Application of DEA for Selecting Most Efficient Information System Project with Imprecise Data

Soroosh Nalchigar[1], Seyed Mohammad Reza Nasserzadeh[2]

[1 & 2] Department of Information Technology Management, Faculty of Management, University of Tehran, Tehran, Iran

nalchigar@ut.ac.ir

naserzdh@ut.ac.ir

Abstract - **Selection of best Information System (IS) project from many competing proposals is a critical business activity which is very helpful to all organizations. While previous IS project selection methods are useful but have restricted application because they handle only cases with precise data. Indeed, these methods are based on precise data with less emphasis on imprecise data. This paper proposes a new integrated Data Envelopment Analysis (DEA) model which is able to identify most efficient IS project in presence of imprecise data. As an advantage, proposed model identifies most efficient IS project by solving only one Mixed Integer Linear Programming (MILP). Applicability of this method is indicated by using data set includes specifications of 8 competing projects in Iran Ministry of Commerce.**

Keywords - **Data envelopment analysis, Information system project selection, imprecise data.**

I. INTRODUCTION

A critical aspect of IT management is the decision whereby the best set of IS projects is selected from many competing proposals [1]. According to [2] an annual or quarterly managerial decision making activity that most IS managers have to perform is the selection of IS projects. Indeed, selecting the right IS project is a critical business activity that has been recognized and repeatedly emphasized by many researchers. The decision can be accomplished using scoring, ranking, decision trees, game theoretic approach, the Delphi technique, Analytical Hierarchy Process (AHP), goal programming, AHP in conjunction with goal programming, dynamic programming, linear 0-1 programming [1]. However, previously proposed selection methods deal with precise data settings with less emphasis on imprecise data. In other words, these typically used methods do not consider real business environment conditions in which calculation of a precise numerical value for some criteria is difficult. Indeed, IS project selection takes place under an incomplete, vague (intangible), and uncertain information environment. For instance, some factors like ''importance to user" are subjective and difficult to measure [3]. The current study attempts to overcome this shortcoming by proposing a new Data Envelopment Analysis (DEA) model which identifies most efficient IS project while considering imprecise data.

II. DEA MODELS

DEA is a non-parametric linear programming based technique for measuring the relative efficiency of a set of similar units, usually referred to as decision making units (DMUs). Since the pioneering work of [4] DEA has demonstrated to be an effective technique for measuring the relative efficiency of a set of DMUs which utilize the same inputs to produce the same outputs. New applications with more variables and more complicated models are being introduced [5].

In many applications of DEA, finding the most efficient DMU is desirable. Assume that there are n DMUs, ($\mathrm{DMU_j} : j = 1,2,...,n$) which consume m inputs ($\mathbf{x}_i : i = 1,2,\ldots,m$) to produce s outputs ($\mathbf{y}_r : r = 1,2,\ldots,s$). Reference [6] proposed Model (1) as an integrated model for finding most efficient DMU.

$$
\begin{aligned}
& M^* = \min M \\
& s.t. \\
& M - d_j \geq 0 \qquad j = 1,2,...,n \\
& \sum_{i=1}^{m} w_i x_{ij} \leq 1 \qquad j = 1,2,...,n \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} w_i x_{ij} + d_j - \beta_j = 0 \qquad j = 1,2,...,n \\
& \sum_{j=1}^{n} d_j = n - 1 \\
& 0 \leq \beta_j \leq 1, d_j \in \{0,1\} \quad j = 1,2,...,n \\
& w_i \geq \varepsilon \quad i = 1,2,...,m \\
& u_r \geq \varepsilon \quad r = 1,2,...,s
\end{aligned} \tag{1}
$$

In Model (1) u_r and w_i are the weights attached to the inputs and outputs. In addition, d_j as a binary variable represents the deviation variable of $\mathrm{DMU_j}$. β_j is considered in the model because of discrete nature of d_j. Indeed, it causes efficiency of units to be less than one. M represents maximum inefficiency which should be minimized. $\mathrm{DMU_j}$ is most efficient unit if and only if

978-1-4244-4870-8/09/$26.00 ©2009 IEEE

$d_j = 0$. Furthermore, [6] proposed a model for finding ε^* which is maximum non-Archimedean epsilon.

