Journal of Mathematical Extension
Vol. 17, No. 11, (2023) (1)1-52
URL: https://doi.org/10.30495/JME.2023.2787
ISSN: 1735-8299
Original Research Paper

Merging Decision-Making Units with Stochastic Data

A. Ghomi
Science and Research Branch, Islamic Azad University

S. Ghobadi*
Khomeinishahr Branch, Islamic Azad University

M. H. Behzadi
Science and Research Branch, Islamic Azad University

M. Rostamy-Malkhalifeh
Science and Research Branch, Islamic Azad University

Kh. Soleimani-Chamkhorami
Najafabad Branch, Islamic Azad University

Abstract. The first problem in merging units is an estimation of the inherited inputs/outputs of the merged unit from merging units and the identification of the least and most achievable efficiency targets from the merged unit is the second one. There are some models to attain a response to these problems. However, these models could not be employed when a deviation from the frontier is observed due to noise and random error in the data. In order to deal with this problem, this paper presents a novel method according to inverse data envelopment analysis for estimating the inherited input/output levels of the merged unit to reach the pre-determined efficiency score in the level of significance

Received: July 2023; Accepted: October 2023
*Corresponding Author

$\alpha \in (0, 1)$. This paper also suggests a stochastic programming model for estimating the least possible efficiency score via the given merging. The practical applicability of these proposed methodologies is further validated through a comprehensive example within the banking sector.

AMS Subject Classification: 90-XX; 90-10; 90C15
Keywords and Phrases: Inverse DEA, Stochastic DEA, Efficiency, Merging Units, Stochastic Multiple-Objective Programming (SMOP)

1 Introduction

Inverse Data Envelopment Analysis (InvDEA) has been recognized as one of the interesting and significant topics in the DEA area. The main concept in InvDEA is to estimate the required level of resources and products for a given unit to achieve a pre-defined efficiency goal. In contrast, the traditional DEA models aim to estimate the efficiency index of a particular unit with specific resources and products. According to the operations research (OR) literature, DEA and InvDEA were initially proposed by Charnes et al. [10] and Zhang and Cui [63], respectively. It is worth noting that in the last two decades, a wide range of operations research studies have been devoted to the DEA field. In fact, this scientific topic is an important tool in operation research and management science according to mathematical programming to evaluate the performance of a unit by comparison with other units. Various applications of this field can be seen in the existing literature.

According to the basic concepts of DEA, it is imperative to possess precise knowledge of both the input and output quantities of the units under evaluation, while this condition may not be realized in some realistic situations. In real-world applications, the quantification of input and output levels frequently presents challenges characterized by imprecision and ambiguity. In response to these challenges, a range of methodologies has been proposed in academic literature. These include the Fuzzy DEA (FDEA) approach, the Bootstrap DEA approach, the Imprecise DEA approach, and the Robust DEA (RDEA) approach. However, due to the inherent complexity and competition of the real world, DEA models often consider random data errors in the production process. In the mentioned approaches, random errors have not been considered in their respective models. Stochastic DEA (SDEA) is utilized for the efficiency assessment of units under uncertain data with random errors. In other words, according to the structure of conventional DEA models, any deviation from the efficiency boundary is considered inefficient, and there is no chance for random noise. This is something to ponder when managers deal with units with inaccurate input and output levels in different real-world situations. Analysts may assume inaccurate input and output levels as stochastic parameters in these cases. It is worth noting that it is possible to identify different data features while working with random variables based on the probability of unforeseen events. This is a crucial feature that encourages the decision-maker to employ SDEA.

On the other hand, the idea of InvDEA has been utilized to solve the merging units problem by Gattoufi et al. [22]. The following important problem is solved in the InvDEA: *If a subset of units is needed for merging and generating a new unit with a particular input/output level and a pre-defined efficiency goal, to what extent should the input/output level of the generated new unit be?* Gattoufi et al. [22] established new InvDEA models to solve this problem. A banking sector application was employed to investigate the credibility and capability of the presented method. According to the DEA literature, various InvDEA models with different theoretical and practical frameworks have been established to estimate the inherited input/output levels of the new generated unit from merging units to achieve the pre-defined efficiency score [1, 2, 29, 62].

Recently, the InvDEA problems have been extended in a stochastic analytical framework [32]. They answered the following important questions in the stochastic InvDEA: among a set of units, to what extent should the input/output levels of the unit increase while maintaining the efficiency score in the level of significance $\alpha \in (0, 1)$, and increasing the output/input levels? Ghomi et al. [32] employed SMOP models to obtain sufficient conditions for input/output estimation in the level of significance α. For this purpose, they introduced two new optimality notions for SMOP problems at the significance level α, with stochastic Pareto (SP) optimality and stochastic weak Pareto (SWP) optimality. Note that the stochastic InvDEA estimates the input/output levels to make the production possibility set (PPS) frontiers, incorporating both inefficiency and stochastic errors, closer to the bulk of the generating

DMUs.

To the best of the author's knowledge, none of the performed studies covers the case of merging units to create a new unit with random data. Due to some modeling limitations in the InvDEA models, these models could not be employed to reach the pre-defined goal in the merging units problem when observing a deviation from the frontier due to the presence of noise and random error in the data. Therefore, it is necessary to incorporate stochastic data into modeling the merging units problem based on the idea of InvDEA. It is worth noting that there are two critical issues in the merging units problem: i) Estimation of the inherited input/output levels of the new generated unit from merging units; ii) Identification of the least possible efficiency score at the significance level α from the new generated unit. The main contribution of this study is to provide a theoretical and practical framework to solve the above critical issues based on stochastic InvDEA. For this purpose, novel InvDEA models are established to answer the significant issue: (i) estimation of the required input/output levels of the new generated unit from merging units to achieve a pre-specified efficiency score at the significance level α. Sufficient conditions are derived for estimating input/output levels using stochastic (weak) Pareto solutions at the significance level α of the stochastic multiple-objective programming (SMOP) problems. The paper also provides an insightful method in the SDEA framework to answer the significant topic (ii). The proposed method indicates the least efficiency target at the significance level α that can be realized by the merged unit. From a managerial point of view, it is crucial to know the lowest performance score that can be achieved at the significance level α for the generated new unit. The managers can employ these models to design useful approaches to improve the efficiency of units. Finally, the validity of the established models has been demonstrated via a banking application for better observation.

The remainder of this paper is structured as follows. Section 2 gives the related literature from SDEA and InvDEA. In section 3, some conventional models in DEA in the presence of negative data are reviewed. Section 4 is dedicated to the main results of this paper and deals with the merging unit problem. A banking sector application is provided in Section 5 to illustrate the efficiency of the proposed method. Section 6 gives a short conclusion. Limitations and future extensions are given in Section 7. Proofs of the stated theorems and application data tables are also presented in the appendices.

2 Literature Review

2.1 Stochastic DEA Literature Review

The chance-constrained DEA (CCDEA) and the semi-parametric stochastic frontier analysis (SFDEA) are two traditional approaches in dealing with SDEA models [52]. The CCDEA approach is employed for random input/output levels with a characterized probability distribution function for the uncertain parameters [14]. The SFDEA approach is a regression analysis-based method that employs econometric methods to predict the generation boundary [46]. Concerning the DEA literature, numerous scholars, including [38, 41, 48, 54], have discussed various theoretical and practical views of the SDEA models. Most of the established models are based on nonlinear programming, and solving these models is challenging for researchers. Some SDEA models were established by extending the conventional production possibility set to a random state [13]. Besides presenting the stochastic efficient units by Cooper et al. [14], they also formulated several generalized DEA models for random data. In the literature, there exist some studies about The efficiency dominance of a unit by probabilistic comparisons of input and output levels with other units based on chance-constrained programming [38, 48]. Behzadi and Mirbolouki [6] conducted a novel and distinctive study to find a symmetric error structure for inputs and outputs and proposed an envelopment stochastic input-oriented DEA model under the constant returns to scale (CRS) assumption of the production technology. The given model is linear. This model can assess efficient DMUs and determine the relative efficiency of units at a significantly specified level. Three copula-stochastic CCR models were proposed by considering the dependency between input and output variables and their simultaneous dependencies [5]. Tsionas [58] proposed a Bayesian alternative to adaptive LASSO model estimation to combine DEA and stochastic frontier models. Besides, some scholars have established SDEA models and provided different outcomes in various frameworks, including [7, 9, 20, 36, 39, 44, 45, 50, 59].

2.2 Inverse DEA Literature Review

The InvDEA literature has been enhanced by various theoretical and practical frameworks, such as sensitivity analysis [40], firms restructuring [3], setting revenue target [49], resource allocation [34], under imprecise or vague data [31], preserving (improving) efficiency values [47]. Some of these studies are given in Tables (1), (2), and (3) with brief descriptions.

The idea of InvDEA has been utilized to solve the merging units problem by Gattoufi et al. [22]. Gattoufi et al. [22] established new InvDEA models to solve this problem. On the other hand, the InvDEA problems have been extended in a stochastic analytical framework [32]. In this paper, a new connection is established between these two issues, which is an important problem from both theoretical and practical points of view.

3 Stochastic DEA

Assume that there are n DMUs, $\{\text{DMU}_j : \ j \in J = \{1, 2, ..., n\}\}$, in which DMU_j consume m inputs $(\tilde{X}_j = (\tilde{x}_{1j}, \tilde{x}_{2j}, \ldots, \tilde{x}_{mj}))$ to produce s outputs $(\tilde{Y}_j = (\tilde{y}_{1j}, \tilde{y}_{2j}, \ldots, \tilde{y}_{sj}))$. Suppose that the input and output levels of the jth DMU are random variables with normal distribution as follows:

$$\tilde{x}_{ij} \sim N(x_{ij}, \sigma^2_{ij}), \quad i = 1, 2, ..., m,$$
$$\tilde{y}_{rj} \sim N(y_{rj}, \psi^2_{rj}), \quad r = 1, 2, ..., s,$$

where x_{ij} and y_{rj} are the inputs and outputs mean of the jth DMU, respectively. Moreover, σ^2_{ij} and ψ^2_{rj} are the inputs and outputs variance of the jth DMU, respectively. Note that if the mean and variance of the data are unknown, they must be estimated. For this purpose, unbiased and efficient estimators should be used. With regard to the stochastic DEA literature, if $\{(\check{x}^k_{ij}, \ \check{y}^k_{rj})| \ i = 1, 2, ..., m, \ r = 1, 2, ..., s, \ k = 1, 2, ..., \kappa\}$ is a random sample with size κ for the DMU_j, $j = 1, 2, ..., n$,

Table 1: Some research in the field of InvDEA.

Reference	Short description
Jahanshahloo et al. [42]	This work has been extended the InvDEA problems for input/output estimation under inter-temporal dependence data addressed in [18].
Hadi-vencheh et al. [35]	In this work, the InvDEA models are extended under interval data.
Dong Joon [16]	Considering the expected changes to the production frontier, an InvDEA model is presented to estimate of the outputs based on decision weights.
Amin and Al-Muharrami [1]	This work introduced new InvDEA models for pre-defined target setting of a merger with negative data.
Amin et al. [2]	This study suggested a new method to anticipate whether a merger in a market is generating a major or a minor consolidation.
Ghiyasi [23]	The InvDEA problems extended in which the technical and cost (revenue) efficiency of all DMUs are maintained simultaneously.
Xiao et al. [61]	A two-stage cost efficiency model provided by minimizing the cost of the new unit while it maintains the sum of the output levels of the merging units.
Eyni et al. [21]	The InvDEA problems extended in the presence of undesirable and desirable data.
Chen et al. [11]	A novel InvDEA model provided for analyzing the investment problem in the presence of undesirable outputs.
Ghobadi [26]	The problems of the InvDEA solved using the non-radial enhanced Russell [53] under fuzzy data.
Emrouznejad et al. [19]	A new InvDEA model is developed to achieve pre-defined goals to allocate CO2 emission quota in different Chinese regions.
Ghiyasi [24]	The problems of InvDEA extended in the presence of controllable and non-controllable inputs. Also, some trade off indexes proposed for tracking the relationship between needed input, producible output and efficiency improvement percentage.
Ghobadi [27]	A general model proposed for answer to the following question in the dynamic DEA framework: among a set of units, to what extent should the inputs/outputs of the unit change if the efficiency score of the unit remains unchanged, yet the outputs/inputs change?

Table 2: Some research in the field of InvDEA.

Reference	Short description
Ghiyasi [25]	This article, the InvDEA problems solved using novel criterion models. This leads to a reduction of computational complexity. Moreover, the proposed models solved some problematical fails of the InvDEA models.
Ghobadi and Jahangiri [30]	A new InvDEA model provided for optimal allocation of resources based on the ideal-solutions.
Wegener and Amin [60]	A novel InvDEA model presented to minimizing greenhouse gas emissions generated by a set of units for producing a certain level of outputs, provided that the units preserve at least their existing efficiency status.
Kalantary and Farzipoor Saen [43]	An InvDEA model with network and dynamic structure is proposed.
Zenodin and Ghobadi [62]	The idea of InvDEA used for input-estimation/output-estimation in merging the DMUs under the inter-temporal dependence assumption of input-output levels.
Soleimani-Chamkhorami et al. [56]	The proposed methods, in this study, respond to the following questions: To what extent outputs/inputs should be incremented in order to preserve the unit under evaluation revenue/cost-efficient, when the inputs/output of a unit increment?
Soleimani-Chamkhorami et al. [57]	The idea of InvDEA utilized for ranking extreme efficient units based on the growth potential.
Shinto and Sushama [55]	An Inverse DEA Model proposed to estimate of the input/output levels in the presence of integer data.
Ghobadi [28]	A novel InvDEA model provided for resource allocation in the dynamic DEA framework.
Hu et al. [37]	At first, this paper shown possible failures of radial InvDEA models due to neglect to slack variables. Then, to rectify this problem, a revised model presented where the investigated unit has no slack variable.
Guijarro et al. [33]	A new method proposed for units' restructuring. In this method, the identification of merging units and estimation of inputs is done by genetic algorithm and InvDEA, respectively.

Table 3: Some research in the field of InvDEA.

