Emerald Insight

Kybernetes

Fuzzy C-means based data envelopment analysis for mitigating the impact of units' heterogeneity

Seyed Hossein Razavi Hajiagha Shide Sadat Hashemi Hannan Amoozad Mahdiraji

Article information:

To cite this document:
Seyed Hossein Razavi Hajiagha Shide Sadat Hashemi Hannan Amoozad Mahdiraji , (2016),"Fuzzy C-means based data envelopment analysis for mitigating the impact of units' heterogeneity", Kybernetes, Vol. 45 Iss 3 pp. 536 - 551
Permanent link to this document:
http://dx.doi.org/10.1108/K-07-2015-0176

Downloaded on: 14 November 2016, At: 22:04 (PT)
References: this document contains references to 38 other documents.
To copy this document: permissions@emeraldinsight.com
The fulltext of this document has been downloaded 44 times since 2016*

Users who downloaded this article also downloaded:

(2016),"An approach to evaluation of emergency plans for unconventional emergency events based on soft fuzzy rough set", Kybernetes, Vol. 45 Iss 3 pp. 461-473 http://dx.doi.org/10.1108/K-03-2014-0055

(2016),"A self-organising network model of decision making by the honey bee swarm", Kybernetes, Vol. 45 Iss 3 pp. 358-370 http://dx.doi.org/10.1108/K-12-2014-0290

Access to this document was granted through an Emerald subscription provided by emerald-srm:563821 []

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/0368-492X.htm

Fuzzy C-means based data envelopment analysis for mitigating the impact of units' heterogeneity

Seyed Hossein Razavi Hajiagha
Department of Management, Khatam Institute of Higher Education, Tehran, Iran
Shide Sadat Hashemi
Faculty of Management and Accounting, Allameh Tabatabaei University, Tehran, Iran, and
Hannan Amoozad Mahdiraji
Department of Management, Kashan Branch, Islamic Azad University, Kashan, Iran

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

Abstract

Purpose – Data envelopment analysis (DEA) is a non-parametric model that is developed for evaluating the relative efficiency of a set of homogeneous decision-making units that each unit transforms multiple inputs into multiple outputs. However, usually the decision-making units are not completely similar. The purpose of this paper is to propose an algorithm for DEA applications when considered DMUs are non-homogeneous.

Design/methodology/approach – To reach this aim, an algorithm is designed to mitigate the impact of heterogeneity on efficiency evaluation. Using fuzzy C-means algorithm, a fuzzy clustering is obtained for DMUs based on their inputs and outputs. Then, the fuzzy C-means based DEA approach is used for finding the efficiency of DMUs in different clusters. Finally, the different efficiencies of each DMU are aggregated based on the membership values of DMUs in clusters.

Findings – Heterogeneity causes some positive impact on some DMUs while it has negative impact on other ones. The proposed method mitigates this undesirable impact and a different distribution of efficiency score is obtained that neglects this unintended impacts.

Research limitations/implications – The proposed method can be applied in DEA applications with a large number of DMUs in different situations, where some of them enjoyed the good environmental conditions, while others suffered from bad conditions. Therefore, a better assessment of real performance can be obtained.

Originality/value – The paper proposed a hybrid algorithm combination of fuzzy C-means clustering method with classic DEA models for the first time.

Keywords Cluster analysis, Data envelopment analysis, Fuzzy C-means algorithm, Heterogeneous units

Paper type Research paper

1. Introduction

Resource scarcity is the main concern of almost all scientific methods in economy and management theories, making the art of proper resource consumption as a critical factor for organizations competitions. Following this phenomenon results in the concept of efficiency, and we confront the challenge that whether organizational resources assignment is working properly or not. The concept of production function is introduced as a tool to appraise efficiency, and a majority of related methods are based on the approximation of this function. Considering the role of production function,

Kybernetes
Vol. 45 No. 3, 2016
pp. 536-551
© Emerald Group Publishing Limited
0368-492X
DOI 10.1108/K-07-2015-0176

we can classify the efficiency evaluation methods into two types: first, parametric methods seeking to approximate the production function (Kumbhakar and Knox Lovell, 2003); and second, non-parametric methods which indirectly approximate this function. Data envelopment analysis (DEA) is one of the most well-known and widely accepted methods in non-parametric class (Ray, 2004).

Farrell (1957) introduced a method of efficiency evaluation known as the origin of DEA. He decomposed the efficiency of each unit into two technical and assignment components. Later, Charnes *et al.* (1978) developed the DEA method based on the Farrell's model. The first DEA model was called CCR model due to its authors. After 1978, the DEA method was widely known and accepted as a permanent paradigm in efficiency evaluation. Emrouznejad *et al.* (2008) and Liu *et al.* (2013) surveyed more than thousands of papers and applications of DEA in different fields.

A DEA problem can be defined as follows: suppose that there are n homogeneous decision-making units DMU_j, $j = 1, 2, \ldots, n$, where each DMU_j used an m-dimensional vector x_j as its inputs to produce an s-dimensional vector y_j as outputs. The DEA seeks to find the best efficiency of DMUs by maximizing each DMU's individual efficiency, while the efficiency of all units is required to be less than unity.

A main advantage of DEA is that it does not require any specific statistical distribution for inputs and outputs. In addition, the form of relation between inputs and outputs is free. Nonetheless, as previously mentioned, DEA is developed under the assumption of DMUs homogeneity, dealing with two aspects: using similar inputs and outputs; and having the same functional and operational characteristics.

Furthermore, the practical applications of DEA are mainly on units with similar nature, like banks, hospitals, and, etc.; nevertheless, heterogeneity seems an inevitable feature of practice. Regarding this fact, some bank branches have been taken into consideration. Although they look like each other, they are influenced by various factors such as socio-economic situation, local culture, size, and so on, being capable of entirely changing their efficiency.

