Available online at http://ijdea.srbiau.ac.ir

Int. J. Data Envelopment Analysis (ISSN 2345-458X)

Vol.4, No.3, Year 2016 Article ID IJDEA-00422, 8 pages
Research Article

International Journal of Data Envelopment Analysis

Science and Research Branch (IAU)

Choosing the Best Bundle of Projects: A DEA Approach

Mahnaz Mirbolouki *

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

Received April 17, 2016, Accepted July 29, 2106

Abstract
One of the many important decisions organizations must make is project selection. Every project includes an initial plan to run, but not every plan can be implemented as a project. In situations where they lack resources or funds, all different plans must first be able to assess profitability in an accurate way, leading to the selection of a combination of proposals to carry out as projects. This paper provides a method to select the most effective set of proposals while considering the maximum use of available resources. Assessing efficiency is considered by a data envelopment analysis (DEA) model. Note that it is assumed that a vector of limited sources is at hand. This vector of resources can be contained human resource, budget, equipment, and facilities. Here, while reviewing some of the models in the selection project field, a common set of weight approach performance evaluation model for assessing the efficiency of the selected proposals is proposed. In this article it has tried to resolve the problems of former models.

Keywords: Data envelopment analysis; Efficiency; Project selection; Stochastic programming.

*. Email: m.mirbolouki@srbiau.ac.ir

1. Introduction

Many researchers investigated the issue of selection proposals in various applications such as construction projects, portfolio selection, R&D proposals, location/ allocation problem, etc. Albino and Gravelli [1] proposed a neural network approach for subcontractor selection. They studied the neural network implementation and the related management and technical innovations by an application case related to an assembly operation in a construction site. Bhattacharyya et al. [2] presents a fuzzy multi-objective programming approach to facilitate decision making in the selection of R&D projects. Park et al. [9] considered preferences in National Institutes of Health (NIH) projects using natural experiment in research funding.

DEA was developed by Charnes et al. [4]. Nowadays, DEA is well-deployed in other industries with many papers published on its utilization for performance measurement and decision making. Several researchers using DEA models proposed several models to choose the best combination of projects with different objectives. Kumar et al. [8] studied six sigma projects and its analysis using DEA to identify projects, which result in maximum benefit. Maximum benefit here provides a Pareto optimal solution based on inputs and outputs directly related to the efficiency of the six sigma projects under study. Tavana et al. [10] have considered the problem of assessment and selection of high-technology projects at the National Aeronautic and Space Administration (NASA). They proposed a DEA model with uncertainty in input and output data, which is modeled with fuzzy sets. Their models are capable of maximizing the satisfaction level of the fuzzy objectives and efficiency scores, simultaneously. Moreover, these models are capable of generating a common set of multipliers for all projects in a single run. Chang and Lee [3] discussed the specific problem of selecting a portfolio of projects that achieves an organization's objectives without exceeding limited capital resources, especially when each project possesses vague input and output data in the selection. In this paper, a DEA knapsack formulation and fuzzy set theory integrated model was proposed. Cook and Green [5] investigated selection problem among a large set of proposals when there are some budget limitations. Their approach treats each subset of the projects that could feasibly be selected within the resource constraints as a single, composite project. These composite projects are then evaluated, by DEA.

Although the proposed model in this paper is similar to Cook and Green [5] paper in some details, it is different in the objective function. Also, in the special case where suggested proposals' income is imprecise stochastic DEA (SDEA) approach is proposed and deterministic equivalent is presented.

In this paper, we consider the problem of choosing appropriate combinations of proposals, which has the best efficiency measure. Assessing efficiency is considered by DEA and SDEA models.

Our approach is organized in five sections. In the following section preliminaries on efficiency evaluation by a DEA model is presented. Section 3 discusses the proposed model for deterministic data and section 4 is the extension of this problem for imprecise outputs of proposals. Section 5 is the conclusion.

2. Preliminaries

DEA is a non-parametric multi-factor productivity analysis model that evaluates the relative efficiencies of a homogenous set of decision-making units (DMUs) in the presence of multiple input and output factors.

