Data driven approach for stochastic data envelopment analysis

Hengki Tamando Sihotang1, Syahril Efendi2, Muhammad Zarlis2, Herman Mawengkang3
1Doctoral Program, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia
2Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia
3Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara, Medan, Indonesia

Article Info

ABSTRACT

Decision making based on data driven deals with a large amount of data will evaluate the process's effectiveness. Evaluate effectiveness in this paper is measure of performance efficiency of data envelopment analysis (DEA) method in this study is the approach with uncertainty problems. This study proposed a new method called the robust stochastic DEA (RSDEA) to approach performance efficiency in tackling uncertainty problems (i.e., stochastic and robust optimization). The RSDEA method develops to combine the stochastics DEA (SDEA) formulation method and Robust Optimization. The numerical example demonstrates the performance efficiency of the proposed formulation method, with the results performing confirmed that the efficiency value is 89%.

This is an open access article under the CC BY-SA license.

Corresponding Author:
Hengki Tamando Sihotang
Doctoral Program, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara
Jl. Dr. T. Mansur No. 9, Padang Bulan, Kota Medan, Sumatera Utara 20222, Indonesia
Email: hengkitamando@students.usu.ac.id

1. INTRODUCTION

Today, the development of information technology in data driven decision making is an important things in business world. The idea of the data driven decision making is that every decision based on data must be generalized from an effective set of primary information. Decision making is not only related with data but also related to measure of performance efficiency. Performance efficiency measurement deals with data envelopment analysis (DEA). In classic concept, DEA is non-parametric technique first proposed by [1] called Charnes, Cooper and Rhodes (CCR) model. This model using linear programming technique that produce a single efficiency measurement. Base of DEA is a linear programming technique to evaluate the relative efficiency of multiinput and multi multioutput decision making unit (DMU) on the observed data [2] by comparing DMU one with others that the utilities is same resources to produce the same output [3].

The main goal of DEA is intended as a method for performance evaluation and benchmarking [4]. DEA techniques have been widely used by many researchers for various field, including: measurement of efficiency radiotherapy treatment [5]; hotel performance evaluation [6]; evaluation of efficiency of logistic sustainability performance [7]; evaluation of teaching performance in university [8]; and efficiency analysis in emergency departments [9], [10]. The efficiency evaluation process often involves a stochastic approach because of uncertainty inherent in the data in real life problems [11], [12], the stochastic DEA model using input and output data with stochastic variations, and discussed develop an algorithm to help organization for evaluating performance in stochastic problems. Khodabakhshi [13], describes the superefficiency problem based on the input relaxation model in the DEA stochastic model. The relaxation superefficiency input model was developed through stochastic DEA (SDEA), deterministic equivalence, and also nonlinear program derivatives. Khodabakhshi and Asgharian [14], describes an input relaxation model and a stochastic version
into DEA that uses more flexibility of change that using combination of inputs to find the maximum possible output. The optimization models, such as stochastic programming, chance constraint programming, and robust optimization (RO) have been proposed to support decision making in uncertainty problems [15].

Motivation this study is how to determinate of evaluation performance based on data with DEA framework with combine SDEA model and robust optimization model. The SDEA formulation methods in this studies proposed by [16]. Determination efficiency problems with DEA contains of large data is difficult. This related to the completion of liner programming with large data. Therefore, this level of difficulty increase with uncertainty input and output [17]. Hence, the authors proposed the methods robust stochastic data envelopment analysis as an approaches model to evaluate performance efficiency in uncertainty problems. To emphasise performance of efficiencies the proposed model, a comparison with SDEA model is made.

This study is organized into sections as: section 2 presents methods and materials, explains the based DEA method, stochastic DEA (SDEA), robust optimization, and develops formulation robust stochastic DEA (RSDEA) and a numerical example. Section 3 explains the result and discussion. Section 4 refers to the conclusion about studies.

