CLUSTERING IN A DATA ENVELOPMENT ANALYSIS USING BOOTSTRAPPED EFFICIENCY SCORES

Joseph G. Hirschberg and Jenny N. Lye

Economics, University of Melbourne

Melbourne, 3010

Australia

July 2001

This paper explores the insight from the application of cluster analysis to the results of a Data Envelopment Analysis of productive behaviour.

Cluster analysis involves the identification of groups among a set of different objects (individuals or characteristics). This is done via the definitions of a distance matrix that defines the relationship between the different objects, which then allows the determination of which objects are most similar into clusters. In the case of DEA, cluster analysis methods can be used to determine the degree of sensitivity of the efficiency score for a particular DMU to the presence of the other DMUs in the sample that make up the reference technology to that DMU. Using the bootstrapped values of the efficiency measures we construct two types of distance matrices. One is defined as a function of the variance covariance matrix of the scores with respect to each other. This implies that the covariance of the score of one DMU is used as a measure of the degree to which the efficiency measure for a single DMU is influenced by the efficiency level of another. An alternative distance measure is defined as a function of the ranks of the bootstrapped efficiency. An example is provided using both measures as the clustering distance for both a one input one output case and a two input two output case.

1. Introduction

Data envelopment analysis (DEA) is a widely used method in the area of efficiency measurement. Charnes, Cooper and Rhodes have written numerous papers and monographs since the origin of this technique with their 1978 publication. The 1995 bibliography of DEA compiled by Tim Anderson at the University of Oregon (which can be found at www.emp.pdx.edu) lists over 360 papers that use DEA methods through up until then and there has been much more growth in the rate of use since then. The 2000/4 EconLit CD references over 230 journal articles that use some form of DEA in the economics literature alone.

DEA has been applied in a number of different areas that have previously been very hard to assess. These areas include health care (hospitals, doctors), education (schools, universities), banks, manufacturing, benchmarking, management evaluation, energy efficiency, fast food restaurants, and retail stores. A recent reference on the subject can be found in Cooper et al (2000).

In its most commonly used form, DEA is used to compute a score which defines the relative efficiency of a particular decision making unit (DMU) versus all other DMUs observed in the sample. A DMU could be any level of operation that has a distinct set statistics that describe its inputs and outputs. However, unlike the traditional stochastic frontier methods of production and cost function estimation as proposed in Aigner et al (1977), DEA does not require monetary valued inputs, a single output, nor does it rely on assumptions of a particular functional form or a particular statistical distribution. Thus for example, one can measure outputs as the number of a certain kind of patients and inputs as the number of hospital beds without the need to establish a market prices or an algebraic formulation of inputs that generated outputs.

The objective of this paper is to investigate the use of the bootstrapped DEA efficiency scores in the interpretation of the results of a particular analysis. Typically DEA efficiency scores are reported and used in summarisations with no corresponding measures of statistical reliability. Ferrier and Hirschberg (1997, 1999) show that DEA efficiency scores can be bootstrapped so that one can compute statistics relating to the individual DEA scores. The matrix of bootstrapped DEA scores will be used to develop new techniques for the interpretation and graphic display of DEA results through the methods of Cluster analysis.

2. DEA

We can define the DEA process as follows (see Ali and Seiford 1993, Färe, Grosskopf and Lovell 1994, Cooper Seiford and Tone 2000). Assume that there is a sample of T DMUs (eg. organisations, facilities, etc.), each producing an m-dimensional vector of outputs, y, from an n-dimensional vector of inputs, x. Technology governs the transformation of inputs into outputs; the reference technology relative to which efficiency is assessed is given by the input requirement set $L(y) = \{x\text{: } x \text{ can produce } y\}$. Farrell's (1957) input-based measure of technical efficiency for each observation $t = 1,...,T$ is given by:

$$\mathrm{TE}_t\left(x_t, y_t\right) = \min\left\{ \mu_t : \mu_t \bullet x_t \in L\left(y_t \right) \right\};$$

Thus the t^{th} DMU's observed input vector (x_t) is reduced by a scalar ($0 \leq \mu_t \leq 1$) until it is still just able to produce the observed level of output (y_t). The solution, $\mathrm{TE}_t = \mu_t^*$, gives the proportion of the observation's actual input vector that is technologically necessary to produce its observed output vector given the best-practice technology as revealed by the observed data. The vector $x_t^* = \mu_t^* \cdot x_t$ would give the technically efficient ("optimal") input vector for the t^{th} DMU.

One way to calculated this measure of technical efficiency is by solving the following linear programming problem once for each DMU $t = 1,...,T$ (see Färe, Grosskopf and Lovell (1985)):

$$\begin{gathered} \text{Min } z, \mu_t, \text{ subject to :} \\ z \bullet Y \;\geq\; y_t, z \bullet X \;\leq\; \mu \bullet x_t, \\ z_s \;\geq\; 0,\;\; s = 1,\ldots,T,\; \sum_{s=1}^{T} z_s \;=\; 1, \end{gathered}$$

Where Y is the m by T matrix of the observed outputs of all DMUs, X is the n by T matrix of the observed inputs for all DMUs, and z is a T dimensional vector of weights. These weights form a convex combination of observed DMUs relative to which the subject DMU's efficiency is evaluated. The constraints in this problem simply describe the input requirement set as given by the observed data (ie. the best-practice technology). This specification is the variable rate of return case for minimising the cost. Increasingly DEA has been used for a number of alternative forms such as: maximising output, imposing constant returns to scale, modelling panel data, and allowing for the inclusion of fixed inputs.

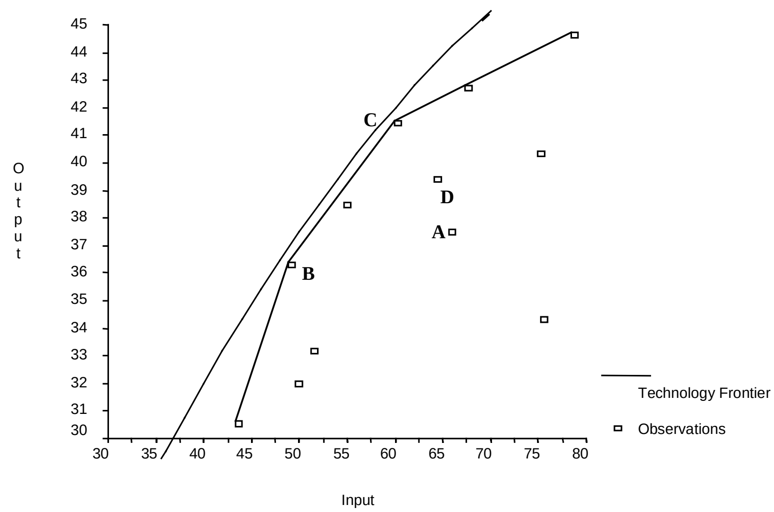

Figure 1. A one input one output technology

In the case where there is one input (x_t), one output (y_t) and cost is minimised, the DEA measure of inefficiency (μ_t^*) is the reduction in input that would furnish the same level of output. A DEA analysis for a particular DMU is performed using the (y_t, x_t) combination to define the technology of each DMU in the sample. DEA can be demonstrated using a graphic solution to the one input one output case. From Figure 1 one can see an actual production the frontier as the smooth relationship and the data for a set of DMUs.

In the case of variable returns to scale, DEA approximates the production technology with a piece-wise linear function based on the observed DMUs that are the most efficient as shown in Figure 1. Thus the computation of the DEA score for DMU **A** (y_A, x_A) is defined by the ratio of the hypothetical value of the input ($\hat{x}_A$) if **A** was on the estimated frontier to the actual value of x_A. For DMU **A**, **B** and **C** act as the reference DMUs that define the efficient technology and $\hat{x}_A$ is defined by value of x on the line segment connecting DMUs **B** (y_B, x_B) and **C** (y_C, x_C) at y_A

(thus $y_A \approx 37.4$, $x_A \approx 66.0$, and $\hat{x}_A \approx 50.5$). Note that DMU **B** is drawn from those cases where $y_B \leq y_A$ and DMU **C** is drawn from those cases where $y_C > y_A$. Thus the estimated efficiency score is formed as a ratio $\hat{\mu}_A = \hat{x}_A / x_A$ and in this case $\hat{\mu}_A \approx 50.5/66 = .765$. This implies that by reducing inputs by 15.5 units DMU **A** would be on the efficiency frontier. In the case of multiple inputs the definition of $\hat{\mu}$ involves the equivalent contraction of all inputs in equal proportions so that they touch the equivalent piece-wise technology.

