

INVERSE DATA ENVELOPMENT ANALYSIS WITH STOCHASTIC DATA

ALI GHOMI¹, SAEID GHOBADI^{2,*}, MOHAMMAD HASSAN BEHZADI¹ AND
MOHSEN ROSTAMY-MALKHALIFEH³

Abstract. The inverse Data Envelopment Analysis (InvDEA) is an exciting and significant topic in the DEA area. Also, uncertain data in various real-life applications can degrade the efficiency results. The current work addresses the InvDEA in the presence of stochastic data. Under maintaining the efficiency score, the inputs/outputs-estimation problem is investigated when some or all of its outputs/inputs increase. A novel optimality concept for multiple-objective programming problems, stochastic (weak) Pareto optimality in the level of significance $\alpha \in [0, 1]$, is introduced to derive necessary and sufficient conditions for input/output estimation. Furthermore, the performance of the developed theory in a banking sector application is verified.

Mathematics Subject Classification. 90C05, 90C29, 90C39, 90C90, 90B50, 47N10.

Received March 10, 2021. Accepted August 27, 2021.

1. INTRODUCTION

Data envelopment analysis (DEA), as an excellent nonparametric efficiency measurement method, has been employed to determine the efficiency scores of a class of homogeneous decision-making units (DMUs) under several inputs and outputs [7, 13]. DEA has been first proposed by Charnes *et al.* [4] and is based on Farrell [16] studies. DEA has been extensively utilized in several areas and employed in various applications, including water, energy, power, banking, agriculture, mining, finance, transportation, communication, sport, and education [33]. It is worth noting that although DEA is one of the most powerful tools used for efficiency evaluation, some researchers debate on the three main components of the DEA technique as follows [3, 34, 35]: (i) DEA is a nonparametric technique, so the usual conclusions comparing with a parametric function are impossible. In other words, formulation for any parametric function related to the production including cost, or profit function evaluation of marginal products, marginal costs, and partial elasticities is impossible almost. (ii) The conventional DEA is based on linear programming (LP). One of the important assumptions of these models is the explicit definition of constraints and the positive Lagrangian coefficients. But regression analysis is a linear model based on defined probability space with the estimated parameters and without any other constraints on the multipliers. (iii) According to the concepts of DEA, there is no chance for random noise in nature. In other

Keywords. Data envelopment analysis, inverse DEA, stochastic data, efficiency, multiple-objective programming.

¹ Department of Statistics, Science and Research Branch, Islamic Azad University, Tehran, Iran.

² Department of Mathematics, Khomeinishahr Branch, Islamic Azad University, Isfahan, Iran.

³ Department of Mathematics, Science and Research Branch, Islamic Azad University, Tehran, Iran.

*Corresponding author: ghobadi@iauakhsh.ac.ir

words, when a deviation from the frontier is observed, it is called inefficiency in the traditional DEA. The third component is the main reason for our study.

The inverse DEA (InvDEA) is a primary and significant problem in the DEA area, where its idea has been initially introduced by Zhang and Cui [47]. In this work, the DMU's input increments are determined for its corresponding output increments in the presence of CCR efficiency-fixed constraints. In InvDEA, an efficiency level is chosen as a strategic purpose, and the primary purpose is to estimate the necessary input and output levels to attain this desired efficiency level. In contrast, in conventional DEA models, the DMU's efficiency score is usually determined under given input and output levels. Based on Ghobadi and Jahangiri [22] studies, the multiple-objective programming (MOP) tools have been utilized to estimate inputs (outputs) under growing outputs (inputs) and maintaining the efficiency score [23, 45]. The previous studies on DEA verified different theoretical and practical perspectives of the InvDEA, such as merging units [17, 46], restructuring units [1], and preserving efficiency [42].

Data uncertainty is a primary and significant problem in employing the traditional DEA. Although conventional DEA models assess the DMUs with crisp data, the DMUs' evaluation under imprecise or vague data by an analytical DEA structure called uncertain DEA (UDEA) has been recently considered. Data reliability in conventional DEA models cannot be attained in some practical situations. Indeed, a variety of real-life applications encounter uncertain data, which can affect the efficiency results. Different strategies have been presented in the literature to deal with these inaccurate and ambiguous data. According to Peykani *et al.* [38] studies, the uncertain theory-based approaches in the UDEA could be categorized into the following five classes: (i) The fuzzy DEA (FDEA) approach; (ii) The bootstrap DEA approach; (iii) The imprecise DEA approach; (iv) The robust DEA (RDEA) approach; and (v) The stochastic DEA (SDEA) approach. Fuzzy DEA (FDEA) is utilized under the vague inputs/outputs situation. Namely, the FDEA approach is employed when input/output data contain fuzzy numbers or linguistic variables [37]. The FDEA literature is divided into six categories, including the tolerance method, the fuzzy ranking method, the α -level based method, the possibility method, the fuzzy arithmetic method, and the fuzzy random/type-2 fuzzy set method [12]. Bootstrap DEA strategy is utilized in cases where the number of DMUs is not adequate, and the estimated efficient frontier is different from the actual one [41]. The imprecise DEA approach is utilized in situations where inputs/outputs include inaccurate data in the form of the interval, ratio bounded, and weak and strong ordinals [48]. As a novel strategy in the UDEA literature, RDEA is based on the RO method, an appropriate and common approach employed to deal with discrete and continuous optimization problems. This strategy does not require notable historical data and probability distribution function and ensures the solution feasibility for all feasible amounts of uncertain variables in the considered convex uncertainty set [38]. Stochastic DEA (SDEA) is the last UDEA method utilized for the performance evaluation of DMUs under uncertain data. In various realistic situations, managers deal with units with inaccurate data. In the mentioned cases, analyzers may assume inaccurate data as stochastic parameters. However, various data features can be discovered while working with stochastic variables assuming the possibility of unpredicted events. The primary benefit of working with stochastic data in DEA is the ability to predict the upcoming efficiencies. According to the DEA literature, different theoretical and practical perspectives of the SDEA models have been investigated by various researchers, such as Huang and Li [25], Jahanshahloo *et al.* [26], Li [31] and Sengupta [39]. It should be noted that most of the constructed models are based on nonlinear programming. In a new and different study, Behzadi and Mirbolouki [2] considered symmetric error structure for inputs and outputs and presented an envelopment stochastic input-oriented DEA model under constant returns to scale (CRS) assumption of the production technology. The main advantage of the mentioned model is its linearity. This model can calculate the relative efficiency of units at a significant specific level and identify efficient DMUs.

It is notable that the aim of a conventional InvDEA model is to estimate the input and output levels are deterministic types, which does not take the random errors of data into account in the production process. Due to the inherent complexity and competition of the real world, InvDEA problems often involve some random or ordinal variables. However, previous InvDEA studies have been undertaken in a deterministic environment, which cannot answer such situations. Therefore, it is necessary to incorporate the stochastic data into InvDEA.

It is worth noting that the stochastic InvDEA estimates inputs and outputs such that production frontiers that incorporate both inefficiency and stochastic error, which moves the frontiers closer to the bulk of the producing units. Then, the estimated inputs and outputs of DMUs are improved comparing to the deterministic InvDEA models. The main contribution of this paper is to provide a theoretical and practical framework of the stochastic InvDEA with random data.

This study presents a theoretical and practical framework of InvDEA under random data and focuses on the key thematic areas of resource allocation management, handling investment analysis problems, the implication of new performance strategies, and the motivations. In other words, InvDEA models are proposed in this study to manage resource allocation and investment analysis issues. Our proposed models are constructed to estimate inputs and output changes to keep the decision-making unit's efficiency score and other DMUs unchanged. InvDEA models are constructed for the first time to deal with stochastic data. The managers can employ these models to detect efficient DMUs from inefficient ones and design useful approaches to improve the efficiency score. These approaches can be considered as new approaches in resource allocation management and handling investment analysis problems. It is worth noting that these approaches do not change the efficiency scores. Moreover, a novel optimality principle is presented in this study for stochastic MOP (SMOP) problems, called stochastic (weak) Pareto optimality in the level of significance $\alpha \in [0, 1]$. The mentioned points can be utilized to solve the assumed InvDEA problems. A banking sector application is provided to demonstrate the credibility and capability of the presented approach in a practical situation. The main contributions of the current work are summarized as follows:

- For the first time, an InvDEA model is constructed under random data.
- For the first time, the inverse input-oriented SDEA model is constructed and employed for resource allocation problems.
- Similarly, an inverse output-oriented SDEA model is constructed to deal with investment analysis problems.
- A novel optimality principle is presented for SMOP, named stochastic (weak) Pareto optimality in the level of significance $\alpha \in [0, 1]$.
- Stochastic (weak) Pareto optimality in the level of significance $\alpha \in [0, 1]$ is utilized in InvDEA.

The article's remainder is organized as follows: Section 2 introduces the literature review on SDEA. Section 3 presents some principles of SDEA. The fundamental results of this work are stated in Section 4. The mentioned section offers new inverse models. A banking sector example is provided in Section 5 to demonstrate the performance of the developed theory. Finally, a short conclusion and future work aspects are presented in Section 6.

2. LITERATURE REVIEW ON SDEA

In the conventional DEA models, deterministic data are considered. However, in practical applications, we may face some stochastic data, including costs, prices, and production, depending on the financial crisis, social and political features, inflation, and natural components. In other words, due to the intrinsic uncertainties in many practical applications, stochastic estimations are inevitable in the efficiency evaluation procedure. Therefore, the expansion of DEA models can be considered a necessity. SDEA contains efficiency analysis strategies based on production economic principles like convexity or ray unboundedness of the included stochastic inputs and outputs. Moreover, they employ statistical principles or distributional considerations to estimate stochastic inefficiency rather than a deterministic or a stochastic production possibility set (PPS).

According to Olesen and Petersen [36] studies, the SDEA theory-based approaches can be classified into the following two categories: (i) The semi-parametric stochastic frontier analysis (SFDEA) approach; and (ii) The chance-constrained DEA (CCDEA) approach. The stochastic frontier analysis is a regression analysis-based method in which econometric models like Cobb–Douglas production functions are utilized to predict the production frontier [30]. CCDEA approach is usually utilized for stochastic inputs/outputs with a determined probability distribution function for the uncertain parameters [5]. To deal with stochastic data in DEA, Thore

[43] developed DEA-based chance-constrained programming. Cooper *et al.* [6] explicitly modeled a stochastic PPS to propose an SDEA method. Cooper *et al.* [5] generalized DEA models with inputs and outputs described with normal random variables. Moreover, they presented stochastic efficient DMU. Li [31] and Huang and Li [25] introduced the efficiency dominance of a DMU using probabilistic comparisons of inputs and outputs with other DMUs, which are evaluated through solving a chance-constrained programming problem. Then, some DEA models have been constructed with stochastic data by various researchers, such as Khodabakhshi [29], Behzadi *et al.* [3], Lotfi *et al.* [34] and Hosseinzadeh Lotfi *et al.* [24].

Behzadi and Mirbolouki [2] introduced the symmetric error structure and proposed a linear form of stochastic CCR model. In the current work, the inverse version of the SDEA model addressed in Behzadi and Mirbolouki [2] is considered to extract necessary and sufficient conditions for input and output estimation under preserving efficiency score.

CCDEA and semi-parametric SFDEA seems to be competing methods within the SDEA field. SFDEA's proponents may claim that this method is superior to CCDEA due to its consistency with essential principles in production theory and its strong statistical basis, including a basic data production procedure. In contrast, CCDEA's proponents may claim this method's advantage since DMU-specific distributions of noise and inefficiency can be easily incorporated, and multiple inputs/outputs can be explicitly handled. Moreover, a recent study performed by Simar [40] provides a data production procedure for a multivariate, cross-sectional, non-parametric stochastic frontier model employed in the CCDEA context. According to Olesen and Petersen [36] studies, although the stochastic PPS estimators in the context of SFDEA and CCDEA have different features in the modeling of noise and stochastic inefficiency, all have one output and multiple inputs.

