

Research Article

A Data Envelopment-Based Clustering Approach for Public Sugar Factories in Privatizing Process

Ezgi A. Demirtas

Department of Industrial Engineering, Eskisehir Osmangazi University, 26480 Eskisehir, Turkey

Correspondence should be addressed to Ezgi A. Demirtas, eaktar@ogu.edu.tr

Received 31 March 2011; Revised 17 June 2011; Accepted 19 June 2011

Academic Editor: Ezzat G. Bakhoun

Copyright © 2011 Ezgi A. Demirtas. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Turkish Sugar Inc., a public enterprise including 25 factories, is the first corporation of Turkish industry. According to the government policy, public sugar factories (PSFs) will be privatized as geography-based 6 portfolio groups in two years. As performance measures of PSF affect government, sugar producers, and several unions in privatizing process, a systematic approach is necessary to measure efficiencies and grouping factories. This paper uses a new DEA- (Data Envelopment Analysis-) based clustering approach for measuring efficiency scores of PSF and grouping them instead of geography- based portfolio groups. This new approach can help decision makers in privatizing process. At the same time, target values obtained by dual model can be used to eliminate inefficiencies of some PSFs.

1. Introduction

Sugar factories are the first corporations of Turkish industry. The first sugar factory was established by the direction of Kemal Ataturk in Alpullu in 1926. Annual sugar demand of Turkey which is supplied by three different sugar producers is 2.3 million ton. These producers are Turkish Sugar Inc., a public enterprise including 25 factories, Pankobirlik (beet producers union) with 6 factories, and starch-based sugar producers with 5 factories. The market share of these producers is 70%, 20%, and 10%, respectively. Turkish sugar Inc. and Pankobirlik use beet to produce sugar instead of starch.

According to the government policy, sugar factories of Turkish Sugar Inc. will be privatized as geography-based 6 portfolio groups in two years: A (Kars, Ercis, Agri, Mus, and Erzurum), B (Elazig, Malatya, Erzincan, and Elbistan), C (Kastamonu, Kirsehir, Turhal, Yozgat, Corum, and Carsamba), D (Bor, Eregli, and Ilgin), and E (Usak, Alpullu, Burdur, and Afyon). Some Turkish and foreign corporations aspire to buy sugar factories like Pankobirlik, Keskinlik, Torunlar (Turkish), Sudzucker (German), Saint Louys Sucre (French), and British Sugar.

According to the privatizing supporters, average production season length is shorter in TR (64 days) than that in EU (120 days) and the average number of personnel per factory in TR and EU are 500 and 200, respectively. The number of average personnel per factory is too large in contrast with Europe. On the other hand, usage of noneconomic beets with low polar sugar value brings high costs. According to the opponents, the reasons of high prices are high starch-based sugar quota (15% in TR and 2% in EU) and outdated technology. If the starch based-sugar quota is decreased and new technology is used, PSFs become profitable. At the same time, to prevent external dependency and rural-urban migration and to protect sector, state of council should stop enforcement decision about privatizing. In the final situation, state of council stopped enforcement decision because of the following reasons: (i) the specification contains 5-year production obligation and 50-million-dollar assurance, (ii) it does not guarantee supply-demand balance and stability, and (iii) it does not guarantee production sustainability, and it creates external dependency.

Because of the reasons mentioned above, performance measurement of PSF is an important task which affects government, sugar producers, and several unions in privatizing process. A systematic approach is necessary to measure efficiencies of PSF. This paper is the first real-life application of DEA-based clustering approach developed by Po et al. [1] for measuring efficiency scores of PSF and grouping them instead of geography-based portfolio groups. This new approach can help decision makers in privatizing process of PSF. At the same time, target values obtained by DEA model can be used to eliminate inefficiencies and to make inefficient factories profitable. Additionally, to the best of the author's knowledge, there is no scientific study for efficiency measurement of public sugar factories in our country or elsewhere.

The rest of this paper is organized as follows: Section 2 discusses DEA and DEA-based clustering approach which is developed by Po et al. [1]. In this section, the focus is why and how piecewise production functions drawn from DEA models are employed to cluster data. Section 3 illustrates the proposed DEA-based clustering approach for measuring efficiency scores of PSF and grouping them to help decision makers in privatizing process. The results obtained by DEA-based clustering approach are compared with geographic based portfolio groups and target values obtained by CCR model are given to eliminate the inefficiencies of some PSF. Finally, conclusions are stated in Section 4.

2. DEA-Based Clustering Approach

Conventionally, most clustering algorithms are procedures that minimize total dissimilarity; examples of such algorithms are given in the paper of Po et al. [1].

