as a data reduction technique. Component scores are represented in a graphical form, highlighting the similarities and differences between the DMUs. The second subsection highlights the relationship between DMUs and models, and between models, by means of Property Fitting, a regression-based technique. The third subsection completes the graphical representation by superimposing the results of a hierarchical cluster analysis.

3. 1 PCA

Models in Table 3 have been treated as variables and efficiency ratings as observations, and a PCA exercise has been performed. The minimum value for eigenvalue extraction has been set to 0.8, in line with Joliffe's (1972) recommendation. Three eigenvalues exceeded the 0.8 limit, indicating that three components are sufficient to describe the structure of the data. The first component was by far the most important, accounting for 71.9% of the variability in the data. The addition of the second component increases this percentage to 91.9%, and the addition of the third one takes it to 96.5%. This dominance of the first component is typical of highly correlated variables; Dunteman (1989). For the purposes of this study, the first two components provide an adequate representation of the data. The results are given in Table 4. The relevance of the third component was also studied, but found to add little to the present argument, and is not presented here.

Table 4 about here

Models have been ranked according to their loading in the first principal component. The results are shown in Table 5. All the models have positive loadings in this component. It is to be noticed that the model with the highest loading in the first component is AB123, the complete model that includes two inputs and three outputs. In this kind of situation the first principal component is often taken to be an overall measure of strength of the relationship. It is clear that this component can be interpreted as an overall measure of efficiency. Ranking of DMUs on this component will produce a ranking of all DMUs in terms of efficiency; this ranking includes both efficient and inefficient DMUs.

Table 5 about here

Turning to the second component, it is to be noticed that the only models that load highly on it are B3, AB3 and A3. All these models contain a single output in their specification, Y3. It is interesting to notice that all the models that contain output Y3 have positive loadings in the second component, while those models that exclude Y3 have negative loadings. The second component is clearly associated with the ability that DMUs have of generating output Y3.

Using similar reasoning, it can be argued that the third principal component is related to the efficient use of inputs. Models A2, A23, AB23, AB2, A123, AB123, A12, AB12, A3, and AB3 have positive loadings in the third components. All these models contain input X1 or both inputs. Models B1, B13, AB1, B12, B123, A1, AB13, B23, B2, B3, and A13 have negative loadings in the third component. All but two models contain input X2 or both inputs.

In summary, the first principal component gives an overall measure of efficiency; the second principal component is related to output Y3; and the third principal component is a contrast between input X1 and input X2. For the purposes of this paper, it is sufficient to concentrate on the first two components, and this is what will be done.

For each DMU, component scores for the first and second principal component have been calculated and plotted in a graph. This graph can be seen in Figure 1. The DMUs that achieve efficiency scores of 100% are to be found at the extreme right hand side of the first principal component, as expected. DMU 2 shows its reliance on output Y2 by finding its way to the top of the second principal component. The fact that DMU 2 achieves efficiency by concentrating on output Y3 is now clear. The isolated position of DMU 2 in the figure suggests that we are dealing with a “maverick”. In general, once meaning has been attached to the various components, extreme points can be analysed, particularly efficient extreme points, as this may indicate that the relevant DMUs use an unusual mix of inputs and outputs to achieve efficiency, and this may reveal maverick behaviour. At the other extreme of the first principal component we find DMUs 14, 15, and 18. These DMUs achieve low efficiencies under most models.

Figure 1 about here

3. 2 Property Fitting

The relationship between components and models can be displayed graphically by using the regression-based technique of Property Fitting, or Pro-Fit for short; Schiffman et al (1981). In this technique, vectors are drawn in such a way that, for a particular DEA model, the value of the efficiency derived from the model increases in the direction of the vector. The direction of the vector is calculated as a result of a regression analysis in which the efficiencies derived from the particular model are the dependent variables and the component scores are the independent variables. This technique has the advantage of highlighting up to what point two models are similar, since the angle between any two vectors is related to the correlation between the efficiencies generated by the two models concerned. All the vectors are represented through the centre of coordinates in Figure 1. This is the compass that we use as a guide to interpret the configuration. Vectors are only drawn if the coefficient of determination, R^2, in the relevant regression exceeds a particular value. In this case all models achieved very high values of R^2, the lowest value achieved being 0.81, and all models were represented. All vectors pointed towards the positive side of the first component, forming an open fan, something that indicates that the various ways of achieving efficiency are positively correlated. Table 6 shows the results of all Pro-Fit regressions, and gives directional cosines, γ_1 γ_2 γ_3 ,with their level of significance. The vectors can be seen in Figure 2.