Consider n homogeneous DMUs $DMU_j; j=1,\dots,n$, where each DMU_j uses input vector $x_j \in R^m$ to produce output vector $y_j \in R^s$. According to the DEA assumptions it is assumed that $x_j \underset{\neq}{\geq} 0, y_j \underset{\neq}{\geq} 0$. Then the efficiency of $DMU_o; o \in \{1,\dots,n\}$; by considering input weights $v_i, i=1,\dots,m$ and output weights $u_r, r=1,\dots,s$ is defined as follows:

$$E_o = \frac{u_1 y_{1o} + u_2 y_{2o} + \dots + u_s y_{so}}{v_1 x_{1o} + v_2 x_{2o} + \dots + v_m x_{mo}}$$

To calculate input and output weights so that maximum relative efficiency of DMU_o yields, Charnes et al. [4] proposed the following model:

$$\begin{aligned} \max \ & \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}} \\ s.t. \ & \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \le 1, \\ & j = 1,\ldots,n, \\ & u_r \ge \varepsilon,\ v_i \ge \varepsilon, \\ & r = 1,\ldots,s,\ i = 1,\ldots,m. \end{aligned} \quad (1)$$

where $\varepsilon > 0$ is an infinitesimal value to avoid vanishing the weights. This linear fractional programming problem can be reduced to a non-ratio format in the usual manner of Charnes et al. [4] using transformation $\sum_{i=1}^{m} v_i x_{io} = 1$. Thus model (1) can be expressed as:

$$\begin{aligned} \max \ & \sum_{r=1}^{s} u_r y_{ro} \\ s.t. \ & \sum_{i=1}^{m} v_i x_{io} = 1, \\ & \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0, \\ & j = 1,\ldots,n, \\ & u_r \ge \varepsilon,\ v_i \ge \varepsilon, \\ & r = 1,\ldots,s,\ i = 1,\ldots,m. \end{aligned} \quad (2)$$

The optimal objective of the above model, which is considered as the related efficiency of DMU_o, ranges between zero and one. DMU_o is called an efficient DMU if it receives a score of one. Through using the method, there is no need to have the initial weight for the related inputs and outputs of every DMU. In the other words, the best input and output weights of each DMU are achieved through solving a linear programming problem in order to get a higher efficiency. The related efficiency of each DMU calculated by the method is higher than the actual real value. To overcome this difficulty, the following model was proposed, which determines a set of optimal weights for all DMUs:

$$\begin{aligned} \max \ & \left\{ \frac{\sum_{r=1}^{s} u_r y_{r1}}{\sum_{i=1}^{m} v_i x_{i1}}, \frac{\sum_{r=1}^{s} u_r y_{r2}}{\sum_{i=1}^{m} v_i x_{i2}}, \ldots, \frac{\sum_{r=1}^{s} u_r y_{rn}}{\sum_{i=1}^{m} v_i x_{in}} \right\} \\ s.t. \ & \frac{\sum_{r=1}^{s} u_r y_{rn}}{\sum_{i=1}^{m} v_i x_{in}} \le 1 \quad j = 1,\ldots,n, \\ & u_r \ge \varepsilon, \qquad r = 1,\ldots,s \\ & v_i \ge \varepsilon, \qquad i = 1,\ldots,m \end{aligned} \quad (3)$$

The aim of this model is to determine a common set of weights to get the highest efficiency of all DMUs simultaneously. Model (3) is a multiple objective problem. There are some approaches to solve this model. Hosseinzadeh-Lotfi et al. [7] linearized model (3) using a goal programming approach, which minimizes the sum of deviations from the efficiency level. This model is as follows:

$$\begin{aligned} \min \ & \sum_{j=1}^{n} \varphi_j \\ s.t. \ & \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \varphi_j = 0, \\ & j = 1,\ldots,n, \\ & \varphi_j \ge 0, \qquad j = 1,\ldots,n, \\ & u_r \ge \varepsilon, \qquad r = 1,\ldots,s, \\ & v_i \ge \varepsilon, \qquad i = 1,\ldots,m. \end{aligned} \quad (4)$$

Suppose (u^*, v^*, φ^*) is an optimal solution of model (4), then the efficiency score of DMU_j can be calculated using the following expression:

$$\theta_j^* = \frac{\sum_{r=1}^{s} u_r^* y_{rj}}{\sum_{i=1}^{m} v_i^* x_{ij}} = 1 - \frac{\varphi_j^*}{\sum_{i=1}^{m} v_i^* x_{ij}}, \ j = 1,\ldots,n \quad (5)$$

3. Choosing the best subset of proposals

In this section we assume every proposal, which requests m types of resources to produce s types of outputs, as a DMU. Also, a manager encounter with the problem of choosing the most efficient subset of proposals with the following limitation on the available resources:

(a) There are at most b_i units of resource (input) *i*.