3. STOCHASTIC DEA

Consider a set of n DMUs, $\{\text{DMU}_j : j \in J = \{1, 2, \dots, n\}\}$, such that DMU_j produces output level $\tilde{Y}_j = (\tilde{y}_{1j}, \tilde{y}_{2j}, \dots, \tilde{y}_{sj})^t$ by consuming input level $\tilde{X}_j = (\tilde{x}_{1j}, \tilde{x}_{2j}, \dots, \tilde{x}_{mj})^t$. Let the input and output levels of DMU_j , $j \in J$, be normal-stochastic numbers as follows:

$$\begin{aligned} \tilde{x}_{ij} &\sim N(x_{ij}, \sigma_{ij}^2), & \forall i \in I = \{1, 2, \dots, m\}, \\ \tilde{y}_{rj} &\sim N(y_{rj}, \psi_{rj}^2), & \forall r \in O = \{1, 2, \dots, s\}, \end{aligned}$$

where x_{ij} and σ_{ij}^2 for all $i \in I$ and $j \in J$, are mean and variance of inputs, respectively. Also, y_{rj} and ψ_{rj}^2 for all $r \in O$ and $j \in J$, are mean and variance of outputs, respectively. It is worth noting that the mean and variance of data are rarely unknown. Therefore, we try to estimate these parameters. For this purpose, it is a suitable estimator, if the estimator is an unbiased and efficient estimator. According to the stochastic DEA literature, if $\{(\hat{x}_{ij}^t, \hat{y}_{rj}^t) \mid i \in I, r \in O, t = 1, 2, \dots, p\}$ is a random sample with size p for the DMU_j , $j \in J$, then

$$\begin{aligned} \bar{x}_{ij} &= \frac{1}{p} \sum_{t=1}^p \hat{x}_{ij}^t, \\ \bar{\sigma}_{ij}^2 &= \frac{1}{p-1} \sum_{t=1}^p (\hat{x}_{ij}^t - \bar{x}_{ij})^2, \end{aligned}$$

are unbiased estimators for " x_{ij} " and " σ_{ij}^2 ", and

$$\begin{aligned} \bar{y}_{rj} &= \frac{1}{p} \sum_{t=1}^p \hat{y}_{rj}^t, \\ \bar{\psi}_{rj}^2 &= \frac{1}{p-1} \sum_{t=1}^p (\hat{y}_{rj}^t - \bar{y}_{rj})^2, \end{aligned}$$

are unbiased estimators for “ y_{rj} ” and “ ψ_{rj}^2 ”. Clearly, if the sample size is large enough, then \bar{x}_{ij} (\bar{y}_{rj}) and $\bar{\sigma}_{ij}^2$ ($\bar{\psi}_{rj}^2$) are the best estimators for mean and variance of inputs (outputs).

The following input-oriented SDEA problem is considered to estimate the relative efficiency in the level of significance α of the unit under assessment DMU_o , $o \in J$ [8]:

$$\begin{aligned} \theta_o(\alpha) = \min \quad & \theta \\ \text{s.t.} \quad & P \left\{ \sum_{j \in J} \mu_j \tilde{X}_j \leq \theta \tilde{X}_o \right\} \geq 1 - \alpha, \\ & P \left\{ \sum_{j \in J} \mu_j \tilde{Y}_j \geq \tilde{Y}_o \right\} \geq 1 - \alpha, \\ & \mu \in \Omega, \end{aligned} \quad (3.1)$$

where

$$\Omega = \left\{ \mu \mid \mu = (\mu_1, \dots, \mu_n), \delta_1 \left(\sum_{j \in J} \mu_j + \delta_2 (-1)^{\delta_3} \nu \right) = \delta_1, \nu \geq 0, \mu_j \geq 0, j \in J \right\}.$$

In this model, $(\mu, \theta) \in \mathbb{R}^n \times \mathbb{R}$ is the variables vector. Also, $\alpha \in [0, 1]$ is a predefined error level and P means “Probability”. Moreover, $\delta_1, \delta_2, \delta_3 \in \{0, 1\}$. Clearly, the production technology is under the following conditions:

- (i) Constant returns to scale (CRS) if $\delta_1 = 0$;
- (ii) Variable returns to scale (VRS) if $\delta_1 = 1, \delta_2 = 0$;
- (iii) Non-increasing returns to scale (NIRS) if $\delta_1 = \delta_2 = 1$ and $\delta_3 = 0$;
- (iv) Non-decreasing returns to scale (NDRS) if $\delta_1 = \delta_2 = \delta_3 = 1$.

According to the DEA literature [8], if Φ is cumulative standard normal distribution function, then the stochastic model (3.1) can be simplified to the following non-linear problem:

$$\begin{aligned} \theta_o(\alpha) = \min \quad & \theta \\ \text{s.t.} \quad & \sum_{j \in J} \mu_j x_{ij} - \Phi^{-1}(\alpha) \sigma_i^I(\mu, \theta) \leq \theta x_{io}, \quad \forall i \in I, \\ & \sum_{j \in J} \mu_j y_{rj} + \Phi^{-1}(\alpha) \sigma_r^O(\mu) \geq y_{ro}, \quad \forall r \in O, \\ & \mu \in \Omega, \end{aligned} \quad (3.2)$$

where

$$\begin{aligned} (\sigma_i^I(\mu, \theta))^2 &= \sum_{j \in J} \sum_{k \in J} \mu_j \mu_k \text{Cov}(\tilde{x}_{ij}, \tilde{x}_{ik}) + \theta^2 \text{Var}(\tilde{x}_{io}) - 2\theta \sum_{j \in J} \mu_j \text{Cov}(\tilde{x}_{ij}, \tilde{x}_{io}), \quad \forall i \in I, \\ (\sigma_r^O(\mu))^2 &= \sum_{j \in J} \sum_{k \in J} \mu_j \mu_k \text{Cov}(\tilde{y}_{rj}, \tilde{y}_{rk}) + \text{Var}(\tilde{y}_{ro}) - 2 \sum_{j \in J} \mu_j \text{Cov}(\tilde{y}_{rj}, \tilde{y}_{ro}), \quad \forall r \in O. \end{aligned}$$

Also, it is clear that if all the input and output levels of the units are uncorrelated, then:

$$\begin{aligned} (\sigma_i^I(\mu, \theta))^2 &= \sum_{j \in J} \mu_j^2 \text{Var}(\tilde{x}_{ij})^2 + \theta \text{Var}(\tilde{x}_{io})^2 (\theta - 2\mu_o), \quad \forall i \in I, \\ (\sigma_r^O(\mu))^2 &= \sum_{j \in J} \mu_j^2 \text{Var}(\tilde{y}_{rj}) + \text{Var}(\tilde{y}_{ro})^2 (1 - 2\mu_o), \quad \forall r \in O. \end{aligned}$$

The output-version of the model (3.1) in the level of significance α is as follows [8]:

$$\begin{aligned} \varphi_o(\alpha) = \max \quad & \varphi \\ \text{s.t.} \quad & P \left\{ \sum_{j \in J} \mu_j \tilde{X}_j \leq \tilde{X}_o \right\} \geq 1 - \alpha, \\ & P \left\{ \sum_{j \in J} \mu_j \tilde{Y}_j \geq \varphi \tilde{Y}_o \right\} \geq 1 - \alpha, \\ & \mu \in \Omega. \end{aligned} \quad (3.3)$$

$(\mu, \varphi) \in \mathbb{R}^n \times \mathbb{R}$ is the variables vector in model (3.3). Considering Φ as cumulative standard distribution function, the stochastic model (3.3) can be simplified to model (3.4).

$$\begin{aligned} \varphi_o(\alpha) = \max \quad & \varphi \\ \text{s.t.} \quad & \sum_{j \in J} \mu_j x_{ij} - \Phi^{-1}(\alpha) \sigma_i^I(\mu) \leq x_{io}, \quad \forall i \in I, \\ & \sum_{j \in J} \mu_j y_{rj} + \Phi^{-1}(\alpha) \sigma_r^O(\mu, \varphi) \geq \varphi y_{ro}, \quad \forall r \in O, \\ & \mu \in \Omega, \end{aligned} \quad (3.4)$$

where

$$\begin{aligned} (\sigma_i^I(\mu))^2 &= \sum_{j \in J} \sum_{k \in J} \mu_j \mu_k \text{Cov}(\tilde{x}_{ij}, \tilde{x}_{ik}) + \text{Var}(\tilde{x}_{io}) - 2 \sum_{j \in J} \mu_j \text{Cov}(\tilde{x}_{ij}, \tilde{x}_{io}), \quad \forall i \in I, \\ (\sigma_r^O(\mu, \varphi))^2 &= \sum_{j \in J} \sum_{k \in J} \mu_j \mu_k \text{Cov}(\tilde{y}_{rj}, \tilde{y}_{rk}) + \varphi^2 \text{Var}(\tilde{y}_{ro}) - 2\varphi \sum_{j \in J} \mu_j \text{Cov}(\tilde{y}_{rj}, \tilde{y}_{ro}), \quad \forall r \in O. \end{aligned}$$

If $\varphi_o(\alpha) = 1$ (resp. $\theta_o(\alpha) = 1$), then DMU_o is called stochastic output (resp. input-) oriented weakly efficient in the level of significance α . Also, Cooper *et al.* [8] established the following theorem.

Theorem 3.1. Consider models (3.2) and (3.4):

- (i) These models are feasible for all the levels of significance α .
- (ii) If $\alpha \leq 0.5$ then $0 < \theta_o(\alpha) \leq 1$ and $\varphi_o(\alpha) \geq 1$.
- (iii) If $\alpha \leq \alpha'$ then $\theta_o(\alpha') \leq \theta_o(\alpha)$ and $\varphi_o(\alpha') \geq \varphi_o(\alpha)$.

Remark 3.2. Assume that the input and output levels of the DMUs have a symmetric error structure. In fact, the following structure is considered for the inputs and outputs of DMU_j , $j \in J$:

$$\begin{aligned} \tilde{x}_{ij} &= x_{ij} + a_{ij} \tilde{\varepsilon}_{ij}, & \forall i \in I, \\ \tilde{y}_{rj} &= y_{rj} + b_{rj} \tilde{\xi}_{rj}, & \forall r \in O, \end{aligned} \quad (3.5)$$

where $a_{ij} \in \mathbb{R}_{\geq 0}$, $b_{rj} \in \mathbb{R}_{\geq 0}$, $\tilde{\varepsilon}_{ij} \sim N(0, \bar{\sigma}^2)$, and $\tilde{\xi}_{rj} \sim N(0, \bar{\sigma}^2)$. In fact, $\tilde{\varepsilon}_{ij}$ and $\tilde{\xi}_{rj}$ are errors of input and output levels in conflict with the mean values, respectively. If the input and output vectors of the units are uncorrelated and $\tilde{\varepsilon}_{ij} = \tilde{\varepsilon}_i$ and $\tilde{\xi}_{rj} = \tilde{\xi}_r$ for all $i \in I$ and $r \in O$, then models (3.2) and (3.4) can be transformed to linear programming (LP) models (3.6) and (3.7), respectively [2]:

$$\theta_o(\alpha) = \min \quad \theta \quad (3.6)$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{j \in J} \mu_j x_{ij} - \Phi^{-1}(\alpha) \bar{\sigma}(p_i^+ + p_i^-) \leq \theta x_{io}, & \forall i \in I, \\ & \sum_{j \in J} \mu_j y_{rj} + \Phi^{-1}(\alpha) \bar{\sigma}(q_r^+ + q_r^-) \geq y_{ro}, & \forall r \in O, \\ & \sum_{j \in J} \mu_j a_{ij} - \theta a_{io} = p_i^+ - p_i^-, & \forall i \in I, \\ & \sum_{j \in J} \mu_j b_{rj} - b_{ro} = q_r^+ - q_r^-, & \forall r \in O, \\ & \mu \in \Omega, p_i^+ \geq 0, p_i^- \geq 0, q_r^+ \geq 0, q_r^- \geq 0, & \forall i \in I, \forall r \in O, \end{aligned}$$

$$\varphi_o(\alpha) = \max \quad \varphi \quad (3.7)$$

$$\begin{aligned} \text{s.t.} \quad & \sum_{j \in J} \mu_j x_{ij} - \Phi^{-1}(\alpha) \bar{\sigma}(p_i^+ + p_i^-) \leq x_{io}, & \forall i \in I, \\ & \sum_{j \in J} \mu_j y_{rj} + \Phi^{-1}(\alpha) \bar{\sigma}(q_r^+ + q_r^-) \geq \varphi y_{ro}, & \forall r \in O, \\ & \sum_{j \in J} \mu_j a_{ij} - a_{io} = p_i^+ - p_i^-, & \forall i \in I, \\ & \sum_{j \in J} \mu_j b_{rj} - \varphi b_{ro} = q_r^+ - q_r^-, & \forall r \in O, \\ & \mu \in \Omega, p_i^+ \geq 0, p_i^- \geq 0, q_r^+ \geq 0, q_r^- \geq 0, & \forall i \in I, \forall r \in O. \end{aligned}$$

In the models (3.6) and (3.7), φ , θ , μ_j , p_i^+ , p_i^- , q_r^+ , and q_r^- are variables for all indices.