Table 6 about here

Figure 2 about here

The vectors in Figure 2 confirm the conclusions obtained when component loadings were discussed. Vectors group neatly into three groups. One group is formed by models AB3, B3, and A3. All these models achieve their highest value in DMU 2 and contain only output 3, indicating that DMU 2 achieves 100% efficiency by attaching high weights to output 3. The remaining vectors split into two groups, both of them pointing in the positive direction of component 1. One group points towards the positive side of component 2 and the other group points towards the negative side of component 2. The difference between the two groups concerns the presence or absence or output 3 in the specification. The models that do not contain output 3 point downwards, and those that contain output 3 point upwards. Thus, output 3 is crucial in the modelling procedure. The average vector -labelled Mean- has also been calculated and represented, and almost coincides with the axis associated with the first component. In other practical situations one would also look at the projection on other principal components, and this may reveal the different reasons why DMUs achieve a given level of efficiency.

The procedure to select a model is now clear. If the directional vectors form a closed fan, model selection is very simple, as this is an indication that all models are equivalent. In this case one would select the most parsimonious model. If the fan is wide open, we need to explore any groups that may exist and base our model selection on economic considerations as well as on statistical principles. Thus, the fan is the wind rose that guides the DEA sailor through the sea of models. In the present case it is to be first decided whether output 3 should or should not be included in the specification. This is a crucial decision. Models AB3, B3, and A3 do not appear to be reasonable since they favour a maverick DMU, DMU 2, and show the remaining DMUs in bad light, a fact that can be confirmed by inspecting Table 3. If it is decided to leave output 3, in the specification, then any model amongst B23, A13, B13, AB23, B123, A23, B123, A123, AB123 could be chosen. Parsimony would probably favour A23, as it plots in the middle of the fan and, contains only one input and two outputs.

We can now see in which way DMU 1 is different from DMU 16. They both achieve the same efficiency score under the complete model AB123, and have almost identical projections on the first principal component. But DMU 1 plots on the positive side of the second component, indicating that it values output Y3, while DMU 16 plots towards the negative side of the second principal component, indicating that models that ignore output Y3 will favour this DMU. If output Y3 was to be considered important by decision makers, DMU 1 would be preferred to DMU 16.

As far as DMU ranking is concerned, it could be argued that no single model should contribute to the position of a DMU in the list, and that the ranking should take into account all possible specifications. Thus, the ranking along the first principal component would be appropriate. We think that only the first principal component should be involved in the ranking, and not all of them weighted according to the variance they explain, as done by Zhu (1998). The ranking based on the first principal component would produce the following ordering of DMUs: 10, 6, 2, 5, 9, 13, 12, 4, 8, 1, 16, 7, 11, 17, 3, 15, 18, 14. It is to be noticed that this procedure allows for the ranking of all DMUs and not only the efficient ones.

3. 3 Clustering DMUs

A complementary way of analysing the data in Table 3 is to use Cluster Analysis. It is good practice to supplement the results obtained from graphical representations of multivariate data with the superimposition of cluster countours; Arabie et al. (1987). Clusters were obtained using Ward's method, that maximises within group homogeneity and between group heterogeneity. The dendogram is shown in Figure 3. Cluster Analysis shows the presence of five main clusters, one of them containing only DMU 2, which appears yet again as a case on its own. At a higher level of clustering the DMUs divide neatly into two groups, one of them containing DMUs that reach 100% efficiency and the other one containing the DMUs that never do. The clusters are shown in Figure 1.