A general clustering method is to find c cluster centers z_1, z_2, \dots, z_c so that the total dissimilarity measure $J_s(z)$ with $J_s(z) = \sum_{i=1}^c \sum_{j=1}^n a_{ij} f(d(x_j, z_i))$ is minimized. $J_s(z)$ is usually defined as a distance-based function, and the problem here is to select a useful and reasonable distance measure $d(x_j, z_i)$.

On the other hand, the stated clustering approaches can be seen as a feature analysis technique. An assumption of the underlying feature analysis is to regard the feature items A_1, A_2, \dots, A_s as multiple features so that the minimization of $J_s(z)$ presents the closer of data among their features and makes it more possible for these DMUs to be classified into the same cluster. However, the clustering results derived from the minimization of the total feature dissimilarity $J_s(z)$ may not be helpful in some cases of clustering DMUs, especially in production units. In these cases, we use their production data to cluster them. Suppose that the production data have feature items $A_1, A_2, \dots, A_k, A_{k+1}, \dots, A_s$ with A_1 to A_k being input items and

A_{k+1} to A_s being output items. The clustering information obtained from the conventional clustering approaches can only reveal that DMUs are more similar to another one. However, the more important information we want to know is the production feature (functions) implied from the production data of all DMUs. That is, $f_j(A_1, A_2, \dots, A_k, A_{k+1}, A_{k+2}, \dots, A_s) = 0$. From these derived production functions, f_1, f_2, \dots , all DMUs are classified into different clusters (production functions). Therefore, each DMU knows not only the cluster that it belongs to but also knows the production function type that it confronts. Each DMU can compare its production feature with the other production functions so that the combination of its input resources or the combination of inputs and outputs can be readjusted. That is, for the case of data feature with input and output items, the cluster derived from production functions is more valuable than that derived from feature dissimilarity measures.

The idea of Po et al.'s study [1] is to employ the production functions to cluster production data. The method supporting this idea is DEA, as initiated and developed by Charnes et al. [2]. The DEA is a data-oriented method for evaluating the relative efficiency of DMUs where each DMU is an entity responsible for converting multiple inputs into multiple outputs. Since the fundamental of DEA uses the nonparametric mathematical programming approach to estimate piecewise frontiers and envelop the DMU data sets, in this study, each piecewise frontier is regarded as one cluster of production functions. Therefore, we use all piecewise frontiers as a base to cluster production data. That is, they give up traditional clustering approaches of feature dissimilarity and propose a new approach by adopting the production functions revealed by the observation data to cluster all DMUs.

DEA is a nonparametric method for the estimation of production frontiers. It is a useful tool for evaluating the relative efficiency for a group of DMUs. Up to now, DEA has been widely studied and applied in various areas for 30 years since Charnes et al. [2] first proposed the DEA method with the CCR model. Among them, the main forms of DEA models and their extensions include those of BCC model [3], the additive model, [4] and the imprecise DEA models [5, 6]. Modifications and extensions are the assurance region models [7, 8], superefficiency models [9, 10], cone ratio models [11, 12]. Stochastic and chance-constrained extensions are considered by some authors [13–17]. Taxonomy and general model frameworks for DEA can be found in [18, 19]. The CCR is the original model of DEA (see the M1 model) and is used in this study to explain the DEA-based clustering approach.

The DEA model generalizes the usual input/output ratio measure of efficiency for a given unit in terms of a fractional linear program formulation. According to the economic notion of Pareto optimality, the DEA method states that a DMU is considered inefficient if some other DMUs or some combinations of other DMUs produce at least the same amount of output with less of the same resources input and not more of any other resources. Conversely, a DMU is considered Pareto efficient if the above is not possible. Suppose that there are n DMUs to be evaluated, x_{ij} is the noted amount of the i th input for the j th DMU and y_{rj} is the noted amount of the r th output for the j th DMU. Output multipliers are u_1, u_2, \dots, u_s (one for each item of output) and input multipliers are v_1, v_2, \dots, v_m (one for each item of input). The mathematical formulation of the method is summarized next, where the relative efficiency of the DMU _{k} is determined [20]. See the M1 model.