(b) The remaining resources after allocation to the final selected set; S^*, must be too little to allow other proposals into S^*.

For this purpose, assume binary variable k_j which take the value 1 if proposal j is selected to implement and zero otherwise. Consider the following constraints which are proposed by Cook and Green [5]:

$$\sum_{j=1}^{n} k_j x_{ij} + s_i = b_i, \ \forall i \tag{6}$$

$$(1-k_j)x_{ij} + Mk_j + Mh_{ij} \geq s_i + \frac{1}{M} \quad \forall i, \forall j \tag{7}$$

$$\sum_{i=1}^{m} h_{ij} \leq m-1 \qquad \forall j \tag{8}$$

Where M is considered as a big number. The constraints (6)-(8) satisfy both assumptions (a) and (b). If $k_j = 1$ then constrains (7) are obviously satisfied. In the situation that $k_j = 0$ and $x_{ij} \leq s_i$, constraints (6) and (7) result $h_{ij} = 1$, on the other hand $x_{ij} > s_i$ results $h_{ij} = 0$.

Constraints (8) lead to condition (b), since if $k_j = 0$, this constraint assure that at least one of $h_{1j}, h_{2j}, \ldots, h_{mj}$ gets zero value, which means that a not selected proposal could not be implemented by the remaining resources.

Now, consider the modified version of model (3), which contains the binary variables k_j's and resource constraints (6)-(8):

$$\max\left\{k_1\frac{\sum_{r=1}^{s}u_r y_{r1}}{\sum_{i=1}^{m}v_i x_{i1}}, k_2\frac{\sum_{r=1}^{s}u_r y_{r2}}{\sum_{i=1}^{m}v_i x_{i2}}, \ldots, k_n\frac{\sum_{r=1}^{s}u_r y_{rn}}{\sum_{i=1}^{m}v_i x_{in}}\right\}$$

$$\begin{aligned}
s.t.\ & k_j\left(\frac{\sum_{r=1}^{s}u_r y_{rj}}{\sum_{i=1}^{m}v_i x_{ij}}\right) \leq 1, \quad j=1,\ldots,n,\\
& \sum_{j=1}^{n} k_j x_{ij} + s_i = b_i, \quad i=1,\ldots,m, \ (9.1)\\
& (1-k_j)x_{ij} + Mk_j + Mh_{ij} \geq s_i + \frac{1}{M}\\
& i=1,\ldots,m, \ j=1,\ldots,n, \quad (9.2)\\
& \sum_{i=1}^{m} h_{ij} \leq m-1, \ j=1,\ldots,n\\
& k_j, h_{ij} \in \{0,1\},\\
& i=1,\ldots,m, \ j=1,\ldots,n, \ (9.4)\\
& s_i \geq 0, \ i=1,\ldots,m, \ (9.5)\\
& u_r, v_i \geq \varepsilon, \ r=1,\ldots,s, \ i=1,\ldots,m.
\end{aligned} \tag{9}$$

Model (9) is a nonlinear MIP model. This model can be converted to the following linear MIP model:

$$\begin{aligned}
\max\ & \sum_{j=1}^{n}\varphi_j\\
s.t.\ & \sum_{r=1}^{s}k_j u_r y_{rj} - \sum_{i=1}^{m}k_j v_i x_{ij} + \varphi_j = 0, \ j=1,\ldots,n\\
& \varphi_j \geq 0, \quad j=1,\ldots,n,\\
& u_r \geq \varepsilon, \quad r=1,\ldots,s,\\
& v_i \geq \varepsilon, \quad i=1,\ldots,m,\\
& (9.1)-(9.5).
\end{aligned} \tag{10}$$

The above model, which calculates the most efficient proposals, can be applied on many applications. Despite the applicability of the model, it cannot be applied when inputs and outputs of proposals are imprecise. In the next section, the extension of model (10) into a chance constrained programming model is presented where deterministic inputs against imprecise outputs are assumed. This assumption may be accrued in many real word applications such as construction projects, portfolio selections, urban development and etc. Each proposal has an opportunity to participate in the 'production possibility set' as well as to combine with other projects to be evaluated against possible technology and to be selected.

