

Mixed Stochastic Input Oriented Data Envelopment Analysis Model

Nahia Mourad, Assem Tharwat

Abstract: Data envelopment analysis (DEA) is a mathematical tool used to evaluate relative efficiency of decision-making units (DMUs). It is a benchmarking method for these units. To measure this relative efficiency, data related to a set of inputs and outputs are provided from all the DMUs under analysis, and then implemented in a suitable DEA model. Stochastic DEA allows the inputs or outputs to be stochastic random variables. In this article, we consider combination of deterministic and stochastic inputs following Normal and/or Poisson distribution. To the best of our knowledge, variables following Poisson distribution are not yet considered in these methods. We introduce the stochastic input-oriented data envelopment analysis (SIODEA) model. The random inputs, following either normal or Poisson distributions, are controlled by chance constrained. Using functional analysis techniques, the chance constrained with Poisson variables is replaced by difference of Marcum functions evaluated at points related to the parameters of these variables. Consequently, we formulate a deterministic equivalent model with mixed random inputs. Finally, a numerical example is presented and the efficiencies of different DMUs are calculated using the obtained equivalent model to test its validity.

Keywords: Mathematical Programming, Performance, Efficiency, Decision Making, Data Envelopment Analysis, Input Oriented DEA, Stochastic Inputs, Chance Constrained

1 INTRODUCTION

Data envelopment analysis has become a powerful tool for measuring product or organizational performance [14]. It was first introduced by Charnes and Cooper [7], in 1978. DEA, which is a non-parametric method, uses mathematical programming to analyze the performance of comparable organizations. These organizations are known by the decision-making units. Information regarding predetermined inputs and outputs are gathered from different DMUs and implemented in a suitable DEA model. Using this model, one can compute the maximum relative efficiency of each DMU under analysis. DEA method is not restricted to business and economics problems. It has a wide range of applications in different fields. This method is used to compute the efficiency in agriculture [1, 5, 20], ecological development [8], consumption and production of energy [20, 27, 28], etc. It is especially used in the studies of sustainable energy [12, 21]. More applications of DEA, in various fields, can be found in the literature. The efficiency score is the ratio of the weighted output to the weighted input [29]. This produces a DEA ratio problem. This last is transformed from fractional programming to linear programming, forming either input or output oriented model. Finally, the dual of this last formulation is generated (for more details refer to [2]). In this paper, we will consider input oriented data envelopment analysis model. There are many types of DEA models. We differentiate between variable return to scale (VRS), and constant return to scale (CRS). The former is stemmed from a model set by Banker, Charnes and Cooper (BCC), while the last is based on a model described by Charnes, Cooper and Rhodes (CCR) [4]. In this paper, we consider the VRS DEA models, for which the change in inputs is not necessarily proportional to the change in outputs [23].

Introducing the stochastic data envelopment analysis [26] is a trial to improve the DEA approach. The obtained model allows the possible presence of stochastic inputs and outputs [11]. That is, this possible input or output can be a random variable following specific probability distribution. In that case, the constrained with random variables must be maintained at a prescribed probability level. This is known by chance constrained [10, 11]. Stochastic DEA model with random variable (typically input or output) following normal distribution, is already introduced in the literature [19]. In fact, this random variable need not necessarily follows normal distribution. For example, if the random variable measures the rate of occurrence of a specific event with a known average. It, under specific conditions, may follow Poisson distribution. If the average is not relatively big, this random variable can't be approximated by normal distribution. In this article, we introduce the possibility of allowing inputs to follow Poisson distributions. Performance assessment of organization(s) is an important management requirement. Managerial decisions should focus on clear evidence and decision science techniques. To increase the economic impact of the industry, the economic benefits in organization environment need to be investigated and discussed using a Data Envelopment Analysis (DEA) model. DEA is an optimization technique which has the potential to compare DMUs through efficiency evaluation. It generates ranking of DMUs, which will facilitate a rationale decision making for managerial actions. This technique has become a popular management science tool in assessing performance of organizations [6]. In this paper, we extended the type of distribution the inputs are following; thus, the random variable input won't be restricted to normal. This maximize the economic output of organization environment and provide decision makers a flexibility to choose the appropriate distribution of the stochastic variables being studied. This paper is organized as follows: In Section 2, we recall the general input-oriented DEA model. We presented the corresponding minimization problem with inputs and outputs are all deterministic in Section 2.1. Following, in Section 2.2, the stochastic input-oriented DEA (SIODEA) minimization problem with chance constrained is set. This allows the inputs to be either deterministic or stochastic variables. In Section 3, using

- Nahia Mourad is currently an assistant professor at American University in the Emirates, Dubai, UAE, E-mail: nahia.mourad@aue.ae
- Assem Tharwat is currently a full professor at American University in the Emirates, Dubai, UAE, E-mail: assem.tharwat@aue.ae

functional analysis, some formulas for the probability of comparing two normal and two Poisson distributions are generated. The main findings are presented in Section 4, where we developed a novel deterministic model equivalent to the SIODEA one, which allows inputs to be possibly Poisson and/or normal random variables. Finally, in Section 5, we presented an application, which is solved using the obtained model to test its validity.

