An Approach in Solving Data Envelopment Analysis with Stochastic Data

Amir Hossein Nafei[1,2(✉)],
Seyed Mohammad Esmaeil Pourmohammad Azizi[1,2],
and Rajab Ali Ghasempour[1,2]

[1] Department of Mathematics, Semnan University, Semnan, Iran
amir208plg@yahoo.com
[2] Department of Mathematics, Allameh Tabatabae'i University, Tehran, Iran
{e.azizi, r.ghasempour}@atu.ac.ir

Abstract. The importance and necessity of the data envelopment analysis as a relevant and effective instrument in investigation of the performance of units studied such as banks and so forth is an evident issue. One of the main issues we encounter is envelope analysis of data with random variable values. In this research, by explaining the general model of the data envelopment analysis models and inspired by how to work on interval input/output, there would a solution for random state provided. In fact, the interval the random variable varies in is considered and using envelope analysis on interval data, we will determine the effective unit.

Keywords: Data envelopment analysis · Stochastic programing · Stochastic input and output · Interval data

1 Introduction

Data envelopment analysis is a decision making instrument and approach about the organizations performance. In this structure, the efficiency and effectiveness of entities would be studied. Throughout the world either in developed or developing countries, in a long-term planning for the future, there have always been 3 substantial principles of the efficiency increasing and enhancement so that they can at least reach their minimum economic growth. As theories developed, getting together by economies and elimination of their distance caused a competition in production and world trading. In this structure, undoubtedly, making use of different data in line with best selecting the entities for investment is a necessary issue. In this regard, lack of how to select the best companies would be eliminated by data envelopment analysis. Hence, the data envelopment analysis is called a mathematical model or planning provides the decision maker with the best selection based on available data. The importance of study in this area can be seen in different situations. For example, in a portfolio containing the risk assets such as stocks, the chance in structure of each model planning is subjected to issues such as calculations. In this article, we use available approaches in planning the data envelopment analysis issues for envelope analysis modeling using random data. In fact in this paper, we study the random data in data envelopment analysis modeling

B.-Y. Cao (ed.), *Fuzzy Information and Engineering and Decision*,
Advances in Intelligent Systems and Computing 646, DOI 10.1007/978-3-319-66514-6_17

structure. Studies conducted on data envelopment analysis has begun since 1950 and early 1960s. In fact, the first model of planning was announced by Farrell in 1975 whose structure was non-parametric and had one input and output. After Farrell, one can mention the Cupper, Charles and Rhodes (1978). This research fixated the studies in the area of data envelopment analysis. Their model was the expansion of Farrell's model with input/output variables. Their model is called CCR [1]. Using the definition of efficiency (P) as:

$$P = \frac{U * Y}{V * X}. \tag{1}$$

Where U, Y, V and X are output weight, output value, input weight and input value, respectively, they planned the chance constraint programming in case of a certain event occurrence [2]. This programming holds each chance in an event like λ. Land et al. extended this model to obtain the systems' performance with the only random output [3]. Olsen and Petersen used the Land et al.'s idea in CCR model [4]. Kwakernaak was the first to use the random concept with mixed fuzzy in data envelopment analysis [5]. Other articles in this regard are Jing Liang et al.'s research worked on a random environment model DEA which can measure the environmental performance under random conditions [6]. In next chapter, the CCR model, the interval solutions, random problem and necessary tools are investigated. Then the proposed state is provided and the interval solution and averaging procedures are proposed. Finally, the numerical results are analyzed.

2 The Basic Methods

In this chapter, 2 mathematical expectation and interval envelope analysis are investigated which are instruments to solve the random problems used.

2.1 Data Envelopment Analysis

The first data overage analysis model as a comprehensive one was proposed by Edward Roders in his PhD dissertation. In this thesis, the achievement of the students of Carnegie in USA was investigated and the CCR model was provided for the teacher's performance [7]:

$$\begin{aligned} & Min\, \theta \\ & s.t. \\ & \sum_{t=1}^{n} \lambda_t X_t \leq \theta X_k \; t = 1, \ldots, n \quad t \neq k, \\ & \sum_{t=1}^{n} \lambda_t Y_t \geq Y_k \; t = 1, \ldots, n \quad t \neq k, \\ & \lambda_t \geq 0 \quad t = 1, \ldots, n. \end{aligned} \tag{2}$$

In this structure, the efficiency value of unit k (DMU_k) is obtained. After CCR programming, the BCC model was build adding $\sum_{t=1}^{n} \lambda_t = 1$ to above model which is used by Charles used in order to make the efficiency fixed relative to the scale [8]. Above problem is as follow:

$$\begin{aligned} & MAX\ U^T Y_k \\ & s.t. \\ & V^T X_k = 1, \\ & U^T Y_t - V^T X_t \leq 0 \quad t = 1, \ldots, n, t \neq k, \\ & U, V \geq 0. \end{aligned} \tag{3}$$