2. METHOD AND MATERIALS

This section will introduce the concept of the based DEA model as a base of the framework in this research and preparation structured to build SDEA and the new RSDEA method.

2.1. Based method of DEA

Suppose there are 𝑛 DMUs, each of them have varying amounts of 𝑚 different inputs to produce 𝑠 different outputs denoted by 𝑥𝑖𝑗, ..., 𝑥𝑚𝑗 and 𝑦𝑖𝑗, ..., 𝑦𝑠𝑗, as the respective input and output vectors of 𝐷𝑀𝑈𝑗.

This study using the based method of DEA provided by [1] as:

\[
\begin{align*}
\max & \quad \text{maximize } u, y \sum_{r=1}^{s} u_r y_{r0} \\
\text{Subject to:} & \\
& \sum_{i=1}^{m} v_i x_{i0} = 1 \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} u_i x_{ij} \leq 0 \\
& j = 1, \ldots, n \\
& v_i \geq 0 i = 1, \ldots, m \\
& u_r \geq 0 r = 1, \ldots, s
\end{align*}
\]

2.2. SDEA

The following development SDEA methods assume a stochastic framework with all input variables transformed to output variables. Let variable of 𝐷𝑀𝑈ₖ(𝑠 = 1, 2, 3, ..., 𝑟) which all of them have random behavior. Let us further assume that \( \tilde{u}_r = (\tilde{u}_{1r}, \tilde{u}_{2r}, \ldots, \tilde{u}_{sr}) \) for each \( \tilde{u}_{rj} (r = 1, \ldots, s) \) denoting output components has probability distribution. By maximizing the expected of efficiency evaluated DMU, as:

MaxE(\( y^t \tilde{u}_0 \))

Subject to:

\[
\begin{align*}
\text{prob}(y^t \tilde{u}_j \leq \beta_j) & \geq 1 \\
u & \geq 0
\end{align*}
\]
In (2) transformed into a deterministic DEA form by considered from stochastic DEA. In [18], model Chance constraint programming have new variables with standard normal distribution, as following model:

$$\tilde{z}_j = \frac{y^j(\tilde{u}_j-\bar{u}_j)}{\text{var}_j}, j = 1,2,\ldots,N$$

(3)

By substituted (2), (3) as:

$$\text{prob} \left( \tilde{z}_j \leq \frac{\beta_j-y^j\bar{u}_j}{\text{var}_j} \right) \geq 1 - \alpha_j$$

(4)

Furthermore, with simply converted into following form:

$$\frac{\beta_j-y^j\bar{u}_j}{\text{var}_j} \geq \phi^{-1}(1 - \alpha_j), j = 1,2,\ldots,N$$

(5)

From (2), (5), simply transformation as following equivalent linear programming model as follows:

$$\text{Maximize} \ y^j\bar{u}_0$$

Subject to:

$$\beta_j - y^j\bar{u}_j \geq y^j\beta \phi^{-1}(1 - \alpha_j), j = 1,2,\ldots,N$$

(6)

2.3. Robust optimization

Robust optimization model when decision makers have to choose a strategy without knowing the exact value taken from an uncertain parameter. Robust optimization become popular approach as a robust methodology for solving the model optimization in uncertainty problems by providing a feasible solution in it’s the objective function. This approach to optimized mathematical uncertainty and maintain tractability of the model as proposed by [19]. Robust Optimization Model has been used to make decisions in dynamic environments where future decisions depends on the realization of data such as inventory issues [20], contracts [21] and project expansion planning [22]. Uncertainty has two forms: (i) estimation for constant value parameter, (ii) stochastic random variables. Optimization developed based on 2 principles [23]–[25], a modern practice in operations uncertainty management: i) estimating with often fail point and must be replaced with range estimates; ii) estimating with aggregate are more accurate than individual estimates. To present the robust framework in mathematics way [26] consider following linier programming problem:

$$\text{Minimize} \ c'x$$

Subject to: $Ax \geq b,$

$$x \in X$$

(8)