2 INPUT ORIENTED DATA ENVELOPMENT ANALYSIS MODEL

Data envelopment analysis is an approach used to compare the efficiency of different decision-making units. The efficiency score, which is the ratio of linear combination of output(s) and input(s), of each DMU is set for maximization. This formulate the DEA ratio model of the DMU under analysis. A minimization problem with constrained is obtained by transforming the DEA ratio model to a linear input-oriented model. The dual of this last formulation is generated [2] to obtain what is known by input oriented data envelopment analysis model (IODEA). In the following, we consider n DMUs are under examination with m inputs and s deterministic outputs for each DMU denoted by $(y_{kj})_{1 \leq k \leq s, 1 \leq j \leq n} \in \mathbb{R}^{s \times n}$.

2.1 Deterministic Input(s)

In this section, all the inputs and outputs are assumed to be deterministic. That is, the m inputs, for all the DMUs, $(x_{ij})_{1 \leq i \leq m, 1 \leq j \leq n} \in \mathbb{R}^{m \times n}$. Let $p \in \{1, \dots, n\}$, the VRS DEA model which evaluates the efficiency of the p -th DMU is as follows:

$$\begin{aligned} & \text{Minimum}(\theta_p) \\ & \text{s.t.} \\ & \sum_{j=1}^n \lambda_j x_{ij} \leq \theta_p x_{ip}, \quad i = 1, \dots, m \\ & \sum_{j=1}^n \lambda_j y_{rj} \geq y_{rp}, \quad r = 1, \dots, s \\ & \sum_{j=1}^n \lambda_j = 1 \\ & \lambda_j \geq 0, \quad 1 \leq j \leq n \end{aligned} \quad (\text{P-I})$$

where the optimal $\theta_p \in [0,1]$ is the maximum efficiency score of the p -th DMU, λ_j 's are the weights given for the j 's DMU, which can be seen as the Lagrange Multipliers corresponding to the constrained of the dual problem [2]. Minimization problem (P-I) is solved for each $p \in \{1, \dots, n\}$ to calculate the efficiency scores of all the DMUs under analysis.

2.2 Stochastic Input(s)

The stochastic input-oriented data envelopment analysis (SIODEA) model, allows any input to be stochastically variable [11]. Each constrained with stochastic variables must be maintained at prescribed probability level. This is known by chance constrained programming (CCP) [10, 11].

Let J_S be the set of indices i for which $(x_{ij})_{1 \leq j \leq n} \in \mathcal{S}^n$, where \mathcal{S} is the set of random variables.

Let J_D be the set of indices i for which $(x_{ij})_{1 \leq j \leq n} \in \mathbb{R}^n$ is a vector of deterministic inputs. We have

$$J_D \cup J_S = \{1, \dots, m\}.$$

The following VRS SIODEA model determine the optimal efficiency score of the p -th DMU, with m inputs being either deterministic or random variables and s deterministic outputs,

$$\begin{aligned} & \text{Minimum}(\theta_p) \\ & \text{s.t.} \\ & P \left\{ \sum_{j=1}^n \lambda_j x_{ij} \leq \theta_p x_{ip} \right\} \geq 1 - \epsilon, \quad i \in J_S \\ & \sum_{j=1}^n \lambda_j x_{ij} \leq \theta_p x_{ip}, \quad i \in J_D \\ & \sum_{j=1}^n \lambda_j y_{rj} \geq y_{rp}, \quad r = 1, \dots, s \\ & \sum_{j=1}^n \lambda_j = 1 \\ & \lambda_j \geq 0, \quad 1 \leq j \leq n \end{aligned} \quad (\text{P-II})$$

where $\epsilon \in [0,1]$ is a small prescribed number. Similarly, the optimal $\theta_p \in [0,1]$, solving the minimization problem (P-II), is the maximum efficiency score of the p -th DMU. One should solve problem (P-II) for all $p \in \{1, \dots, n\}$ to obtain all the efficiency scores for the n DMUs under analysis.

3 PRELIMINARIES

3.1 Normal Random variables

In this section, we prove a useful equivalency property for linear combination of normal random variables.

Lemma 1. Let X be normally distributed random variable with mean μ and standard deviation σ , (i.e $X \sim \mathcal{N}(\mu, \sigma)$). Then

$$P\{X \leq 0\} \geq 1 - \epsilon$$

is equivalent to

$$\mu \leq e \sigma,$$

where e is defined by $\Phi_Z(e) = P\{Z \leq e\} = \epsilon$ and Φ_Z is the cumulative density function of $Z \sim \mathcal{N}(0,1)$.

