Input/Output Variables Selection in Data Envelopment Analysis: A Shannon Entropy Approach

Pejman Peykani 1, Fatemeh Sadat Seyed Esmaeili 2, Mirpouya Mirmozaffari 3,*, Armin Jabbarzadeh 4 and Mohammad Khamechian 5

1 School of Industrial Engineering, Iran University of Science and Technology, Tehran 1684613114, Iran; pejman.peykani@yahoo.com
2 Department of Mathematics, Faculty of Science, Science and Research Branch, Islamic Azad University, Tehran 1477893855, Iran; f.seyedesmaeeli@srbiau.ac.ir
3 Department of Industrial Engineering, Dalhousie University, 5269 Morris Street, Halifax, NS B3H 4R2, Canada
4 Department of Systems Engineering, École de Technologie Supérieure (ETS), University of Quebec, 1100 Notre-Dame Street West, Montreal, QC H3C 1K5, Canada; armin.jabbarzadeh@etsmtl.ca
5 Department of Industrial and Manufacturing Engineering, University of Wisconsin-Milwaukee, Milwaukee, WI 53211, USA; khamech@uwm.edu

* Correspondence: mr828394@dal.ca

Abstract: The purpose of this study is to provide an efficient method for the selection of input–output indicators in the data envelopment analysis (DEA) approach, in order to improve the discriminatory power of the DEA method in the evaluation process and performance analysis of homogeneous decision-making units (DMUs) in the presence of negative values and data. For this purpose, the Shannon entropy technique is used as one of the most important methods for determining the weight of indicators. Moreover, due to the presence of negative data in some indicators, the range directional measure (RDM) model is used as the basic model of the research. Finally, to demonstrate the applicability of the proposed approach, the food and beverage industry has been selected from the Tehran stock exchange (TSE) as a case study, and data related to 15 stocks have been extracted from this industry. The numerical and experimental results indicate the efficacy of the hybrid data envelopment analysis–Shannon entropy (DEASE) approach to evaluate stocks under negative data. Furthermore, the discriminatory power of the proposed DEASE approach is greater than that of a classical DEA model.

Keywords: data envelopment analysis; Shannon entropy; input/output selection; discriminatory power; stock market; negative data

1. Introduction

Data envelopment analysis (DEA) is a non-parametric approach based on mathematical programming and multi-criteria decision making (MCDM) that is capable of evaluating the performance, ranking, classification, and benchmarking of a set of homogeneous decision-making units (DMUs), according to the desired inputs and outputs [1–6]. The DEA approach is one of the most powerful, applicable, and effective methods in the field of performance evaluation among researchers and is widely used in various fields such as agriculture, airline, airport, bank, gas, hospital, hotel, information technology, insurance, manufacturing, mutual fund, power, production system, research and development, school, sport, stock exchange, supply chain, university, water, etc. [7–27].

The main advantages of the DEA approach are as follows: it needs no knowledge about the production function and its constraints; able to use multiple inputs and multiple outputs simultaneously; needs no knowledge about the weight of each input and output indicator; able to use various inputs and outputs with different measurement scales; compares inefficient DMUs with reference sets directly; ranks decision-making units; and benchmarks
for inefficient DMUs. The main disadvantages of the DEA approach can be summarized as follows: it measures relative efficiency instead of absolute efficiency; finds it difficult to solve large problems due to high computational value; experiences many deviations in results due to measurement error; potential change in performance evaluation results due to change in type and number of inputs and outputs; difficulty of statistical tests including hypothesis test due to its non-parametric nature; fragility of performance obtained due to the sensitivity of the results to sample change [28–35].

In addition to the above advantages and disadvantages, the lack of general agreement on the selection and determination of input and output variables is one of the most important challenges in applying the DEA method in various applications [36–41]. Another important point that should be considered in using the DEA approach is to increase the discriminatory power of the model in evaluating the performance of DMUs, and to differentiate as much as possible their performance results. Accordingly, in the current study, the hybrid data envelopment analysis–Shannon entropy (DEASE) approach is proposed. Notably, the proposed DEASE can be employed under negative data and values.

The rest of this paper is organized as follows. The concepts, definitions, and explanations of the Shannon entropy technique are introduced in Section 2. Then, the steps of the hybrid data envelopment analysis–Shannon entropy approach as a proposed approach of the current research is presented in Section 3. In the following, the proposed DEASE approach is implemented in a real-world case study from the Tehran stock exchange (TSE), and the experimental results are analyzed in Section 4. Finally, the conclusions and future research directions are discussed in Section 5.