One of based issues of (6), the uncertainty problem about feasibility. In particular, the decision maker will ensure that every constraint is included for the possible value of A within a certain specified uncertainty of A. The set A is assumed each $a_{ij}$ coefficient of matrix $A$ is an uncertainty constraint, and all coefficients are independent. The decision maker knows the various estimates for all the uncertain parameters, in particular the parameter $a_{ij}$ belongs to a symmetrical interval $[\tilde{a}_{ij} - \bar{a}_{ij}, \tilde{a}_{ij} + \bar{a}_{ij}]$ centered on the approximate point $a_{ij}$ when measuring the accuracy of estimate.

2.4. Developed RSDEA method

In this study, the formulation of the proposed RSDEA aims to maximize the efficiency of the multi-uncertainty output of the stochastic DEA (SDEA) methods and its robust optimization. The formulation method is an equivalent deterministic model to be solved using linear programming.

Let for each variable of $\text{DMU}(s = 1,2,3,\ldots,r)$ has random with expressed by $\tilde{u}_s = (\tilde{u}_{1r}, \tilde{u}_{2r}, \ldots, \tilde{u}_{pr})$, for each $\tilde{u}_{tr}(t = 1,2,3,\ldots,s)$ also suppose all in variable are jointly probability distributed. The uncertainty of DEA for $(\tilde{x}_{ij})$ input or $(\tilde{y}_{ij})$ output are data wich is developed by the error. By the maximize the objective of (6), then the Robust DEA method be given as:
Maximize \( c'x \)
Subject to:
\[
\sum_{j=1}^{n} c'x \leq b_i
\]
\[
\sum_{i=1}^{m} \tilde{v}_i \tilde{x}_{i0} = 1
\]
\[
\sum_{r=1}^{s} \tilde{a}_r \tilde{y}_{rj} - \sum_{i=1}^{m} \tilde{u}_i \tilde{x}_{ij} \leq 0; j = 1, \ldots, n
\]
\[
\tilde{v}_i \geq 0; i = 1, \ldots, n
\]
\[
\tilde{a}_r \geq 0; r = 1, \ldots, s
\] (8)

Substitue (8) with (2) as:

Maximize: \( \sum_{r=1}^{t} \tilde{a}_r \tilde{y}_{r0} \) (9)
Subject to: \( \sum_{r=1}^{s} \tilde{a}_r \tilde{y}_{r0} \leq b_i \)
\[
\sum_{i=1}^{m} \tilde{v}_i \tilde{x}_{i0} = 1
\]
\[
\sum_{r=1}^{s} \tilde{a}_r \tilde{y}_{rj} - \sum_{i=1}^{m} \tilde{u}_i \tilde{x}_{ij} \leq 0
\]
\[
\tilde{v}_i \geq 0
\]

Since (9), at the first constraint \( \sum_{r=1}^{s} \tilde{a}_r \tilde{y}_{r0} \leq b_i \), reformulated to giving the score of super efficiency. In (9) will change with the constraint at (7) in order to the constraint not deviation on surplus variable \( M \). Further at the constraint (7) there is an inequality greater than equal to, so the constraint (7) reformulated as:
\[
y^t(b_j\phi^{-1}(1 - \alpha_j) + \tilde{u}_j) \leq \beta_j \quad j = 1, \ldots, N
\] (10)

Let \( b_j = 1 \) and all of input convert to be output with \( \alpha = \) risk level and also \( \beta = \) aspiration level. Then the formulation (9) and (10) called RSDEA can be formulated as:

Maximize \( \sum_{r=1}^{s} \tilde{a}_r \tilde{y}_{r0} \) (11)
Subject to: \( y^t(b_j\phi^{-1}(1 - \alpha_j) + \tilde{u}_j) \leq \beta_j \quad j = 1, \ldots, N \) (12)
\[
\sum_{i=1}^{m} \tilde{v}_i \tilde{x}_{i0} = 1
\]
\[
\sum_{r=1}^{s} \tilde{a}_r \tilde{y}_{rj} - \sum_{i=1}^{m} \tilde{u}_i \tilde{x}_{ij} \leq 0 \quad j = 1, \ldots, n
\]
\[
(\tilde{v}_i, \tilde{u}_r) \geq 0 \quad i = 1, \ldots, m; r = 1, \ldots, s
\]

In other words, the formulation formed above produces a method formulation with the objective function of maximizing the expectation of DMU evaluation on the number of outputs by considering the following constraints: a. multiplying the cumulative distribution function of the output is less than or equal to the value of the aspiration level, b. the result of adding up all the inputs must be equal one, c. the result of the difference between the number of outputs and inputs must be less than zero.
2.4. Numerical example
To understand the performance of the proposed RSDEA, we implement proposed (11) to (14) to analyze the efficiency performance of the college. For estimating the efficiency performance measure with this RSDEA, this study adopts 12 DMUs with four inputs and four outputs. In this case, given value of \( \alpha = 0.2, \beta = 0.9 \) and \( \phi^{-1} = 1.0625 \). Table 1 gives the information data of inputs and outputs. All of the data inputs are the mean of the number of lecturers, students, research of funding, employees. As well as the outputs data which is the mean of the number of alumni, research publications, and LPM.

<table>
<thead>
<tr>
<th>Table 1. Information data of DMUs with 4 inputs and outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMU</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>DMU₁</td>
</tr>
<tr>
<td>DMU₂</td>
</tr>
<tr>
<td>DMU₃</td>
</tr>
<tr>
<td>DMU₄</td>
</tr>
<tr>
<td>DMU₅</td>
</tr>
<tr>
<td>DMU₆</td>
</tr>
<tr>
<td>DMU₇</td>
</tr>
<tr>
<td>DMU₈</td>
</tr>
<tr>
<td>DMU₉</td>
</tr>
<tr>
<td>DMU₁₀</td>
</tr>
<tr>
<td>DMU₁₁</td>
</tr>
<tr>
<td>DMU₁₂</td>
</tr>
</tbody>
</table>

3. RESULTS AND DISCUSSION
This study aims to develop the new robust stochastic DEA (RSDEA) methods to tackle uncertainty problems (i.e., stochastic and robust optimization. The objective function of the RSDEA is to maximize the efficiency of multi-output from the uncertain stochastic model and robust optimization. In this case, input variable denotes \( v \) and output variable denotes \( u \). The numerical example as performed the proposed methods to determine performance efficiency, we use the case study to maximize the output to get efficiency performance of the college. On the Table 1 show 12 DMUs with 4 inputs and 4 outputs. Furthermore, based on a case, we solved the formulation 12 times according to the number of DMUs, with each formulation having the same constraints. Developing this formula was extended from the formulation of the proposed method for DMU₁ as according to this study case as:

Maximize \( 610u₁ + 5u₂ + 5u₃ + 5u₄ \) \hspace{1cm} (15)

