Journal of Mathematical Extension
Vol. 17, No. 12, (2023) (3)1-44
URL: https://doi.org/10.30495/JME.2023.2724
ISSN: 1735-8299
Original Research Paper

Investigating Models of Stochastic Data Envelopment Analysis

S. Farzam
Shahid Rajaee Teacher Training University

F. Hosseinzadeh Lotfi*
Science and Research Branch, Islamic Azad University

H. Mesgarani
Shahid Rajaee Teacher Training University

Abstract. Using the right tools to support decision-making for managers is very important. Today, using mathematical methods to support managers in making decisions is prevalent. Data envelopment analysis is one of the mathematical methods used in the mentioned field. In recent years, researchers in this field have developed data envelopment analysis methods to a great extent. Many data organizations use to evaluate performance and other purposes are imprecise and uncertain. To solve this problem, one should use methods whose results are reliable. Stochastic, fuzzy-stochastic, fuzzy, robust, etc. models can be mentioned among these methods. Among non-deterministic methods, stochastic models are of particular importance. This article only reviews the existing stochastic methods for dealing with non-deterministic data. We have divided stochastic methods into four categories to understand the subject better. Also, the fit between the models and the data has been reviewed in the articles. The process mining method is suggested to determine the appropriate model for the available data. Finally, the weaknesses of some previous models have been introduced. Suggestions

Received: April 2023; Accepted: October 2023
*Corresponding author

have been made to fix these weaknesses. The current research results provide researchers in this field with appropriate information regarding stochastic data envelopment analysis.

AMS Subject Classification: MSC code1; MSC code 2; more
Keywords and Phrases: Data envelopment analysis, Network DEA, Stochastic DEA

1 Introduction

Data envelopment analysis is a non-parametric method used to evaluate organizational units' efficiency and ranking. In recent years, the use of stochastic modeling for real-world problems has become prevalent. To achieve more realistic results, researchers try to examine issues under uncertainty. For this reason, non-deterministic models have recently been widely used in new data envelopment analysis models.In this article, only stochastic models have been investigated among non-deterministic models. Banker (1986, 1993)[5][6] and Sengupta (1982, 1987)[43][44] first integrated statistical elements into DEA. Charnes and Cooper (1961) investigated and presented deterministic equations for stochastic models in the framework of chance constraints.The existence of "deterministic equations" in the form of convex programming problems for a general class of linear decision rules is divided into three categories: (1) maximum expected value ("E model"). (2) Minimum variance ("model V"); and (3) maximum likelihood (T model). Finally, various explanations and interpretations of these results and other aspects of constrained stochastic programming are presented. Sengupta (1987) [43] used a data envelopment analysis model in the context of chance constraint programming. In the research, each DMU has one output and m inputs, and the optimal weight vector p form inputs is determined from the DEA method, based on which each reference unit can be compared in terms of efficiency. Also, the exponential probability density function is used in the final linear model. Land et al. (1993)[33], in order to provide a numerical illustration of the DEA chance constraints, first reviewed the DEA application of the "Program Tracking" study by Charnes et al. (1981). The proposed model assumes that the standard probability distribution of all outputs is reasonable and uses a geometric mean instead of an arithmetic mean.Cooper et al. (1995)[20]

