Developing a Stochastic Input Oriented Data Envelopment Analysis (SIODEA) Model

Basma E. El-Demerdash

Teaching Assistant,
Department of Operations
Research and Decision Support,
Faculty of Computers and
Information, Cairo University,
Egypt

Ihab A. El-Khodary

Associate Professor,
Department of Operations
Researchand Decision Support,
Faculty of Computers and
Information, Cairo University,
Egypt

Assem A. Tharwat

Professor, Dean of the Higher
Canadian International College for
Engineering and Business,
Canadian International College,
Egypt

Abstract—Data Envelopment Analysis (DEA) is a powerful quantitative tool that provides a means to obtain useful information about efficiency and performance of firms, organizations, and all sorts of functionally similar, relatively autonomous operating units, known as Decision Making Units (DMU). Usually the investigated DMUs are characterized by a vector of multiple inputs and multiple outputs. Unfortunately, not all inputs and/or outputs are deterministic; some could be stochastic. The main concern in this paper is to develop an algorithm to help any organization for evaluating their performance given that some inputs are stochastic. The developed algorithm is for a Stochastic Input Oriented Model based on the Chance Constrained Programming, where the stochastic inputs are normally distributed, while the remaining inputs and all outputs are deterministic.

Keywords—Data Envelopment Analysis; Stochastic Variables; Input Oriented; Performance Measure; Efficiency Measurement.

I. INTRODUCTION

Measuring the efficiency of organizations has become an important and appealing research area in recent years. Many public organizations depend on their income from public funds, thus making it essential in interest of accountability, to measure the efficiency of such institutions. These organizations are "non-profit"; thus there is an absence of output and input prices while producing multiple outputs from multiple inputs. This poses a challenge in measuring their efficiencies.

An assortment of methodological approaches has been employed in an effort to resolve the problem of efficiency measurement in this context. These include the deterministic frontier approach, the stochastic frontier approach and the mathematical programming approach. The latter approach differs from both statistical frontier approaches in that it is fundamentally non-parametric and from the stochastic frontier approach in that it is non-stochastic.

The mathematical programming approach differs from both statistical frontier approaches in that it is fundamentally non-parametric, and from the stochastic frontier approach in that is non-stochastic. There are a number of benefits implicit in the programming approach that makes it attractive on a theoretical level. Simulation studies have indicated that the piecewise linear production frontier formulated by DEA is generally more flexible in approximating the true production frontier than even the most flexible parametric function form [1].

DEA as originally proposed by A. Charnes, et al [2] is a non-parametric frontier estimation methodology based on linear programming for measuring relative efficiencies and performance of a collection of related comparable entities called Decision Making Units (DMUs) which transform multiple inputs into multiple outputs. The objective of a DEA study is to assess the efficiency of each DMU in relation to its peers. The result of a DEA study is a classification of all DMUs as either "efficient" or "inefficient". Not only classifying the entities, but also determining the source of inefficiency (input) and the corresponding level or amount to enhance the performance. Unfortunately, the available DEA models and most of the previous DEA applications considering input variables of deterministic nature although some might have a random nature. Therefore in this study, we are looking forward to apply a Stochastic DEA since some of our input parameters have a random nature. Accordingly, we are proposing a modification to the existing stochastic DEA model which is chance constrained output oriented in order for the DEA model to be chance constrained input oriented.

The rest of the paper is organized as follows. The coming section discusses the methodology of the general DEA model and that for the stochastic DEA model. The third section includes the proposed stochastic chance constrained input oriented DEA model, the proposed solving algorithm, and a hypothetical illustrative example. The paper will end with the customary conclusions and implications for the future.

II. GENERAL MODEL OF DEA

The basic DEA model for 'n' DMUs with 'm' inputs and 's' outputs was first proposed by A. Charnes, et al [2]. The model determines the relative efficiency score for the different DMUs. The model depends on maximizing a production function estimated by DEA. This function is a deterministic frontier. For any inputs, the value of the DEA estimate defines the maximum output producible from inputs under all circumstances. On the other hand, for any outputs, the value of the DEA estimate defines the minimum input producing a

given output under all circumstances. In this sense, it is comparable to the parametric frontier with one-sided deviations estimated using mathematical programming methods.