Figure 3 about here

The extreme position of DMUs 14, 15, and 18 is to be noted. Zhu (1998) comments on these three cities as follows: "these three DMUs were declared by Chinese government as model for economic reforms and developments". Considering the low efficiency levels achieved by these three cities, any directed economic effort has great opportunities for success.

4 CONCLUSIONS

This paper has presented a new method for model selection in DEA based on multivariate statistical analysis. The methodology requires evaluating efficiencies for all possible input/output combinations. It is clear that such methodology produces much redundancy, but also generates valuable information. The matrix of efficiencies by models is then analysed by means of data reduction techniques, such as Principal Components Analysis. Further understanding of the data can be gained by applying Hierarchical Cluster Analysis in this data set.

It has been shown that there are advantages with calculating efficiencies under all possible specifications of the DEA model, and then performing multivariate analysis on the results obtained. Principal Components Analysis has been the chosen technical approach, although Multidimensional Scaling would have been equally appropriate. This methodology permits the joint graphical representation of models and DMUs, and thus it makes it possible to explain up to what point two models are equivalent, and if they are not equivalent, why they are not equivalent. The relationship between models and DMUs becomes clarified. By supplementing the representations with the results of Property Fitting techniques, it is possible to assess why a particular DMU achieves high efficiency scores under some models and low efficiency scores under other models. Maverick DMUs are easily identified. Finally, the method permits the ranking of DMUs. Such ranking includes both efficient and inefficient DMUs.

REFERENCES

Adler, N. and Golany, B. (2001): Evaluation of deregulated airline networks using data envelopment analysis combined with principal components analysis with an application to Western Europe. European Journal of Operational Research, 132, 260-273.

Andersen P. and Petersen N.C. (1993): A procedure for ranking efficient units in data envelopment analysis. Management Science, 39, 1261-1264.

Arabie P., Carroll J.D. and De Sarbo W.S. (1987): Three way scaling and clustering. Sage University Paper Series on Quantitative Applications in the Social Sciences. Number 07-065. Beverley Hills. Sage Publications.

Bradley, S., Johnes, G, and Millington, J (2001): The effect of competition on the efficiency of secondary schools in England. European Journal of Operational Research, 135, 545-568.

Doyle J.R. and Green R (1994): Efficiency and cross-efficiency in DEA: derivations, meanings and uses. Journal of the Operational Research Society, 45, 567-578.

Dunteman G.H. (1989): Principal Component Analysis. Sage Publications Ltd. London, UK.

Friedman L. and Sinuany-Stern Z (1997): Scaling units via the canonical correlation analysis in the DEA context. European Journal of Operational Research, 100, 629-637.

Joliffe. I.T. (1972): Discarding variables in Principal Components Analysis. Applied Statistics, 21, 160-173.

Mancebon, M.J. and Mar Molinero, C. (2000): Performance in primary schools. Journal of the Operational Research Society, 51, 843-854.

Nath, P. (2001): “Efficiency Analysis of Indian Banking Industry - An Exploratory Study”, APORS’2000, The Fifth Conference of the Association of Asian-Pacific Operations Research Societies within IFORS, 5-7 July 2000, Singapore

Norman, M. and Stocker, B. (1991): Data Envelopment Analysis: the assessment of performance. John Wiley and Sons. Chichester. UK.

Pedraja Chaparro. F.; Salinas Jimenez J.. and Smith. P. (1999): On the quality of the Data Envelopment Analysis model. Journal of the Operational Research Society. 50. 636-645.

Premachandra, I M (2001): A note on DEA vs principal component analysis: An improvement to Joe Zhu's approach, European Journal of Operational Research, (132): 3, pp: 553-560

Raveh A. (2000a): The Greek banking system: reanalysis of performance. European Journal of Operational Research, 120, 525-534.

Raveh A. (2000b): Co-plot: a graphic display method for geometrical representations of MCDM. European Journal of Operational Research, 125, 670-678.