3 Bootstrapping DEA

The bootstrap is a method by which repeated resampling of a single data set is done to construct an empirical distribution for the target statistic (Efron 1979). Artificial, or "pseudo-samples" are created from the actual observed data in such a way that they are representative of the observed data in dimension and distribution. Then the statistic of interest is recalculated on the basis of each pseudo-sample. The resulting bootstrapped values of the statistic are then used to construct a sampling distribution for the statistic of interest. Because the procedure is not based on the assumption of a particular distribution and is created solely by using the observed observations the investigator literally "pulls themselves up by their own bootstraps". The recent literature on the bootstrap includes a number of monographs Efron and Tibshirani 1993, Davidson and Hinkley 1997, and Chernick 1999.

The most widely suggested type of bootstrap for DEA is a form of the bootstrap commonly used in the analysis of regression equations referred to as a "conditional" bootstrap. A conditional bootstrap makes a model assumption first and thus the resampling is done once a part of the data generating process has been assumed. In the bootstrap of regression this means the equation specification is determined first and the resampling only involves the estimated errors or residuals (see Freedman and Peters 1984). In the case of DEA one creates the pseudo-sample for the reference technology first (see Ferrier and Hirschberg 1997, 1999). In the one-input-one-output variable-returns-to-scale cost minimizing case, as shown in Figure 1, this is done by first making all the DMUs efficient, except the one of interest, by reducing their inputs to the levels that would place them on the piecewise-estimated frontier. Then the pseudo-technology is created by picking an efficiency score for each DMU from the actual scores with replacement to generate new levels of inputs. Thus the pseudo-technology could look exactly like the original observed technology if for each DMU one happened to choose the score for that DMU. Once the pseudo-technology is defined the efficiency score for the particular DMU in question is computed where the inputs for this DMU remain at the observed levels. A matrix (**Q**) of dimension $B \times T$ of bootstrapped efficiency scores is generated from this process (i.e. B pseudo-samples are generated for each of the T observations in the data set).

A modification to this conditional approach has been proposed by Simar and Wilson (1998) in which a smoothed distribution of efficiency scores is drawn from instead of the actual distribution. This modification has been suggested to reduce the discontinuous nature of the distribution of efficiency scores especially in small samples however it relies on the need to assume smoothness properties that may be inappropriate (see Ferrier and Hirschberg 1999 for more details). Löthgren and Tambour (1999) also propose a conditional bootstrap method however in their model not only does the reference technology change with each bootstrap subsample but the technology of the DMU of interest changes as well. The advantage to this modification is that they are able to make probability statements about those DMUs that are at the edges of the frontier which neither of the other methods are able to furnish. However this method implies that the observed level of inputs for the DMU of interest is of no informational value.

As in regression, an alternative to conditional bootstrapping of DEA is nonconditional or the "resampling of cases" method. In the resampling of cases the DEA is performed on pseudo-samples formed by simply drawing with replacement from the rows of the data matrix, where the values of inputs and outputs are recorded on the columns, to form a new matrix of exactly the same length as the original. As opposed to the conditional method where a new score for each DMU is obtained in each replication, in the resampling of cases no score may be computed for any particular DMU in any particular replication. In the resampling of cases method it is inevitable to draw multiple copies of a particular DMU's row from the original data while not drawing any rows from other DMUs. However, with enough replications the resampling of cases method will produce a set of representative value for each DMU. Xue and Harker (1999) use such a method in the case when the primary interest is not in individual scores but in aggregates values such as parameters in used in post-DEA regressions. Post-DEA regressions are often run when the individual DMUs are sampled and additional information is also collected that can be used as regressors when the efficiency score is used as the dependent variable.

Note that all of the resampling methods allow for the assumption of different models used in the DEA - such as whether constant or variables returns are assumed or if the DMUs are cost minimising or output maximising.

4. Clustering of the DMUs Based on Bootstrap Replications and Other Distance Measures.

Once the DEA analysis has been performed it may be difficult to interpret the scores obtained for each DMU. One method for obtaining some measure of the interrelationship between the DMUs was to construct what has been referred to as the "envelope map" which is a *T* by *T* matrix with checks for the case when the technology of a DMU is used as a reference for any other (see Cooper Seiford and Tone 2000 chapter 2). Unfortunately, only the reference technologies are included and thus this says nothing about the DMUs that are both off the frontier. Cluster analysis applied to the matrix of bootstrapped DEA scores will be used to improve on this deficiency.

Cluster analysis involves the identification of groups among a set of different objects (individuals or characteristics) see Kaufman and Rousseeuw (1990) for an overview of these methods. This is done via the definitions of a distance measure between the different objects which then allows the determination of which objects are most similar. Clustering has been used for a variety of applications in economics. In some cases the groupings consider the combination of observation units such as individuals, industries, nations, or time periods (for example see Hirschberg and Aigner 1987, Hirschberg and Slottje 1994, Hirschberg and Dayton 1996 and

Borland et al 2000). While in other cases this method is used to combine characteristics as measured across a set of observations such as measures of economic welfare (see Hirschberg Maasoumi and Slottje 1991, 2000a, 2000b, and Slottje et al 1991). Once these distances are defined the clustering method can operate in a stepwise manner to form agglomerative clusters or divisive clusters (this method is most widely applied and is referred to as hierarchical clustering). Alternatively, if the number of clusters is known then an iterative process can be used to find the group memberships that lead to the most homogeneous clusters (a method referred to as k-means clustering). Although the hierarchical method can lead to clusters that are predetermined by the previous clusters used to form them it lends itself to the ability to construct the dendrogram or tree diagram that provides the genealogy of the clusters as they form. Thus providing a method for the interpretation of a matrix of distances.

In the case of DEA, cluster analysis methods can be used in two ways. In the first, it can show the degree of sensitivity of the efficiency score for a particular DMU to the presence of the other DMUs in the sample that make up the reference technology to that DMU. In the one input - one output case variable returns cost minimizing case, as shown in Figure 1, the interrelationships between the various DMUs is rather obvious from the graphic display. By inspection it can be seen that DMUs **B** and **C** make up the reference technology of a number of DMUs such as **A** and **D**. However, when the DEA is performed in the case of multiple inputs and outputs it is far harder to establish the relative interrelationships especially between DMUs that are not either on or off the frontier. To show this relationship the distance matrix can be made a function of the variance covariance matrix of the scores (Ω). This implies that the covariance of the score of one DMU to another provides a measure of the degree to which the efficiency measure for a single DMU is influenced by the efficiency level of another. This type of analysis results in a way of showing the interactions between the DMUs that distinguishes the technologies. The distance we use in this case is defined as a function of the correlation matrix that is defined as $D_c(i, j) = \left(1 - \left|\hat{\rho}_{ij}\right|\right)$. Figure 2 plots the inverted "v" shape of the relationship between the correlation and the distance metric.

Note that most large correlations between bootstrapped DEA scores that are positive.

