INDIAN INSTITUTE OF MANAGEMENT CALCUTTA

WORKING PAPER SERIES

WPS No. 736/ October 2013

Heuristic for efficiency ranking with fuzzy data

by

Sanjeet Singh
Associate Professor, Indian Institute of Management Calcutta, Joka, Kolkata 700104

&

Surya Sarathi Majumdar
Doctoral student, IIM Calcutta D. H. Road, Joka, Kolkata 700104, India

Heuristic for efficiency ranking with fuzzy data

Sanjeet Singh and Surya Sarathi Majumdar

Indian Institute of Management Calcutta
Diamond Harbor Road, Joka, Kolkata 700104, India.

Abstract

In this paper, we have developed a heuristic that uses Data Envelopment Analysis (DEA) to rank Decision Making Units (DMUs) with fuzzy input and output values in the order of their probable efficiency. The heuristic is designed to be used with the raw samples of the fuzzy data, and does not require knowledge of the fuzzy data's probability distribution. It is meant to be a simple method of differentiating fuzzy DMUs, rather than a replacement for more complex and rigorous methods of analysis. Working of the heuristic has been demonstrated using numerical examples from existing works on fuzzy DEA.

Keywords: DEA, fuzzy DEA, efficiency, linear programming, heuristic, efficiency ranking

1. Introduction

Data Envelopment Analysis (DEA) is a widely applied non-parametric mathematical programming technique to calculate the relative efficiency of firms/organizations or Decision Making Units (DMUs) operating in a similar environment and utilizing multiple inputs to produce multiple outputs. Based on Farrell's (1957) work on productive efficiency, DEA was first introduced by Charnes, Cooper, and Rhodes (1978). Efficiency obtained using DEA, in its simplest form, is the comparison of the weighted output to weighted input ratio of the observing DMU with that of the best practice in the group. Measurement of efficiency is important to shareholders, managers, and investors for any future course of action. DEA has been extensively applied to a wide spectrum of practical problems. Examples include financial institutions (Lamb and Tee 2012), bank failure prediction (Barr, Seiford and Sims 1994), electric utilities evaluation (Goto and Tsutsui 1998), textile industry performance (Chandra et al. 1998), portfolio evaluation (Murthi, Choi and Desai 1997).

The basic DEA models start with the Charnes Cooper Rhodes (CCR) models (1978). The CCR models assume that the production function exhibits Constant Returns to Scale (CRS). These models use linear programming to determine the optimum virtual weights for the parameters in order to calculate the best possible efficiency score for each DMU in comparison to the other DMUs. These models can be oriented to either minimizing the inputs while keeping outputs constant (input oriented model) or vice versa (output-oriented). The Banker-Charnes-Cooper (BCC) models (Cooper, Seiford and Tone 2007) are an extension of the basic CCR model, designed to be used when the DMUs are under Variable Returns to Scale (VRS). Other models include the Additive models (Tone 2001), which focus directly on the slacks in the system as a means of determining efficiency. A further difference is that CCR and BCC focus on radial projection whereas additive models look at non-radial projection. The two approaches are combined in the Hybrid model (Cooper, Seiford and Tone 2007). Further developments in DEA involve generalizing these models for wider application under different sets of assumptions.

All standard DEA models have one thing in common – they are designed for the use with crisp input and output values of the various DMUs. However, in real life, it is often impossible to obtain exact values of the parameters, such as in Charnes, Gallegos, and Li's (1996) study on Latin American airlines. This led to the development of various methods to deal with fuzziness in the data. Methods include stochastic frontier methods (Bauer 1990) to provide crisp efficiency results. Other methods involve fuzzy linear programming models with predefined possibility levels to provide fuzzy efficiency values (Kao and Liu 2000, Guo and Tanaka 2001).