According to the assumptions relating the change in outputs as a result of the change in inputs, the DEA model can be classified as having either constant returns to scale (CRS) or variable returns to scale (VRS). Under CRS models the outputs are not affected by the size of the DMU, rather they change in direct proportion to the change in inputs assuming that the scale of operation does not influence efficiency; therefore, in the CRS models the output and input oriented measures of efficiency are equal. Under VRS models, changes in outputs are not necessarily proportional to the changes in the inputs; therefore In the VRS models the output and input oriented measures of efficiency scores are not equal for inefficient units [3]. In this paper we concerned about input oriented VRS model, the model is as follows:

$$\text{Min} \quad Z_p = \theta$$

$$s.t.$$

$$\sum_{i=1}^{n} \lambda_i x_i \leq \theta x_p \qquad , \forall j = 1, \dots m$$

$$\sum_{i=1}^{n} \lambda_i y_i \geq y_p \qquad , \forall k = 1, \dots s \qquad (1)$$

$$\sum_{i=1}^{n} \lambda_i = 1$$

$$\lambda_i \geq 0 \qquad , i = 1, \dots n$$

III. THE CHANCE CONSTRAINED OUTPUT ORIENTED DATA ENVELOPMENT ANALYSIS MODEL

In a typical DEA model, the production function estimated by DEA is deterministic. Subhash Ray [4] modified the standard DEA model to measure relative efficiency in the presence of random variation in the all outputs produced from given inputs. For any input bundle, the value of the DEA estimate defines the maximum output producible from inputs under all circumstances. In the stochastic output oriented model, the inputs are assumed to be deterministic while all outputs are random, each output y_k is normally distributed with mean μ_p and variance σ_p^2 and the relation between the same stochastic output variable through different DMUs is independent, this means $cov(y_k, y_p) = 0$. The latter restriction regarding the output quantities in the DEA model translates into a random inequality that may at times be violated. Because an inequality involving a number of random variables can never be imposed with certainty, the strategy in CCP is to ensure that the probability that the inequality holds for a random sample of these variables does not fall below a certain level. Accordingly, the chance constrained output oriented model that measures the efficiency level of DMUp is as follow:

$$\text{Max} \, Z_p = \emptyset$$

$$s.t.$$

$$\sum_{i=1}^{n} \lambda_i \mu_i - \emptyset\mu_p \geq e \sqrt{\sum_{\substack{i=1 \\ i \neq p}}^{n} \lambda_i^2 \sigma_i^2 + (\lambda_p - \emptyset)^2 \sigma_p^2} \quad , \forall k = 1, \dots \dots s$$

$$\sum_{i=1}^{n} \lambda_i x_i \leq x_p \qquad , \forall j = 1, \dots \dots m \qquad (2)$$

$$\sum_{i=1}^{n} \lambda_i = 1$$

$$\lambda_i \geq 0, (i = 1,2, \dots, n)$$

where e is Significance level

IV. DEVELOPED INPUT ORIENTED STOCHASTIC DEA MODEL

Since interested in evaluating the performance of public HEIs in order to assure the quality given that some of the input variables might have as stochastic nature, it was necessary to develop a stochastic input oriented model. Therefore, in this section we present our modification to the standard DEA model (Deterministic DEA) in order to measure technical efficiency in the presence of random variation in some of the inputs. Our developed Stochastic Input Oriented DEA model which is also based on the CCP method is provided below [5].