Ruiz Gomez, J.L; Pastor, J. and Sirvent, I. (2001): A Statistical Test for Nested Radial DEA Models. Operations Research, forthcoming.

Schiffman, J.F., Reynolds, M.L. and Young, F.W. (1981): Introduction to Multidimensional Scaling: Theory, Methods and Applications. Academic Press, London.

Serrano Cinca C., Mar Molinero C., and Fuertes Callen Y. (2001): The path to efficiency in DEA . Multidimensional Scaling as a tool for post-optimality analysis. Discussion papers in Management, number M01-1. School of Management. University of Southampton. UK.

Sinuany-Stern Z. and Friedman L. (1998): DEA and the discriminant analysis of ratios for ranking units. European Journal of Operational Research, 111, 470-478.

Vargas, S.C. and Bricker, D (2000): Combining DEA and Factor Analysis to improve the evaluation of academic departments given uncertainty about the output constructs. Research paper, Department of Engineering, University of Iowa, Iowa City, USA.

Zhu, Joe (1998): Data envelopment analysis vs. principal component analysis: An illustrative study of economic performance of Chinese cities European Journal of Operational Research, (111): 1, pp: 50-61

DMU	*x1*	*x2*	*y1*	*y2*	*y3*
1	2874.8	16738	160.89	80800	5092
2	946.3	691	21.14	18172	6563
3	6854	43024	375.25	144530	2437
4	2305.1	10815	176.68	70318	3145
5	1010.3	2099	102.12	55419	1225
6	282.3	757	59.17	27422	246
7	17478.6	116900	1029.09	351390	14604
8	661.8	2024	30.07	23550	1126
9	1544.2	3218	160.58	59406	2230
10	428.4	574	53.69	47504	430
11	6228.1	29842	258.09	151356	4649
12	697.7	3394	38.02	45336	1555
13	106.4	367	7.07	8236	121
14	4539.3	45809	116.46	56135	956
15	957.8	16947	29.2	17554	231
16	1209.2	15741	65.36	62341	618
17	972.4	23822	54.52	25203	513
18	2192	10943	25.24	40267	895

Table 1: Values taken by two inputs and three outputs on 18 Chinese cities. Source: Zhu (1998) and Premachandra (2001).

DMU	*INPUT*	*OUTPUT*
A1	X1	Y1
A12	X1	Y1 Y2
A123	X1	Y1 Y2 Y3
A13	X1	Y1 Y3
A23	X1	Y2 Y3
A2	X1	Y2
A3	X1	Y3
B1	X2	Y1
B12	X2	Y1 Y2
B123	X2	Y1 Y2 Y3
B13	X2	Y1 Y3
B23	X2	Y2 Y3
B2	X2	Y2
B3	X2	Y3
AB1	X1 X2	Y1
AB12	X1 X2	Y1 Y2
AB123	X1 X2	Y1 Y2 Y3
AB13	X1 X2	Y1 Y3
AB23	X1 X2	Y2 Y3
AB2	X1 X2	Y2
AB3	X1 X2	Y3

