Available online at http://ijim.srbiau.ac.ir/
Int. J. Industrial Mathematics (ISSN 2008-5621)
Vol. 14, No. 2, 2022 Article ID IJIM-1469, 14 pages
DOI: http://dx.doi.org/10.30495/ijim.2022.19278
Research Article

Stochastic Efficiency Based on a Common Set of Weights in Data Envelopment Analysis

M. Ghasemi *, M. R. Mozaffari †‡, M. Rostamy Malkhalifeh §, M. H. Behzadi ¶

Received Date: 2020-10-29 Revised Date: 2021-03-13 Accepted Date: 2021-08-11

Abstract

Data envelopment analysis (DEA) is a method for evaluating the performance of a set of homogenous decision-making units (DMUs) with multiple inputs and multiple outputs. There are various models for efficiency calculation in DEA, one of which is based on determining a common set of weights (CSW), which is used extensively by DEA experts. In classical data envelopment analysis, all input and output values are clearly specified. However, this assumption might not always be true in real-world applications. One of the important methods for dealing with imprecise data is to explore the use of stochastic data in DEA. This manuscript extends the CSW model to stochastic inputs and outputs. Next, the stochastic CSW model is transformed into a nonlinear model, and then, the deterministic model is transformed into a quadratic programming model. The efficiency obtained using stochastic data is called stochastic efficiency. Finally, the concept presented in this article is demonstrated through a numerical example involving a number of Iranian banks.

Keywords : Data envelopment analysis; Stochastic efficiency; Common set of weights; Normal distribution; Quadratic Programming.

1 Introduction

Up to the year 1978, a large body of research was conducted on measuring the efficiency of DMUs in a system. The majority of these studies led to the development of non-parametric methods. Farrell [13] was one of the researchers trying to overcome these issues by introducing a production function, but he was unable to extend the method to multiple outputs. However, he introduced an efficiency measure that came to be famously known as Farrells measure of efficiency. Data envelopment analysis is an extension of Farrells idea with regard to efficiency calculation using a production function. Twenty years after Farrells distinguished accomplishment, Charnes, Cooper, and Rhodes [4] introduced a model based on previous works called the CCR model. The CCR model was presented as a single-objective

*Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran.
†Corresponding author. Mozaffari854@yahoo.com, Tel:+98(917)041338.
‡Department of Mathematics, Shiraz Branch, Islamic Azad University, Shiraz, Iran.
§Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran.
¶Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran Tehran, Iran.

model for calculating the relative efficiency of decision-making units with multiple inputs and multiple outputs. Later on, Banker, Charnes, and Cooper [2] developed the CCR model under the assumption of variable returns to scale (VRS) and called it the BCC model. To calculate the relative efficiency in DEA, the ratio of a weighted set of outputs to a weighted set of inputs is maximized. It is worth noting that to measure the efficiency of each unit, the model should be solved again with respective modifications. In this respect, different weights are obtained for the inputs and outputs of each DMU in each iteration of the model. Flexibility in weight assignment can be both considered a strength and a weakness of this method. The advantage is that if a DMU was found inefficient with this method, it will also be inefficient using any other method. The flaw to the method is that it allows any unit to assign input and output weights that would result in the maximum level of efficiency in comparison with other units, and hence, there is the concern that a unit might only be found efficient because of the assigned weights. In addition, when different weights are obtained, it gives rise to the question that which weight is more suitable. Due to these issues and the incapability of basic models in resolving them, calculating a common set of weights (CSW) became a method of interest. In the CSW method, the ratio of weighted outputs to weighted inputs is simultaneously maximized for all DMUs. Thompson et al. [27, 26], Cook et al. [6], Charnes et al. [5], and Roll et al.[23, 24] have studied some of the CSW concepts. Jahanshahloo et al. [14, 17] introduced a multi-objective model for efficiency measurement through the CSW method in one study, and using the infinity norm, proposed a nonlinear method for solving the CSW model in another. Furthermore, Cook and Zhu [7] utilized goal programming to solve the CSW model, and Davoodi and Rezai [12] suggested a linear method for the same purpose. Despite the many capabilities of input and output parameters in DEA, which have been mentioned by many researchers, there are major flaws to them as well. One of the most critical flaws is dependence on information from a time period in which the units under study have actually been operating. Therefore, the results provided as solutions by these models are based on past information. One way to resolve this issue is to develop a model that would provide the possibility of prediction by taking into account the incidence odds and applying the predicted values. One of the critical methods for dealing with imprecise data is the application of stochastic data in DEA.

Cooper et al. [8] were the first to introduce the concept of chance constrained programming. Hosseinzadeh-Lotfi et al. [21, 17] used the concept of stochastic data to rank DMUs and achieve centralized resource allocation. Azadeh et al. [1] used a stochastic DEA model to calculate the efficiency of electric power companies. Cooper et al. [11, 10, 9] and Huang and Lee [16] were also among researchers that utilized the concept of chance constrained programming. Khodabakhshi et al. [20] developed an additive model for estimating returns to scale in stochastic DEA. Furthermore, Khodabakhshi [19, 18] used this concept to develop input- and output-oriented super-efficiency models in DEA. There are various models of efficiency measurement in DEA that are extended to interval and fuzzy data, but there are no models suited to determine efficiency in the presence of stochastic data using a common set of weights. For instance, consider the numerous branches of a bank. In each branch, the personnel salaries, administrative costs, rental fees, and costs of movable properties form the inputs of the branch. Meanwhile, the deposits (including current accounts, savings accounts, short-term and long-term investment accounts, etc.), loans (including all loans extended to economic, industrial, mining, agricultural, housing, commercial, and service sectors), and interbank services (including incoming/outgoing wire transfers and checks) are the outputs of the branch. Numerous factors, such as macroeconomic and political factors, unexpected governmental decisions and their impact on the banking system, and other similar environmental factors, have a completely random and uncontrollable effect on the behavior of depositors and borrowers. Therefore, we