used a DEA model and stochastic frontier analysis for their assessments. Also introduced a new performance measure as a set of concepts to advance the DEA. Cooper (1996) [21], examined Simon's satisfactory models and developed Simon's satisfactory method based on CCP models (Chance Constrained Programming) and DEA concepts. In some modeling, the inputs and outputs are stochastic; in others, only the outputs are stochastic.Sueyoshi (2000)[49] proposed a stochastic data envelopment analysis model for the restructuring strategy planning of a Japanese oil company. In this model, CCP (Chance Constraint Programming) and PERT/CPM (Program Evaluation and Review Technique/Critical Path Method) are used. Huang and Li (2001). [27] obtained deterministic equations for both situations of multivariate symmetric random disturbances and one random factor in production relations. They believe that using the single-factor assumption of the random variable has at least two advantages: (1) the number of parameters is significantly reduced compared to general stochastic models. (b) The dual form for the definite equivalent to and is explicitly formulated. Post (2001) [42] proposes a mean-variance framework to control the uncertainty of inputs, outputs, and risk aversion in DEA. Chen (2002)used chance-constrained data envelopment analysis (CCDEA) and stochastic border analysis (SFA) to measure the technical performance of 39 banks in Taiwan.Cooper et al.(2002). [23] introduced conventional DEA models with a stochastic counterpart in a series of CCP(chance-constrained programming) models. They believe that these models can predict changes in efficient and inefficient behaviors. In fact, instead of using data generated in the past, these models help prevent dysfunctional behaviors by predicting and generating data values for the future. Tsionas (2003)[53] used a combination of stochastic frontier models and data envelopment analysis methods to measure the efficiency of US airlines. The proposed model uses the Bayesian theorem, and Monte Carlo methods have been developed to perform Bayesian experimental inference in the new model. Gibbs sampling is also organized to reinforce the data. Cooper et al. (2004) [22] described Congestion treatment models in data envelopment analysis. They developed CCP-based models to solve this problem. Using statistically correct hypotheses, replaced the definite models with random models. Olsen (2006) [37] compared two different models (LLT) and (OP) and developed a new hybrid model. The new model combines the attractive features of each of the two models. This integrated model can show the effect of the correlation between DMUs and the correlation between inputs and outputs. In the global business environment, which is full of threats from natural, political, economic, and technical sources, enterprise risk management (ERM) is essential. ERM has been developed in supply chain management by showing how to use risk modeling. Choosing the right vendor is one of the most critical decisions in supply chain management. Internal errors in organizations include human error, fraud, system failure, production disruption, and other errors. In risk management, they often look for methods to control and identify risk caused by internal and external factors, and systems are designed for this. These systems sometimes produce incorrect data for various reasons (Schaefer and Cassidy, (2006) [45]).In a paper, Simar (2007) [47] first proposed a way to improve the performance of DEA/FDH estimators for frontiers in the presence of noise. The research used DEA/FDH methods for random frontier analysis in the presence of statistical noises. Margari et al. (2007) [34] investigated the impact of regulatory and environmental factors and statistical noise on the efficiency of Italian public transportation systems. Indeed, the factors mentioned above significantly affect the data produced by the designed systems and cause ambiguity. They used a combined DEA-SFA approach to decompose the desired criteria and determine the efficiency frontier for the given data.Bruni et al. (2009)[14] propose a stochastic data envelopment analysis model based on the theory of chance-constrained programming, which can be used with general multivariate distribution functions. The main assumption is that a discrete approximation of a general non-normal multivariate continuous distribution is available in the form of scenarios, which may be very specific in scope. Tsionas and Papadakis (2009)[52] used a Bayesian approach to data envelopment analysis to evaluate the performance of Greek banks. In fact, the Bayesian approach has been used for statistical inference in the random DEA. In a paper using a statistical simulation, Kao and Tai Liu (2009)[31] examined how to obtain the efficiency distribution of each DMU. They claim that the statistical simulations performed better than the SDEA models. The statistical simulation will perform excellently if the data distribution function is suitable. Wu and Olson (2010). [56] believe that the combination of stochastic data coverage analysis techniques and value at risk (VaR) is a suitable tool to control the risk caused by internal factors of the organization. Khodabakhshi (2010)[32] used CCP approaches to develop an output-oriented super-efficiency model in SDEA(stochastic data envelopment analysis). Wu et al. (2010)[57] examined the issue of pricing of residential complexes in conditions of uncertainty. This study's data envelopment analysis model considers random variables and ordinal data simultaneously. Simar and Zelenyuk (2011). [46] developed the Simar model (J Product Ananl 28:183–201, 2007) to introduce noise in non-parametric frontier models. The new approach models multi-input-multi-output technologies without imposing parametric assumptions on the production relationship. The proposed model uses non-parametric DEA and Free Disposal Hull (FDH) methods to analyze the random boundary, including outlier noises.One of the major problems in many organizations is allocating part of their tasks to individual units, which is called decentralization. These allocations will have appropriate results if all the individuals perform efficiently. Centralized resource allocation (CRA) is a method in which all efficient DMUs are predicted by solving just one DEA model. CRA models have been introduced in the DEA framework to allocate resources to thematic units optimally.Hosseinzadeh Lotfi et al. (2012)[25] provided a random DEA framework for CRA. Assaf (2012)[4] presented an innovative method based on the combination of random frontier and data envelopment analysis in the Bayesian framework. Wu et al. (2013)[55] examined and evaluated environmental performance in China. The research presented a new model of data envelopment analysis to deal simultaneously with uncertain data and undesirable outputs. This model is based on the definition of stochastic performance and the concept of risk.Bruni et al. (2014)[15], In the context of stochastic programming, used a stochastic data envelopment analysis model to predict the potential financial problems of potential borrowers. Jin et al.(2014)[28] presented a stochastic data envelopment analysis model to evaluate environmental performance with undesirable outputs. A new stochastic DEA model has been developed to define stochastic performance based on CCP and expected value. A stochastic net environmental performance index (SPEI) is defined based on the models. The proposed random environmental DEA model was used to assess the environmental performance of APEC members in 2010. In the framework of stochastic programming, Beraldi and Bruni (2014) show the output parameters by discrete randomly distributed random variables. They also offer two different models based on a neutral and risk-averse perspective. Tavana et al. (2014). [50] proposed a chance-constrained DEA model with bi-random input and output data. They used a super-efficiency model with bi-random constraints and a nonlinear deterministic equivalent model to solve the super-efficiency model. Branda and Kopa (2014)[8] used two data envelopment analysis (DEA), and second-order stochastic dominance (SSD) approaches to evaluate portfolio performance and compare them. Following Brenda and Kopa (2012a)[9], they also designed DEA risk models based on classic DEA models in which CVaR is used as input or output at some levels. The cumulative probabilities of the scenarios give CVaR levels. They enriched their models with more reward criteria as output. In addition, they introduced a new VRS DEA risk model based on input-output, which is equivalent to testing the performance of a pairwise SSD portfolio. Unlike the DEA equivalent risk models for convex SSD efficiency and SSD portfolio performance presented in Branda and Kopa (2012b)[10], the new DEA risk model allows only binary weights. Sinuany-Stern and Friedman (2016)[48] examined different statistical methods in the DEA. Branda et al. (2016)[12] introduced data envelopment analysis (DEA) models equivalent to efficiency tests concerning the N-th order stochastic dominance (NSD). Park et al.(2017)[39]. Provided a new framework for evaluating vendors. This method uses stochastic discrete simulation and DEA; each supply chain is considered a vendor. Discrete stochastic simulation is used to simulate the supply chain and generate reliable data; then, a deterministic DEA is used to evaluate vendors.Chen et al.(2017)[17] evaluated the performance of 13 Chinese airlines from 2006 to 2014. This research uses the DEA stochastic network (SNDEA) to calculate adverse random outputs, flight delays, and CO_2 emissions.Aleskerov et al. (2017)[1] reviewed the empirical research works evaluating the efficiency scores of universities and their structural units of different levels (from faculties to small groups and research programs).Zhou et al.(2017)[62] examined the performance of 16 commercial banks in China in a study. This model proposes the possibility of random generation and network structure together.Nasseri et al. (2018)[36] proposed a fuzzy stochastic DEA model with Undesirable outputs. Three fuzzy DEA models have been applied according to the constraints of probability–possibility, probability–necessity, and probability–credibility. Wen et al. (2018)[59] proposed a stochastic data envelopment analysis (DSEA) model to solve the optimization problem of various spare parts in uncertain conditions. Also, a factor system is proposed in the product life-cycle process, containing five design indexes, four operation indexes, and five support indexes.