These fuzzy and stochastic DEA methods all rely on being able to assign some sort of probability distribution to the fuzzy values. Guo and Tanaka (2001), for example, assign each fuzzy value a symmetrical triangular distribution. These probability distributions are generally arrived at by polling and sampling the DMUs, and such extensive sampling can be both expensive and time-consuming. Gathering enough samples to develop an accurate distribution is often impossible, such as any study which needs to be conducted in a fixed amount of time. The heuristic proposed in this paper is a method to rank fuzzy DMUs in order of their efficiencies without having to first deduce or assume a probability distribution for the fuzzy data. Instead, the raw sampling data can be utilized by this heuristic to rank the various DMUs. The heuristic is meant to be used in situations where the user is unwilling or unable to extrapolate probability distributions for the fuzzy data, such as in cases with small number of samples. The proposed heuristic's results can be utilized in various different ways, depending on the user's preferences. For example, this heuristic can be used to rank DMUs in the order of efficiency, or to simply determine which DMUs are more likely to be efficient in the best-case scenario.

2. The Heuristic Algorithm

Following notations have been used throughout this study:

θ_k : Efficiency of the kth DMU.

θ_{ik} : Efficiency of the kth VDMU generated by the ith DMU

p_{ik}: Probability that the kth VDMU generated by the ith DMU will occur.

u_r : The weight assigned to the rth output.

v_j : The weight assigned to the jth input.

u_0 : Value representing the "variable" part of variable returns to scale DEA models like BCC.

x_{ij} : jth input of the ith DMU.

y_{ir} : rth output of the ith DMU.

n : The number of DMUs.

m : The number of inputs.

s : The number of outputs.

This heuristic is designed for those scenarios where the data is fuzzy, but not sufficient information exists to determine the distribution function for the data. This heuristic can be applied using the direct samples made to determine the data values of the DMUs, and can be applied to a mixture of crisp and fuzzy data.

Step 1: The first step of the heuristic is to generate virtual DMUs (VDMUs) representing every possible scenario for every DMU that is under consideration. If, for example, the decision maker has three possible values for an output of a particular DMU, and two possible input values, then there will be 3 x 2 = 6 VDMUs, one each corresponding to each possible combination of input and output values, generated by the DMU.

Step 2: After the VDMUs are generated in step 1, we apply the following DEA model to determine the efficient frontier of all these VDMUs.

$$\theta_k = \text{Max} \frac{\sum_{r=1}^{s} u_r y_{kr} + u_0}{\sum_{j=1}^{m} v_j x_{kj}}$$

subject to

$$\frac{\sum_{r=1}^{s} u_r y_{ir} + u_0}{\sum_{j=1}^{m} v_j x_{ij}} \leq 1, \text{ for } i = 1, 2, \ldots, n,$$

v_j, $u_r \geq 0$ *for all* $j=1,\ldots m$, $r = 1,\ldots,s$; u_0 *free in sign.*

This formulation is the standard Banker-Charnes-Cooper (BCC) model and is designed for situations where the DMUs are under variable returns to scale (VRS). If the DMUs being studied are under Constant Returns to Scale (CRS) instead, then we use the CCR model by adding the constraint $u_0 = 0$ to the above model.

Step 3: The results of step 2 are analyzed to identify the most efficient DMUs. This is done by determining the probability of the efficient VDMUs occurring, then the probabilities of the efficient VDMUs which belong to the same DMU are added together to get the probability of that DMU giving an efficient outcome. Eg:

Let the k^{th} DMU have generated d VDMUs.
Let VDMUs 1, ..., c be on the efficient frontier, and $c+1,\ldots, d$ be inefficient.

Then probability of the k^{th} DMU being efficient = $\sum_{i=1}^{c} p_{ki}$

This will allow us to rank investments from highest to lowest probability of efficient returns. We may also rank the DMU using other criteria. Alternative ranking methods are described in the next section.