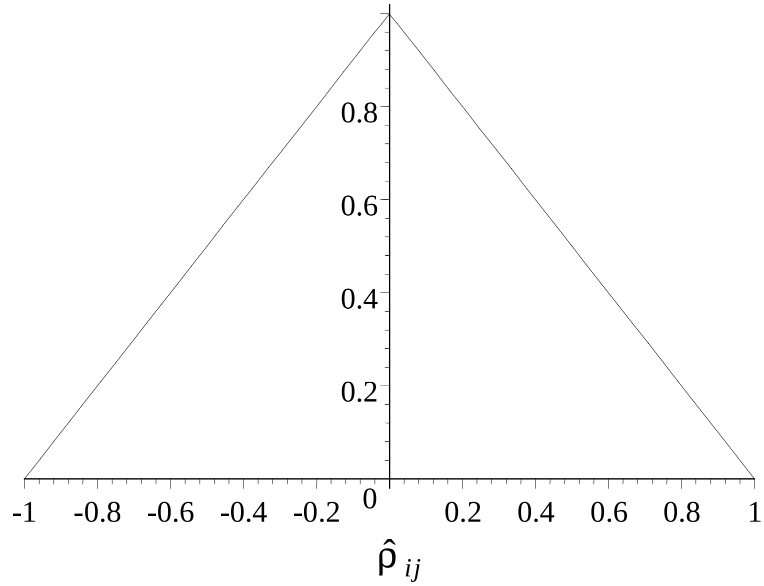

Figure 2. $D_c(i, j)$ by $\hat{\rho}_{ij}$.

The second form of cluster analysis that can be applied to DEA results involves the use of cluster analysis to compare the efficiency score of each DMU to the other scores. In the typical application of DEA to DMUs that are identified the analyst is interested in the relative rank of the DMUs. The point estimates for each DMU are ranked, however the robustness of these ranks is not well established. No analysis is done to determine if a change in the reference technology would change the rankings. The bootstrap results in a series of scores that can then be ranked. One method to proceed would be to determine average rankings over the bootstrap subsamples and test for the differences in rankings via a test of paired differences or a form of sign test. The formal tests are of use only if one is interested in those cases where one can "reject the null hypothesis" that two DMUs are equivalent. However it may be more reasonable to determine gradations of value that show how probable the ordering between two DMUs may change. We can use the results of the bootstrap to construct a distance metric based on the probability that the order of two DMUs may be reversed. We have defined a simple measure as $D_e(i, j) = 2\left(\left|P(\mu_i > \mu_j) - ½\right|\right)$, where the μ_j is the efficiency scores for DMU *j* and μ_i is the efficiency scores for DMU *i*. Note that $D_e(i, j) = 0$ when there is zero probability that $\mu_i > \mu_j$ or a probability of one that $\mu_i > \mu_j$. And

$D_e(i, j) = 1$ when there is a .5 probability that $\mu_i > \mu_j$. Thus the distance is a "v" shaped relationship of $P(\mu_i > \mu_j)$ as shown in Figure 2.

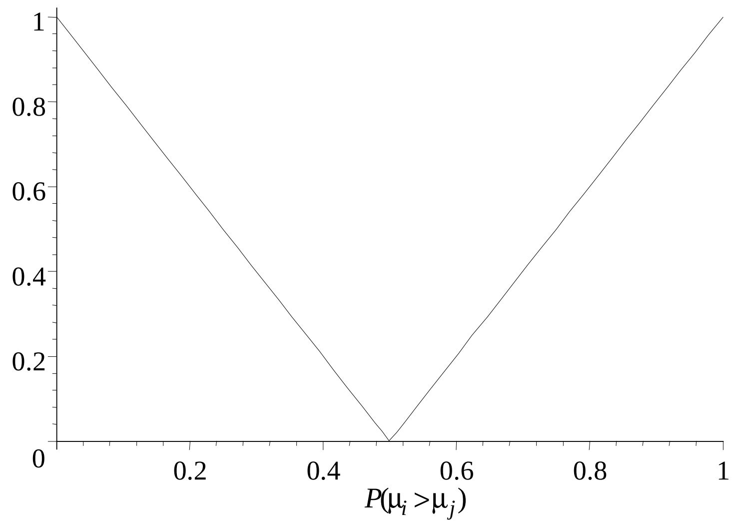

Figure 3. $D_e(i, j)$ by $P(\mu_i > \mu_j)$.

5. An Example Application

The example application is drawn from the data reported in Xue, M. and P. T. Harker (1999), which lists the levels of two outputs (the number of patient days, number of patient discharges) and two inputs (Number of full-time employees and total expenses in $) for a set of 100 hospitals. In these applications 1000 bootstrap pseudo-samples are drawn. The bootstrap pseudo-samples were drawn using the balanced bootstrap sampling proposed by Davidson et al (1986) which insures that each DMU has an equal probability of being drawn over the set of 1000 replications.

5.1 One Input - One Output Technology.

In the first example we apply the DEA to a case in which we have one output (PTDAYS the number of patient days) and one input (FTE the number of full-time employees). Figure 4 is a plot of the DMUs for this case. Note that DMUs 3, 11, and 19 define the frontier for the case where we assume variable returns to scale and because they are at the extremes of the technology and there are no reference DMUs to them. The draw back of this form of bootstrap is that we cannot infer any more information for these DMUs.

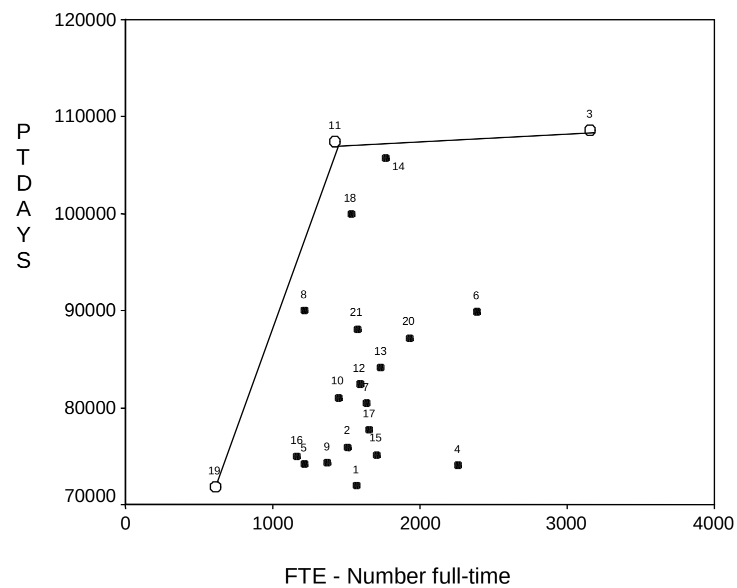

Figure 4. The 1 input 1 output technology.

5.1.1 Clusters Based on the Correlation Based Metric.

Using the correlation based distance metric 100 x $D_c(i,j)$ we compute a distance matrix listed in Table 1 based on the bootstrapped scores.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
1	0	8	92	3	3	92	36	89	3	41	97	51	63	91	6	5	16	94	31	81	85
2	8	0	88	3	2	78	19	76	2	23	92	32	45	96	1	1	3	98	41	65	70
3	92	88	0	90	90	76	82	82	90	81	95	80	78	90	89	89	85	89	94	77	77
4	3	3	90	0	0	88	29	85	0	34	96	44	57	93	1	1	10	96	37	76	81
5	3	2	90	0	0	87	29	85	0	33	95	44	57	93	1	1	9	97	37	76	80
6	92	78	76	88	87	0	44	4	87	39	72	29	17	56	83	84	64	37	91	5	3
7	36	19	82	29	29	44	0	45	28	0	83	4	11	91	24	25	7	78	56	29	34
8	89	76	82	85	85	4	45	0	84	41	68	32	22	53	81	81	63	32	87	11	8
9	3	2	90	0	0	87	28	84	0	33	95	43	56	93	1	1	9	97	37	75	80
10	41	23	81	34	33	39	0	41	33	0	81	2	8	88	28	29	10	75	58	25	30
11	97	92	95	96	95	72	83	68	95	81	0	79	76	47	94	94	88	57	94	74	73
12	51	32	80	44	44	29	4	32	43	2	79	0	3	83	38	39	18	68	64	16	20
13	63	45	78	57	57	17	11	22	56	8	76	3	0	76	51	52	29	60	72	6	10
14	91	96	90	93	93	56	91	53	93	88	47	83	76	0	94	94	100	24	96	67	64
15	6	1	89	1	1	83	24	81	1	28	94	38	51	94	0	0	6	99	39	71	76
16	5	1	89	1	1	84	25	81	1	29	94	39	52	94	0	0	7	99	39	72	76
17	16	3	85	10	9	64	7	63	9	10	88	18	29	100	6	7	0	90	45	50	55
18	94	98	89	96	97	37	78	32	97	75	57	68	60	24	99	99	90	0	99	50	46
19	31	41	94	37	37	91	56	87	37	58	94	64	72	96	39	39	45	99	0	84	87
20	81	65	77	76	76	5	29	11	75	25	74	16	6	67	71	72	50	50	84	0	1
21	85	70	77	81	80	3	34	8	80	30	73	20	10	64	76	76	55	46	87	1	0