A. The Chance Constrained Input Oriented Data Envelopment Analysis Model

The restriction involving some of input quantities in the DEA model will be a random inequality that may at times be violated. Because an inequality involving a number of random variables can never be imposed with certainty, the strategy in CCP is to ensure that the probability that the inequality holds for a random sample of these variables does not fall below a certain level. The chance-constrained input oriented model for measuring the efficiency level of DMUp is as follow:

$$\text{Min} \, Z_p = \theta$$

$$s.t.$$

$$pr\left\{\sum_{i=1}^{n} \lambda_i x_i \leq \theta x_p\right\} \geq (1 - \alpha_j) \qquad , \forall j = 1, \dots \dots J_S$$

$$\sum_{i=1}^{n} \lambda_i x_i \leq \theta x_p \quad , \forall j = 1, \dots \dots J_D$$

$$\sum_{i=1}^{n} \lambda_i y_i \geq y_p \qquad , \forall k = 1, \dots \dots s \qquad (3)$$

$$\sum_{i=1}^{n} \lambda_i = 1$$

$$\lambda_i \geq 0, (i = 1,2, \dots, n)$$

After that, we need to know nature of the relation between each DMU for each stochastic input variable, through statistical measure which is covariance. Covariance (*cov*) is a statistical measure of correlation of the fluctuations of two

different quantities. The value of the covariance is interpreted as follows:

Positive covariance: implies that one variable is above (below) its mean value when the other variable is above (below) its mean value.

Negative covariance: implies that one variable is above (below) its mean value when the other variable is below (above) its mean value.

Zero covariance: if the two random variables are independent, the covariance will be zero. However, a covariance of zero does not necessarily mean that the variables are independent.

Assume that:

some of inputs are random variables and remaining inputs are deterministic variables;

each input $x_j, j \in J_S$ is normally distributed with mean μ_p and variance σ_p^2; and

the relation between the same stochastic input variable through different DMUs is dependent, this means $cov(x_i, x_p) \neq 0$.

Then, we can define a random variable u:

$$u = \sum_{i=1}^{n} \lambda_i x_i - \theta x_p \qquad (4)$$

with mean:

$$E(u) = \sum_{i=1}^{n} \lambda_i \mu_i - \theta \mu_p \equiv \mu_u \qquad (5)$$

and with variance:

$$var(u) = \sum_{\substack{i=1 \\ i \neq p}}^{n} \lambda_i^2 \sigma_i^2 + (\lambda_p - \theta)^2 \sigma_p^2 + 2cov(x_i, x_p)$$

$$\equiv \sigma_u^2 \qquad (6)$$

Since the $x_i's$ are normally distributed with mean μ_u and variance σ_u^2, therefore the variable u can be transformed into its equivalent standardized normal value z, as follows:

$$z = \frac{u - \mu_u}{\sigma_u} \qquad (7)$$

Hence,

$$pr\left\{\sum_{i=1}^{m} \lambda_i x_i \leq \theta x_p\right\} = pr\{u \leq 0\}$$

$$= pr\left\{z \leq \frac{-\mu_u}{\sigma_u}\right\} \qquad (8)$$

Given the symmetric property of the normal distribution, then:

$$pr\left\{z \leq \frac{-\mu_u}{\sigma_u}\right\} = pr\left\{z \geq \frac{\mu_u}{\sigma_u}\right\}$$

$$= 1 - \varphi\left(\frac{\mu_u}{\sigma_u}\right) \qquad (9)$$

where $\varphi(\ )$ is the cumulative standard distribution function.

The random inequality restriction in the chance constrained DEA problem (CCDEAP) can be replaced by the equivalent restriction:

$$1 - \varphi\left(\frac{\mu_u}{\sigma_u}\right) \geq (1 - \alpha) \qquad (10)$$

$$-\varphi\left(\frac{\mu_u}{\sigma_u}\right) \geq -\alpha \qquad (11)$$

$$\varphi\left(\frac{\mu_u}{\sigma_u}\right) \leq \alpha \qquad (12)$$

$$\varphi\left(\frac{\mu_u}{\sigma_u}\right) \leq \varphi(e) \qquad (13)$$

$\varphi(e)$ is obtained from the table of standard normal distribution. Hence equation (13) can be written as,

$$\mu_u \leq e\sigma_u \qquad (14)$$

Substitute equations (5) and (6) in equation (14). i.e.,

$$\sum_{i=1}^{n} \lambda_i \mu_i - \theta \mu_p$$

$$\leq e \sqrt{\sum_{\substack{i=1 \\ i \neq p}}^{n} \lambda_i^2 \sigma_i^2 + (\lambda_p - \theta)^2 \sigma_p^2 + 2cov(x_i, x_p)} \qquad (15)$$