4. STOCHASTIC INVDEA

The DMU's efficiency score is usually determined under given input and output levels in conventional DEA models. In contrast, in InvDEA, an efficiency target is chosen as a strategic purpose, and the primary purpose is to find the necessary input and output vectors to attain this desired efficiency target. Initially, the following significant problem has been studied in the framework of InvDEA [23, 45]:

Question. If the efficiency index of DMU_o remains unchanged with respect to other DMUs, but the input/output levels increase, to what extent should the output/input levels of DMU_o increase?

This Question was solved using MOP problems by Wei *et al.* [45] and Hadi-Vencheh *et al.* [23]. Different theoretical and practical perspectives of the InvDEA has been studied in the DEA literature, such as setting revenue target [10], investment analysis [44], and resource allocation [14] under inter-temporal dependence of data [19, 20, 28], and fuzzy data [18].

Nevertheless, the constructed models in the mentioned works could not solve the issue of input-estimation and output-estimation in the presence of stochastic data. Therefore, Sections 4.1 and 4.2 attempt to estimate the input/output levels in the presence of stochastic data.

4.1. Estimation of input levels

This sub-section extends the following question (proposed by [23]) in a stochastic framework.

Question 1. If the efficiency index of DMU_o remains unchanged with respect to other DMUs, but the output levels increase, to what extent should the input levels of DMU_o increase?

Suppose that the output levels of DMU_o are increased from \tilde{Y}_o to $\tilde{\beta}_o = \tilde{Y}_o + \Delta\tilde{Y}_o$, such that $\Delta\tilde{Y}_o \geq 0$, $\Delta\tilde{Y}_o \neq 0$. The goal of Question 1 is to find the input levels $\tilde{\alpha}_o = (\tilde{\alpha}_{1o}, \tilde{\alpha}_{2o}, \dots, \tilde{\alpha}_{mo})$ on the condition that the efficiency score of DMU_o in the level of significance α is still $\theta_o(\alpha)$. More accurately, $\tilde{\alpha}_o = \tilde{X}_o + \Delta\tilde{X}_o$, where $\Delta\tilde{X}_o \geq 0$.

Assume DMU_{n+1} indicates the entity generated after increasing the input and output levels of DMU_o . The following model estimates the efficiency index of DMU_{n+1} in the level of significance α :

$$\begin{aligned} \theta_{n+1}(\alpha) &= \min \quad \theta & (4.1) \\ \text{s.t.} \quad & P \left\{ \sum_{j \in J} \mu_j \tilde{X}_j + \mu_{n+1} \tilde{\alpha}_o \leq \theta \tilde{\alpha}_o \right\} \geq 1 - \alpha, \\ & P \left\{ \sum_{j \in J} \mu_j \tilde{Y}_j + \mu_{n+1} \tilde{\beta}_o \geq \tilde{\beta}_o \right\} \geq 1 - \alpha, \\ & \mu \in \Omega^+, \\ & \Omega^+ = \left\{ \mu \mid \mu = (\mu_1, \dots, \mu_n, \mu_{n+1}), \delta_1 \left(\sum_{j \in J} \mu_j + \mu_{n+1} + \delta_2 (-1)^{\delta_3} \nu \right) = \delta_1, \nu \geq 0, \mu_j \geq 0, j \in J \cup \{n+1\} \right\}. \end{aligned}$$

If the optimal values of the problems (3.1) and (4.1) are the same in the level of significance α , it is said that the efficiency index remains unchanged, *i.e.*

$$\text{eff}(\tilde{\alpha}_o, \tilde{\beta}_o)(\alpha) = \text{eff}(\tilde{X}_o, \tilde{Y}_o)(\alpha)^1.$$

The following SMOP problem is proposed in the level of significance α to answer the above question or find the input levels.

$$\begin{aligned} \min & (\tilde{\alpha}_{1o}, \tilde{\alpha}_{2o}, \dots, \tilde{\alpha}_{mo}) & (4.2) \\ \text{s.t.} \quad & P \left\{ \sum_{j \in J} \mu_j \tilde{X}_j \leq \theta_o(\alpha) \tilde{\alpha}_o \right\} \geq 1 - \alpha, \\ & P \left\{ \sum_{j \in J} \mu_j \tilde{Y}_j \geq \tilde{\beta}_o \right\} \geq 1 - \alpha, \\ & P \left\{ \tilde{\alpha}_o \geq \tilde{X}_o \right\} \geq 1 - \alpha, \\ & \mu \in \Omega. \end{aligned}$$

In this model, $(\mu, \tilde{\alpha}_o)$ is the variables vector. $\theta_o(\alpha)$ is a constant, which represents the optimal value of variable θ in the problem (3.1). Additionally, $\tilde{\beta}_o$ is known and fixed.

A novel optimality concept is defined for the SMOP (4.2) based on the special structure of this problem.

Definition 4.1. A feasible solution $\Delta = (\mu^*, \tilde{\alpha}_{1o}^*, \tilde{\alpha}_{2o}^*, \dots, \tilde{\alpha}_{mo}^*)$ is called a stochastic Pareto (SP) solution in the level of significance α to SMOP (4.2) if there is no other feasible solution $\Gamma = (\mu, \tilde{\alpha}_{1o}, \tilde{\alpha}_{2o}, \dots, \tilde{\alpha}_{mo})$ such that

$$P \{ \tilde{\alpha}_{io} - \tilde{\alpha}_{io}^* \leq 0 \} \geq 1 - \alpha \quad \text{for each } i = 1, 2, \dots, m,$$

¹After this, the notation “eff” is employed instead of “efficiency” for simplicity.

$$P \{ \tilde{\alpha}_{io} - \tilde{\alpha}_{io}^* \leq -\epsilon \} \geq 1 - \alpha \quad \text{for some } i = 1, 2, \dots, m,$$

where ϵ is a non-archimedean infinitesimal.

Definition 4.2. A feasible solution $\Delta = (\mu^*, \tilde{\alpha}_{1o}^*, \tilde{\alpha}_{2o}^*, \dots, \tilde{\alpha}_{mo}^*)$ is called a stochastic weak Pareto (SWP) solution in the level of significance α to SMOP (4.2) if there is no other feasible solution $\Gamma = (\mu, \tilde{\alpha}_{1o}, \tilde{\alpha}_{2o}, \dots, \tilde{\alpha}_{mo})$ such that

$$P \{ \tilde{\alpha}_{io} - \tilde{\alpha}_{io}^* \leq -\epsilon \} \geq 1 - \alpha \quad \text{for each } i = 1, 2, \dots, m,$$

where ϵ is a non-Archimedean infinitesimal.

Theorem 4.3 shows how the model SMOP 4.2 can be applied to answer Question 1 in the stochastic framework.

Theorem 4.3. Suppose that $\Delta = (\mu^*, \tilde{\alpha}_o^*)$ is a SP solution in the level of significance α to model (4.2) provided that one of the following conditions holds:

- (i) $P \{ |\tilde{\alpha}_o^* - \tilde{X}_o| < \epsilon \} \geq 1 - \alpha$ where ϵ is a non-Archimedean infinitesimal.
- (ii) There exists at least $l \in I$ such that $P \{ \tilde{\alpha}_{lo}^* - \tilde{x}_{lo} \geq \epsilon \} \geq 1 - \alpha$ where ϵ is a non-Archimedean infinitesimal.

$$\text{Then } \text{eff} \left(\tilde{\alpha}_o^*, \tilde{\beta}_o \right) (\alpha) = \text{eff} \left(\tilde{X}_o, \tilde{Y}_o \right) (\alpha).$$

Proof. See Appendix A. □

Theorem 4.4 is a converse version of Theorem 4.3.

Theorem 4.4. Let $\Pi = (\bar{\mu}, \bar{\alpha}_o = (\bar{\alpha}_{1o}, \bar{\alpha}_{2o}, \dots, \bar{\alpha}_{mo}))$ be a feasible solution to SMOP (4.2) such that $\text{eff} \left(\bar{\alpha}_o, \tilde{\beta}_o \right) (\alpha) = \text{eff} \left(\tilde{X}_o, \tilde{Y}_o \right) (\alpha)$. Then, Π is a SWP solution in the level of significance α to SMOP (4.2).

Proof. If Π not be a SWP solution in the level of significance α to problem (4.2), then there is another feasible solution to problem (4.2), $\Delta = (\mu, \tilde{\alpha}_o = (\tilde{\alpha}_{1o}, \tilde{\alpha}_{2o}, \dots, \tilde{\alpha}_{mo}))$ in which,

$$P \{ \tilde{\alpha}_o - \bar{\alpha}_o \leq -\epsilon \} \geq 1 - \alpha, \quad (4.3)$$

where ϵ is a non-Archimedean infinitesimal. Feasibility of Δ for SMOP (4.2), implies

$$P \left\{ \sum_{j \in J} \mu_j \tilde{X}_j - \theta_o(\alpha) \tilde{\alpha}_o \leq 0 \right\} \geq 1 - \alpha, \quad (4.4)$$

$$P \left\{ \sum_{j \in J} \mu_j \tilde{Y}_j \geq \tilde{\beta}_o \right\} \geq 1 - \alpha, \quad (4.5)$$

$$\mu \in \Omega. \quad (4.6)$$

By equations (4.3) and (4.4), we get

$$P \left\{ \sum_{j \in J} \mu_j \tilde{X}_j - \theta_o(\alpha) (\bar{\alpha}_o + \epsilon) \leq 0 \right\} \geq 1 - \alpha. \quad (4.7)$$

Therefore, there is a positive scalar $0 < k_o < 1$ in which

$$P \left\{ \sum_{j \in J} \mu_j \tilde{X}_j - (k_o \theta_o(\alpha)) \bar{\alpha}_o \leq 0 \right\} \geq 1 - \alpha. \quad (4.8)$$

According to equations (4.5), (4.6), and (4.8), the vector $((\mu, \mu_{n+1} = 0), k_o \theta_o(\alpha))$ is a feasible solution to problem (4.1). The value of the goal function of problem (4.1) in this feasible solution is equal to $k_o \theta_o(\alpha)$. Then, $\theta_{n+1}(\alpha) \leq k_o \theta_o(\alpha) < \theta_o(\alpha)$. This contradicts the assumption and therefore completes the proof of the theorem. \square

We close this sub-section by discussing when to start using InvDEA. Suppose that the output levels of DMU_o are increased from \tilde{Y}_o to $\tilde{\beta}_o$. We consider the following assumption in the level of significance α :

$$\begin{cases} H_0 : & \theta_{n+1}(\alpha) = \theta_o(\alpha), \\ H_1 : & \theta_{n+1}(\alpha) \neq \theta_o(\alpha), \end{cases}$$

or equivalently,

$$\begin{cases} H_0 : & \tilde{\alpha}_o = \tilde{X}_o, \\ H_1 : & \tilde{\alpha}_o \neq \tilde{X}_o. \end{cases}$$