In a study, Davtalab-Olyaie et al.(2019)[24] Presented models for combining all the data in the ranking. They introduced two ranking methods, partial and linear, to evaluate performance in SDEA using the reliability function of the scores. The proposed partial ranking is based on random order, while the linear order is the weighted average of the reliability of the performance scores. Hosseini et al. (2019)[26] developed the Malmacquist index within the framework of the SNDEA(stochastic network data envelopment analysis). They used it to examine the productivity changes of public network-generating units with stochastic data. Kao and Liu (2019)[30] presented a stochastic data envelopment analysis model for performance evaluation using correlated data. The presented model uses a simulation technique to obtain a random efficiency distribution. They use the random number generation technique to generate and simulate random data. The main idea of this model is to transform correlated random variables that follow a multivariate normal distribution into a standard normal random variable. Jradi and Ruggiero (2019)[29] introduced the concept of the most probable quantity. They also developed the Bunker Quantitative Random Data Envelopment Analysis Model to find the most probable Quantum and a new stochastic data envelopment analysis model to estimate production boundaries. Chen and Zhou (2019)[18] investigated the computability of data envelopment analysis models with chance constraints. Their main goal is to provide equivalent simple and tractable models for stochastic models. In the present study, chance-constrained DEA forms are first classified under Gaussian distribution and uncorrelation assumptions, convex tractable