are faced with stochastic data in the efficiency evaluation of bank branches.
Given that stochastic inputs and outputs are observed in various areas, it seems necessary to extend the previous definitions and models to this type of data. In this paper, utilizing the concept of chance constrained programming, we first introduce a stochastic CSW model along with probability restrictions. Then, we transform the stochastic CSW model into a deterministic model, and following that, the deterministic model is transformed into a quadratic programming model. Determining a level of error would cause the results to have a probable nature, and increasing the level of error would reduce our confidence in results such as efficiency scores. The rest of this manuscript is organized as follows:
Section 2 presents the definitions and concepts relating efficiency and the CSW model, and a stochastic CSW model and its deterministic equivalent are proposed in section 3. In section 4, the application of the proposed model is demonstrated using a numerical example, and finally, our conclusions and suggestions for future research are presented in the end.

2 Background

Assume that we have n DMUs each consuming m inputs to produce s outputs. Also, let x_{ij} and y_{rj} denote the ith and rth values of DMU_j, respectively. The following shows the fractional model for calculating the relative efficiency of DMU_o, where $o \in \{1, \ldots, n\}$:

$$\theta_o = \max \frac{\sum_{r=1}^{s} u_r y_{ro}}{\sum_{i=1}^{m} v_i x_{io}}$$

$$s.t: \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, j = 1, \ldots, n \tag{2.1}$$

$$u_r, v_i \geq 0, i = 1, \ldots, m, r = 1, \ldots, s$$

u_r and v_i represent the relative importance of input and output vectors, respectively. Using the Charnes-Cooper transformation [3], the fractional model above is transformed into the following linear model, which is known as the multiplier CCR model:

$$\theta_o = \max \sum_{r=1}^{s} u_r y_{ro}$$

$$s.t: \sum_{i=1}^{m} v_i x_{i0} = 1$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, j = 1, \ldots, n \tag{2.2}$$

$$u_r, v_i \geq 0, i = 1, \ldots, m, r = 1, \ldots, s$$

To calculate the efficiency of each unit using the multiplier CCR model, the model should be iterated. By doing so, different input and output weights are produced for each DMU in each iteration of the model. To resolve this issue and determine a single weight for all DMUs, the CSW method can be used as follows [17]:

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

$$s.t: \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, j = 1, \ldots, n \tag{2.3}$$

$$u_r, v_i \geq 0, i = 1, \ldots, m, r = 1, \ldots, s$$

Various methods have been proposed for solving this multi-objective fractional model. Goal programming (GP) is one of these methods [25]. In goal programming, the decision maker specifies certain levels for achieving the objectives. Moreover, the decision maker allows deviation from the goals, and therefore, creates flexibility in the decision-making process. Also, the objective function seeks to minimize the undesirable deviations. Based on the GP method, Model (2.3) an be transformed into the following nonlinear model in order to identify a set of common weights:

$$\min \sum_{j=1}^{n} \left(s_j^- + s_j^+ \right)$$

Table 1: The average input-output data for 15 banks.

	Personnel scores	paid interests	received interests	loans	bank fees
DMU01	12.1657143	3042777312	4101560290	171017942	238815365.7
DMU02	18.6157143	4554095611	6218453905	440984913	1519490550
DMU03	18.1257143	5420614923	6323510346	420481191	2191018454
DMU04	13.1728571	2899451169	5109351561	475870573	1355977200
DMU05	11.3671429	3426429642	4821842767	196013953	328262665.9
DMU06	13.57	5120803446	7315158710	586040178	2127950910
DMU07	12.78	5780459280	5651786897	232289072	574769441.4
DMU08	8.45428571	3472026870	2740483351	255702740	1165573151
DMU09	15.0014286	4332428054	4107394280	249633733	766701171.3
DMU10	15.2585714	5000232886	6518731217	624622398	5457856304
DMU11	12.65	4370662601	5591460382	444093713	1370700809
DMU12	10.5314286	2526027536	3504779713	658320635	977485353
DMU13	9.82714286	3107371152	3332608236	131463968	164765249.7
DMU14	12.6314286	3389075306	3220094121	286234437	1240881842
DMU15	11.6128571	3619785956	4550116513	281302473	857893696.4

s.t:

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} + s_j^- - s_j^+ = A_j, j = 1, \dots, \quad (2.4)$$

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, j = 1, \dots, n$$

$$s_j^-, s_j^+ \geq 0, j = 1, \dots, n$$

$$u_r, v_i \geq \varepsilon, i = 1, \dots, m, r = 1, \dots, s$$

where A_j denotes the goal of the jth objective. s_j^- and s_j^+ are negative and positive deviations from the objective, respectively. Technically, the deviational variables s_j^- and s_j^+ help the objective function j to achieve the goal $A_j = 1$. Thereby, the positive deviation is equal to zero here, i.e. $s_j^+ = 0$. Thus, the first constraint can be reformulated as follows:

$$\sum_{r=1}^{s} u_r y_{rj} + s_j^- \sum_{i=1}^{m} v_i x_{ij} = \sum_{i=1}^{m} v_i x_{ij},$$

$$j = 1, \dots, n$$

Considering the nonlinear constraint above, Model (2.4) cannot be transformed into a linear model. In order to achieve the goal of DMU_j (efficiency scores of one), the numerator should increase in the fraction $\frac{\sum_{r=1}^{S} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}}$ while the denominator is decreased. Therefore, Model (2.3) can be reformulated as follows:

$$\min \sum_{j=1}^{n} \left(s_j^- + s_j^+ \right)$$

s.t:

$$\frac{\sum_{r=1}^{s} u_r y_{rj} + s_j^+}{\sum_{i=1}^{m} v_i x_{ij} - s_j^-} = 1, j = 1, \dots, n \quad (2.5)$$

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, j = 1, \dots, n$$

$$s_j^-, s_j^+ \geq 0, j = 1, \dots, n$$

$$u_r, v_i \geq \varepsilon, i = 1, \dots, m, r = 1, \dots, s$$

Obviously, based on the first constrai(2.5), the second constraint is redundant and can be removed from the model. Therefore, the model can

Table 2: Covariance of the inputs.