M1 Model:

$$\begin{aligned} \text{Max}_{u,v} \quad \text{Eff}_k &= \frac{\sum_{r=1}^s u_r y_{rk}}{\sum_{i=1}^m v_i x_{ik}}, \\ \text{Subject to} \quad &\frac{\sum_{r=1}^s u_r y_{rk}}{\sum_{i=1}^m v_i x_{ik}} \leq 1, \quad u_r \geq \varepsilon, \quad v_i \geq \varepsilon \quad \forall r, i. \end{aligned} \quad (2.1)$$

The DEA model is essentially a fractional programming problem with a ratio of a weighted sum of outputs to a weighted sum of inputs where the weights for both inputs and outputs are to be selected in a manner that calculates the efficiency of the evaluated unit. Therefore, the original form of the DEA model is both nonlinear and nonconvex problem. Charnes et al. [21] proved that fractional programming problem can be transformed into linear programming formulations. The first formulation is “input based,” constraining the weighted sum of outputs to be unity and minimizes the inputs that can then be obtained. The second formulation is “output based,” constraining the weighted sum of inputs to be unity and maximizes the outputs that can then be obtained (see the M2 model). Given constant returns to scale assumption, the result from the input-based model is the reciprocal of that from output-based model. If variable returns to scale are assumed, there is no direct relation can be found between these two models.

For the clustering approach used in this study, the results can be different for those PSFs which are not on the production frontier according to the way that input-based or output based model is applied. The choice of using an input-based or output-based model depends on the production process characterizing the firm (i.e., minimize the use of inputs to produce a given output or maximize the output with given levels of inputs). The objective of this study is to find the set of coefficients associated with each output and input that will give the PSF being evaluated the highest possible efficiency by using the M2 model. Then, target values are calculated by using this model to eliminate the inefficiencies of some PSFs.

M2 Model:

$$\begin{aligned} \text{Max}_{u,v} \quad \text{Eff}_k &= \sum_{r=1}^s u_r y_{rk}, \\ \text{Subject to} \quad &\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0 \quad j = 1, \dots, n, \\ &\sum_{i=1}^m v_i x_{ik} = 1 \quad u_r \geq \varepsilon, v_i \geq \varepsilon \quad \forall r, i. \end{aligned} \quad (2.2)$$

DEA differs from the production theory of economics in that it is nonparametric. In economics, the production function is a function that summarizes the process of converting multiple inputs into a single output. Thus, a general mathematical form for the production function in economics can be expressed as $y = f(x_1, x_2, x_3, \dots, x_n)$, where y is a quantity of output and $x_1, x_2, x_3, \dots, x_n$ are quantities of inputs. However, DEA is a nonlinear programming model for evaluating a process converting multiple inputs into multiple outputs, that is, $g(y_1, y_2, y_3, \dots, y_n) = f(x_1, x_2, x_3, \dots, x_n)$. Most previous studies had mentioned and discussed the properties of production function that are hidden in DEA methods [8–10, 14, 15, 17, 22].

Since the number of DMUs is usually much larger than the number of inputs, we prefer to express the linear programming in its duality form. Further, the duality form can interpret the geometric meaning of DEA and provide information about conservation of resources or expansion of outputs to have DMUs from inefficiency to efficiency.

If Eff_k^* is the optimal value of Eff_k , the DMU_k is said to be efficient if and only if $\text{Eff}_k^* = 1$. If Eff_k^* is less than 1, DMU_k is inefficient. According to the efficiency ratio, DMUs may be grouped as good ($\text{Eff}_k^* = 1$) and poor ($\text{Eff}_k^* < 1$) performers or clustered by assigning different efficiency ratio grades [23–27]. Although clustering by efficiency ratio gives some information about the rationality of output/input, it does not reveal the intrinsic relationship

between the input and output production features. Therefore, this study adopts piecewise production functions derived from the DEA method to cluster data.

In M2 model, it is obvious that the constraint $\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0$ is an inequality formula of production functions. Solving M2 model yields the virtual multipliers u_r^* and v_i^* . Thus, $\sum_{r=1}^s u_r^* y_{rj} - \sum_{i=1}^m v_i^* x_{ij} = 0$ is derived. Running M2 model for $k = 1$ to n gives all production functions. Then, all DMUs are classified into different clusters by these piecewise production functions. Thus, a clustering method using production functions via the DEA method is implemented. Po et al. [1] find that there is less consideration in using these production functions as a reference to classify evaluated DMUs, and they propose a clustering approach according to the properties of DEA and its production possibility set such that they can use these production functions as a reference to classify evaluated DMUs. The details about the algorithm used in which the DEA-based clustering method is applied are given in their paper.

3. DEA-Based Clustering of PSF

In this study, we have an efficiency evaluation problem with 25 PSFs (DMU), each PSF with three inputs and one output obtained by 2009-2010 annual activity reports. Actually processed beet quantity (PBQ), fuel consumption (FC), number of total personnel (TP), sugar production (SP), and molasses production (MP) data are placed in annual activity reports of PSF, and all of them are real and correct. PBQ, FC, and TP are considered as inputs. Only SP is selected as output because it is correlated with MP.