Table 1. Distance matrix in based on 100 x $D_c(i,j)$

Using the distance matrix in Table 1 based on $D_c(i,j)$ we perform a hierarchical cluster analysis using the total linkage method. This means that once the distance matrix is formed the closest DMUs are combined to form the first cluster then the next two closest. Once a cluster is formed it is necessary to define a distance to the cluster from other clusters and other individual DMUs. In the case of total linkage this distance is defined as the maximum distance between any two DMUs in the clusters under consideration for combination. The hierarchical clustering algorithm proceeds to form clusters until all the DMUs are included in one cluster. The distance required to form the next cluster is measured on the horizontal axis. Thus we find that DMUs 4, 5 and 9 are the closest. From Figure 4 one can see that these DMUs all line up as having very similar levels of output thus implying that the reference technology would be similar for all of these. Other such clusters made up of DMUs with close proximity would be those made up of DMUs #15 and #16, as well as DMUs #7 and #10.

As mentioned above, an advantage of the hierarchical clustering method is that it allows for the examination of the relationship between the DMUs using a dendrogram. The dendrogram shows the genealogy of the clusters as they are formed. Figure 5 is the dendrogram for the cluster analysis using $D_c(i,j)$ and it shows how the clusters are formed as well as the distance needed to form the clusters. From this diagram we note that the three clusters that we could find by inspection of the distance matrix are all given as clusters with low distances to form by the dendrogram.

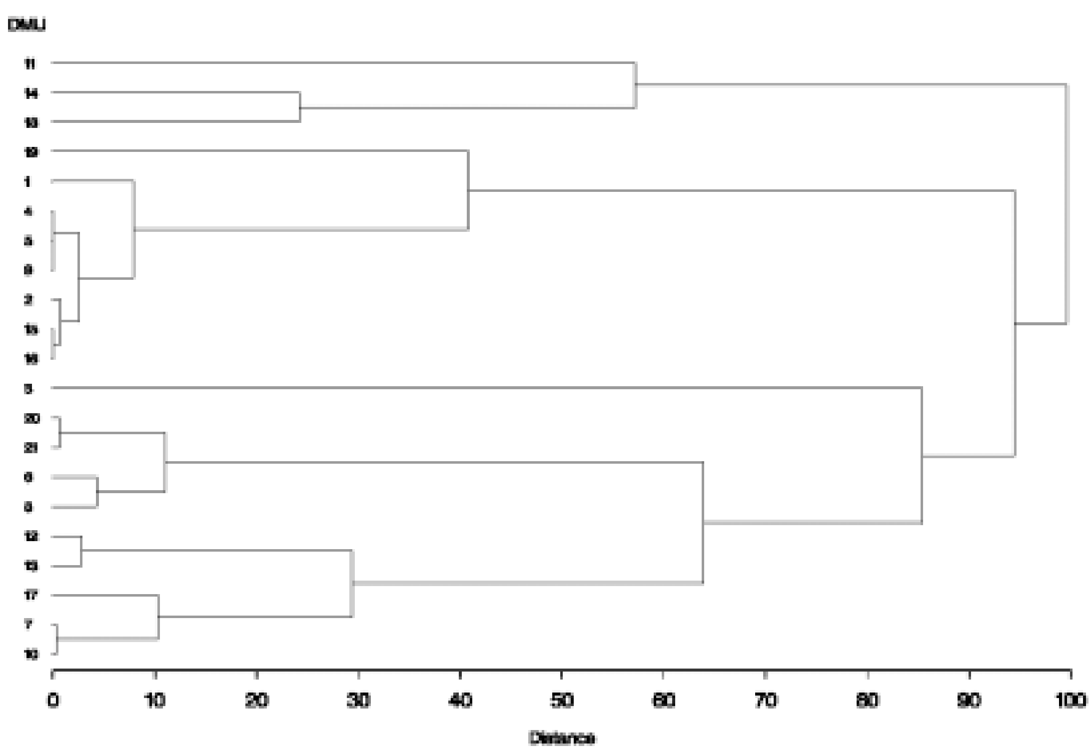

Figure 5. Dendrogram based on $D_c(i,j)$ for the 1 input 1 output case.

In order to have a better view of the distances needed to form the clusters we can plot the distance to form the last cluster as a function of the number of clusters. When each DMU is the only member of its cluster then we have 21 clusters. When we have 1 cluster the distance is the maximum. The shape of this relationship provides an indication of the number of clusters that may be formed before the distances or dissimilarities become "too great". How great is "too" may need to be determined case by case.

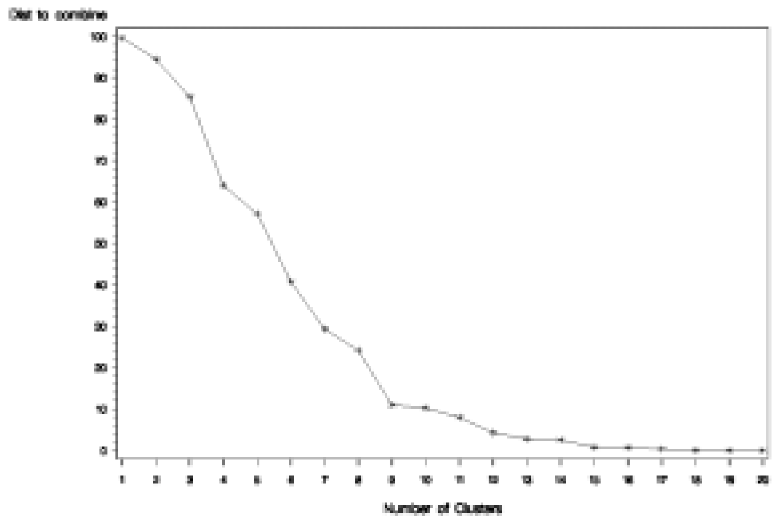

Figure 6. Change in $D_c(i,j)$ by cluster number for the 1 input 1 output case.

One method proposed is to view Figure 6 in the same way one would view the "scree diagram" of eigen values when performing a principal component analysis. Thus one examines the data for an elbow in the graph which in this case appears to occur between 8 and 9 clusters. This would imply that one might stop at 9 clusters. At 9 clusters the distance to the next cluster is just slightly above .1 which indicates that the correlation of approximately .9 (since all the correlations over .5 are positive in this case). From Figure 5 this would mean that the clusters listed in Table 2 can be defined from the bottom of Figure 5 to the top.

Cluster	**Members**
1	10, 7 , 17
2	12, 13
3	8, 6, 21, 20
4	3
5	16, 15, 2, 9, 5, 4, 1
6	19
7	18
8	14
9	11

Table 2. Cluster Membership of the 9 Clusters using $D_c(i,j)$.

Referring back to Figure 4 one can see that the DMUs that make up these clusters are those that one could have predicted by examining which DMUs are similar in levels of output. However, in a higher dimensional DEA it would be much more difficult to make the relationship.