Subject to:
\begin{align*}
610.85u₁ + 5.85u₂ + 5.85u₃ + 5.85u₄ + 17.85u₅ + 588.85u₆ + 200.85u₇ + 5.85u₈ & \leq 0.9 \hspace{1cm} (16) \\
533.85u₁ + 5.85u₂ + 5.85u₃ + 5.85u₄ + 26.85u₅ + 747.85u₆ + 250.85u₇ + 6.85u₈ & \leq 0.9 \hspace{1cm} (17) \\
195.85u₁ + 5.85u₂ + 4.85u₃ + 5.85u₄ + 15.85u₅ + 396.85u₆ + 150.85u₇ + 4.85u₈ & \leq 0.9 \hspace{1cm} (18) \\
300.85u₁ + 5.85u₂ + 5.85u₃ + 5.85u₄ + 17.85u₅ + 467.85u₆ + 200.85u₇ + 3.85u₈ & \leq 0.9 \hspace{1cm} (19) \\
252.85u₁ + 5.85u₂ + 7.85u₃ + 5.85u₄ + 25.85u₅ + 348.85u₆ + 250.85u₇ + 2.85u₈ & \leq 0.9 \hspace{1cm} (20) \\
224.85u₁ + 5.85u₂ + 5.85u₃ + 5.85u₄ + 23.85u₅ + 499.85u₆ + 200.85u₇ + 2.85u₈ & \leq 0.9 \hspace{1cm} (21) \\
326.85u₁ + 5.85u₂ + 6.85u₃ + 5.85u₄ + 19.85u₅ + 420.85u₆ + 220.85u₇ + 1.85u₈ & \leq 0.9 \hspace{1cm} (22) \\
273.85u₁ + 5.85u₂ + 5.85u₃ + 5.85u₄ + 17.85u₅ + 689.85u₆ + 200.85u₇ + 4.85u₈ & \leq 0.9 \hspace{1cm} (23) \\
284.85u₁ + 5.85u₂ + 4.85u₃ + 5.85u₄ + 34.85u₅ + 822.85u₆ + 350.85u₇ + 5.85u₈ & \leq 0.9 \hspace{1cm} (24) \\
204.85u₁ + 5.85u₂ + 4.85u₃ + 5.85u₄ + 10.85u₅ + 501.85u₆ + 100.85u₇ + 7.85u₈ & \leq 0.9 \hspace{1cm} (25) \\
273.85u₁ + 5.85u₂ + 8.85u₃ + 5.85u₄ + 11.85u₅ + 719.85u₆ + 100.85u₇ + 4.85u₈ & \leq 0.9 \hspace{1cm} (26) \\
183.85u₁ + 5.85u₂ + 4.85u₃ + 5.85u₄ + 11.85u₅ + 262.85u₆ + 100.85u₇ + 2.85u₈ & \leq 0.9 \hspace{1cm} (27) \\
17v₁ + 588v₂ + 200v₃ + 5v₄ & = 1 \hspace{1cm} (28)
\end{align*}
The formulation of the RSDEA can be solved using the LINDO software as shown in Figure 1. The calculation is carried out 12 times, according to the number of DMUs to the case. Each DMU (this case the DMU is defined as major of department) has a different objective function with the same constraints.

This study proposed a new method called RSDEA to evaluate the effectiveness of performance to tackle uncertainty problems with deal-driven input and output. Assess the efficiency of the proposed model; a confirmed case of the college with 12 departments of majors, 4 inputs, and 4 outputs was studied, as shown in Table 1. The data inputs are mean values of lecturers, students, research of funding, employees. As well as the outputs data which is the mean values of alumni, research publications, and LPM. In this case, given value of \( \alpha = 0.2 \). Furthermore, the RSDEA develops by extended SDEA with the robust optimization method. In (15) is only to maximize the output without changing the inputs. For (16) until (27) are the constraints of the SDEA formula where all inputs change with outputs, then there are eight outputs.
(28) until (40) also the constraints from the DEA framework. The result of RSDEA shows in Table 2. The higher value of the optimal efficiency from Table 2 is 0.8987476 or represents 89% at DMU1. That means the optimal performance efficiency with the RSDEA method obtained 89% only at DMU1.

