Advances in Mathematical Finance
& Applications, 5 (1), (2020), 53-67
DOI: 10.22034/amfa.2019.1868389.1223

Published by IA University of
Arak, Iran
Homepage: www.amfa.iau-arak.ac.ir

Reduction of DEA-Performance Factors Using Rough Set Theory: An Application of Companies in the Iranian Stock Exchange

Maryam Joulaei, Mahnaz Mirbolouki*

Department of Mathematics, Yadegar-e-Imam Khomeini (RAH) Shahre Rey Branch ,Islamic Azad University, Tehran, Iran

ARTICLE INFO

Article history:
Received 11 May 2019
Accepted 24 July 2019

Keywords:
Portfolio
Data envelopment analysis
Rough set theory
Quick-reduct algorithm
Efficiency

ABSTRACT

The financial management field has witnessed significant developments in recent years to help decision makers, managers and investors, to made optimal decisions. In this regard, the institutions investment strategies and their evaluation methods continuously change with the rapid transfer of information and access to the financial data. When information is available as several inputs and output factors, the data envelopment analysis (DEA) applies to calculate the efficiency of companies. Distinguishing efficient companies from inefficient ones, makes it possible for the financial managers to select suitable portfolios. The discriminating power of DEA depends on the number of companies under evaluation and the number of inputs and outputs. When the number of inputs and outputs are high compared to the number of units, most of the units will be evaluated as efficient, thus the discriminating power of DEA decreases and the results are not reliable. To deal with this problem, the Quick-Reduct algorithm of the rough set theory (RST) was used in this study to reduce inputs or outputs. It should be noted that the advantage of this algorithm is its ability to use negative data.

1 Introduction

Portfolio management assists the investors to select optimal portfolio through analyzing the past and present status of companies and identifying the most efficient companies based on some criteria. Due to its importance, some models have been developed in the field of portfolio selection according to the conducted research, which was gradually replaced by new models after identifying their shortcomings. Several criteria such as the return on investment, earnings per share, price to profit per share ratio, and risk taking have been identified and applied in different models of portfolio evaluation in the conducted researches to identifying the factors and criteria affecting the stock portfolio selection. Hence, it is necessary to adopt the methods including different variables for decision making in portfolio selection. Beside, competition between traders eliminates asset pricing, so each stock is always properly priced and it has been tried to overcome random stock selection. So, it is difficult to select stocks with high expected returns for market failure. It is also difficult to assess the company's performance with different inputs and outputs; however, the DEA method combines multiple inputs and outputs of a company with the unit's overall performance criterion. As a result, DEA can be considered as a useful tool for selecting the shares of capital owners in the financial services industry. The

**** Corresponding author Tel.:*** *+989122736202*
E-mail address: *m.mirbolouki@srbiau.ac.ir*

estimated performance for each company (decision making unit) is affected by its inputs and outputs. Most of the units are evaluated as efficient and the dimensions of production space increase when the number of inputs and outputs are higher than the number of units; but instead, the discriminating power of DEA decreases. In this case, the results of the measurements are not reliable (See [1]). The input and output variables that are less important than others and have less impact on the performance, may be reduced by adopting appropriate scientific methods to deal with this problem. Several studies have been carried out so far on stock portfolio selection based on the DEA models. As the preliminary assumption of DEA and to correctly determine the efficiency in each of these studies, it is generally assumed that three times the sum of efficiency assessment factors is less than the assessed units. However, this assumption is ignored in some studies and therefore, the assessments are accompanied by drawbacks. In the papers [2,3], there is a concern about the impact of the number of input and output factors in DEA models. Cooper et al. [1] introduced an empirical rule that expresses the relationship between the number of units and the number of performance measures. Although removing any factor will affect relative performance, the selection set should inevitably be selected to have fewer effects on the performance of the units. There are several methods to discriminate units. According to [4,5], the most popular approaches are super-efficiency models [6-9]. These approaches do not attempt to reduce dimension to increase discrimination of units and using all the information, include additional models rank the units. Despite the applicability of these approaches on some issues, researchers in the opposite view emphasize on observing the thumb rule and maintaining reasonable dimensions in DEA models [1,10]. In the literature, there are approaches such as Delphi, Fuzzy Delphi and Shannon's entropy that have been applied to reduce the number of measurement factors in DEA [11-13]. The Delphi method requires knowledgeable and expert contributors to individually respond to questions and submit the results to a central coordinator. Nevertheless, not all problems of the expert survey can be avoided; Interviewing several people creates further restrictions. Shannon's Entropy is used on determining importance weights of influencing factors based on a decision matrix. Constructing such a decision matrix needs a deep statistical knowledge in data, although it may be constructed subjectively like what is done in [13]. Principle component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables. The combination of PCA and DEA models is suggested when there is an excessive number of inputs and outputs in relation to the number of decision-making units [14-15]. In this approach none of inputs and outputs are removed, but they are with a weight will be combined together. It must be noted that PCA-DEA models did not developed when inputs or outputs contain negative values. Since several factors should be examined when investing in the stock exchange, especially in allocating their funds in stock portfolio, it is important to obtain the vital performance measures of each industry in order to calculate its efficiency. By examining the financial statements of the companies listed in the stock exchange of Iran in this study, fifteen factors affecting the efficiency (nine inputs and six outputs) were considered corresponding to each industry. Since the number of companies in the Iranian stock exchange is not proportional to the number of assessment factors in many industries, so a method is needed to select the necessary factors in assessing the efficiency of each industry. The Rough Sets Theory (RST) and Quick-Reduct (QR) algorithm were used in this regard. RST seeks to identify the pattern of data. based on RST and discernibility criteria, the dependency degree of each input(output) to a data set can be identified. Therefore, independent factors can be removed with minimal damage to data [16]. Then, it is possible to determine the efficiency of companies affiliated with various industries and consider in stock portfolio evaluation using the negative DEA method.

The rest of the paper is structured as follows: Some previous studies on the portfolio evaluation and selection based on DEA models are reviewed in the second section. Section 3 provides the basic definitions of DEA, negative DEA, and RST. Section 4 has dedicated to how to evaluate the efficiency of Iranian companies and finally, the conclusion is presented in the fifth section.