	Covariance	11	12
DMU01	11	28.1037619	2130753830
	12	2130753830	2.88E + 18
DMU02	11	343.2171619	86734091712
	12	86734091712	2.52E + 19
DMU03	11	141.6761619	24151949825
	12	24151949825	1.55E + 19
DMU04	11	16.95742381	4133633431
	12	4133633431	2.03E + 18
DMU05	11	37.35045714	20172661533
	12	20172661533	1.44E + 19
DMU06	11	13.0552	9714084720
	12	9714084720	2.51E + 19
DMU07	11	33.54252381	13930523397
	12	13930523397	1.16E + 19
DMU08	11	15.71349524	2819236832
	12	2819236832	3.06E + 18
DMU09	11	47.42264762	11860619917
	12	11860619917	5.27E + 18
DMU10	11	83.96704762	27107378091
	12	27107378091	1.21E + 19
DMU11	I1	15.47603333	8094851174
	12	8094851174	9.15E + 18
DMU12	11	13.54364762	4521995985
	12	4521995985	3.72E + 18
DMU13	11	10.14172381	1563242334
	12	1563242334	4.08E + 18
DMU14	11	43.92771429	−4296514482
	12	−4296514482	1.06E + 18
DMU15	11	17.61259048	4375450930
	12	4375450930	7.18E + 18

be formulated in the following linear form:

$$\min \sum_{j=1}^{n} \left(s_j^- + s_j^+\right)$$

s.t:

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + \left(s_j^+ + s_j^-\right) = 0 \qquad (2.6)$$

$$s_j^-, s_j^+ \geq 0, j = 1, \ldots, n$$

$$u_r, v_i \geq \varepsilon, i = 1, \ldots, m, r = 1, \ldots, s$$

By setting $s_j^- + s_j^+ = s_j$, the model is transformed as follows:

$$min \sum_{j=1}^{n} s_j$$

$$\text{s.t:} \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + s_j = 0 \qquad (2.7)$$

$$s_j \geq 0, j = 1, \ldots, n$$

$$u_r, v_i \geq \varepsilon, i = 1, \ldots, m, r = 1, \ldots, s$$

Model (2.7) can have multiple optimal solutions. In this case, using the model presented in [22], a unique optimal solution can be obtained. Model (2.7) can be reformulated as follows:

$$\text{Max} \sum_{j=1}^{n} \left(\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \right)$$

s.t:

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, j = 1, \ldots, n \qquad (2.8)$$

Table 3: Covariance of the outputs.

	Covariance	O1	O2	O3
3* DMU01	O1	3.97E + 18	1.58E + 17	2.24E + 17
	O2	1.58E + 17	7.54E + 15	1.10E + 16
	O3	2.24E + 17	1.10E + 16	6.72E + 16
3* DMU02	O1	5.74E + 19	5.21E + 18	2.40E + 19
	O2	5.21E + 18	5.18E + 17	2.44E + 18
	O3	2.40E + 19	2.44E + 18	1.16E + 19
3* DMU03	O1	1.42E + 19	5.34E + 17	5.82E + 18
	O2	5.34E + 17	1.21E + 17	7.70E + 17
	O3	5.82E + 18	7.70E + 17	6.81E + 18
3* DMU04	O1	1.25E + 19	9.48E + 17	7.59E + 18
	O2	9.48E + 17	2.44E + 17	1.30E + 18
	O3	7.59E + 18	1.30E + 18	8.35E + 18
3* DMU05	O1	4.48E + 19	8.04E + 17	3.40E + 18
	O2	8.04E + 17	2.41E + 16	6.21E + 16
	O3	3.40E + 18	6.21E + 16	2.84E + 17
3* DMU06	O1	1.54E + 20	1.16E + 19	4.96E + 19
	O2	1.16E + 19	9.02E + 17	3.88E + 18
	O3	4.96E + 19	3.88E + 18	1.68E + 19
3* DMU07	O1	2.27E + 19	9.08E + 16	1.65E + 17
	O2	9.08E + 16	2.57E + 16	1.39E + 17
	O3	1.65E + 17	1.39E + 17	9.42E + 17
3* DMU08	O1	4.40E+18	2.49E+17	3.65E+18
	O2	2.49E+17	4.03E+16	3.61E+17
	O3	3.65E+18	3.61E+17	4.86E+18
3* DMU09	O1	8.81E+18	2.00E+17	1.26E+18
	O2	2.00E+17	2.17E+16	8.64E+16
	O3	1.26E+18	8.64E+16	5.48E+17
3* DMU10	O1	5.07E+19	5.44E+18	7.73E+19
	O2	5.44E+18	6.76E+17	9.00E+18
	O3	7.73E+19	9.00E+18	1.29E+20
3* DMU11	O1	1.00E+19	2.09E+17	-5.41E+17
	O2	2.09E+17	6.52E+16	4.34E+17
	O3	-5.41E+17	4.34E+17	3.62E+18
3* DMU12	O1	8.39E+18	3.33E+18	5.91E+18
	O2	3.33E+18	1.75E+18	3.07E+18
	O3	5.91E+18	3.07E+18	5.38E+18
3* DMU13	O1	4.67E+18	-1.05E+16	-1.92E+17
	O2	-1.05E+16	2.68E+15	2.57E+15
	O3	-1.92E+17	2.57E+15	3.37E+16
3* DMU14	O1	8.25E+18	4.18E+17	7.18E+17
	O2	4.18E+17	7.97E+16	5.45E+17
	O3	7.18E+17	5.45E+17	4.64E+18
3* DMU15	O1	1.38E+19	5.12E+17	7.02E+1
	O2	5.12E+17	7.62E+16	4.30E+17
	O3	7.02E+17	4.30E+17	3.04E+18