Table 2. The 21 DEA models used in the study

DMU	*A1*	*A12*	*A123*	*A13*	*A23*	*A2*	*A3*	*B1*	*B12*	*B123*	*B13*	*B23*	*B2*	*B3*	*AB1*	*AB12*	*AB123*	*AB13*	*AB23*	*AB2*	*AB3*
1	27	28	47	47	44	25	26	10	10	12	12	8	6	3	27	28	47	47	44	25	26
2	11	17	100	100	100	17	100	33	33	100	100	100	32	100	33	33	100	100	100	32	100
3	26	26	28	28	21	19	5	9	9	9	9	4	4	1	26	26	28	28	21	19	5
4	37	37	50	50	41	28	20	17	17	19	19	10	8	3	37	37	50	50	41	28	20
5	48	54	63	59	58	49	17	52	52	53	53	34	32	6	59	59	63	63	58	49	17
6	100	100	100	100	88	88	13	84	84	84	84	44	44	3	100	100	100	100	88	88	13
7	28	28	36	36	26	18	12	9	9	10	10	4	4	1	28	28	36	36	26	18	12
8	22	33	50	41	49	32	25	16	16	19	19	17	14	6	22	33	50	41	49	32	25
9	50	50	63	63	48	35	21	53	53	55	55	26	22	7	60	60	66	66	48	35	21
10	60	100	100	66	100	100	14	100	100	100	100	100	100	8	100	100	100	100	100	100	14
11	20	24	30	27	28	22	11	9	9	10	10	7	6	2	20	24	30	27	28	22	11
12	26	59	79	52	79	59	32	12	16	19	15	19	16	5	26	59	79	52	79	59	32
13	32	70	75	43	75	70	16	21	27	28	22	28	27	3	32	70	75	43	75	70	16
14	12	13	14	14	12	11	3	3	3	3	3	2	1	0	12	13	14	14	12	11	3
15	15	18	19	16	17	17	3	2	2	2	2	1	1	0	15	18	19	16	17	17	3
16	26	46	47	30	47	46	7	4	5	5	4	5	5	0	26	46	47	30	47	46	7
17	27	27	31	31	27	23	8	2	2	2	2	1	1	0	27	27	31	31	27	23	8
18	5	17	20	10	20	17	6	2	4	5	3	5	4	1	5	17	20	10	20	17	6

Table 3: DEA efficiencies for DMU under the 21 DEA models. Efficiencies vary between 0 and 100.

Component	*Eigenvalue*	*% of variance*	*Cumulative%*
PC1	15.10	71.88	71.88
PC2	4.21	20.05	91.93
PC3	.95	4.51	96.45
PC4	.61	2.91	99.37

Table 4. PCA results.

	Component		
	PC1	*PC2*	*PC3*
AB123	.964	.141	.197
A123	.963	.143	.205
AB13	.962	.173	-.122
B123	.942	.204	-.246
AB23	.937	.187	.290
A23	.937	.187	.290
B13	.932	.207	-.284
B12	.923	-.263	-.256
B1	.913	-.256	-.303
B2	.894	-.170	-.100
AB12	.893	-.411	.146
A13	.891	.281	-
AB1	.882	-.366	-.266
B23	.874	.345	-.116
AB2	.872	-.374	.288
A12	.827	-.525	.195
A2	.815	-.487	.293
A1	.730	-.530	-.187
B3	.397	.896	-
AB3	.438	.890	-
A3	.438	.890	-

Table 5. Component loadings. Models are ordered on the first component.

	Directional cosines			*F*	*Adj R2*
	g_1	g_2	g_3		
A1	0.79 (7.04)**	-0.58 (-5.11)**	-0.20 (-1.80)	26.3	0.817
A12	0.83 (59.91)**	-0.53 (-38.03)**	0.20 (14.14)**	1745.0	0.997
A123	0.97 (36.88)**	0.14 (5.47)**	0.21 (7.86)**	484.0	0.988
A13	0.95 (9.57)**	0.30 (3.02)**	-0.08 (-0.82)	33.8	0.853
A23	0.94 (81.34)**	0.19 (16.22)**	0.29 (25.20)**	2504.6	0.998
A2	0.82 (28.19)**	-0.49 (-16.85)**	0.30 (10.15)**	393.8	0.986
A3	0.44 (16.90)**	0.89 (34.36)**	0.09 (3.30)**	492.2	0.989
B1	0.92 (35.58)**	-0.26 (-9.98)**	-0.30 (-11.80)**	501.6	0.989
B12	0.93 (30.74)**	-0.26 (-8.75)**	-0.26 (-8.52)**	364.8	0.985
B123	0.95 (33.57)**	0.21 (7.29)**	-0.25 (-8.79)**	419.2	0.987
B13	0.94 (38.74)**	0.21 (8.60)**	-0.29 (-11.82)**	571.4	0.990
B23	0.92 (10.14)**	0.36 (4.00)**	-0.12 (-1.35)	40.2	0.874
B2	0.98 (8.31)**	-0.19 (-1.58)	-0.11 (-0.93)	(24.2	0.803
B3	0.40 (8.27)**	0.91 (18.65)**	-0.08 (-1.65)	139.6	0.961
AB1	0.89 (25.27)**	-0.37 (-10.49)**	-0.27 (-7.63)**	268.9	0.979
AB12	0.90 (29.62)**	-0.41 (-13.64)**	0.15 (4.84)**	362.3	0.985
AB123	0.97 (32.79)**	0.14 (4.81)**	0.20 (6.69)**	381.1	0.99
AB13	0.98 (21.09)**	0.18 (3.80)**	-0.12 (-2.68)*	155.5	0.96
AB23	0.94 (81.34)**	0.19 (16.22)**	0.29 (25.20)**	2504.6	1.00
AB2	0.88 (25.36)**	-0.38 (-10.89)**	0.29 (8.38)**	277.4	0.98
AB3	0.44 (16.90)**	0.89 (34.36)**	0.09 (3.30)**	492.2	0.99