5.1.2 Clusters Based on the Efficiency Score.

Table 3 lists the efficiency scores by DMU ordered from least efficient to most. From this table we note that the least efficient DMU is #4 which the score indicates could produce the same output with 29% of the input used if it was producing on the frontier. We also note that DMU #8 is the most efficient hospital not on the frontier. In this case we find that it could produce the same output on the frontier with 85% of the input used. Note however that these scores could well be off due to the inefficiency of those DMUs on the frontier and due to the misspecification of the technology. The misspecification of the technology could be accounted for by the inclusion of additional inputs and outputs.

DMU	*4*	*1*	*15*	*6*	*17*	*2*	*9*	*7*	*20*	*13*	*12*	*5*	*10*	*16*	*21*	*14*	*18*	*8*	*3*	*11*	*19*
score	0.29	0.39	0.4	0.43	0.45	0.47	0.48	0.49	0.5	0.51	0.53	0.55	0.57	0.58	0.62	0.78	0.82	0.85	1	1	1

Table 3. Scores by DMU for the 1 input 1 output case.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
1	0	100	100	100	100	32	81	100	100	100	100	94	85	100	47	100	64	100	100	80	99
2	100	0	100	100	100	40	56	100	95	100	100	84	70	100	100	100	91	100	100	64	97
3	100	100	0	100	100	100	100	93	100	100	79	100	100	86	100	100	100	87	80	100	100
4	100	100	100	0	100	99	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
5	100	100	100	100	0	82	90	100	100	41	100	65	66	98	100	100	100	99	100	60	71
6	32	40	100	99	82	0	67	100	56	94	100	90	91	100	60	89	30	100	100	98	100
7	81	56	100	100	90	67	0	100	14	100	100	100	82	100	100	96	100	100	100	70	100
8	100	100	93	100	100	100	100	0	100	100	80	100	100	15	100	100	100	0	70	100	100
9	100	95	100	100	100	56	14	100	0	80	100	57	52	100	100	100	94	100	100	36	94
10	100	100	100	100	41	94	100	100	80	0	100	97	94	100	100	72	100	100	100	80	95
11	100	100	79	100	100	100	100	80	100	100	0	100	100	100	100	100	100	100	80	100	100
12	94	84	100	100	65	90	100	100	57	97	100	0	83	100	100	80	100	100	100	70	100
13	85	70	100	100	66	91	82	100	52	94	100	83	0	100	97	83	93	100	100	57	100
14	100	100	86	100	98	100	100	15	100	100	100	100	100	0	100	98	100	23	4	100	100
15	47	100	100	100	100	60	100	100	100	100	100	100	97	100	0	100	100	100	100	92	100
16	100	100	100	100	100	89	96	100	100	72	100	80	83	98	100	0	100	98	100	77	56
17	64	91	100	100	100	30	100	100	94	100	100	100	93	100	100	100	0	100	100	85	100
18	100	100	87	100	99	100	100	0	100	100	100	100	100	23	100	98	100	0	22	100	100
19	100	100	80	100	100	100	100	70	100	100	80	100	100	4	100	100	100	22	0	100	100
20	80	64	100	100	60	98	70	100	36	80	100	70	57	100	92	77	85	100	100	0	100
21	99	97	100	100	71	100	100	100	94	95	100	100	100	100	100	56	100	100	100	100	0

Table 4. Distance matrix in based on 100 x $D_e(i,j)$.

Table 4 is the distance matrix based on the efficiency score based comparisons defined by 100 x $D_e(i,j)$. The 100's indicate those DMUs that never switch in rank over any of the bootstrapped DEA scores.

Figure 7 is the dendrogram for this case and Figure 8 is the plot of the distances to cluster for this clustering. Note that there are 9 clusters that never have overlapping scores. The relationship between the DMUs in these clusters is much less obvious when examining the plot of the data in Figure 4. The DMUs that are the closest in this metric are those where the slope of the line between them is parallel to the reference technology.

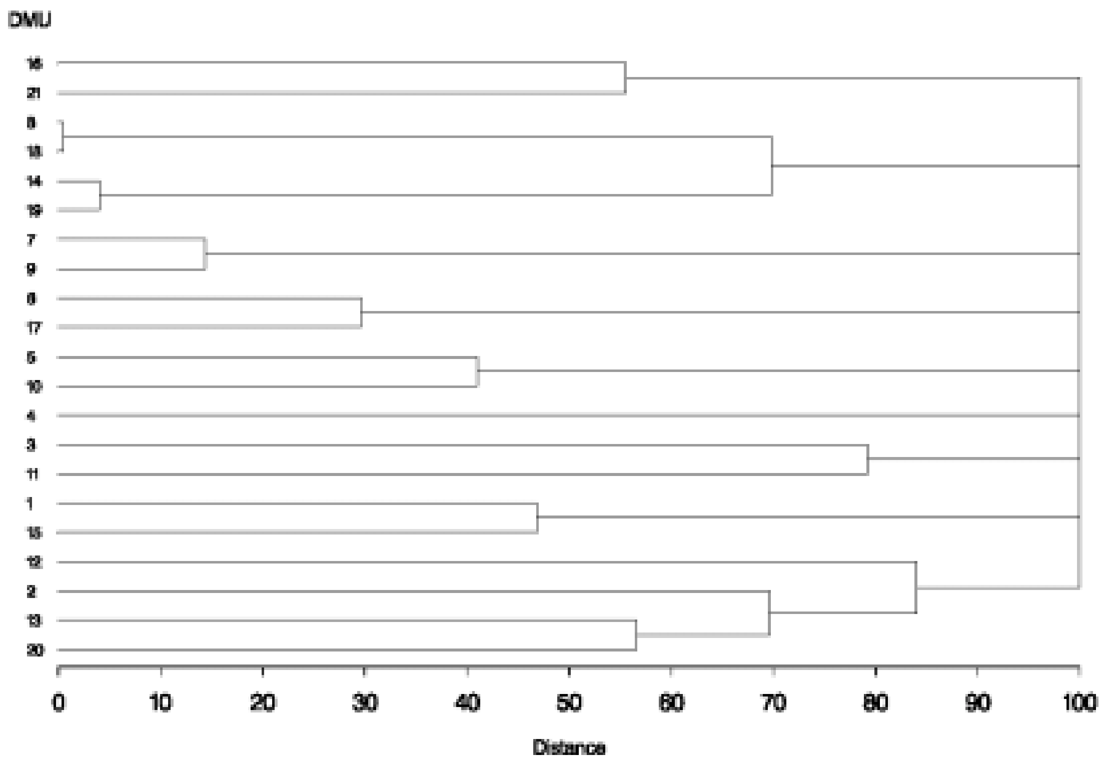

Figure 7. Dendrogram based on $D_e(i,j)$ for the 1 input 1 output case.

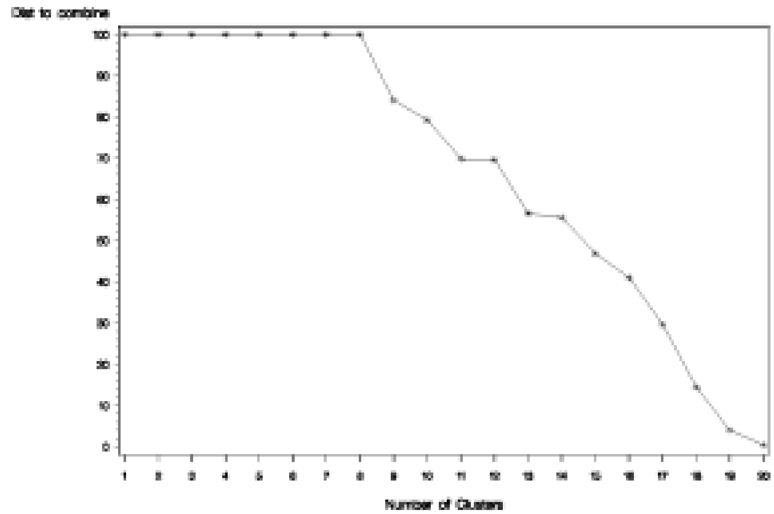

Figure 8. Change in $D_e(i,j)$ by cluster number for the 1 input 1 output case.