$$u_r, v_i \geq \varepsilon, \quad r = 1, 2, \ldots, s; i = 1, 2, \ldots, m$$

If we let (u_r^*, v_i^*) be the optimal solution of this model, the CSW-efficiency of DMU_j are calculated as follows:

$$\theta_j^* = \frac{\sum_{r=1}^{s} u_r^* y_{rj}}{\sum_{i=1}^{m} v_i^* x_{ij}}$$

Definition 2.1. *DMU_j is CWA-efficient if* $\theta_j^* =$

1. *Otherwise, DMU_j is CWA-inefficient.*

3 Efficiency measurement using the CSW model with stochastic inputs and outputs

Assume that $\tilde{X}_j=(\tilde{x}_{1j},\ldots,\tilde{x}_{mj})^T$ and $\tilde{Y}_j=(\tilde{y}_{1j},\ldots,\tilde{y}_{sj})^T$ are the stochastic input and output vectors of $DMU_j(j=1,\ldots,n)$, respectively. Moreover, let $\mathrm{X}_j=(x_{1j},\ldots,x_{mj})^T$ and $\mathrm{Y}_j=(y_{1j},\ldots,y_{sj})^T$ be the expected value vectors for the inputs and outputs of DMU_j, respectively. Also, assume that the inputs and outputs have normal distribution as follows:

$$\tilde{x}_{ij}\sim N\left(x_{ij},\sigma_{ij}^2\right),\quad \tilde{y}_{rj}\sim N\left(y_{rj},\sigma_{rj}^2\right)$$

Thus, the stochastic version of Model (2.8) with probability constraints would be as follows:

$$\operatorname{Max} E\left(\sum_{j=1}^{n}\left(\sum_{r=1}^{s}u_r\tilde{y}_{rj}-\sum_{i=1}^{m}v_i\tilde{x}_{ij}\right)\right)$$

s.t:

$$P\left(\sum_{r=1}^{s}u_r\tilde{y}_{rj}-\sum_{i=1}^{m}v_i\tilde{x}_{ij}\leq 0\right)\geq 1-\alpha,\qquad (3.9)$$

$$j=1,\ldots,n$$

$$u_r,v_i\geq\varepsilon,\quad r=1,2,\ldots,s;i=1,2,\ldots,m$$

where P stands for probability and the level of error, which is determined by the manager, lies between zero and one. If we let (u_r^*,v_i^*) be the optimal solution of the model above, the efficiency scores of DMU_j will be calculated as follows:

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

Stochastic efficiency can be defined as follows through Model (3.9).

Definition 3.1. *If $\theta_o^*=1$, DMU_o is stochastically efficient, and if $\theta_o^*<1$, DMU_o is stochastically inefficient.*

3.1 Deterministic equivalent of the stochastic CSW model

Now, we formulate the deterministic equivalent of Model (3.9). The objective function of the model is transformed into a deterministic expression using an expected value as follows:

$$E\left(\sum_{j=1}^{n}\left(\sum_{r=1}^{s}u_r\tilde{y}_{rj}-\sum_{i=1}^{m}v_i\tilde{x}_{ij}\right)\right)=$$

$$\sum_{j=1}^{n}\sum_{r=1}^{s}u_ry_{rj}-\sum_{j=1}^{n}\sum_{i=1}^{m}v_ix_{ij}$$

Based on chance-constrained programming approaches with probability constraints [8], we obtain the deterministic form of the models stochastic constraint. For this purpose, consider the probability constraint of unit j :

$$P\left(\sum_{r=1}^{s}u_r\tilde{y}_{rj}-\sum_{i=1}^{m}v_i\tilde{x}_{ij}\leq 0\right)\geq 1-\alpha,$$

By defining the slack variable ε_j, the inequality above will be transformed into the following equation:

$$P\left(\sum_{r=1}^{s}u_r\tilde{y}_{rj}-\sum_{i=1}^{m}v_i\tilde{x}_{ij}\leq 0\right)=1-\alpha+\varepsilon_j,$$

$$j=1,\ldots,n$$

Remark 3.1. *Let X be a random variable and a, b and c constant numbers, if $P(X\leq a)=c$ and $b\leq a$ then there exists $d\leq c$ such that $P(X\leq b)=d$.*

Using the point described above, there exists a non-negative variable s_j, such that:

$$P\left(\sum_{r=1}^{s}u_r\tilde{y}_{rj}-\sum_{i=1}^{m}v_i\tilde{x}_{ij}\leq -s_j\right)=1-\alpha,\qquad (3.10)$$

$$j=1,\ldots,n$$

We define:

$$\tilde{h}_j=\sum_{r=1}^{s}u_r\tilde{y}_{rj}-\sum_{i=1}^{m}v_i\tilde{x}_{ij}$$

Table 4: Covariance of the inputs with outputs.

