Available online at http://jnrm.srbiau.ac.ir
Vol.1, No.2, Summer 2015

Journal of New Researches in Mathematics

Science and Research Branch (IAU)

Classifying Flexible Factors Using Fuzzy Concept

S.Saati [a*], N.Imani [b]

[(a,b)] *Department of Mathematics, Tehran North Branch, Islamic Azad University, Tehran, Iran*

Received Spring 2015, Accepted Summer 2015

Abstract

In Data Envelopment Analysis (DEA), it is assumed that the role of each factor is known as input or output. However, in some cases, there are shared factors that their input versus output status is not clearly known. These are flexible measures. In such cases, determining whether a factor is input or output is ambiguous. Therefore, using fuzzy concept seems to be necessary.

In this paper, a two phase procedure is proposed to fuzzy classification of flexible measures. In the first phase, applying the existing classification methods, an orientation of flexible measures to aid in the definition of inputs and outputs is achieved. Through defining a membership function in second phase, the input versus output status of a factor is expressed by fuzzy notion. By the proposed method, the efficiency of a decision mating unit is defended by a membership degree. We illustrate the proposed model in a practical problem setting.

Keywords: *Data envelopment analysis, flexible measures, fuzzy sets, input, output membership function*

*. Corresponding Author: s_saatim@iau-tnb.ac.ir

1. Introduction

Data Envelopment Analysis (DEA) is a set of concepts and methods which provides a framework for evaluating the efficiency and inefficiency of each decision making unit (DMU) with congruent inputs and outputs. This topic was begun from 1978 by Edward Rhodes' thesis, who was guided by Cooper.

The results of Rhode's studies cooperating with Cooper and Charnes led into CCR paper (1978). The CCR paper extended the study of Farrell (1957) to several inputs and outputs to determine the efficiency of decision-making units converting multi-inputs and multi-outputs to one virtual input and one virtual output using Mathematical programming problem.

In 1984, an article was published by Banker, Charnes, and Cooper known as BCC. The difference between the BCC and CCR models was their returns to scale (RTS).

In the CCR model, RTS is constant, while in the BCC model it is variable. In data envelopment analysis, any decision-making unit is known by its inputs and outputs, and it is assumed that with a set of useable measures, the input versus output status of each of the chosen measures is clearly known. Nevertheless, in some cases, there are shared factors that the input versus output status of them is not known. These are flexible measures.

In a study which was conducted by Beasley (1990) at universities, research income was considered as an input as well as output. Also, in a conventional study about the efficiency of bank branches operations, such as discussed in Cook et al. (2000), and Cook and Hababou (2001), the standard counter transactions such as deposits and with drywalls were the outputs and resources such as various staff types were the inputs. Now, assume that one wishes to evaluate the efficiency of each DMU to attract investments. In this case, factor such as the number of ''high value'' customers, could serve as either an input or an output. On the one hand, such a measure may play the role of proxy for future investment; hence it can be reasonably classified as an output. On the other hand, it can legitimately be considered as an environmental input that aids the branch in generating its existing investment portfolio.

Nurse trainees and medical interns have a similar interpretation about the evaluation of hospital efficiency. It is important to deal with these flexible measures. Otherwise, undesirable results may be occurred. In particular, the efficiency of a decision-making unit may be the same or different when flexible measures to be considered as input or output.

Using flexible measures, Bala and Cook (2003) offered an advanced measurement tool for the assessment of banking industry.

Cook and Zhu (2007) modified DEA model with the ratio of efficiency to standard fixed production scale and introduced a modified mixed integer linear programming problem for utilizing flexibility measures including a large positive number.

Since the inclusion of a large positive number in Cook and Zhu (2007) model, in some cases, lead to inaccurate efficiency rate, so Toloo (2009) changed this model to the new Mixed Integer Linear Programming (MILP) problem that did not require such a large positive number. Toloo (2012) introduced a new classifying model that identified the shared cases and adjusts non-shared cases of flexible measures simultaneously in one step.