2 DEA Background on Stock Portfolio Evaluation

There has been a lot of discussion about how to select the best stock portfolio. The main stock portfolio model was introduced by Markowitz [17] in 1952. The Markowitz Mean-Variance model is the most common and popular approach in investment selection problem. The Markowitz Mathematical Modeling Program is also the most effective tool for selecting the optimal portfolio. Markowitz derived the expected returns and risk for the asset portfolios for the first time and showed that the rate of return deviation is a proper measure for the stock portfolio risk under a set of reasonable assumptions and explained a method for calculating the risk of stock portfolios. Following to the Markowitz method, Lee and Lerro [18] selected 10 industries including 61 companies as their research statistical population during 1959-1968. They considered Beta, returns, Ci, Di and the risk of each share criteria in 61selected companies, for creating the optimal portfolio. Finally, they presented an ideal programming model for building the optimal portfolio by prioritizing the criteria of their research population. Murthi et al. [19] introduced an index in portfolio performance based on DEA method. They noted that this index addresses the shortcomings of the former indices that were Jensen's alpha and the Sharpe index. They showed the benefits of the proposed approach and assessed the performance of mutual funds and resulted that the mutual funds are all approximately mean-variance efficient. Bowlin [2] investigated the financial performance of the business units of the US Department of Defense using DEA. Financial performance of commercial sectors related to defense affairs was assessed in comparison with non-defense sectors from 1983 to 1992 using DEA in this research. He compared the DEA and financial ratios analysis, and concluded that the above methods are complementary. DEA was also used by [20] to identify the efficient and inefficient mutual fund companies listed in the Morning Star 500 inventory. They also identified financial variables that differ significantly between efficient and inefficient investment companies and identify the nature of these relationships. 80 out of 84 mutual fund companies were analyzed and companies were divided into three categories: efficient, with lowest efficacy and inefficient companies. Among these companies, 27, 22 and 31 companies were efficient, with lowest efficacy and inefficient, respectively. Edirisinghe and Zhang [21] developed a generalized DEA model to analyse a firm's financial statements over time in order to determine a relative financial strength indicator that is predictive of firm's stock price returns. Their model is based on maximizing the correlation between the DEA-based score of financial strength and the stock exchange performance. In a research conducted on the Taiwan Stock Exchange, Hung Chen [22] has used DEA models to select a portfolio of the most efficient companies in the eight Taiwan Stock Exchange industries. The historical data was prepared in this research from the second quarter of 2004 to the third quarter of 2007 as quarterly periods and consisted separate portfolios by CCR and BCC models. The portfolios obtained from the two models were compared with the portfolio resulted from selecting based on the size of the company and also the average market returns. The researcher has used Sharp's ratio to compare the returns adjusted with portfolio risk. The portfolio made by the size of the company yields lowest returns among the rest, but the portfolios made by DEA models have been accepted, and finally, it could yield higher returns than the average industry return and make an optimal stock portfolio by DEA method and considering several input and output, simultaneously. Nowadays, there are many studies about the applications of DEA on portfolio performance evaluation and portfolio selection. Few of preliminaries are listed above. We can mention [23-31] to refer to the recent studies.

3 Preliminaries

3.1 Data Envelopment Analysis

DEA is a non-parametric approach based on mathematical programming for assessment the relative performances of many different kinds of entities, commonly referred to as decision-making units (DMUs), where the attendance of multiple inputs and outputs makes collation difficult. DEA was initially introduced by [32] (CCR model) and extended by [33] (BCC model). Assume that there are n DMUs $(\mathrm{DMU}_j; j \in \{1, \dots, n\})$ each using m semi-positive inputs $x_j = (x_{1j}, \dots, x_{mj})$ to produce s semi-positive outputs $y_j = (y_{1j}, \dots, y_{sj})$. Consider The following BCC model to assess $\mathrm{DMU}_o; o \in \{1, \dots, n\}$:

$$
\begin{array}{ll}
\theta_o = \min \theta & \\
s.t. & \\
\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{io} & i = 1, \dots, m \\
\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro} & r = 1, \dots, s \\
\sum_{j=1}^{n} \lambda_j = 1 & \\
\lambda_j \geq 0 & j = 1, \dots, n
\end{array} \tag{1}
$$

where $0 < \theta_o \leq 1$ is the efficiency score of DMU_o. If $\theta_o = 1$ then DMU_o is efficient, otherwise, it is inefficient. With the aim of generalizing of the existing distance functions and introducing a non-oriented approach, Chambers et al. [34] suggested following generic directional distance (DD) model, which assess DMU_o under VRS technology:

$$
\begin{array}{ll}
e_o = \max \theta & \\
s.t. & \\
\sum_{j=1}^{n} \lambda_j x_{ij} \leq x_{io} - \theta g_{x_i} & i = 1, \dots, m \\
\sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro} + \theta g_{y_r} & r = 1, \dots, s \\
\sum_{j=1}^{n} \lambda_j = 1 & \\
\lambda_j \geq 0 & j = 1, \dots, n
\end{array} \tag{2}
$$

where the nonzero directional vector $(\boldsymbol{g}_x, \boldsymbol{g}_y) = (g_{x_1}, \dots, g_{x_m}, g_{y_1}, \dots, g_{y_s})$ enables the model to contract inputs and expand outputs simultaneously. Using model (2), the inefficient unit depicts on the efficiency frontier. In fact, e_o is the inefficiency score of the under evaluation unit . Therefore, DMU_o is efficient if in model (2) $e_o = 0$. Otherwise it is inefficient and the corresponding number is $1 - e_o$. As mentioned above, in the classical models of DEA, it is assumed that the input and output vectors are non-negative, while in applications such as the assessment of the performance of companies in the stock exchange, some inputs and outputs may have negative values. There are several studies on negative data in DEA [35-40]. Allahyar and Rostami-Malkhlife [41] were pioneered. They have proposed a non-oriented model which permits inputs and outputs to capture both positive and negative values and yields an efficiency measure. To assess DMU_o, they considered the direction vector $(\boldsymbol{g}_x, \boldsymbol{g}_y) = (|x_{1o}|, \dots, |x_{mo}|, |y_{1o}|, \dots, |y_{so}|)$ by using the absolute values of data and formulated the following model:

$$\begin{aligned} & \theta_o^* = \max\theta \\ & s.t. \\ & \sum_{j=1}^{n}\lambda_j x_{ij} \le x_{io} - \theta|x_{io}| \qquad i = 1, \dots, m \\ & \sum_{j=1}^{n}\lambda_j y_{rj} \ge y_{ro} + \theta|y_{ro}| \qquad r = 1, \dots, s \\ & \sum_{j=1}^{n}\lambda_j = 1 \\ & \lambda_j \ge 0 \qquad j = 1, \dots, n \end{aligned} \tag{3}$$