Proof. The random variable Z , defined by

$$Z = \frac{X - \mu}{\sigma},$$

is a standard normal distribution, with mean $\mu_Z = 0$ and standard deviation $\sigma_Z = 1$. We have

$$P\{X \leq 0\} = P\left\{Z \leq \frac{-\mu}{\sigma}\right\} = \Phi_Z\left(\frac{-\mu}{\sigma}\right) \geq 1 - \epsilon$$

The symmetry property of the standard normal distribution implies that

$$\Phi_Z\left(\frac{-\mu}{\sigma}\right) = 1 - \Phi_Z\left(\frac{\mu}{\sigma}\right).$$

Therefore,

$$\Phi_Z\left(\frac{\mu}{\sigma}\right) \leq \epsilon = \Phi_Z(e). \quad (1)$$

Applying the inverse function Φ_Z^{-1} to (1), we get

$$\frac{\mu}{\sigma} \leq e. \quad (2)$$

Proposition 1. For $1 \leq k \leq N$, let X_k be normally distributed random variables with mean μ_k and standard deviation σ_k , i.e

$$X_k \sim \mathcal{N}(\mu_k, \sigma_k),$$

then, for $(a_k)_{1 \leq k \leq N} \in \mathbb{R}^N$,

$$P\left\{\sum_{k=1}^N a_k X_k \leq 0\right\} \geq 1 - \epsilon$$

is equivalent to

$$\sum_{k=1}^N a_k \mu_k \leq e \sqrt{\sum_{k=1}^N a_k^2 \sigma_k^2 + 2 \sum_{i < k} a_i a_k \text{Cov}(X_i, X_k)},$$

where $\Phi_Z(e) = \epsilon$.

Proof. The random variable U defined by

$$U := \sum_{k=1}^N a_k X_k$$

is normally distributed [16] with mean

$$\mu_U = \sum_{k=1}^N a_k \mu_k$$

and variance

$$\sigma_U^2 = \sum_{k=1}^N a_k^2 \sigma_k^2 + 2 \sum_{i < k} a_i a_k \text{Cov}(X_i, X_k).$$

We have

$$P\{U \leq 0\} \geq 1 - \epsilon$$

Lemma 1 implies that

$$\mu_U \leq e \sigma_U.$$

That is

$$\sum_{k=1}^N a_k \mu_k \leq e \sqrt{\sum_{k=1}^N a_k^2 \sigma_k^2 + 2 \sum_{i < k} a_i a_k \text{Cov}(X_i, X_k)}.$$

3.2 Poisson Random Variables

It is known that the sum of independent Poisson distributions is a Poisson distribution [22]. However, the subtraction of two independent Poisson distributions follows Skellam distribution [18]. In this section, we derive a relation between the probability comparing independent Poisson random variables, that is the probability of Skellam variable below or above zero, and Marcum function.

Definition 1. [3] The modified Bessel functions I_k of the first kind of order $k \in \mathbb{N}$ are defined by

$$I_k(x) := \sum_{m=0}^{\infty} \frac{1}{m! (m+k)!} \left(\frac{x}{2}\right)^{2m+k}. \quad (3)$$

Remark 1. The modified Bessel function of the first kind of zero order can be computed by integration as follows

$$I_0(x) := \sum_{m=0}^{\infty} \frac{1}{m!^2} \left(\frac{x}{2}\right)^{2m} = \frac{1}{\pi} \int_0^{\pi} \exp(x \cos \theta) d\theta \quad x \geq 0. \quad (4)$$

Definition 2. [25] The Skellam distribution with parameters ν_1 and ν_2 , denoted by

$$S \sim \text{Skellam}(\nu_1, \nu_2),$$

is defined by its probability function as follows

$$P(S = s) = e^{-(\nu_1 + \nu_2)} \left(\frac{\nu_1}{\nu_2}\right)^{s/2} I_{|s|}(2\sqrt{\nu_1 \nu_2})$$

for all $s \in \mathbb{Z}$.

In fact, for $X \sim \text{Poisson}(\nu)$ and $Y \sim \text{Poisson}(\mu)$, if $\text{Cov}(X, Y) = 0$, i.e X and Y are independent, then $X - Y \sim \text{Skellam}(\nu, \mu)$ [15].

Definition 3. The Marcum Q function is defined by

$$Q(a, b) := \int_b^{\infty} x \exp\left(-\frac{x^2 + a^2}{2}\right) I_0(ax) dx$$

for all $a, b \geq 0$, where I_0 is the modified Bessel function of zero order given by (4).

Note that, the Marcum function is a built-in routine in MATLAB and can be called by $Q = \text{marcumq}(a, b)$.

Lemma 2. Let X and Y be independent random variables following Poisson distribution with means ν and μ , respectively; then

$$P\{X \leq Y\} = 1 - \exp(-\mu) \int_0^{\nu} \exp(-t) I_0(2\sqrt{\mu t}) dt,$$

where I_0 is the modified Bessel function of order 0, defined in (4).