The 9 clusters based on the $D_e(i,j)$ metric are given in the Table 5. Interestingly it can be seen that there is no cluster membership that is shared with the earlier cluster set found using $D_c(i,j)$.

Cluster	Members
1	20, 13, 2, 12
2	15, 1
3	11,3
4	4
5	10, 5
6	9, 7
7	14, 19
8	8, 18
9	16, 21

Table 5. Cluster Membership of the 9 Clusters from $D_e(i,j)$.

5.2 Two Input - Two Output Technology.

In this example we use the same DMUs as before but now we define the technology in the DEA by two outputs (the number of patient days y_1, number of patient discharges y_2) and two inputs (Number of full-time employees x_1 and total expenses in \$ x_2). Because this technology is defined in $\mathbb{R}_4^+$ space and we no longer have a simple method for graphing the relationship between the inputs and outputs the results of the cluster analysis become one of the only ways we can evaluate the results. Figures 10 and 11 provide scatter plots for these inputs and outputs

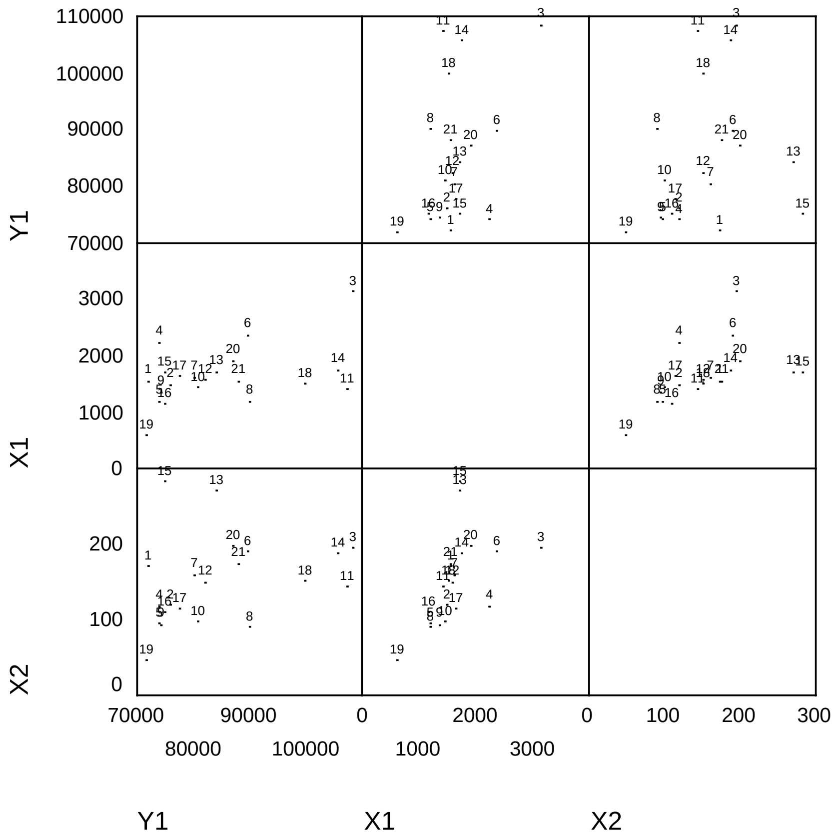

Figure 10. The two inputs Number of full-time employees x_1 and total expenses in \$ x_2) via the first output (the number of patient days y_1).

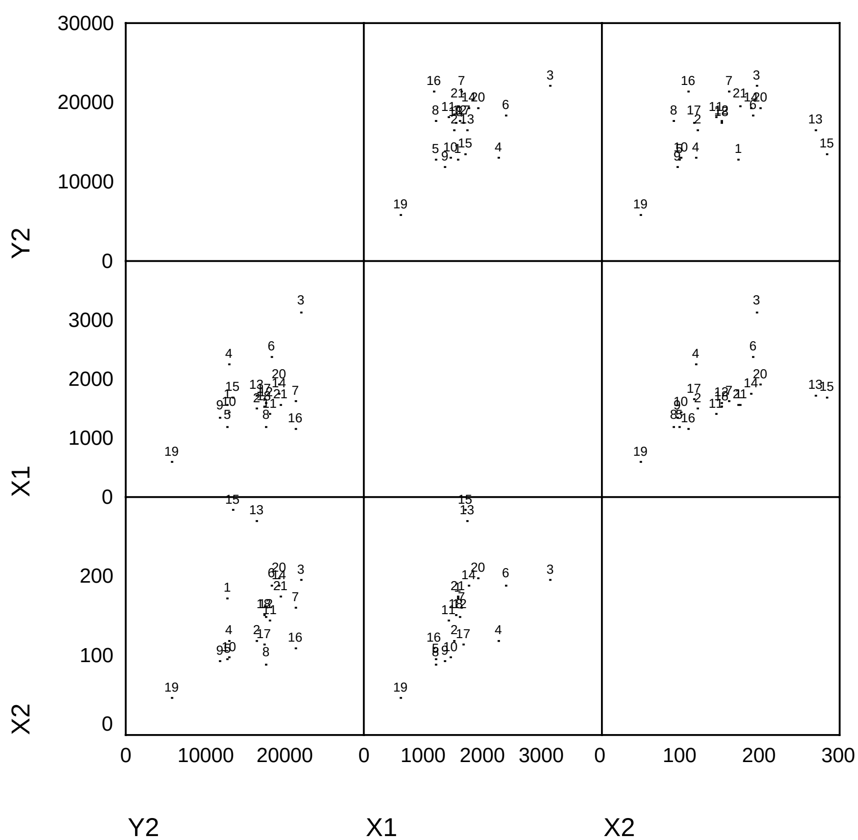

Figure 11. The two inputs Number of full-time employees x_1 and total expenses in \$ x_2) via the second output (number of patient discharges y_2).

Figure 12 provides the dendrogram for the $D_c(i,j)$ metric in the 2 input 2 output case. This result is much less clear cut as the 1 input 1 output case dealt with above. In this case the correlations are not as extreme as the case given above. Here almost all the DMUs scores are more interrelated due to the change in dimensionality. Figure 13 lists the changes in distance to combine the clusters in this case and from this plot there is little to suggest a stopping point for the clustering after 15 clusters. This implies that there are a small number of DMUs that are highly correlated with the rest with smaller interactions.

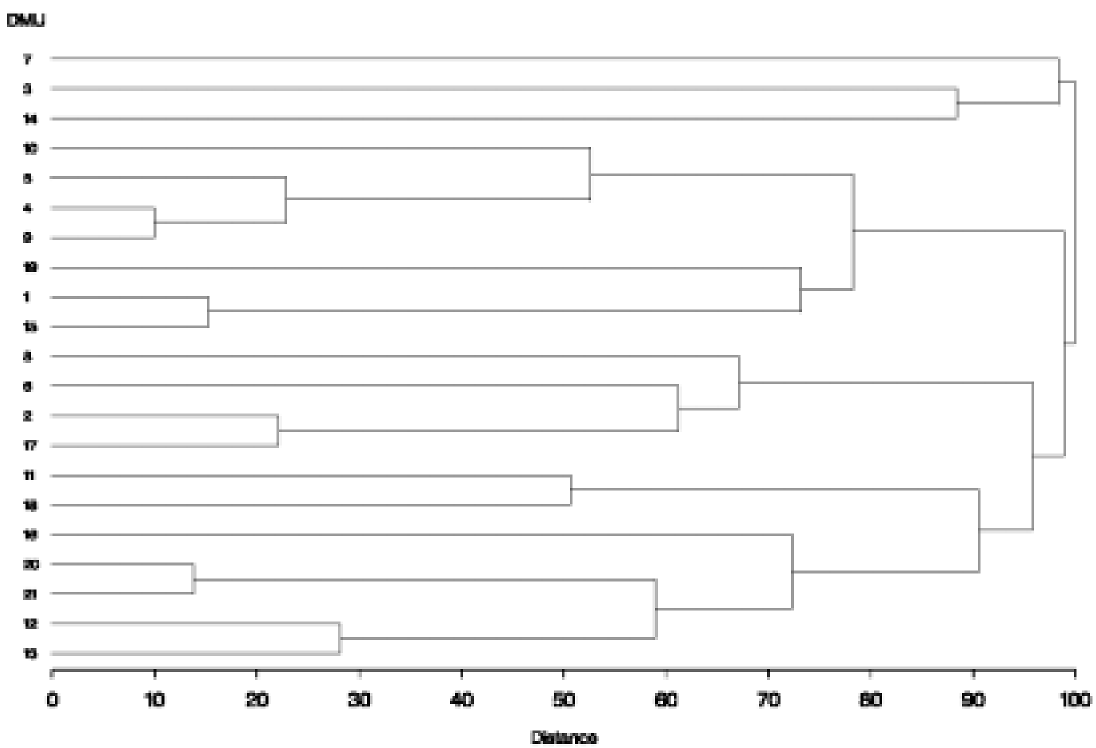

Figure 12. Dendrogram based on 100 x $D_c(i,j)$ for the 2 input 2 output case.