Finally, from the above mathematical manipulation, the new presentation for the SIODEA model provided in (3) is as shown below:

$$\text{Min } Z_p = \theta$$

$$s.t.$$

$$\sum_{i=1}^{n} \lambda_i \mu_i - \theta \mu_p \leq \sum_{i=1}^{n} \lambda_i x_i \leq \theta x_p \ , \forall j = 1, \ldots \ldots . J_D$$

$$\sum_{i=1}^{n} \lambda_i y_i \geq y_p \ , \quad \forall k = 1, \ldots \ldots . m \qquad (16)$$

$$\sum_{i=1}^{n} \lambda_i = 1$$

$$\lambda_i \geq 0, (i = 1, 2, \ldots ., n)$$

B. The Algorithm of Developing SIODEA Model

From the previous section, we reached to the final form of the mathematical stochastic input oriented DEA model. Therefore the algorithm and its related flow chart (Figure 1) for the developing model will be as follows:

Step1: **Input**: n, s, J_D, J_S, e

Step2: **Set** $i = 1$

    while $i \leq n$

        a. Input: deterministic inputs, parameters of stochastic inputs, deterministic outputs

        b. Set $\theta_i = 0, \lambda_i = 0$

        c. $i = i+1$

    endwhile

Step3: **Set** $i = 1$

    while $i \leq n$

        a. Formulate model for DMU_i

        b. Calculate θ_i

        c. **If** $\theta_i = 1$ is true

            Print "DMU_i: $\theta_i * 100\%$, Efficient"

          else

            Print" DMU_i: $\theta_i * 100\%$, Inefficient"

          endif

        d. $i = i+1$

    endwhile
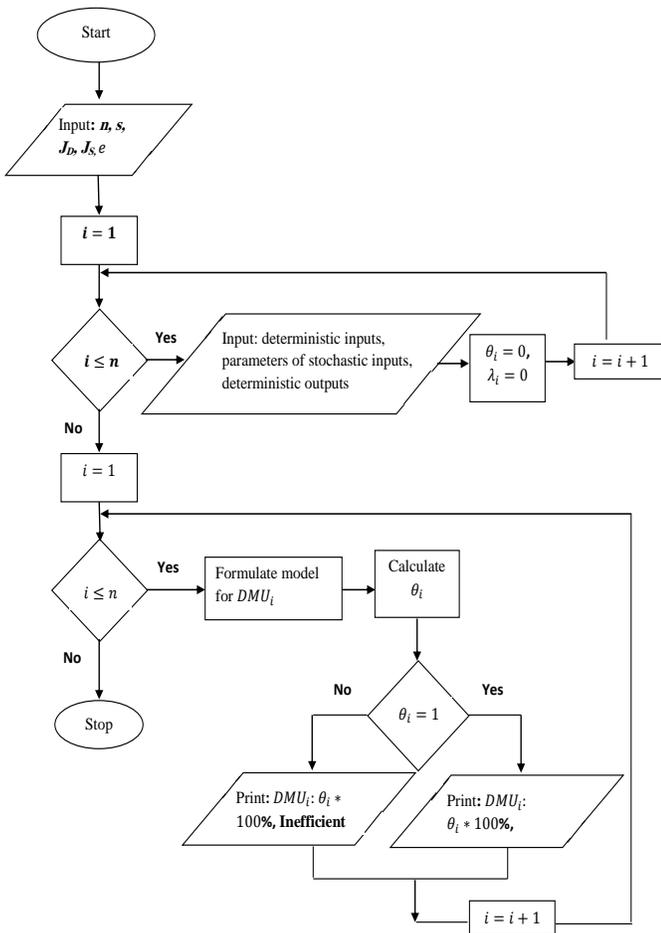
Fig.1.   The flow chart for SIODEA Model