DEA is a set of linear programming-based methods to evaluate the efficiency of a group of homogeneous units by using a set of inputs to produce a set of outputs. DEA considers the efficiency of each unit as the ratio between its weighted sums of outputs to the weighted sums of inputs. In contrast with the classical methods of constant weights, DEA allows each unit to take its variable weights in such a way that its efficiency is maximized, while the efficiency of all units is constrained to be less than one. It can be concluded that the DEA weights are closely related to their inputs and outputs data, and a small swing in units' data will have a great influence on the DEA results. Dyson *et al.* (2001) reviewed the basic assumption of classic DEA method; moreover, Brown (2006) emphasized the pitfall of this assumption. Homogeneity of DMUs is one of these assumptions dealing with homogeneity in the activities and sources of DMUs. However, a wide variety of practical applications were included, considering a set of non-homogeneous or heterogeneous units. This heterogeneity can be raised from the scale of units' activity, i.e. two different bank branches with different sizes, or different types of activities, i.e. different departments of a university. If the heterogeneous DMUs are assessed by DEA without any modifications, the DEA yields biased performance scores and inaccurate analyses (Sharma and Jin, 2011). Dyson *et al.* (2001) argued that classic DEA models should be modified to deal with heterogeneous units.

Some approaches are proposed to deal with the problem of heterogeneous DMUs. Haas and Murphy (2003) compared three different methods to compensate for the non-homogeneity. These three methods include the two-sage method of Sexton *et al.* (1994)

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

along with two additional methods of the magnitude of error and the ratio of actual to forecast. In practice, they advised trying all methods or a basket of methods and comparing the results to one's knowledge of the actual situation. Sengupta (2005) investigated two types of heterogeneity including: first, the problem of heteroscedasticity that arises when data set comprises several clusters rather than one and the variances are not constant across clusters; and second, the different size of DMUs. Two sets of transformations are introduced, one which reduces heterogeneity of the data set by choosing the appropriate model, for instance, quadratic or log-linear cost frontier and then by applying a smoothing technique, and the other one which applies the standard statistical tests of heteroscedasticity to the regression equations using DEA results and testing the pattern of variations of the squared residuals. Farzipoor Saen *et al.* (2005) proposed a modification of DEA for slightly non-homogeneous units. They indicated that after inserting the missing values by series mean, the weights of DMUs are measured by analytic hierarchy process and then, the relative efficiency of DMUs is computable by chance-constrained DEA.

An interesting idea to deal with heterogeneous DMUs is to use the clustering analysis approach. Samoilenko and Osei-Bryson (2008) proposed a three-step methodology that allows an increase in discriminatory power of DEA in the presence of the heterogeneity. First of all, the cluster analysis (CA) is applied to test for the presence of the naturally occurring subsets in the sample. In the second phase, DEA is performed to calculate the relative efficiency of the DMUs, as well as averaged relative efficiency of each subset identified in the previous phase. Eventually, decision tree is used to examine the subset-specific nature of the relative efficiency of the DMUs in the sample. Samoilenko and Osei-Bryson (2010) proposed a five-step algorithm and augmented DEA with CA and neural networks to determine whether the difference in the scores of scale heterogeneous DMUs is due to the heterogeneity of the levels of inputs and outputs, or it is caused by the conversion efficiency of inputs into outputs.

Exact CA had a binary characteristic in which a DMU could be in a specific cluster or not. Such a dichotomous nature is difficult in practice. Admittedly, when DMUs are classified in different groups, it seems more realistic that a membership degree be assigned to each DMU in different groups. This situation can be characterized as a fuzzy clustering analysis (FCA). This paper proposed the augmentation of DEA with FCA to handle the heterogeneous DMUs. Therefore, the main advantage of the current paper is to develop a fuzzy framework to deal with DMUs heterogeneity, in which DMUs are classified in different groups according to their similarity, while a membership degree is attained to each DMU in each cluster which shows its belongingness to that cluster. These belongingness degrees are then applied in efficiency appraisal of DMUs to lower the impact of situational differences in their efficiencies. The rest of the paper is organized as follows. Section 2 includes an overview of research techniques including DEA and FCA. The research algorithm is explained in Section 3. In Section 4, the application of algorithm is investigated in a practical case. Finally, the conclusion is presented in the last section.

2. Overview of research techniques

2.1 DEA

DEA is proposed as a linear programming-based technique to measure the relative efficiency of a group of homogeneous decision-making units, employing an m-dimensional input vector to produce an s-dimensional output vector. The initial CCR model of Charnes, Cooper, and Rhodes is developed based on the characteristics of production possibility sets. Besides, DEA models are classified based on their orientations: input oriented, output oriented, and base oriented (Charnes *et al.*, 1994).

Regardless of their types, DEA models generally try to find the optimal weights of inputs and outputs of a unit in such a way that its relative efficiency is maximized. Suppose that there are n decision-making units, $DMU_j, j=1, 2, \ldots, n$, where the jth unit and DMU_j used the input vector $x_j=(x_{1j}, x_{2j}, \ldots, x_{mj})$ to produce the output factor $y_j=(y_{1j}, y_{2j}, \ldots, y_{sj})$. Following the original DEA model, the CCR model aggregated the multiple inputs and multiple outputs and constituted a ratio as follows:

$$E_j = \frac{\sum_{r=1}^{S} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \tag{1}$$

where u_r, $r=1, 2, \ldots, s$ represents the output value and v_i, $i=1,2, \ldots, m$ illustrates the input value in determination of the relative efficiency of DMU_j. Adding the normalization constraint that the relative efficiency of each unit is bounded above to 1, and using the Charnes and Cooper (1962) transformation technique, the following multiplier form of input-oriented CCR model is obtained to evaluate the relative efficiency of $DMU_0, 0 \in \{1, 2, \ldots, n\}$:

$$Max \sum_{r=1}^{s} u_r y_{r0} \sum_{i=1}^{m} v_r x_{i0} = 1 \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_r x_{ij} \leqslant 0,\ j=1,2,\ \ldots,\ n$$

$$u_r \geqslant 0,\ r=1,2,\ \ldots,\ s\, v_r \geqslant 0,\ i=1,2,\ \ldots,\ m \tag{2}$$

where, $(x_{10}, x_{20}, \ldots, x_{m0})$ and $(y_{10}, y_{20}, \ldots, y_{s0})$ are the observed inputs and outputs of DMU_0. The dual form of above model called envelopment model as follows:

$$\begin{aligned} & Min\ \theta_0 \\ & \sum_{j=1}^{n} \lambda_j x_{ij} \leqslant \theta_0 x_{i0},\ i=1,2,\ \ldots,\ m \\ & \sum_{j=1}^{n} \lambda_j y_{rj} \geqslant y_{r0},\ r=1,2,\ \ldots,\ s \\ & \lambda_j \geqslant 0,\ j=1,2,\ \ldots,\ n \\ & \theta_0 \text{ unrestricted in sign} \end{aligned} \tag{3}$$

where θ_0 is the input-oriented CCR efficiency of DMU_0. Solving Equation (3) for different DMUs, we determined their relative efficiency.