First constraint of Model (1) implies that *M* is equal to maximum inefficiency. Second constraint shows input-oriented nature of the Model (1). Considering d_j as a measure of inefficiency, third constraint cause efficiency of all units is less that 1. Forth constraint ($\sum_{j=1}^{n} d_j = n-1$, with the binary variables $d_j,\ j = 1,2,...,n$) implies "among all the DMUs for only most efficient unit" [6], say DMU_p, which has $d_p^* = 0$ in any optimal solution.

The objective function of Model (1) attempts to minimize the maximum inefficiency of DMUs.

It should be noted that Model (1) is based on CCR model and identify most CCR-efficient DMU. Indeed, Model (1) is not applicable for situations in which DMUs operating in variable return to scale. To overcome this drawback, [7] proposed an integrated model which is able to find most BCC-efficient DMU. However, these DEA models are applicable in situations in which data of DMUs are precise. In the next section, a new DEA model is proposed which is able to find most efficient DMU while considering imprecise data.

III. PROPOSED MODEL

References [8] and [9] discussed that some of the outputs and inputs are imprecise data in the forms of bounded data, ordinal data, and ratio bounded data as follows:

Bounded data:

$$\underline{y}_{rj} \le y_{rj} \le \overline{y}_{rj} \quad \text{and} \quad \underline{x}_{ij} \le x_{ij} \le \overline{x}_{ij} \quad \text{for } r \in \text{BO}, i \in \text{BI} \qquad (2)$$

where $\underline{y}_{rj}$ and $\underline{x}_{ij}$ are the lower bands and $\overline{y}_{rj}$ and $\overline{x}_{ij}$ are the upper bounds, and BO and BI represent the associated sets containing bounded outputs bounded inputs, respectively.

Weak ordinal data:

$$y_{rj} \le y_{rk} \quad \text{and} \quad x_{ij} \le x_{ik} \quad \text{for } j \ne k,\ r \in \text{DO}, i \in \text{DI}$$

Or to simplify the presentation,

$$y_{r1} \le y_{r2} \le ... \le y_{rk} \le ... \le y_{rn} \quad (r \in \text{DO}), \qquad (3)$$

$$x_{i1} \le x_{i2} \le ... \le x_{ik} \le ... \le x_{in} \quad (i \in \text{DI}), \qquad (4)$$

where DO and DI represent the associated sets containing weak ordinal outputs and inputs, respectively.

Strong ordinal data:

$$y_{r1} < y_{r2} < ... < y_{rk} < ... < y_{rn} \quad (r \in \text{SO}), \qquad (5)$$

$$x_{i1} < x_{i2} < ... < x_{ik} < ... < x_{in} \quad (i \in \text{SI}), \qquad (6)$$

where SO and SI represent the associated sets containing strong ordinal outputs and inputs, respectively.

Ratio bounded data:

$$L_{rj} \le \frac{y_{rj}}{y_{rj_o}} \le U_{rj} \qquad (j \ne j_o)\ (r \in \text{RO}) \qquad (7)$$

$$G_{ij} \le \frac{x_{rj}}{x_{ij_o}} \le H_{ij} \qquad (j \ne j_o)\ (i \in \text{RI}) \qquad (8)$$

where L_{rj} and G_{ij} represent the lower bounds, and U_{rj} and H_{ij} represent the upper bounds. RO and RI represent the associated sets containing ratio bounded outputs and inputs, respectively.