By referring to the concepts of statistics, a $(1 - \alpha)\%$ confidence interval for the mean inputs and the mean outputs are as follows:

$$\begin{aligned} x_{ij} &\in \left[\bar{x}_{ij} - t_{(\frac{\alpha}{2})} \frac{\bar{\sigma}_{ij}}{\sqrt{p}}, \bar{x}_{ij} + t_{(\frac{\alpha}{2})} \frac{\bar{\sigma}_{ij}}{\sqrt{p}} \right], & \forall i \in I, j \in J, \\ \beta_{rj} &\in \left[\bar{\beta}_{rj} - t_{(\frac{\alpha}{2})} \frac{\bar{\psi}_{rj}}{\sqrt{p}}, \bar{\beta}_{rj} + t_{(\frac{\alpha}{2})} \frac{\bar{\psi}_{rj}}{\sqrt{p}} \right], & \forall r \in O, j \in J, \end{aligned}$$

where \bar{x}_{ij} and $\bar{\sigma}_{ij}^2$ are the mean and variance of inputs, respectively. Also, $\bar{\beta}_{rj}$ and $\bar{\psi}_{rj}^2$ are the mean and variance of outputs, respectively. Moreover, $t_{(\frac{\alpha}{2})}$ is a random variable with distribution T and degree of freedom $p - 1$. Therefore, the lower and upper bounds of inputs and outputs are as follows:

$$\begin{aligned} x_{ij}^l(\alpha) &= \bar{x}_{ij} - t_{(\frac{\alpha}{2})} \frac{\bar{\sigma}_{ij}}{\sqrt{p}}, & x_{ij}^u(\alpha) &= \bar{x}_{ij} + t_{(\frac{\alpha}{2})} \frac{\bar{\sigma}_{ij}}{\sqrt{p}}, & \forall i \in I, j \in J, \\ \beta_{rj}^l(\alpha) &= \bar{\beta}_{rj} - t_{(\frac{\alpha}{2})} \frac{\bar{\psi}_{rj}}{\sqrt{p}}, & \beta_{rj}^u(\alpha) &= \bar{\beta}_{rj} + t_{(\frac{\alpha}{2})} \frac{\bar{\psi}_{rj}}{\sqrt{p}}, & \forall r \in O, j \in J. \end{aligned}$$

According to the pessimistic and optimistic viewpoints [9, 15, 21], the following models proposed to estimate the upper and lower bounds of efficiency of DMU_{n+1} with $(1 - \alpha)\%$ confidence interval:

$$\theta_{n+1}^l = \min \theta \quad (4.9)$$

$$\text{s.t.} \quad \sum_{j \in J} \lambda_j x_{ij}^l(\alpha) + \lambda_{n+1} x_{in+1}^u(\alpha) \leq \theta x_{in+1}^u(\alpha), \quad \forall i \in I,$$

$$\sum_{j \in J} \lambda_j \beta_{rj}^u(\alpha) + \lambda_{n+1} \beta_{rn+1}^l(\alpha) \geq \beta_{rn+1}^l(\alpha), \quad \forall r \in O,$$

$$\lambda \in \Omega^+,$$

$$\theta_{n+1}^u = \min \theta \quad (4.10)$$

$$\text{s.t.} \quad \sum_{j \in J} \lambda_j x_{ij}^u(\alpha) + \lambda_{n+1} x_{in+1}^l(\alpha) \leq \theta x_{in+1}^l(\alpha), \quad \forall i \in I,$$

$$\sum_{j \in J} \lambda_j \beta_{rj}^l(\alpha) + \lambda_{n+1} \beta_{rn+1}^u(\alpha) \geq \beta_{rn+1}^u(\alpha), \quad \forall r \in O,$$

$$\lambda \in \Omega^+.$$

It is obvious that $\theta_{n+1}(\alpha) \in [\theta_{n+1}^l(\alpha), \theta_{n+1}^u(\alpha)]$. With regard to model (4.1) (setting $\tilde{\alpha}_o = \tilde{X}_o$), if the optimal value of this model does not belong to $[\theta_{n+1}^l(\alpha), \theta_{n+1}^u(\alpha)]$, then the proposed method in this sub-section can be applied to estimating of the input levels such that the efficiency score of DMU_o remains unchanged in the level of significance α ($\theta_o(\alpha)$).

Remark 4.5. According to Remark 3.2, if all input and output levels of the units have a symmetric error structure, then the same method for converting problem (3.2) to the problem (3.6) can be employed to convert the stochastic models (4.1) and (4.2) to the LP and MOLP problems, respectively.

4.2. Estimation of output levels

This sub-section is devoted to extending the following question (provided by [45]) in a stochastic framework.

Question 2. If the efficiency index of DMU_o remains unchanged with respect to other DMUs, while the input levels increase, to what extent should the output levels of DMU_o increase?

Assume the input levels of DMU_o are increased from \tilde{X}_o to $\tilde{\alpha}_o = \tilde{X}_o + \Delta\tilde{X}_o$, where $\Delta\tilde{X}_o \geq 0$, $\Delta\tilde{X}_o \neq 0$. The goal of the study is to estimate the output levels $\tilde{\beta}_o = (\tilde{\beta}_{1o}, \tilde{\beta}_{2o}, \dots, \tilde{\beta}_{so})$ such that the efficiency index of DMU_o in the level of significance α is still $\varphi_o(\alpha)$. More accurately, $\tilde{\beta}_o = \tilde{Y}_o + \Delta\tilde{Y}_o$, where $\Delta\tilde{Y}_o \geq 0$.

Suppose DMU_{n+1} indicates the entity obtained after increasing the input and output levels of DMU_o . The following problem estimates the efficiency score of DMU_{n+1} in the level of significance α :

$$\begin{aligned} \varphi_{n+1}(\alpha) = \max \quad & \varphi \\ \text{s.t.} \quad & P \left\{ \sum_{j \in J} \mu_j \tilde{X}_j + \mu_{n+1} \tilde{\alpha}_o \leq \tilde{\alpha}_o \right\} \geq 1 - \alpha, \\ & P \left\{ \sum_{j \in J} \mu_j \tilde{Y}_j + \mu_{n+1} \tilde{\beta}_o \geq \varphi \tilde{\beta}_o \right\} \geq 1 - \alpha, \\ & \mu \in \Omega^+. \end{aligned} \tag{4.11}$$

If the optimal values of models (3.3) and (4.11) are equal, it is said that the efficiency index in the level of significance α remains unchanged, namely

$$\text{eff}(\tilde{\alpha}_o, \tilde{\beta}_o)(\alpha) = \text{eff}(\tilde{X}_o, \tilde{Y}_o)(\alpha).$$

To answer Question 2, namely to finding the output levels, the following SMOP model is suggested:

$$\begin{aligned} \max \quad & (\tilde{\beta}_{1o}, \tilde{\beta}_{2o}, \dots, \tilde{\beta}_{so}) \\ \text{s.t.} \quad & P \left\{ \sum_{j \in J} \mu_j \tilde{X}_j \leq \tilde{\alpha}_o \right\} \geq 1 - \alpha, \\ & P \left\{ \sum_{j \in J} \mu_j \tilde{Y}_j \geq \varphi_o(\alpha) \tilde{\beta}_o \right\} \geq 1 - \alpha, \\ & P \left\{ \tilde{\beta}_o \geq \tilde{Y}_o \right\} \geq 1 - \alpha, \\ & \mu \in \Omega, \end{aligned} \tag{4.12}$$

where $(\mu, \tilde{\beta}_o)$ is the variables vector in the above problem. $\varphi_o(\alpha)$ is a constant value, which describes the optimal value of variable φ in the problem (3.3). Moreover, $\tilde{\alpha}_o$ is known and fixed.

Theorem 4.6 shows how the model SMOP (4.12) can be applied to answer Question 2 in the stochastic framework.

Theorem 4.6. *If $\Delta = (\mu^*, \tilde{\beta}_o^* = (\tilde{\beta}_{1o}^*, \dots, \tilde{\beta}_{so}^*))$ is a SWP solution in the level of significance α to model (4.12), then $\text{eff}(\tilde{\alpha}_o, \tilde{\beta}_o^*)(\alpha) = \text{eff}(\tilde{X}_o, \tilde{Y}_o)(\alpha)$.*

Proof. See Appendix B. □

Theorem 4.7 is the converse version of Theorem 4.6.

Theorem 4.7. *Let $\Pi = (\bar{\mu}, \bar{\beta}_o = (\bar{\beta}_{1o}, \bar{\beta}_{2o}, \dots, \bar{\beta}_{so}))$ be a feasible solution to SMOP (4.12) such that $\text{eff}(\alpha_o, \bar{\beta}_o) = \text{eff}(\tilde{X}_o, \tilde{Y}_o)$. Then, Π is a SWP solution in the level of significance α to SMOP (4.12).*

Proof. If Π not be a SWP solution in the level of significance α to problem (4.12), then there is another feasible solution to model (4.12), $\Delta = (\mu, \tilde{\beta}_o = (\tilde{\beta}_{1o}, \tilde{\beta}_{2o}, \dots, \tilde{\beta}_{so}))$ in which,

$$P \left\{ \tilde{\beta}_o - \bar{\beta}_o \geq \epsilon \right\} \geq 1 - \alpha, \quad (4.13)$$

where ϵ is a non-Archimedean infinitesimal. Feasibility of Δ for SMOP (4.12), implies

$$P \left\{ \sum_{j \in J} \mu_j \tilde{X}_j \leq \alpha_o \right\} \geq 1 - \alpha, \quad (4.14)$$

$$P \left\{ \sum_{j \in J} \mu_j \tilde{Y}_j \geq \varphi_o(\alpha) \tilde{\beta}_o \right\} \geq 1 - \alpha, \quad (4.15)$$

$$\mu \in \Omega. \quad (4.16)$$

By equations (4.13) and (4.15), we get

$$P \left\{ \sum_{j \in J} \mu_j \tilde{Y}_j - \varphi_o(\alpha) (\bar{\beta}_o + \epsilon) \geq 0 \right\} \geq 1 - \alpha. \quad (4.17)$$

Therefore, there is a positive scalar $k_o > 1$ where

$$P \left\{ \sum_{j=1}^n \mu_j \tilde{Y}_j - (k_o \varphi_o(\alpha)) \bar{\beta}_o \geq 0 \right\} \geq 1 - \alpha. \quad (4.18)$$

According to equations (B.20), (4.16), and (4.18), the vector $((\mu, \mu_{n+1} = 0), k_o \varphi_o(\alpha))$ is a feasible solution to model (4.11). The value of the aim function of model (4.11) in this feasible solution is equal to $k_o \varphi_o(\alpha)$. Hence, $\varphi_{n+1}(\alpha) \geq k_o \varphi_o(\alpha) > \varphi_o(\alpha)$. This contradicts the assumption and so completes the proof of the theorem. □

In a similar discussion to the final part of Section 4.2, assume the input levels of DMU_o are increased from \tilde{X}_o to $\tilde{\alpha}_o$. We consider the following statistical assumption in the level of significance α :

$$\begin{cases} H_0 : & \varphi_{n+1}(\alpha) = \varphi_o(\alpha), \\ H_1 : & \varphi_{n+1}(\alpha) \neq \varphi_o(\alpha), \end{cases}$$

or equivalently,

$$\begin{cases} H_0 : & \tilde{\beta}_o = \tilde{Y}_o, \\ H_1 : & \tilde{\beta}_o \neq \tilde{Y}_o. \end{cases}$$