Reference	Short description
Ghobadi [29]	The concept of InvDEA utilized to identify of the inherited input/output levels of the merged unit from merging units to obtain a pre-defined efficiency score.
Amin and Ibn Boamah [4]	To estimate potential gains from bank mergers, the two-stage InvDEA models extended in this study.
Daryani et al. [15]	According to InvDEA concept, a four-stage method the proposed to estimate of the inputs and outputs with a two-stage network structure method.

then the estimators $\bar{x}_{ij}$ and $\bar{\sigma}^2_{ij}$ are unbiased estimators for "x_{ij}" and "σ^2_{ij}", respectively, as follows:

$$\bar{x}_{ij} = \frac{1}{\kappa}\sum_{k=1}^{\kappa}\check{x}^k_{ij}, \qquad i = 1, 2, ..., m,$$
$$\bar{\sigma}^2_{ij} = \frac{1}{\kappa - 1}\sum_{i=1}^{\kappa}(\check{x}^k_{ij} - \bar{x}_{ij})^2, \qquad i = 1, 2, ..., m.$$

Moreover, the estimators $\bar{y}_{rj}$ and $\bar{\psi}^2_{rj}$ are unbiased estimators for "y_{rj}" and "ψ^2_{rj}", respectively, as follows:

$$\bar{y}_{rj} = \frac{1}{\kappa}\sum_{k=1}^{\kappa}\check{y}^t_{rj}, \qquad r = 1, 2, ..., s,$$
$$\bar{\psi}^2_{rj} = \frac{1}{\kappa - 1}\sum_{k=1}^{\kappa}(\check{y}^t_{rj} - \bar{y}_{rj})^2 \qquad r = 1, 2, ..., s.$$

It is clear that the best estimators for mean and variance of outputs/inputs are $\bar{y}_{rj}$ ($\bar{x}_{ij}$) and $\bar{\psi}^2_{rj}$ ($\bar{\sigma}^2_{ij}$), respectively provided that the sample size ($\kappa \geq 30$) is large enough.

The stochastic input-oriented model evaluates the efficiency of DMU_o in the level of significance $\alpha \in (0, 1)$, DMU under consideration, by solving the following stochastic program [12]:

$$
\begin{aligned}
\theta_o(\alpha) = \min \quad & \theta && (1)\\
s.t. \qquad & P\{\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} \leq \ \theta \tilde{x}_{io}\} \geq 1-\alpha, \quad i=1,2,...,m,\\
& P\{\sum_{j=1}^{n} \lambda_j \tilde{y}_{rj} \geq \tilde{y}_{ro}\} \geq 1-\alpha, \quad r=1,2,...,s,\\
& \lambda = (\lambda_1, \ldots, \lambda_n) \in \Omega,
\end{aligned}
$$

where

$$
\begin{aligned}
\Omega = \{\lambda | \lambda = (\lambda_1, \ldots, \lambda_n),\ \delta_1(\sum_{j=1}^{n} \lambda_j + \delta_2(-1)^{\delta_3}\nu) = \delta_1,\\
\nu \geq 0,\ \lambda_j \geq,\ j=1,2,...,n\}.
\end{aligned}
$$

In this model, $\delta_1, \delta_2, \delta_3 \in \{0,\ 1\}$. It is obvious that if $\delta_1 = 0$, then the production technology is constant returns to scale (CRS); if $\delta_1 = 1, \delta_2 = 0$, then the production technology is variable returns to scale (VRS); if $\delta_1 = \delta_2 = 1$ and $\delta_3 = 0$, then the production technology is non-increasing returns to scale (NIRS); if $\delta_1 = \delta_2 = \delta_3 = 1$, then the production technology is non-decreasing returns to scale (NDRS). In addition, P means "Probability", $\alpha \in (0,1)$ is a pre-determined error level. In fact, (1-α) is the permissible error level of the constraints' violation defined by the manager. Also, $\theta, \lambda_1, ..., \lambda_n$ are decision variables. If $\theta_o(\alpha) = 1$, then DMU_o is called stochastic input-oriented weakly efficient in the level of significance α.

The following stochastic model is version of output-oriented of the model (1) in the level of significance α [12]:

$$
\begin{aligned}
\varphi_o(\alpha) = \max \quad & \varphi && (2)\\
s.t. \qquad & P\{\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} \leq \tilde{x}_{io}\} \geq 1-\alpha, \quad i=1,2,...,m,\\
& P\{\sum_{j=1}^{n} \lambda_j \tilde{y}_{rj} \geq \varphi \tilde{y}_{ro}\} \geq 1-\alpha, \quad r=1,2,...,s,\\
& \lambda = (\lambda_1, \ldots, \lambda_n) \in \Omega.
\end{aligned}
$$

In the above model, $\varphi, \lambda_1, ..., \lambda_n$ are decision variables. If $\varphi_o(\alpha) = 1$, then DMU_o is called stochastic output-oriented weakly efficient in the level of significance α. Let Φ be cumulative standard normal distribution function. Then, Models (1) and (2) can be converted to Models (3) and (6), respectively [12]:

$$\begin{aligned} \theta_o(\alpha) = \min \quad & \theta \\ s.t. \quad & \sum_{j=1}^{n} \lambda_j x_{ij} - \Phi^{-1}(\alpha)\sigma_i^I(\lambda, \theta) \leq \ \theta x_{io}, \quad i = 1, 2, ..., m, \\ & \sum_{j=1}^{n} \lambda_j y_{rj} + \Phi^{-1}(\alpha)\sigma_r^O(\lambda) \geq y_{ro}, \quad r = 1, 2, ..., s, \\ & \lambda = (\lambda_1, \ldots, \lambda_n) \in \Omega, \end{aligned} \tag{3}$$

where

$$\begin{aligned} \left(\sigma_i^I(\lambda, \theta)\right)^2 = & \textstyle\sum_{j=1}^{n} \sum_{k=1}^{n} \lambda_j \lambda_k Cov(\tilde{x}_{ij}, \tilde{x}_{ik}) + \theta^2 Var(\tilde{x}_{io}) \\ & -2\theta \textstyle\sum_{j=1}^{n} \lambda_j Cov(\tilde{x}_{ij}, \tilde{x}_{io}), \quad i = 1, 2, ..., m, \end{aligned} \tag{4}$$

$$\begin{aligned} \left(\sigma_r^O(\lambda)\right)^2 = & \textstyle\sum_{j=1}^{n} \sum_{k=1}^{n} \lambda_j \lambda_k Cov(\tilde{y}_{rj}, \tilde{y}_{rk}) + Var(\tilde{y}_{ro}) \\ & -2 \textstyle\sum_{j=1}^{n} \lambda_j Cov(\tilde{y}_{rj}, \tilde{y}_{ro}), \quad r = 1, 2, ..., s, \end{aligned} \tag{5}$$

$$\begin{aligned} \varphi_o(\alpha) = \max \quad & \varphi \\ s.t. \quad & \sum_{j=1}^{n} \lambda_j x_{ij} - \Phi^{-1}(\alpha)\sigma_i^I(\lambda) \leq \ x_{io}, \quad i = 1, 2, ..., m, \\ & \sum_{j=1}^{n} \lambda_j y_{rj} + \Phi^{-1}(\alpha)\sigma_r^O(\lambda, \varphi) \geq \varphi y_{ro}, \qquad r = 1, 2, ..., s, \\ & \lambda = (\lambda_1, \ldots, \lambda_n) \in \Omega, \end{aligned} \tag{6}$$

where

$$\begin{aligned} \left(\sigma_i^I(\lambda)\right)^2 = & \textstyle\sum_{j=1}^{n} \sum_{k=1}^{n} \lambda_j \lambda_k Cov(\tilde{x}_{ij}, \tilde{x}_{ik}) + Var(\tilde{x}_{io}) \\ & -2 \textstyle\sum_{j=1}^{n} \lambda_j Cov(\tilde{x}_{ij}, \tilde{x}_{io}), \quad i = 1, 2, ..., m, \end{aligned} \tag{7}$$

$$\left(\sigma_r^O(\lambda,\varphi)\right)^2 = \sum_{j=1}^n \sum_{k=1}^n \lambda_j \lambda_k Cov(\tilde{y}_{rj},\tilde{y}_{rk}) + \varphi^2 Var(\tilde{y}_{ro}) \\ -2\varphi \sum_{j=1}^n \lambda_j Cov(\tilde{y}_{rj},\tilde{y}_{ro}), \quad r=1,2,...,s. \tag{8}$$

It is worth noting that usually the *ith*-input (*rth*-output) in different units ($j = 1, 2, ..., n$) are uncorrelated. To clarify the discussion, consider a set of universities as decision-making units. Each university consumes different inputs, including the tuition fee, facilities (personnel and the educational spaces), and equipment to produce different outputs, including the number of graduates and the number of Ph.D. awards. It is clear that the facilities in different units are uncorrelated, as also the tuition fee and equipment. In addition, it is clear that the number of graduates in different units are uncorrelated, as also the number of Ph.D. awards. Therefore, in the relations of (4) and (7), the assumption of $Cov(\tilde{x}_{ij},\tilde{x}_{ik}) = 0$ $(\forall j, k, j \neq k)$ is not an unconventional assumption. Then, Eqs. (4) and (7) can be converted to Eqs. (9) and (11), respectively. Similarly, in the relations of (5) and (8), the assumption of $Cov(\tilde{y}_{rj},\tilde{y}_{rk}) = 0$ $(\forall j, k, j \neq k)$ is a correct assumption. Hence, Eqs. (5) and (8) can be converted into Eqs. (10) and (12), respectively.

$$\left(\sigma_i^I(\lambda,\theta)\right)^2 = \sum_{j=1}^n \lambda_j^2 Var(\tilde{x}_{ij})^2 + \theta Var(\tilde{x}_{io})^2(\theta - 2\lambda_o),\ i=1,2,...,m, \tag{9}$$

$$\left(\sigma_r^O(\lambda)\right)^2 = \sum_{j=1}^n \lambda_j^2 Var(\tilde{y}_{rj}) + Var(\tilde{y}_{ro})^2(1 - 2\lambda_o),\ r=1,2,...,s, \tag{10}$$

$$\left(\sigma_i^I(\lambda,\theta)\right)^2 = \sum_{j=1}^n \lambda_j^2 Var(\tilde{x}_{ij})^2 + Var(\tilde{x}_{io})^2(1 - 2\lambda_o),\ i=1,2,...,m, \tag{11}$$

$$\left(\sigma_r^O(\lambda)\right)^2 = \sum_{j=1}^n \lambda_j^2 Var(\tilde{y}_{rj}) + \varphi Var(\tilde{y}_{ro})^2(\varphi - 2\lambda_o),\ r=1,2,...,s. \tag{12}$$

Cooper et al. [12] showed that the models (3) and (6) are feasible for all the levels of significance $\alpha \in (0,1)$. Also, $0 < \theta_o(\alpha) \leq 1$ and $\varphi_o(\alpha) \geq 1$ for each $\alpha \in (0,\ 0.5]$. Moreover, if $\alpha \leq \alpha'$ then $\theta_o(\alpha') \leq \theta_o(\alpha)$ and $\varphi_o(\alpha') \geq \varphi_o(\alpha)$.

We close this section with a discussion on the case where models (3) and (6) can be converted linearly. Assume that the inputs and outputs of the units have the following structure:

$$\begin{aligned} \tilde{x}_{ij} &= x_{ij} + u_{ij}\tilde{\eta}_{ij},\ \tilde{\eta}_{ij} \sim N(0,\bar{\sigma}^2), \quad i = 1,2,...,m, \\ \tilde{y}_{rj} &= y_{rj} + v_{rj}\tilde{\xi}_{rj},\ \tilde{\xi}_{rj} \sim N(0,\bar{\sigma}^2) \quad r = 1,2,...,s, \end{aligned} \tag{13}$$

where u_{ij} and v_{rj} are non-negative real numbers. Also, $\tilde{\eta}_{ij}$ and $\tilde{\xi}_{rj}$ are errors of inputs and outputs in conflict with the mean values x_{ij} and y_{rj}, respectively. The data in the form of (13) are called data with a symmetric error structure. Suppose $\tilde{\eta}_{ij} = \tilde{\eta}_i$, $\tilde{\xi}_{rj} = \tilde{\xi}_r$, and the inputs and outputs of the DMUs are uncorrelated. Then, Models (3) and (6) can be converted to linear programming (LP) problems (14) and (15), respectively [6]:

$$\begin{aligned} \theta_o(\alpha) = \min \quad & \theta \\ s.t. \quad & \sum_{j=1}^{n} \lambda_j x_{ij} - \Phi^{-1}(\alpha)\bar{\sigma}(p_i^+ + p_i^-) \leq \ \theta x_{io}, \quad i = 1,2,...,m, \\ & \sum_{j=1}^{n} \lambda_j y_{rj} + \Phi^{-1}(\alpha)\bar{\sigma}(q_r^+ + q_r^-) \geq y_{ro}, \quad r = 1,2,...,s, \\ & \sum_{j=1}^{n} \lambda_j u_{ij} - \theta u_{io} = p_i^+ - p_i^-, \quad i = 1,2,...,m, \\ & \sum_{j=1}^{n} \lambda_j v_{rj} - v_{ro} = q_r^+ - q_r^-, \quad r = 1,2,...,s, \\ & \lambda \in \Omega, p_i^+ \geq 0, p_i^- \geq 0, q_r^+ \geq 0, q_r^- \geq 0, \forall i,r, \end{aligned} \tag{14}$$

$$\begin{aligned} \varphi_o(\alpha) = \max \quad & \varphi \\ s.t. \quad & \sum_{j=1}^{n} \lambda_j x_{ij} - \Phi^{-1}(\alpha)\bar{\sigma}(p_i^+ + p_i^-) \leq \ x_{io}, \quad i = 1,2,...,m, \\ & \sum_{j=1}^{n} \lambda_j y_{rj} + \Phi^{-1}(\alpha)\bar{\sigma}(q_r^+ + q_r^-) \geq \varphi y_{ro}, \quad r = 1,2,...,s, \end{aligned} \tag{15}$$

$$\sum_{j=1}^{n} \lambda_j u_{ij} - u_{io} = p_i^+ - p_i^-, \quad i = 1, 2, ..., m,$$
$$\sum_{j=1}^{n} \lambda_j v_{rj} - \varphi v_{ro} = q_r^+ - q_r^-, \quad r = 1, 2, ..., s,$$
$$\lambda \in \Omega, p_i^+ \geq 0, p_i^- \geq 0, q_r^+ \geq 0, q_r^- \geq 0, \forall i, r.$$

In the models (14) and (15), $\varphi,\ \theta,\ \mu_j,\ p_i^+,\ p_i^-,\ q_r^+$, and q_r^- are variables for all indices.