3. Alternative methods of analysis

Step 3 of the heuristic in section 2 applies a simple probability calculation to rank the various investments by the probability of the investment having an efficient outcome. However, there are situations where this method does not give satisfactory results. One such situation would be when there are investments which have low-probability high-return scenarios. The DMUs representing these scenarios would define the efficiency frontier, but being low-probability scenarios, it would mean no investment has a high probability of being efficient. Thus, a number of alternate approaches to step 3 of the heuristic has been described below.

Elimination of outliers: If a few low-probability high-return scenarios are dominating the efficiency frontier, then one approach is to eliminate all the VDMUs on the efficient frontier from consideration, and reapply the DEA model. The probabilities attached to the eliminated VDMUs are directly added to their parent DMUs' probability of being efficient. This allows us to account for the presence of the outlier without having it interfere with the DEA formulation. Eg:
Let the k^{th} DMU have generated d VDMUs.
Let VDMUs 1, ..., c be on the efficient frontier, and $c+1,\ldots, d$ be inefficient.
Eliminating VDMUs 1 to c from consideration, we reapply the DEA model. Then, let VDMUs $c+1,\ldots, e$ be efficient and VDMUs $e+1,\ldots, d$ be inefficient

At this point, the probability of the k^{th} DMU being efficient is $\sum_{i=1}^{c} p_{ki} + \sum_{i=c+1}^{e} p_{ki} = \sum_{i=1}^{e} p_{ki}$

Determine efficiency slabs for all VDMUs: The VDMUs of each DMU can be differentiated on what efficiency slab they fall on (e.g. θ_{ik} between 75% and 80%, above 90%). Using this method instead of focusing solely on the efficiency frontier allows a broader picture and to look at the overall performance of a DMU instead of only its best-case scenarios.

Rank by mean and standard deviation of VDMUs: The DMUs can be ranked by the mean efficiency and standard deviation of the VDMUs generated by it. Eg:
Let there be d VDMUs generated by the k^{th} DMU.

Mean Efficiency (ME) = $\frac{1}{d}\sum_{i=1}^{d}\theta_{ki}$

Standard deviation = $\sqrt{\frac{\sum_{i=1}^{d}(\theta_{ki} - ME)^2}{d}}$

A high mean efficiency with a low standard deviation indicates a DMU of consistent efficiency.

Identify the investments which are efficient with a minimum p% probability: If the decision maker is interested in those DMUs which have a certain minimum percentage probability of being on the efficiency frontier. If no DMU meets this criterion, then we will apply the same method as described under 'Elimination of outliers'. We will then continue to reapply this method until the VDMUs for a DMU on the efficiency frontier reach the minimum acceptable probability of being efficient.

4. Numerical Illustration

To demonstrate the heuristic, we will use data from existing works on fuzzy DEA (Kao and Liu 2000; Guo and Tanaka 2001). This will allow us to compare the heuristic results with the results from these works. The efficiency calculations are performed using Coelli's (1996) DEA software.

4.1 Single-input single-output

The data is obtained from Kao and Liu's (2000) work on fuzzy DEA, and is shown in the following table.

Table 1. Single-input single-output fuzzy data

DMU	Possible Inputs	Possible Outputs
A	11, 12, 14	10
B	30	12, 13, 14, 16
C	40	11
D	45, 47, 52, 55	12, 15, 19, 22

According to step 1 of the heuristic, VDMUs are generated corresponding to each DMU for all possible combinations of input and output values of that DMU. Thus, DMU A gives 3 VDMUs, B gives 4, C gives 1, and D gives 16. The data is considered to be under VRS, so at step 2, the DEAP software is configured for VRS efficiency calculation. The efficiency calculation (as displayed in table 2) shows that 1 VDMU from A (p_{11}= 1/3) and 1 VDMU from D (p_{41}= 1/16) are on the efficiency frontier. Of the other DMUs, the efficiency of C is below every possible efficiency value of B. At this point, the heuristic shows that A has the highest probability of being efficient (1/3), followed by D (1/16). This is in line with the results of Kao and Liu (2000), where the results were A>D>B>C in terms of efficiency. The full results of the heuristic are displayed in Fig. 1 and Table 2.