By referring to the concepts of statistics, a $(1 - \alpha)\%$ confidence interval for the mean inputs and the mean outputs are as follows:

$$\begin{aligned} \bar{\alpha}_{ij} &\in \left[\bar{\alpha}_{ij} - t_{(\frac{\alpha}{2})} \frac{\bar{\sigma}_{ij}}{\sqrt{p}}, \bar{\alpha}_{ij} + t_{(\frac{\alpha}{2})} \frac{\bar{\sigma}_{ij}}{\sqrt{p}} \right], & \forall i \in I, j \in J, \\ \bar{y}_{rj} &\in \left[\bar{y}_{rj} - t_{(\frac{\alpha}{2})} \frac{\bar{\psi}_{rj}}{\sqrt{p}}, \bar{y}_{rj} + t_{(\frac{\alpha}{2})} \frac{\bar{\psi}_{rj}}{\sqrt{p}} \right], & \forall r \in O, j \in J, \end{aligned}$$

where $\bar{\alpha}_{ij}$ and $\bar{\sigma}_{ij}^2$ are the mean and variance of inputs, respectively. Also, \bar{y}_{rj} and $\bar{\psi}_{rj}^2$ are the mean and variance of outputs, respectively. Moreover, $t_{(\frac{\alpha}{2})}$ is a random variable with distribution T and degree of freedom $p - 1$. Therefore, the lower and upper bounds of inputs and outputs are as follows:

$$\begin{aligned} \alpha_{ij}^l(\alpha) &= \bar{\alpha}_{ij} - t_{(\frac{\alpha}{2})} \frac{\bar{\sigma}_{ij}}{\sqrt{p}}, & \alpha_{ij}^u(\alpha) &= \bar{\alpha}_{ij} + t_{(\frac{\alpha}{2})} \frac{\bar{\sigma}_{ij}}{\sqrt{p}}, & \forall i \in I, j \in J, \\ y_{rj}^l(\alpha) &= \bar{y}_{rj} - t_{(\frac{\alpha}{2})} \frac{\bar{\psi}_{rj}}{\sqrt{p}}, & y_{rj}^u(\alpha) &= \bar{y}_{rj} + t_{(\frac{\alpha}{2})} \frac{\bar{\psi}_{rj}}{\sqrt{p}}, & \forall r \in O, j \in J. \end{aligned}$$

According to the pessimistic and optimistic viewpoints [9, 15, 21], the following models proposed to estimate the upper and lower bounds of efficiency of DMU_{n+1} with $(1 - \alpha)\%$ confidence interval:

$$\begin{aligned} \varphi_{n+1}^u &= \max \quad \varphi & (4.19) \\ \text{s.t.} \quad \sum_{j \in J} \lambda_j \alpha_{ij}^l(\alpha) + \lambda_{n+1} \alpha_{in+1}^u(\alpha) &\leq \alpha_{in+1}^u(\alpha), & \forall i \in I, \\ \sum_{j \in J} \lambda_j y_{rj}^u(\alpha) + \lambda_{n+1} y_{rn+1}^l(\alpha) &\geq \varphi y_{rn+1}^l(\alpha), & \forall r \in O, \\ \lambda &\in \Omega^+, \end{aligned}$$

$$\begin{aligned} \varphi_{n+1}^l &= \max \quad \varphi & (4.20) \\ \text{s.t.} \quad \sum_{j \in J} \lambda_j \alpha_{ij}^u(\alpha) + \lambda_{n+1} \alpha_{in+1}^l(\alpha) &\leq \alpha_{in+1}^l(\alpha), & \forall i \in I, \\ \sum_{j \in J} \lambda_j y_{rj}^l(\alpha) + \lambda_{n+1} y_{rn+1}^u(\alpha) &\geq \varphi y_{rn+1}^u(\alpha), & \forall r \in O, \\ \lambda &\in \Omega^+. \end{aligned}$$

It is obvious that $\varphi_{n+1}(\alpha) \in [\varphi_{n+1}^l(\alpha), \varphi_{n+1}^u(\alpha)]$. Considering model (4.11) (setting $\tilde{\beta}_o = \tilde{Y}_o$), if the optimal value of this model does not belong to $[\varphi_{n+1}^l(\alpha), \varphi_{n+1}^u(\alpha)]$, then the presented approach in this sub-section can be used to finding of the outputs under preserving the efficiency index of DMU_o in the level of significance α ($\varphi_o(\alpha)$).

Remark 4.8. According to Remark 3.2, if all input and output levels of the units have a symmetric error structure, then the same method for converting problem (3.4) to the problem (3.7) can be employed to convert the stochastic models (4.11) and (4.12) to the LP and MOLP problems, respectively.

TABLE 1. The efficiency index of 20 bank branches.

DMU	$\alpha = 0.5$	$\alpha = 0.2$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.01$
DMU01	0.58725	0.62227	0.64101	0.65714	0.68820
DMU02	1.00000	1.00000	1.00000	1.00000	1.00000
DMU03	1.00000	1.00000	1.00000	1.00000	1.00000
DMU04	0.27658	0.29342	0.30369	0.31264	0.33123
DMU05	0.48755	0.55616	0.59581	0.62879	0.68847
DMU06	0.91961	0.96016	0.98250	1.00000	1.00000
DMU07	0.48060	0.51515	0.52854	0.53974	0.56020
DMU08	1.00000	1.00000	1.00000	1.00000	1.00000
DMU09	1.00000	1.00000	1.00000	1.00000	1.00000
DMU10	1.00000	1.00000	1.00000	1.00000	1.00000
DMU11	1.00000	1.00000	1.00000	1.00000	1.00000
DMU12	0.80731	0.84684	0.86039	0.87167	0.89264
DMU13	0.90849	0.92999	0.93994	0.94777	0.96112
DMU14	0.69302	0.70699	0.71478	0.72232	0.74068
DMU15	1.00000	1.00000	1.00000	1.00000	1.00000
DMU16	1.00000	1.00000	1.00000	1.00000	1.00000
DMU17	0.93143	0.97501	0.99782	1.00000	1.00000
DMU18	0.56322	0.58407	0.59493	0.60403	0.62068
DMU19	1.00000	1.00000	1.00000	1.00000	1.00000
DMU20	0.47566	0.52995	0.56430	0.59605	0.68959

5. AN APPLICATION

The current section provides an application of the presented method in the banking sector. A dataset containing 20 branches of an Iranian commercial bank is considered to verify the realization of the research goals. The dataset containing three inputs and five outputs is reproduced from Behzadi and Mirbolouki [2] and presented in Appendix C (see Tabs. C.1 and C.2). Nevertheless, the proposed method can be utilized for other fields.

There are two main approaches for selecting input and output factors in the literature: the production approach and the intermediation approach. In this study, the intermediation approach is employed. The personal rate (\tilde{x}_1), payable benefits (\tilde{x}_2), and delayed requisitions (\tilde{x}_3) are considered as inputs while the facilities (\tilde{y}_1), amount of deposits (\tilde{y}_2), received benefits (\tilde{y}_3), received commission (\tilde{y}_4), and other resources of deposits (\tilde{y}_5) are considered as the outputs. The weighted combination of personal qualifications, quantity, education, and other bank branches is considered the cost input. The payable benefits of all deposits to customers in each branch are considered as the payable benefits. Delays in repaying loans and other facilities in each branch are considered as delay claims. The sum of business and single loans in each branch is considered as the facilities. The value of various deposits, including current, and short/long duration accounts in each branch, is considered the number of deposits. The received benefits from the total loans and facilities are considered as the received benefits. The sum received commission of all banking actions, issuance guaranty, money transfer, and others in each branch is considered the received commission.

The model (3.7) under CRS and $\bar{\sigma} = 1$ assumptions is employed to obtain the branches' efficiency score in the levels of significance $\alpha = 0.01, 0.05, 0.1, 0.2, 0.5$ (see Tab. 1).

According to Table 1, for instance, it is evident that DMU02 is efficient in the level of significance $\alpha = 0.01$ ($\theta_{02}(0.01) = 1$). An InvDEA case is described as follows: Suppose that DMU02 describes targets for some criteria, which should be realized in the future. Namely, consider that DMU02 should verify this problem: among all branches, how the input levels, including the personal rate, payable benefits, and delayed requisitions, can increase if its efficiency index in the level of significance $\alpha = 0.01$ is kept unchanged, yet the output levels, including facilities, number of deposits, received benefits, received commission, and other resources of

TABLE 2. The percentage of expected increases in the output levels DMU02.

Outputs	Facilities		Amount of deposits		Received benefits		Received commission		Other resources of deposits	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Old output (\tilde{Y}_{02})	50.772	8.011	73.132	13.11	1.815	0.02	3.035	0.68	5.823	0.689
New output ($\tilde{\beta}_{02}$)	55.772	9.011	80.132	14.61	2.015	0.03	3.335	0.70	6.323	0.709
The expected increase ($\Delta\tilde{Y}_{02}$)	5.00	1.00	7.00	1.50	0.20	0.01	0.300	0.02	0.500	0.04
Percentage changes	9.85%	12.48%	9.57%	11.44%	11.02%	50.00%	9.88%	2.94%	8.59%	2.90%

TABLE 3. The percentage of necessary increases in the inputs DMU02.

Inputs	Personal rate ($\tilde{\alpha}_{02}^1$)		Payable benefits ($\tilde{\alpha}_{02}^2$)		Delayed requisitions ($\tilde{\alpha}_{02}^3$)	
	Mean	Variance	Mean	Variance	Mean	Variance
Old input (\tilde{X}_2)	10.59	0.53	44.32	24.1	1121	0.02
The first SP solution ($\tilde{\alpha}_2$)	11.20	0.63	44.32	24.6	1122.27	0.03
The expected increase ($\Delta\tilde{X}_{02}$)	0.61	0.10	0.00	0.50	1.27	0.01
Percentage increases	5.76%	18.87%	0.0%	2.07%	0.11%	50.00%
The second SP solution ($\tilde{\alpha}_2$)	11.72	0.73	48.97	24.5	1121.30	0.04
The expected increase ($\Delta\tilde{X}_{02}$)	1.13	0.20	4.65	0.40	0.30	0.02
Percentage increases	10.67%	37.74%	10.49%	1.66%	0.06%	100.00%

deposits increase from $\tilde{Y}_{02} = (\tilde{y}_{02}^1, \tilde{y}_{02}^2, \tilde{y}_{02}^3, \tilde{y}_{02}^4, \tilde{y}_{02}^5)$, to $\tilde{\beta}_{02} = (\tilde{\beta}_{02}^1, \tilde{\beta}_{02}^2, \tilde{\beta}_{02}^3, \tilde{\beta}_{02}^4, \tilde{\beta}_{02}^5)$ as shown in Table 2. More accurately, the percentage of expected increases in the facilities (\tilde{y}_{02}^1), amount of deposits (\tilde{y}_{02}^2), received benefits (\tilde{y}_{02}^3), received commission (\tilde{y}_{02}^4), and other resources of deposits (\tilde{y}_{02}^5) are given in Table 2.

Considering SMOP (4.2) in the level of significance $\alpha = 0.01$ corresponding to DMU02 and employing the weight-sum approach [11], two SP solutions (two scenarios to increase input levels under preserving the efficiency index) are obtained to estimate the input vector (personal rate, payable benefits, and delayed requisitions) as presented in Table 3. More accurately, the percentage of the required increase in the input levels in each of two scenarios increases input levels under preserving the efficiency index in the level of significance $\alpha = 0.01$ (personal rate, payable benefits, and delayed requisitions) should be similar to Table 3. More accurately, the percentage of the required increase in the input levels in each of two scenarios increases the input levels under preserving the efficiency index in the level of significance $\alpha = 0.01$ (personal rate, payable benefits, and delayed requisitions), like to Table 3.

Table 3 proposes two patterns to the decision-maker to choose the best decision to generalize DMU02. This means that the decision-maker can perform necessary operations by selecting an appropriate method for extending DMU02. Consider that, based on Theorem 4.3 the efficiency score for the two patterns are equal to one in the level of significance $\alpha = 0.01$. Table 3 shows that if the output levels of branch DMU02 increase from \tilde{Y}_{02} to $\tilde{\beta}_{02}$, then the following operations should be performed to retain the efficiency score of this bank:

- (i) If the first scenario is selected, then the mean and variance of the personal rate input level (input 1) should increase to 5.76% and 18.87%, respectively. In contrast, if the second scenario is selected, the mean and variance of this input should increase to 10.67% and 37.74%, respectively.
- (ii) If the second scenario is selected, then the mean and variance of the second input factor should increase to 10.49% and 1.66%, respectively, while if the first scenario is selected, without increasing the mean of the payable benefits level, only the variance of this input should increase to 2.07%.

TABLE 4. The percentage of expected increases in the outputs DMU04.