The simplified production data of PSF are shown in Table 1. This table shows the required quantity of inputs to produce one unit of (one metric ton) sugar. For example, PSF 22 uses 9.96 ton beet, 0.419 ton fuel, and 0.0213 personnel to produce one ton sugar according to Table 1.

The objective is to find the set of coefficients u 's associated with each output and v 's associated with each input that will give the PSF being evaluated the highest possible efficiency. By using the M2 model for each PSF its efficiency ratio Eff_k and the solution of virtual multipliers, $v_1^*, v_2^*, v_3^*, u_1^*$ are obtained. The multipliers v_1^*, v_2^*, v_3^* are measure of the relative increase in efficiency with each unit

reduction of input value, where u_1^* is a measure of the relative decrease in efficiency with each unit reduction of output value.

The analytical results are shown in Table 2.

By selecting the set of virtual multipliers $v_1^*, v_2^*, v_3^*, u_1^*$ to be all nonzero, four frontiers of production functions (PFs) are found:

$$y = 0.142x_1 + 0.094x_2 + 5.36x_3, \quad (\text{PF1})$$

$$y = 0.101x_1 + 1.181x_2 + 1.415x_3, \quad (\text{PF2})$$

$$y = 0.119x_1 + 0.619x_2 + 5.36x_3, \quad (\text{PF3})$$

$$y = 0.109x_1 + 1.018x_2 + 0.978x_3. \quad (\text{PF4})$$

PSFs with (*), (**), (***) in Table 2 confront the degenerative frontier. Po et al. [1] suggest that they should be reclassified into the nearest effective frontier (the frontier with nonzero virtual multipliers). In this application, it is observed that PSFs with (*) confront the nearest effective frontier $y = 0.142x_1 + 0.09x_2 + 5.36x_3$, thus their efficiency

Table 1: Production data of PSF.

PSF no.	PSF	PBQ (ton)	FC (ton)	TP	SP (ton)
1	Afyon	6.602	0.3384	0.0058	1
2	Agri	7.7604	0.3406	0.0335	1
3	Alpullu	7.6027	0.3708	0.0216	1
4	Ankara	7.4387	0.2882	0.0217	1
5	Bor	7.3062	0.3063	0.0124	1
6	Burdur	6.5925	0.3174	0.008	1
7	Carsamba	9.6162	0.5149	0.032	1
8	Corum	6.6725	0.2634	0.0082	1
9	Elazig	7.6171	0.3585	0.0312	1
10	Elbistan	7.2789	0.3477	0.0207	1
11	Ercis	6.6769	0.2289	0.0367	1
12	Eregli	7.0098	0.3773	0.0042	1
13	Erzincan	6.807	0.2692	0.018	1
14	Erzurum	6.1656	0.3018	0.018	1
15	Eskisehir	6.8239	0.2763	0.0076	1
16	Ilgın	6.8072	0.3484	0.0061	1
17	Kars	7.336	0.3341	0.0455	1
18	Kastamonu	7.1647	0.3345	0.015	1
19	Kirsehir	6.8418	0.3623	0.0095	1
20	Malatya	8.2043	0.3356	0.019	1
21	Mus	6.6949	0.2525	0.0172	1
22	Susurluk	9.96	0.4191	0.0213	1
23	Turhal	7.4756	0.3609	0.0104	1
24	Uşak	6.8136	0.308	0.0201	1
25	Yozgat	6.4054	0.2818	0.012	1

ratio will be reevaluated by this frontier. However, in complicated applications (with more data items of input and output), it is impossible to judge the nearest effective frontier by observation. Hence, for PSF 7 (Carsamba), we follow the procedure of Po et al. [1], taking $(x_1, x_2, x_3, y) = (9.6162t(i), 0.5149t(i), 0.032t(i), 1)$, $i = 1, 2, 3, 4$ into (PF1), (PF2), (PF3), and (PF4), respectively. The $t(i)$ value is calculated, giving $t(1) = 0.6307$, $t(2) = 0.6155$, $t(3) = 0.6118$ and $t(4) = 0.6236$. By taking the maximal value, the efficiency ratio for PSF 7 is re-evaluated as 0.6307. In addition, PSF 7 is classified into the cluster determined by the corresponding envelope $y = 0.142x_1 + 0.09x_2 + 5.36x_3$ (PF1).