2. Shannon Entropy Technique

Determination of the relative weights of indicators in multi-criteria decision making (MCDM) is always one of the basic and required steps in the problem-solving process. It should be noted that among the well-known and widely used methods in determining the weights of indicators, expert opinions-based approaches, least squares method, special vector technique, and Shannon entropy can be mentioned. In the following, the Shannon entropy technique is introduced as one of the most important methods for determining the weight of criteria.

Entropy in information theory is a measure of the amount of uncertainty, and it is expressed by a discrete probability distribution [42–47]. Notably, in entropy method, more fluctuations and scattering in the values of criterion indicate its greater importance factor and weight [48–53]. Accordingly, the steps of the Shannon entropy technique to determine the weights of the indices are as follows:

Step (1) First, the decision matrix is created with \(m\) alternatives and \(n\) criteria in the form of Equation (1), which \(x_{ij}\) is the value of \(i^{th}\) alternative in terms of \(j^{th}\) criterion.

\[
X = \begin{bmatrix}
x_{11} & x_{12} & \cdots & x_{1n} \\
x_{21} & x_{22} & \cdots & x_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
x_{m1} & x_{m2} & \cdots & x_{mn}
\end{bmatrix}
\]

(1)

Step (2) The decision matrix is normalized using Equation (2). By dividing the value of each column by the sum of its column, the normalized value \(p_{ij}\) is obtained as follows.

\[
p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{m} x_{ij}} \quad \forall i,j
\]

(2)

Step (3) The entropy of each criterion \(E_j\) is calculated using Equation (3). A constant value keeps the value of \(E_j\) between 0 and 1.
\[ E_j = -\omega \sum_{i=1}^{m} [(P_{ij})\ln(P_{ij})] \quad \forall j, \quad k = \frac{1}{\ln(m)} \]  

**Step (4)** The degree of deviation \( d_j \) from the information that is generated for \( j \)th criterion is calculated from Equation (4). The degree of deviation indicates the amount of useful information that the relevant criterion provides to the decision maker.

\[ d_j = 1 - E_j \quad \forall j \]  

**Step (5)** Finally, the weight \( w_j \) is calculated from Equation (5), in which the weight of \( j \)th criterion is obtained by dividing \( d_j \) by the sum of \( d_j \).

\[ w_j = \frac{d_j}{\sum_{j=1}^{n} d_j} \quad \forall j \]  

Thus, the criterion with more weight \( w_j \) is chosen, because less weight indicates that the effect of the criterion is almost the same for all the alternatives.

3. The Proposed Approach

In this section, the process of proposing and implementing the hybrid data envelopment analysis–Shannon entropy approach is presented to the input/output selection to improve the discriminatory power of the model for performance measurement of the DMUs. In order to be more comprehensive and applicable to the proposed approach, the basic steps of this approach, by assuming the presence of negative values and data, are presented as follows:

**Step (1) Modifying Indicators with Negative Values**: As observed in the steps of the Shannon entropy technique, the existence of a function \( \ln \) for computation of \( p_{ij} \) means the entropy method is used only for positive indicators and quantities. Therefore, to solve this problem, a suggested method is used in this research. In this way, first, in the criterion column whose values are negative for some alternatives, the largest and smallest numbers are determined, and then their difference from each other \((\text{Max} - \text{Min})\) is calculated. Then, the value of \( \text{Max} - \text{Min} + 1 \) and 1 are assigned to the largest column number and the smallest column number, respectively. The values of the other column numbers are obtained using the relation \( \text{Value} - \text{Min} + 1 \). Thus, using the suggested method, all the values that are related to the criterion are presented in a positive value.

**Step (2) Implementing the Shannon Entropy Technique**: After modifying and reviewing to change the amplitude of negative values in the previous step and preparing a new data structure, the Shannon entropy approach for each of the input and output groups that have similar indicators is implemented and calculated.

**Step (3) Selecting of Inputs and Outputs**: According to the values of the Shannon entropy technique, in each of the input and output groups, among the similar indicators in each group, the criterion with more weight as the final input or output of the DEA model is selected.

**Step (4) Checking the Isotonicity Relations between Inputs and Outputs**: Since the inputs and outputs used in DEA should satisfy the condition that greater quantities of inputs provide increased output, the appropriateness of the inputs and outputs that were included in the previous step was tested by conducting an isotonicity test [54–59]. An isotonicity test involves the calculation of all inter-correlations between inputs and outputs for identifying whether increasing amounts of inputs lead to greater outputs [60–66].