Outputs	Facilities		Amount of deposits		Received benefits		Received commission		Other resources of deposits	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
Old output (\tilde{Y}_{02})	137.51	21.65	44972	13.78	4923	1.65	4212	3.84	63.61	33.73
New output ($\tilde{\beta}_{02}$)	152.51	22.85	44980	14.18	4924	1.75	4213	3.94	70.61	33.93
The expected increase ($\Delta\tilde{Y}_{02}$)	15.00	1.20	8.00	0.40	1.00	0.10	1.00	0.10	7.00	0.20
Percentage changes	10.91%	5.54%	0.02%	2.90%	0.02%	6.06%	0.02%	2.60%	11.00%	0.59%

TABLE 5. The percentage of required increases in the input levels DMU04.

Inputs	Personal rate ($\tilde{\alpha}_{04}^1$)		Payable benefits ($\tilde{\alpha}_{04}^2$)		Delayed requisitions ($\tilde{\alpha}_{04}^3$)	
	Mean	Variance	Mean	Variance	Mean	Variance
Old input (\tilde{X}_4)	11.91	0.31	17.43	12.2	59.47	5.99
The first SP solution ($\tilde{\alpha}_4$)	12.80	0.34	18.97	13.50	59.47	6.59
The expected increase ($\Delta\tilde{X}_{04}$)	0.89	0.03	1.54	1.20	0.00	0.60
Percentage increases	7.47%	9.68%	8.84%	9.84%	0.00%	10.02%
The second SP solution ($\tilde{\alpha}_2$)	12.21	0.41	22.09	12.3	147.88	0.95
The expected increase ($\Delta\tilde{X}_{04}$)	0.20	0.04	4.66	0.90	88.41	0.40
Percentage increases	1.68%	12.90%	9.67%	7.38%	148.66%	6.68%

(iii) If the first scenario is selected, then the mean and variance of the delayed requisitions input level (input 3) should increase to 0.11% and 50.00%, respectively, while if the second scenario is selected, then the mean and variance of this input should increase to 0.06% and 100.00%, respectively.

As another intended performance level, another InvDEA is described as follows: Suppose that DMU04 describes goals for some criteria, which should be attained in the future. Namely, consider that DMU04 should study this problem: among all branches, how the input levels (personal rate, payable benefits, and delayed requisitions) increase if its efficiency index in the level of significance $\alpha = 0.01$ remains unchanged ($\theta_{02}(0.01) = 0.33122$), yet the output levels increase from \tilde{Y}_{04} , to $\tilde{\beta}_{04}$ as presented in Table 4. More accurately, the percentage of intended increases in the facilities (\tilde{y}_{04}^1), amount of deposits (\tilde{y}_{04}^2), received benefits (\tilde{y}_{04}^3), received commission (\tilde{y}_{04}^4), and other resources of deposits (\tilde{y}_{04}^5) are given in Table 4.

Considering SMOP (4.2) in the level of significance $\alpha = 0.01$ corresponding to DMU04 and adopting the weight-sum approach [11], two SP solutions are produced to estimate the input levels as presented in Table 5. More accurately, the percentage of the required increase in the input levels in each of two scenarios to increase input levels under preserving the efficiency score in the significance level $\alpha = 0.01$ must be like Table 5.

Table 5 proposes two patterns to the decision-maker to select the optimal decision to extend DMU04. This means that the decision-maker can perform required operations by selecting appropriate strategies to spread DMU04. Based on Theorem 4.3 the efficiency score for the two patterns are equal to “0.33122” in the level of significance $\alpha = 0.01$.

At first, under maintaining the efficiency score, the inputs (outputs)-estimation problem is solved by Hadi-Vencheh *et al.* [23] and Wei *et al.* [45] when some or all of its outputs/inputs increase. Afterward, different theoretical and practical perspectives of the InvDEA have been verified in the previous studies, such as Lin [32], Dong Joon [10], Hadi-Vencheh *et al.* [23], Ghobadi [20], Emrouznejad *et al.* [14], Jahanshahloo *et al.* [27], Wegener and Amin [44], Jahanshahloo *et al.* [28], Gattoufi *et al.* [17], Ghobadi [21] and Amin *et al.* [1]. Nevertheless, the constructed models in the mentioned works could not solve the issue of input-estimation and

output-estimation in the presence of stochastic data. Thus, there is no available approach to compare our results with that.

6. CONCLUSIONS

In the current work, the InvDEA problems are generalized for input/output estimation in the presence of stochastic data. We used MOP problems to derive the necessary and sufficient conditions for input/output estimation. A novel optimality concept (stochastic (weak) Pareto optimality in the level of significance $\alpha \in [0, 1]$) for MOP problems was introduced to attain this goal. Its performance in attaining the desired efficiency in a banking sector application is investigated to describe the presented InvDEA strategy.

The obtained results give novel relationships among DEA and MOP problems. Besides, the mentioned results can be utilized in real applications, such as resource allocation, sensitivity analysis, preserving (or improving) efficiency values, determining revenue targets, merging the banks, and firm restructuring. Identification of the type of distribution of the efficiency random variable and also obtaining similar models in dynamic and network DEA frameworks can be considered as future research aspects.

APPENDIX A.

Proof of Theorem 4.3. To prove the theorem, $\theta_o(\alpha) = \theta_{n+1}(\alpha)$ should be shown, considering $\tilde{\alpha}_o^* = (\tilde{\alpha}_{1o}^*, \tilde{\alpha}_{2o}^*, \dots, \tilde{\alpha}_{mo}^*)$ in model (4.1). Because Δ is a feasible solution for SMOP (4.2), we get

$$P \left\{ \sum_{j \in J} \mu_j^* \tilde{X}_j \leq \theta_o(\alpha) \tilde{\alpha}_o^* \right\} \geq 1 - \alpha, \quad (\text{A.1})$$

$$P \left\{ \sum_{j \in J} \mu_j^* \tilde{Y}_j \geq \tilde{\beta}_o \right\} \geq 1 - \alpha, \quad (\text{A.2})$$

$$P \left\{ \tilde{\alpha}_o^* \geq \tilde{X}_o \right\} \geq 1 - \alpha, \quad (\text{A.3})$$

$$\mu^* \in \Omega. \quad (\text{A.4})$$

According to (A.1)–(A.4), the vector $(\mu = (\mu^*, 0), \theta = \theta_o(\alpha))$ is a feasible solution to model (4.1) (setting $\tilde{\alpha}_o^*$ in stead of $\tilde{\alpha}_o$). Therefore, $\theta_{n+1}(\alpha) \leq \theta_o(\alpha)$.

Suppose that $(\mu^{**}, \mu_{n+1}^{**}, \theta_{n+1}(\alpha) = \theta^{**})$ is an optimal solution of model (4.1), we have

$$P \left\{ \sum_{j \in J} \mu_j^{**} \tilde{X}_j + \mu_{n+1}^{**} \tilde{\alpha}_o^* \leq \theta^{**} \tilde{\alpha}_o^* \right\} \geq 1 - \alpha, \quad (\text{A.5})$$

$$P \left\{ \sum_{j \in J} \mu_j^{**} \tilde{Y}_j + \mu_{n+1}^{**} \tilde{\beta}_o \geq \tilde{\beta}_o \right\} \geq 1 - \alpha, \quad (\text{A.6})$$

$$\mu^{**} \in \Omega^+. \quad (\text{A.7})$$

Since $0 < \theta_o(\alpha) \leq 1$, by equation (A.1) we get

$$P \left\{ \sum_{j \in J} \mu_j^* \tilde{X}_j \leq \tilde{\alpha}_o^* \right\} \geq 1 - \alpha. \quad (\text{A.8})$$

Considering (A.5) and (A.8), we have

$$P \left\{ \sum_{j \in J} \mu_j^{**} \tilde{X}_j + \mu_{n+1}^{**} \left(\sum_{j \in J} \mu_j^* \tilde{X}_j \right) \leq \theta^{**} \tilde{\alpha}_o^* \right\} \geq 1 - \alpha, \quad (\text{A.9})$$

then,

$$P \left\{ \sum_{j \in J} (\mu_j^{**} + \mu_{n+1}^{**} \mu_j^*) \tilde{X}_j \leq \theta^{**} \tilde{\alpha}_o^* \right\} \geq 1 - \alpha. \quad (\text{A.10})$$

Let $\bar{\mu}_j = \mu_j^{**} + \mu_{n+1}^{**} \mu_j^*$, and write (A.10) as

$$P \left\{ \sum_{j \in J} \bar{\mu}_j \tilde{X}_j - \theta^{**} \tilde{\alpha}_o^* \leq 0 \right\} \geq 1 - \alpha. \quad (\text{A.11})$$

Using a similar method (by Eqs. (A.2) and (A.6)) we obtain

$$P \left\{ \sum_{j \in J} \bar{\mu}_j \tilde{Y}_j - \tilde{\beta}_o \geq 0 \right\} \geq 1 - \alpha. \quad (\text{A.12})$$

Also, it is obvious that

$$\bar{\mu} = (\bar{\mu}_1, \dots, \bar{\mu}_n) \in \Omega. \quad (\text{A.13})$$

If condition (i) holds, then by (A.11), we get

$$P \left\{ \sum_{j \in J} \bar{\mu}_j \tilde{X}_j - \theta^{**} \tilde{X}_o^* \leq 0 \right\} \geq 1 - \alpha. \quad (\text{A.14})$$

Also, by definition of $\tilde{\beta}_o$ and (A.12), we have

$$P \left\{ \sum_{j \in J} \bar{\mu}_j \tilde{Y}_j - \tilde{Y}_o \geq 0 \right\} \geq 1 - \alpha. \quad (\text{A.15})$$

Therefore, equations (A.13)–(A.15) imply that $(\bar{\mu}, \theta^{**})$ is a feasible solution to model (3.1). This implies $\theta_o(\alpha) \leq \theta^{**} = \theta_{n+1}(\alpha)$. Hence, with regard to inequality $\theta_{n+1}(\alpha) \leq \theta_o(\alpha)$, we have $\theta_o(\alpha) = \theta_{n+1}(\alpha)$ and completes the proof under condition (i).

If condition (ii) holds, then $P\{\tilde{x}_{lo} - \tilde{\alpha}_{lo}^* \leq -\epsilon\} \geq 1 - \alpha$. There exists non-negative scalar $\varepsilon \geq 0$ such that $P\{\tilde{x}_{lo} - \tilde{\alpha}_{lo}^* \leq -\epsilon\} = 1 - \alpha + \varepsilon$. Then, there exists positive scalar $\kappa > 0$ such that

$$P\{\tilde{x}_{lo} - \tilde{\alpha}_{lo}^* \leq -\kappa\} = 1 - \alpha. \quad (\text{A.16})$$

Now, by contradiction assume that $\theta_{n+1}(\alpha) = \theta^{**} < \theta_o(\alpha)$. By equation (A.11) and contradiction assume $(\theta_{n+1}(\alpha) = \theta^{**} < \theta_o(\alpha))$, we obtain

$$P \left\{ \sum_{j \in J} \bar{\mu}_j \tilde{X}_j < \theta_o(\alpha) \tilde{\alpha}_o^* \right\} \geq 1 - \alpha. \quad (\text{A.17})$$

Then

$$P \left\{ \sum_{j \in J} \bar{\mu}_j \tilde{X}_j - \theta_o(\alpha) \tilde{\alpha}_o^* < 0 \right\} \geq 1 - \alpha. \quad (\text{A.18})$$

There exists non-negative vector $\epsilon_o = (\epsilon_{1o}, \epsilon_{2o}, \dots, \epsilon_{mo})$ such that

$$P \left\{ \sum_{j=1}^n \bar{\mu}_j \tilde{X}_j - \theta_o(\alpha) \tilde{\alpha}_o^* < 0 \right\} = 1 - \alpha + \epsilon_o. \quad (\text{A.19})$$

Therefore, there exists positive vector $S_o = (s_{1o}, s_{2o}, \dots, s_{mo})$ such that

$$P \left\{ \sum_{j \in J} \bar{\mu}_j \tilde{X}_j - \theta_o(\alpha) \tilde{\alpha}_o^* \leq -S_o \right\} = 1 - \alpha. \quad (\text{A.20})$$