By adding Eqs. (2)-(8) added to Model (1), there will be:

$$
\begin{aligned}
& M^* = \min M \\
& s.t. \\
& M - d_j \ge 0 && j = 1,2,...,n \\
& \sum_{i=1}^{m} w_i x_{ij} \le 1 && j = 1,2,...,n \\
& \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} w_i x_{ij} + d_j - \beta_j = 0 && j = 1,2,...,n \\
& \sum_{j=1}^{n} d_j = n-1 \\
& 0 \le \beta_j \le 1, d_j \in \{0,1\} && j = 1,2,...,n \\
& (x_{ij}) \in \Theta_i^- \\
& (y_{rj}) \in \Theta_r^+ \\
& w_i \ge \varepsilon && i = 1,2,...,m \\
& u_r \ge \varepsilon && r = 1,2,...,s
\end{aligned} \qquad (9)
$$

where $(x_{ij}) \in \Theta_i^-$ and $(y_{rj}) \in \Theta_r^+$ represent any or all of Eqs. (2)-(8).

Clearly, Model (9) is nonlinear and non-convex, because some of the outputs and inputs become unknown decision variables. Since Model (9) is nonlinear and non-convex, consequently local optimum is produced and global optimum may remains unknown.

To convert Model (9) into the linear program (LP), some approaches exist in literature. For instance, [10] proposed a generalized model which is capable to handle both interval and ordinal data. Its approach is to transform a non-linear DEA model to a linear programming equivalent, on the basis of the original data set, by applying transformations only on the variables. In addition, [11] developed a simple approach by defining

$$X_{ij} = w_i x_{ij} \qquad \forall i,j$$
$$Y_{ij} = u_r y_{rj} \qquad \forall r,j \qquad (10)$$

In this paper, by adopting [11]'s approach, Model (9) is converted to Model (11) which is a MILP.

$$\begin{aligned}
& M^* = \min M \\
& s.t. \\
& M - d_j \geq 0 && j = 1,2,\dots,n \\
& \sum_{i=1}^{m} X_{ij} \leq 1 && j = 1,2,\dots,n \\
& \sum_{r=1}^{s} Y_{rj} - \sum_{i=1}^{m} X_{ij} + d_j - \beta_j = 0 && j = 1,2,\dots,n \\
& \sum_{j=1}^{n} d_j = n-1 \\
& 0 \leq \beta_j \leq 1, d_j \in \{0,1\} && j = 1,2,\dots,n \\
& X_{ij} \in \widetilde{D}_i^- \\
& Y_{rj} \in \widetilde{D}_i^+ \\
& X_{ij} \geq \varepsilon^* && \forall\ i,j \\
& Y_{rj} \geq \varepsilon^* && \forall\ r,j
\end{aligned} \qquad (11)$$

where Θ_i^- and Θ_r^+ are replaced by $\widetilde{D}_i^-$ and $\widetilde{D}_i^+$ with:

Bounded data:
$\underline{y}_{rj} u_r \leq Y_{rj} \leq \overline{y}_{rj} u_r$, $\underline{x}_{ij} w_i \leq X_{ij} \leq \overline{x}_{ij} w_i$.

Ordinal data:
$Y_{rj} \leq Y_{rk}$, $X_{ij} \leq X_{ik}$ $\forall j \neq k$ for some r,i.

Ratio bounded data:
$L_{rj} \leq \dfrac{Y_{rj}}{Y_{rj_o}} \leq U_{rj}$ and $G_{ij} \leq \dfrac{X_{rj}}{X_{ij_o}} \leq H_{ij}$ $(j \neq j_0)$.

Cardinal data:
$Y_{rj} = \hat{y}_{rj} u_r$ and $X_{ij} = w_i \hat{x}_{ij}$, where $\hat{y}_{rj}$ and $\hat{x}_{ij}$ represent cardinal data.