		O1	O2	O3
2* DMU01	L1	1719671889	1.27E+18	1719671889
	L2	99614471.18	6.95E+16	99614471.18
2* DMU02	L1	924243294.5	2.24E+17	924243294.5
	L2	1.38E+11	3.65E+19	137947000000
2* DMU03	L1	12350825587	3.54E+18	12350825587
	L2	56936293137	1.68E+19	56936293137
2* DMU04	L1	32129081261	4.49E+18	32129081261
	L2	2201601751	1.19E+18	2201601751
2* DMU05	L1	12278464628	8.49E+18	12278464628
	L2	10430640443	3.67E+18	10430640443
2* DMU06	L1	1175340556	4.14E+17	1175340556
	L2	8523015768	2.58E+18	8523015768
2* DMU07	L1	29019474287	2.38E+19	29019474287
	L2	555764038.4	4.11E+17	555764038.4
2* DMU08	L1	2786719373	1.94E+18	2786719373
	L2	17526781593	5.52E+19	17526781593
2* DMU09	L1	1763118129	4.33E+18	1763118129
	L2	7624782536	1.83E+19	7624782536
2* DMU10	L1	27050849785	1.28E+19	27050849785
	L2	41068395.33	2.77E+17	41068395.33
2* DMU11	L1	-251693086.1	1.96E+18	-251693086.1
	L2	1766141441	7.51E+17	1766141441
2* DMU12	L1	276551928.8	2.28E+17	276551928.8
	R3	-569067395.9	1.61E+18	-569067395.9
2* DMU13	L2	14436066250	6.39E+18	14436066250
	L1	669579880.9	1.85E+17	669579880.9
2* DMU14	L1	4721358438	1.06E+18	4721358438
	L2	52782510218	2.36E+19	52782510218
2* DMU15	L1	6921174483	2.79E+18	6921174483
	L2	93220772175	3.88E+19	93220772175

Since any linear combination of normally distributed stochastic variables has a normal distribution itself, we therefore have:

$$\tilde{h}_j \sim N\left(h_j, \sigma_j^2(u,v)\right)$$

$$h_j = E\left(\tilde{h}_j\right) = E\left(\sum_{r=1}^{s} u_r \tilde{y}_{rj} - \sum_{i=1}^{m} v_i \tilde{x}_{ij}\right) = \sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij}$$

$$\begin{aligned}
{\sigma_j}^2(u,v) &= \operatorname{Var}\left(\tilde{h}_j\right) \\
&= \operatorname{Var}\left(\sum_{r=1}^{s} u_r \tilde{y}_{rj} - \sum_{i=1}^{m} v_i \tilde{x}_{ij}\right) \\
&= \operatorname{Var}\left(\sum_{r=1}^{s} u_r \tilde{y}_{rj}\right) + \operatorname{Var}\left(\sum_{i=1}^{m} v_i \tilde{x}_{ij}\right) - 2\operatorname{cov}\left(\sum_{r=1}^{s} u_r \tilde{y}_{rj}, \sum_{i=1}^{m} v_i \tilde{x}_{ij}\right) \\
&= \operatorname{Var}\left(\sum_{r=1}^{s} u_r \tilde{y}_{rj}\right) + \operatorname{Var}\left(\sum_{i=1}^{m} v_i \tilde{x}_{ij}\right) - 2\sum_{r=1}^{s}\sum_{i=1}^{m} u_r v_i \operatorname{Cov}\left(\tilde{y}_{rj}, \tilde{x}_{ij}\right) \\
&= \sum_{k=1}^{s}\sum_{r=1}^{s} u_r u_k \operatorname{Cov}\left(\tilde{y}_{rj}, \tilde{y}_{kj}\right) + \sum_{k=1}^{m}\sum_{i=1}^{m} v_i v_k \operatorname{Cov}\left(\tilde{x}_{ij}, \tilde{x}_{kj}\right) - 2\sum_{r=1}^{s}\sum_{i=1}^{m} u_r v_i \operatorname{Cov}\left(\tilde{y}_{rj}, \tilde{x}_{ij}\right)
\end{aligned}$$

Taking into account the stochastic variable $\tilde{h}_j$, Equation (3.10) is reformulated as follows:

$$P\left(\tilde{h}_j \leq -s_j\right) = 1-\alpha, j = 1,\dots,n$$

$$P\left(\frac{\tilde{h}_j - h_j}{\sigma_j(u,v)} \leq \frac{-h_j - s_j}{\sigma_j(u,v)}\right) = 1-\alpha,$$
$$j = 1,\dots,n$$

On the other hand, by setting $\tilde{Z}_j = \frac{\tilde{h}_j - h_j}{\sigma_j(u,v)}$ and taking into account the fact that $\tilde{Z}_j$ has a standard normal distribution, we have:

$$P\left(\tilde{Z}_j \leq \frac{-h_j - s_j}{\sigma_j(u,v)}\right) = 1-\alpha$$

$$P\left(\tilde{Z}_j \leq \frac{h_j + s_j}{\sigma_j(u,v)}\right) = \alpha$$

$$\phi\left(\frac{h_j + s_j}{\sigma_j(u,v)}\right) = \alpha \rightarrow \frac{h_j + s_j}{\sigma_j(u,v)} = \phi^{-1}(\alpha)$$

Therefore, the deterministic for of the probability constraint will be as follows:

$$h_j + s_j - \sigma_j(u,v)\phi^{-1}(\alpha) = 0$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + s_j - \sigma_j(u,v)\phi^{-1}(\alpha) = 0$$

The deterministic form of Model (3.9) with probability restrictions is as follows:

$$\text{Max} \left(\sum_{j=1}^{n}\sum_{r=1}^{s} u_r y_{rj} - \sum_{j=1}^{n}\sum_{i=1}^{m} v_i x_{ij}\right)$$

s.t:

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + s_j - \sigma_j(u,v)\phi^{-1}(\alpha) = 0, \tag{3.11}$$
$$u_r, v_i \geq \varepsilon, r = 1,2,\dots,s; i = 1,2,\dots,m$$
$$s_j \geq 0, j = 1,\dots,n$$

Now, defining the non-negative variable λ_j, Model (3.11) can be transformed into a quadratic programming model:

$$\text{Max} \left(\sum_{j=1}^{n}\sum_{r=1}^{s} u_r y_{rj} - \sum_{j=1}^{n}\sum_{i=1}^{m} v_i x_{ij}\right)$$

s.t:

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} + s_j - \lambda_j\phi^{-1}(\alpha) = 0, \forall j \tag{3.12}$$

$$\begin{aligned}\lambda_j^2 = & \sum_{k=1}^{s}\sum_{r=1}^{s} u_r u_k \operatorname{Cov}\left(\tilde{y}_{rj}, \tilde{y}_{kj}\right) \\ & + \sum_{k=1}^{m}\sum_{i=1}^{m} v_i v_k \operatorname{Cov}\left(\tilde{x}_{ij}, \tilde{x}_{kj}\right) \\ & - 2\sum_{r=1}^{s}\sum_{i=1}^{m} u_r v_i \operatorname{Cov}\left(\tilde{y}_{rj}, \tilde{x}_{ij}\right)\end{aligned}$$

$$u_r, v_i \geq \varepsilon, s_j, \lambda_j \geq 0,$$
$$r = 1,\dots,s; i = 1,\dots,m, j = 1\dots,n$$

Theorem 3.1. *Model (3.12) is always feasible for any level of error α.*

Proof. Put

$$u_r = \varepsilon \quad r = 1,\dots,s$$
$$v_i = \varepsilon \quad i = 1,\dots,m$$

We get λ_j from the second constraint. From the first constraint, we obtain s_j as follows:

$$s_j = \lambda_j\phi^{-1}(\alpha) - \sum_{r=1}^{s} u_r y_{rj} + \sum_{i=1}^{m} v_i x_{ij} \quad \forall j$$

Therefore, the model is always feasible for any level of error α. □

Theorem 3.2. *For any $0 \leq \alpha \leq 0.5$, the optimal solution of Model (3.12) is always a number between zero and one.*

Proof. Since $\alpha \leq 0.5$, we therefore have: $\phi^{-1}(\alpha) \leq 0$ and $s_j, \lambda_j \geq 0$. Thus,

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0 \rightarrow$$

$$\sum_{r=1}^{s} u_r y_{rj} \leq \sum_{i=1}^{m} v_i x_{ij} \rightarrow$$

$$\frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1$$

and the proof is complete. □

Table 5: Stochastic efficiency scores.

	$\alpha = 0.5$	$\alpha = 0.4$	$\alpha = 0.3$	$\alpha = 0.2$	$\alpha = 0.1$
DMU01	0.588	0.498	0.443	0.403	0.301
DMU02	0.713	0.604	0.537	0.488	0.373
DMU03	0.654	0.554	0.493	0.448	0.373
DMU04	0.95	0.805	0.716	0.65	0.48
DMU05	0.619	0.524	0.466	0.424	0.345
DMU06	0.777	0.658	0.585	0.532	0.493
DMU07	0.443	0.376	0.334	0.304	0.289
DMU08	0.476	0.403	0.358	0.326	0.314
DMU09	0.469	0.398	0.354	0.321	0.262
DMU10	1	0.847	0.753	0.685	0.603
DMU11	0.672	0.57	0.507	0.46	0.412
DMU12	0.808	0.684	0.608	0.553	0.457
DMU13	0.463	0.393	0.349	0.317	0.263
DMU14	0.556	0.471	0.419	0.381	0.308
DMU15	0.624	0.528	0.47	0.427	0.361

If the level of error becomes greater than 0.5, the theorem above will not hold. We will now discuss the matter through the following example.

Example 3.1. *Assume that there are two DMUs each having one input and one output, such that*

$$x_{11} \sim N(1,1), \quad y_{11} \sim N(4,1),$$

$$x_{12} \sim N(2,1), \quad y_{12} \sim N(6,1)$$

In the evaluation of DMU_2, the model has the following restrictions:

$$4u_1 - v_1 + s_1 - \phi^{-1}(\alpha)\lambda_1 = 0$$

$$6u_2 - 2v_1 + s_2 - \phi^{-1}(\alpha)\lambda_2 = 0$$

Let the inputs and outputs of this DMU be uncorrelated; therefore, we have:

$$\lambda_1^2 = u_1^2 + v_1^2$$

$$\lambda_2^2 = u_1^2 + v_1^2$$

If $\alpha = 0.97$, then $\phi^{-1}(\alpha) = 2$; thereby, the solution

$$u_1 = 3, v_1 = 4, s_1 = 2, s_2 = 0$$

will apply to the above mentioned restrictions and will be a feasible solution for the model in the evaluation of DMU_2, given that it has an efficiency score of

$$\frac{u_1 y_{12}}{v_1 x_{12}} = \frac{18}{8} > 1$$

This is due to the fact that the considered level of error is greater than 0.5.

Theorem 3.3. *For any level of error $\alpha \leq \alpha'$ in Model (3.12), we have: $Z^*(\alpha) \leq Z^*(\alpha')$.*

Proof. Let (Z^*, U^*, V^*) be the optimal solution of Model (3.12) in the evaluation of DMU_p at the level of error α'. Since $\phi^{-1}(\alpha)$ is an increasing function, the inequality $\phi^{-1}(\alpha) \leq \phi^{-1}(\alpha')$ holds. Hence, we have:

$$\begin{aligned} 0 &\geq \sum_{r=1}^{s} u_r^* y_{rj} - \sum_{i=1}^{m} v_i^* x_{ij} - \lambda_j \phi^{-1}(\alpha) \\ &\geq \sum_{r=1}^{s} u_r^* y_{rj} - \sum_{i=1}^{m} v_i^* x_{ij} - \lambda_j \phi^{-1}(\alpha') \end{aligned}$$

The expression above shows that (Z^*, U^*, V^*) is a feasible solution for Model (3.12) in the evaluation of DMU_p at the level of error α'. Given that the model is a maximization model, the theorem holds. □

4 Numerical Example

In this section, through an applied example, we explore the application of the proposed stochastic CSW model with probability constraints. Data are related to 15 Iranian banks between the years 2010-2016.