4 Merging DMUs in the Presence of Stochastic Data

One of the managers' approaches to improving the technical efficiency of units is to merge them. In business environments, the most common consolidations occur between banking units, which naturally implies improving their performance. In this regard, Gattoufi et al. [22] provided a novel method to reach the pre-defined target based on the InvDEA concept. However, the provided method could not be employed under stochastic data due to some modeling limitations in the traditional InvDEA models. Therefore, this section proposes a new method to deal with stochastic data in the problem of merging units. A set of units, $\{DMU_j,\ j \in \Lambda \subset J\}$, is assumed, where their performance is improved through merging. This set of units generates a new unit through synergy to reach the pre-defined efficiency score at the significance level α. Suppose the new unit and the set of units not participating in the merger are denoted by DMU_q and $\Pi = J - \Lambda$, respectively. The following assumptions are considered in the proposed approach: i) The set of units involved in the merger is deleted after the merger; ii) The new unit searches for the highest (lowest) possible output (input) level of the merged units to attain the pre-defined performance level while maintaining the sum of the input (output) levels of these units.

In the subsections 4.1 and 4.2, the proposed approach is provided to determine the lowest and highest possible inherited inputs and outputs, respectively.

4.1 Estimation of Lowest the Possible Inherited Inputs to Reach the Pre-defined Performance Level

According to the assumption (ii), the output levels of the new unit should be $\tilde{y}_{rq} = \sum_{j\in\Lambda} \tilde{y}_{rj}$ for each $r = 1, 2, ..., s$. In fact, $\tilde{y}_{rq}$ is the sum of the output levels of units involved in the merger process to achieve the pre-defined efficiency score in the level of significance α, $\bar{\theta}_q(\alpha)$. Suppose $\tilde{x}_{iq}(i = 1, 2, ..., m)$ represents the inputs obtained of the new unit after estimating them. The following model measures the efficiency score of DMU_q in the level of significance α:

$$\begin{aligned}
&\theta^*_q(\alpha) = \min \quad \theta && (16)\\
&s.t. \qquad P\{\sum_{j\in\Pi} \lambda_j \tilde{x}_{ij} + \lambda_q \tilde{x}_{iq} \leq \theta \tilde{x}_{iq}\} \geq 1-\alpha, \quad i = 1, 2, ..., m, &&\\
&\qquad\qquad P\{\sum_{j\in\Pi} \lambda_j \tilde{y}_{rj} + \lambda_q \tilde{y}_{rq} \geq \tilde{y}_{rq}\} \geq 1-\alpha, \quad r = 1, 2, ..., s, &&\\
&\qquad\qquad \lambda = (\lambda_j;\ j \in \Pi \cup \{q\}) \in \Omega_q. &&
\end{aligned}$$

where

$$\begin{aligned}
\Omega_q = \{\lambda | \lambda = (\lambda_j;\ j \in \Pi \cup \{q\}),\ &\delta_1(\sum_{j\in\Pi} \lambda_j + \lambda_q + \delta_2(-1)^{\delta_3}\nu) = \delta_1,\\
&\nu \geq 0,\ \lambda_j \geq 0,\ \forall j \in \Pi \cup \{q\}\}.
\end{aligned}$$

The variables vector in the above model is $(\lambda_j;\ j \in \Pi \cup \{q\},\ \theta)$. If the optimal value of Model (16) is equal $\bar{\theta}_q(\alpha)$, we say that the expected efficiency score of the generated unit has been achieved.

The following stochastic multiple-objective programming (SMOP) model proposed to estimate of the lowest possible input level of the new unit (DMU_q):

$$\begin{aligned}
&\min\ (\tilde{\alpha}_{ij};\ i = 1, 2, ..., m,\ \forall j \in \Lambda) && (17)\\
&s.t. \quad P\{\sum_{j\in\Pi} \lambda_j \tilde{x}_{ij} + \lambda_q(\sum_{j\in\Lambda} \tilde{\alpha}_{ij}) \leq \bar{\theta}_q(\alpha) \sum_{j\in\Lambda} \tilde{\alpha}_{ij}\} \geq 1-\alpha,\ i = 1, 2, ..., m, &&\\
&\qquad P\{\sum_{j\in\Pi} \lambda_j \tilde{y}_{rj} + \lambda_q \tilde{y}_{rq} \geq \tilde{y}_{rq}\} \geq 1-\alpha, \quad r = 1, 2, ..., s, &&
\end{aligned}$$

$$
\begin{aligned}
&P\{\tilde{\alpha}_{ij} \leq \tilde{x}_{ij}\} \geq 1-\alpha, \quad i=1,2,...,m,\ \forall j \in \Lambda,\\
&P\{\tilde{\alpha}_{ij} \geq 0\} \geq 1-\alpha, \qquad\quad i=1,2,...,m,\ \forall j \in \Lambda,\\
&\lambda = (\lambda_j;\ j \in \Pi \cup \{q\}) \in \Omega_q.
\end{aligned}
$$

In the above model $(\lambda_j;\ j \in \Pi \cup \{q\}, \tilde{\alpha}_{ij};\ i=1,2,...,m,\ \forall j \in \Lambda)$ is the variables vector. $\bar{\theta}_q(\alpha)$ is the expected efficiency goal in the level of significance α for the new unit (DMU_q). In this model, $\tilde{\alpha}_{ij}$ shows the amount of the ith input inherited by DMU_q from the jth unit of participating in the merger process. The goals of the InvDEA model (17) guarantees that the inherited input levels by DMU_q from units involved in the merger process are minimized to achieve efficiency score in the level of significance α, $\bar{\theta}_q(\alpha)$. In fact, saving resources is the main target to achieve the most benefit through merging. The first, second, and final sets of constraints in Model (17) ensure that the generated new unit has the expected performance in the level of significance α. The third set of constraints in the model (17) ensures that the amount of resources received by the new unit does not exceed from the amount of resources available to each of the units involved in the integration process.

Due to the special structure Model (17), a long with [32], stochastic (weak) Pareto solutions are defined as follows:

Definition 4.1. A feasible solution $\Delta = (\lambda^*, \tilde{\alpha}^*_{ij};\ i=1,2,...,m,\ \forall j \in \Lambda)$ is called a stochastic Pareto (SP) solution in the level of significance α to SMOP (17) if there is no other feasible solution $\Gamma = (\lambda, \tilde{\alpha}_{ij};\ i = 1,2,...,m,\ \forall j \in \Lambda)$ such that

$$
\begin{aligned}
&P\{\tilde{\alpha}_{ij} - \tilde{\alpha}^*_{ij} \leq 0\} \geq 1-\alpha \quad \text{for each} \quad i=1,2,\ldots,m,\ \forall j \in \Lambda,\\
&P\{\tilde{\alpha}_{ij} - \tilde{\alpha}^*_{ij} \leq -\epsilon\} \geq 1-\alpha \quad \text{for some} \quad i=1,2,\ldots,m,\ \&j \in \Lambda,
\end{aligned}
$$

where ϵ is a non-archimedian infinitesimal.

Definition 4.2. A feasible solution $\Delta = (\lambda^*, \tilde{\alpha}^*_{ij};\ i=1,2,...,m,\ \forall j \in \Lambda)$ is called a stochastic weak Pareto (SWP) solution in the level of significance α to SMOP (17) if there is no other feasible solution $\Gamma = (\lambda, \tilde{\alpha}_{ij};\ i=1,2,...,m,\ \forall j \in \Lambda)$ such that

$$
P\{\tilde{\alpha}_{ij} - \tilde{\alpha}^*_{ij} \leq -\epsilon\} \geq 1-\alpha \quad \text{for each} \quad i=1,2,\ldots,m,\ \forall j \in \Lambda,
$$

where ϵ is a non-Archimedian infinitesimal.

In practical, the most common merger of units occur to improve their respective performances. Therefore, it can be assumed that the units involved in the integration process are not technically efficient. According to the concepts of DEA, this assumption implies that the elimination of units involved in the integration process does not change efficiency frontiers in the level of significance α. Accordingly, if the new unit is within the PPS in the level of significance α, then DMU_q can be presented by some of the other efficient units in the level of significance α. This is achieved based on a similar process by converting the stochastic model (1) to the linear programming model (14) and assuming that the new unit is placed in the current PPS. In this case, in each SP solution in the level of significance α to SMOP-model (17), we have $\lambda_q^* = 0$.

According to the above discussion, model (17) can be converted to the following revised SMOP (RSMOP) model:

$$\begin{aligned}
\min \quad & (\tilde{\alpha}_{ij};\ i = 1, 2, ..., m,\ \forall j \in \Lambda) \\
s.t. \quad & P\{\sum_{j \in \Pi} \lambda_j \tilde{x}_{ij} \leq \bar{\theta}_q(\alpha) \sum_{j \in \Lambda} \tilde{\alpha}_{ij}\} \geq 1 - \alpha, \quad i = 1, 2, ..., m, \\
& P\{\sum_{j \in \Pi} \lambda_j \tilde{y}_{rj} \geq \tilde{y}_{rq}\} \geq 1 - \alpha, \quad r = 1, 2, ..., s, \\
& P\{\tilde{\alpha}_{ij} \leq \tilde{x}_{ij}\} \geq 1 - \alpha, \quad i = 1, 2, ..., m,\ \forall j \in \Lambda, \\
& P\{\tilde{\alpha}_{ij} \geq 0\} \geq 1 - \alpha, \quad i = 1, 2, ..., m,\ \forall j \in \Lambda, \\
& \lambda = (\lambda_j;\ \forall j \in \Pi) \in \overline{\Omega}_q,
\end{aligned} \tag{18}$$

where

$$\overline{\Omega}_q = \{\lambda | \lambda = (\lambda_j;\ \forall j \in \Pi),\ \delta_1(\sum_{j \in \Pi} \lambda_j + \delta_2(-1)^{\delta_3}\nu) = \delta_1,\ \nu \geq 0,\ \lambda_j \geq 0,\ \forall j \in \Pi\}.$$

It is worth noting that the current study is confined to the case of the merging where the generated new unit is inside or on the frontier of efficiency of the current PPS. Therefore, the discussed topic in this paper without consideration this restriction can be a worthwhile direction for further research, though this paper does not pursue it. Let us to consider the virtual unit (DMU_v), in which DMU_v consumes multiple stochastic inputs $\tilde{x}_{iv} = \sum_{j\in\Lambda} \tilde{x}_{ij}$ $(i = 1, 2, ..., m)$ to produce multiple stochastic outputs $\tilde{y}_{rv} = \sum_{j\in\Lambda} \tilde{y}_{rj}$ $(r = 1, 2, ..., s)$. According to the goals of Models (17) and (18), efforts are made to maintain the minimum level of the sources of the virtual unit. As a result, necessary and sufficient conditions for $DMU_q \in PPS$ is that $DMU_v \in PPS$.

According to the following theorem, the input-oriented model (18) can be used for determining the input levels of the produced new unit (DMU_q).

Theorem 4.3. *Suppose that deleting units participating in the merger process does not change the PPS. Also, let $\Delta = (\lambda^*, \tilde{\alpha}^*_{ij} : \ i = 1, 2, ..., m, \forall j \in \Lambda)$ be a SP solution to Model (18). If $\tilde{x}_{iq} = \sum_{j\in\Lambda} \tilde{\alpha}^*_{ij}$ $(i = 1, 2, ..., m)$, then the optimal value of Model (16) is equal to $\bar{\theta}_q(\alpha)$.*

Proof. See Appendix A. □

We close this subsection with a discussion on minimum achievable efficiency target in the level of significance α of the merged DMU. It is worth noting that knowing minimum achievable efficiency target is necessary for the decision maker deliberating about engaging in the merging process. The lowest efficiency score realized through a merging could be determined via the following theorem.