** Significant at the 0.01 level

* Significant at the 0.05 level

Table 6. Pro-Fit Analysis. Linear regression results

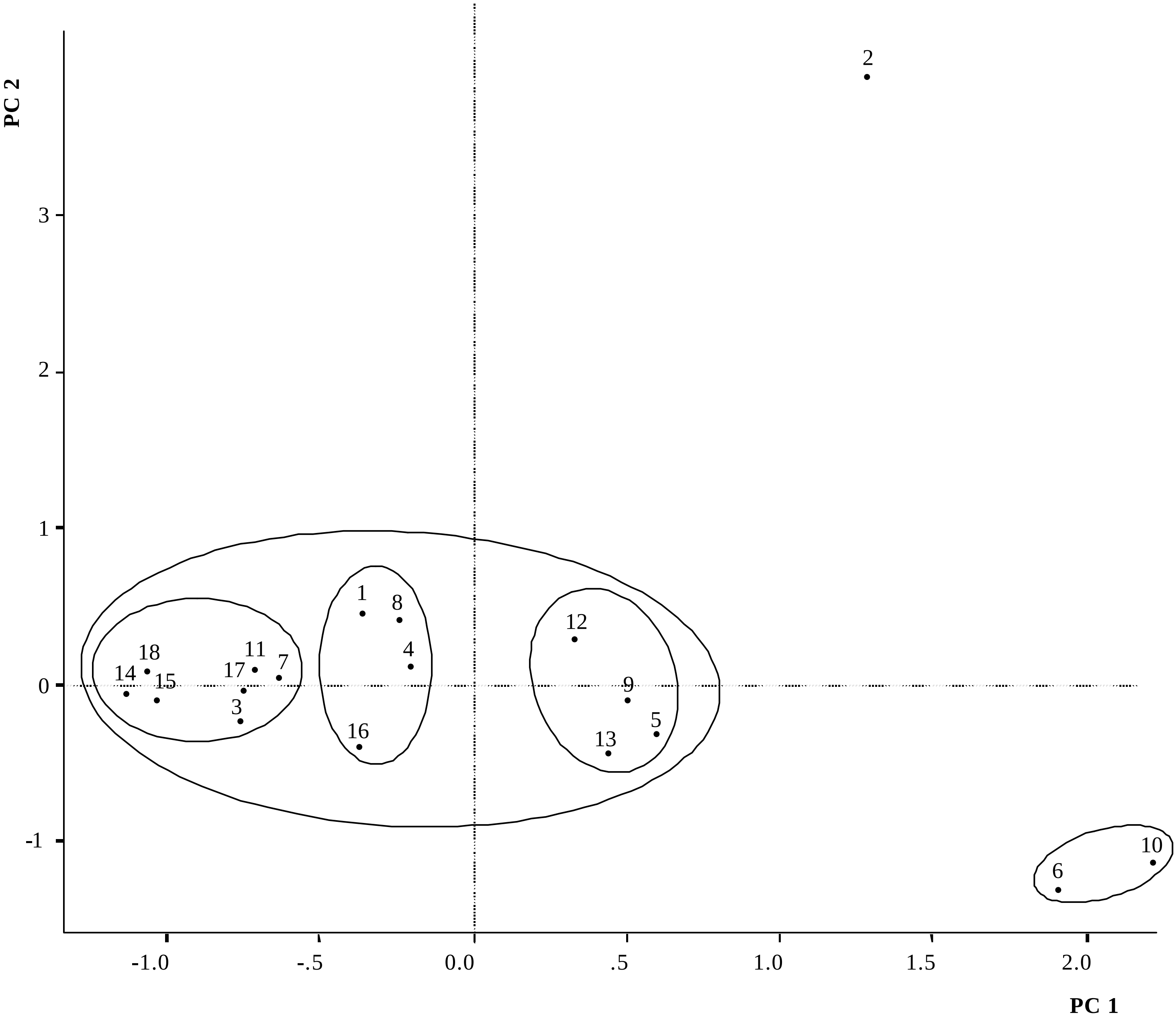