Model (3) yields the efficiency rating θ^*corresponding to each DMU it is evident that $\theta^* \ge 0$. To have an efficiency score between 0 and 1 for inefficient units, and may consider $E_o = 1 - \frac{\theta_o^*}{\bar{\theta}_o}$ where $\bar{\theta}_o$ is defined as:

$$\bar{\theta}_o = \min\left\{\min_i\left\{\frac{x_{io} - x_{iI}}{|x_{io}|}, x_{io} \ne 0\right\}, \min_r\left\{\frac{y_{rK} - y_{ro}}{|y_{ro}|}, y_{ro} \ne 0\right\}, i = 1, \dots, m, r = 1, \dots, s\right\}$$

where $I = \min_j \{x_{ij}, i = 1, \dots, m\}$ and $\mathrm{K} = \max_j \{y_{rj}, r = 1, \dots, s\}$.

If $\bar{\theta}_o = 0$ then $\theta_o^* = 0$, otherwise $0 < E_o \le 1$. DMU_o is an efficient unit if $E_o = 1$.

3.1 Rough Set Theory and Feature Selection

Rough set theory (RST) was first developed by Pawlak [42] as an extension of conventional set theory to deal with imprecise, incomplete and inconsistent information and knowledge. The RST has two steps. First step is forming the concepts and rules by classification of the relational database. The second step is discovering the knowledge through the classification of the equivalence relation and classification for the approximation of the target. In the field of data analysis and processing theory, the RST emerged after probability theory, fuzzy set theory, and evidence theory to deal with uncertain information. As a novel thinking method, RST has become an important information processing tool in the field of intelligent information processing due to its easy operation. It has also used in many other research fields including machine learning, knowledge discovery, data mining and decision support and analysis. We refer the readers to [43,44] to learn more about RST. In the RST, classification of the domain of interest into disjoint categories are a pair of supported approximations, called as lower and upper approximations. Description of the domain objects which are known with certainty to belong to the subset of interest was considered as the lower approximation while the upper approximation is a description of the objects which possibly belong to the subset. [45] introduced two main approaches to find rough set reducts.

Let $IS = (\mathbb{U}, \mathbb{A})$be an information system, where $\mathbb{U}$ is a non-empty set of finite objects (the universe) and $\mathbb{A}$ is a non-empty finite set of attributes such that $f_a: \mathbb{U} \to V_a$ for every $a \in \mathbb{A}$. V_a is the set of values that attribute a may take. With any $P \subseteq \mathbb{A}$ there is an associated equivalence relation $IND(P)$:

$$IND(P) = \{(x_i, x_j) \in \mathbb{U}^2 | \forall\, a \in A، a(x_i) = a(x_j)\} \tag{4}$$

The partition of $\mathbb{U}$, generated by $IND(P)$ is denoted $U/IND(P)$ (or U/P) and can be calculated as follows:

$$U/IND(P) = \otimes\, \{U/Ind\{a\} | a \in P\} \tag{5}$$

in which

$$A \otimes B = \{X \cap Y | X \in A, Y \in B, X \cap Y \neq \emptyset\}.$$

If $(x, y) \in IND(P)$, then x and y are indiscernible by attributes from P. The equivalence classes of the P -indiscernibility relation are denoted $[x]_P$. Let $X \subseteq \mathbb{U}$. X can be approximated using only the information contained within P by constructing the P -lower and P -upper approximations of X:

$$\underline{P}X = \{x_i \in \mathbb{U} | [x_i]_P \subseteq X\} \quad (6)$$

$$\overline{P}X = \{x_i \in \mathbb{U} | [x_i]_P \cap X \neq \emptyset\} \quad (7)$$

Let P and Q be equivalence relations over $\mathbb{U}$, then the positive region can be defined as:

$$POS_P(Q) = \cup_{X \in \mathbb{U}/Q} \underline{P}X \quad (8)$$

The positive region contains all objects of $\mathbb{U}$ that can be classified to classes of U/Q using the information in attributes P. Using this definition of the positive region, the rough set degree of dependency of a set of attributes Q on a set of attributes P is defined in the following way:

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|\mathbb{U}|} \quad (9)$$

For $P, Q \subseteq A$ it is said that Q depends on P in a degree of k $(0 \leq k \leq 1)$.

3.1.1 RST based reduct generation algorithms

Categorization of objects in the universe needs a minimal subset of features. Reduct is a minimal and sufficient subset applied in the Quick-Reduct algorithm (QR algorithm) introduced in [46] as a RST based reduct generation algorithm. The reduct computation process starts with an empty reduct set and the attributes recursively add to it, one by one to produce a maximum possible value. QR-forward, QR-backward and Improved-QR generation are three variations of QR algorithm based on the criteria of adding features. The QR-forward algorithm, starts by the first feature and successive selection of next feature and checking for any improvement in the metric i.e., degree of dependency. The QR-Backward algorithm starts from the last feature in the feature set and successively adds the features from the last one and similarly observes for improvement and iteratively repeats the process and terminates when the degree of dependency of the reduct set is equals to the degree of dependency of the full conditional attribute set. The drawback of these algorithms is that whenever the stopping criteria meets, the algorithm terminates without examining all features. The Improved QR algorithm (see Figure 1), examines the remaining attributes and from this it selects the one with maximum improvement in the metric. The following example clearly explains the variants of RST based reduct generation algorithms.