optimization problems, and non-convex unsolvable optimization problems. Al-Khasawneh (2020)[2] examined the Total productivity and cost efficiency dynamics of US merging banks from 1992 to 2003. This study used the Malmacquist index's development in the DEA framework. They used statistical bootstrap to generate random data to overcome the problem's uncertainty. Mehdizadeh.S et al (2020)[35] proposed a random network DEA model with a two-stage structure. This two-stage stochastic network DEA model is formulated based on P-models of chance constraint programming and leader-follower concepts. In addition, the relationship between the stages of the title of leader or follower has been investigated.Beraldi and Bruni (2020)[13] propose a stochastic DEA approach and calculate mean tail risk. A definite equivalent with the assumption of discrete distributions is also presented. While the expected value (E-model) or the most probable value (P-model) of the efficiency level is commonly used in SDEA models, the proposed formula includes a risk measurement. In an article, Tavassoli et al.(2020)[51] Evaluated and measured suppliers' sustainability. They claim this study provides the first integrated approach and flexibility to assess supplier sustainability in supply chain management. This research uses a fuzzy-stochastic model for evaluation, and the alpha-cut method is used to solve the fuzzy-stochastic model. Zhou et al. (2021)[63] used Stochastic leader-follower DEA models to measure the performance of a set of Chinese banks. Amirteimoori.A et al. (2022) measured value-based scale elasticity (SE) using random data envelopment analysis. They introduced a stochastic value-based efficiency measure in the chance constraint programming framework to develop a SE-based value measure.
In the next section (uncertainty), we tried to answer main questions 3, 4, and 5.

2 Review Methodology

This study is a systematic literature review investigating stochastic models in data envelopment analysis. In addition, the uncertainty space investigation has been considered in the issues raised in the coverage analysis of random data. The main purpose of the review is to answer the following main questions:

RQ1: stochastic DEA models are divided into how many categories?
RQ2: Is the fit between the data and the model presented in previous research related to stochastic data envelopment analysis established?
RQ3: How to investigate the uncertainty of issues in the real world?
RQ4: How many categories are uncertainty data divided?
RQ5: On what basis is the type of data uncertainty determined?
In recent years, research on stochastic DEA has attracted the attention of many researchers. To show the necessity of this article, after reviewing the previous articles in this field, the main features of the articles are presented in Tables 1 and 2.

2.1 Search strategy

The research conducted in this study is based on the main research questions. For this research, several databases have been used, including the main academic and international databases. First, the main keywords are searched. Keyword searches cover all subject areas without specifying a course. Then, studies related to stochastic DEA are reviewed. Mainly, the reviewed articles are published from 1963 to November 2022. To achieve this goal, all relevant abstracts of relevant articles are reviewed and keywords are refined in terms of purpose and review.A large volume of literature has been reviewed in order to achieve a deeper understanding of uncertainty in data envelopment analysis. Many published articles have dealt with the fields of fuzzy, robust, meta-heuristic, and other non-deterministic models. A search was performed in well-known academic databases, such as Elsevier, Springer, Emerald, and IEEE Transactions. When searching for keywords, approximately 180 articles from different publishers were found. Out of 180 articles, 52 have dealt with methods and applications of stochastic DEA. Due to the large volume of articles in non-deterministic data envelopment analysis models, only random models have been reviewed. The search process is applied in two dimensions. First, attention has been paid to investigating the most well-known random data envelopment analysis models. Second, different classes of stochastic DEA are identified in different sources.