<table>
<thead>
<tr>
<th>DMUs</th>
<th>Result of RSDEA formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMU1</td>
<td>0.8987476</td>
</tr>
<tr>
<td>DMU2</td>
<td>0.7379925</td>
</tr>
<tr>
<td>DMU3</td>
<td>0.7274454</td>
</tr>
<tr>
<td>DMU4</td>
<td>0.7136675</td>
</tr>
<tr>
<td>DMU5</td>
<td>0.7692308</td>
</tr>
<tr>
<td>DMU6</td>
<td>0.7692308</td>
</tr>
<tr>
<td>DMU7</td>
<td>0.7692308</td>
</tr>
<tr>
<td>DMU8</td>
<td>0.6718858</td>
</tr>
<tr>
<td>DMU9</td>
<td>0.4292363</td>
</tr>
<tr>
<td>DMU10</td>
<td>0.7692308</td>
</tr>
<tr>
<td>DMU11</td>
<td>0.8316246</td>
</tr>
<tr>
<td>DMU12</td>
<td>0.7692308</td>
</tr>
</tbody>
</table>

4. CONCLUSION

We have presented perform the proposed RSDEA method to evaluate the efficiency with uncertainty multi-inputs and multi-outputs. The development of RSDEA methods carries out using linear programming techniques. RSDEA methods have the optimal solution to evaluate the effectiveness of performance to tackle uncertainty problems. Recommendation for the implementation of further research considering extending this approach by examining the characteristics of data uncertainty at each input and output by providing the best scenario in constraints for maximizing or minimizing the objectives of the RSDEA.

ACKNOWLEDGEMENTS

The author would like to thank the Faculty of Computer Science and Information Technology, University of North Sumatra, and my supervisor. The author would also like to thank the anonymous reviewers for their valuable and constructive feedback.

REFERENCES


Data driven approach for stochastic data envelopment analysis (Hengki Tamando Sihotang)


BIOGRAPHIES OF AUTHORS

Hengki Tamando Sihotang is a student of the Doctoral program in computer science, University of North Sumatra 2018, an author who completed his undergraduate education at the Informatics Engineering Study Program, STMIK Pelita Nusantara, Completed his Studies Computer Master's Program from STMIK Eresha Jakarta, and is currently completing his Computer Science Doctor's Studies at the University of North Sumatera, Research interest in the fields of operation research and information science, where he is the author/co-author of more than 13 research publications H-index 4. He can be reached by email: hengkitamandos26@gmail.com.

Syahril Efendi is an Associate Professor at the Faculty of Computer Science and Information Technology, University of North Sumatra, From 2021, he also became Secretary of the Doctoral Study Program of Computer Science University of North Sumatra. is an author who completed his undergraduate education at the Faculty of Mathematics, University of North Sumatera, Medan. Completed Master of Computer education in computer science study program at Universitas Kebangsaan Malaysia (UKM), and completed Doctorate Program education at Faculty of Mathematics, University of North Sumatera, Medan. Research interest in the fields of operation research and information science, where he is the author/co-author of more than 105 research publications. He can be reached by email: syahril@usu.ac.idis.

Muhammad Zarlis is a Professor at the Faculty of Computer Science and Information Technology at Universitas Sumatera Utara, Medan, Indonesia. Completed his Bachelor of Physics education at the Faculty of Mathematics, Universitas Sumatera Utara, Medan. Completed his Masters in Computer Science in the Computer Science study program at the Faculty of Computer Science, Universitas Indonesia, Jakarta, and completed his Doctorate program in Computer Science at Universiti Sains Malaysia (USM), Penang. Current research interests are in the fields of intelligent systems, computer security and computation, where he is the author/co-author of more than 187 research publications. He can be reached by email: m.zarlis@usu.ac.id.

Herman Mawengkang is a Professor at the Faculty of Mathematics, Universitas Sumatera Utara, Medan, Indonesia. Completed his Bachelor of Mathematics education at the Faculty of Mathematics, Universitas Sumatera Utara, Medan in 1974. Completed his Doctorate Program in Mathematics education at the University of New South Wales, Sydney, Australia. Current research is conducted in the fields of computer science and applied mathematics, where he is the author/co-author of more than 153 research publications. He can be reached by email: hmawengkang@yahoo.com, hmawengkang@usu.ac.id.