Example 1. Suppose that there are 12 DMUs each of them consuming four inputs to produce five outputs. Table 1 shows the related data. If we calculate efficiency score of these units applying DEA model (3), it would be revealed that all units are efficient that means a weakness of DEA. To have a suitable assessment framework by DEA, the number of DMUs must be greater than three times of the sum of number of inputs and outputs [1]. Here, using variants of reduct generation algorithms, the number of inputs and outputs are reduced. Let conditional attributes (inputs) be $C_I = \{I_1, I_2, I_3, I_4\}$ and (outputs) $C_O = \{O_1, O_2, O_3, O_4, O_5\}$. In this example, we use k-mean algorithm to divide the dataset into k partitions. The goal of k-mean algorithm is clustering data such that the sum of squared deviations from the cluster's centroids is minimal. The decision attributes that are identified by k-mean algorithm are D_I for inputs and D_O for outputs.

$R \leftarrow \{\,\}$
do
$T \leftarrow R$
$\forall x \in (C - R)$
    if $\gamma_{R\cup\{x\}}(D) > \gamma_T(D)$
  $T \leftarrow R \cup \{x\}$
$R \leftarrow T$
until $\gamma_R(D) = \gamma_C(D)$
Return R

Fig 1: The Improved QR algorithm

Table 1: Data of example.

DMU	I_1	I_2	I_3	I_4	O_1	O_2	O_3	O_4	O_5	D_I	D_O
U1	1	3	9	13	4	8	-2	13	0	d_3	d'_4
U2	2	3	9	14	1	6	10	-3	-1	d_3	d'_2
U3	2	5	9	10	3	6	10	-3	-1	d_3	d'_2
U4	-1	5	-2	10	3	5	12	11	15	d_1	d'_3
U5	2	5	-2	10	2	5	9	13	0	d_1	d'_4
U6	-1	3	11	14	1	6	10	13	-1	d_3	d'_4
U7	-1	7	-2	10	2	5	-2	13	0	d_1	d'_4
U8	-1	3	9	10	3	7	-2	-3	0	d_3	d'_1
U9	4	8	9	0	3	7	-2	-3	0	d_2	d'_1
U10	4	5	-2	14	1	6	10	-3	-1	d_1	d'_2
U11	-1	8	12	10	3	7	-2	14	16	d_4	d'_3
U12	-1	3	12	0	1	6	-2	13	-1	d_2	d'_4

The equivalence classes generated for the input conditional attributes are given below:

$$\mathbb{U}/I_1 = \{\{U1\},\{U2,U3,U5\},\{U4,U6,U7,U8,U11,U12\},\{U9,U10\}\}$$
$$\mathbb{U}/I_2 = \{\{U1,U2,U6,U8,U12\},\{U3,U4,U5,U10\},\{U7\},\{U9,U11\}\}$$
$$\mathbb{U}/I_3 = \{\{U1,U2,U3,U8,U9\},\{U4,U5,U7,10\},\{U11,U12\},\{U6\}\}$$
$$\mathbb{U}/I_4 = \{\{U1\},\{U2,U6,U10\},\{U3,U4,U5,U7,U8,U11\},\{U9,U12\}\}$$
$$\mathbb{U}/C_I = (\mathbb{U}/I_1)\otimes(\mathbb{U}/I_2)\otimes(\mathbb{U}/I_3)\otimes(\mathbb{U}/I_4)$$
$$= \{\{U1\},\{U2\},\{U3\},\{U4\},\{U5\},\{U6\},\{U7\},\{U8\},\{U9\},\{U10\},\{U11\},\{U12\}\}$$
$$\mathbb{U}/D_I = \{\{U1,U2,U3,U6,U8\},\{U4,U5,U7,U10\},\{U11\},\{U9,U12\}\}$$

The degree of dependency of the attributes can be calculated using equation (9) as follows:

$$\gamma_{C_I}(D_I) = \frac{|\{\{U1\},\{U2\},\{U3\},\{U4\},\{U5\},\{U6\},\{U7\},\{U8\},\{U9\},\{U10\},\{U11\},\{U12\}\}|}{|\mathbb{U}|} = 1$$

$$\gamma_{I_1}(D_I) = \frac{|\{U1\}|}{|\mathbb{U}|} = 0.083$$
$$\gamma_{I_2}(D_I) = \frac{|U7|}{|\mathbb{U}|} = 0.083$$
$$\gamma_{I_3}(D_I) = \frac{|\{U4, U5, U7, U10\} \cup \{U6\}|}{|\mathbb{U}|} = 0.41$$
$$\gamma_{I_4}(D_I) = \frac{|\{U1\} \cup \{U9, U12\}|}{|\mathbb{U}|} = 0.25$$

<u>QR-forward algorithm</u>

Initially, $Reduct = \emptyset$ and $\gamma_{Reduct}(D_I) = 0$.

Select the first feature i.e., I_1 and add it to the reduct set. $Reduct = \{I_1\}$ $\gamma_{Reduct}(D_I) = 0.083$.

Now, add the second attribute I_2 to the reduct and continue the process until the degree of dependency of the reduct equals to the dependency of all conditional attributes. $Reduct = \{I_1, I_2\}$ و $\gamma_{Reduct}(D_I) = \frac{7}{12} = 0.58$. Then, $Reduct = \{I_1, I_2, I_3\}$ و $\gamma_{Reduct}(D_I) = 1$.

The degrees of dependency of the attributes $\{I_1, I_2, I_3\}$ are equals to the degrees of dependency of the full set of attributes. Then algorithm is terminated and the reduct is $\{I_1, I_2, I_3\}$.

<u>QR-backward algorithm</u>

Initially, $Reduct = \emptyset$ and $\gamma_{Reduct}(D_I) = 0$.

Select the last feature i.e., I_4 and add it to the reduct. $Reduct = \{I_4\}$ و $\gamma_{Reduct}(D_I) = 0.25$.