Proof. Let $X \sim \text{Poisson}(x)$ for $x \in \mathbb{R}_+$ independent of Y . Using the probability density function of the Poisson random variable, for $k \in \mathbb{N}$, we have

$$P\{X = k\} = \frac{\exp(-x) x^k}{k!}.$$

Differentiating both sides with respect to x , we get for $k > 0$,

$$\begin{aligned} \frac{d}{dx} P\{X = k\} &= \frac{1}{(k-1)!} \exp(-x) x^{k-1} \\ &- \frac{1}{k!} \exp(-x) x^k = P\{X \\ &= k-1\} - P\{X = k\}, \end{aligned} \quad (5)$$

however, for $k = 0$,

$$\frac{d}{dx} P\{X = 0\} = -\exp(-x) = -P\{X = 0\}. \quad (6)$$

Using equations (5) and (6), we get for $y \in \mathbb{N}$,

$$\frac{d}{dx} P\{X \leq y\} = \frac{d}{dx} \sum_{k=0}^y P\{X = k\} = -P\{X = y\}. \quad (7)$$

We have

$$P\{X \leq Y\} = \sum_{k=0}^{\infty} P\{X \leq y \mid Y = y\}.$$

Since X and Y are independent random variables, then

$$P\{X \leq Y\} = \sum_{y=0}^{\infty} P\{X \leq y\} P\{Y = y\}.$$

Using Fubini theorem on differentiation [17], we get

$$\begin{aligned} \frac{d}{dx} P\{X \leq Y\} &= \frac{d}{dx} \left(\sum_{y=0}^{\infty} P\{X \leq y\} P\{Y = y\} \right) \\ &= \sum_{y=0}^{\infty} \frac{d}{dx} P\{X \leq y\} P\{Y = y\}. \end{aligned} \quad (8)$$

Using (7) and (8), we obtain

$$\frac{d}{dx} P\{X \leq Y\} = - \sum_{y=0}^{\infty} P\{X = y\} P\{Y = y\}.$$

Hence

$$\begin{aligned} \frac{d}{dx} P\{X \leq Y\} &= - \sum_{y=0}^{\infty} \exp(-(x + \mu)) \frac{x^y \mu^y}{(y!)^2} \\ &= - \exp(-(x + \mu)) I_0(2\sqrt{x\mu}). \end{aligned}$$

The Fundamental Theorem of Calculus implies that

$$P\{X \leq Y\} - P\{0 \leq Y\} = - \int_0^x \exp(-(t + \mu)) I_0(2\sqrt{t\mu}) dt.$$

In particular, for $x = v$,

$$P\{X \leq Y\} = 1 - \int_0^v \exp(-(t + \mu)) I_0(2\sqrt{t\mu}) dt.$$

Proposition 2. Let X and Y be independent random variables following Poisson distributions with means v and μ , respectively; then

$$P\{X > Y\} = Q(\sqrt{2\mu}, 0) - Q(\sqrt{2\mu}, \sqrt{2v})$$

where Q is the Marcum function.

Proof. Lemma 2 implies that

$$P\{X \leq Y\} = 1 - \exp(-\mu) \int_0^v \exp(-t) I_0(2\sqrt{t\mu}) dt.$$

Opposite events property implies that

$$P\{X > Y\} = \exp(-\mu) \int_0^v \exp(-t) I_0(2\sqrt{t\mu}) dt.$$

Substituting x by $\sqrt{2t}$ in the integral of the right-hand side, we get

$$\begin{aligned} P\{X > Y\} &= \exp(-\mu) \int_0^{\sqrt{2v}} x \exp\left(-\frac{x^2}{2}\right) I_0(\sqrt{2\mu}x) dx \\ &= \int_0^{\sqrt{2v}} x \exp\left(-\frac{x^2 + (\sqrt{2\mu})^2}{2}\right) I_0(\sqrt{2\mu}x) dx. \end{aligned}$$

Using Definition 3 of the Marcum function, we get

$$P\{X > Y\} = Q(\sqrt{2\mu}, 0) - Q(\sqrt{2\mu}, \sqrt{2v}).$$

4 FORMULATION OF SIODEA MODEL

In this section, we develop a novel DEA model, which is equivalent to the VRS SIODEA model (P-II) with deterministic outputs and mixed inputs. Each input from all the DMUs can be either

1. deterministic variable or
2. normal random variable or
3. Poisson random variable.

The new DEA model is a minimization problem with deterministic nonlinear constrained replacing the chance constrained in problem (P-II).

Let J_N be the set of indices i for which $x_{ij} \in \mathcal{S}$ is normally distributed with mean μ_{ij} and standard deviation σ_{ij} for any $j \in \{1, \dots, n\}$, i.e.

$$J_N := \{i \in \{1, \dots, m\} \mid x_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}), \forall j = 1, \dots, n\}. \quad (9)$$

Similarly define J_P as the set of indices i for which $x_{ij} \in \mathcal{S}$ follows poisson distribution with mean v_{ij} , i.e.

$$J_P := \{i \in \{1, \dots, m\} \mid x_{ij} \sim \text{Poisson}(v_{ij}), \forall j = 1, \dots, n\}. \quad (10)$$

We have

$$J_S = J_N \cup J_P.$$

Recall that J_D is the set of indices i for which $x_{ij} \in \mathbb{R}$ is a deterministic input for any $j = 1, \dots, n$.

4.1 Deterministic equivalence of chance constrained with normal inputs

In this section, we construct a deterministic equivalent condition to the probabilistic one given in the optimization problem (P-II), in case the inputs are normally distributed. In this last condition the means and the standard deviations of the normal random variables are compared to a parameter corresponding to the standard normal distribution.