In this study, some PSFs achieve 100 percent efficiency and are referred to as the relatively efficient units, whereas other units with efficiency ratings of less than 100 percent are referred to as inefficient units. According to the results of Table 2, there are six efficient (PSF1, PSF8, PSF11, PSF14, PSF21, and PSF25) and 19 inefficient PSFs. The 5 PSFs out of 19 have greater than or equal to 0.95 efficiency ratio. Additionally, the net revenues of PSF are supported by the DEA results. According to four different production functions, 25 PSFs are classified into four clusters. Clustering results are shown in Table 3.

By considering (PF1), (PF2), and (PF3), TP is the most critical input for the PSF in clusters 1, 2, and 3. The multiplier of TP has the biggest value for these clusters. The relative increase in efficiency is 5.36 with each number reduction of TP for the inefficient PSF placed

Table 2: Analytical results derived from M2 model for PSF.

PSF no.	PSF Name	v_1^*	v_2^*	v_3^*	u_1^*	Evaluated by the frontier of			
1	Afyon	0.142	0.094	5.36	1	0.142	$x_1 + 0.094$	$x_2 + 5.36$	x_3
2***	Agri	0.099	0.691	0	0.816	0.109	$x_1 + 1.018$	$x_2 + 0.978$	x_3
					0.816				
3	Alpullu	0.115	0.076	4.358	0.813	0.142	$x_1 + 0.094$	$x_2 + 5.36$	x_3
4	Ankara	0.09	1.049	1.256	0.888	0.101	$x_1 + 1.181$	$x_2 + 1.415$	x_3
5	Bor	0.091	1.054	1.262	0.892	0.101	$x_1 + 1.181$	$x_2 + 1.415$	x_3
6	Burdur	0.1407	0.093	5.312	0.991	0.142	$x_1 + 0.094$	$x_2 + 5.36$	x_3
7*	Carsamba	0.104	0	0	0.641	0.142	$x_1 + 0.094$	$x_2 + 5.36$	x_3
					0.631				
8	Corum	0.119	0.619	5.36	1	0.119	$x_1 + 0.619$	$x_2 + 5.36$	x_3
9***	Elazig	0.099	0.692	0	0.818	0.109	$x_1 + 1.018$	$x_2 + 0.978$	x_3
					0.816				
10	Elbistan	0.093	0.87	0.836	0.854	0.109	$x_1 + 1.018$	$x_2 + 0.978$	x_3
11**	Ercis	0	4.369	0	1	0.101	$x_1 + 1.181$	$x_2 + 1.415$	x_3
					1				
12*	Eregli	0.139	0	5.36	0.952	0.142	$x_1 + 0.094$	$x_2 + 5.36$	x_3
					0.949				
13	Erzincan	0.098	1.142	1.367	0.967	0.101	$x_1 + 1.181$	$x_2 + 1.415$	x_3
14***	Erzurum	0.121	0.847	0	1	0.109	$x_1 + 1.018$	$x_2 + 0.978$	x_3
					1				
15	Eskisehir	0.116	0.605	5.24	0.978	0.119	$x_1 + 0.619$	$x_2 + 5.36$	x_3
16	Ilgin	0.138	0.091	5.195	0.969	0.142	$x_1 + 0.094$	$x_2 + 5.36$	x_3
17***	Kars	0.103	0.725	0	0.856	0.109	$x_1 + 1.018$	$x_2 + 0.978$	x_3
					0.844				
18	Kastamonu	0.126	0.083	4.748	0.886	0.142	$x_1 + 0.094$	$x_2 + 5.36$	x_3
19*	Kirsehir	0.139	0	4.981	0.948	0.142	$x_1 + 0.094$	$x_2 + 5.36$	x_3
					0.947				
20	Malatya	0.081	0.94	1.126	0.796	0.101	$x_1 + 1.181$	$x_2 + 1.415$	x_3
21**	Mus	0	3.659	4.428	1	0.101	$x_1 + 1.181$	$x_2 + 1.415$	x_3
					1				
22	Susurluk	0.066	0.769	0.921	0.651	0.101	$x_1 + 1.181$	$x_2 + 1.415$	x_3
23	Turhal	0.123	0.082	4.657	0.869	0.142	$x_1 + 0.094$	$x_2 + 5.36$	x_3
24	Uzak	0.101	0.944	0.907	0.927	0.109	$x_1 + 1.018$	$x_2 + 0.978$	x_3
25	Yozgat	0.142	0.094	5.36	1	0.142	$x_1 + 0.094$	$x_2 + 5.36$	x_3

*They are reevaluated by $y = 0.142x_1 + 0.094x_2 + 5.36x_3$, because of zero virtual multipliers.