If θ_0^* is the optimal objective value of Equation (2) and s_i^{-*}, $i=1,2,\ \ldots,\ m$ and s_r^{-*}, $r=1,2,\ \ldots,\ s$ are its input constraints and output constraints slacks, respectively, a DMU can be classified in one of the following classes (Cooper *et al.*, 2007):

(1) If $\theta_0^*=1$ and all the input and output slacks are zero, the considered DMU is known as strong efficient.

(2) If $\theta_0^*=1$ and at least one of the input and output slacks are positive, the considered DMU is known as weak efficient.

(3) If $\theta_0^*<1$, the considered DMU is known as inefficient.

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

The envelopment form of CCR output-oriented model is presented as follows:

$$\begin{aligned} & Max\ \rho_0 \\ & \sum_{j=1}^{n} \lambda_j x_{ij} \leqslant x_{i0},\ i = 1, 2,\ \ldots,\ m \\ & \sum_{j=1}^{n} \lambda_j y_{rj} \geqslant \rho_0 y_{r0},\ r = 1, 2,\ \ldots,\ s \\ & \lambda_j \geqslant 0,\ j = 1, 2,\ \ldots,\ n \\ & \rho_0 \text{ unrestricted in sign} \end{aligned} \quad (4)$$

where ρ_0 is the output-oriented CCR efficiency of DMU_0. The result of output-oriented model classifies DMUs like input-oriented model as discussed above.

In this paper, the set of input-oriented or output-oriented BCC models are employed.

2.2 FCA

CA is a technique for partitioning or classification (Everitt *et al.*, 2011; Mirkin, 2012).

Different clustering methods are proposed. A general classification of clustering techniques is based on hard fuzzy clustering (Bezdek, 1981; Dave, 1992). Hard (crisp) clusters are defined by Boolean indicator function in which a specific objective deterministically belongs to a given cluster or not. On the other hand, fuzzy clusters are defined by fuzzy indicator functions, where each objective belongs to a given cluster with a degree between 0 and 1 (Mirkin, 1996). Inspired by fuzzy set theory, the FCA does not consider a specific and hard border between clusters; hence, units can be considered as members of different clusters with their corresponding membership degrees.

In this paper, the fuzzy C-means algorithm is used to cluster the DMUs of a DEA study. Among different FCA algorithms fuzzy C-means is the most well-known method due to having the advantage of robustness for ambiguity and for maintaining much more information than any hard clustering methods (Pham, 1999).

The fuzzy C-means algorithm is a type of objective function-based algorithm (Bezdek, 1981). If $X = \{x_1, x_2, \ldots, x_n\}$, where $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$, $i = 1,2, \ldots, n$, is the collection of data, and c, $2 \leqslant c \leqslant n$, is an integer number, the fuzzy C-means algorithm seeks to compute a fuzzy c partition of X, represented by $U = [u_{ij}] \in W_{c \times n}$, where $0 \leqslant u_{ij} \leqslant 1$ is the membership of x_i in the jth cluster, with the following conditions:

$$\sum_{j=1}^{c} u_{ij} = 1,\ \ i = 1, 2,\ \ldots,\ n \quad (5)$$

and:

$$0 < \sum_{i=1}^{n} u_{ij} < n \quad (6)$$

Fuzzy C-means uses iterative optimization to approximate minima of an objective function. The fuzzy C-means function is defined as:

$$J_m(U, v) = \sum_{i=1}^{n} \sum_{j=1}^{c} \left(u_{ij}\right)^m \left(d_{ij}\right)^2 \quad (7)$$

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

where:

$$d_{ij}^2 = \|x_i - v_j\|^2 \qquad (8)$$

And ||·|| any inner product norm metric, and $m \in [1,\infty)$, where $m = 1$ is the non-fuzzy C-means algorithm. Wu (2012) in his parameter selection analysis suggested that $m \in [1.5,4]$. In this paper, m is equal to two (Zimmermann, 2001; Dembélé and Kastner, 2003; Bai, Dhavale and Sarkis, 2014). The fuzzy C-means algorithm is represented in *Algorithm 1* (Cannon *et al.*, 1986):

Algorithm 1. Fuzzy C-means algorithm.

1. Fix the number of clusters c, $2 \leqslant c \leqslant n$ where $n =$ number of data items. Fix m, $1 < m < \infty$. Choose any inner product induced norm metric ||.||,
2. Initialize the fuzzy c partition $U^{(0)}$,
3. At step b, $b = 1, 2, \ldots$,
4. Calculate the c cluster centers $\left\{v_j^{(b)}\right\}$ with $U^{(b)}$ and the formula for the jth cluster center:

$$v_{lj} = \frac{\sum_{i=1}^{n} \left(u_{ij}\right)^m x_{ij}}{\sum_{i=1}^{n} \left(u_{ij}\right)^m}, \quad l = 1, 2, \ \ldots, \ p$$

5. Update $U^{(b)}$: calculate the memberships in $U^{(b+1)}$ as follows. For $k = 1$ to n,
   a) Calculate I_k and $\tilde{I}_k$:
      $I_j = \{j | 1 \leqslant j \leqslant c, d_{ij} = x_i - v_j = 0\}$,
      $\tilde{I}_j = \{1, 2, \ \ldots, \ c\} - I_j$
   b) For data item j, compute new membership values:
      i) If $I_j = \varnothing$,

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}}$$

      ii) Else $u_{ij} = 0$ for all $j \in \tilde{I}_j$ and $\sum_{j \in I_j} u_{ik} = 1$;
6. Compare $U^{(b)}$ and $U^{(b+1)}$ in a convenient matrix norm; if $\|U^{(b)} - U^{(b+1)}\| < \varepsilon$, stop; otherwise, set $b = b + 1$, and go to step 4.