There is an improvement in the dependency after adding the attribute I_4. Now, add the second attribute I_3 to the reduct and continue the process until the degree of dependency of the Reduct equals to the dependency of all conditional attributes. $Reduct = \{I_3, I_4\}$ و $\gamma_{Reduct}(D_I) = 1$.

The degrees of dependency of the attributes $\{I_3, I_4\}$ are equals to the degrees of dependency of the full set of attributes. Then algorithm is terminated and the reduct is $\{I_3, I_4\}$.

<u>Improved-QR algorithm</u>

Initially, $Reduct = \emptyset$ and $\gamma_{Reduct}(D_I) = 0$

The attributes with the maximum dependency degree is I_3. So, first selection is I_3 and $\gamma_{Reduct}(D_I) = 0.41$. Next attribute is I_4. Then, $Reduct = \{I_3, I_4\}$ و $\gamma_{Reduct}(D_I) = 1$.

The degrees of dependency of the attributes $\{I_3, I_4\}$ are equals to the degrees of dependency of the full set of attributes. Then algorithm is terminated and the reduct is $\{I_3, I_4\}$.

Similarly, we identify these algorithms for the output attributes $C_O = \{O_1, O_2, O_3, O_4, O_5\}$ relative to decision attributes D_o and determining the reducts.

- QR-forward algorithm results $Reduct = \{O_1, O_2, O_3, O_4\}$.
- QR-backward algorithm results $Reduct = \{O_4, O_5\}$.
- Improved-QR algorithm results $Reduct = \{O_4, O_5\}$

Here the reducts resulted from QR-backward and Improved-QR algorithms for both input and output sets are equal. Table 2 contains the DEA-efficiency score calculated by model (3) in the presence of negative data and related to each reduct. Results show the discrimination of efficiency scores when the number of selected factors are relatively less than DMUs; $n \geq 3(m+s)$.

Table 2: Results of efficiency scores calculated by model (3).

DMU	QR-forward reduct	QR-backward reduct	Improved-QR reduct
U1	1	0.2512	0.2512
U2	1	0.3689	0.3689
U3	1	0.45	0.45
U4	1	1	1
U5	1	1	1
U6	1	0.1238	0.1238
U7	1	1	1
U8	1	0.45	0.45
U9	0.1894	1	1
U10	1	1	1
U11	1	1	1
U12	1	1	1

4 Evaluating the Active Companies Listed in the Iranian Stock Exchange

The stock exchange or the stock market is a place where stocks of various manufacturing, service and investment companies are traded. The securities refer to the shares of the companies, and to the place where exchanges are conducted, the exchange hall. Providing two groups of people with the legal right to engage in an economic activity and benefit from this activity are the main task of the stock exchange. These two groups are called investor and investee; investors have money and capital or savings but cannot work with it, and investees group have the ability to carry out economic activities but do not have enough capital and the exchange makes it possible that these two groups legally trade. The efficiency assessment of the companies and industries in the Iranian stock exchange are presented in this section, considering efficiency in three consecutive periods. It should be noted that each industry uses Improved QR algorithm to reduce the evaluation factors and then efficiencies of the companies in the related industry can be evaluated applying the model (3). According to the audited balance sheet and the profit (loss) statement information of the companies active in the Iranian stock exchange[1], 15 criteria were calculated including input parameters (liquidity ratio, leverage ratio, activity ratio) and output parameters (profitability ratio, valuation ratio) according to the following formulas. It must be noted that we just considered companies and industries that their financial statement audit for the year 2014 have been reported.

<u>Financial Ratios Formulas:</u>

1. Debt to Equity Ratio: Total non-current debt divided by shareholder equity.
2. Quick Ratio: (Non-current assets, inventories of goods and materials) divided by current liabilities.
3. Current Ratio: total current assets divided by current liabilities.
4. Solvency Ratio-I: total liabilities divided by total assets.
5. Solvency Ratio-II: total liabilities divided by shareholder equity.
6. Leverage Ratio: total assets divided by total shareholder equity.
7. Asset Turnover: Income from sales of services and goods divided total assets.
8. Inventory Turnover: income from goods sold divided by inventory and goods.

[1] Securities and Exchange Organization web site: www.codal.ir

9. Receivables Turnover: Income from sales of services and goods divided into trade accounts receivable.
10. Return on Assets: Net Profit (loss) divided by total assets.
11. Earnings Per Share (EPS): (Net profit (loss) - preferred dividends) divided by (capital/1000).
12. Net Profit Margin: Net profit (loss) divided by income from sales of services and goods.
13. Return on Equity: Net profit (loss) divided by equity.
14. Price to Earnings (PE) Ratio: (present value divided by (Capital * 1000)) divided by EPS.
15. Price to Book Ratio: (present value divided by (capital*1000)) divided by ((equity) * 1,000,000) divided by (capital*1000)).

Table 3: Inputs and outputs of Iranian stock exchange.

Inputs	Liquidity Ratios	Debt to Equity Ratio	I1
		Quick Ratio	I2
		Current Ratio	I3
	Leverage Ratios	Solvency Ratio-I	I4
		Solvency Ratio-II	I5
		Leverage Ratio	I6
	Asset Utilization	Asset Turnover	I7
		Inventory Turnover	I8
		Receivables Turnover	I9
Outputs	Profitability Ratio	Return on Assets	O1
		Earnings Per Share (EPS)	O2
		Net Profit Margin	O3
		Return on Equity	O4
	Valuation Ratio	Price to Earnings (PE) Ratio	O5
		Price to Book Ratio	O6