In this respect, the personnel scores (weighted set of personnel demographics including employee count, education level, work experience, etc.) and paid interests (to all bank accounts) are considered as inputs and received interests, loans, bank fees, and other resources are parameters regarded as outputs. Table 1 presents the data related to these 15 banks expressed as means. To calculate stochastic efficiency using Model (3.12), we require the input covariance, output covariance, and input-output covariance. All the covariance values were calculated using EXCEL. Table 2 two provides the input covariance values. Table 3 pertains to output covariance values. Table 4 shows the input-output covariance values.

The efficiency of each bank was calculated for different values using the stochastic CSW model (3.12). The results are presented in Table 5. The values of $\varphi^{-1}(\alpha)$ equal $0, -0.25, -0.52, -0.84$, and -1.28 for $\alpha = 0.5, \alpha = 0.4, \alpha = 0.3$, $\alpha = 0.2$, and $\alpha = 0.1$. For different α values, different efficiency scores are obtained for each DMU. The results are provided in columns two to six of Table 5. To analyze the results better, a related diagram was drawn using Matlab software. The resulting diagram is presented in fig. 1. The

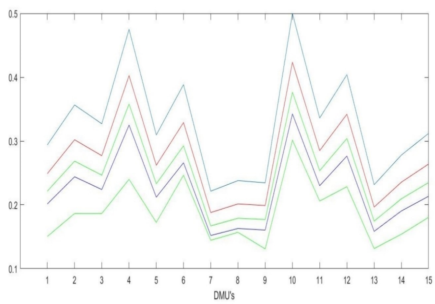

Figure 1

diagram above illustrates the efficiency of each DMU based on various α values. The horizontal axis is related to the DMUs and the vertical axis represents different α values. The highest efficiency scores pertain to DMUs $10, 4, 12$ and 6, in that order. The only efficiency score of one belongs to DMU 10 at $\alpha = 0.5$, and there are no other efficiency scores of one for other α values. The lowest efficiency scores are related to DMUs $7, 9, 13$, and 8 . For $\alpha = 0.1$, DMU 9 has a lower efficiency than DMU 7, while at other levels of error, DMU 7 is less efficient than DMU 9. Thereby, the highest and lowest efficiency scores pertain to DMU 10 at $\alpha = 0.5$ and DMU 7 at $\alpha = 0.1$, respectively. A reduction in the α value will result in a lower efficiency score, because the lower the value of α, the larger the number of constraints that will hold, and thus, the smaller our feasible region. Therefore, since we are dealing with a maximization problem, there will be poorer optimal solutions with every stage. Hence, there will be lower efficiency scores as the α value is reduced. As can be observed in the diagram, with increased α values, the efficiency scores increase as well. Therefore, efficiency scores are dependent on the level of error (α). The level of error α, which is predetermined by the manager, indicates the level by which the problem constraints do not hold. Thereby, any changes in this level would lead to different results. If $\alpha = 0.5$, then $\varphi^{-1}(\alpha) = 0$. Thus, the results obtained in stochastic DEA with this level of error are similar to the results obtained in DEA with deterministic data.

5 Conclusion

In many applied problems, managers are faced data that are imprecise and stochastic. The present study discussed efficiency measurement using a common set of weights in stochastic data envelopment analysis. In this regard, we extended the CSW model to cases with stochastic data and obtained its deterministic equivalent, which can be transformed into a quadratic programming model. Furthermore, by determining a level of errorα , we considered a probability of incidence for unexpected situations. This level of error should be specified by the manager from the start based on his or her sensitivity toward the results. The produced results will be dependent on the level of error and any changes in this level would lead to different results. As an empirical example, we calculated the efficiency of a number of Iranian banks. Results showed that among the units under study, unit 10 had the highest efficiency for all values ofα and unit 7 had the lowest efficiency for all levels of error except$\alpha = 0.1$. The stochastic model allowed a level of error for the data; therefore, in cases where the data are relatively imprecise and an approximate estimation is required, the stochastic model would be preferred. Nonetheless, in case of having precise data, the deterministic model will be more suitable. In this study, the data were assumed to be normally distributed. For future research, we recommend considering other forms of statistical distribution. Moreover, in our model, the line pertaining to the objective function was under the influence of mean values. In this regard, a model can be presented in the future in which the objective function is a function of the mean and variance of data.

References

[1] A. Azadeh, S. Motevali Haghighi, M. Zarrin, S. Khaefi, Performance evaluation of Iranian electricity distribution units by using stochastic data envelopment analysis, *International Journal of Electrical Power & Energy Systems* 73 (2015) 919-931.

[2] R. D. Banker, A. Charnes, W. W. Cooper, Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, *Management Science* 30 (1984) 1078-1092.

[3] A. Charnes, W. W. Cooper, Programming with linear fractional functionals, *Naval Research logistics quarterly* 9 (1962) 181-186.

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

[5] A. Charnes, W. Cooper, Q. L. Wei, Z. Huang, Cone ratio data envelopment analysis and multi-objective programming, *International Journal of Systems Science* 20 (1989) 1099-1118.

[6] W. D. Cook, Y. Roll, A. Kazakov, A Dea Model For Measuring The Relative Eeficiency Of Highway Maintenance Patrols, *INFOR: Information Systems and Operational Research* 28 (1990) 113-124.

[7] W. D. Cook, J. Zhu, Within-group common weights in DEA: An analysis of power plant efficiency, *European Journal of Operational Research* 178 (2007) 207-216.

[8] W. W. Cooper, H. Deng, Z. Huang, S. X. Li, Chance constrained programming approaches to congestion in stochastic data envelopment analysis, *European Journal of Operational Research* 155 (2004) 487-501.

[9] W. W. Cooper, H. Deng, Z. Huang, S. X. Li, Chance constrained programming approaches to technical efficiencies and inefficiencies in stochastic data envelopment analysis, *Journal of the Operational Research Society* 53 (2002) 1347-1356.