For different values of c, different clustering schemes are obtained. Therefore, a way is required to determine the best clustering. Some authors have proposed several indexes to evaluate the validity of fuzzy clustering (Bezdek, 1981; Wu and Yang, 2005; Zhang *et al.*, 2008; Arbelaitz *et al.*, 2013). In this paper, the partition Entropy index is used to find the most validated clustering (Bezdek, 1973, 1974):

$$V_{PE} = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij} \cdot \log\left(u_{ij}\right) \qquad (9)$$

A minimum value of V_{PE} determines the best clustering.

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

3. FCA-based DEA algorithm to decrease units heterogeneity impact

As stated in Section 1, the problem of heterogeneous DMUs can be raised in DEA problems. This heterogeneity might be in different inputs or outputs of DMUs, in their dissimilar activities, or in their size and scale. The considered problem of heterogeneity in this paper is as third type. Therefore, there are a set of n DMUs in different sizes and scales, using m inputs to produce s outputs.

The proposed algorithm to decrease the impact of heterogeneity on efficiency appraisal includes four steps as is elaborated in this section. *Algorithm 2* presents the FCA-based DEA algorithm to mitigate the impact of heterogeneity. These steps are detailed in the following subsections:

Algorithm 2. FCA-based DEA algorithm.

1. Define the problem: identify DMUs and define inputs and outputs. Then, data are gathered and input-output matrix $[X,Y]$ is constructed.
2. Use *Algorithm 1* for clustering of DMUs based on their input-output matrix and applying *Algorithm 1*.
3. Find the best fuzzy clustering by V_{PE} index.
4. Construct the DMUs membership matrix U (Equation (11)).
5. Compute the efficiency of each DMU in each cluster using Equation (12).
6. Construct the efficiency scores membership matrix, Equation (13).
7. Compute the unified efficiency scores using Equation (14).
8. Rank the DMUs based on $\overline{\theta}_j,\ j = 1, 2, \ \ldots, \ n$ values.

3.1 Initialization

The algorithm is initiated by identifying DMUs, and defining inputs and outputs. And afterwards data will be gathered on DMUs inputs and outputs. Accordingly, a $n \times (m + s)$ matrix is formed at the end of this step:

$$[X, Y] = \begin{matrix} DMU_1 \\ DMU_2 \\ \vdots \\ DMU_n \end{matrix} \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{m1} & y_{11} & y_{21} & \cdots & y_{s1} \\ x_{12} & x_{22} & \cdots & x_{m2} & y_{21} & y_{22} & \cdots & y_{s2} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} & y_{1n} & y_{2n} & \cdots & y_{sn} \end{bmatrix} \tag{10}$$

where $X_j = (x_{1j}, x_{2j}, \ldots, x_{mj})$ is the input vector received by $\mathrm{DMU_j}$ and $Y_j = (y_{1j}, y_{2j}, \ldots, y_{sj})$ is the output vector that is produced by it.

3.2 Fuzzy clustering

The main purpose of this step is to mitigate the impact of heterogeneity over DMUs on DEA results. Therefore, a fuzzy C-means algorithm, *Algorithm 1*, is performed to classify the DMUs. To achieve this aim, CA is applied for different values of c, $2 \leqslant c \leqslant n$. For each value of c, a partition Entropy index V_{PE} is computed and the clustering with minimum index is chosen. Finally, a $n \times c$ matrix of DMUs membership degree for different clusters is obtained:

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1c} \\ u_{21} & u_{22} & \cdots & u_{2c} \\ \vdots & \vdots & \cdots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nc} \end{bmatrix} \tag{11}$$

In which u_{ij} denotes the membership of ith DMU in jth cluster of the best chosen clustering, such that $\sum_{k=1}^{c} u_{jk} = 1, \forall j,\ j \in \{1, 2, \ldots, n\}$.

3.3 FCA-based DEA

The proposed algorithm for FCA-based DEA includes evaluating DMUs efficiency in different clusters to compose a fuzzy set of efficiencies. Consider the input-output matrix $[X,Y]$ in Equation (10). The membership degrees of DMUs in kth cluster are noted as $(u_{1k}, u_{2k}, \ldots, u_{nk})^T$ in matrix U of Equation (11). To find the efficiency of DMUs in kth cluster, the inputs and outputs of DMUs are justified. For this purpose, the inputs vector $x_j = (x_{1j}, x_{2j}, \ldots, x_{mj})$ and the outputs vector $y_j = (y_{1j}, y_{2j}, \ldots, y_{sj})$ of DMU_j, $j = 1, 2, \ldots, n$ are replaced with $x_j^k = u_{jk}x_j$ and $y_j^k = u_{jk}y_j$. Juxtaposing these justified vectors, the justified input-output matrix $[X^k, Y^k]$ is constituted. Any forms of radial models including CCR or BCC are applicable to evaluate efficiency of DMUs. Suppose that an input-oriented CCR model is used; consequently, the model used to evaluate the relative efficiency of DMU_0 is as follows:

$$\begin{aligned} & Min\ \theta_0^k \\ & \sum_{j=1}^{n} \lambda_j^k x_{ij}^k \leqslant \theta_0^k x_{i0}^k,\ i = 1, 2, \ldots, m \\ & \sum_{j=1}^{n} \lambda_j^k y_{rj}^k \geqslant y_{r0}^k,\ r = 1, 2, \ldots, s \\ & \lambda_j^k \geqslant 0,\ j = 1, 2, \ldots, n \\ & \theta_0^k \text{ unrestricted in sign} \end{aligned} \tag{12}$$

The above model is solved and the efficiency of DMUs in kth cluster is determined. Integrating the efficiency of θ_j^k, $j = 1, 2, \ldots, n$ with membership degrees of $(u_{j1}, u_{j2}, \ldots, u_{jk})^T$, a fuzzy vector of $((\theta_j^1, u_{j1}), (\theta_j^2, u_{j2}), \ldots, (\theta_j^k, u_{jk}))^T$ is composed, where (θ_j^l, u_{jl}) indicates that the efficiency of DMU_j regarding lth cluster is appraised to be θ_j^l. In fact, u_{jl} is interpreted as the membership of θ_j^l in cluster k of fuzzy efficiency set.