2.1.1 Interval Data Envelopment Analysis

In interval data envelopment analysis method, the value of each data is in an interval and can be variable in this interval. If each of n units available uses m input to produce s output, then K^th unit performance is DMU_K which is from $\{X_{jk} | j = 1, \ldots, n\}$ in order to generate $\{Y_{ik} | i = 1, \ldots, s\}$. Now, if input and output are interval ones and have lower and upper bounds, then for unit k, the input j is as $\left[X_{jk}^L,\ X_{jk}^U\right]$ and output i is as $\left[Y_{ik}^L,\ Y_{ik}^U\right]$ denoted where L is lower bound and U is upper bound of interval and $X_{jk}^L,\ Y_{ik}^L \geq 0$ cannot be neglected. The model considered for interval state is written in 2 forms [9]. For upper bound calculation it is given as:

$$\begin{aligned} & MAX \sum_{i=1}^{s} U_i^T Y_{ik}^U \\ & s.t. \\ & \sum_{j=1}^{m} V_j^T X_{jk}^L = 1, \\ & \sum_{i=1}^{s} U_i^T Y_{it}^U - \sum_{j=1}^{m} V_j^T X_{jt}^L = 1\, t = 1, \ldots, n, t \neq k, \\ & U_i, V_j \geq 0\, i = 1, \ldots, m j = 1, \ldots, s. \end{aligned} \tag{4}$$

And, for lower bound calculation it is given as:

$$\begin{aligned} & MAX \sum_{i=1}^{s} U_i^T Y_{ik}^L \\ & s.t. \\ & \sum_{j=1}^{m} V_j^T X_{jk}^L = 1, \\ & \sum_{i=1}^{s} U_i^T Y_{it}^U - \sum_{j=1}^{m} V_j^T X_{jt}^L = 1\, t = 1, \ldots, n\, t \neq k, \\ & U_i, V_j \geq 0\, i = 1, \ldots, m j = 1, \ldots, s. \end{aligned} \tag{5}$$

2.2 Mathematical Expectation

The set S including all possible events of an experiment is called the sample space. This set can be finite or infinite, countable or uncountable. Suppose F is a member of subsets of S with following features:

1. If $A \in F$, then F includes the complement $A\,(A^c)$ too.
2. $S \in F$,
3. $\cup_{i \in I} A_i \in F$ in which I is the index set and for $i \in I$, we have $A_i \in F$.

The mapping $P : F \to [01]$ is called a probable value if:

1. $P(S) = 1$,
2. $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ in which for each i, $A_i \in F$ for $i \neq j$ and $A_i \neq A_j$.

The pair $(S,\ F)$ is called measure space and (S,F,P) is probable space.

In probable space, the mapping $X :\ (S, F) \to (\mathbb{R}, \beta)$ in which β is the set of all open sets in real digits set and is called random variable when $X^{-1}(B) \in F$ in which $B \in \beta$. Considering the probable space (S, F, P) and random variable X, the mathematical expectation, average and expected values or considered value expected from a random variable equals sum of multiplication results in their probability which is denoted as $E[X]$. In finite discrete state, the expectancy is denoted as $E[X] = \sum_{i=1}^{n} \chi_i p(\chi_i)$. In continuous state we will have $E[X] = \int \chi dP(\chi)$ if $f(x)$ is the distribution function of this random variable, $f(x) * dx = dP(x)$. Therefore, we will have $E[X] = \int \chi f(\chi) d\chi$ [10].

In this chapter, at first the data envelopment analysis and programming like envelope and multiple ones are investigated and then the interval model is introduced and using the classical approach, solving methodology would be provided and finally, the mathematical expectation is defined. All mentioned issues are solutions for programing problem solving with random data in data envelopment analysis which are explained in next chapter.