Theorem 4.4. *Consider a merging with $\bar{\theta}_q(\alpha)$ as the efficiency target in the level of significance α for the merged DMU.*

i) If model (18) is feasible, then it remains feasible for each efficiency target $\bar{\bar{\theta}}_q(\alpha)$, with $\bar{\theta}_q(\alpha) \le \bar{\bar{\theta}}_q(\alpha) \le 1$.

ii) Suppose that model (18) admits a stochastic Pareto solution. If

$$
\begin{aligned}
\theta^*_q(\alpha) = \min \quad & \theta \\
s.t. \quad & P\{\sum_{j\in\Pi} \lambda_j \tilde{x}_{ij} \le \sum_{j\in\Lambda} \tilde{\alpha}_{ij}\} \ge 1-\alpha, \quad i = 1, 2, ..., m, \\
& P\{\sum_{j\in\Pi} \lambda_j \tilde{y}_{rj} \ge \sum_{j\in\Lambda} \tilde{y}_{rj}\} \ge 1-\alpha, \quad r = 1, 2, ..., s, \\
& P\{\sum_{j\in\Lambda} \tilde{\alpha}_{ij} \le \theta \sum_{j\in\Lambda} \tilde{x}_{ij}\} \ge 1-\alpha, \quad i = 1, 2, ..., m, \\
& P\{\tilde{\alpha}_{ij} \le \tilde{x}_{ij}\} \ge 1-\alpha, \quad i = 1, 2, ..., m, \ \forall j \in \Lambda, \\
& P\{\tilde{\alpha}_{ij} \ge 0\} \ge 1-\alpha, \quad i = 1, 2, ..., m, \ \forall j \in \Lambda,
\end{aligned}
$$

$$\bar{\lambda} = (\bar{\lambda}_j;\ j \in \Pi) \in \overline{\Omega}_q,\ \theta \geq 0,$$

then $\bar{\theta}_q(\alpha) \geq \theta^*_q(\alpha)$ *(The minimum achievable efficiency target in the level of significance* α *of the merged DMU is* $\theta^*(\alpha)$*).*

Proof. According to the model (18), the proof of part (i) is obvious. By contradiction assume that $\bar{\theta}_q(\alpha) < \theta^*(\alpha)$. Let $\Delta = (\mu^*, \tilde{\alpha}^*_{ij} : i = 1, 2, ..., m, \forall j \in \Lambda)$ be a stochastic Pareto solution to model (18). Feasibility of Δ to model (18), implies:

$$P\{\tilde{\alpha}^*_{ij} \leq \tilde{x}_{ij}\} \geq 1 - \alpha, \quad i = 1, 2, ..., m,\ \forall j \in \Lambda. \tag{19}$$

By (19) and $0 < \bar{\theta}_q(\alpha) \leq 1$, we get

$$P\{\bar{\theta}_q(\alpha)\tilde{\alpha}^*_{ij} \leq \tilde{x}_{ij}\} \geq 1 - \alpha, \quad i = 1, 2, ..., m,\ \forall j \in \Lambda.$$

Now, define $\bar{\tilde{\alpha}}_{ij} := \bar{\theta}_q(\alpha)\tilde{\alpha}^*_{ij}$ for each $i = 1, 2, ..., m$, and $j \in \Lambda$. Therefore,

$$P\{\sum_{j\in\Lambda} \bar{\tilde{\alpha}}_{ij} \leq \theta_q(\alpha) \sum_{j\in\Lambda} \tilde{x}_{ij}\} \geq 1 - \alpha, \quad i = 1, 2, ..., m. \tag{20}$$

By (20) and feasibility of Δ to model (18), it is obvious that $\Psi = (\mu^*,\ \bar{\theta}_q(\alpha),\ \bar{\tilde{\alpha}}_{ij} : i = 1, 2, ..., m,\ \forall j \in \Lambda)$ is a feasible solution to model (19). Therefore, the optimal value of model (19) is less than or equal to $\bar{\theta}_q(\alpha)$. This contradicts the assumption that θ^* is the optimal value of model (19), and so the proof of part (ii) is completed. □

The following remarks discuss a special case of linearization of stochastic models (16), (18), and (19).

Remark 4.5. *If the input and output levels have a normal distribution with a symmetric error structure, then the same method for converting model (1) to model (14) can be employed to convert the proposed stochastic models (16), (18), and (19) into linear programming models. According to the operations research literature, there are various methods to solve these problems, including the weight sum method [17]. In this study, we also use this method because it allows the decision maker to pursue different goals of merging (for example, saving more inputs from a particular unit).*

4.2 Estimation of Highest the Possible Inherited Outputs to Reach the Pre-defined Performance Level

In this subsection, we assume that the generated new unit keeps the amount of input levels of set of the merging units ($\{DMU_j,\ j \in \Lambda \subset J\}$) and searches for the maximum amount of outputs of these DMUs to achieve the desired efficiency score in the level of significance α, $\bar{\varphi}_q(\alpha)$. In fact, assumption (ii) is satisfied. Therefore, the input levels of the new generated DMU should be $\tilde{x}_{iq} = \sum_{j\in\Lambda} \tilde{x}_{ij}$ for each $i = 1, 2, ..., m$. In other words, $\tilde{x}_{iq}$ is the sum of the input levels of the participating units in the merger process in order to reach the desired given efficiency target in the level of significance α, $\bar{\varphi}_q(\alpha)$. If $\tilde{y}_{rq}(r = 1, 2, ..., s)$ represents the generated outputs of the new unit after estimating them, then following model can be measures the efficiency score of DMU_q in the level of significance α:

$$\begin{aligned} \varphi_q^*(\alpha) = \max \quad & \varphi \qquad\qquad (21)\\ s.t. \quad & P\{\sum_{j\in\Pi} \lambda_j \tilde{x}_{ij} + \lambda_q \tilde{x}_{iq} \leq \tilde{x}_{iq}\} \geq 1-\alpha, \quad i = 1, 2, ..., m,\\ & P\{\sum_{j\in\Pi} \lambda_j \tilde{y}_{rj} + \lambda_q \tilde{y}_{rq} \geq \varphi \tilde{y}_{rq}\} \geq 1-\alpha, \quad r = 1, ..., s,\\ & \lambda = (\lambda_j;\ j \in \Pi \cup \{q\}) \in \Omega_q. \end{aligned}$$

In the above model, the variables vector is $(\lambda_j;\ j \in \Pi \cup \{q\},\ \varphi)$. If the optimal value of problem (21) is equal $\bar{\varphi}_q(\alpha)$, we say that the expected efficiency score of the new generated DMU has been realized.

To estimate of the highest possible output levels of the new unit (DMU_q), we consider the following SMOP model:

$$\begin{aligned} & \max\ (\tilde{\beta}_{rj};\ r = 1, 2, ..., s,\ \forall j \in \Lambda) \qquad\qquad (22)\\ & s.t.\ P\{\sum_{j\in\Pi} \lambda_j \tilde{x}_{ij} + \lambda_q \tilde{x}_{iq} \leq \tilde{x}_{iq}\} \geq 1-\alpha, \quad i = 1, 2, ..., m,\\ & P\{\sum_{j\in\Pi} \lambda_j \tilde{y}_{rj} + \lambda_q (\sum_{j\in\Lambda} \tilde{\beta}_{rj}) \geq \bar{\varphi}_q(\alpha) \sum_{j\in\Lambda} \tilde{\beta}_{rj}\} \geq 1-\alpha,\ r = 1, ..., s,\\ & P\{\tilde{\beta}_{rj} \geq \tilde{y}_{rj}\} \geq 1-\alpha, \quad r = 1, 2, ..., s,\ \forall j \in \Lambda,\\ & \lambda = (\lambda_j;\ j \in \Pi \cup \{q\}) \in \Omega_q. \end{aligned}$$

In Model (22), $(\lambda_j;\ j \in \Pi \cup \{q\}, \tilde{\beta}_{rj};\ r = 1,2,...,s,\ \forall j \in \Lambda)$ is the variables vector. $\bar{\varphi}_q(\alpha)$ is the expected efficiency score in the level of significance α for the generated new unit (DMU_q). In the above model, $\tilde{\beta}_{rj}$ shows the amount of the rth output inherited by the new unit DMU_q from the jth DMU of participating in the merger process. The objectives of the InvDEA model (22) ensures that the inherited output level by DMU_q from units involved in the merger process are maximized in order to achieve the desired efficiency goal in the level of significance α, $\bar{\varphi}_q(\alpha)$. In other words, in here, the most profits are made through mergers based on the production of more products. Also, the third set of constraints in the model (22) ensures that the amount of output generated by the new unit is greater than or equal to the existing output of each of the units involved in the integration process. In a method similar with convert Model (17) to (18), we can convert the InvDEA model (22) to the following revised SMOP (RSMOP) model:

$$
\begin{aligned}
\max \quad & (\tilde{\beta}_{rj};\ r = 1,2,...,s,\ \forall j \in \Lambda) \\
s.t. \quad & P\{\sum_{j \in \Pi} \lambda_j \tilde{x}_{ij} \leq \tilde{x}_{iq}\} \geq 1 - \alpha, \quad i = 1,2,...,m, \\
& P\{\sum_{j \in \Pi} \lambda_j \tilde{y}_{rj} \geq \bar{\varphi}_q(\alpha) \sum_{j \in \Lambda} \tilde{\beta}_{rj}\} \geq 1 - \alpha, \quad r = 1,2,...,s, \\
& P\{\tilde{\beta}_{rj} \geq \tilde{y}_{rj}\} \geq 1 - \alpha, \quad r = 1,2,...,s,\ \forall j \in \Lambda, \\
& \lambda = (\lambda_j;\ j \in \Pi) \in \overline{\Omega}_q.
\end{aligned}
\tag{23}
$$

To estimate of the output levels of the generated new unit, the following theorem can be used.

Theorem 4.6. *Assume that the integration process is such that deleting participating units does not change the PPS. Let the generated new unit aims to be equal to $\bar{\varphi}_q(\alpha)$. In addition, let $\Delta = (\lambda^*, \tilde{\beta}^*_{rj} :\ r = 1,2,...,s,\ \forall j \in \Lambda)$ be a SP solution to Model (23). If $\tilde{y}_{rq} = \sum_{j \in \Lambda} \tilde{\beta}^*_{rj}$ $(r = 1,2,...,s)$, then the optimal value of Model (21) is equal to $\bar{\varphi}_q(\alpha)$.*

Proof. See Appendix B. □

It is worth noting that knowing highest receivable efficiency score has an decisive role for assessment of the manager's about engaging in the integration process. To estimate the highest receivable efficiency score of the generated new unit, the following theorem can be used.

Theorem 4.7. *Consider a merging with $\bar{\varphi}_q$ as the efficiency score in the level of significance α for the generated new unit.*

i) If model 23 is feasible, then it remains feasible for each efficiency target $\bar{\bar{\varphi}}_q(\alpha)$, with $1 \leq \bar{\bar{\varphi}}_q(\alpha) \leq \bar{\varphi}_q(\alpha)$.

ii) Suppose that model 23 admits a SP solution. If

$$\begin{aligned}
\varphi_q^*(\alpha) = \max \quad & \varphi \\
s.t. \quad & P\{\sum_{j\in\Pi} \lambda_j \tilde{x}_{ij} \leq \sum_{j\in\Lambda} \tilde{x}_{ij}\} \geq 1-\alpha, \quad i=1,2,...,m, \\
& P\{\sum_{j\in\Pi} \lambda_j \tilde{y}_{rj} \geq \sum_{j\in\Lambda} \tilde{\beta}_{rj}\} \geq 1-\alpha, \quad r=1,2,...,s, \\
& P\{\sum_{j\in\Lambda} \tilde{\beta}_{rj} \geq \varphi \sum_{j\in\Lambda} \tilde{y}_{rj}\} \geq 1-\alpha, \quad r=1,2,...,s, \\
& P\{\tilde{\beta}_{rj} \geq \tilde{y}_{rj}\} \geq 1-\alpha, \quad r=1,2,...,s, \\
& \lambda = (\lambda_j,\ \forall j \in \Pi),\ \varphi \geq 1,
\end{aligned} \tag{24}$$

then $\bar{\varphi}_q(\alpha) \leq \varphi^*(\alpha)$ (The maximum achievable efficiency score of the merged DMU is $\varphi^*(\alpha)$).

Proof. The proof of part (i) is straightforward. By contradiction assume that $\bar{\varphi}_q(\alpha) > \varphi^*(\alpha)$. Let $\Delta = (\lambda^*, \tilde{\beta}^*_{rj} :\ r=1,2,...,s, \forall j \in \Lambda)$ be a SP solution to model (23). Feasibility of Δ to model (23), implies:

$$P\{\tilde{\beta}^*_{rj} \geq \tilde{y}_{rj}\} \geq 1-\alpha, \quad r=1,2,...,s,\ \forall j \in \Lambda. \tag{25}$$

By (25) and $\bar{\varphi}_q(\alpha) \geq 1$, we get

$$P\{\bar{\varphi}_q(\alpha)\tilde{\beta}^*_{rj} \geq \tilde{y}_{rj}\} \geq 1-\alpha, \quad r=1,2,...,s,\ \forall j \in \Lambda.$$

Now, for all r and $j \in \Lambda$ define $\bar{\tilde{\beta}}_{rj} := \bar{\varphi}_q(\alpha)\tilde{\beta}^*_{rj}$. Therefore,

$$P\{\sum_{j\in\Lambda} \bar{\tilde{\beta}}_{rj} \geq \varphi_q(\alpha) \sum_{j\in\Lambda} \tilde{y}_{rj}\} \geq 1-\alpha, \quad r=1,2,...,s. \tag{26}$$

By (26) and feasibility of Δ to model (23), it is obvious that $\Psi = (\lambda^*,\ \bar{\varphi}_q(\alpha),\ \bar{\tilde{\beta}}_{rj} : r = 1, 2, ..., s,\ \forall j \in \Lambda)$ is a feasible solution to model (24). Therefore, the optimal value of model (24) is greater than or equal to $\bar{\varphi}_q(\alpha)$. This contradicts the assumption that φ^* is the optimal value of model (24), and so the proof of part (ii) is completed. $\square$

Remark 4.8. *If the input and output levels have a normal distribution with a symmetric error structure, then the same method for converting model (2) to model (15) can be employed to convert the proposed stochastic models (21), (23), and (24) into linear programming models.*

Remark 4.9. *It is worth noting that in the real world usually the most common merger of units occur to improve their respective performances. Therefore, it can be assumed that the units involved in the integration process are not technically efficient in the level of significance α. According to the concepts of DEA, this assumption implies that the elimination of units involved in the integration process does not change efficiency frontiers. However, it is clear that the new unit may be within or outside the current PPS in the level of significance α. According to the considered assumption in the article, the our study is limited to the case of the merging where the merged DMU is inside the current PPS. Accordingly, the issue discussed in this study can be worth studying without consideration this assumption as well, although this paper does not pursue it. Also, it is clear that the new generated unit will be within of the current PPS, if and only if the virtual unit $(\sum_{j\in\Lambda} \tilde{x}_{ij}, \sum_{j\in\Lambda} \tilde{y}_{rj})$ for all $i \in I$ and $r \in O$; is inside the PPS in the level of significance α. This stems from the objectives of the SMOP input-(resp.output-) oriented models (17) (resp. (22)) as well as the objectives of the relaxed input-(resp. output-) oriented models (18) (resp. (23)). This condition ensures that the merged DMU is within the PPS when it is inefficient or on the border once it is efficient. Then, the relaxed models (18) and (23) can replace the models (17) and (22), respectively.*

5 An Application

The current section provides and evaluates the InvDEA method through a real-world data set. We considered a dataset containing 20 branches of an Iranian commercial bank and presented in Tables (13) and (14) in Appendix C [6]. For each unit, a random independent sample of size 40 is considered and then the mean and variance of each branch are estimated separately for each of the input and output factors. Moreover, each of the input factors in different units ($j = 1, 2, ..., 20$) and each of the output factors are uncorrelated. According to the literature, two main approaches are proposed to select input and output factors: the intermediation approach and the production approach. In the intermediation approach, banks are considered financial intermediaries between the liability and the fund beneficiaries, while in the production approach, banks are considered service providers to customers. The operational activities are significant in the production approach. In contrast, converting the earned funds into loans is critical in the production approach [8, 51]. Therefore, commissions and loans are considered the output factors, while deposits, employees, and fixed assets are the input factors in the intermediation approach. However, the deposit is treated as the output factor because it is considered a bank's service to its customers. In other words, maintenance of customer deposits, financial transactions, customers' financial document processing, and other bank services are significant in this approach.