Figure 1. Component scores for the first and second principal component

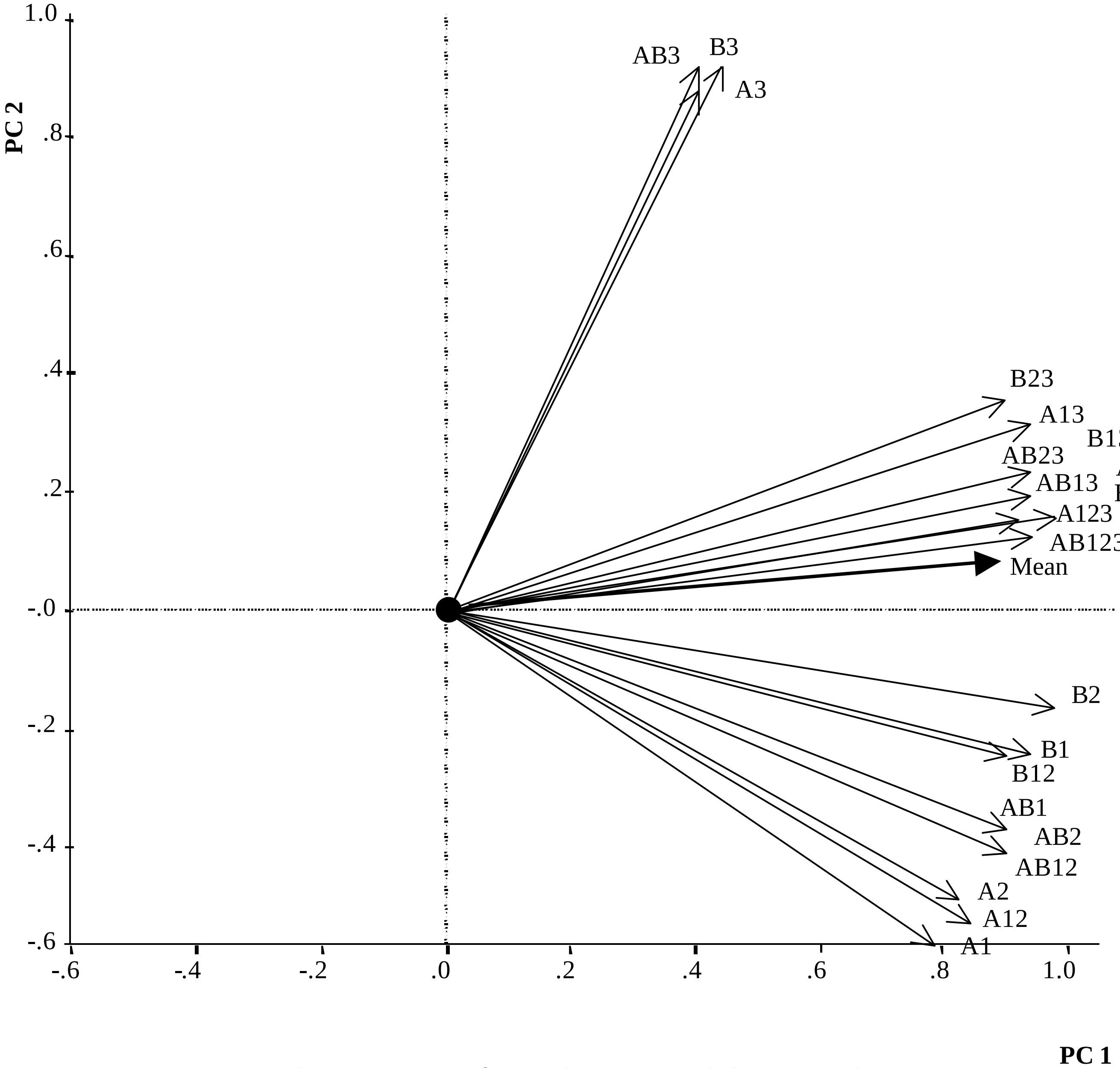