If the above process is repeated for different values of k, $k = 1,2, \ldots, c$, a fuzzy matrix of efficiencies can be created as follows:

$$\tilde{U} = \begin{bmatrix} (\theta_1^1, u_{11}) & (\theta_1^2, u_{12}) & \cdots & (\theta_1^c, u_{1c}) \\ (\theta_2^1, u_{21}) & (\theta_2^2, u_{22}) & \cdots & (\theta_2^c, u_{2c}) \\ \vdots & \vdots & \cdots & \vdots \\ (\theta_n^1, u_{n1}) & (\theta_n^2, u_{n2}) & \cdots & (\theta_n^c, u_{nc}) \end{bmatrix} \tag{13}$$

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

where the jth row of $\tilde{U}$ indicates the fuzzy vector of DMU_j's efficiencies in different clusters.

3.4 Finding aggregated efficiency

At third step, a fuzzy vector of efficiencies is obtained for each DMU, i.e. $((\theta_j^1, u_{j1}), (\theta_j^2, u_{j2}), \ldots, (\theta_j^c, u_{jk}))$. Using center of gravity (COG) index, a unified measure to appraise the efficiency of DMUs is obtained. The unified efficiency of DMU_j is defined as:

$$\overline{\theta}_j = \frac{\sum_{k=1}^{c} u_{jk}\theta_j^k}{\sum_{k=1}^{c} u_{jk}} \tag{14}$$

This measure is considered as the efficiency of DMU_j after eliminating the impacts of heterogeneity among DMUs.

4. A real world case study

To shed more light on what was delineated above in this section, a real world case study using the proposed algorithm is presented. This study is related to analyzing the efficiency of HNI bank, a private bank of Iran in the financial year 2012-2013. Banking is the most applicable area of DEA (Emrouznejad *et al.*, 2008), and it seems interesting to apply the proposed algorithm in this area. The considered problem deals with evaluating the efficiencies of 117 branches of HNI.

A set of seven inputs and six outputs are employed to evaluate the branches efficiency. These inputs and outputs are defined upon the guidelines of Berger and Humphrey (1997) and Luo *et al.* (2012). The inputs and outputs measures are illustrated in Table I.

The location of each branch is rated in a five-point scale by a committee in the bank. The new services deal with the income of the branch, which is obtained by providing new services. Furthermore, activity volume is evaluated based on the time spent to handle the documents and files. Table II presents the descriptive statistics of all 117 branches over financial year 2012-2013.

The next stage is to find the best clustering of branches. To achieve this end, the fuzzy C-means algorithm is run over the branches data, starting with $c=2$. Moreover, m is fixed at 2 and $\varepsilon=0.01$. Table III reveals the resulted V_{PE} index for different values of c. According to these values, the number of clusters is chosen as $c=2$.

The next stage is performing the membership of branches in different clusters to evaluate the efficiency of the banks. Table IV demonstrates the efficiency of units in

Table I.
Input and output measures

Inputs	Outputs
Personnel costs	Sum of deposits
Current and administrative costs	Loans
Current assets	Securities
Cost accounts	New services
Renting cost	Activity volume
Location	Branch income
The ratio of non-current to current receivables	

Variable	No. of branches	Maximum	Minimum	Mean	SD
Input variables					
Personnel costs	117	1,055.75	111.17	323.55	171.07
Current and administrative costs	117	1,321.02	54.07	259.97	221.30
Current assets	117	3,870.02	237.10	1,171.47	615.36
Cost accounts	117	8,337.65	15.14	462.54	1,103.50
Renting cost	117	264	5	38.41	56.79
Location	117	2.16	4.E−06	0.27	0.57
The ratio of non-current to current receivables	117	0.27	0	0.03	0.03
Output variables					
Sum of deposits	117	128,357.16	58.28	993.23	1,625.65
Loans	117	16,248.16	58.28	993.23	1,625.65
Securities	117	18,183.31	0	75.44	86.72
New services	117	539.85	0	75.44	86.72
Activity volume	117	2,563.30	319.71	1,211.05	515.39
Branch income	117	36,335	36	1,798.73	4,275.18

Table II. Descriptive statistics of input and output variables

Number of clusters	V_{PE}
2	0.006
3	0.061
4	0.064
5	0.186

Table III. Fuzzy clusters and partition entropy measure

each cluster. Additionally, the relative efficiency of classic DEA model is represented in this table. DMUs are evaluated using input-oriented BCC model.

4.1 Analysis of the added-value of the proposed method

The main result of this paper is to reduce the impacts of DMUs non-homogeneity on DEA results. In this section, the obtained results are analyzed. Initially, the results of the proposed method are compared with three non-homogeneity adjustment method as reviewed by Haas and Murphy (2003). In these approaches, the SST method referred to Sexton *et al.* (1994) method of adjusting outputs based on the ration of DMUs unadjusted efficiency score to their expected efficiency score. Two other methods, named magnitude of error (A-F) and ratio (A/F) methods are also used in regression analysis to adjust inputs and outputs. In this case, the location measure is used as the measure that is expected to account for non-homogeneity of units. Figure 1 presents the distribution of FCA-based scores with classic DEA and three considered adjustment-based methods.

Since the A/F method resulted in efficiency score of 1 for all the DMUs without any discrimination (the worst case in this example), it is eliminated from further consideration. According to *Algorithm 1*, the SST method coincides completely with the BCC model without any further improvement. In addition, A-F method resulted in efficiency scores between 0.87 and 1.00 in a narrow range. Actually, the distribution of A-F scores is highly dense. Table V illustrates correlation coefficient between different methods.