**They are reevaluated by $y = 0.101x_1 + 1.181x_2 + 1.415x_3$, because of zero virtual multipliers.

***They are reevaluated by $y = 0.109x_1 + 1.018x_2 + 0.978x_3$, because of zero virtual multipliers.

in clusters 1 and 3. Similarly, the relative increase in efficiency is 1.415 for the inefficient PSF placed in cluster 2 (see Table 3). The order of multipliers for other inputs changes. For example, for the PSF in cluster 2, the multipliers of FC and TP are similar and higher than PBQ. On the other hand, FC is the most critical input for the PSF in cluster 4. The relative increase in efficiency is 1.018 with each ton reduction of FC for the inefficient PSF placed in cluster 4 (see Table 3). DEA and geography-based clustering results are compared in Table 4.

As you can see in Table 4, DEA-based clusters contain different geography-based portfolio groups (1-E, C, D), (2-A, B, D), (3-C), and (4-A, B, E). Moreover, the clustering

Table 3: DEA-based PSF clusters.

Cluster (PF) no.	PSF status	PSF number
(PF1)	Efficient	1, 25
	Inefficient	3, 6, 7, 12, 15, 16, 18, 19, 23
(PF2)	Efficient	11, 21
	Inefficient	4, 5, 13, 20, 22
(PF3)	Efficient	8
	Inefficient	15
(PF4)	Efficient	14
	Inefficient	2, 9, 10, 17, 24

Table 4: DEA-/Geography-based PSF clusters.

PSF no.	PSF	DEA-based cluster no.	Geography-based cluster no.
1	Afyon	1	E
2	Agri	4	A
3	Alpullu	1	E
4	Ankara	2	*
5	Bor	2	D
6	Burdur	1	E
7	Carsamba	1	C
8	Corum	3	C
9	Elazig	4	B
10	Elbistan	4	B
11	Ercis	2	A
12	Eregli	1	D
13	Erzincan	2	B
14	Erzurum	4	A
15	Eskisehir	3	*
16	Ilgın	1	D
17	Kars	4	A
18	Kastamonu	1	C
19	Kirsehir	1	C
20	Malatya	2	B
21	Mus	2	A
22	Susurluk	2	*
23	Turhal	1	C
24	Uşak	4	E
25	Yozgat	1	C

* These PSFs are out of geography-based portfolio according to government policy.

results derived from geography-based portfolio may not be helpful in cases of clustering PSF. From the derived production functions (PF1), (PF2), (PF3), and (PF4), all PSFs are classified into four different clusters (production functions). Therefore, each PSF knows the PF type that it confronts. Additionally, each PSF can compare its production feature with the other production functions so that the combination of its input resources or the combination of inputs and outputs can be readjusted. That is, for the case of data feature with input

Table 5: Input target values of PSF.

PSF no.	PSF name	Input targets		
		PBQ	FC	TP
2	Agri	6.333	0.278	0.024
3	Alpullu	6.182	0.302	0.018
4	Ankara	6.605	0.256	0.0193
5	Bor	6.518	0.273	0.011
6	Burdur	6.516	0.314	0.009
7	Carsamba	6.602	0.338	0.0058
9	Elazig	6.227	0.293	0.020
10	Elbistan	6.208	0.297	0.018
12	Eregli	6.602	0.338	0.006
13	Erzincan	6.587	0.261	0.017
15	Eskisehir	6.611	0.2678	0.009
16	Ilgin	6.592	0.338	0.006
17	Kars	6.277	0.286	0.022
18	Kastamonu	6.347	0.296	0.013
19	Kirsehir	6.487	0.329	0.009
20	Malatya	6.525	0.267	0.015
22	Susurluk	6.477	0.273	0.014
23	Turhal	6.503	0.314	0.009
24	Usak	6.313	0.285	0.019

and output items, the cluster derived from production functions is more valuable than that derived from geography-based portfolio groups. It is possible to eliminate inefficiencies by considering DEA-based clustering. For example, inefficient PSF placed in clusters (1), (2), and (3) should give priority to decrease the number of total personnel. In the same manner, inefficient PSF placed in cluster (4) should decrease fuel consumption at first. It is meaningful to support privatizing decisions by DEA-based clustering results than geography-based portfolio groups.

At the same time, target values of inputs are calculated by using slack variables of M2 model and illustrated in Table 5 for the inefficient PSF. Target values can help decision makers to eliminate the inefficiencies. For example, 9.96 ton beet, 0.419 ton fuel, and 0.0213 personnel are used to produce one ton sugar in PSF 22. When 6.477 ton beet, 0.273 ton fuel, and 0.0014 personnel are used to produce one ton sugar, PSF 22 becomes efficient.