Figure 2: ProFit Analysis. Vectors for each DEA model. PC1 and PC2.

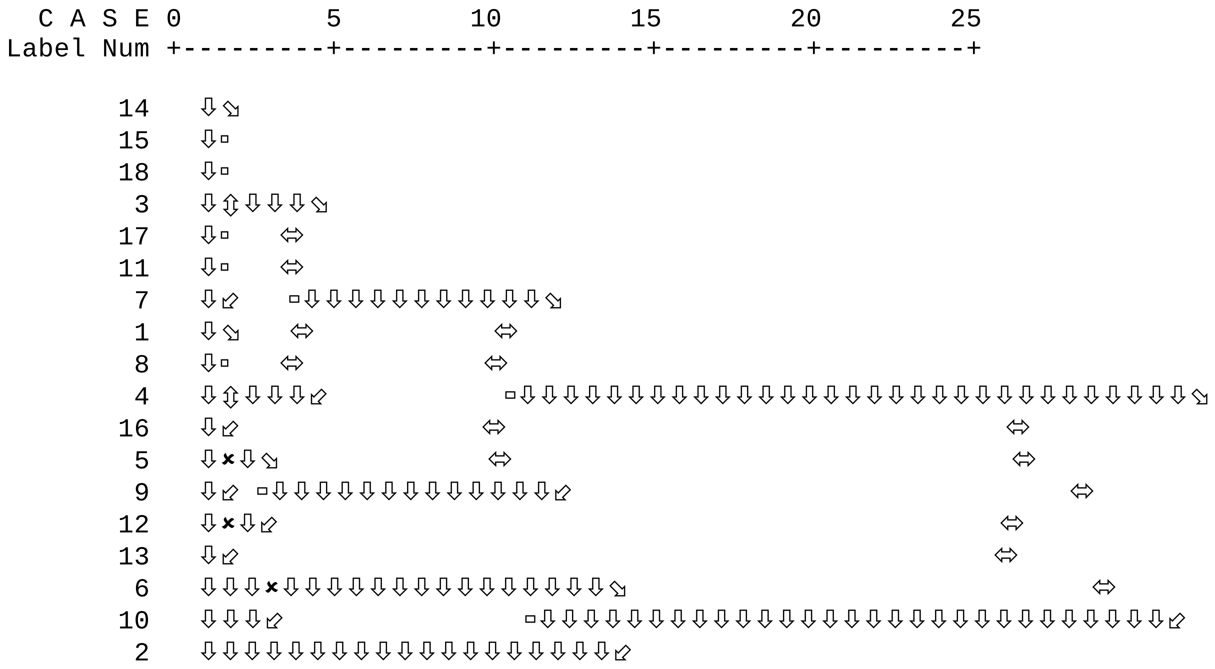

Figure 3. Dendrogram using Ward Method