3 Uncertainty

All kinds of activities in the organization are recorded by process-aware information systems (PAISs). This information can be extracted from PAIS as an event report or a database containing the digital process of the performed operations and recorded as an event.[41] Event reports can be different in form and contain different structured information depending on the information system that implements the data collection in the organization. Although many different event characteristics can be recorded, it is generally assumed that three basic characteristics of an event are included in the log: 1-The time in which the event occurred, 2- the activity performed,3- the case identified to which the event belongs. Uncertain events are recorded activities in a process accompanied by indications of uncertainty in the event's characteristics. Figure1 describes the steps of the mining process in full.[41] There are several reasons for uncertainty: [40][41]
Incorrectness: Occasionally, uncertainty is caused by errors that occurred when the data was recorded.
Coarseness: Some information systems have limitations on how data can be recorded, often related to factors such as the accuracy of the data format, so event data can be considered imprecise. For example, consider an information system that only records the date of an event but not the time.
Ambiguity: Sometimes, the recorded data does not have a specific event attribute identifier. In these cases, the data must be automatically or manually interpreted to obtain a value for the event attribute. Uncertainty may arise if the data is ambiguous and cannot be accurately interpreted. For example, we can refer to images, text, or video data.
These factors cause implicit uncertainty in the event report, and uncertain events produce uncertain data. Uncertain event data is generally divided into two categories:
Strong uncertainty: The possible values for specific attributes are known, but their probabilities are uncertain.
Strong uncertainty: The possible values for specific attributes are known, but their probabilities are uncertain.
Weak uncertainty: An attribute's possible values and associated probabilities are known.

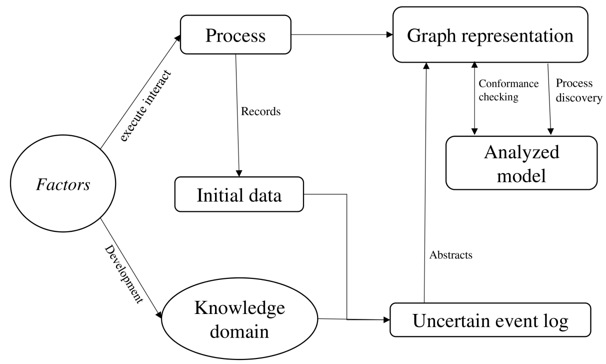

Figure 1: The framework of the mining process on uncertain event data

In the case of a discrete attribute, the concept of strong uncertainty includes the set of possible values assumed for the attribute. In this case, the probability of each possible value is unknown. Conversely, there is a discrete probability distribution over that set of values under weak uncertainty. In the case of a continuous feature, the concept of strong uncertainty can be represented by considering an interval for the variable's value. Note that an interval does not represent a uniform distribution. There is no information about the probability of values in it. Conversely, a probability density function is defined over a given interval in weak uncertainty.We tried to answer questions 1 and 2 in the following two sections

4 Basic Stochastic Models

This section gives some basic stochastic models in DEA to understand the topic better and familiarize ourselves with stochastic modeling.Stochastic programming models are divided into two categories, E-model and P-model, based on the type of objective function.

4.1 E-model

Land et al.(1993)[33] first defined a model based on the concept of expected value.

Definition 4.1. *Expected Value represents the average outcome of a series of random events with identical odds being repeated over a long period.*

Assume that there are n decision making units $DMU_j, (j = 1, ..., n)$ which convert m inputs $x_{ij}, (i = 1, ..., m)$ into s outputs $y_{rj}, (r = 1, ..., s)$ and DMU_o is an under evaluation DMU. Also, suppose that all inputs and outputs are non-negative.Her θ represents the reduction ratio of unit inputs under consideration to improve efficiency. Land et al. (1993) defined the formal form of E-model on the CCR model as follows:

$$\begin{aligned} &\theta_0^* = min\theta \qquad (1)\\ &s.t. \quad p\{\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} \leq \theta \tilde{x}_{io}\} \geq 1 - \alpha \quad i = 1, ..., m \\ &p\{\sum_{j=1}^{n} \lambda_j \tilde{y}_{rj} \leq \theta \tilde{y}_{ro}\} \geq 1 - \alpha \quad r = 1, ..., s \\ &\lambda_j \geq 0 \quad j = 1, ..., n \end{aligned}$$