Furthermore, by extending [5]'s epsilon model, Model (12) is proposed to determine the non-Archimedean epsilon in Model (11).

$$\begin{aligned}
& \varepsilon^* = \max \varepsilon \\
& s.t. \\
& \sum_{i=1}^{m} X_{ij} \leq 1 && j = 1,2,\dots,n \\
& \sum_{r=1}^{s} Y_{rj} - \sum_{i=1}^{m} X_{ij} \leq 0 && j = 1,2,\dots,n \\
& X_{ij} \in \widetilde{D}_i^- \\
& Y_{rj} \in \widetilde{D}_i^+ \\
& X_{ij} - \varepsilon \geq 0 && \forall\ i,j \\
& Y_{rj} - \varepsilon \geq 0 && \forall r,j
\end{aligned} \qquad (12)$$

As mentioned in previous section, Model (1) is based on CCR model and evaluates DMUs in constant return to scale. Hence, Model (11), which is extended version of Model (1), is not applicable for situations in which DMUs operate in variable return to scale. By extending previous work of [7], Appendix A proposes another new DEA model which is able to find most BCC-efficient DMU in existence of imprecise data.

IV. APPLICATION OF PROPOSED MODEL

The manager of IS department at Iran Ministry of Commerce used to perform personally the task of selecting IS projects. The manager had to depend on experience and intuition. From an operational aspect, selecting the best project among large number of projects was difficult to do in order to the lack of a formal process and a clearly defined and transparent way for decision making. It was necessary to choose a model which could handle settings with imprecise data environment and could be used as an objective method for selecting right IS project. DEA was suggested as the best solution which could model the complexity of choosing projects and show that the decisions so made are fair and equitable.

To choose the best IS project from a set of projects that are in competition for limited resources, different criteria, which may be qualitative in nature, must be included in the evaluation. First, IS department of Iran Ministry of Commerce employed a team include several specialists on the field of IS and software engineering. These specialists were asked to develop a set of criteria which capture all aspects of IS projects and to estimate them for each project. It is notable that data of software cost, training cost and support cost values were obtained from proposals and potential risk, time reduction, system accuracy and improvement management capabilities were subjectively scored by specialist. Because of difficulties in measurement of precise numerical value for potential risk and improvement management capabilities, these data are in ordinal and interval format. Table (1) shows estimations of inputs and outputs obtained by specialist.

TABLE 1
DATA OF IS PROJECTS

DMU	Inputs: Software Cost x_{1j}	Training Cost x_{2j}	Support Cost x_{3j}	Potential Risk[a] x_{4j}	Outputs: Time Reduction y_{1j}	System Accuracy y_{2j}	Improvement management capabilities y_{3j}
1	3500	12	24	6	5	5	$[1,15]$
2	455	45	0.83	2	23	5	$[7,20]$
3	695	69	1.5	1	14	5	$[1,15]$
4	513	14	122	4	16	5	$[8,15]$
5	3510	351	16.8	3	23	5	$[5,10]$
6	3725	30	100	7	14	5	$[1,5]$
7	4000	40	50	5	10	5	$[1,5]$
8	2500	250	30	8	13	5	$[4,12]$

[a] Ranking such that 8 ≡ highest rank, …, 1 ≡ lowest rank $(x_{48} > x_{46} > ... > x_{43})$.

In this section, applicability of proposed model is illustrated. Inputs and outputs of IS projects are as follows:

$\Theta_1^- = \{x_{11} = 3500; x_{12} = 455; x_{13} = 695; ...; x_{18} = 2500\}$ (Cardinal data)

$\Theta_2^- = \{x_{21} = 12; x_{22} = 45; x_{23} = 69; ...; x_{28} = 250\}$ (Cardinal data)

$\Theta_3^- = \{x_{31} = 24; x_{32} = 0.83; x_{33} = 1.5; ...; x_{38} = 30\}$ (Cardinal data)

$\Theta_4^- = \{x_{48} > x_{46} > ... > x_{43}\}$ (Ordinal data)

$\Theta_1^+ = \{y_{11} = 5, y_{12} = 23, y_{13} = 14, ..., y_{18} = 13\}$ (Cardinal data)

$\Theta_2^+ = \{y_{21} = 5, y_{22} = 5, y_{23} = 5, ..., y_{28} = 5\}$ (Cardinal data)

$\Theta_3^+ = \{1 \le y_{31} \le 15; 7 \le y_{32} \le 20; 1 \le y_{33} \le 15, ..., 4 \le y_{18} \le 12\}$ (Interval data)