A direct consequence of Proposition 1 is that, for $i \in J_N$ and $p \in \{1, \dots, n\}$, the first constrained in problem (P-II)

$$P\left\{\sum_{j=1}^N \lambda_j x_{ij} \leq \theta_p x_{ip}\right\} \geq 1 - \epsilon$$

is equivalent to

$$\begin{aligned} &\sum_{j=1}^n \lambda_j \mu_{ij} - \theta_p \mu_{ip} \\ &\leq e \sqrt{\sum_{j=1}^n (\lambda_j - \delta_{jp} \theta_p)^2 \sigma_{ij}^2 + 2 \sum_{k < j} (\lambda_k - \delta_{kp} \theta_p)(\lambda_j - \delta_{jp} \theta_p) t} \end{aligned} \quad (1)$$

where δ_{ij} is the Kronecker delta, defined by

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases}$$

4.2 Properties following chance constrained with Poisson inputs

For any $i \in J_P$, assume that, $(x_{ij})_{j=1, \dots, n}$ are independent Poisson distribution random variables. Consider $\theta_p \in [0, 1]$ and $(\lambda_j)_{1 \leq j \leq N} \in [0, 1]^N$ s.t. $\sum_{j=1}^n \lambda_j = 1$. Define the following random variable

$$X_{i,p} := \sum_{j=1}^n \lambda_j x_{ij} - \theta_p x_{ip}. \quad (12)$$

The chance constrained in the stochastic DEA model (P-II) is as follows

$$P\{X_{i,p} \leq 0\} \geq 1 - \epsilon. \quad (13)$$

Remark 2. For $\epsilon \in (0, 1)$, if $X_{i,p}$ satisfies (13), then

$$\begin{cases} 0 \leq \lambda_p < \theta_p \\ \text{or} \\ \theta_p = \lambda_p = 1 \end{cases}$$

since otherwise $X_{i,p} = \sum_{j=1, j \neq p}^n \lambda_j x_{ij} - (\theta_p - \lambda_p)x_{ip}$ is a non-zero random variable following Poisson distribution with $P(X_{i,p} \leq 0) > 0$.

Remark 2 and the fact that the sum of independent Poisson distribution is a Poisson distribution [22] implies that $X_{i,p}$ defined by (12) is either identically zero or

$$X_{i,p} \sim Skellam \left(\sum_{j=1, j \neq p}^n \lambda_j v_{ij}, (\theta_p - \lambda_p)v_{ip} \right).$$

4.3 Deterministic Equivalent Model

The following theorem proves the existence of an equivalent deterministic model to the SIODEA problem given by (P-II). The outputs are deterministic, while the inputs are mixed between deterministic variables, normal random variables and independent Poisson random variables. The chance constrained with normal variables are replaced by a deterministic constrained involving the parameters of the normal distributions. However, this was not trivial for chance constrained with Poisson variable as there is no standard form for Poisson distributions.

Theorem 1. For $i \in \mathcal{J}_p$, assume $(x_{ij})_{1 \leq j \leq n}$ are independent

Poisson distributions, then

the minimization problem of the SIODEA (P-II) is equivalent to

$$\begin{aligned} & \text{Minimum}(\theta_p) \\ & \text{s. t.} \end{aligned}$$

$$\begin{aligned} & \sum_{j=1}^n \lambda_j \mu_{ij} - \theta_p \mu_{ip} \\ & \leq e \sqrt{\sum_{j=1}^n (\lambda_j - \delta_{jp} \theta_p)^2 \sigma_{ij}^2 + 2 \sum_{k < j} (\lambda_k - \delta_{kp} \theta_p)(\lambda_j - \delta_{jp} \theta_p) C_{kj}} \\ & \in \mathcal{J}_N \\ & Q \left(\sqrt{2(\theta_p - \lambda_p)v_{ip}}, 0 \right) \\ & - Q \left(\sqrt{2(\theta_p - \lambda_p)v_{ip}}, \sqrt{2 \sum_{j=1, j \neq p}^n \lambda_j v_{ij}} \right) \leq \epsilon, \quad i \in \mathcal{J}_p \end{aligned} \quad \begin{matrix} \text{(P)} \\ \text{-} \\ \text{III)} \end{matrix}$$

$$\sum_{j=1}^n \lambda_j x_{ij} \leq \theta_p x_{ip}, \quad i \in \mathcal{J}_D$$

$$\sum_{j=1}^n \lambda_j y_{rj} \geq y_{rp}, \quad r = 1, \dots, s$$

$$\sum_{j=1}^n \lambda_j = 1$$

$$\lambda_j \geq 0, \quad 1 \leq j \leq n$$

where δ_{ij} is the Kronecker delta, Q is the Marcum function, and $e = \Phi^{-1}(\epsilon)$, where Φ is the cumulative density function of the standard normal distribution and the sets \mathcal{J}_N and \mathcal{J}_p are defined by (9) and (10).