4. Selecting strategy for proposals with imprecise income

In this section, assuming known amounts of consumption and probabilistic level of income for each project, the extension of model (10) is examined. Here, DMUs with imprecise outputs are considered so that these values are as random variables with normal distribution.

Let's assume related outputs of DMU_j have normal random distributions as,
$\tilde{y}_{rj} \sim N(y_{rj}, \sigma^2_{rj})$
Then, the probabilistic form of the first constraint of model (10) can be stated as the following chance constrained approach:

$$P\left(\sum_{r=1}^{s} u_{rj}\tilde{y}_{rj} - \sum_{i=1}^{m} v_{ij}x_{ij} + \varphi_j = 0\right) \ge 1-\alpha \quad (11)$$

Where α is the level of error between zero and one. According to statistical rules, the left hand-side of inequality (11) gets the zero value. To make this inequality meaningful, by removing variable φ_j it can be changed to the following:

$$P\left(\sum_{r=1}^{s} u_{rj}\tilde{y}_{rj} - \sum_{i=1}^{m} v_{ij}x_{ij} \le 0\right) \ge 1-\alpha \quad (12)$$

Now defining random variable $\tilde{h}_j = \sum_{r=1}^{s} u_{rj}\tilde{y}_{rj}$, would be result the following deterministic constraints:

$$\sum_{r=1}^{s} u_{rj}y_{rj} - \sum_{i=1}^{m} v_{ij}x_{ij} + s_j - \Phi^{-1}(\alpha)\bar{\sigma}_j = 0$$
$$\bar{\sigma}^2_j = \sum_{r=1}^{s}\sum_{t=1}^{s} u_{rj}u_{tj}Cov(\tilde{y}_{rj}, \tilde{y}_{tj}) \quad (13)$$

Where Φ is the cumulative distribution function of the standard normal distribution and $\Phi^{-1}(\alpha)$, is its inverse in level of α. Also, $\Phi^{-1}(\alpha)\bar{\sigma}_j$ indicates the deviation from stochastic frontier with respect to α level of error and s_j is an alternative of φ_j.

Finally, the deterministic equivalent of the related chance constraint model (10) can be written as,

$$
\begin{aligned}
\min\ & \sum_{j=1}^{n}\varphi_j - \Phi^{-1}(\alpha)w_j \\
s.t.\ & \sum_{r=1}^{s} u_{rj}y_{rj} - \sum_{i=1}^{m} v_{ij}x_{ij} + \varphi_j - \Phi^{-1}(\alpha)w_j = 0, \\
& j = 1,\dots,n \\
& \sum_{r=1}^{s} u_{rj} + \sum_{i=1}^{m} v_{ij} + \varphi_j \le k_j M, \quad j = 1,\dots,n \\
& u_{rj} - u_{rt} \le (2-k_j-k_t)M, \quad j,t = 1,\dots,n, r = 1,\dots,s \\
& v_{ij} - v_{it} \le (2-k_j-k_t)M, \quad j,t = 1,\dots,n, i = 1,\dots,m \\
& w_j^2 = \sum_{r=1}^{s}\sum_{t=1}^{s} u_{rj}u_{tj}Cov(\tilde{y}_{rj}, \tilde{y}_{tj}), \quad j = 1,\dots,n, \ (I) \\
& \varphi_j \ge 0, w_j \ge 0, \quad j = 1,\dots,n \\
& u_{rj} \ge \varepsilon k_j, \quad j = 1,\dots,n, r = 1,\dots,s \\
& v_{ij} \ge \varepsilon k_j, \quad j = 1,\dots,n, i = 1,\dots,m \\
& (a)-(e)
\end{aligned}
\quad (14)
$$

In the model (14) constraints (a)-(e) are the same constraints in model (10) and constraints (I) make this model nonlinear.

Example 1. (a numerical example)
Consider eight DMUs with two inputs and one probabilistic output. Table 1 contains the related data of these DMUs. Now we are going to select the best combination of these DMUs in the limited resources condition. In this example it is assumed that the each DMU's output is normally distributed. The example is solved at the levels 5% and 60% of error by the model (14) and the results are shown in table 2. The results in table 2 show that when the error level is decreased the proper use of resources would be gain. Notation '-' in table2 is for unselected DMUs.