Since there is no consensus among the existing methods about decision-making units for determining the input versus output status of each factor, it is impossible to definitively determine the type of the factors. Furthermore, because the results of the performance appraisal models are not definite, efficient or inefficient status of units is also expressed as fuzzy. Fuzzy logic is a form of many-valued logic compared to two-valued logic, in which there is only two answers or two concepts (correct or incorrect, white or black, zero ore one). Using membership function in fuzzy logic, which assigns each member a degree of membership ranging between 0 and 1, we can determine the input versus output status of flexible measures.

In this paper, we intended to use the results of the existing methods to classify shared factors with fuzzy concept and to determine the status of these shared factors. Therefore, we used fuzzy concept to write a membership function that determined the rate of input versus output status of flexible measures. One of the advantages of this method is that we reach desired results comfortably with less calculation.

The second section of this paper contains definitions of basic concepts, Section 3 contains several models that have been presented for flexible measures previously, Section 4 covers the proposed method for determining the input versus output status of the flexible factors using fuzzy concept, and section 5 includes numerical example.

Finally, the last part includes results and references.

2. Basic Definitions

2.1. Fuzzy logic

Suppose X as an arbitrary universal set, set $\widetilde{A}$ which the degree of membership of its members is continuously between the closed interval [0,1] is called fuzzy set. This set is determined completely and uniquely by membership function shown as $\mu_{\widetilde{A}}(x)$ which is given below:

$$\tilde{A}=\left\{ (x,\mu_{\widetilde{A}}(x)) \,\middle|\, x\in X \right\}$$
$$\mu_{\widetilde{A}}(x):\ X\rightarrow[0\,,1]$$

The closer $\mu_{\widetilde{A}}$ to 1, the more x belongs to a fuzzy set and the closer $\mu_{\widetilde{A}}$ to 0, the less x belongs set $\widetilde{A}$. In limit state, If x is completely in $\widetilde{A}$, so $\mu_{\widetilde{A}}(x)=1$ and if x is not in $\widetilde{A}$ at all, so $\mu_{\widetilde{A}}(x)=0$.

Some Membership functions are as follow:

- The triangular membership function is a function that is composed of a variable (x) that depends on three parameters: *a, b,* and *c*.
- The trapezoidal membership function and π membership function are those that are composed of a variable (x) which depends on four parameters: *a, b, c,* and *d*.
- Gaussian membership function is one which depends on two parameters: c,σ And a bell-membership function depends on three parameters: *a, b,* and *c.*
- S-shaped membership function and z-shaped membership function which are composed of one variable (x), depend on two parameters: *a, b.*

To determine the membership function, a few ways have been expressed. One of the Algorithms derived from the nature is the Ant Colony Algorithm (ACO), which was presented by Dorigo, Maniezzo, and Colorni (2012). They proposed a multi-level algorithm based on ant colony algorithm to write a membership function that uses binary (zero or one) codes.

Using CGEs (examples' center of gravity), Hiroshi and Anca (1994) proposed a way to write the membership function. In this method, positive and negative examples as well as the center of gravity of them are used.

2.2. Data envelopment analysis

Data envelopment analysis (DEA), developed by Charnes et al. (1978), provides a nonparametric methodology for evaluating the efficiency of each set of comparable decision making units (DMUs), relative to one another.

Suppose we wish to evaluate the efficiencies of *n* decision making units (DMUs). Each DMU_j, (j = 1,...,n) produces

s different outputs y_{rj} (r= 1,...,s), using m different inputs x_{ij} (i = 1,..., m). Model 1 represents the CCR model in input nature.

$$\max \quad \sum_{r=1}^{s} u_r y_{ro} \tag{1}$$
$$s.t: \quad \sum_{i=1}^{m} v_i x_{io} = 1$$
$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0 \quad ,j=1,\ldots,n$$
$$u_r, v_i \ge 0 \quad r=1,\ldots,s \quad ,i=1,\ldots,m$$

In this model, $x_{io}\ (i=1,\ldots,m)$ and $y_{ro}\ (r=1,\ldots,s)$ are the inputs and outputs of the DMU under consideration.

3. Flexible Factors

In the conventional application of DEA, it is assumed that one can clearly specify which factor will constitute inputs and outputs.

However, in many problematic situations, there are some shared factors, which their input versus output status is not clearly recognizable.