Proof. For $i \in \mathcal{J}_p$, the chance constrained in model (P-II) can be written as

$$P \left\{ \sum_{j=1, j \neq p}^n \lambda_j x_{ij} \leq (\theta_p - \lambda_p)x_{ip} \right\} \geq 1 - \epsilon,$$

which is equivalent to

$$P \left\{ \sum_{j=1, j \neq p}^n \lambda_j x_{ij} > (\theta_p - \lambda_p)x_{ip} \right\} \leq \epsilon. \tag{14}$$

Remark 2 ensures that $(\theta_p - \lambda_p) \geq 0$, then

$$\sum_{j=1, j \neq p}^n \lambda_j x_{ij} \sim \text{Poisson} \left(\sum_{j=1, j \neq p}^n \lambda_j v_{ij} \right)$$

and

$$(\theta_p - \lambda_p)x_{ip} \sim \text{Poisson} \left((\theta_p - \lambda_p)v_{ip} \right).$$

Proposition 2 implies that the condition given in (14) is equivalent to

$$\begin{aligned} & Q \left(\sqrt{2(\theta_p - \lambda_p)v_{ip}}, 0 \right) \\ & - Q \left(\sqrt{2(\theta_p - \lambda_p)v_{ip}}, \sqrt{2 \sum_{j=1, j \neq p}^n \lambda_j v_{ij}} \right) \leq \epsilon. \end{aligned}$$

Therefore, the theorem holds by combining this last inequality with the equivalency given by condition (11).

5 APPLICATION

In this section, we present a hypothetical example which will be solved by model (P-III). Consider a university with three colleges. This university wishes to study the relative performance of its colleges. For that purpose, it examines three inputs: number of professors $(x_{1j})_{1 \leq j \leq 3}$, annual budget $(x_{2j})_{1 \leq j \leq 3}$ and annual students registration $(x_{3j})_{1 \leq j \leq 3}$. The former is a deterministic variable, while the rest are stochastic. It also observed two outputs: the number of bachelors $(y_{1j})_{1 \leq j \leq 3}$ and masters $(y_{2j})_{1 \leq j \leq 3}$ granted to students. The data for the deterministic variables and the parameters for the stochastic random variables for the three colleges are given by Table 1.

	College 1	College 2	College 3
No. of Professors (x_{1j})	6	15	10
No. of Bachelors (y_{1j})	29	27	2
No. of Masters (y_{2j})	23	8	12
Mean of Annual Budget (μ_j)	17	28	12
Variance of Annual Budget (σ_j^2)	1.3	1.4	1.3
Mean of Students registration (v_j)	13	18	10

Table 1: Inputs and outputs data for the three colleges

The annual students' registration for the three colleges are independent, while the annual budget for the three colleges is a dependent variable as they are parts of one university. The dependency is described by the covariance as follows:

$$\begin{aligned} Cov(x_{21}, x_{22}) &= 0.8, \quad Cov(x_{21}, x_{23}) \\ &= 0.7, \quad \text{and} \quad Cov(x_{22}, x_{23}) = 0.6 \end{aligned}$$

Recall that, the aim of this example is to calculate the relative efficiency of the three colleges, within the same university.

Let $\epsilon = 30\%$ and hence $e = \Phi^{-1}(\epsilon) = -0.5244$.

All the below results are obtained by running a MATLAB script, which is available on Github for further simulations [24].

Case 1

In this case, we consider the annual budget follows normal distribution, while the annual students' registration follows Poisson distribution. Then, according to Theorem 1, the relative efficiency score of the p -th college θ_p solves the following nonlinear optimization problem

$$\text{Minimum}(\theta_p)$$

s. t.

$$\begin{aligned} 17\lambda_1 + 28\lambda_2 + 12\lambda_3 - \mu_p \theta_p \\ \leq -0.5244 \left[1.3(\lambda_1 - \delta_{1p} \theta_1)^2 \right. \\ + 1.4(\lambda_2 - \delta_{2p} \theta_2)^2 \\ + 1.3(\lambda_3 - \delta_{3p} \theta_3)^2 \\ + 2(0.8(\lambda_1 - \delta_{1p} \theta_1)(\lambda_2 - \delta_{2p} \theta_2) \\ + 0.7(\lambda_1 - \delta_{1p} \theta_1)(\lambda_3 - \delta_{3p} \theta_3) \\ \left. + 0.6(\lambda_2 - \delta_{2p} \theta_2)(\lambda_3 - \delta_{3p} \theta_3) \right]^{1/2} \end{aligned} \tag{15}$$

$$\begin{aligned} Q \left(\sqrt{2v_p(\theta_p - \lambda_p)}, 0 \right) \\ - Q \left(\sqrt{2v_p(\theta_p - \lambda_p)}, \sqrt{2(13\lambda_1 + 18\lambda_2 + 16\lambda_3 - v_p \lambda_p)} \right) \\ \leq 0.3 \end{aligned}$$

$$6\lambda_1 + 15\lambda_2 + 10\lambda_3 \leq x_{1p} \theta_p$$

$$29\lambda_1 + 27\lambda_2 + 2\lambda_3 \geq y_{1p}$$

$$23\lambda_1 + 8\lambda_2 + 12\lambda_3 \geq y_{2p}$$

$$\begin{aligned} \lambda_1 + \lambda_2 + \lambda_3 &= 1 \\ \lambda_1, \lambda_2, \lambda_3 &\geq 0 \end{aligned}$$

where $\mu_p, v_p, x_{1p}, y_{1p}$ and y_{2p} are varying according to Table 1. Solving problem (15), for $p \in \{1, 2, 3\}$, we obtain

- First college efficiency score: $\theta_1 = 100\%$;
- Second college efficiency score: $\theta_2 = 98.36\%$;
- Third college efficiency score: $\theta_3 = 100\%$.