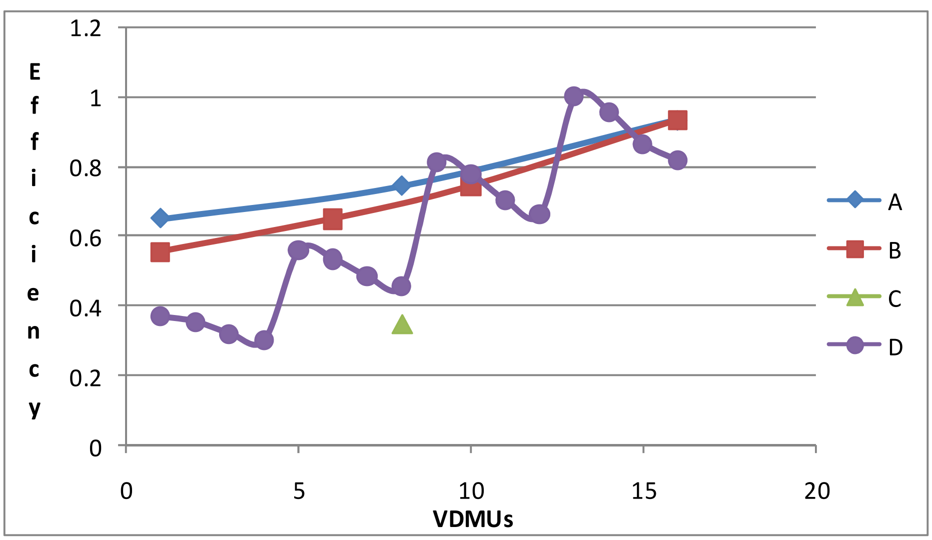


Fig. 1. Efficiency of the VDMUs

Table 2. Efficiency distribution of the VDMUs

Efficiency	VDMU(A)	VDMU(B)	VDMU(C)	VDMU(D)
θ_{ik} = 1	1	0	0	1
1> θ_{ik} >=0.9	1	1	0	1
0.9> θ_{ik} >=0.8	0	0	0	3
0.8> θ_{ik} >=0.7	1	1	0	2
0.7> θ_{ik} >=0.6	-	1	0	1
0.6> θ_{ik} >=0.5	-	1	0	2
θ_{ik} <0.5	-	-	1	6
Mean	0.901	0.721	0.346	0.624
St. Dev.	0.108	0.161	0.0	0.232

Once we consider the overall picture, it is revealed that DMU D is a better performer than B under best-case scenarios with a 5/16 chance of showing efficiency within 80% as opposed to B's 1/4. But overall, D has a lower average and wider variation in its performance. Thus, going by averages, the efficiency ranking of the DMUs is A>B>D>C.

4.2 Multiple inputs multiple outputs

The next example uses two inputs and two outputs. We obtain this data from Guo and Tanaka's (2001) work on fuzzy DEA. The data is given in Table 3, with the fuzzy values expressed as (central value; spread). Eg: (4.1; 0.7) means the value has a center at 4.1 and spread 0.7, meaning it ranges from 3.4 to 4.8.