Table 4: Active industries in the Iranian stock exchange and selected reducts

Industry	Number of companies	Input-reduct	Output-reduct
Extraction of metal ores	10	{I1,I2,I4}	{O1,O3}
Massive real estate	12	{I1,I4}	{O1,O2}
Automobile and parts	32	{I1,I2,I3,I7,I8}	{O1,O2,O3,O5}
Medicine	35	{I4,I6,I8}	{O1,O2,O4}
Electrical appliances	12	{I1,I2}	{O1,O2,O4}
Computer	12	{I1,I2,I3}	{O1,O2}
Gypsum-lime-cement	32	{I2,I4,I8}	{O1,O2,O3,O4,O5}
Chemical	52	{I3,I4,I7,I8,I9}	{O1,O2,O3,O5}
Food except sugar	30	{I1,I3,I7}	{O2,O4,O6}
Basic metals	30	{I2,I4,I7,I9}	{O1,O4,O6}
Sugar	16	{I2,I4}	{O1,O3}
Ceramic Tile	11	{I3,I4}	{O1,O2,O3}
Non-metallic mineral	21	{I2,I3,I4,I7}	{O1,O2,O5,O6}
Rubber and plastic	11	{I2,I3}	{O1,O2,O3}

Table 3 shows input and output classifications of the above 15 criteria. In this study we considered 14 active industries and a number of companies in each industry as stated in Table 4. According to the Table 4, there are totally 316 active companies. All companies are homogeneous because they have the same inputs and outputs. We may want to compare all the companies together and evaluate the efficiency scores. Because three times the total number of inputs and outputs is 45 and this is less than the total number of companies, the resulting efficiency score can be estimated reliably. However, the nature of companies is not necessarily the same, implying an efficient company in an industry cannot be considered as a benchmark of inefficient unit in another industry. Therefore, companies in each industry must be evaluated separately and ultimately their efficiency should be considered for portfolio selection. Except for chemical industry, all other industries contain less than 45 companies that means the number of inputs and outputs are higher than the appropriate proportion with the number of companies. So, we applied Improved QR algorithm introduced in the previous section to reduce the number of performance factors as much as possible to have a reliable efficiency score.

Table 5: Efficiency score of companies in Extraction of metal ores industry

Company	Efficiency score	
	R1: {I1,I2,I4,O1,O3}	R2: {I2,I4,O1,O3}
Bama	1.00	1.00
Supply of Sabian Steel	1.00	1.00
Milad Steel and Iron	0.65	0.59
Chadormalu	1.00	1.00
Gol Gohar	0.67	0.65
Bafgh mines	1.00	1.00
Zinc mines of Iran	1.00	1.00
Taknar Copper mine	1.00	1.00
Manganese mines of Iran	0.59	0.54
Damavand mining	1.00	1.00

Table 4 contains selected reducts for both inputs and outputs that are calculated by Improved-QR algorithm. The number of inputs and outputs are reduced as much as possible, although the thumb rule which is an empirical rule is not valid for serval industries. In these situations, the analyst can select the most dependent criteria to include in the reduct. It should be noted that in case of using the direct DEA method without reducing the evaluation factors, most companies are evaluated as efficient, which indicates that the calculations are not valid. Companies with a good investment risk can be identified based on the efficiencies and the methods of literatures mentioned in the section 2.

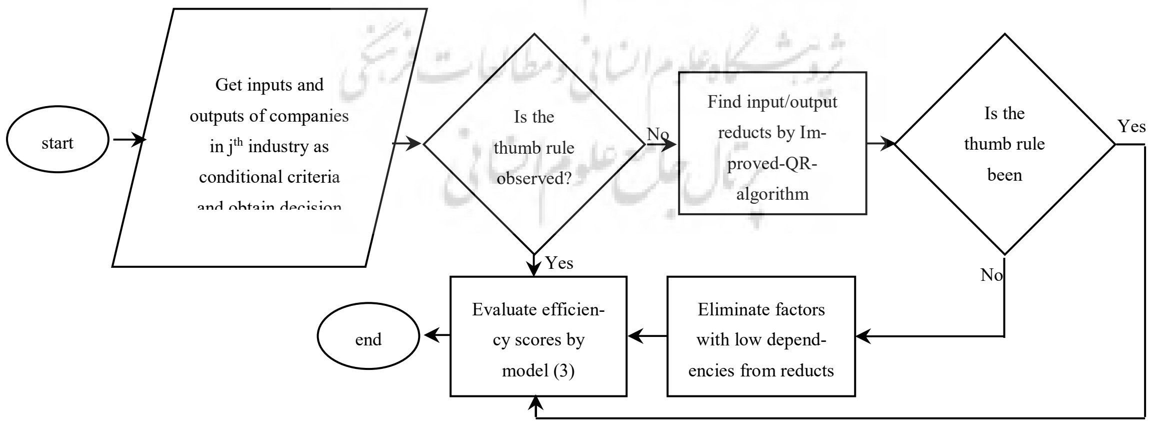

Fig 2: The steps of companies' efficiency evaluation.

Due to the large number of companies, we just report efficiency scores of companies in Extraction of metal ores industry. The efficiency scores in the Table 5 are evaluated based on two types of reducts.

The first reduct (R_1) is related to Improved-QR algorithm and the second reduct (R_2) eliminate I_1 from R_1 because of low independency of I_1 to input-reduct and regarding to the thumb rule. According to Table 5, efficient companies in both reducts are the same. To have a glimpse of the presented method used to calculate the efficiency score of companies, Figure 2 depicted the steps of the jth industry as a flowchart.

5 Conclusions

In the case of existing relatively large number of factors in evaluating the company's efficiency by DEA model, the calculated efficiency is not often valid. Accordingly, various methods have been introduced to eliminate unnecessary factors. In this paper, applying the QR-algorithm of RST, we presented a method to reduce low dependent factors. The introduced method was applied for evaluating efficiency scores of companies and industries listed in Iranian stock exchange. Since there are negative data in this application, former reduction methods such as DEA-PCA methods are not applicable more. We address a negative DEA model in the literatures to resolve this problem. This model along with introduced QR-algorithm can be helpful to estimate the efficiency scores of companies in the Iranian stock exchange. Observing the status of companies during several periods of time and introducing a dynamic method for eliminating low dependent criteria is suggested for next studies.

References

[1] Cooper, W.W., Seiford, L.M., and Tone, K., *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, 2nd ed., Springer, Berlin, 2007. Doi: 10.1007/978-0-387-45283-8.

[2] Bowlin, W.F., *Measuring Performance: An Introduction to Data Envelopment Analysis(DEA)*, The Journal of Cost Analysis, 1998, **15**(2), P. 3-27. Doi: 10.1080/08823871.1998.10462318.