4. Conclusions

This study develops a DEA-based clustering approach for the evaluation of PSF. The proposed approach employs the piecewise production functions derived from the DEA method to cluster the data with input and output items. Compared with geography-based clustering that only considers geographical location of PSF, our proposed approach reveals the input-output relationships hidden in the data items of input and output. Thus, for each evaluated PSF, we know not only the cluster that it belongs to but also the production function type that it confronts. It is very important for managerial decision making where decision

makers are interested in knowing the changes required in combining input resources so that it can be reclassified into a different and desired cluster/class in privatizing process.

The focus of this paper is to examine the CCR model of DEA and then establish the DEA-based clustering. Without loss of generality, while this approach has been carried out for the CCR model, the proposed approach can be easily extended to other DEA models. The clustering results drawn from the DEA-based clustering are unit invariant, meaning that they are not affected by the scale of data.

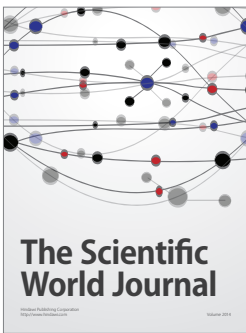
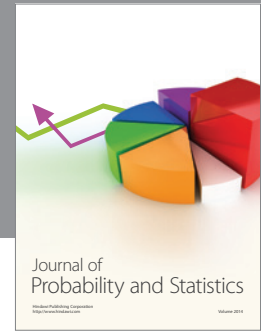
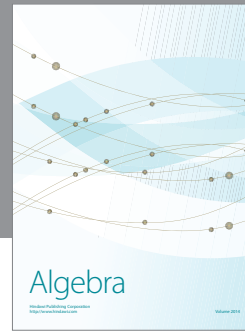
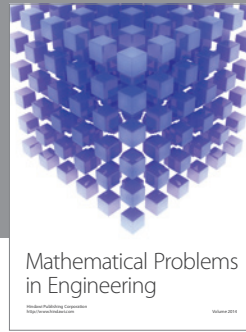
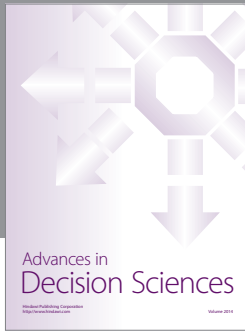
The DEA-based clustering approach is suitable for most clustering problems, where there are inputs-and-outputs or cause-and-effect relationships between the features. For example, we use the proposed approach in the analysis of industry classification, sorting of PSF by input-output data.

In summary, in view of the advantages of the DEA-based clustering approach, it is uniquely poised for clustering problems. We believe that future researches are necessary to unleash the full potential of this DEA-based clustering approach. It thus has tremendous potential to be used for various clustering problems. DEA-based clustering algorithm developed by Po et al. [1] is robust to a slight change in the input and output data sets, but not to outliers. Future researches will consider developing a robust-type DEA-based clustering algorithm.

References

- [1] R.-W. Po, Y.-Y. Guh, and M.-S. Yang, "A new clustering approach using data envelopment analysis," *European Journal of Operational Research*, vol. 199, no. 1, pp. 276–284, 2009.
- [2] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978.
- [3] R. D. Banker, A. Charnes, and W. W. Cooper, "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management Science*, vol. 30, no. 9, pp. 1078–1092, 1984.
- [4] A. Charnes, W. W. Cooper, B. Golany, L. Seiford, and J. Stutz, "Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions," *Journal of Econometrics*, vol. 30, no. 1-2, pp. 91–107, 1985.
- [5] W. W. Cooper, K. S. Park, and G. Yu, "Idea and AR-IDEA: models for dealing with imprecise data in DEA," *Management Science*, vol. 45, no. 4, pp. 597–607, 1999.
- [6] J. Zhu, "Imprecise data envelopment analysis (IDEA): a review and improvement with an application," *European Journal of Operational Research*, vol. 144, no. 3, pp. 513–529, 2003.
- [7] R. G. Thompson, R. D. Singleton, R. M. Thrall, and B. A. Smith, "Comparative site evaluations for locating a high energy physics laboratory in Texas," *Interfaces*, vol. 16, no. 6, pp. 35–49, 1986.
- [8] S. H. Zanakis, C. Alvarez, and V. Li, "Socio-economic determinants of HIV/AIDS pandemic and nations efficiencies," *European Journal of Operational Research*, vol. 176, no. 3, pp. 1811–1838, 2007.
- [9] P. Andersen and N. C. Petersen, "A procedure for ranking efficient units in data envelopment analysis," *Management Science*, vol. 39, no. 10, pp. 1261–1264, 1993.
- [10] S. Li, G. R. Jahanshahloo, and M. Khodabakhshi, "A super-efficiency model for ranking efficient units in data envelopment analysis," *Applied Mathematics and Computation*, vol. 184, no. 2, pp. 638–648, 2007.
- [11] A. Charnes, W. W. Cooper, Q. L. Wei, and Z. M. Huang, "Cone ratio data envelopment analysis and multi-objective programming," *International Journal of Systems Science*, vol. 20, no. 7, pp. 1099–1118, 1989.
- [12] A. Charnes, W. W. Cooper, Z. M. Huang, and D. B. Sun, "Polyhedral cone-ratio DEA models with an illustrative application to large commercial banks," *Journal of Econometrics*, vol. 46, no. 1-2, pp. 73–91, 1990.
- [13] K. C. Land, C. A. K. Lovell, and S. Thore, "Productivity and efficiency under capitalism and state socialism: an empirical inquiry using chance-constrained data envelopment analysis," *Technological Forecasting and Social Change*, vol. 46, no. 2, pp. 139–152, 1994.
- [14] O. B. Olesen and N. C. Petersen, "Chance constrained efficiency evaluation," *Management Science*, vol. 41, no. 3, pp. 442–457, 1995.