Over the last two decades, bank mergers and acquisitions have been occurring at an unprecedented rate. There are four main paths identified that explains the reasons behind the mergers/acquisitions activity. These four paths are related to (1) creating economies of scales, (2) expanding geographically, (3) increasing the combined capital base (size) and product offerings, and (4) gaining market power. As discussed in the introduction, data uncertainty (imprecise or vague data) is a primary and significant problem in employing the traditional DEA. Different strategies have been presented in the literature to deal with these inaccurate and ambiguous data. One of the most important and widely used strategies based on uncertainty theory in UDEA is the SDEA approach. According to the structure of conventional DEA models, any deviation from the efficiency boundary is considered inefficiency and there is no chance for random noise in nature. This is something to ponder when managers deal with units with inaccurate input and output levels in different real-world situations. It is worth noting that it is possible to identify different data features while working with random variables based on the probability of unforeseen events. This is a crucial feature that encourages the decision-maker to employ SDEA. In fact, the primary benefit of working with stochastic data in DEA is the ability to predict the upcoming efficiencies. As mentioned above, in the evaluation of a bank in the evaluation of bank branches, various indicators can be taken into consideration, including the amount of deposits. It is clear that this index is a normal random variable in a period of 30 days. The merger of two branches is a big step that if the randomness approach is included in the estimation of this index for the new branch, managers can better predict the future performance of this branch.

In the present study, the input and output factors are identified based on the intermediation approach and are shown in Table 4. In fact, each branch produces five outputs (the facilities, amount of deposits, received benefits, received commission, and other resources of deposits) using three inputs (the personal rate, payable benefits, and delayed requisitions). It is worth noting that the personal rate is considered based on the weighted combination of personal qualifications, education, quantity, and other items. Also, the payable benefits of all deposits to customers and delays in repaying loans and other facilities in each branch are considered as the payable benefits and delay claims, respectively. Moreover, the sum of business and single loans, the value of various deposits (including current, and short/long duration accounts), the received benefits from the total loans and facilities, and the sum received commission (including of all banking actions, issuance guaranty, money transfer, and others) in each branch are considered as the facilities, the amount of deposits, the received benefits, and the received commission, respectively.

Using the model (14) under CRS and $\bar{\sigma} = 1$ assumptions, we obtained the units' efficiency index in the levels of significance $\alpha = 0.01,\ 0.05,$ $0.1,\ 0.2,\ 0.5$ (see Table 5). Here, because the input and output levels of the units have a symmetric error structure, LP models (14) can be used to estimate the efficiency index at different significant levels.

Table 5 shows that B04, B07, and B18 branches are inefficient in the levels of significance $\alpha = 0.01$. In other words, with a confidence factor of 99%, B04, B07, and B18 branches have the score efficiencies

Table 4: Input and output factors.

Input factors	Output factors
The personal rate ($\tilde{x}_1$)	The facilities ($\tilde{y}_1$)
The payable benefits ($\tilde{x}_2$)	The amount of deposits ($\tilde{y}_2$)
The delayed requisitions ($\tilde{x}_3$)	The received benefits ($\tilde{y}_3$)
	The received commission ($\tilde{y}_4$)
	The other resources of deposits ($\tilde{y}_5$)

Table 5: The efficiency index of 20 bank branches at different significant levels.

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

0.33123, 0.56020, and 0.62068, respectively. Merging these branches can be a practical solution to improve their performance. Assume that these units merge their activities by creating a new unit (B_q), so that its output levels (facilities, amount of deposits, received benefits, received commission, and other resources of deposits) equal to $\tilde{y}_{rq} = \tilde{y}_{r4} + \tilde{y}_{r7} + \tilde{y}_{r18}$ $(r = 1, 2, ..., 5)$ as shown in Tables 6 and 7.

Table 6: The output levels $(\tilde{y}_1,\ \tilde{y}_2,\ \tilde{y}_3)$ of merging branches and new branch.

Outputs	Facilities		Amount of deposits		Received benefits	
DMU	Mean	Variance	Mean	Variance	Mean	Variance
DMU04	137.51	21.65	44.972	13.78	4.923	1.65
DMU07	192.97	14.56	78.015	19.56	7.791	3.56
DMU18	259.21	34.21	81.779	21.65	5.212	1.94
DMUq	589.690	70.420	204.766	54.990	17.926	7.150

Table 7: The output levels $(\tilde{y}_4,\ \tilde{y}_5)$ of merging branches and new branch.

Outputs	Received commission		Other resources of deposits	
DMU	Mean	Variance	Mean	Variance
DMU04	4.212	3.84	63.61	33.73
DMU07	15.89	2.34	7.499	1.268
DMU18	8.021	3.52	107.9	13.07
DMUq	28.123	9.700	179.009	48.068

Here, the set of units intended for process integration is $\Lambda = \{4, 7, 18\}$ and therefore $\Pi = \{1, 2, ..., 20\} - \Lambda$. Using the model (19), the lowest achievable efficiency index in the significance levels of $\alpha = 0.01$ is obtained by the new unit equal to $\theta^*(\alpha) = 0.4627$. Obviously, if this level of efficiency is assessed as satisfactory for the new branch, the merger process can begin. With regard to Theorem 4.4, the model (18) will be feasible, if $\bar{\theta}_q(0.01) \geq 0.4627$. Assume that B04, B07, and B18 branches are merged with two different expected efficiency scores in the levels of significance $\alpha = 0.01$. To obtain these expected performance scores, the outputs of the merged unit must be in accordance with Tables 6 and 7. According to these tables, the new branch should achieve the averages of 589.690 and 204.766 with variances of 70.420 and 54.990 in the indicators of facilities and the amount of deposits. Therefore, the relevant managers should create the conditions for this synergy through the merging branches. First, assume that the efficiency goal of the generated new branch B_q is equal to 0.85 in the levels of significance $\alpha = 0.01$ ($\bar{\theta}_q(\alpha) = 0.85$). The proposed model (18) is employed for this integration, and two SP solutions are generated to estimate the input levels using the weight-sum method [17] as written in Tables 8 and 9. Indeed, to achieve the efficiency goal ($\bar{\theta}_q(\alpha) = 0.85$) in the levels of significance $\alpha = 0.01$, Tables 8 and 9 present two possible scenarios for the inherited input values of the new branch Bq from branches B04, B07 and B18. For example, if the manager chooses the first scenario to form the new unit Bq, then the amount of resources required in each of the input factors the personal rate, payable benefits, and delayed requisitions are equal to N(17.60, 0.50), N(36.69, 82.20), and N(55.15, 49.31), respectively. In fact, to obtain this expected performance score, the new branch must receive averages of 17.60, 36.69, and 55.15 on the Personal Rate, Benefits Payable, and Delayed Requisitions indicators. Therefore, the relevant managers should provide the needed resources to the new branch through the synergy of the merging branches. According to this scenario, the personal rate (the first input) of the new branch Bq should be provided by the personal rate of the three merging branches. According to Table 8, the amount of supply required by each of the B04, B07, and B18 branches is equal to N(2.49, 0.29), N(0.00, 0.01) and N(15.11, 0.20), respectively. Moreover, the share of three branches B04, B07, and B18 in the personal rate of the new branch q based on the first SP solution to achieve the efficiency goal ($\bar{\theta}_q(0.01) = 0.85$) is shown in the figure 1. This figure shows that the share of branch B18 in providing the resources needed of the new branch q in the personal rate index is higher than other branches. Therefore, managers should provide the conditions for this transfer of resources. It is worth noting that the branch B07 has almost no shares. In a similar discussion, the amount of supply required to the input second (payable benefits) of the new branch Bq (N(36.69, 82.20)) by each of the B04, B07, and B18 branches is equal to N(0.00, 18.54), N(0.00, 3.01) and N(36.69, 60.65), respectively. Moreover, the share of three branches B04, B07, and B18 in the payable benefits of the new branch q according to the first scenario to attain the efficiency index

($\bar{\theta}_q(0.01) = 0.85$) is shown in the figure 2. This figure shows that the share of branch B18 in the payable benefits of the new branch q is higher than other branches. Also, branches B04 and B07 have almost no shares. According to figures 1 and 2, branch B18 plays a key role in providing the two primary sources of the new branch. Obviously, branches B04 and B07 have almost no contribution, and the main contribution belongs to branch B18. Also, the provided amount to the input three (delayed requisitions) of the new branch Bq (N(55.15, 49.31)) by each of the B04, B07, and B18 branches is equal to N(0.00, 10.62), N(55.15, 30.02) and N(0.00, 8.67), respectively. Moreover, the share of three branches B04, B07, and B18 in the delayed requisitions of the new branch q according to the first SP solution to achieve the efficiency goal ($\bar{\theta}_q(0.01) = 0.85$) is shown in the figure 3. This figure shows that the share of branch B07 in the delayed requisitions of the new branch q is higher than other branches. Also, branches B04 and B18 have almost no shares. Therefore, branch B07 plays a key role in providing the needed resources for the third index of the new branch. Indeed, figure 3 shows that the main share of the delayed requisitions of new branch q is borne by the B07 branch.

Table 8: The first proposed input levels for the merged branch to achieve the efficiency goal 0.85 in the significance level of 0.01.

Inputs	Personal rate $\tilde{\alpha}^1_{1j}$		Payable benefits $\tilde{\alpha}^1_{2j}$		Delayed requisitions $\tilde{\alpha}^1_{3j}$	
DMU	Mean	Variance	Mean	Variance	Mean	Variance
B04	2.49	0.29	0.00	18.54	0.00	10.62
B07	0.00	0.01	0.00	3.01	55.15	30.02
B18	15.11	0.20	36.69	60.65	0.00	8.67
Bq	17.60	0.50	36.69	82.20	55.15	49.31

Moreover, the share of three branches B04, B07, and B18 in supply inputs of the new branch q based on the second SP solution to achieve the efficiency goal ($\bar{\theta}_q(0.01) = 0.85$) are shown in the figures 4, 5, and 6. According to this figures, it is clear that the share of branch B18 in providing the needed resources related to the personal rate index of the new branch q is higher than other branches. Moreover, share of

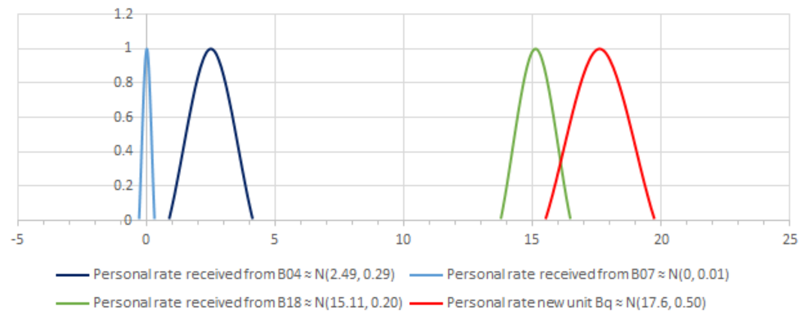

Figure 1: The first proposed share of the merging branches in the personal rate of the new branch.

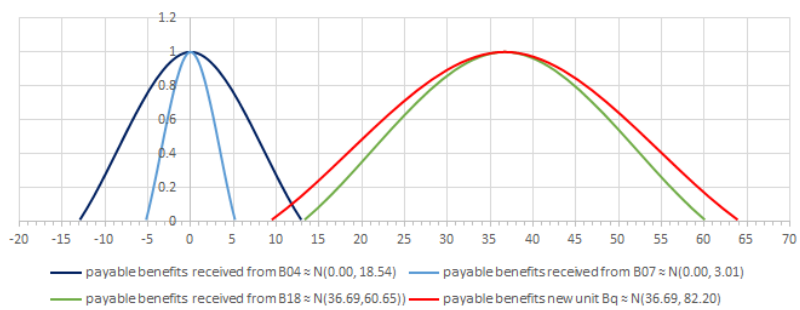

Figure 2: The first proposed share of the merging branches in the payable benefits of the new branch.

Table 9: The second proposed input levels for the merged branch to achieve the efficiency goal 0.85 in the significance level of 0.01.

Inputs	Personal rate $\tilde{\alpha}^2_{1j}$		Payable benefits $\tilde{\alpha}^2_{2j}$		Delayed requisitions $\tilde{\alpha}^2_{3j}$	
DMU	Mean	Variance	Mean	Variance	Mean	Variance
B04	2.00	0.32	28.94	43.52	67.53	29.55
B07	0.00	0.01	4.89	22.18	0.00	7.72
B18	15.11	0.17	0.00	16.50	3.79	12.04
Bq	17.11	0.50	33.83	82.20	71.32	49.31

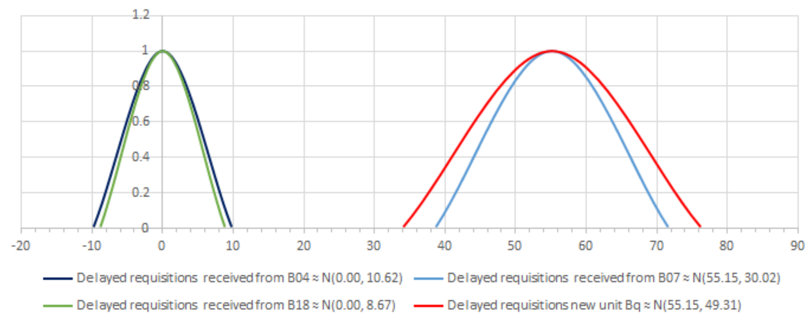

Figure 3: The first proposed share of the merging branches in the delayed requisitions of the new branch.

branch B04 in providing the needed resources related to indexes of the payable benefits and delayed requisitions of the new branch q is higher than other branches. Also, the B07 branch has almost no contribution in providing the resources needed for the new branch.