[3] Raab, R., and Lichty, R., *Identifying sub-areas that comprise a greater metropolitan area: the criterion of country relative efficiency*. Journal of Regional Science, 2002, **42**(3),579–594. Doi:10.1016/j.ijpe .2011.11.010.

[4] Angulo-Meza, L., and Lins, M., *Review of Methods for Increasing Discrimination in Data Envelopment Analysis*, Annals of Operations Research, 2002, **116**(1-4), P. 225–242. Doi:10.1023/A:1021340 616758.

[5] Podinovski, V., and Thanassoulis, E., *Improving discrimination in data envelopment analysis: some practical suggestions*, Journal of Productivity Analysis, 2007, **28**(1-2), P.117–126. Doi: 10.1007/s11123-007-0042-x.

[6] Andersen, P., and Pitersen, N.C., *A Procedure for Ranking Efficient Units in Data Envelopment Analysis,* Management Science, 1993, **39**(10), P. 1261-1264. Doi: 10.1287/mnsc.39.10.1261.

[7] Sexton, T. R., *The Methodology of Data Envelupment Analysis*, New Directions for Program Evaluation, 1986, **32**, P. 7-29. Doi: 10.1002/ev.1438.

[8] Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., Shale, E. A., *Pitfalls and protocols in DEA*, European Journal of Operational Research, 2001, **132**(2), P. 245-259. Doi; 10.1016/S0377-2217(00)00149-1.

[9] Doyle, J., and Green, R., *Efficiency and Cross-efficiency in DEA: Derivations, Meanings and Uses*, Journal of the Operational Research Society, 1994, **45**(5), P. 567–578. Doi:10.1057/jors.1994.84.

[10] Green, R., Doyle, J., Cook, W.D., *Preference voting and project ranking using DEA and cross-evaluation*, European Journal of Operational Research, 1996, **90**, P. 461–472. Doi: 10.1016/0377-2217(95)00039-9.

[11] Arsad, R., Nasir Abdullah, M., Alias, S., and Isa, Z., *Selection Input Output by Restriction Using DEA Models Based on a Fuzzy Delphi Approach and Expert Information*, Journal of Physics: Conference Series, 2017, **892**, 012010. Doi: 10.1088/1742-6596/892/1/012010.

[12]

[13] Sadidi, M., Khalilifar, O., Amiri, M., and Moradi, R., *Use of Partial Least Squares Structural Equation Modeling for Identifying the Most Important Variables via Application of Data Envelopment Analysis*, J Arch Mil Med, 2018, **6**(1) P. 1-9. Doi: 10.5812/jamm.67114.

[14] Xie, Q., Dai, Q., Li, Y., and Jiang, A., *Increasing the Discriminatory Power of DEA Using Shannon's Entropy*, Entropy, 2014, **16**(3), P. 1571–1585. Doi:10.3390/e16031571.

[15] Adler, N., Golany, B., *Including principal component weights to improve discrimination in data envelopment analysis*. Journal of the Operational Research Society, 2002, **53**(9), P. 985-991. Doi: 10.1057/palgrave.jors.2601400.

[16] Azadeh, A., Ghaderi, S.F., and Ebrahimipour, V., *An integrated PCA DEA framework for assessment and ranking of manufacturing systems based on equipment performance*, Engineering Computations (Swansea, Wales), 2007, **24**(4), P. 347-372. Doi: 10.1108/02644400710748689.

[17] Aliakbarpoor, Z., Izadikhah, M., *Evaluation and ranking DMUs in the presence of both undesirable and ordinal factors in data envelopment analysis*, International Journal of Automation and Computing, 2012, **9** (6), 609-615. DOI: 10.1007/s11633-012-0686-5

[18] Nguyen, T.T., and Nguyen, P.K., *Reducing Attributes in Rough Set Theory with the Viewpoint of Mining Frequent Patterns*, International Journal of Advanced Computer Science and Applications, 2013, **4**(4), P. 130-138. Doi: 10.14569/IJACSA.2013.040421.

[19] Markowitz, H.M., *Portfolio selection*, The Journal of Finance, 1952, **7**(1), P. 77–91. Doi: 10.1111/j.1540-6261.1952.tb01525.x.

[20] Lee, S.M., and Lerro, A.J., *Optimaizing the portfolio selection for mutual funds*, The Journal of Finance, 1973, **28**(5), P. 1087-1101. Doi: 10.1007/978-3-642-59132-7_46.

[21] Murthi, B.P.S, Choi, Y.K., and Desai, P., *Efficiency of Mutual Funds and Portfolio Performance Measurement: A Non - Parametric Approach,* European Journal of Operational Research, 1997, **98** (2), P. 408-418. Doi: 10.1016/S0377-2217(96)00356-6.

[22] Haslem, J. M., and Scheraga, C.A., *Data Envelopment Analysis of Morningstar's Large-cap Mutual Funds*, The Journal of Investing, 2003, **12**(4): 41-48. Doi: 10.3905/joi.2003.319566.

[23] Edirisinghe, N.C.P., and Zhang, X., *Generalized DEA model of fundamental analysis and its application to portfolio optimization*, Journal of Banking & Finance, 2007, **31**(11): 3311-3335. Doi: 10.1016/j. jbankfin.2007.04.008.

[24] Hung Chen, H., *Stock selection using data envelopment analysis*. Industrial Management & Data Systems, 2008,**108**(9), P. 1255-1268. Doi: 10.1108/02635570810914928.

[25] Lim, S., Oh, K.W., and Zhu, J., *Use of DEA cross-efficiency evaluation in portfolio selection: An application to Korean stock market,* European Journal of Operational Research, 2014, **236**(1), P. 361-368. Doi: 10.1016/j.ejor.2013.12.002.

[26] Liu, W., Zhoua, Z., Liu, D., and Xiao, H., *Estimation of portfolio efficiency via DEA*, Omega, 2015, **52**, P. 107-118. Doi; 10.1016/j.omega.2014.11.006.