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

Table IV. Efficiency of DMUs in classic and fuzzy C-means based model

	Membership		Efficiency			Membership		Efficiency	
DMU	C1	C2	FCA-based	Classic	DMU	C1	C2	FCA-based	Classic
A1	0.9653	0.0347	1.0000	1.0000	A60	0.0001	0.9999	1.0000	1.0000
A2	0.9149	0.0851	1.0000	1.0000	A61	0.0003	0.9997	0.9999	1.0000
A3	0.0731	0.9269	1.0000	1.0000	A62	0.0001	0.9999	0.9987	1.0000
A4	0.0191	0.9809	0.9992	0.9211	A63	0.0003	0.9997	0.9999	1.0000
A5	0.0049	0.9951	1.0000	1.0000	A64	0.0003	0.9997	0.9992	1.0000
A6	0.0003	0.9997	1.0000	1.0000	A65	0.0001	0.9999	0.8894	0.8895
A7	0.0005	0.9995	0.5429	0.5497	A66	0.0001	0.9999	0.8973	1.0000
A8	0.0004	0.9996	0.9999	1.0000	A67	0.0004	0.9996	0.9998	1.0000
A9	0.0307	0.9693	1.0000	1.0000	A68	0.0003	0.9997	1.0000	1.0000
A10	0.0002	0.9998	0.5078	0.6245	A69	0.0004	0.9996	0.4962	0.4767
A11	0.0003	0.9997	0.6181	0.6621	A70	0.0092	0.9908	1.0000	1.0000
A12	0.0001	0.9999	0.7062	0.7609	A71	0.0001	0.9999	0.6900	0.6897
A13	0.0033	0.9967	1.0000	1.0000	A72	0.0013	0.9987	1.0000	1.0000
A14	0.0011	0.9989	0.6974	0.8927	A73	0.0005	0.9995	1.0000	1.0000
A15	0.0001	0.9999	1.0000	1.0000	A74	0.0006	0.9994	1.0000	1.0000
A16	0.0003	0.9997	0.6759	0.7692	A75	0.0004	0.9996	0.8601	0.8600
A17	0.0406	0.9594	1.0000	1.0000	A76	0.0003	0.9997	1.0000	1.0000
A18	0.0003	0.9997	1.0000	1.0000	A77	0.0003	0.9997	0.9082	0.9084
A19	0.0002	0.9998	0.5728	0.6103	A78	0.0003	0.9997	0.7007	0.6902
A20	0.0001	0.9999	1.0000	1.0000	A79	0.0003	0.9997	0.8402	0.8403
A21	0.0003	0.9997	1.0000	1.0000	A80	0.0002	0.9998	1.0000	1.0000
A22	0.0006	0.9994	1.0000	1.0000	A81	0.0006	0.9994	1.0000	1.0000
A23	0.0002	0.9998	1.0000	1.0000	A82	0.0004	0.9996	0.9994	1.0000
A24	0.0075	0.9925	1.0000	1.0000	A83	0.0005	0.9995	1.0000	1.0000
A25	0.0001	0.9999	1.0000	1.0000	A84	0.0013	0.9987	1.0000	1.0000
A26	0.0002	0.9998	0.9371	0.9372	A85	0.0001	0.9999	1.0000	1.0000
A27	0.0003	0.9997	1.0000	1.0000	A86	0.0159	0.9841	1.0000	1.0000
A28	0.0001	0.9999	0.5497	0.5687	A87	0.0021	0.9979	1.0000	1.0000
A29	0.0011	0.9989	1.0000	1.0000	A88	0.0003	0.9997	0.8709	0.8709
A30	0.0001	0.9999	0.7437	0.7010	A89	0.0004	0.9996	1.0000	1.0000

(*continued*)

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

	Membership		Efficiency			Membership		Efficiency	
DMU	C1	C2	FCA-based	Classic	DMU	C1	C2	FCA-based	Classic
A31	0.0001	0.9999	0.8896	0.8896	A90	0.0002	0.9998	0.2994	0.2994
A32	0.0006	0.9994	1.0000	1.0000	A91	0.0003	0.9997	0.99999	1.0000
A33	0.0003	0.9997	0.7985	0.7985	A92	0.0001	0.9999	0.8449	0.8450
A34	0.0003	0.9997	0.7988	0.7988	A93	0.0003	0.9997	1.0000	1.0000
A35	0.0005	0.9995	1.0000	1.0000	A94	0.0004	0.9996	1.0000	1.0000
A36	0.0003	0.9997	0.5681	0.5698	A95	0.0003	0.9997	1.0000	1.0000
A37	0.0003	0.9997	1.0000	1.0000	A96	0.0003	0.9997	1.0000	1.0000
A38	0.0010	0.9990	1.0000	0.9978	A97	0.0004	0.9996	0.9998	1.0000
A39	0.0002	0.9998	1.0000	1.0000	A98	0.0004	0.9996	1.0000	1.0000
A40	0.0001	0.9999	1.0000	1.0000	A99	0.0002	0.9998	1.0000	1.0000
A41	0.0003	0.9997	0.8029	0.8029	A100	0.0003	0.9997	1.0000	1.0000
A42	0.0002	0.9998	1.0000	1.0000	A101	0.0001	0.9999	0.5966	0.7458
A43	0.0003	0.9997	0.5594	0.5605	A102	0.0005	0.9995	1.0000	1.0000
A44	0.0004	0.9996	0.99997	1.0000	A103	0.0011	0.9989	1.0000	1.0000
A45	0.0003	0.9997	1.0000	1.0000	A104	0.0001	0.9999	1.0000	1.0000
A46	0.0007	0.9993	0.9999	1.0000	A105	0.0000	1.0000	1.0000	1.0000
A47	0.0002	0.9998	0.5722	0.5722	A106	0.0003	0.9997	0.9777	0.9777
A48	0.0001	0.9999	1.0000	1.0000	A107	0.0003	0.9997	0.7620	0.7639
A49	0.0003	0.9997	0.7022	0.8093	A108	0.0002	0.9998	1.0000	1.0000
A50	0.0004	0.9996	0.7424	0.8282	A109	0.0003	0.9997	0.6043	0.6043
A51	0.0004	0.9996	1.0000	1.0000	A110	0.0002	0.9998	0.7888	0.7940
A52	0.0003	0.9997	1.0000	1.0000	A111	0.0003	0.9997	0.8323	0.8323
A53	0.0002	0.9998	0.99998	1.0000	A112	0.0004	0.9996	0.8137	0.8137
A54	0.0003	0.9997	1.0000	1.0000	A113	0.0002	0.9998	0.7214	0.7216
A55	0.0010	0.9990	1.0000	1.0000	A114	0.0002	0.9998	0.7557	0.7557
A56	0.0001	0.9999	0.5873	0.5873	A115	0.0003	0.9997	1.0000	1.0000
A57	0.0003	0.9997	0.7447	0.7448	A116	0.0001	0.9999	1.0000	1.0000
A58	0.0003	0.9997	1.0000	1.0000	A117	0.0001	0.9999	1.0000	1.0000
A59	0.0002	0.9998	0.99997	1.0000					

Table IV.