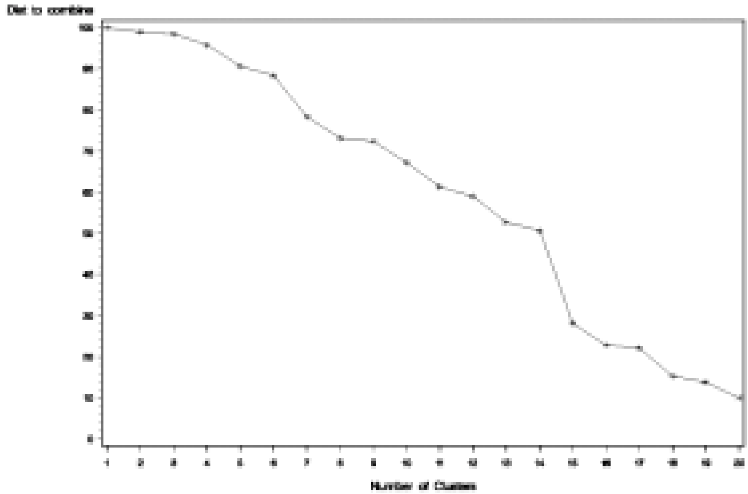

Figure 13. Change in 100 x $D_c(i,j)$ by cluster number for the 2 input 2 output case.

Figure 14 lists the average bootstrap efficiency scores for each of the DMUs in this example. From this we see that on average the inefficiency is less than in the 1 input 1 output case. This is a dimensionality problem with DEA. The average efficiency score for a problem with the same number of DMUs will increase as we include more inputs and more outputs due to a form of overfitting in DEA. Thus this phenomenon is a bit like the increase in R^2 in a multiple regression when the number of parameters increases when the sample size remains fixed. In the case of DEA as the number of inputs and outputs is increased all the DMUs will be on the frontier although the relationship is not as simple as the degrees of freedom computation in regression. Also note that we now have 6 DMUs that have scores of 1.

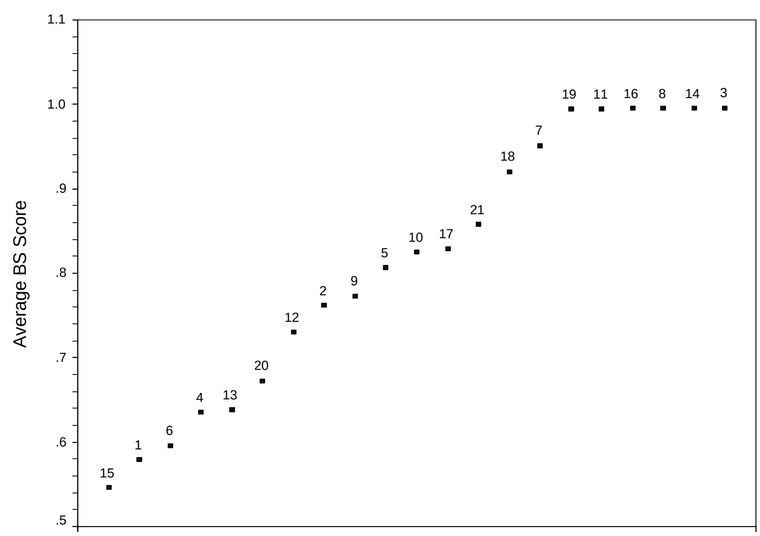

Figure 14. Average BS efficiency scores by DMU for the 2 input - 2 output case.

Figure 15 is the dendrogram we obtain when using the $D_e(i,j)$ distance measure for the 2 input 2 output case. Note that unlike the 1 input 1 output version given in Figure 7, this diagram shows that many more of the DMUs have scores that change in rank. From Figure 16 the plot of the changes in distance by number of clusters we find that there are only 5 clusters of DMUs in which the probability of the rank being reversed is less than 5%. At 14 clusters there is a group of DMUs that have scores that are close together.

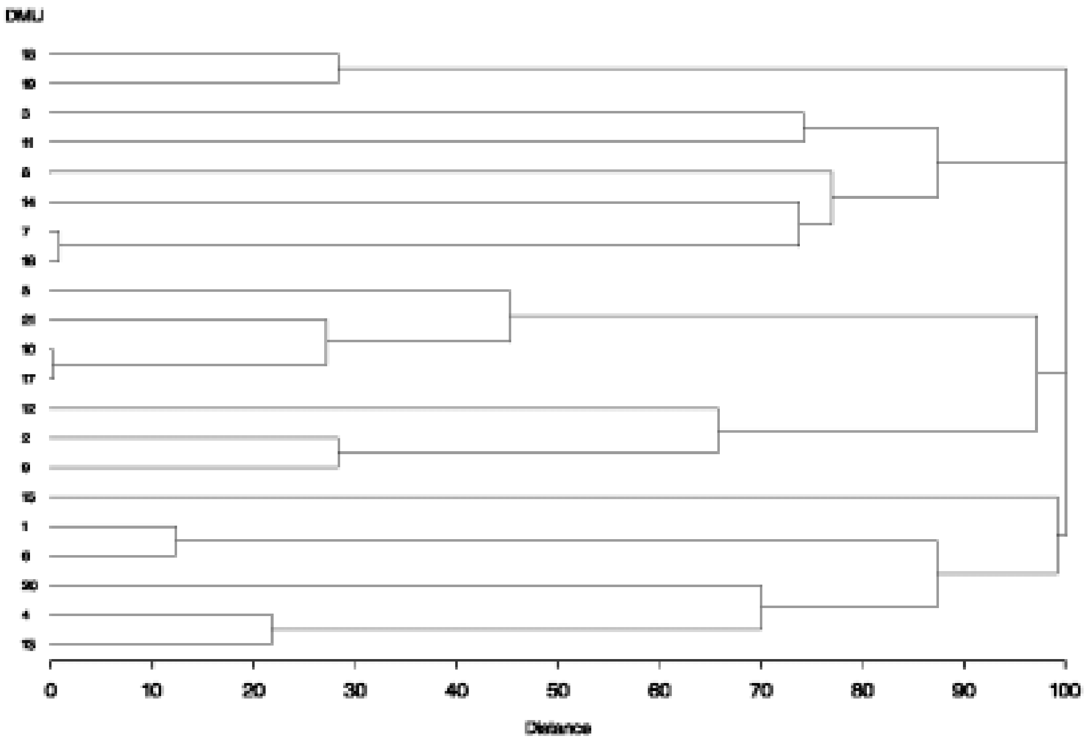

Figure 15. Dendrogram based on 100 x $D_e(i,j)$ for the 2 input 2 output case.