Where $\alpha \in [0, 1]$ is a predetermined value, in this case, the inputs and outputs are considered random parameters, and by estimating the probability distribution governing them, the future efficiency is estimated. Model (1) should be converted into a definite form. For this purpose, the i-th probable constraint model is written as follows:

$$p\{\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} - \theta \tilde{x}_{io} \leq 0\} \geq 1 - \alpha \quad i = 1, ..., m$$

By defining the external auxiliary variable(covariate) $\eta_i \geq 0$, the above inequality becomes the following equality:

$$p\{\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} - \theta \tilde{x}_{io} \leq 0\} \geq 1 - \alpha + \eta_i \quad i = 1, ..., m$$

By defining auxiliary variable $s_i^- \geq 0$, the above inequality is written as follows:

$$p\{\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} - \theta \tilde{x}_{io} \leq -s_i^-\} \geq 1 - \alpha + \eta_i \quad i = 1, ..., m$$

Now, putting $\tilde{h}_i = \sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} - \theta \tilde{x}_{io}$ and considering that every linear combination of normal random variables has a normal distribution, we have:
where in:

$$\tilde{h}_i \sim N(h_i(\sigma_i^I(\lambda,\theta))^2)$$

$$h_i = E[\tilde{h}_i] = E[\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} - \theta \tilde{x}_{io}] = \sum_{j=1}^{n} \lambda_j x_{ij} - \theta x_{io}$$

$$(\sigma_i^I(\lambda,\theta))^2 = Var[\tilde{h}_i] = var[\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} - \theta \tilde{x}_{io}]$$

$$= var[\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij}] + var[\theta \tilde{x}_{io}] - 2cov[\tilde{h}_i] = var[\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} - \theta \tilde{x}_{io}]$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{n} \lambda_j \lambda_k cov(\tilde{x}_{ij}, \tilde{x}_{ik}) + \theta^2 var(\tilde{x}_{io}) - 2\theta \sum_{j=1}^{n} \lambda_j cov(\tilde{x}_{ij}, \tilde{x}_{io})$$

Therefore, relation (2) is written as follows:

$$P(\tilde{h}_i \leq -s_i^-) = 1 - \alpha$$

Now, using the central limit theorem, we can write:

$$P(\frac{\tilde{h}_i - h_i}{\sigma_i^I(\lambda,\theta)} \leq \frac{-s_i^- - h_i}{\sigma_i^I(\lambda,\theta)}) = 1 - \alpha$$

By placing $\tilde{z}_i = \frac{\tilde{h}_i - h_i}{\sigma_i^I(\lambda,\theta)}$ and considering that $\tilde{z}_i \sim N(0,1)$, we have:

$$P(\tilde{z}_i \leq \frac{-s_i^- - h_i}{\sigma_i^I(\lambda,\theta)}) = 1 - \alpha$$

$$\Rightarrow P(\tilde{z}_i \leq \frac{s_i^- + h_i}{\sigma_i^I(\lambda,\theta)}) = \alpha$$

$$\Rightarrow \phi(\frac{s_i^- + h_i}{\sigma_i^I(\lambda,\theta)}) = \alpha$$

In the above expression, ϕ is the standard normal cumulative distribution function, and according to its invertibility, we have:

$$\frac{s_i^- + h_i}{\sigma_i^I(\lambda,\theta)} = \phi^{-1}(\alpha)$$
$$\Rightarrow s_i^- + h_i - \phi^{-1}(\alpha)\sigma_i^I(\lambda,\theta) = 0$$

By inserting h_i and sorting, the deterministic constraint of the ith input becomes as follows.

$$\sum_{j=1}^{n}\lambda_j\tilde{x}_{ij} + s_i^- - \phi^{-1}(\alpha)\sigma_i^I(\lambda,\theta) = \theta\tilde{x}_{io}$$

Similarly, the deterministic form of the rth probable output constraint will be as follows:

$$\sum_{j=1}^{n}\lambda_j\tilde{y}_{rj} - s_r^- + \phi^{-1}(\alpha)\sigma_r^o(\lambda) = \tilde{y}_{ro}$$

where in:

$$(\sigma_r^o(\lambda))^2 = Var[\tilde{h}_i] = var[\sum_{j=1}^{n}\lambda_j\tilde{y}_{rj}\tilde{y}_{ro}]$$
$$= \sum_{j=1}^{n}\sum_{k=1}^{n}\lambda_j\lambda_k cov(\tilde{y}_{rj},\tilde{y}_{rk}) + var(\tilde{y}_{io})$$
$$-2\sum_{j=1}^{n}\lambda_j cov(\tilde{y}_{rj},\tilde{y}_{ro}), \quad r = 1,...,s$$