This means that first and third colleges are efficient while the second one is not efficient.

Case 2

In this case, the annual budget and the student registration inputs are assumed to follow normal distributions. Thus, we have two normal inputs and no Poisson inputs (i.e. $|J_N| = 2$ and $|J_P| = 0$). The variance of student registration variable corresponding to each college is equal to its mean and with no correlation among the colleges (i.e. $Cov(x_{3j_1}, x_{3j_2}) = 0$ for

$j_1 \neq j_2$). Then, according to Theorem 1, the relative efficiency score of the p -th college θ_p solves the following nonlinear optimization problem

$$\text{Minimum}(\theta_p)$$

s. t.

$$\begin{aligned} 17\lambda_1 + 28\lambda_2 + 12\lambda_3 - \mu_p \theta_p \\ \leq -0.5244 \left[1.3(\lambda_1 - \delta_{1p} \theta_1)^2 \right. \\ + 1.4(\lambda_2 - \delta_{2p} \theta_2)^2 \\ + 1.3(\lambda_3 - \delta_{3p} \theta_3)^2 \\ + 2(0.8(\lambda_1 - \delta_{1p} \theta_1)(\lambda_2 - \delta_{2p} \theta_2) \\ + 0.7(\lambda_1 - \delta_{1p} \theta_1)(\lambda_3 - \delta_{3p} \theta_3) \\ \left. + 0.6(\lambda_2 - \delta_{2p} \theta_2)(\lambda_3 - \delta_{3p} \theta_3) \right]^{1/2} \end{aligned} \tag{16}$$

$$\begin{aligned} 13\lambda_1 + 18\lambda_2 + 16\lambda_3 - v_p \theta_p \\ \leq -0.5244 \left[13(\lambda_1 - \delta_{1p} \theta_1)^2 \right. \\ + 18(\lambda_2 - \delta_{2p} \theta_2)^2 \\ \left. + 16(\lambda_3 - \delta_{3p} \theta_3)^2 \right]^{1/2} \end{aligned}$$

$$6\lambda_1 + 15\lambda_2 + 10\lambda_3 \leq x_{1p} \theta_p$$

$$29\lambda_1 + 27\lambda_2 + 2\lambda_3 \geq y_{1p}$$

$$23\lambda_1 + 8\lambda_2 + 12\lambda_3 \geq y_{2p}$$

$$\begin{aligned} \lambda_1 + \lambda_2 + \lambda_3 &= 1 \\ \lambda_1, \lambda_2, \lambda_3 &\geq 0 \end{aligned}$$

where $\mu_p, v_p, x_{1p}, y_{1p}$ and y_{2p} are given in Table 1. Solving problem (16), for $p \in 1,2,3$, we obtain

- First college efficiency score: $\theta_1 = 100\%$;
- Second college efficiency score: $\theta_2 = 87.28\%$;
- Third college efficiency score: $\theta_3 = 100\%$.

This means that first and third colleges are efficient while the second one is not efficient.

Note that the efficiency score of the second college differs in the above two cases. Assuming the student registration variable is normally distributed leads to different results than assuming it following Poisson distribution. Therefore, the above results show the assumption that any random variable under consideration in a stochastic DEA model is following normal distribution will lead to misleading measurement for the DMUs efficiency. Interpreting the efficiency scores, while assuming all the random variables are normally distributed, may lead to wrong management decision. Therefore, it is recommended to choose the appropriate distribution for the random variables (inputs or outputs) before using data envelopment analysis approach.

6 CONCLUSION AND FUTURE WORK

DEA is a powerful tool, in management science, using mathematical programming to evaluate the performance, which is determined by the efficiency score, of different organizations, firms, products, etc. This efficiency measurement takes place in different applications like education, energy, agriculture, etc. Stochastic DEA models allow the inputs or outputs to be stochastic random variables. Only normal variables in such models were introduced in the literature. Variables which can be described by rate of flow usually follow Poisson or exponential distributions. These

distributions have not been yet considered in stochastic DEA models. In this paper, the probability comparing two Poisson random variables is determined by difference of Marcum functions evaluated at different points related to the means of these variables. As a result, a deterministic model with non-identical random inputs, varying between either normal or Poisson distributions, is developed. In reality, this developed model gives flexibility in choosing the appropriate distribution for the input variables under consideration. The approach, used to formulate this novel model, can be used for output oriented stochastic DEA. Lastly, it is shown that misleading efficiency score will result from assuming all the stochastic inputs are normally distributed. For future work, inputs and outputs being random variables following distributions different than normal and Poisson can be considered. Moreover, the assumption on the dependency of Poisson random variables among the DMUs under analysis can be addressed in future studies.