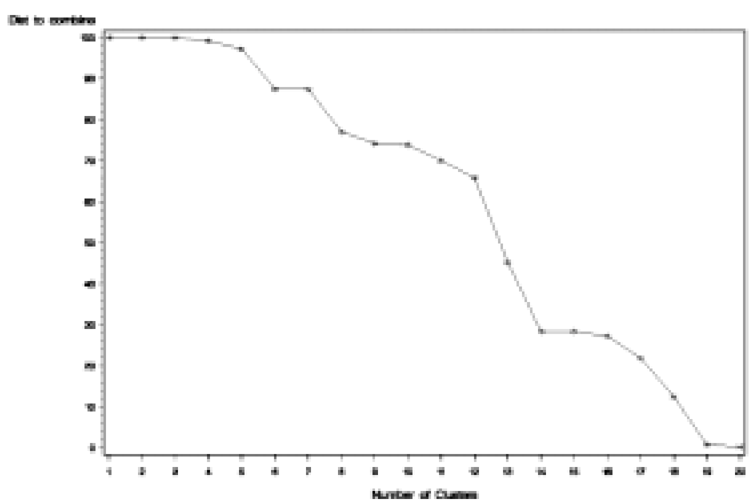

Figure 16. Change in 100 x $D_e(i,j)$ by cluster number for the 2 input 2 output case.

6. Conclusion

This paper shows that the use of cluster analysis to assess DEA scores can provide two dimensions on the relationship between DMUs. The first is the relationship based on common reference technologies. Clusters based on a correlation based criterion result in groupings of DMUs that are most similar in their production characteristics. In the second method the ranking of the DMUs is considered for the definition of groups. Those DMUs that are most similar in their efficiency score are combined in to similar groups.

The cluster analysis proposed in this paper are a first step in the development of other methods for the interpretation of the results of DEA. The future area of research in this area is to examine the possibility of using different measures for the clustering criteria..

References

Aigner, D., C. A. K. Lovell and P. Schmidt (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models," *Journal of Econometrics* 6, 21-37.

Ali, A. I. and L. M. Seiford (1993), "The Mathematical Programming Approach to Efficiency Analysis," in H. O. Fried, C.A.K. Lovell and S. S. Schmidt, eds., *The Measurement of Productive Efficiency: Techniques and Applications*. New York: Oxford University Press, 120-159.

Borland, J. , J. G. Hirschberg, and J. N. Lye, (2001), "Data reduction of discrete responses: an application of cluster analysis", *Applied Economics Letters*, 8, 149-153.

Charnes, A., W. W. Cooper, and E. Rhodes (1978), "Measuring the Efficiency of Decision Making Units," *European Journal of Operations Research* 2, 429-444.

Cooper, William W., Lawrence M. Seiford, and Kaoru Tone (2000), *Data Envelopment Analysis*, Kluwer Academic Publishers.

Chernick, Michael R., (1999), *Bootstrap Methods: A Practitioner's Guide*, John Wiley & Sons.

Davison, A. C., and D. V. Hinkley, (1997), *Bootstrap Methods and Their Application*, Cambridge University Press.

Davison, A. C., D. V. Hinkley, and E. Schechtman (1986), "Efficient bootstrap simulation", *Biometrika*, 73, 555-566.

Efron, B. (1979), "Bootstrapping Methods: Another Look at the Jackknife," *Annals of Statistics* 7, 1-26.

Efron, B, and R. J. Tibshirani, (1993), *An Introduction to the Bootstrap*, Chapman and Hall.

Färe, R., S. Grosskopf and C. A. K. Lovell (1985), *The Measurement of Efficiency and Production*. Boston, MA: Kluwer-Nijhoff Publishing.

Färe, R., S. Grosskopf and C. A. K. Lovell (1994), *Production Frontiers*. Cambridge: Cambridge University Press.

Farrell, M. J. (1957), "The Measurement of Productive Efficiency," *Journal of the Royal Statistical Society*, Series A (General), 120, pt. 3, 253-281.

Ferrier, G. D., and J. G. Hirschberg, (1992), "Climate Control Efficiency", *The Energy Journal*, 13, 37-54.

Ferrier, G. D., and J. G. Hirschberg, (1997), "Bootstrapping Confidence Intervals for Linear Programming Efficiency Scores: With an Illustration Using Italian Banking Data", *Journal of Productivity Analysis*, 8, 19-33.

Ferrier, G. D., and J. G. Hirschberg, (1999), "Can We Bootstrap DEA Scores?", *Journal of Productivity Analysis*, 11, 81-92.

Freedman, D. A. and S. C. Peters, (1984), "Bootstrapping a regression equation: Some empirical results", *Journal of the American Statistical Association*, 79, 97-106.

Hayes, K. J., S. Grosskopf, and J. G. Hirschberg, (1995), "Fiscal Stress and the Production of Public Safety: a Distance Function Approach", *Journal of Public Economics*, 57, 277-296

Hirschberg, J. G. (1992), "A Computationally Efficient Method for Bootstrapping Systems of Demand Equations: A Comparison to Traditional Techniques", *Statistics and Computing*, 2, 19-24.

Hirschberg, J. G. (2000), "Modelling time of day substitution using the second moments of demand", (forthcoming) *Applied Economics.*

Hirschberg, J. G. and D. J. Aigner (1987) "A Classification for Medium and Small Firms by Time-of-Day Electricity Usage", *Papers and Proceedings of the Eight Annual North American Conference of the International Association of Energy Economists*, Cambridge, MA, November 19-21, 1986, 253-257.

Hirschberg, J. G., and J. R. Dayton, (1996), ``Detailed patterns of intra-industry trade in processed food," in *Industrial Organization and Trade in the Food Industries*, I M. Sheldon and P. C. Abbott eds., Westview Press, Boulder, Colorado, 141-159.

Hirschberg, J. G. and P. J. Lloyd (2000), "An Application Of Post-DEA Bootstrap Regression Analysis To The Spill Over Of The Technology Of Foreign-Invested Enterprises In China", Department of Economics Working Paper #732, The University of Melbourne. (found at http://www.ecom.unimelb.edu.au/ecowww/research/732.pdf)

Hirschberg, J. G., E. Maasoumi and D. J. Slottje, (1991), "Cluster analysis for measuring welfare and quality of life across countries," *Journal of Econometrics*, 50, 131-150.

Hirschberg, J. G., E. Maasoumi and D. J. Slottje, (2000a forthcoming), "Clusters of Attributes and Well-Being in the US", *Journal of Applied Econometrics.*

Hirschberg, J. G., E. Maasoumi and D. J. Slottje, (2000b forthcoming), "The Environment and the Quality of Life in the United States Over Time", *Environmental Modelling and Software*.

Hirschberg, J. G. and D. J. Slottje, (1989), "Remembrance of Things Past: The Distribution of Earnings Across Occupations and the Kappa Criterion", 1989, *Journal of Econometrics,* 42, No 1., 121-130.

Hirschberg, J. G. and D. J. Slottje, (1994) "An Empirical Bayes Approach to Analyzing Earnings Differentials for Various Occupations and Industries", *Journal of Econometrics*, 61, 65-79.

Jensen, U., (forthcoming in 2000), "Is it efficient to analyse efficiency rankings?", *Empirical Economics*.

Kaufman, L., and P. J. Rousseeuw, (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley & Sons, New York.

Löthgren, M. and M. Tambour (1999), "Bootstrapping the data envelopment analysis Malmquist productivity index", *Applied Economics*, 31, 417-425.

McGee, V. E. and W. T. Carlton, (1970), "Piecewise Regression," *Journal of the American Statistical Association*, 65, 1109-1124

Norman, M. and B. Stoker, (1991), *Data Envelopment Analysis: The Assessment of Performance,* John Wiley & Sons, Chichester.

Simar, L. and P. L. Wilson, (1998), "Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Non-parametric Frontier Models", *Management Science*, 44, 49-61.

Xue, Mei. and Patrick. T. Harker, (1999), "Overcoming the Inherent Dependency of DEA Efficiency Scores: A Bootstrap Approach", Working paper Department of Operations and Information Management, University of Pennsylvania. Can be found at http://opim.wharton.upenn.edu/~harker/DEAboot.pdf.