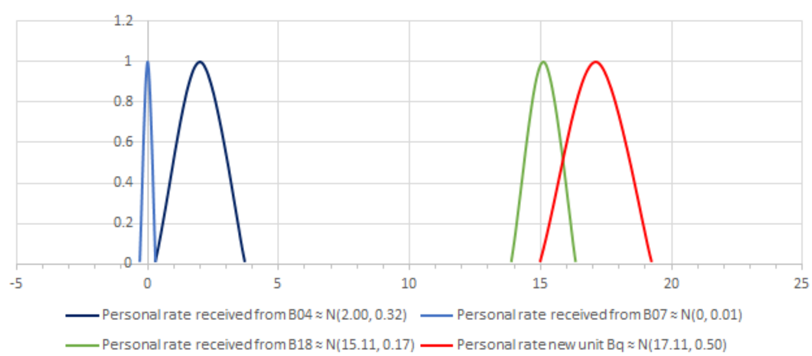

Figure 4: The second proposed share of the merging branches in the personal rate of the new branch.

In fact, Tables 8 and 9, and Figures 1-6 are shown the contribution of three branches B04, B07, and B18 in the personal rate, payable benefits, and delayed requisitions of the new branch q based on two merger scenarios, respectively. In fact, Tables 8 and 9 give the inherited contributions of branches B04, B07, and B18 to generate outputs. In other words,

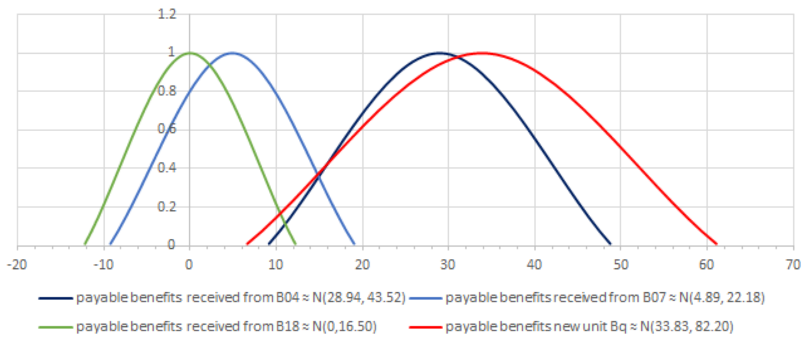

Figure 5: The second proposed share of the merging branches in the payable benefits of the new branch.

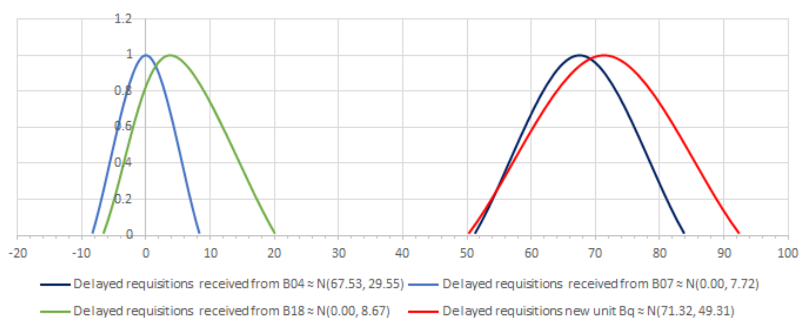

Figure 6: The second proposed share of the merging branches in the delayed requisitions of the new branch.

i) If the first scenario is selected to merge, then the share of two branches B04 and B18 in the personal rate of the new branch q are approximately 14% and 86%, respectively while branch B07 has no role. However, if the two scenario is selected to merge, then the share of two branches B04 and B18 in the personal rate of the new branch q are approximately 12% and 88%, respectively while in this scenario, branch B07 still has no role.

ii) If the first scenario is selected to merge, then the payable benefits are fully supplied by B18 and branches B04 and B07 have no role. While if the second scenario is selected to merge, then the share of two branches B04 and B07 in the payable benefits of the new branch q are approximately 12% and 88%, respectively and branch B18 has no role.

iii) If the second scenario is selected to merge, then the share of two branches B04 and B18 in the delayed requisitions of the new branch q are approximately 95% and 5%, respectively and branch B07 has no role. However, if the first scenario is selected to merge, then the delayed requisitions is fully supplied by B07 and branches B04 and B18 have no role.

In order to investigate the effect of error in estimating the new unit inputs, the proposed model (18) is employed for this integration with the expected efficiency goal 0.85 in the levels of significance $\alpha = 0.01, 0.05, 0.1$, and the results are written in Table 10. According to the proposed approach in this study, it is expected that with the increase of the error, the amount of resources required in each of the input factors the personal rate, payable benefits, and delayed requisitions will decrease or at most remain constant. With regard to Table 10, this expectation is covered. Table 10 shows that the mean of first input ($\tilde{x}_1$) in the levels of significance $\alpha = 0.01$, $\alpha = 0.01$, and $\alpha = 0.1$ is equal to 17.59, 17.43, and 17.34. In fact, by increasing the error level from 0.01 to 0.05, the mean value has decreased to 0.16. Also, by increasing the error level from 0.05 to 0.10, the mean value decreased to 0.09. Table 10 shows a similar reduction under increasing error for the second input as well. However, the third input remains unchanged as the error increases. Moreover, figure 7 shows the discussed topic well. In other words, this figure shows that to obtain the expected performance score (0.85) with decrease the confidence factor, the new branch must receive less resources, which is consistent with the generality of the study.

As another intended performance level, suppose that the merged branch B_q aims to be fully efficient in the level of significance $\alpha = 0.01$, $\bar{\theta}_q(\alpha) = 1$. Consider that the merger of three merging branches B04, B07, and B18 should achieve an optimistic target level, that is equal to one in the level of significance $\alpha = 0.01$. By using Model (18) that correspond with this merger, the following two SP solutions are generated for this model as presented in Tables 11 and 12.

It is worth noting that the InvDEA idea is employed to solving the merging DMUs problem by some researchers, including [1, 2, 22, 29, 62].

Table 10: The mean proposed input levels for the merged branch to achieve the efficiency goal 0.85 in the significance levels of 0.01, 0.05, 0.1.

Inputs	Personal rate $\tilde{x}_{1q}$	Payable benefits $\tilde{x}_{2q}$	Delayed requisitions $\tilde{x}_{3q}$	Pre-defined error level
DMU	Mean	Mean	Mean	α
Bq	17.59	36.69	55.15	0.01
Bq	17.43	33.57	55.15	0.05
Bq	17.34	31.87	55.15	0.10

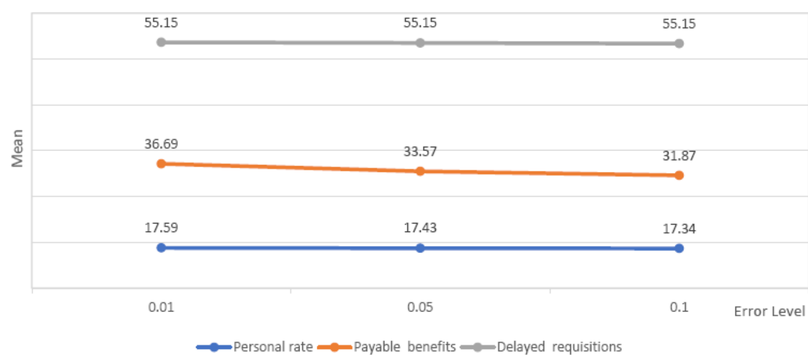

Figure 7: The mean proposed input levels of the merged branch to achieve the efficiency goal 0.85 in the levels of significance 0.01, 0.05, 0.1.

Table 11: The first proposed input levels for the merged branch to achieve fully efficient in the significance levels of 0.01.

Inputs	Personal rate $\tilde{\alpha}^1_{1j}$		Payable benefits $\tilde{\alpha}^1_{2j}$		Delayed requisitions $\tilde{\alpha}^1_{3j}$	
DMU	Mean	Variance	Mean	Variance	Mean	Variance
B04	0.00	0.15	0.00	20.12	0.00	13.91
B07	0.00	0.01	0.00	26.43	49.43	23.56
B18	14.68	0.34	34.25	35.65	0.00	11.84
Bq	14.68	0.50	34.25	82.20	49.93	49.31

However, these proposed solutions fail in the presence of stochastic data. Therefore, there is no available approach based on the InvDEA idea to

Table 12: The second proposed input levels for the merged branch to achieve fully efficient in the significance levels of 0.01.

Inputs	Personal rate $\tilde{\alpha}_{1j}^2$		Payable benefits $\tilde{\alpha}_{2j}^2$		Delayed requisitions $\tilde{\alpha}_{3j}^2$	
DMU	Mean	Variance	Mean	Variance	Mean	Variance
B04	13.77	0.35	28.93	45.73	49.47	39.11
B07	0.93	0.14	5.31	23.70	0.00	7.13
B18	0.00	0.01	0.00	12.67	0.00	3.07
Bq	14.70	0.50	34.24	82.20	49.47	49.31

compare our results with that.

6 Conclusions

This paper proposes new InvDEA models to highlight the potential gains to improving efficiency in M&A by producing some extra product(s) or decreasing some utilized source(s). These InvDEA models are proposed using the SP solutions of SMOP problems to identify the merged unit's inherited input/output levels from merging units to attain a predetermined efficiency score at the significance level α. These models can be employed in applications with more than one input/output dataset for each unit. The results indicate that the merged unit can attain any pre-determined efficiency target at the significance level α if the proposed input-oriented (or output-oriented) stochastic InvDEA models are solved. Also, this paper proposes a stochastic programming model to identify the lowest efficiency goal at the significance level α that the merged entity can achieve. It is critical for managers to know the lowest receivable efficiency score at the significance level α to the new unit for the merging program. Managers can use the results obtained from these models to design useful approaches based on the merging of units to improve their efficiency scores when production frontiers include inefficiency and random error. The proposed models can be transformed into linear programming problems, provided that the input and output levels have a normal distribution with a symmetric error structure. Besides, the validity of the developed theory is demonstrated through a use case of the banking sector. The provided results are important because they can be employed in various applications such as resource allocation and investment analysis.

7 Limitations and Suggestions for Future Research

The proposed models can be transformed into linear programming problems, provided that the input and output levels have a normal distribution with a symmetric error structure. Our study is limited to the merging case where the generated new unit is inside or on the frontier of the current PPS at the significance level α. Clearly, the generated new unit will be inside or on the frontier of the current PPS at the significance level α, if and only if the virtual unit $(\sum_{j\in\Lambda} \tilde{X}_j,\ \sum_{j\in\Lambda} \tilde{Y}_j)$ is within or on frontier of the current PPS at the significance level α. As a result, the subject studied in this article can be worth discussing without considering this limitation as well. As a suggestion, considering the multivariate normal distribution of the data can be a valuable avenue for further research to face these limitations. Also, extending the proposed methods to dynamic and network DEA frameworks can be a valuable route for further research. Moreover, our study is limited to the case where the sample size is large enough ($\kappa \geq 30$), otherwise ($\kappa < 30$) $\bar{x}$ has T-student distribution. Therefore, the discussed topic in this paper without consideration this restriction can be a worthwhile direction for further research. It is worth noting that, in the current work, the problem of merging units is investigated based on the input/output-oriented models. Nevertheless, this issue can be worth studying based on the mix-oriented models as well.

Conflict of interest. The authors declare that they have no conflict of interest.

Acknowledgment. The authors would like to thank the editor of the journal and anonymous reviewers for their insightful and constructive comments and suggestions.

References

[1] G. R. Amin and S. Al-Muharrami. A new inverse data envelopment analysis model for mergers with negative data. *IMA Journal of Management Mathematics*, 00:1–13, 2016.

[2] G. R. Amin, A. Emrouznejad, and S. Gattoufi. Minor and major consolidations in inverse dea: Definition and determination. *Computers Industrial Engineering*, 103:193–200, 2017.

[3] G. R. Amin, A. Emrouznejad, and S. Gattoufi. Modelling generalized firms' restructuring using inverse dea. *Journal of Productivity Analysis*, 48:51–61, 2017.

[4] G. R. Amin and M. Ibn Boamah. A two-stage inverse data envelopment analysis approach for estimating potential merger gains in the us banking sector. *Manage Decis Econ.*, 42:1454– 1465, 2021.

[5] S. Balak, M. H. Behzadi, and A. Nazari. Stochastic copula-dea model based on the dependence structure of stochastic variables: An application to twenty bank branches. *Economic Analysis and Policy*, 72:326–341, 2021.

[6] M. H. Behzadi and M. Mirbolouki. Symmetric error structure in stochastic dea. *Int. J. Industrial Mathematics*, 4(4):335–343, 2012.

[7] M. H. Behzadi, N. Nematollahi, and M. Mirbolouki. Ranking efficient dmus with stochastic data by considering inefficient frontier. *International Journal of Industrial Mathematics*, 1(3):219–226, 2009.

[8] Allen N. Berger and David B. Humphrey. Measurement and efficiency issues in commercial banking. In *Output measurement in the service sectors*, pages 245–300. University of Chicago Press, 1992.

[9] M. S. Campos, M. A. Costa, T. S. Gontijo, and A. L. Lopes-Ahn. Robust stochastic frontier analysis applied to the brazilian electricity distribution benchmarking method. *Decision Analytics Journal*, 3:100051, 2022.

[10] A. Charnes, W. W. Cooper, and E. Rhodes. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2:429–444, 1978.

[11] L. Chen, Y. Wang, F. Lai, and F. Feng. An investment analysis for china's sustainable development based on inverse data envelopment analysis. *Journal of Cleaner Production*, 142(4):1638–1649, 2017.

[12] 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. *Journal of the Operational Research Society*, 53:1347–1356, 2002.

[13] 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. *Journal of Productivity Analysis*, 9(1):53–80, 1998.

[14] W. W. Cooper, Z. Huang, and S. Li. Satisficing dea models under chance constraints. *The Annals of Operations Research*, 66:259–279, 1996.