Table 3. Two input two output fuzzy data for 5 DMUs

DMUs	A	B	C	D	E
Input 1	(4.0; 0.5)	(2.9; 0.0)	(4.9; 0.5)	(4.1; 0.7)	(6.5; 0.6)
Input 2	(2.1; 0.2)	(1.5; 0.1)	(2.6; 0.4)	(2.3; 0.1)	(4.1; 0.5)
Output 1	(2.6; 0.2)	(2.2; 0.0)	(3.2; 0.5)	(2.9; 0.4)	(5.1; 0.7)
Output 2	(4.1; 0.3)	(3.5; 0.2)	(5.1; 0.8)	(5.7; 0.2)	(7.4; 0.9)

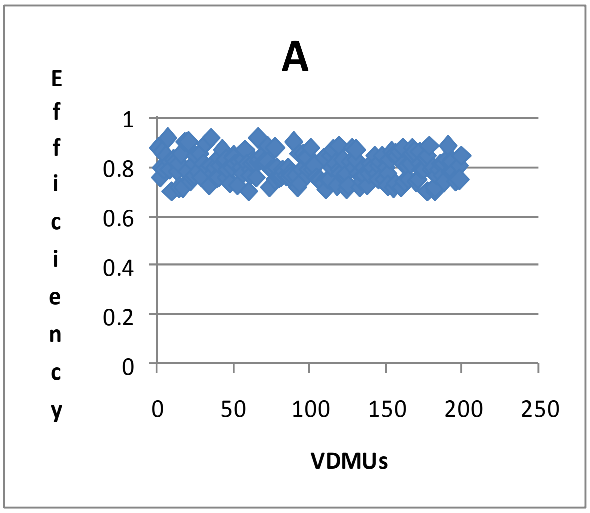
We use a Monte Carlo simulation to generate possible input and output values. We generate 200 VDMUs for each of the 5 DMUs, the input and output values being randomly determined from the given fuzzy ranges. Each VDMU has p_{ki} = 1/200 = 0.005. After applying the heuristic, the number of efficient VDMUs generated by each DMU and the probability of efficiency is as shown in Table 4:

Table 4. Result of Heuristic applied to table 3 data

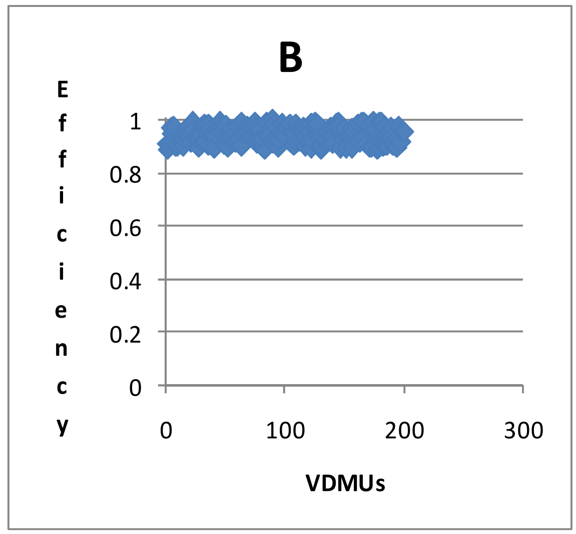
DMU	A	B	C	D	E
No. of efficient VDMUs (out of 200)	0	0	3	6	3
Probability of efficiency $\sum_{i=1}^{c} p_{ki}$	0	0	0.015	0.03	0.015

It is clear from table 4, that under best-case scenarios, D is the most efficient DMU. However, none of the DMUs are represented on the efficiency frontier in great numbers, with D having the highest probability of 3%. Thus while D has a clear advantage, it is difficult to rank the others. To gain a better view of the overall picture, the efficiency distribution of the VDMUs of each DMU is displayed in Fig. 2 (a)-(e).