According to [11]'s approach:

$$\tilde{D}_1^- = \begin{Bmatrix} X_{11} = 3500w_1; X_{12} = 455w_1; X_{13} = 695w_1, ..., \\ X_{18} = 216w_1 \end{Bmatrix}$$

$$\tilde{D}_2^- = \begin{Bmatrix} X_{21} = 12w_2; X_{22} = 45w_2; X_{23} = 69w_2, ..., \\ X_{28} = 250w_2 \end{Bmatrix}$$

$$\tilde{D}_3^- = \begin{Bmatrix} X_{31} = 24w_3; X_{32} = 0.83w_3; X_{33} = 1.5w_3, ..., \\ X_{38} = 300w_3 \end{Bmatrix}$$

$$\tilde{D}_4^- = \{X_{48} > X_{46} > .... > X_{43}\}$$

$$\tilde{D}_1^+ = \{Y_{11} = 5\mu_1, Y_{12} = 23\mu_1, Y_{13} = 14\mu_1, ..., Y_{18} = 13\mu_1\}$$

$$\tilde{D}_2^+ = \{Y_{21} = 5\mu_2, Y_{22} = 5\mu_2, Y_{23} = 5\mu_2, ..., Y_{18} = 5\mu_2\}$$

$$\tilde{D}_3^+ = \begin{Bmatrix} \mu_3 \le Y_{31} \le 15\mu_3; 7\mu_3 \le Y_{32} \le 20\mu_3; \\ \mu_3 \le Y_{33} \le 15\mu_3; ...; 4\mu_3 \le Y_{38} \le 12\mu_3 \end{Bmatrix}$$

Using these equations and solving Model (11) for data presented in Table (1), (with considering suitable value for epsilon, equal to 0.0067) DMU_2 is easily identified as most CCR-efficient IS project ($d_2^* = 0,\ d_{j\ne2}^* = 1$). Detailed results of proposed model are presented in Table (2).

TABLE 2
RESULTS OF PROPOSED MODEL

Variable	Optimum Value
d_j^*	$d_2^* = 0,\ d_{j\ne2}^* = 1$
w_1^*	0.00001
w_2^*	0.00050
w_3^*	0.00803
u_1^*	0.00120
u_2^*	0.00120
u_3^*	0.00120

So DMU_2 is most efficient IS project and is proposed to allocate resource. Using the proposed method, IS department at Iran Ministry of Commerce is able to identify most efficient IS objectively. In brief, some of the clear benefits of the proposed model are:

- IT manager could find most efficient IS project by solving only one MILP, so it is computationally efficient.
- The model finds most efficient IS project objectively and there is no need for determining weights and pair-wise comparison for all criteria and projects.
- The model is applicable for settings in which data are imprecise.

V. CONCLUSION

Selecting best IS projects is a critical aspect of IT management and has been recognized and repeatedly emphasized by many researchers. Indeed, the optimal selection process is a significant strategic resource allocation decision that can engage an organization in substantial long-term commitments. Although there are numerous methods for IS project selection, prior researches have ignored the presence of imprecise data. Therefore, this paper developed upon the work conducted on IS project selection considering the cases in which data of projects are imprecise. This paper proposed a new DEA model which identifies most efficient IS project by solving only one MILP. As an advantage, this model is complex enough to handle cases with imprecise data well and accurately and yet simple enough to be understood by the IT managers' community. Finally, applicability of proposed method illustrated in real case of Iran Ministry of Commerce.

REFERENCES

[1] M. A. Badri, , D. Davis, D. Davis, "A comprehensive 0-1 goal programming model for project selection," *International Journal of Project Management*, 19, 243-252, 2001.

[2] M. J. Schniederjans, R. Santhanam, "A multi-objective constrained resource information system project selection method," *European Journal of Operational Research*, 70, 244-253, 1993.