**Step (5) Choosing the Data Envelopment Analysis Model**: After determining the inputs and outputs, the data envelopment analysis model should be selected. In this study, due to the presence of negative data, the range directional measure (RDM) model [67] is used. Now, suppose that there are \( g \) homogeneous decision-making units that con-
vert \upsilon inputs \theta_k = (\theta_{1k}, \theta_{2k}, \ldots, \theta_{\upsilon k}) into \upsilon outputs \varphi_k = (\varphi_{1k}, \varphi_{2k}, \ldots, \varphi_{\upsilon k}). Finally, the envelopment form of the RDM model is Model (6).

\[
\begin{align*}
\text{Max } & \Psi \\
\text{S.t. } & \sum_{k=1}^{g} \lambda_k \theta_{ak} \leq \theta_{aq} - \Psi \xi_{aq}, \quad \forall a \\
& \sum_{k=1}^{g} \lambda_k \varphi_{bk} \geq \varphi_{bq} + \Psi \xi_{bq}, \quad \forall \beta \\
& \sum_{k=1}^{g} \lambda_k = 1 \\
& \lambda_k \geq 0
\end{align*}
\]

\(\Psi^*\) expresses the measurement of inefficiency, and the efficiency of the DMU under evaluation is equal to \(1 - \Psi^*\) and \(1/(1 + \Psi^*)\) in the input-oriented model and output-oriented model, respectively. Moreover, \(\xi_{aq}\) and \(\xi_{bq}\) in Model (6) are the range of possible improvements for the DMU under evaluation, which are defined as Equations (7) and (8), respectively.

\[
\xi_{aq} = \theta_{aq} - \min_k \{\theta_{ak}; k = 1, \ldots, g\}, \quad \forall a
\]

(7)

\[
\xi_{bq} = \max_k \{\varphi_{bk}; k = 1, \ldots, g\} - \varphi_{bq}, \quad \forall \beta
\]

(8)

**Step (6) Calculating the Efficiency Scores of DMUs:** Finally, the research data envelopment analysis model is implemented for the extracted data and related to the selected input and output indicators, and the performance assessment results of the decision-making units are calculated and analyzed.

4. **Case Study**

In order to implement the proposed DEASE approach, 15 stocks from the food and beverage industry of the Tehran stock exchange are selected as a case study for the research. It should be noted that in order to evaluate the stocks fundamentally and comprehensively, 15 indicators in the form of five groups (including liquidity, asset utilization, leverage, profitability, and growth) have been considered [68–73]. By applying the Microsoft Excel Software, the Shannon entropy technique is implemented, and the obtained results are presented in Table 1. A description of all the financial parameters is introduced in Table 2.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Perspective</th>
<th>Financial Parameters</th>
<th>(E_j)</th>
<th>(d_j)</th>
<th>(w_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Liquidity</td>
<td>Current ratio (CUR)</td>
<td>0.980</td>
<td>0.020</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Quick ratio (QUI)</td>
<td>0.952</td>
<td>0.048</td>
<td>0.218</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cash ratio (CAR)</td>
<td>0.846</td>
<td>0.154</td>
<td>0.694</td>
</tr>
<tr>
<td>2</td>
<td>Asset utilization</td>
<td>Inventory turnover (INT)</td>
<td>0.983</td>
<td>0.017</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Receivable turnover ratio (RTR)</td>
<td>0.895</td>
<td>0.105</td>
<td>0.766</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total assets turnover (TAT)</td>
<td>0.985</td>
<td>0.015</td>
<td>0.108</td>
</tr>
<tr>
<td>3</td>
<td>Leverage</td>
<td>Solvency ratio-I (SRI)</td>
<td>0.981</td>
<td>0.019</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Solvency ratio-II (SRII)</td>
<td>0.788</td>
<td>0.212</td>
<td>0.635</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Leverage ratio (LER)</td>
<td>0.897</td>
<td>0.103</td>
<td>0.307</td>
</tr>
<tr>
<td>4</td>
<td>Profitability</td>
<td>Net profit to sales (NPS)</td>
<td>0.937</td>
<td>0.063</td>
<td>0.273</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Return on assets (ROA)</td>
<td>0.927</td>
<td>0.073</td>
<td>0.314</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Return on equity (ROE)</td>
<td>0.904</td>
<td>0.096</td>
<td>0.413</td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>Groups</th>
<th>Perspective</th>
<th>Financial Parameters</th>
<th>(E_j)</th>
<th>(d_j)</th>
<th>(w_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Growth</td>
<td>Earnings per share growth rate (EPSGR)</td>
<td>0.871</td>
<td>0.129</td>
<td>0.336</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total revenue growth rate (TRGR)</td>
<td>0.930</td>
<td>0.070</td>
<td>0.183</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Profit margin growth rate (PMGR)</td>
<td>0.815</td>
<td>0.185</td>
<td>0.482</td>
</tr>
</tbody>
</table>