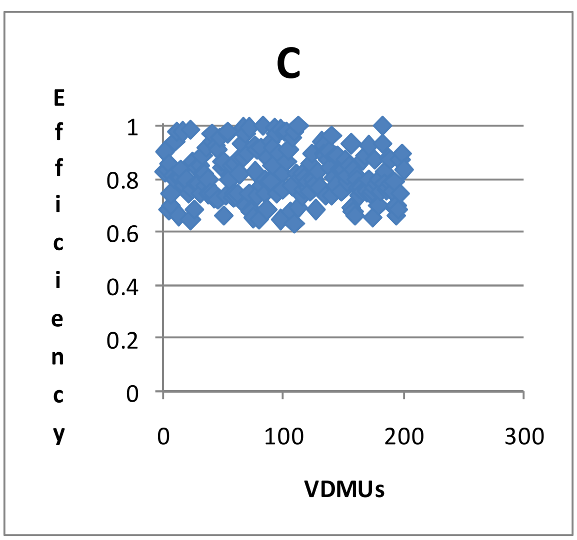
Fig 2. Efficiency chart for the VDMUs of the 5 DMUs



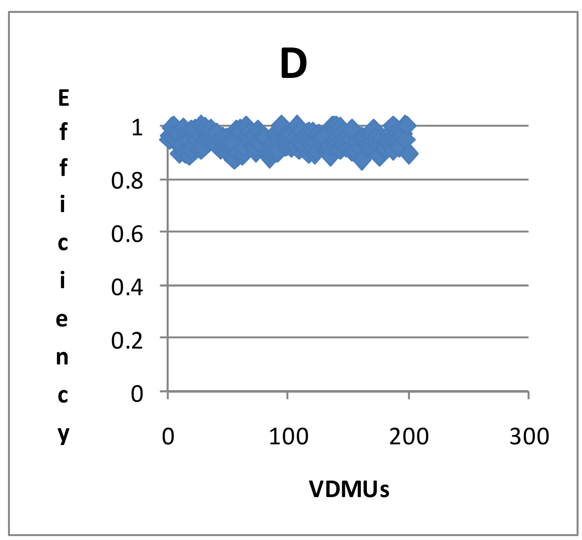
(a)



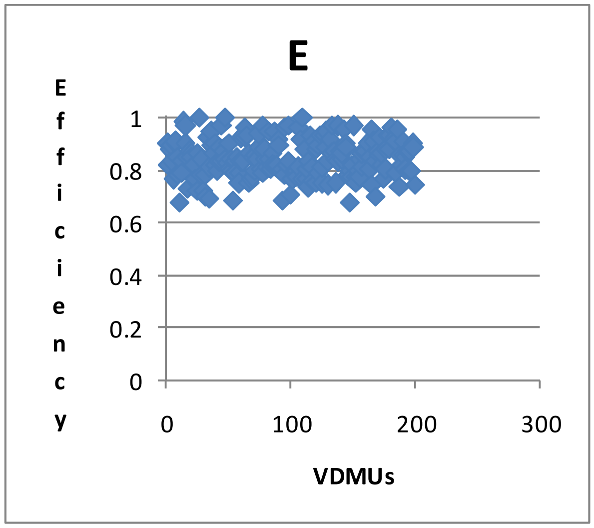
(b)



(c)



(d)



(e)

Table 5. Efficiency distribution of VDMUs in 2-input 2-output example

Efficiency	VDMU(A)	VDMU(B)	VDMU(C)	VDMU(D)	VDMU(E)
θ_{ik} = 1	0	0	3	6	3
1> θ_{ik} >=0.9	6	178	38	175	44
0.9> θ_{ik} >=0.8	87	22	62	19	95
0.8> θ_{ik} >=0.7	107	-	73	-	51
θ_{ik} <0.7	-	-	24	-	7
Mean	0.799	0.942	0.817	0.941	0.844
St. Dev.	0.051	0.030	0.092	0.030	0.073

Analyzing the results in Table 5, there is hardly any difference between D and B as both show similar performance. The difference is that DMU D shows up more often on the efficiency frontier, meaning it has a better chance of outperforming all other DMUs. Thus, D is ranked higher than B. Between C and E, E has a small but measurable advantage in overall performance (higher average, lower variance), and so is ranked higher, with A in last place. The ranking, according to our heuristic, is D>B>E>C>A.