Table 1. inputs and output

	inputs		Output	
	input1	input2	mean	Variance
DMU1	7	9	12	9
DMU2	5	8	18	25
DMU3	10	4	8	1
DMU4	12	7	9	9
DMU5	5	11	13	9
DMU6	14	3	11	4
DMU7	7	6	19	36
DMU8	8	13	15	16
Resource	**35**	**25**		

Table 2. Computationally results of model (14)

	5% error		60% error	
	input1	input2	input1	input2
DMU1	-	-	7	9
DMU2	-	-	5	8
DMU3	10	4	-	-
DMU4	-	-	-	-
DMU5	-	-	-	-
DMU6	14	3	-	-
DMU7	-	-	7	6
DMU8	8	13	-	-
Remained Resource	**3**	**2**	**19**	**2**

The efficiency of the selected DMUs in each level of error is demonstrated in table 3. The results of this table indicate more efficient selection in lower error.

Table 3. efficiency of each selection

	5% error	60% error
DMU1	-	1
DMU2	-	1
DMU3	1	-
DMU4	-	-
DMU5	-	-
DMU6	1	-
DMU7	-	0.92
DMU8	1	-

5. Conclusion

In this paper, a MIP model is offered in order to optimize the selection of the most efficient and feasible subset of proposals. In this way, the amount of deviation from efficiency level is minimized. The issue of the choice of this efficient bunch is one of the issues raised in many of the organizations that their practical projects encounter with the lots of suggestions, while the lack of funds prevents them from implementing all those projects. In this paper the selection project is considered in two conditions. First considering all inputs and outputs deterministic and second considering imprecise income for projects. The second assumption is more compatible with many real world applications such as portfolio selection where the result of investment never is clear. The results of the numerical example in section 4 confirm the maximum use of existing resources and the efficient selected bundle.

Acknowledgment

The authors would like to extend their gratitude to Yadegar-e-Imam Khomeini (RAH) Shahr-e Rey Branch, Islamic Azad University because of their financial support to carry out a research project related to this article.

References

[1] Albino, V., and Gravelli, A., A Neural Network Application to Subcontractor Rating in Construction Firms. *Int. J. Proj. Manag.,* 1998, **16 (1)**, 9-14.

[2] Bhattacharyya, R., Kumar, P., and Kar, S., Fuzzy R&D portfolio selection of interdependent projects. *Comput. Math. Appl.,* 2011, **62(10)**, 3857-3870.

[3] Chang, P. T., and Lee, J. H., A fuzzy DEA and knapsack formulation integrated model for project selection. *Comput. Oper. Res.,* 2012, **39**, 112-125.

[4] Charnes, A., Cooper, W. W., and Rhodes, E., Measuring the efficiency of decision making units. *Eur. J. Oper. Res.*, 1978, **2**, 429-444.

[5] Cook, W. D., and Green, R. H., Project prioritization: a resource-constrained data envelopment analysis approach. *Socio Econ. Plan. Sci.,* 2000, 34, 85-99.

[6] El-Mashaleh, M. S., A Construction Subcontractor Selection Model. *Jordan J. Civil Eng.*, 2009, **3(4)**, 375-383.

[7] Hosseinzadeh-Lotfi, F., Hatami-Marbini, A., Agrell, P.J., Aghayi, N., Gholami, K. Allocating fixed resources and setting targets using a common-weights DEA approach, *Comput. Ind. Eng.*, 2013, **64**, 631-640.

[8] Kumar, U. D., Saranga, H., Ramírez-Márquez, J. E., and Nowicki, D., Six sigma project selection using data envelopment analysis, *TQM Mag.*, 2007**, 19(5)**, 419-441.

[9] Park, H., Lee, J., and Kim, B. C., Project selection in NIH: A natural experiment from ARRA. *Res. Policy*, 2015, **44**, 1145-1159.

[10] Tavana, M., Khalili-Damghani, K., and Sadi-Nezhad, S., A fuzzy group data envelopment analysis model for high-technology project selection: A case study at NASA. *Comput. Ind. Eng.*, 2013, **66**, 10-23.