Table 2. The description of all financial indicators in the current study.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUR</td>
<td>Total current assets divided by total current liabilities</td>
</tr>
<tr>
<td>QUR</td>
<td>Subtract inventory from total current assets divided by total current liabilities</td>
</tr>
<tr>
<td>CAR</td>
<td>Cash and marketable securities divided by total current liabilities</td>
</tr>
<tr>
<td>INT</td>
<td>Revenues for the period divided by inventories</td>
</tr>
<tr>
<td>RTR</td>
<td>Net receivable sales divided by average net receivables</td>
</tr>
<tr>
<td>TAT</td>
<td>Revenues for the period divided by total assets</td>
</tr>
<tr>
<td>SRI</td>
<td>Total liability divided by total assets</td>
</tr>
<tr>
<td>SRII</td>
<td>Total liability divided by shareholders equity</td>
</tr>
<tr>
<td>LER</td>
<td>Total assets divided by shareholders equity</td>
</tr>
<tr>
<td>NPS</td>
<td>Net profit after tax divided by sales</td>
</tr>
<tr>
<td>ROA</td>
<td>Net income divided by the total assets</td>
</tr>
<tr>
<td>ROE</td>
<td>Net income generated per unit of common shareholders’ equity</td>
</tr>
<tr>
<td>EPSGR</td>
<td>Current quarter’s EPS divided by the previous quarter’s EPS minus one</td>
</tr>
<tr>
<td>TRGR</td>
<td>Current quarter’s total revenue divided by the previous quarter’s total revenue minus one</td>
</tr>
<tr>
<td>PMGR</td>
<td>Current quarter’s profit margin divided by the previous quarter’s profit margin minus one</td>
</tr>
</tbody>
</table>

Then, in each of the five groups related to different financial ratios, the criterion with more weight is selected from the three criteria. Concerning the Shannon entropy technique, the inputs and outputs of the data envelopment analysis model are presented in Figure 1. In the DEA approach, performance metrics can be classified as the larger the better for the outputs, and the smaller the better for the inputs. In other words, positively connoted (the more the better) factors are used as outputs; conversely, negatively connoted (the fewer the better) factors are classified as inputs [74,75].

![Figure 1. Selected inputs and outputs of RDM model for stock evaluation.](image-url)
Now, after determining the selected variables of the DEA model, the information related to the mentioned indicators in the form of two inputs and three outputs for 15 stocks that are active in the food and beverage industry is presented in Table 3. Pearson’s correlation is taken to test the isotonicity relationship between the chosen input and output parameters. Accordingly, the inter-correlations of all the indicators are positive and significant, suggesting that the specification of the DEA model is valid.

Table 3. Data set for 15 stocks from Tehran stock exchange.

<table>
<thead>
<tr>
<th>Stocks</th>
<th>Inputs</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I (1)</td>
<td>O (1)</td>
</tr>
<tr>
<td>Stock 01</td>
<td>4.98</td>
<td>54.47</td>
</tr>
<tr>
<td>Stock 02</td>
<td>44.74</td>
<td>44.82</td>
</tr>
<tr>
<td>Stock 03</td>
<td>8.87</td>
<td>22.93</td>
</tr>
<tr>
<td>Stock 04</td>
<td>8.35</td>
<td>48.34</td>
</tr>
<tr>
<td>Stock 05</td>
<td>11.89</td>
<td>55.86</td>
</tr>
<tr>
<td>Stock 06</td>
<td>97.36</td>
<td>205.27</td>
</tr>
<tr>
<td>Stock 07</td>
<td>42.98</td>
<td>12.23</td>
</tr>
<tr>
<td>Stock 08</td>
<td>68.52</td>
<td>33.72</td>
</tr>
<tr>
<td>Stock 09</td>
<td>60.87</td>
<td>46.72</td>
</tr>
<tr>
<td>Stock 10</td>
<td>40.46</td>
<td>8.64</td>
</tr>
<tr>
<td>Stock 11</td>
<td>19.77</td>
<td>71.89</td>
</tr>
<tr>
<td>Stock 12</td>
<td>71.00</td>
<td>57.86</td>
</tr>
<tr>
<td>Stock 13</td>
<td>117.57</td>
<td>37.01</td>
</tr>
<tr>
<td>Stock 14</td>
<td>54.46</td>
<td>32.55</td>
</tr>
<tr>
<td>Stock 15</td>
<td>16.87</td>
<td>53.31</td>
</tr>
<tr>
<td>Min</td>
<td>4.98</td>
<td>0.29</td>
</tr>
<tr>
<td>Max</td>
<td>117.57</td>
<td>12.25</td>
</tr>
</tbody>
</table>

Finally, using the range directional measure model and LINGO Software, the performance of all 15 stocks is calculated based on the data that are extracted from the Tehran stock market. The results can be seen in Table 4.