In Guo and Tanaka's (2001) study, the ranking according to their method is either B>D>E>C>A or B>E>C>D>A, depending on their evaluation method. Thus, apart from the position of D, our heuristic agrees with their analysis. This discrepancy most likely arises from the fact that they are using fuzzy numbers with a particular probability distribution, and the different evaluation methods in their paper represent different possibility levels, whereas the heuristic in this paper automatically uses the highest possibility level.

5. Conclusion

In this paper, a heuristic is described for ranking the efficiency of DMUs containing fuzzy data. The heuristic is designed to work with raw sampling data. This gives it an advantage over other forms of fuzzy DEA which normally require some knowledge as to the probability distribution of the fuzzy data. The heuristic is designed to be quick and simple method of differentiating among fuzzy DMUs, so as to allow decision makers to focus on the DMUs which are of the most interest. The heuristic can work with VRS or CRS, and can work with a mixture of fuzzy and crisp data. We adapt two numerical examples from two different studies on fuzzy DEA, and compare the results of the heuristic with the results of the original studies. The results of the heuristic are mostly in keeping with the original results, but there are some deviations. These differences are most likely the result of the way the original examples have been adapted for use by this heuristic, but a closer study may be conducted.

References

Barr R.S., Seiford L.M., Sims T.F., 1994. *Forecasting Bank Failure: A Non-Parametric Frontier Estimation Approach*, Louvain Economic Review, Vol. 60, No. 4, pp. 417-429.

Bauer P.W., 1990. *Recent developments in the econometric estimation of frontiers*, Journal of Econometrics, Vol. 46, pp. 39-56.

Chandra P., Cooper W.W., Li S., Rahman A., 1998. *Using DEA To evaluate 29 Canadian textile companies — Considering returns to scale*, International Journal of Production Economics, Vol. 54, Issue 2, pp. 129-141.

Charnes A., Cooper W.W., Rhodes E., 1978. *Measuring the Efficiency of Decision Making Units*, European Journal of Operational Research, Vol. 17, Issue 1, pp. 35-44.

Charnes A., Gallegos A., Li H., 1996. *Robustly efficient parametric frontiers via multiplicative DEA for domestic and international operations of the Latin American airline industry*, European Journal of Operational Research, Vol. 88, pp. 525-536.

Coelli, T., 1996. *A Guide to DEAP Version 2.1: A DataEnvelopment Analysis (Computer) Program*, CEPA Working Paper 96/08, Centre for Efficiency and Productivity Analysis, University of New England.

Cooper W.W., Seiford L.M., Tone K., 2007. *Data Envelopment Analysis 2nd Ed.*, Springer Business & Media LLC, New York.

Farrell M.J., 1957. *The Measurement of Productive Efficiency,* Journal of the Royal Statistical Society, Series A, General, Vol. 120, pp. 253-282.

Goto M., Tsutsui M., 1998. *Comparison of Productive and Cost Efficiencies among Japanese and US Electric Utilities*, Omega, Vol. 26, Issue 2, pp. 177-194.

Guo P., Tanaka H., 2001. *Fuzzy DEA: A Perceptual Evaluation Method*, Fuzzy Sets and Systems, Vol. 119, pp. 149-160.

Kao C., Liu S., 2000. *Fuzzy Efficiency Measures in Data Envelopment Analysis*, Fuzzy Sets and Systems, Vol. 113, pp. 427-437.

Lamb J.D., Tee K.H., 2012. *Data envelopment analysis models of investment funds*, European Journal of Operational Research, Vol. 216, Issue 3, pp. 687-696.

Murthi B.P.S., Choi Y.K., Desai P., 1997. *Efficiency of mutual funds and portfolio performance measurement: A non-parametric approach*, European Journal of Operational Research, Vol. 98, Issue 2, pp. 408-418.

Tone K., 2001. *A slacks-based measure of efficiency in data envelopment analysis,* European Journal of Operational Research, Vol. 130, Issue 3, pp. 498-509.