Clearly, there exists positive vector $\bar{S}_o = (\bar{s}_{1o}, \bar{s}_{2o}, \dots, \bar{s}_{mo})$ such that $S_o = \theta_o(\alpha) \bar{S}_o$. Then, by equation (A.20), we get

$$P \left\{ \sum_{j \in J} \bar{\mu}_j \tilde{X}_j - \theta_o(\alpha) \tilde{\alpha}_o^* \leq -\theta_o(\alpha) \bar{S}_o \right\} = 1 - \alpha. \quad (\text{A.21})$$

Therefore,

$$P \left\{ \sum_{j \in J} \bar{\mu}_j \tilde{X}_j - \theta_o(\alpha) (\tilde{\alpha}_o^* - \bar{S}_o) \leq 0 \right\} = 1 - \alpha. \quad (\text{A.22})$$

Defining

$$-\bar{k} = \max \{ -\kappa, \max \{ -\bar{s}_{io}, \quad i = 1, 2, \dots, m \} \}, \quad (\text{A.23})$$

we have $\bar{k} > 0$. Now, for each $i \in I$, define

$$\bar{\alpha}_{io} := \begin{cases} \tilde{\alpha}_{io}^* & \text{if } i \neq l, \\ \tilde{\alpha}_{io}^* - \bar{k} & \text{if } i = l. \end{cases} \quad (\text{A.24})$$

By (A.16) and (A.22)–(A.24) we have

$$P \left\{ \sum_{j \in J} \bar{\mu}_j \tilde{x}_{ij} - \theta_o(\alpha) \bar{\alpha}_{io} \leq 0 \right\} \geq 1 - \alpha, \quad \forall i \in I, \quad (\text{A.25})$$

$$P \{ \tilde{x}_{lo} - \bar{\alpha}_{lo} \leq 0 \} \geq 1 - \alpha. \quad (\text{A.26})$$

Also,

$$P \{ \tilde{x}_{io} - \bar{\alpha}_{io} \leq 0 \} \geq 1 - \alpha \quad \forall i \in I, \quad i \notin l. \quad (\text{A.27})$$

Since $\bar{\mu} \in \Omega$, because of (A.12) and (A.25)–(A.27), $(\bar{\mu}, \bar{\alpha}_o)$ is a feasible solution to problem (4.2), where $P \{ \bar{\alpha}_{io} - \tilde{\alpha}_{io}^* \leq 0 \} \geq 1 - \alpha$ for all $i \in I - \{l\}$, and $P \{ \bar{\alpha}_{lo} - \tilde{\alpha}_{lo}^* \leq -\bar{k} \} \geq 1 - \alpha$. But this is impossible because Δ is a SP solution in the level of significance α to SMOP (4.2). \square

APPENDIX B.

Proof of Theorem 4.6. To prove the theorem, $\varphi_o(\alpha) = \varphi_{n+1}(\alpha)$ should be shown, setting $\tilde{\beta}_o^* = (\tilde{\beta}_{1o}^*, \tilde{\beta}_{2o}^*, \dots, \tilde{\beta}_{so}^*)$ in model (4.11). Because Δ is a feasible solution for SMOP (4.12), we have

$$P \left\{ \sum_{j \in J} \mu_j^* \tilde{X}_j \leq \tilde{\alpha}_o \right\} \geq 1 - \alpha, \quad (\text{B.1})$$

$$P \left\{ \sum_{j \in J} \mu_j^* \tilde{Y}_j \geq \varphi_o(\alpha) \tilde{\beta}_o^* \right\} \geq 1 - \alpha, \quad (\text{B.2})$$

$$P \left\{ \tilde{\beta}_o^* \geq \tilde{Y}_o \right\} \geq 1 - \alpha, \quad (\text{B.3})$$

$$\mu^* \in \Omega. \quad (\text{B.4})$$

According to (B.1)–(B.4), the vector $(\mu = (\mu^*, 0), \varphi = \varphi_o(\alpha))$ is a feasible solution to model (4.11) (setting $\tilde{\beta}_o^*$ in stead of $\tilde{\beta}_o$). Therefore, $\varphi_{n+1}(\alpha) \geq \varphi_o(\alpha)$.

Suppose that $(\mu^{**}, \mu_{n+1}^{**}, \varphi_{n+1}(\alpha) = \varphi^{**})$ is an optimal solution of model (4.11), we have

$$P \left\{ \sum_{j \in J} \mu_j^{**} \tilde{X}_j + \mu_{n+1}^{**} \tilde{\alpha}_o \leq \tilde{\alpha}_o \right\} \geq 1 - \alpha, \quad (\text{B.5})$$

$$P \left\{ \sum_{j \in J} \mu_j^{**} \tilde{Y}_j + \mu_{n+1}^{**} \tilde{\beta}_o^* \geq \varphi^{**} \tilde{\beta}_o^* \right\} \geq 1 - \alpha, \quad (\text{B.6})$$

$$\mu^{**} \in \Omega^+. \quad (\text{B.7})$$

Considering (B.1) and (B.5), we have

$$P \left\{ \sum_{j \in J} \mu_j^{**} \tilde{X}_j + \mu_{n+1}^{**} \left(\sum_{j=1}^n \mu_j^* \tilde{X}_j \right) \leq \tilde{\alpha}_o \right\} \geq 1 - \alpha, \quad (\text{B.8})$$

then,

$$P \left\{ \sum_{j \in J} (\mu_j^{**} + \mu_{n+1}^{**} \mu_j^*) \tilde{X}_j \leq \tilde{\alpha}_o \right\} \geq 1 - \alpha. \quad (\text{B.9})$$

Let $\bar{\mu}_j = \mu_j^{**} + \mu_{n+1}^{**} \mu_j^*$, and write (B.9) as

$$P \left\{ \sum_{j \in J} \bar{\mu}_j \tilde{X}_j \leq \tilde{\alpha}_o \right\} \geq 1 - \alpha. \quad (\text{B.10})$$

Also, it is obvious that

$$\bar{\mu} = (\bar{\mu}_1, \dots, \bar{\mu}_n) \in \Omega. \quad (\text{B.11})$$

Since $\varphi_o(\alpha) \geq 1$ and equation (B.2), we have

$$P \left\{ \sum_{j \in J} \mu_j^* \tilde{Y}_j \geq \tilde{\beta}_o^* \right\} \geq 1 - \alpha. \quad (\text{B.12})$$

By contradiction assume that $\varphi_{n+1}(\alpha) = \varphi^{**} > \varphi_o(\alpha)$. By (B.6), we get

$$P \left\{ \sum_{j \in J} \mu_j^{**} \tilde{Y}_j + \mu_{n+1}^{**} \tilde{\beta}_o^* > \varphi_o(\alpha) \tilde{\beta}_o^* \right\} \geq 1 - \alpha. \quad (\text{B.13})$$

By (B.12) and (B.13), we have

$$P \left\{ \sum_{j=1}^n \mu_j^{**} \tilde{Y}_j + \mu_{n+1}^{**} \left(\sum_{j=1}^n \mu_j^* \tilde{Y}_j \right) > \varphi_o(\alpha) \tilde{\beta}_o^* \right\} \geq 1 - \alpha. \quad (\text{B.14})$$

Then

$$P \left\{ \sum_{j \in J} \left(\mu_j^{**} + \mu_{n+1}^{**} \left(\sum_{j=1}^n \mu_j^* \right) \right) \tilde{Y}_j > \varphi_o(\alpha) \tilde{\beta}_o^* \right\} \geq 1 - \alpha. \quad (\text{B.15})$$

According to definition of $\bar{\mu}_j$ for $j = 1, 2, \dots, n$, and equation (B.15), we have

$$P \left\{ \sum_{j=1}^n \bar{\mu}_j \tilde{Y}_j > \varphi_o(\alpha) \tilde{\beta}_o^* \right\} \geq 1 - \alpha. \quad (\text{B.16})$$

Therefore,

$$P \left\{ \varphi_o(\alpha) \tilde{\beta}_o^* - \left(\sum_{j \in J} \bar{\mu}_j \tilde{Y}_j \right) < 0 \right\} \geq 1 - \alpha. \quad (\text{B.17})$$

There exists non-negative vector $\epsilon_o = (\epsilon_{1o}, \epsilon_{2o}, \dots, \epsilon_{so})$ such that

$$P \left\{ \varphi_o(\alpha) \tilde{\beta}_o^* - \left(\sum_{j \in J} \bar{\mu}_j \tilde{Y}_j \right) < 0 \right\} = 1 - \alpha + \epsilon_o. \quad (\text{B.18})$$

Therefore, there exists positive vector $S_o = (s_{1o}, s_{2o}, \dots, s_{so})$ such that

$$P \left\{ \varphi_o(\alpha) \tilde{\beta}_o^* - \left(\sum_{j \in J} \bar{\mu}_j \tilde{Y}_j \right) \leq -S_o \right\} = 1 - \alpha. \quad (\text{B.19})$$

Clearly, there exists positive vector $\bar{S}_o = (\bar{s}_{1o}, \bar{s}_{2o}, \dots, \bar{s}_{so})$ such that $S_o = \varphi_o(\alpha) \bar{S}_o$. Defining

$$-\bar{k} = \max \{ -\bar{s}_{ro}, \quad r = 1, 2, \dots, s \}, \quad (\text{B.20})$$

we have $\bar{k} > 0$. Now, for each $r \in O$, define $\tilde{\beta}_{ro} := \tilde{\beta}_{ro}^* + \bar{k}$. By equation (B.19), we get

$$P \left\{ \sum_{j \in J} \bar{\mu}_j \tilde{Y}_j \geq \varphi_o(\alpha) \tilde{\beta}_o \right\} \geq 1 - \alpha. \quad (\text{B.21})$$

Also, according to equation (B.3), we have

$$P \left\{ \tilde{\beta}_o \geq \tilde{Y}_o \right\} \geq 1 - \alpha. \quad (\text{B.22})$$

According to equations (B.10), (B.11), (B.21), and (B.22), $(\bar{\mu}, \tilde{\beta}_o)$ is a feasible solution to problem (4.12) where $P \left\{ \tilde{\beta}_{ro} - \tilde{\beta}_{ro}^* \geq \bar{k} \right\} \geq 1 - \alpha$ for each $r \in O$. But this is impossible because Δ is a SWP solution in the level of significance α to SMOP (4.12). \square

APPENDIX C.

See Tables C.1 and C.2.

TABLE C.1. The inputs of 20 bank branches.

DMU	Input 1		Input 2		Input 3	
	Mean	Variance	Mean	Variance	Mean	Variance
DMU01	9.131	0.05	18.79	8.81	7.228	0.58
DMU02	10.59	0.53	44.32	24.1	1.121	0.02
DMU03	6.712	0.86	19.73	27.7	19.21	0.47
DMU04	11.91	0.31	17.43	12.2	59.47	5.99
DMU05	7.012	0.02	10.38	2.12	12.23	0.85
DMU06	18.99	0.88	16.67	10.8	568.6	28.1
DMU07	11.16	0.01	25.46	18.6	552.8	43.2
DMU08	15.05	0.48	123.1	42.6	14.78	0.06
DMU09	8.787	0.38	36.16	38.4	361.8	23.2
DMU10	19.88	0.25	46.41	53.1	12.81	0.38
DMU11	18.92	0.17	36.88	54.5	24.43	0.01
DMU12	20.45	0.42	100.8	31.8	115.2	19.4
DMU13	12.41	0.12	20.19	10.6	78.02	24.1
DMU14	8.051	0.79	33.21	24.3	115.3	15.6
DMU15	18.48	0.92	45.36	92.6	57.52	12.8
DMU16	10.35	0.27	11.16	3.32	43.32	36.1
DMU17	9.511	0.01	31.49	38.5	173.3	3.13
DMU18	13.71	0.18	40.32	51.4	10.88	0.12
DMU19	11.69	0.26	26.44	26.2	31.22	0.05
DMU20	7.823	0.58	17.74	10.1	13.06	8.88

TABLE C.2. The outputs of 20 bank branches.