[15] Z. D. Daryani, Gh. Tohidi, B. Daneshian, Sh. Razavyan, and F. Hosseinzadeh Lotfi. Inverse dea in two-stage systems based on allocative efficiency. *J. Intell. Fuzzy Syst.*, 40:591–603, 2021.

[16] L. Dong Joon. Inverse dea with frontier changes for new target setting. *European Journal of Operational Research*, 254:510–516, 2016.

[17] M. Ehrgott. *Multicriteria optimization*. Springer Berlin, 2005.

[18] A. Emrouznejad and E. Thanassoulis. A mathematical model for dynamic efficiency using data envelopment analysis. *Applied Mathematics and Computation*, 160:363–378, 2005.

[19] 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. *Journal of the Operational Research Society*, 0(0):1–12, 2018.

[20] M. Esteve, J. Aparicio, J. J. Rodriguez-Sala, and J. Zhu. Random forests and the measurement of super-efficiency in the context of free disposal hull. *European Journal of Operational Research*, 304(2):729–744, 2023.

[21] M. Eyni, G. Tohidi, and S. Mehrabeian. Applying inverse dea and cone constraint to sensitivity analysis of dmus with undesirable inputs and outputs. *Journal of the Operational Research Society*, 68(1):34–40, 2017.

[22] S. Gattoufi, G. R. Amin, and A. Emrouznejad. A new inverse dea method for merging banks. *IMA Journal of Management Mathematics*, 25:73–87, 2014.

[23] M. Ghiyasi. Inverse dea based on cost and revenue efficiency. *Computers & Industrial Engineering*, 114:258–263, 2017.

[24] M. Ghiyasi. Efficiency improvement and resource estimation:a tradeoff analysis. *International Journal of Productivity and Quality Management*, 25(2):151–169, 2018.

[25] M. Ghiyasi. Novel criterion models in the inverse dea problem. *Int. J. Operational Research*, 35(1):20–36, 2019.

[26] S. Ghobadi. Inverse dea using enhanced russell measure in the presence of fuzzy data. *Int. J. Industrial Mathematics*, 10(2):1–16, 2018.

[27] S. Ghobadi. A generalized dea model for inputs (outputs) estimation under inter-temporal dependence. *RAIRO-Operations Research*, 53(5):1791 –1805, 2019.

[28] S. Ghobadi. A dynamic dea model for resource allocation. *Int. J. of Mathematics in Operational Research*, 17(1):50–77, 2020.

[29] S. Ghobadi. Merging decision-making units with interval data. *RAIRO-Operations Research*, 55:1605–1630, 2021.

[30] S. Ghobadi and S. Jahangiri. Optimal allocation of resources using the ideal-solutions. *Journal of New Researches in Mathematics*, 5(20):121–134, 2019.

[31] S. Ghobadi, G. R. Jahanshahloo, F. Hoseinzadeh Lotfi, and M. Rostami-malkhalifeh. Dynamic inverse dea in the presence of fuzzy data. *Advances in Environmental Biology*, 8(24):139–151, 2014.

[32] A. Ghomi, S. Ghobadi, M. H. Behzadi, and M. Rostamy-Malkhalifeh. Inverse data envelopment analysis with stochastic data. *RAIRO-Operations Research*, 55(5):2739 – 2762, 2021.

[33] F. Guijarro, M. Martinez-Gomez, and D. Visbal-Cadavid. A model for sector restructuring through genetic algorithm and inverse dea. *Expert Syst. Appl.*, 154:113422, 2020.

[34] A. Hadi-Vencheh, A. A. Foroughi, and M. Soleimani-Damaneh. A dea model for resource allocation. *Economic Modelling*, 25(5):983–993, 2008.

[35] A. Hadi-vencheh, A. Hatami-marbini, Z. Ghelej Beigi, and Gholami K. An inverse optimization model for imprecise data envelopment analysis. *Opimization*, 64(11):2441–2452, 2015.

[36] F. Hosseinzadeh, L., N. Nematollahi, M. H. Behzadi, M. Mirbolouki, and Z. Moghaddas. Centralized resource allocation with stochastic data. *Journal of Computations and Applied Mathematics*, 236:1783–1788, 2012.

[37] X. Hu, J. Li, X. Li, and J. Cui. A revised inverse data envelopment analysis model based on radial models. *Mathematics*, 8(5):1–17, 2020.

[38] Z. Huang and SX. Li. Stochastic dea models with different types of input-output disturbances. *Journal of Productivity Analysis*, 15:95–113, 2001.

[39] M. Izadikhah. A fuzzy stochastic slacks-based data envelopment analysis model with application to healthcare efficiency. *Healthcare Analytics*, 2:100038, 2022.

[40] G. R. Jahanshahloo, F. H. Lotfi, N. Shoja, G. Tohidi, and S. Razavyan. Sensitivity of efficiency classifications in the inverse dea models. *Applied Mathematics and Computation*, 169(2):905–916, 2005.

[41] G. R. Jahanshahloo, M. H. MH. Behzadi, and M. Mirbolouki. Ranking stochastic efficient dmus based on reliability. *International Journal of Industrial Mathematics*, 2:263–270, 2010.

[42] G. R. Jahanshahloo, M. Soleimani-damaneh, and S. Ghobadi. Inverse dea under inter-temporal dependence using multiple-objective programming. *European Journal of Operational Research*, 240:447–456, 2015.

[43] M. Kalantary and R. Farzipoor Saen. Assessing sustainability of supply chains: An inverse network dynamic dea model. *Computers & Industrial Engineering*, 135:1224–1238, 2019.

[44] M. Khodabakhshi. Estimating most productive scale size with stochastic data in data envelopment analysis. *Economic Modelling*, 26:968–973, 2009.

[45] M. Koengkan, J. A. Fuinhas, E. Kazemzadeh, F. Osmani, N. K. Alavijeh, A. Auza, and M. Teixeira. Measuring the economic efficiency performance in latin american and caribbean countries: An empirical evidence from stochastic production frontier and data envelopment analysis. *International Economics*, 169:43–54, 2022.

[46] T. Kuosmanen and M. Kortelainen. Stochastic non-smooth envelopment of data: Semi-parametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis*, 38(1):11–28, 2012.

[47] S. Lertworasirikul, P. Charnsethikul, and S. C. Fang. Inverse data envelopment analysis model to preserve relative efficiency values: The case of variable returns to scale. *Computers and Industrial Engineering*, 61(4):1017–1023, 2011.

[48] SX. Li. Stochastic models and variable returns to scales in data envelopment analysis. *European Journal of Operational Research*, 104:532–548, 1998.

[49] H. T. Lin. An efficiency-driven approach for setting revenue target. *Decision Support Systems*, 49:311–317, 2010.

[50] F. Hosseinzadeh Lotfi, N. Nematollahi, M.H. Behzadi, and M. Mirbolouki. Ranking decision making units with stochastic data by using coefficient of variation. *Mathematical and Computational Applications*, 15(1):148–155, 2010.

[51] M. Z. Mahmoudabadi and A. Emrouznejad. Comprehensive performance evaluation of banking branches: A three-stage slacks-based measure (sbm) data envelopment analysis. *International Review of Economics & Finance*, 64:359–376, 2019.

[52] O. B. Olesen and N. C. Petersen. Stochastic data envelopment analysisa review. *European Journal of Operational Research*, 251(1):2–21, 2016.

[53] J. T. Pastor, J. L. Ruiz, and I. Sirvent. An enhanced dea russell graph efficiency measure. *European Journal of Operational Research*, 115:596–607, 1999.

[54] J. K. Sengupta. Efficiency analysis by stochastic data envelopment analysis. *Applied Economics Letters*, 7:379–383, 2002.

[55] K. G. Shinto and C. M. Sushama. An inverse dea model for input/output estimation with integer restriction. *Journal of Mathematics and Statistical Science*, 6(6):187–194, 2020.

[56] Kh. Soleimani-Chamkhorami, F. H. Lotfi, G. R. Jahanshahloo, and M. Rostamy-Malkhalifeh. Preserving cost and revenue efficiency through inverse data envelopment analysis models. *INFOR: Information Systems and Operational Research*, 58(4):561–578, 2020.

[57] Kh. Soleimani-Chamkhorami, F. H. Lotfi, G. R. Jahanshahloo, and M. Rostamy-Malkhalifeh. A ranking system based on inverse data envelopment analysis. *IMA Journal of Management Mathematics*, 31(3):367–385, 2020.

[58] M. G. Tsionas. Combining data envelopment analysis and stochastic frontiers via a lasso prior. *European Journal of Operational Research*, 304(3):1158–1166, 2023.

[59] M. G. Tsionas. Performance estimation when the distribution of inefficiency is unknown. *European Journal of Operational Research*, 304(3):1212–1222, 2023.

[60] M. Wegener and G. R. Amin. Minimizing greenhouse gas emissions using inverse dea with an application in oil and gas. *Expert Systems with Applications*, 122(5):369–375, 2019.

[61] S. Xiao, Y. Li, A. Emrouznejad, J. Xie, and L. Liang. Estimation of potential gains from bank mergers: A novel two-stage cost efficiency dea model. *Journal of the Operational Research Society*, 68(9):1045–1055, 2017.

[62] E. Zenodin and S. Ghobadi. Merging decision-making units under inter-temporal dependence. *IMA Journal of Management Mathematics*, 31(2):139–16, 2020.

[63] X. Zhang and J. Cui. a project evaluation system in the state economic information system of china: An operation research practice in public sectore. *International Transactions in Operational*, 6:441–452, 1999.

Ali Ghomi
Phd Student of Statistics
Department of Statistics
Science and Research Branch, Islamic Azad University
Tehran, Iran
E-mail: Alighomi@pco.iaun.ac.ir

Saeid Ghobadi
Assistant Professor of Applied Mathematics
Department of Mathematics
Khomeinishahr Branch, Islamic Azad University
Isfahan, Iran
E-mail: Ghobadi@iaukhsh.ac.ir

Mohammad Hassan Behzadi
Associate Professor of Statistics
Department of Statistics
Science and Research Branch, Islamic Azad University

Tehran, Iran
E-mail: Behzadi@srbiau.ac.ir

Mohsen Rostamy-Malkhalifeh
Associate Professor of Applied Mathematics
Department of Mathematics
Science and Research Branch, Islamic Azad University
Tehran, Iran
E-mail: $Mohsen\_rostamy$@yahoo.com

Khosro Soleimani-Chamkhorami
Assistant Professor of Applied Mathematics
Department of Mathematics, Faculty of Computer Engineering
Najafabad Branch, Islamic Azad University
Najafabad, Iran
E-mail: Soleimani@iaun.ac.ir

Appendix A

Proof of Theorem 4.3. To prove the theorem, we should show that $\theta_q^*(\alpha) = \bar{\theta}_q(\alpha)$, considering $\tilde{x}_{iq} = \sum_{j\in\Lambda} \tilde{\alpha}_{ij}^*$ $(i = 1, 2, ..., m)$. Since $\tilde{x}_{iq} = \sum_{j\in\Lambda} \tilde{\alpha}_{ij}^*$ and Δ is a feasible solution for model (18), we have

$$P\{\sum_{j\in\Pi} \lambda_j^* \tilde{x}_{ij} \leq \ \bar{\theta}_q(\alpha)\tilde{x}_{iq}\} \geq 1-\alpha, \qquad i = 1, 2, ..., m, \tag{27}$$

$$P\{\sum_{j\in\Pi} \lambda_j^* \tilde{y}_{rj} \geq \ \tilde{y}_{rq}\} \geq 1-\alpha, \qquad r = 1, 2, ..., s, \tag{28}$$

$$P\{\tilde{\alpha}_{ij}^* \leq \tilde{x}_{ij}\} \geq 1-\alpha, \qquad i = 1, 2, ..., m, \ \forall j \in \Lambda, \tag{29}$$

$$P\{\tilde{\alpha}_{ij}^* \geq 0\} \geq 1-\alpha, \qquad i = 1, 2, ..., m, \ \forall j \in \Lambda, \tag{30}$$

$$\lambda^* = (\lambda_j^*; \ \forall j \in \Pi) \in \overline{\Omega}_q. \tag{31}$$

According to Eqs. (27), (28), and (31), it is obvious that $(\bar{\lambda} = (\lambda_j^*, j \in \Pi; \mu_q = 0), \ \theta = \bar{\theta}_q(\alpha))$ is a feasible solution to problem (16). Therefore, the optimal value of model (16) is less than or equal to $\bar{\theta}_q(\alpha)$. In other words, $\theta_q^*(\alpha) \leq \bar{\theta}_q(\alpha)$.