[3] C. T. Chen, H. L. Cheng, "A comprehensive model for selecting information system project under fuzzy environment," *International Journal of Project Management,* 27, 389–399, 2009.

[4] A. Charnes, W. W. Cooper, & E. Rhodes, "Measuring the efficiency of decision-making units," *European Journal of Operational Research*, 2, 429–444, 1978.

[5] A. Emrouznejad, G. Tavares, & B. Parker, "A bibliography of data envelopment analysis (1978–2003)" *Socio-Economic Planning Sciences*, 38, 159-229, 2007.

[6] G. R. Amin, M. Toloo, "Finding the most efficient DMUs in DEA : An improved integrated model," *Computers & Industrial Engineering*, 52, 71-77, 2007.

[7] M. Toloo, S. Nalchigar, "A new integrated DEA model for finding most BCC-efficient DMU," *Applied Mathematical Modelling*, 33, 597-604, 2009.

[8] W. W. Cooper, K. S. Park, G. Yu, "IDEA and AR-IDEA: Models for dealing with imprecise data in DEA," *Management Science*, 45, 597–607, 1999.

[9] S. H. Kim, C. G. Park, K. S. Park, "An application of data envelopment analysis in telephone offices evaluation with partial data," *Computers and Operations Research*, 26, 59– 72, 1999.

[10] D.K. Despotis, Y.G. Smirlis, "Data envelopment analysis with imprecise data," *European Journal of Operational Research*, 140, 24–36, 2002.

[11] J. Zhu, "Imprecise data envelopment analysis (IDEA): A review and improvement with an application," *European Journal of Operational Research,* 144, 513–529, 2003.

APPENDIX A

Reference [7] presented Model (13) as a new integrated model for finding the most BCC-efficient DMU.

$$
\begin{aligned}
& M^* = \min M \\
& s.t. \\
& M - d_j \ge 0 \qquad j = 1,2,\ldots,n \\
& \sum_{i=1}^{m} w_i x_{ij} \le 1 \qquad j = 1,2,\ldots,n \\
& \sum_{r=1}^{s} u_r y_{rj} - u_0 - \sum_{i=1}^{m} w_i x_{ij} + d_j - \beta_j = 0 \qquad j = 1,2,\ldots,n \\
& \sum_{j=1}^{n} d_j = n-1 \\
& 0 \le \beta_j \le 1, d_j \in \{0,1\} \quad j = 1,2,\ldots,n \\
& M, u_0 \quad free \\
& w_i \ge \varepsilon^* \quad i = 1,2,\ldots,m \\
& u_r \ge \varepsilon^* \quad r = 1,2,\ldots,s
\end{aligned} \tag{13}
$$

Model (13) is computationally efficient and also has wider range of application than Model (1), because is capable for situation in which return to scale is variable. Incorporating Eqs. (2)-(8) to Model (13), and converting non-linear model to a LP (based on [11]'s approach), there will be:

$$
\begin{aligned}
& M^* = \min M \\
& s.t. \\
& M - d_j \ge 0 \qquad j = 1,2,\ldots,n \\
& \sum_{i=1}^{m} X_{ij} \le 1 \qquad j = 1,2,\ldots,n \\
& \sum_{r=1}^{s} Y_{rj} - u_0 - \sum_{i=1}^{m} X_{ij} + d_j - \beta_j = 0 \qquad j = 1,2,\ldots,n \\
& \sum_{j=1}^{n} d_j = n-1 \\
& 0 \le \beta_j \le 1, d_j \in \{0,1\} \qquad j = 1,2,\ldots,n \\
& X_{ij} \in \tilde{D}_i^- \\
& Y_{rj} \in \tilde{D}_i^+ \\
& X_{ij} \ge \varepsilon^* \qquad \forall\, i, j \\
& Y_{rj} \ge \varepsilon^* \qquad \forall\, r, j \\
& M, u_0 \quad free
\end{aligned} \tag{14}
$$

Model (14) finds most BCC-efficient DMU in presence of both cardinal and ordinal data. This model has wider range of application than Model (11).