These measures are considered as flexible measures. Cook et al (2007) modified DEA model with the ratio of efficiency to fixed standard production scale and introduced a mixed integer linear programming problem which includes a large positive number for using flexible measures.

$$\max \quad \sum_{r=1}^{s} \mu_r y_{ro} + \sum_{l=1}^{L} \delta_l w_{lo}$$
$$\text{s.t:} \quad \sum_{i=1}^{m} v_i x_{io} + \sum_{l=1}^{L} \gamma_l w_{lo} - \sum_{l=1}^{L} \delta_l w_{lo} = 1$$
$$\sum_{r=1}^{s} \mu_r y_{rj} + 2\sum_{l=1}^{L} \delta_l w_{lj} - \sum_{i=1}^{m} v_i x_{ij} - \sum_{l=1}^{L} \gamma_l w_{lj} \le 0, \quad j=1,\ldots,n$$
$$0 \le \delta_l \le M d_l \quad l=1,\ldots,L \tag{2}$$
$$\delta_l \le \gamma_l \le \delta_l + M(1-d_l) \quad l=1,\ldots,L$$
$$d_l \in \{0,1\}\ , \delta_l, \gamma_l \ge 0\ , \quad l=1,\ldots,L$$
$$\mu_r, v_i \ge 0\ , r=1,\ldots,s\ , \quad i=1,\ldots,m$$

In this model, there exist $w_{lo}\ (l=1,\ldots,L)$ ‘‘flexible measures’’, whose input/ output status is unknown. We denote the weights of these measures as γ_l for DMU$_j$ (l = 1,..., L). For each measure l, Cook (2007) introduced a binary variable $d_l\ (l=1,\ldots,L)$, in such a way that if d_l=1, factor l is designated as an output and if d_l=0, it is designated as an input.

However in model (2)

$\delta_l = d_l \gamma_l, (l=1,\ldots,L)$

The proposed model by Toloo (2009), is a MILP that does not require a large number M and its multiples are less than or equal to 1 as follow.

$$\max \quad \sum_{r=1}^{s} \mu_r y_{ro} + \sum_{l=1}^{L} \delta_l w_{lo} \tag{3}$$
$$\text{s.t:} \quad \sum_{i=1}^{m} v_i x_{io} + \sum_{l=1}^{L} \gamma_l w_{lo} - \sum_{l=1}^{L} \delta_l w_{lo} = 1$$
$$\sum_{r=1}^{s} \mu_r y_{rj} + 2\sum_{l=1}^{L} \delta_l w_{lj} - \sum_{i=1}^{m} v_i x_{ij} - \sum_{l=1}^{L} \gamma_l$$
$$j=1,\ldots,n$$
$$0 \le \delta_l \le d_l \quad l=1,\ldots,L$$
$$\delta_l \le \gamma_l \le \delta_l + (1-d_l) \quad l=1,\ldots,L$$
$$\mu_r \ge 0\ , v_i \le 1\ , r=1,\ldots,s\ , i=1,\ldots,m$$

Table(1), Results obtained from Model (2) using Beasley data

Decision-making unit	d	efficiency
University 1	1	1
University2	0	1
University3	0	0.837244
University4	1	0.685697
University5	0	1
University6	0	1
University7	1	1
University8	1	0.811941
University9	0	1
University10	1	0.906595
University11	0	0.890126
University12	1	0.709313
University13	0	0.803249
University14	0	0.767744
University15	0	0.704214
University16	0	0.54274
University17	1	0.819451
University18	1	0.627824
University19	1	1
University20	0	1
University21	0	0.699625
University22	1	0.716738
University23	0	0.617112
University24	0	1
University25	1	1
University26	0	1
University27	1	0.855471
University28	0	1
University29	1	0.824968
University30	0	1
University31	1	0.775853
University32	0	0.896402
University33	1	1
University34	0	1
University35	1	1
University36	0	0.8369
University37	1	0.830789
University38	0	0.833414
University39	0	0.791219
University40	1	0.741404
University41	1	1
University42	0	0.847172
University43	0	0.920638
University44	0	1
University45	0	1
University46	0	1
University47	1	0.688445
University48	0	0.938878
University49	0	1
University50	0	0.841683

4. The Proposed Method

Because the input versus output status of some of the factors is not definitely known, the results of the performance appraisal models are not definite and efficiency or inefficiency status of units is also expressed as fuzzy.