REFERENCES

- [1] T. Baležentis, I. Kriščiukaitienė, & A. Baležentis "A nonparametric analysis of the determinants of family farm efficiency dynamics in Lithuania," *Agricultural economics*, 45(5), 589-599, 2014.
- [2] R. D. Banker, A. Charnes, & W. W. Cooper "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management science*, 30(9), 1078-1092. *Management Science*. (1984) 30: 1078-1092.
- [3] A. Baricz, "Bounds for modified Bessel functions of the first and second kinds," *Proceedings of the Edinburgh Mathematical Society*, 53(3), 575-599, 2010.
- [4] J. Benicio & J. C. S. De Mello "Productivity analysis and variable returns of scale: DEA efficiency frontier interpretation," *Procedia Computer Science*, 55(I tqm), 341-349, 2015.
- [5] E. Bolandnazar, A. Keyhani & M. Omid "Determination of efficient and inefficient greenhouse cucumber producers using Data Envelopment Analysis approach," a case study: Jiroft city in Iran. *J Clean Prod*; 79:108-15, 2014.
- [6] V. Charles & M. Kumar (Eds.) "Data envelopment analysis and its applications to management" Cambridge Scholars Publishing, 2013.
- [10] 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(12), 1347-1356, 2002.
- [11] 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(2), 487-501, 2004.
- [12] A. O. Costa, L. B. Oliveira, M. P. E. Lins, A. C. M. Silva, M. S. M. Araujo, A. O. Pereira Jr, & L. P. Rosa "Sustainability analysis of biodiesel production: A review on different resources in Brazil," *Renewable and Sustainable Energy Reviews*, 27, 407-412, 2013.
- [13] B. E. El-Demerdash, I. A. El-Khodary, & A. A. Tharwat "Developing a stochastic input oriented data envelopment analysis (SIODEA) model," *International Journal of Advanced Computer Science and Applications*, 4(4), 2013.
- [14] A. Emrouznejad, & G. L. Yang "A survey and analysis of the first 40 years of scholarly literature in DEA: 1978-2016," *Socio-Economic Planning Sciences*, 61, 4-8, 2018.
- [15] H. L. Gan, & E. D. Kolaczyk, "Approximation of the difference of two Poisson-like counts by Skellam," *Journal of Applied Probability*, 55(2), 416-430, 2018.
- [16] S. J. Janke, & F. Tinsley, "Introduction to linear models and statistical inference," John Wiley & Sons, 2005.
- [17] F. Jones "Lebesgue Integration on Euclidean Space," Jones and Bartlett publishers, pp. 527-529, 2001.
- [18] D. Karlis, & I. Ntzoufras "Bayesian analysis of the differences of count data," *Statistics in medicine*, 25(11), 1885-1905, 2006.
- [19] M. Khodabakhshi, M. Asgharian,] & G. N. Gregoriou "An input-oriented super-efficiency measure in stochastic data envelopment analysis: Evaluating chief executive officers of US public banks and thrifts," *Expert Systems with Applications*, 37(3), 2092-2097, 2010.
- [20] B. Khoshnevisan, H. M. Shariati, S. Rafiee & H. Mousazadeh "Comparison of energy consumption and GHG emissions of open field and greenhouse strawberry production" *Renewable and Sustainable Energy Reviews*, 29, 316- 324, 2014.
- [21] S. K. Lee, G. Mogi & K. S. Hui "A fuzzy analytic hierarchy process (AHP)/data envelopment analysis (DEA) hybrid model for efficiently allocating energy R&D resources: In the case of energy technologies against high oil prices," *Renewable and Sustainable Energy Reviews*, 21, 347-355, 2013.
- [22] E. L. Lehmann *Testing Statistical Hypotheses* second edition John Wiley and sons, 1986.
- [23] D. G. T. Reddy "Comparison and Correlation Coefficient between CRS and VRS models of OC Mines," *International Journal of Ethics in Engineering & Management Education*, 2(1), 2348-4748, 2015.
- [24] N. Mourad SIODEA with Normal and Poisson variables Source code, (Version 1.0), <https://github.com/nahiamourad/DEA>, 2019.
- [25] J. G. Skellam "The frequency distribution of the difference between two Poisson variates belonging to different populations," *Journal of the Royal Statistical Society. Series A (General)*, 109(Pt 3), 296-296, 1946.
- [26] L. Simar & V. Zelenyuk "Stochastic FDH/DEA estimators for frontier analysis," *Journal of Productivity Analysis*. 36 (1): 1-20, 2011.
- [27] T. Song, Z. Yang, & T. Chahine "Efficiency evaluation of material and energy flows, a case study of Chinese cities," *Journal of Cleaner Production*, 112, 3667-3675, 2016.
- [28] X. P. Zhang, Y. X. Zhang, R. Rao & Z. P. Shi "Exploring the drivers to energy-related carbon emissions changes at Chinas provincial levels," *Energy Efficiency*, 8(4), 699-712, 2015.
- [29] J. Zhu *Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheets* (Vol. 213). Springer, 2014.