3 Solving Data Envelopment Analysis with Stochastic Data

In previous sections, the approaches used for random programming in data envelopment analysis were investigated. At first, we state the problem. Suppose in a structure with n unit, each unit needs for producing s output ad m input. In fact the data of this programming is as follow (Table 1):

Table 1. Input and output of system

DMU_1	DMU_2	·	$DMU_{(n-1)}$	DMU_n
X_{11}	X_{12}	·	$X_{1(n-1)}$	X_{1n}
X_{21}	X_{22}	·	$X_{2(n-1)}$	X_{2n}
X_{31}	X_{32}	·	$X_{3(n-1)}$	X_{3n}
·	·		·	·
·	·		·	·
·	·		·	·
$X_{(m-2)1}$	$X_{(m-2)2}$	·	$X_{(m-2)(n-1)}$	$X_{(m-2)n}$
$X_{(m-1)1}$	$X_{(m-1)2}$	·	$X_{(m-1)(n-1)}$	$X_{(m-1)n}$
X_{m1}	X_{m2}	·	$X_{m(n-1)}$	X_{mn}
Input	Input		Input	Input
Y_{11}	Y_{12}	·	$Y_{1(n-1)}$	Y_{1n}
Y_{21}	Y_{22}	·	$Y_{2(n-1)}$	Y_{2n}
Y_{31}	Y_{32}	·	$Y_{3(n-1)}$	Y_{3n}
·	·		·	·
·	·		·	·
·	·		·	·
$Y_{(s-2)1}$	$Y_{(s-2)2}$	·	$Y_{(s-2)(n-1)}$	$Y_{(s-2)n}$
$Y_{(s-1)1}$	$Y_{(s-1)2}$	·	$Y_{(s-1)(n-1)}$	$Y_{(s-1)n}$
Y_{s1}	Y_{s2}	·	$Y_{s(n-1)}$	Y_{sn}
Output	Output		Output	Output

In above table, information of n unit of firms is available. Now, the random state is explained. At first, the input variable X_{ij} is considered. This variable is random and has distribution function of $f_{ij}(x)$ or in discrete state, suppose there is a natural digit w for each X_{ij} such that (Table 2):

Table 2. Probability of system's input

		$P^{ij}{}_{1}\left(X^{1}{}_{ij}\right)=\alpha^{ij}1$
		$P^{ij}{}_{2}\left(X^{2}{}_{ij}\right)=\alpha^{ij}2$
	$\rightarrow$	. . .
		$P^{ij}{}_{(w-1)}\left(X^{(w-1)}{}_{ij}\right)=\alpha^{ij}(w-1)$
		$P^{ij}{}_{w}\left(X^{w}{}_{ij}\right)=\alpha^{ij}w$

Also, there is a natural digit z for each Y_{ij} such that (Table 3):

Table 3. Probability of system's Output

		$P^{ij}{}_{1}\left(Y^{1}{}_{ij}\right)=\gamma^{ij}1$
		$P^{ij}{}_{2}\left(Y^{2}{}_{ij}\right)=\gamma^{ij}2$
Y_{ij}	$\rightarrow$	. . .
		$P^{ij}{}_{(z-1)}\left(Y^{(z-1)}{}_{ij}\right)=\gamma^{ij}(z-1)$
		$P^{ij}{}_{z}\left(Y^{z}{}_{ij}\right)=\gamma^{ij}z$

After statement of the problem, these are solved. Now, we will explain 2 solutions for these problems based on input/output variables. The first one is to use mathematical expectation. In fact, we calculate the average value of these input/output random variables and them we will solve the problem and obtain the best DMU. The discrete state for each of X_{ij} is given as:

$$X_{ij} = \sum_{t=1}^{w} P(X_{ij}^t)X_{ij}^t = \sum_{t=1}^{w} \alpha_{ij}^t X_{ij}^t. \tag{6}$$

And for each Y_{ij} is given as:

$$Y_{ij} = \sum_{t=1}^{z} P\left(Y_{ij}^t\right) Y_{ij}^t = \sum_{t=1}^{z} \gamma^{ij} t * Y_{ij}^t. \tag{7}$$

In continuous state with distribution function $f_{ij}(x)$ for each X_{ij}, the random state would be certain as follow:

$$X_{ij} = \int x f_{ij}(x) dx. \tag{8}$$

Also for outputs, each would be as follow with its distribution function:

$$Y_{ij} = \int y f_{ij}(y) dy. \tag{9}$$

By this, the problem would be solves as reference model (abovementioned). In mathematical expectation model, we consider other solutions due to large volume of the calculations particularly in continuous form. The other approach considered for this problem is interval approach. In this structure, we consider the minimum and maximum possible for each input/output random variable. Then, using these value whose lower and upper bounds are minimum and maximum random variables, respectively, we will solve the problem by converting using the interval approach. In discrete finite state, for each X_{ij} the intervals are given as:

$$a_{ij} = Min\left\{X_{ij}^1, X_{ij}^2, X_{ij}^3, \ldots, X_{ij}^{(w-2)}, X_{ij}^{(w-1)}, X_{ij}^w\right\}, \tag{10}$$

$$b_{ij} = Max\left\{X_{ij}^1, X_{ij}^2, X_{ij}^3, \ldots, X_{ij}^{(w-2)}, X_{ij}^{(w-1)}, X_{ij}^w\right\}. \tag{11}$$