Since there is no consensus among the existing methods about decision-making units for determining the input versus output status of a factor, the status of a factor is not definitely known. So using fuzzy concept is a necessity.

Model (2) does not definitely determine the input versus output status of a factor among all units. Moreover, because it has a large positive number M, the improper selection for M can cause the efficiencies obtained is incorrect.

Model (3) can measure correctly the efficiency of decision-making units despite flexible measures, but incorrect results may be obtained due to other optimal solutions.

In this study, using results of the existing methods and defining a membership function, the input versus output status of a factor is expressed as fuzzy. The purpose of this paper was to determine the input/output status of flexible factors. In our case study, we faced sets, which determining the membership of their members was ambiguous. Thus, considering the aforementioned reasons, we decided to generate a membership function using fuzzy function and a two-valued function. This membership function determines the input versus output status of flexible measures. This function is as bellow:

$$\mu_{\tilde{A}}(d)=\begin{cases}\left(\dfrac{n-k}{n}\right)^{d}\left(\dfrac{k}{n}\right)^{1-d} & d=0\,or\,1\\ 0 & otherwise\end{cases}$$

In the above function, n indicates the total number of decision-making units and k indicates the number of states of a factor recognized as input by model 2. Where, d=1, indicates the amount of output, and d= 0, indicates the amount of input.

5. Numerical Example

With regard to the data presented in Table 1 which was obtained by running model 2 with Beasley data (1990), the number of states that the flexible factor "research cost" was recognized as input was 30 and the total number of data was 50. It means $k=30$ and $n-k=20$ if $d=1$ we have:

$$\mu_{\tilde{A}}(d)=\frac{n-k}{n}=\frac{20}{50}=0.4$$

Table (2), the efficiency of Beasley data, research cost as input or output factor

Decision-making unit	Efficiency, research method as Input factor	Efficiency, research method as output factor
University1	1	1
University2	0.615	0.640
University3	0.837	0.663
University4	0.645	0.686
University5	1	0.893
University6	1	1
University7	1	1
University8	0.750	0.812
University9	1	0.658
University10	0.892	0.907
University11	0.890	0.747
University12	0.691	0.709
University13	0.803	0.772
University14	0.768	0.702
University15	0.704	0.688
University16	0.543	0.520
University17	0.536	0.819
University18	0.593	0.628
University19	1	1
University20	0.858	0.898
University21	0.700	0.669
University22	0.664	0.717
University23	0.617	0.560
University24	0.484	1
University25	0.952	1
University26	0.425	0.565
University27	0.853	0.855
University28	1	0.809
University29	0.775	0.825
University30	0.831	0.930
University31	0.728	0.776
University32	0.896	0.841
University33	1	1
University34	1	1
University35	1	1
University36	0.837	0.735
University37	0.782	0.831
University38	0.833	0.806
University39	0.791	0.789
University40	0.740	0.741
University41	1	1
University42	0.847	0.835
University43	0.921	0.643
University44	1	1
University45	0.883	0.889
University46	0.848	0.851
University47	0.655	0.688
University48	0.939	0.883
University49	1	0.637
University50	0.842	0.835

It means that the amount of being output for the flexible factor is equal to 0.4. And if $d = 0$, so:

$$\mu_{\tilde{A}}(d) = \frac{k}{n} = \frac{30}{50} = 0.6$$

It means that the amount of being input for the flexible factor is equal to 0.6.
According to Beasley data (1990), the efficiency of universities was calculated in two states and the obtained results were provided in Table 2. Column 2 in Table 2 presents the efficiency of the universities when considering the flexible factor (research costs) as an input one and column 3 in Table 2 presents the efficiency of the universities when the research cost is an output factor.

Now, using the measures in Table 2 and the written membership function, we want to evaluate the efficiency of these units.
Unit 1 is definitely efficient If flexible factor is as an input or output factor. Considering the obtained efficiency values for Unit 2, it is clear that this unit is deficient whether flexible factor is considered as an output or an input, but the efficiency of this unit is higher when the flexible factor is considered as an output factor rather than as an input.