C. Illustrative Example

Step1:

The following hypothetical example considers three universities with two input variables which are number of professors in each university (deterministic) and annual budget (stochastic), and three deterministic outputs which are the number of diplomas, bachelors, and masters granted in year by each university. Budget variable is normally distributed with different mean and variance for each university (DMU). The data for the deterministic variable is provided in Table I, while that for the parameters of stochastic variable is assumed in Table II.

Step2:

TABLE I.   HYPOTHETICAL DATA FOR THE INPUT AND OUTPUTS FOR THE THREE UNIVERSITIES

University	Deterministic Inputs	Outputs		
	No. of Professors	*No. of Diploma*	*No. of Bachelors*	*No. of Masters*
A	5	9	4	16
B	8	5	7	10
C	7	4	9	13

TABLE II.   HYPOTHETICAL DATA FOR THE STOCHASTIC INPUT (BUDGET) FOR THE THREE UNIVERSITIES

University	μ	σ^2	
A	14	1.4	$cov(x_A, x_B) = 0.9$
B	15	1.5	$cov(x_A, x_C) = 0.6$
C	12	1.2	$cov(x_B, x_C) = 0.7$

The aim of this problem is to determine the relative efficiency of the DMUs with respect to each other using the SIODEA model. Assume for the problem that the level of significance for the problem is 5%, and hence e will be 1.96.

Step3:

As has been explained earlier in model (16), a NLP formulation for each university has to be provided in order to measure the relative efficiency.

To evaluate the relative efficiency of **university A**, we need to solve the following NLP problem:

$\text{Min } Z_A = \theta$

$s.t.$

$14\lambda_A + 15\lambda_B + 12\lambda_C - 14\theta \leq$
$1.96\left[\sqrt{1.5\lambda_B^2 + 1.4(\lambda_A - \theta)^2 + 2 * 0.9} + \sqrt{1.2\lambda_C^2 + 1.4(\lambda_A - \theta)^2 + 2 * 0.6}\right]$

$5\lambda_A + 8\lambda_B + 7\lambda_C \leq 5\theta$

$9\lambda_A + 5\lambda_B + 4\lambda_C \geq 9$

$4\lambda_A + 7\lambda_B + 9\lambda_C \geq 4$

(17)

$16\lambda_A + 10\lambda_B + 13\lambda_C \geq 16$

$\lambda_A + \lambda_B + \lambda_C = 1$

$\lambda_A, \lambda_B, \lambda_C \geq 0$

The first constraint is responsible for the representation of the stochastic input variable constraint, where the left hand side represents summation of means for each DMU minus the mean of university A and the right hand side represents the product of the level of significance and square root of summation of the terms variance of each DMU except university A, variance of university A, and twice covariance between university A and other DMUs. The second constraint is responsible for the representation of the deterministic input variable constraints, where the left hand side is the summation of input value for each DMU and the right hand side represents the product of input value of university A and efficiency of university A. The third, fourth and fifth constraints are responsible for representing the output variables constraints, where the left hand side of each constraint represents the summation of output value for each DMU and the right hand side represents the output value of university A. Finally, the sixth constraint ensures that the total weights for all DMUs equals to 1.

Similarly, the relative efficiency models for university B and C respectively are provided in (18) and (19) below.