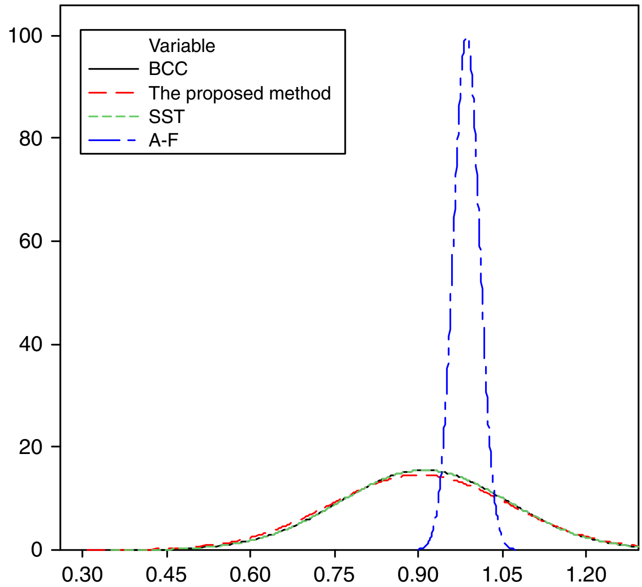

Figure 1.
Distribution of FCA-based scores and its comparison with classic DEA and SST, A-F, and A/F adjustment methods

Table V.
Correlation coefficients among different methods

	BCC	The proposed method	SST	A-F
BCC	–	0.98	1.00	0.511
The proposed method	0.98	–	0.98	0.499
SST	1.00	1.00	–	0.511
A-F	0.511	0.499	0.511	–

According to Table V, it is clear that the proposed method has the highest correlation with BCC and SST, while the correlation between A-F method with other methods is lower. On the other hand, the BCC and SST methods result in 75 DMUs with efficiency of 1.00, while the proposed method leads to 62 full efficient units. The A-F method produced 53 full efficient methods with the highest discrimination power, but in a narrow range.

A shortcoming of SST, A-F, and A/F methods is their dependency on the variables used as independent variables in regression analysis. In fact, some questions are remained for these methods: first, can it be proved that the considered methods are the main causes of non-homogeneity?, second, what is the impact of any change in the considered independent variables on the final results?, and third, what is the impact of regression error on the results? and, etc. Furthermore, the rational explanation for the adjustment methods is somewhat unclear; however, the proposed method does not require any external factors to remove heterogeneity using the observed inputs and outputs to find the pattern of DMUs in different clusters.

To capture a better understanding of fuzzy clustering impact on DEA results, it is worth noting here that the first quartile of classic DEA scores including 30 DMUs with the lowest efficiency are taken into consideration. Using a paired *t*-test, the 95 percent confidence interval of differences among efficiencies after and before clustering is

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

computed as (0.00780, 0.04865). De facto, by applying fuzzy clustering improved the efficiency of these 30 units between 0.78 and 4.86 percent. This finding remarks a significant increase in these branches efficiency after fuzzy clustering. The fuzzy clustering does not have a significant impact on the third quartile DMUs. Admittedly, the efficiency of these DMUs, including 30 DMUs with highest efficiencies, remained indifferent. DMUs between the first and the third quartile also change insignificantly. In these DMUs, including 57 DMUs, the impact of fuzzy clustering is between 0.41 percent decreasing to a 0.5 percent increasing (−0.41, 0.5).

5. Conclusion

Efficiency is an important managerial measure showing the ability of an organizational unit in using its resources to produce the intended outputs. DEA is an accepted and widely used approach to evaluate the relative efficiency of a group of similar units. This similarity lies in the resources that are used, the products which are produced, and the activities of the units. However, this similarity is always violated in reality. This non similarity arises from different conditions which DMUs performed within. Those, if the non-homogeneity of DMUs were not implied in efficiency appraisal, can cause some bias in evaluation, having a positive effect on some DMUs and negative effects on some others. The need for an approach to deal with DMUs heterogeneity is understood by some authors. In this paper, an algorithm based on fuzzy clustering concept is proposed to mitigate the impacts of heterogeneity over DEA results. While heterogeneity can make some high or low impacts on efficiency scores of DMUs, the proposed method can apply these impacts on results. The proposed algorithm consists of clustering DMUs by applying fuzzy C-means algorithm. Accordingly, DMUs will be classified in c clusters with different membership degrees. These membership degrees are then applied in appraising DMUs efficiencies in different groups. Eventually, aggregating different efficiency scores, an aggregated efficiency is computed for each DMU. Application of the proposed method is shown in a real world case study of evaluating relative efficiency of 117 bank branches. Meanwhile, the number of DMUs with unit efficiency decreased from 75 in classic and SST models to 62 in the proposed method, meaning an improvement in discrimination power of the model. A posteriori analysis of the obtained results indicated the improvement of DMUs in lower quartile, while the DMUs in upper quartile showed a slight decrease in their efficiency scores. By and large, regardless of the level of present heterogeneity between the units, the findings of numerical analysis proved impacts of the proposed algorithm on justifying the DEA results in problems with heterogeneous units.

References

Arbelaitz, O., Gurrtxaga, I., Muguerza, J., Perez, J.M. and Perona, I. (2013), "An extensive comparative study of cluster validity indices", *Pattern Recognition*, Vol. 46 No. 1, pp. 243-256.

Bai, C., Dhavale, D. and Sarkis, J. (2014), "Integrating fuzzy C-means and TOPSIS for performance evaluation: an application and comparative analysis", *Expert Systems with Applications*, Vol. 41 No. 9, pp. 4186-4196.

Berger, A.N. and Humphrey, D.B. (1997), "Efficiency of financial institutions: international survey and directions for future research", *European Journal of Operational Research*, Vol. 98 No. 2, pp. 175-212.

Bezdek, J. (1981), *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, NY.

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)

Bezdek, J.C. (1973), "Cluster validity with fuzzy sets", *Journal of Cybernetics*, Vol. 3 No. 3, pp. 58-73.

Bezdek, J.C. (1974), "Numerical taxonomy with fuzzy sets", *Journal of Mathematical Biology*, Vol. 1 No. 1, pp. 57-71.