Therefore, the efficiency of this unit can be shown as (0.4, 0.640) when this unit is considered as an output factor and (0.615, 0.6) when this unit is considered as an input factor.

Now, paying attention to the efficiency obtained for Unit 3, we realize that this unit is inefficient in both states. But the efficiency of this unit is more when the flexible factor is considered as an input factor rather than as an output. Hence, the flexible factor is better to be considered as an input factor. In other words, the efficiency of this unit can be shown as (0.6, 0.837).

It is evident that the efficiency of Unit 5 is 0.6degree of membership when the flexible factor is considered as an input. In other words, we are not sure about the efficiency of Unit 5, so the efficiency of this unit can be shown as (0.6,1).

Regarding Unit 24, the efficiency is obvious when the flexible factor is considered as an input, but the degree of confidence about the efficiency of this unit is 0.4. So, the efficiency of this unit can be shown as (1, 0.4).

6. Conclusion

In conventional DEA models, the input versus output state of factors must be determined. However, in some cases, there are some flexible factors that their states are not determined. Some models have been proposed which were compatible with these flexible measures. In this paper, using the results of the existing models and through classifying shared factors with fuzzy concept; we determined the states of these common factors.

Writing the membership function is not always easy, for example, if we use ant algorithm to write a membership function, it is necessary to calculate fitness measures and calculating these measures is exhausting and time consuming.

Other ways such as writing a membership function with CGEs (center of gravity of examples) may seem suitable, but this way also has a lot of calculations.

Considering the aforementioned facts into account, we decided to write a membership function similar to the Bernoulli distribution function that is a two-valued one. The advantage of using this method is that we reach the desired results comfortably with less calculation.

It should be noted that different data was frequently tested in the presented function and the obtained results showed the efficiency of the discussed issues.

Acknowledgment

Useful comments from anonymous referee are gratefully acknowledged.

References

[1] Banker, R.D., Charnes, A., & Cooper, W.W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, *30*, 1078-1092.

[2] Bala, K., Cook, W. D. (2003). Performance measurement with classification information: An enhanced additive DEA model. *Omega, 31*(6), 439–450.

[3] Beasley, J. (1990). Comparing university departments. *Omega*, *8*(2), 171-183.

[4] Beasley, J.(1995). Determining teaching and research efficiencies. *Journal of the Operational Research Society*, *46*, 441-452.

[5] Charnes, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision making unit. *European journal of Operational Research*, *2*(6), 428-444.

[6] Cook, W.D., Hababou, M., & Tuenter, H. (2000). Multicomponent efficiency measurement and shared inputs in data envelopment analysis: An application to sales and service performance in bank branches. *Journal of Productivity Analysis, 14*(3), 209-224.

[7] Cook, W.D., & Hababou, M. (2001). Sales performance measurement in bank branches. *Omega*, *29*, 299-307.

[8] Cook, W.D., & Zhu, J. (2007). Classifying inputs and outputs in data envelopment analysis. *European Journal of Operational Research, 180*, 692-699.

[9] Dorigo, M., Maniezzo, V., & Colorni, A. (2012). Ant system: Optimization by a colony of cooperating agent. *IEEE Transaction on System, Man, & Cyberneties* –part B, *26*(1), 29-41.

[10] Farrell, M.J. (1957).The measurement of productive efficiency. *Journal of the Royal Statistical Society,* (A, general), *120,*253-281.

[11] Hiroshi, N., & Anca, L. R.(1994). An alternative method for inducting a membership function of a category. *International Journal of Approximate Reasoning, 11*, 1-27.

[12] Toloo, M. (2009). On classifying inputs and outputs in DEA: A revised model. *European Journal of Operational Research, 198*, 358-360.

[13] Toloo, M. (2012). Alternative solutions for classifying inputs and outputs in data envelopment analysis. *Computers and Mathematics with Applications, 63*, 1104-1110.

[14] Hong, T. P., Tung, Y. F., Wang, S. L., Wu, Y. L., & Wu, M. T. (2012). A multi-level ant-colony mining algorithm for membership function, *Information Sciences, 182*, 3-14.