- [15] W. W. Cooper, Z. Huang, and S. X. Li, "Satisficing DEA models under chance constraints," *Annals of Operations Research*, vol. 66, no. 4, pp. 279–295, 1996, Extensions and new developments in data envelopment analysis.
- [16] R. Lahdelma and P. Salminen, "Stochastic multicriteria acceptability analysis using the data envelopment model," *European Journal of Operational Research*, vol. 170, no. 1, pp. 241–252, 2005.
- [17] W. W. Cooper, H. Deng, Z. Huang, and S. X. Li, "Chance constrained programming approaches to technical efficiencies and inefficiencies in stochastic data envelopment analysis," *Journal of the Operational Research Society*, vol. 53, no. 12, pp. 1347–1356, 2002.
- [18] S. Gattoufi, M. Oral, and A. Reisman, "A taxonomy for data envelopment analysis," *Socio-Economic Planning Sciences*, vol. 38, no. 2-3, pp. 141–158, 2004.
- [19] A. Kleine, "A general model framework for DEA," *Omega—The International Journal of Management Science*, vol. 32, no. 1, pp. 17–23, 2004.
- [20] W. W. Cooper, L. M. Seiford, and K. Tone, *Data Envelopment Analysis: A Comprehensive Text, with Models, Applications, References and DEA Solver Software*, Springer, New York, NY, USA, 2nd edition, 2007.
- [21] A. Charnes, W. W. Cooper, and E. Rhodes, "Evaluating program and managerial efficiency: an application of data envelopment analysis to program follow through," *Management Science*, vol. 27, no. 6, pp. 668–697, 1981.
- [22] W. W. Cooper, J. L. Ruiz, and I. Sirvent, "Choosing weights from alternative optimal solutions of dual multiplier models in DEA," *European Journal of Operational Research*, vol. 180, no. 1, pp. 443–458, 2007.
- [23] G. Yu, Q. Wei, P. Brockett, and L. Zhou, "Construction of all DEA efficient surfaces of the production possibility set under the Generalized Data Envelopment Analysis Model," *European Journal of Operational Research*, vol. 95, no. 3, pp. 491–510, 1996.
- [24] R. G. Thompson, E. J. Brinkmann, P. S. Dharmapala, M. D. Gonzalez-Lima, and R. M. Thrall, "DEA/AR profit ratios and sensitivity of 100 large US banks," *European Journal of Operational Research*, vol. 98, no. 2, pp. 213–229, 1997.
- [25] G. R. Jahanshahloo, F. H. Lotfi, N. Shoja, G. Tohidi, and S. Razavyan, "A one-model approach to classification and sensitivity analysis in DEA," *Applied Mathematics and Computation*, vol. 169, no. 2, pp. 887–896, 2005.
- [26] A. Bick, F. Yang, S. Shandalov, and G. Oron, "Data envelopment analysis for assessing optimal operation of an immersed membrane bioreactor equipped with a draft tube for domestic wastewater reclamation," *Desalination*, vol. 204, no. 1–3, pp. 17–23, 2007.
- [27] W. D. Cook and K. Bala, "Performance measurement and classification data in DEA: input-oriented model," *Omega—The International Journal of Management Science*, vol. 35, no. 1, pp. 39–52, 2007.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