By contradiction assume that $\Phi = (\lambda_j^{**}, j \in \Pi; \lambda_q^{**}, \theta^{**})$ is an optimal solution to model (16) in which $\theta_q^*(\alpha) = \theta^{**} < \bar{\theta}_q(\alpha) \leq 1$. Therefore, feasibility of Φ for Model (16), implies

$$P\{\sum_{j\in\Pi} \lambda_j^{**}\tilde{x}_{ij} + \lambda_q^{**}\tilde{x}_{iq} \leq \theta^{**}\tilde{x}_{iq}\} \geq 1-\alpha, \qquad i = 1, 2, ..., m, \tag{32}$$

$$P\{\sum_{j\in\Pi} \lambda_j^{**}\tilde{y}_{rj} + \lambda_q^{**}\tilde{y}_{rq} \geq \tilde{y}_{rq}\} \geq 1-\alpha, \qquad r = 1, 2, ..., s, \tag{33}$$

$$\lambda^{**} = (\lambda_j^{**};\ j \in \Pi \cup \{q\}) \in \Omega_q.$$

Since $0 < \bar{\theta}_q(\alpha) \leq 1$, by Eq. (27) we get

$$P\{\sum_{j\in\Pi} \lambda_j^{*}\tilde{x}_{ij} \leq \ \tilde{x}_{iq}\} \geq 1-\alpha, \qquad i = 1, 2, ..., m. \tag{34}$$

By using inequalities (32) and (34), the following result is obtained:

$$P\{\sum_{j\in\Pi} \lambda_j^{**}\tilde{x}_{ij} + \lambda_q^{**}(\sum_{j\in\Pi} \lambda_j^{*}\tilde{x}_{ij})\tilde{x}_{iq} \leq \theta^{**}\tilde{x}_{iq}\} \geq 1-\alpha,\ i = 1, 2, ..., m,$$

then,

$$P\{\sum_{j\in\Pi}(\lambda_j^{**} + \lambda_q^{**}\lambda_j^{*})\tilde{x}_{ij} \leq \theta^{**}\tilde{x}_{iq}\} \geq 1-\alpha, \qquad i = 1, 2, ..., m, \} \geq 1-\alpha.$$

We have,

$$P\{\sum_{j\in\Pi} \overline{\lambda}_j\tilde{x}_{ij} - \theta^{**}\tilde{x}_{iq} \leq 0\} \geq 1-\alpha, \qquad i = 1, 2, ..., m, \tag{35}$$

where

$$\overline{\lambda}_j = \lambda_j^{**} + \lambda_q^{**}\lambda_j^{*}, \qquad \forall j \in \Pi. \tag{36}$$

Similarly, considering Eqs. (28), (33), and (36), we get

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{y}_{rj}-\tilde{y}_{rq}\geq 0\}\geq 1-\alpha, \qquad r=1,2,...,s. \tag{37}$$

By Eq. (31) and feasibility of Φ to model (16), we have

$$\bar{\lambda}=(\bar{\lambda}_j;\ j\in\Pi)\in\overline{\Omega}_q. \tag{38}$$

According to Eq. (30) and the definition of $\tilde{x}_{iq}$ $(i=1,2,...,m,)$, there exists at least one $p\in\{1,2,...,m,\}$ and at least one $k\in\Lambda$ such that $P\{\tilde{\alpha}^*_{pk}>0\}\geq 1-\alpha$. There exists non-negative scalar $\varepsilon\geq 0$ and positive scaler $\epsilon>0$ such that $P\{\tilde{\alpha}^*_{pk}\geq\epsilon\}=1-\alpha+\varepsilon$. Then, there exists positive scalar $\kappa>0$ such that

$$P\{\tilde{\alpha}^*_{pk}\geq\kappa\}=1-\alpha. \tag{39}$$

By contradiction assume ($\theta^{**}<\bar{\theta}_q(\alpha)$) and Eq. (35), we obtain

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{x}_{ij}-\bar{\theta}_q(\alpha)\tilde{x}_{iq}<0\}\geq 1-\alpha, \qquad i=1,2,...,m. \tag{40}$$

Considering p^{th}-inequality, there exists non-negative scaler ξ_p, such that

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{x}_{pj}-\bar{\theta}_q(\alpha)\tilde{x}_{pq}<0\}=1-\alpha+\xi_p.$$

Therefore, there exists positive scaler s_p, such that

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{x}_{pj}-\bar{\theta}_q(\alpha)\tilde{x}_{pq}\leq -s_p\}=1-\alpha. \tag{41}$$

Clearly, there exists positive scaler $\overline{s}_p$, such that $s_p=\bar{\theta}_q(\alpha)\overline{s}_p$. Then, by Eq. (41), we get

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{x}_{pj}-\bar{\theta}_q(\alpha)\tilde{x}_{pq}\leq -\bar{\theta}_q(\alpha)\overline{s}_p\}=1-\alpha.$$

Therefore,

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{x}_{pj}-\bar{\theta}_q(\alpha)(\tilde{x}_{pq}-\overline{s}_p)\leq 0\}=1-\alpha. \tag{42}$$

Defining

$$\overline{k}=\min\{\kappa,\ \overline{s}_p\}, \tag{43}$$

we have $\overline{k}>0$. Now, for each $i=1,2,...,m$, define

$$\bar{\tilde{\alpha}}_{ij}:=\begin{cases}\tilde{\alpha}^*_{ij}-\overline{k} & if \quad i=p\ \&\ j=k,\\ \\ \tilde{\alpha}^*_{ij} & \quad otherwise.\end{cases} \tag{44}$$

By (39) and (42)-(44) we have

$$P\{\sum_{j\in J}\overline{\lambda}_j\tilde{x}_{pj}-\bar{\theta}_q(\alpha)\sum_{j\in\Lambda}\bar{\tilde{\alpha}}_{pj}\leq 0\}\geq 1-\alpha, \tag{45}$$

$$P\{\bar{\tilde{\alpha}}_{pk}=\tilde{\alpha}^*_{pk}-\overline{k}\geq 0\}\geq 1-\alpha. \tag{46}$$

In addition, by Eqs. (29), (30), (44) and (46), we get

$$P\{\bar{\tilde{\alpha}}_{ij}\leq\tilde{x}_{ij}\}\geq 1-\alpha,\quad \forall i\in\{1,2,...,m\}-\{p\},\ \forall j\in\Lambda, \tag{47}$$

$$P\{\bar{\tilde{\alpha}}_{ij}=\tilde{\alpha}^*_{ij}\geq 0\}\geq 1-\alpha,\quad \forall i\in\{1,2,...,m\}-\{p\},\ \forall j\in\Lambda-\{k\}. \tag{48}$$

Moreover, by Eq. (40), we have

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{x}_{ij}-\bar{\theta}_q(\alpha)\tilde{x}_{iq}\leq 0\}\geq 1-\alpha,\quad \forall i\in\{1,2,...,m\}-\{p\}. \tag{49}$$

According to (36), (37), (38), (45-49), $(\bar{\lambda},\bar{\tilde{\alpha}}_{ij}: i=1,2,...,m,\ \forall j\in\Lambda)$ is a feasible solution to problem (18), such that $P\{\bar{\tilde{\alpha}}_{ij}-\tilde{\alpha}^*_{ij}\leq 0\}\geq 1-\alpha$, $\forall i\in\{1,2,...,m\}-\{p\}$ and $j\in\Lambda$, and $P\{\bar{\tilde{\alpha}}_{pk}-\tilde{\alpha}^*_{pk}\leq-\overline{k}\}\geq 1-\alpha$. But this is impossible because Δ is a SP solution in the level of significance α to SMOP (18).□

Appendix B

Proof of Theorem 4.6. To prove the theorem, we should show that $\varphi_q^*(\alpha) = \bar{\varphi}_q(\alpha)$, considering $\tilde{y}_{rq} = \sum_{j\in\Lambda} \tilde{\beta}_{rj}^*$ $(r = 1, 2, ..., s)$. Since $\tilde{y}_{rq} = \sum_{j\in\Lambda} \tilde{\beta}_{rj}^*$ and Δ is a feasible solution for model (23), we have

$$P\{\sum_{j\in\Pi} \lambda_j^* \tilde{x}_{ij} \leq \tilde{x}_{iq}\} \geq 1-\alpha, \quad i = 1, 2, ..., m, \tag{50}$$

$$P\{\sum_{j\in\Pi} \lambda_j^* \tilde{y}_{rj} \geq \bar{\varphi}_q(\alpha)\tilde{y}_{rq}\} \geq 1-\alpha, \quad r = 1, 2, ..., s, \tag{51}$$

$$P\{\tilde{\beta}_{rj}^* \geq \tilde{y}_{rj}\} \geq 1-\alpha, \quad r = 1, 2, ..., s, \ \forall j \in \Lambda, \tag{52}$$

$$\lambda = (\lambda_j;\ \forall j \in \Pi) \in \overline{\Omega}_q. \tag{53}$$

According to Eqs. (50), (51), and (53), it is obvious that $(\bar{\lambda} = (\lambda_j^*,\ \forall j \in \Pi; \lambda_q = 0),\ \varphi = \bar{\varphi}_q(\alpha))$ is a feasible solution to problem (21). Therefore, the optimal value of model (21) is greater than or equal to $\bar{\varphi}_q(\alpha)$. In other words, $\varphi_q^*(\alpha) \geq \bar{\varphi}_q(\alpha)$.

By contradiction assume that $\Phi = (\lambda_j^{**}, j \in \Pi; \lambda_q^{**}, \varphi^{**})$ is an optimal solution to model (21) in which $\varphi_q^*(\alpha) = \varphi^{**} > \bar{\varphi}_q(\alpha) \geq 1$. Therefore, feasibility of Φ for Model (21), implies

$$P\{\sum_{j\in\Pi} \lambda_j^{**} \tilde{x}_{ij} + \lambda_q^{**}\tilde{x}_{iq} \leq \tilde{x}_{iq}\} \geq 1-\alpha, \quad i = 1, 2, ..., m, \tag{54}$$

$$P\{\sum_{j\in\Pi} \lambda_j^{**} \tilde{y}_{rj} + \lambda_q^{**}\tilde{y}_{rq} \geq \varphi^{**}\tilde{y}_{rq}\} \geq 1-\alpha, \quad r = 1, 2, ..., s, \tag{55}$$

$$\lambda^{**} = (\lambda_j^{**};\ \forall j \in \Pi) \in \overline{\Omega}_q.$$

Since $\bar{\varphi}_q(\alpha) \geq 1$, by Eq. (51) we get

$$P\{\sum_{j\in\Pi} \lambda_j^* \tilde{y}_{rj} \geq \tilde{y}_{rq}\} \geq 1-\alpha, \quad r = 1, 2, ..., s. \tag{56}$$

By using inequalities (55) and (56), the following result is obtained:

$$P\{\sum_{j\in\Pi}\lambda_j^{**}\tilde{y}_{rj}+\lambda_q^{**}(\sum_{j\in\Pi}\lambda_j^{*}\tilde{y}_{rj})\geq\varphi^{**}\tilde{y}_{rq}\}\geq 1-\alpha,\quad r=1,2,...,s,$$

then,

$$P\{\sum_{j\in\Pi}(\lambda_j^{**}+\lambda_q^{**}\lambda_j^{*})\tilde{y}_{rj}\geq\varphi^{**}\tilde{y}_{rq}\}\geq 1-\alpha,\qquad r=1,2,...,s,\}\geq 1-\alpha.$$

We have,

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{y}_{rj}-\varphi^{**}\tilde{y}_{rq}\geq 0\}\geq 1-\alpha,\quad r=1,2,...,s, \tag{57}$$

where

$$\overline{\lambda}_j=\lambda_j^{**}+\lambda_q^{**}\lambda_j^{*},\qquad\forall j\in\Pi. \tag{58}$$

Similarly, considering Eqs. (50), (54), and (58), we get

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{x}_{ij}-\tilde{x}_{iq}\leq 0\}\geq 1-\alpha,\quad i=1,2,...,m. \tag{59}$$

By Eq. (53) and feasibility of Φ to model (21), we have

$$\overline{\lambda}=(\overline{\lambda}_j;\ \forall j\in\Pi). \tag{60}$$

By contradiction assume that $\varphi_q^*(\alpha)=\varphi^{**}>\bar{\varphi}_q(\alpha)$. By (57), we get

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{y}_{rj}-\bar{\varphi}_q(\alpha)\tilde{y}_{rq}>0\}\geq 1-\alpha,\quad r=1,2,...,s. \tag{61}$$

Without loss of generality, we consider p^{th}-inequality. There exists non-negative scaler ξ_p, such that

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{y}_{pj}-\bar{\varphi}_q(\alpha)\tilde{y}_{pq}>0\}=1-\alpha+\xi_p.$$

Then, there exists positive scaler ε_p, in which

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{y}_{pj}-\bar{\varphi}_q(\alpha)\tilde{y}_{pq}\geq\varepsilon_p\}=1-\alpha. \tag{62}$$

Therefore, there is $\overline{s}_p>0$ such that $\varepsilon_p=\bar{\varphi}_q(\alpha)\overline{s}_p$. By Eq. (62), we get

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{y}_{pj}-\bar{\varphi}_q(\alpha)(\tilde{y}_{pq}+\overline{s}_p)\geq 0\}=1-\alpha. \tag{63}$$

With regard to definition of $\tilde{y}_{pq}=\sum_{j\in\Lambda}\tilde{\beta}^*_{pj}$, without loss of generality, for each $r(r=1,2,...,s)$ and $j\in\Lambda$, we define

$$\overline{\overline{\tilde{\beta}}}_{rj}:=\begin{cases}\tilde{\beta}^*_{rj}+\overline{s}_p & if \quad r=p \ \&\ j=k,\\ \\ \tilde{\beta}^*_{rj} & otherwise.\end{cases} \tag{64}$$

Now, we consider $\overline{\overline{\tilde{y}}}_{rq}:=\sum_{j\in\Lambda}\overline{\overline{\tilde{\beta}}}_{rj}$ for each $r=1,2,...,s$. By (52), (59), (60), (61), (63), and (64), we obtain

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{x}_{ij}-\tilde{x}_{iq}\leq 0\}\geq 1-\alpha,\quad i=1,2,...,m. \tag{65}$$

$$P\{\sum_{j\in\Pi}\overline{\lambda}_j\tilde{y}_{rj}-\bar{\varphi}_q(\alpha)\overline{\overline{\tilde{y}}}_{rq}\geq 0\}\geq 1-\alpha,\quad r=1,2,...,s. \tag{66}$$

$$P\{\overline{\overline{\tilde{\beta}}}_{rj}\geq\tilde{y}_{rj}\}\geq 1-\alpha,\quad r=1,2,...,s,\ \forall j\in\Lambda, \tag{67}$$

$$\sum_{j\in\Pi}\overline{\lambda}\in(\overline{\lambda}_j:\ \forall j\in\Pi). \tag{68}$$

According to (65-68), $(\overline{\lambda},\overline{\overline{\tilde{\beta}}}_{rj}:r=1,2,...,s,\ \forall j\in\Lambda)$ is a feasible solution to problem (23), such that

$$P\{\tilde{\beta}^*_{rj}-\overline{\overline{\tilde{\beta}}}_{rj}\leq 0\}\geq 1-\alpha,\quad\forall r\in\{1,2,...,s\}-\{p\},\ \forall j\in\Lambda-\{k\},$$

$$P\{\tilde{\beta}^*_{pk}-\overline{\overline{\tilde{\beta}}}_{pk}\leq -\overline{s}_p\}\geq 1-\alpha.$$

But this is impossible because Δ is a SP solution in the level of significance α to SMOP (23).□

Appendix C

Table 13: The inputs of 20 bank branches.

	Input1		Input2		Input3	
DMU	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 14: The outputs of 20 bank branches.

	output1		output2		output3		output4		output5	
DMU	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance	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