Now, we substitute X_{ij} with $\left[a_{ij}, b_{ij}\right]$. For Y_{ij}, we will have:

$$a'_{ij} = Min\left\{Y_{ij}^1, Y_{ij}^2, Y_{ij}^3, \ldots, Y_{ij}^{(z-2)}, Y_{ij}^{(z-1)}, Y_{ij}^z\right\}, \tag{12}$$

$$b'_{ij} = Max\left\{Y_{ij}^1, Y_{ij}^2, Y_{ij}^3, \ldots, Y_{ij}^{(z-2)}, Y_{ij}^{(z-1)}, Y_{ij}^z\right\}. \tag{13}$$

Here as above, Y_{ij} is substituted with $\left[a'_{ij}, b'_{ij}\right]$.

These intervals include all possible states and based on classical interval approach of data envelopment analysis, the efficiency obtained will be for all points. Theo other significant case is the random processes; i.e. each input/output follows a random variable. In this case, using the approaches related to each, the maximum and minimum would be obtained and the interval considered is determined. An approach for estimation of maximum and minimum proposed I that using historical data related to those processes (such as past prices of stocks which follows a particular process) and discretization of time, at first the interval considered is determined until the end and finding the maximum and minimum in each section of time and averaging them and the problem is solved. It is necessary to note that this is a proposed approach and one can use other approached to make intervals based on the nature and structure of process. Now, the tables (reference) include the intervals instead of random variables and using the model (reference) DMUs considered are obtained and the units are analyzed. In following, the numerical results are presented.

3.1 Numerical Result

Here, we defied several random paths (Brownian motion multiple-process absolute value) as input and output of each unit and then, based on above approach, we investigate the efficiency of each unit (Table 4).

Table 4. Estimate of input and output with end method of last part

$DMU\,1$		$DMU\,2$		$DMU\,3$	
[0.0001	1.0219]	[0.0014	0.6167]	[0.0046	1.4008]
[0.0009	1.0534]	[0.0012	0.7511]	[0.0111	1.7530]
Input	Input	Input	Input	Input	Input
[0.0014	1.4095]	[0.0011	0.6177]	[0.0001	0.9569]
Output	Output	Output	Output	Output	Output

After solving the obtained models, the efficiency interval $[E_k^L, E_k^U]$ is obtained. Based on this, there is no unit being efficient for all values available in interval. As well, units 1 and 3 would not be efficient for none of the values in these intervals.

4 Conclusion

Based on studies on random data envelopment analysis and considering the classical approaches, we solved the data envelopment analysis problem. In this model, we had random input and output. Then by defining the probable space and mathematical expectation, we proposed an approach for solving the random problem by changing it to certain one. Also, for solving these problems the interval approach was used in which the beginning and end of the maximum and minimum intervals was considered for variables and solved with classical interval approach. In further studies, one can analyze the sensitivity by investigating the intervals and their effect on solution.

Acknowledgements. Thanks to the support by Professor S. H. Nasseri.

Recommender: 2016 International workshop on Mathematics and Decision Science, Dr. Hadi Nasseri of University of Mazandaran in Iran.

References

1. Cooper, W.W., Seiford, L.M., Tone, K.: Data Envelopment Analysis. Kluwer Academic C Pulishers, Boston (2000)
2. Charnes, A., Cooper, W.W.: Chance-constrained programming. Manage. Sci. **6**, 73–79 (1959)
3. Land, K., Lovell, C.A.K., Thore, S.: Chance- constrained data envelopment analysis. Manag. Decis. Econ. **14**, 541–554 (1994)
4. Olesen, O.B., Petersen, N.C.: Chance constrained efficiency evaluation. Manage. Sci. **41**, 442–457 (1995)
5. Kwakernaak, H.: Fuzzy random variables. Part I: definitions and theorems. Inf. Sci. **15**(1), 1–29 (1978)
6. Jingliang, J., Dequn, Z., Peng, Z.: Measuring environmental performance with stochastic environmental DEA: the case of APEC economies. Econ. Model. **38**, 80–86 (2014)
7. Charnes, A., Cooper, W.W., Rhodes, E.: Measuring the efficiency of decision-making units. Eur. J. Oper. Res. **2**, 429–444 (1978)
8. Banker, R.D., Charnes, A., Cooper, W.W.: Some models for estimating technical and scale efficiencies in data envelopment analysis. Manage. Sci. **30**, 1078–1092 (1984)
9. Despotis, D.K., Smirlis, Y.G.: Data envelopment analysis with imprecise data. European J. Oper. Res. **140**, 24–36 (2002)
10. Ross, S.: A First Course in Probability. Pearson Prentice Hall, Boston (2010)