DMU	Output 1		Output 2		Output 3		Output 4		Output 5	
	Mean	Variance								
DMU01	149.85	48.51	49.621	18.01	4.701	0.41	4.748	0.41	30.09	24.34
DMU02	50.772	8.011	73.132	13.11	1.815	0.02	3.035	0.68	5.823	0.689
DMU03	259.91	29.59	108.04	22.56	6.016	0.01	10.06	2.98	2.721	0.4392
DMU04	137.51	21.65	44.972	13.78	4.923	1.65	4.212	3.84	63.61	33.73
DMU05	95.901	2.521	31.633	18.94	2.718	0.44	9.024	7.23	64.55	31.25
DMU06	112.58	3.562	71.958	17.05	13.19	9.15	41.89	13.2	291.3	92.52
DMU07	192.97	14.56	78.015	19.56	7.791	3.56	15.89	2.34	7.499	1.268
DMU08	724.38	66.02	219.69	37.56	35.32	15.3	23.98	10.4	361.6	48.35
DMU09	548.15	41.86	86.225	28.31	17.64	3.67	86.23	17.2	565.2	175.1
DMU10	1229.1	69.02	194.58	17.35	25.91	12.3	86.76	22.1	600.6	86.34
DMU11	11557	718.1	155.32	49.13	166.6	14.1	8.142	3.11	119.9	14.89
DMU12	1132.1	35.35	248.16	23.89	46.88	26.3	31.85	4.1	96.21	44.47
DMU13	438.39	17.41	104.41	25.71	10.68	1.04	30.22	3.77	331.9	17.16
DMU14	260.82	15.32	87.369	23.44	8.415	3.52	6.101	2.86	36.93	5.777
DMU15	11190	1214	166.44	10.69	65.12	16.1	132.7	13.9	919.1	133.1
DMU16	709.85	42.19	159.48	32.78	36.89	6.28	12.15	2.51	79.66	10.52
DMU17	308.11	43.72	107.03	29.39	11.87	1.76	13.63	1.57	342.3	134.9
DMU18	259.21	34.21	81.779	21.65	5.212	1.94	8.021	3.52	107.9	13.07
DMU19	381.33	57.38	72.993	29.26	5.165	1.79	50.32	7.47	577.2	113.4
DMU20	399.25	10.41	40.985	15.93	11.51	0.99	6.432	3.25	82.13	9.891

Acknowledgements. The authors would like to thank Professor Nelson Maculan (Editor-in-Chief) and anonymous reviewers for their insightful and constructive comments and suggestions, as results the paper has been improved. We also thank Dr. Soleimani-Chamkhoram and Dr. Saadat Melli for their kind Consultations.

REFERENCES

- [1] G.R. Amin, A. Emrouznejad and S. Gattoufi, Modelling generalized firms' restructuring using inverse DEA. *J. Prod. Anal.* **48** (2017) 51–61.
- [2] M.H. Behzadi and M. Mirbolouki, Symmetric error structure in stochastic DEA. *Int. J. Ind. Math.* **4** (2012) 335–343.
- [3] M.H. Behzadi, N. Nematollahi and M. Mirbolouki, Ranking efficient DMUS with stochastic data by considering inefficient frontier. *Int. J. Ind. Math.* **1** (2009) 219–226.
- [4] A. Charnes, W.W. Cooper and E. Rhodes, Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* **2** (1978) 429–444.
- [5] W.W. Cooper, Z. Huang and S. Li, Satisficing DEA models under chance constraints. *Ann. Oper. Res.* **66** (1996) 259–279.
- [6] W.W. Cooper, Z. Huang, V. Lelas, S. Li and O.B. Olesen, Chance constrained programming formulations for stochastic characterizations of efficiency and dominance in DEA. *J. Prod. Anal.* **9** (1998) 53–80.
- [7] W.W. Cooper, L.M. Seiford and K. Tone, Data Envelopment Analysis: A Comprehensive Text With Models, Applications, References and DEA-Solver Software. Kluwer Academic Publisher (1999).
- [8] W.W. Cooper, H. Deng, Z.M. Huang and S.X. Li, Change constrained programming approaches to technical efficiencies and inefficiencies in stochastic data envelopment analysis. *J. Oper. Res. Soc.* **53** (2002) 1347–1356.
- [9] D. Despotis and Y. Smirlis, Data envelopment analysis with imprecise data. *Eur. J. Oper. Res.* **140** (2002) 24–36.
- [10] L. Dong Joon, Inverse DEA with frontier changes for new target setting. *Eur. J. Oper. Res.* **254** (2016) 510–516.
- [11] M. Ehrgott, Multicriteria Optimization. Springer Berlin (2005).
- [12] A. Emrouznejad and M. Tavana, Performance Measurement with Fuzzy Data Envelopment Analysis. Springer (2014).
- [13] A. Emrouznejad and G. Yang, A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *J. Soc.-Econ. Plan. Sci.* **61** (2018) 4–8.
- [14] A. Emrouznejad, G.-L. Yang and G.R. Amin, A novel inverse DEA model with application to allocate the CO₂ emissions quota to different regions in Chinese manufacturing industries. *J. Oper. Res. Soc.* **70** (2019) 1079–1090.
- [15] T. Entani, Y. Maeda and H. Tanaka, Dual models of interval DEA and its extension to interval data. *Eur. J. Oper. Res.* **136** (2002) 32–45.
- [16] M.J. Farrell, The measurement of productive efficiency. *J. R. Stat. Soc.: Ser. A (General)* **120** (1957) 253–281.
- [17] S. Gattoufi, G.R. Amin and A. Emrouznejad, A new inverse DEA method for merging banks. *IMA J. Manage. Math.* **25** (2014) 73–87.
- [18] S. Ghobadi, Inverse DEA using enhanced russell measure in the presence of fuzzy data. *Int. J. Ind. Math.* **10** (2018) 1–16.
- [19] S. Ghobadi, A generalized DEA model for inputs (outputs) estimation under inter-temporal dependence. *RAIRO:OR* **53** (2019) 1791–1805.
- [20] S. Ghobadi, A dynamic DEA model for resource allocation. *Int. J. Math. Oper. Res.* **17** (2020) 50–77.
- [21] S. Ghobadi, Merging decision-making units with interval data. *RAIRO:OR* **55** (2021) 1605–1630.
- [22] S. Ghobadi and S. Jahangiri, Inverse DEA: review, extension and application. *Int. J. Technol. Decis. Making* **14** (2015) 805–824.
- [23] A. Hadi-Vencheh, A.A. Foroughi and M. Soleimani-Damaneh, A DEA model for resource allocation. *Econ. Model.* **25** (2008) 983–993.
- [24] F. Hosseinzadeh Lotfi, N. Nematollahi, M.H. Behzadi, M. Mirbolouki and Zohreh Moghaddas, Centralized resource allocation with stochastic data. *J. Comput. Appl. Math.* **236** (2012) 1783–1788.
- [25] Z. Huang and S. Li, Stochastic DEA models with different types of input-output disturbances. *J. Prod. Anal.* **15** (2001) 95–113.
- [26] G.R. Jahanshahloo, M.H. Behzadi and M. Mirbolouki, Ranking stochastic efficient dmus based on reliability. *Int. J. Ind. Math.* **2** (2010) 263–270.
- [27] G.R. Jahanshahloo, F. Hoseinzadeh Lotfi, M. Rostami-malkhalifeh and S. Ghobadi, Using enhanced russell model to solve inverse data envelopment analysis problems. *Sci. World J.* **2014** (2014) 1–10.
- [28] G.R. Jahanshahloo, M. Soleimani-damaneh and S. Ghobadi, Inverse DEA under inter-temporal dependence using multiple-objective programming. *Eur. J. Oper. Res.* **240** (2015) 447–456.
- [29] M. Khodabakhshi, Estimating most productive scale size with stochastic data in data envelopment analysis. *Econ. Model.* **26** (2009) 968–973.
- [30] T. Kuosmanen and M. Kortelainen, Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *J. Prod. Anal.* **38** (2012) 11–28.
- [31] S. Li, Stochastic models and variable returns to scales in data envelopment analysis. *Eur. J. Oper. Res.* **104** (1998) 532–548.
- [32] H.T. Lin, An efficiency-driven approach for setting revenue target. *Decision Support Syst.* **49** (2010) 311–317.
- [33] J.S. Liu, L.Y. Lu, W.M. Lu and B.J. Lin, A survey of DEA applications. *Omega* **41** (2013) 893–902.
- [34] F.H. Lotfi, N. Nematollahi, M. Behzadi and M. Mirbolouki, Ranking decision making units with stochastic data by using coefficient of variation. *Math. Comput. App.* **15** (2010) 148–155.

- [35] H. Naseri, S.E. Najafi and A. Saghaei, DEA model considering outputs with stochastic noise and a heavy-tailed (stable) distribution. *INFOR: Inf. Syst. Oper. Res.* **58** (2020) 87–108.
- [36] O.B. Olesen and N.C. Petersen, Stochastic data envelopment analysis – a review. *Eur. J. Oper. Res.* **251** (2016) 2–21.
- [37] P. Peykani, E. Mohammadi, M.S. Pishvaaee, M. Rostamy-Malkhalifeh and A. Jabbarzadeh, A novel fuzzy data envelopment analysis based on robust possibilistic programming: Possibility, necessity and credibility-based approaches. *RAIRO:OR* **52** (2018) 1445–1463.
- [38] P. Peykani, E. Mohammadi, R.F. Saen, S.J. Sadjadi and M. Rostamy-Malkhalifeh, Data envelopment analysis and robust optimization: a review. *Expert Syst.* **37** (2020) e12534.
- [39] J.K. Sengupta, Efficiency analysis by stochastic data envelopment analysis. *Appl. Econ. Lett.* **7** (2002) 379–383.
- [40] L. Simar, How to improve the performance of DEA/FDH estimators in the presence of noise. *J. Prod. Anal.* **28** (2007) 1833–201.
- [41] L. Simar and P.W. Wilson, A general methodology for bootstrapping in non-parametric frontier models. *J. Appl. Stat.* **27** (2000) 779–802.
- [42] K. Soleimani-Chamkhorami, F.H. Lotfi, G.R. Jahanshahloo and M. Rostamy-Malkhalifeh, Preserving cost and revenue efficiency through inverse data envelopment analysis models. *INFOR: Inf. Syst. Oper. Res.* **58** (2020) 561–578.
- [43] S. Thore, Chance-constrained activity analysis. *Eur. J. Oper. Res.* **30** (1987) 267–269.
- [44] M. Wegener and G.R. Amin, Minimizing greenhouse gas emissions using inverse DEA with an application in oil and gas. *Expert Syst. Appl.* **122** (2019) 369–375.
- [45] Q.L. Wei, J.Z. Zhang and X.S. Zhang, An inverse DEA model for inputs/outputs estimate. *Eur. J. Oper. Res.* **121** (2000) 151–163.
- [46] E. Zenodin and S. Ghobadi, Merging decision-making units under inter-temporal dependence. *IMA J. Manage. Math.* **31** (2020) 139–16.
- [47] X. Zhang and J. Cui, A project evaluation system in the state economic information system of China: an operation research practice in public sector. *Int. Trans. Oper.* **6** (1999) 441–452.
- [48] J. Zhu, Imprecise DEA via standard linear DEA models with a revisit to a Korean mobile telecommunication company. *Oper. Res.* **52** (2004) 323–329.

Subscribe to Open (S2O)

A fair and sustainable open access model



This journal is currently published in open access under a Subscribe-to-Open model (S2O). S2O is a transformative model that aims to move subscription journals to open access. Open access is the free, immediate, online availability of research articles combined with the rights to use these articles fully in the digital environment. We are thankful to our subscribers and sponsors for making it possible to publish this journal in open access, free of charge for authors.

Please help to maintain this journal in open access!

Check that your library subscribes to the journal, or make a personal donation to the S2O programme, by contacting subscribers@edpsciences.org

More information, including a list of sponsors and a financial transparency report, available at: <https://www.edpsciences.org/en/math-s2o-programme>