University B:

$$\text{Min } Z_B = \theta$$

$$s.t.$$

$$14\lambda_A + 15\lambda_B + 12\lambda_C - 15\theta \leq$$
$$1.96\left[\sqrt{1.4\lambda_A^2 + 1.5(\lambda_B - \theta)^2 + 2 * 0.9} + \sqrt{1.2\lambda_C^2 + 1.5(\lambda_B - \theta)^2 + 2 * 0.7}\right]$$

$$5\lambda_A + 8\lambda_B + 7\lambda_C \leq 8\theta$$

$$9\lambda_A + 5\lambda_B + 4\lambda_C \geq 5$$

$$4\lambda_A + 7\lambda_B + 9\lambda_C \geq 7 \tag{18}$$

$$16\lambda_A + 10\lambda_B + 13\lambda_C \geq 10$$

$$\lambda_A + \lambda_B + \lambda_C = 1$$

$$\lambda_A, \lambda_B, \lambda_C \geq 0$$

University C:

$$\text{Min } Z_C = \theta$$

$$s.t.$$

$$14\lambda_A + 15\lambda_B + 12\lambda_C - 12\theta \leq$$
$$1.96\left[\sqrt{1.4\lambda_A^2 + 1.2(\lambda_C - \theta)^2 + 2 * 0.6} + \sqrt{1.5\lambda_B^2 + 1.2(\lambda_C - \theta)^2 + 2 * 0.7}\right]$$

$$5\lambda_A + 8\lambda_B + 7\lambda_C \leq 7\theta$$

$$9\lambda_A + 5\lambda_B + 4\lambda_C \geq 4$$

$$4\lambda_A + 7\lambda_B + 9\lambda_C \geq 9 \tag{19}$$

$$16\lambda_A + 10\lambda_B + 13\lambda_C \geq 13$$

$$\lambda_A + \lambda_B + \lambda_C = 1$$

$$\lambda_A, \lambda_B, \lambda_C \geq 0$$

We then used the GAMS programming language software to solve the above 3 models for each university independently. After running the software the relative efficiency of each university is:

DMU A: 100%,

DMU B: 82.7%, and

DMU C: 100%.

The results reveal that both universities A and C are efficient, while university B is inefficient. In other words, the outputs generated by university B are low given the high inputs for the university (8 professors and 15 units of budget). As noticed from Table I and II, university B has the highest inputs among the three universities and the lowest outputs.

V. CONCLUSION AND FUTURE WORK

Data Envelopment Analysis is an excellent tool for the evaluation of performance. It has the advantage over alternative methods that it can be applied in a multiple inputs and outputs production context. A review of DEA applications revealed that most (if not all) has used the standard deterministic DEA model. They based their selection on the fact that the input and output variables are deterministic by nature, although some might be stochastic in nature. A new model, SIODEA, from the standard DEA model to handle random input variables was developed. The model considers that some of the inputs are stochastic following a normal distribution and the remaining inputs and all outputs are deterministic. Also, it is assumed that the covariance between the different DMUs within the stochastic input variable is not equal to zero. Through the example provided, the SIODEA showed promising results, and the model needs to be applied on actual studies.

As part of the future work, it is the intention of the authors to apply the developed SIODEA model to calculate and compare the efficiency of some public Egyptian universities. Further future work, is to develop a stochastic input oriented DEA model, where the limitations imposed on the stochastic variable are eliminated (i.e. the variables could follow any probability distribution). The model could then be expanded to include both stochastic input and output variables.

REFERENCES

[1] Worthington, A., "An Empirical Survey of Frontier Efficiency Measurement Techniques in Education." Education Economics (2001) 9(3): 245-268.

[2] Charnes A., Cooper W., and Rhodes E., "Measuring the efficiency of efficiency of decision-making units". European Journal of Operational Research. (1978) 2(6): 429-444.

[3] Banker R., Charnes A., and Cooper W., Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science. (1984) 30: 1078-1092.

[4] Subhash C. Ray, Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research. Cambridge University Press. (2004), pp. 307-325.

[5] El-Khodary, I., El-Demerdash, B., and Tharwat, A. "An Algorithm for Evaluating the Performance of Higher Education Organizations in Egypt Using a Stochastic DEA." Proceedings of the 8th International Conference on Data Envelopment Analysis (DEA2010) - Performance Management and Measurement, American University of Beirut, Beirut, Lebanon. (2010).