Table 4. The final results of the implementation of range directional measure model.

<table>
<thead>
<tr>
<th>Stocks</th>
<th>Inefficiency</th>
<th>Efficiency</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \Psi' )</td>
<td>( 1 - \Psi' )</td>
<td>( 1/(1+\Psi') )</td>
</tr>
<tr>
<td>Stock 01</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Stock 02</td>
<td>0.161</td>
<td>0.839</td>
<td>0.861</td>
</tr>
<tr>
<td>Stock 03</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Stock 04</td>
<td>0.021</td>
<td>0.979</td>
<td>0.979</td>
</tr>
<tr>
<td>Stock 05</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Stock 06</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Stock 07</td>
<td>0.124</td>
<td>0.876</td>
<td>0.890</td>
</tr>
<tr>
<td>Stock 08</td>
<td>0.219</td>
<td>0.781</td>
<td>0.820</td>
</tr>
<tr>
<td>Stock 09</td>
<td>0.206</td>
<td>0.794</td>
<td>0.829</td>
</tr>
<tr>
<td>Stock 10</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Stock 11</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Stock 12</td>
<td>0.062</td>
<td>0.938</td>
<td>0.942</td>
</tr>
<tr>
<td>Stock 13</td>
<td>0.194</td>
<td>0.806</td>
<td>0.838</td>
</tr>
<tr>
<td>Stock 14</td>
<td>0.179</td>
<td>0.821</td>
<td>0.848</td>
</tr>
<tr>
<td>Stock 15</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
According to stock market experts’ views, the proposed DEASE approach is an efficient, applicable, and powerful approach with the ability to calculate performance and evaluate all stocks in the presence of negative data. Similar to the findings of Xie et al., [47], the experimental results also indicate the acceptable discriminatory power of the hybrid data envelopment analysis–Shannon entropy approach. Notably, the results that are obtained from the DEASE approach can be applied for the construction of desirable investment portfolios in the stock market by recognizing good stocks and filtering bad stocks.

5. Conclusions and Future Research Directions

Humans have always been interested in performance measurement. Efficiency is a criterion for performance measurement. There are two main methods for efficiency measurement: parametric methods that use different statistical methods and econometrics in order to estimate a certain production function, and non-parametric methods, which do not need any production function. DEA is a non-parametric method that calculates the relative efficiency of a set of DMUs. The DEA method uses the inputs and outputs of these units, and they are classified as efficient and inefficient. In this paper, the hybrid data envelopment analysis–Shannon entropy approach is presented with the aim of input/output selection and increasing the discriminatory power of DEA models, in order to evaluate the performance of DMUs under negative data. Notably, the main limitation of the study is that the proposed DEASE approach is not capable to be used under data uncertainty. Accordingly, for future research, the DEASE approach can be proposed under uncertain data, including fuzzy [76–90], stochastic [91–102], and interval [103–113] data. Additionally, the DEASE approach can be combined with machine learning approaches for the prediction of input and output data, and consequently, evaluation of the future performance of DMUs [114–125].


Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in the study are available from the authors and can be obtained upon reasonable requests.

Acknowledgments: The authors would like to thank the anonymous reviewers and the editor-in-chief for their constructive comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References
5. Liu, J.S.; Lu, L.Y.; Lu, W.M. Research fronts in data envelopment analysis. Omega 2016, 58, 33–45. [CrossRef]


23. Lee, H.S.; Choi, Y. Environmental performance evaluation of the Korean manufacturing industry based on sequential DEA. *Sustainability* 2019, 11, 874. [CrossRef]


64. Hwang, Y.G.; Park, S.; Kim, D. Efficiency analysis of official development assistance provided by Korea. *Sustainability* **2018**, *10*, 2697. [CrossRef]
115. Mirmozaffari, M.; Shadkam, E.; Khalili, S.M.; Yazdani, M. Developing a novel integrated generalised data envelopment analysis (DEA) to evaluate hospitals providing stroke care services. *Bioengineering* 2021, 8, 207. [CrossRef]