[10] W. W. Cooper, Z. Huang, V. Lelas, S. X. Li, O. B. Olesen, Chance constrained programming formulations for stochastic characterizations of efficiency and dominance in DEA, *Journal of Productivity Analysis* 9 (1998) 53-79.

[11] W. W. Cooper, Z. Huang, S. X. Li, Satisficing DEA models under chance constraints, *Annals of operations research* 66 (1996) 279-295.

[12] A. Davoodi, H. Z. Rezai, Common set of weights in data envelopment analysis: a linear programming problem, *Central European Journal of Operations Research* 20 (2012) 355-365.

[13] M. J. Farrell, The Measurement of Productive Efficiency, *Journal of the Royal Statistical Society. Series A (General)* 120 (1957) 253-290.

[14] F. Hosseinzadeh Lot, G. R. Jahanshahloo, A. Memariani, A Method for Finding Common Set of Weights by Multiple Objective programming in Data Envelopment Analysis, *Southwest Journal of Pure and Applied Mathematics* 1 (2000) 44-54.

[15] F. Hosseinzadeh Lotfi, N. Nematollahi, M. H. Behzadi, M. Mirbolouki, Z. Moghaddas, Centralized resource allocation with stochastic data, *Journal of Computational and Applied Mathematics* 236 (2012) 1783-1788.

[16] Z. Huang, S. X. Li, Dominance stochastic models in data envelopment analysis, *European Journal of Operational Research* 95 (1996) 390-403.

[17] G. R. Jahanshahloo, A. Memariani, F. H. Lotfi, H. Z. Rezai, A note on some of DEA models and finding efficiency and complete ranking using common set of weights, *Applied Mathematics and Computation* 166 (2005) 265-281.

[18] M. Khodabakhshi, An output oriented super-efficiency measure in stochastic data envelopment analysis: Considering Iranian electricity distribution companies, *Computers & Industrial Engineering* 58 (2010) 663-671.

[19] M. Khodabakhshi, Super-efficiency in stochastic data envelopment analysis: An input relaxation approach, *Journal of Computational and Applied Mathematics* 235 (2011) 4576-4588.

[20] M. Khodabakhshi, Y. Gholami, H. Kheirollahi, An additive model approach for estimating returns to scale in imprecise data envelopment analysis, *Applied Mathematical Modelling* 34 (2010) 1247-1257.

[21] F. H. Lotfi, N. Nematollahi, M. Behzadi, M. Mirbolouki, Ranking decision making units with stochastic data by using coefficient of variation, *Mathematical and Computational Applications* 15 (2010) 148-155.

[22] S. Ramezani-Tarkhorani, M. Khodabakhshi, S. Mehrabian, F. Nuri-Bahmani, Ranking decision-making units using common weights in DEA, *Applied Mathematical Modelling* 38 (2014) 3890-3896.

[23] Y. Roll, W. D. Cook, B. Golany, Controlling factor weights in data envelopment analysis, *IIE transactions* 23 (1991) 2-9.

[24] Y. Roll, B. Golany, Alternate methods of treating factor weights in DEA, *OMEGA-OXFORD-PERGAMON PRESS* 21 (1993) 99-99.

[25] M. Tamiz, D. Jones, C. Romero, Goal programming for decision making: An overview of the current state-of-the-art, *European Journal of Operational Research* 111 (1998) 569-581.

[26] R. Thompson, W. Lesso, R. R. Pettit, C. Samson, L. H. Sterm, F. D. Singleton, B. A. Smith, R. M. Thrall, Systems Task Force Comparative Site Evaluations for the Texas SSC Project: vol, E and Appendices, *Hoston Area Research Center, Woodlands, TX.* 1985.

[27] R. G. Thompson, F. Singleton Jr, R. M. Thrall, B. A. Smith, Comparative site evaluations for locating a high-energy physics lab in Texas, *Interfaces* 16 (1986) 35-49.

Marzieh Ghasemi was born in1981 in Lamerd, Iran. She received bachelors degree in Applied Mathematics from Shiraz University in 2003, and Master degree in optimal control from Isfahan University in 2006. She is currently a PhD student in Operations Research at the Islamic Azad University, Science and Research Branch. Her main interest areas include data envelopment analysis and optimal control.

Mohammad Reza Mozaffari was born in1978 in Firoozabad, Iran. He got his bachelors degree in Applied Mathematics from Shiraz University in the year 2000. In2009, he managed to obtain his PhD in Applied Mathematics with a specialization in operaetions research at the Islamic Azad University, Branch of Science and Research. Dr Mozaffari has showed much interest in researching related to DEA and Multi-Objective Programming. As a scientific board member, he currently works in the Department of Mathematics at the Islamic Azad University of Shiraz. Mohammad Reza Mozaffari is the corresponding author can be contacted at: Mozaffari854@yahoo.com

Mohsen Rostamy- Malkhalifeh is an Associate professor at the Applied Mathematics Operations Research group in Islamic Azad University in Science and Research Branch in Tehran, Iran. His research interests include performance management, efficiency, Fuzzy and especially those based on the broad set of methods known as Data Envelopment Analysis (DEA). Dr. Rostamy- Malkhalifeh research studies are published in such journals as Applied Mathematics and Computation, International Journal of Fuzzy Systems, International Journal of Applied Decision Sciences, European Journal of Operational.

Mohammad Hasan Behzadi was born in1979 in Iran. He received the B. S degree in statistics from Razi University of Iran in 2002, and M. S degree in pure statistics from Amirkabir University of Technology of Iran in 2004, Ph. D degree in applied statistics from Islamic Azad University, Branch of Science and Research in 2009, respectively. He is currently an Associate Professor in department of statistics, Science and Research, Islamic Azad University, Tehran, Iran. He is researching in fuzzy regression, Random optimization, Statistical inference. He has published some paper in this area.