Brown, R. (2006), "Mismanagement or mismeasurement? Pitfalls and protocols for DEA studies in the financial services sector", *European Journal of Operational Research*, Vol. 174 No. 2, pp. 110-1116.

Cannon, R.L., Dave, J.V. and Bezdek, J.C. (1986), "Efficient implementation of the fuzzy C-means clustering algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence, TPAMI*, Vol. 8 No. 2, pp. 248-255.

Charnes, A. and Cooper, W.W. (1962), "Programming with linear fractional functions", *Naval Research Logistics Quarterly*, Vol. 9 Nos 3/4, pp. 181-186.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978), "Measuring efficiency of decision making units", *European Journal of Operational Research*, Vol. 2 No. 6, pp. 429-444.

Charnes, A., Cooper, W.W., Lewin, A.Y. and Seiford, L.M. (1994), *Data Envelopment Analysis: Theory, Methodology and Applications*, Kluwer Academic Publishers, Norwell, MA.

Cooper, W.W., Seiford, L.M. and Tone, K. (2007), *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, Springer, New York, NY.

Dave, R.N. (1992), "Generalized fuzzy C-shells clustering and detection of circular and elliptic boundaries", *Pattern Recognition*, Vol. 25 No. 7, pp. 713-722.

Dembélé, D. and Kastner, P. (2003), "Fuzzy C-means method for clustering microarray data", *Bioinformatica*, Vol. 19 No. 8, pp. 973-980.

Dyson, R.G., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico, C.S. and Shale, E.A. (2001), "Pitfalls and protocols in DEA", *European Journal of Operational Research*, Vol. 132 No. 2, pp. 245-259.

Emrouznejad, A., Parker, B.R. and Tavares, G. (2008), "Evaluation of research in efficiency and productivity: a survey and analysis of the first 30 years of scholarly literature in DEA", *Socio-Economic Planning Sciences*, Vol. 42 No. 3, pp. 151-157.

Everitt, B.S., Landau, S., Leese, M. and Stahl, D. (2011), *Cluster Analysis*, John Wiley & Sons, West Sussex.

Farrell, M.J. (1957), "The measurement of productive efficiency", *Journal of the Royal Statistical Society*, Vol. 120 No. 3, pp. 253-290.

Farzipoor Saen, R., Memariani, A. and Hosseinzadeh Lotfi, F. (2005), "Determining relative efficiency of slightly non-homogeneous decision making units by data envelopment analysis: a case study in IROST", *Applied Mathematics and Computation*, Vol. 165 No. 2, pp. 313-328.

Haas, D.A. and Murphy, F.H. (2003), "Compensating for non-homogeneity in decision-making units in data envelopment analysis", *European Journal of Operational Research*, Vol. 144 No. 3, pp. 530-544.

Kumbhakar, S.C. and Knox Lovell, C.A. (2003), *Stochastic Frontier Analysis*, Cambridge University Press, Cambridge.

Liu, S.J., Lu, L.Y.Y., Lu, W.M. and Lin, B.J.Y. (2013), "Data envelopment analysis 1978-2010: a citation-based literature survey", *Omega*, Vol. 41 No. 1, pp. 3-15.

Luo, Y., Bi, G. and Liang, L. (2012), "Input/output indicator for DEA efficiency evaluation: an empirical study of Chinese commercial banks", *Expert Systems with Applications*, Vol. 39 No. 1, pp. 1118-1123.

Mirkin, B. (1996), *Mathematical Classification and Clustering*, Kluwer Academic Publishers, Dordrecht.

Mirkin, B. (2012), *Clustering: A Data Recovery Approach*, CRC Press, Boca Raton, FL.

Pham, D. (1999), "An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity in homogeneities", *Pattern Recognition Letters*, Vol. 20 No. 1, pp. 57-68.

Ray, S. (2004), *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*, Cambridge University Press, Cambridge.

Samoilenko, S. and Osei-Bryson, K.M. (2008), "Increasing the discriminatory power of DEA in the presence of the sample heterogeneity with cluster analysis and decision tree", *Expert Systems with Applications*, Vol. 34 No. 2, pp. 1568-1581.

Samoilenko, S. and Osei-Bryson, K.M. (2010), "Determining sources of relative inefficiency in heterogeneous samples: methodology using cluster analysis, DEA and neural networks", *European Journal of Operational Research*, Vol. 206 No. 2, pp. 479-487.

Sengupta, J.K. (2005), "Data envelopment analysis with heterogeneous data: an application", *Journal of the Operational Research Society*, Vol. 56, pp. 676-686.

Sexton, T.R., Sleeper, S. and Taggart, R.E. (1994), "Improving pupil transportation in North Carolina", *Interfaces*, Vol. 24 No. 1, pp. 87-103.

Sharma, M.J. and Jin, Y.S. (2011), "Self-organizing maps infusion with data envelopment analysis", in Mwasiagi, J.I. (Ed.), *Self Organizing Maps – Applications and Novel Algorithm Design*, Chapter 5, Intech, Rijeka, pp. 89-94.

Wu, K.L. (2012), "Analysis of parameter selection for fuzzy C-means", *Pattern Recognition*, Vol. 45 No. 1, pp. 407-415.

Wu, K.L. and Yang, M.S. (2005), "A cluster validity index for fuzzy clustering", *Pattern Recognition Letters*, Vol. 26 No. 9, pp. 1275-1291.

Zhang, Y., Wang, W., Zhang, X. and Li, Y. (2008), "A cluster validity index for fuzzy clustering", *Information Sciences*, Vol. 178 No. 4, pp. 1205-1218.

Zimmermann, H.J. (2001), *Fuzzy Set Theory – and its Applications*, Kluwer Academic Publishers, Boston, MA.

Further reading

Buckley, J.J. and Eslami, E. (2002), *An Introduction to Fuzzy Logic and Fuzzy Sets*, Phisica-Verlag, New York, NY.

Hatami-Marbini, A., Emrouznejad, A. and Tavana, M. (2011), "A taxonomy and review of the fuzzy data envelopment analysis literature: two decades in the making", *European Journal of Operational Research*, Vol. 214 No. 3, pp. 457-472.

Corresponding author

Seyed Hossein Razavi Hajiagha can be contacted at: s.hossein.r@gmail.com

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 22:04 14 November 2016 (PT)