[27] Choi, H.S., and Min, D., *Efficiency of well-diversified portfolios: Evidence from data envelopment analysis*, Omega, 2017, **73**, P. 104-113. Doi: 10.1016/j.omega.2016.12.008.

[28] Esfandiar, M., Saremi M., and Jahangiri Nia, H., *Assessment of the efficiency of banks accepted in Tehran Stock Exchange using the data envelopment analysis technique,* Advances in Mathematical Finance and Applications,2018, **3**(2), P. 1-11. Doi: 10.22034/AMFA.2018.540815.

[29] Izadikhah, M., *Improving the Banks Shareholder Long Term Values by Using Data Envelopment Analysis Model.* Advances in Mathematical Finance and Applications, 2018, **3**(2), P. 27-41. Doi: 10.22034/AMFA.2 018.540829.

[30] Tarnaud, A.C., and Leleu, H., *Portfolio analysis with DEA: Prior to choosing a model*, Omega, 2018, **75**, P. 57-76. Doi: 10.1016/j.omega.2017.02.003.

[31] Zhang, Y.J., and Chen, M.Y., *Evaluating the dynamic performance of energy portfolios: Empirical evidence from the DEA directional distance function*, European Journal of Operational Research, 2018, **269**(1), P. 64-78. Doi: 10.1016/j.ejor.2017.08.008.

[32] Peykani, P., Mohammadi, E., Rostamy-Malkhalifeh, M., and Hosseinzadeh Lotfi, F., *Fuzzy Data Envelopment Analysis Approach for Ranking of Stocks with an Application to Tehran Stock Exchange,* Advances in Mathematical Finance and Applications,2019, **4**(1), P. 31-43. Doi: 10.22034/AMFA.20 19.581412.1155.

[33] Peykani, P., Mohammadi, E., Pishvaee, M.S., Rostamy-Malkhalifeh, M. and Jabbarzadeh, A., *A novel fuzzy data envelopment analysis based on robust possibilistic programming: possibility, necessity and credibility-based approaches*, RAIRO-Operations Research, 2018, **52**(4), P.1445-1463. Doi: 10.1051/ro/2018019.

[34] Charnes, A., Cooper, W.W., and Rhodes, E., *Measuring the efficiency of decision making units,* European Journal of Operational Research, 1978, **2**(6), P. 429–444. Doi: 10.1016/0377-2217(78)90138-8.

[35] Banker, R.D., Charnes, A., and Cooper, W.W., *Some models for estimating technical and scale efficiencies in date envelopment analysis*, Management Science, 1984, **30**, P. 1078-1092. Doi: 10.1287/ mnsc.30.9.1078.

[36] Chambers, R.G., Chung, Y., and Färe, R., Profit, *Directional Distance Functions, and Nerlovian Efficiency*, Journal of Optimization Theory and Applications, 1998, **98**(2), P. 351–364. Doi: 10.1023/A:10 22637501082.

[37] Khoveyni, M., Eslami, R., Yang, G.l., *Negative data in DEA: Recognizing congestion and specifying the least and the most congested decision making units*, Computers & Operations Research, 2017, **79**, P. 39-48. Doi: 10.1016/j.cor.2016.09.002.

[38] Tavana, M., Izadikhah, M., Di Caprio, D., Saen, R. F., *A new dynamic range directional measure for two-stage data envelopment analysis models with negative data*, Computers & Industrial Engineering, 2018, **115**, P. 427-448. Doi: 10.1016/j.cie.2017.11.024.

[39] Lin, R., Yang, W., Huang, H., *A modified slacks-based super-efficiency measure in the presence of negative data*, Computers & Industrial Engineering, 2019, **135**, P. 39-52. Doi: 10.1016/j.cie.201 9.05.030.

[40] Izadikhah, M., Saen, R.F., *Evaluating sustainability of supply chains by two-stage range directional measure in the presence of negative data*, Transportation Research Part D: Transport and Environment, 2016, **49**, P.110-126. Doi: 10.1016/j.trd.2016.09.003.

[41] Mehdiloozad, M., Zhu, J., Sahoo, B. K., *Identification of congestion in data envelopment analysis under the occurrence of multiple projections: A reliable method capable of dealing with negative data*, European Journal of Operational Research, 2018, **265**(2), P. 644-654. Doi: 10.1016/j.ejor.2017.07.065.

[42] Izadikhah, M., Saen, R.F., Roostaee, R., *How to assess sustainability of suppliers in the presence of volume discount and negative data in data envelopment analysis?*, Annals of Operations Research, 2018, **269**(1-2),P. 241-267. Doi:10.1007/s10479-018-2790-6.

[43] Allahyar, M., and Rostamy-malkhalifeh, M., *Negative data in data envelopment analysis: Efficiency analysis and estimating returns to scale*. Computers & Industrial Engineering, 2014, **82**, P.78-81. Doi: 10.1016/j.cie.2015.01.022.

[44] Pawlak, Z., *Rough Sets*, International Journal of Computer and Information Science, 1982, **11**(5), P. 341-356. Doi: 10.1007/BF01001956.

[45] Pawlak, Z., *Rough sets and intelligent data analysis*, Information Sciences, 2002, **147**(1), P. 1-12. Doi: 10.1016/S0020-0255(02)00197-4.

[46] Zhang, Q., Xie, Q., and Wang G., *A Survey on Rough Set Theory and Its Applications*, CAAI Transactions on Intelligence Technology, 2016, **1**(4), P. 323-333. Doi: 10.1016/j.trit.2016.11.001.

[47] Jensen, R., Shen, Q., and Tuson, A., *Finding Rough Set Reducts with SAT. In: Ślęzak D., Wang G., Szczuka M., Düntsch I., Yao Y. (eds) Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*. RSFDGrC 2005. Lecture Notes in Computer Science, vol 3641, Springer, Berlin, Heidelberg, 2005. Doi: 10.1007/11 548669_21.

[48] Peters, J.F., and Skowron, A., *Transactions on Rough Sets XVII,* Springer-Verlag Berlin Heidelberg, 2014. Doi: 10.1007/978-3-642-54756-0.