As a result, the deterministic form of the input-oriented random CCR model is as follows:

$$\begin{aligned}
& min\theta \\
& s.t. \quad \sum_{j=1}^{n}\lambda_j\tilde{x}_{ij} + s_i^- - \phi^{-1}(\alpha)\sigma_i^I(\lambda,\theta) = \theta\tilde{x}_{io} \\
& \sum_{j=1}^{n}\lambda_j\tilde{y}_{rj} - s_r^- + \phi^{-1}(\alpha)\sigma_r^o(\lambda) = \tilde{y}_{ro} \\
& s_i^- \geq 0, \quad s_r^+ \geq 0, \quad \lambda_j \geq 0 \\
& i = 1,...,m, \quad j = 1,...,n, \quad r = 1,...,s
\end{aligned}$$

which, by defining non-negative variables $u_r = \sigma_r^o(\lambda)$ and $v_i = \sigma_i^I(\lambda, \theta)$, turns into the following quadratic model:

$$
\begin{aligned}
& min\theta \\
& s.t. \quad \sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} + s_i^- - v_i \phi^{-1}(\alpha) = \theta \tilde{x}_{io} \\
& \sum_{j=1}^{n} \lambda_j \tilde{y}_{rj} - s_r^- + u_r \phi^{-1}(\alpha) = \tilde{y}_{ro} \\
& v_i = \sum_{j=1}^{n} \sum_{k=1}^{n} \lambda_j \lambda_k cov(\tilde{x}_{ij}, \tilde{x}_{ik}) + \theta^2 var(\tilde{x}_{io}) \\
& -2\theta \sum_{j=1}^{n} \lambda_j cov(\tilde{x}_{ij}, \tilde{x}_{io}) \\
& u_r = \sum_{j=1}^{n} \sum_{k=1}^{n} \lambda_j \lambda_k cov(\tilde{y}_{rj}, \tilde{y}_{rk}) + var(\tilde{y}_{io}) \\
& -2 \sum_{j=1}^{n} \lambda_j cov(\tilde{y}_{rj}, \tilde{y}_{ro}) \\
& s_i^- \geq 0, \quad s_r^+ \geq 0, \quad \lambda_j \geq 0 \\
& u_r, v_i \geq 0 \\
& i = 1, ..., m, \quad j = 1, ..., n, \quad r = 1, ..., s
\end{aligned}
$$

Definition 4.2. *DMU_o is called stochastic efficiency in CCR input-oriented stochastic model if and only if $\theta^* = 1$.*

In this section, the random variable error structure, presented by Cooper et al. (1998), is introduced to transform the problem with the probable constraint into a deterministic linear form. Of course, Sharp (1963) and Kahane (1977) used the single-factor assumption of variables in economics and finance. Assume that the inputs and outputs have a single-factor symmetric distribution as shown below.

$$
\begin{aligned}
\tilde{x}_{ij} &= x_{ij} + a_{ij} \tilde{\xi}_{ij} \\
\tilde{y}_{rj} &= y_{rj} + b_{rj} \tilde{\epsilon}_{rj}
\end{aligned} \tag{2}
$$

Where b_{rj}, a_{ij}, y_{rj} and x_{ij} are non-negative, and $\tilde{\xi}_{ij}$ and $\tilde{\epsilon}_{rj}$ are independent random variables with standard normal distribution, which are called the error of input and output random variables. Due to the symmetric nature of the normal distribution, structure four is called a symmetric error structure. Also, x_{ij} and y_{rj} are the expected value, and a_{ij} and b_{rj} are the standard deviations of the random variables $\tilde{x}_{ij}$ and $\tilde{y}_{rj}$, respectively. Furthermore, assume that all inputs and outputs are independent. That is, for each $j \neq k$:

$$cov(\tilde{x}_{ij}, \tilde{x}_{ik}) = 0$$
$$cov(\tilde{y}_{rj}, \tilde{y}_{rk}) = 0$$

It follows from relation (2):

$$\tilde{x}_{ij} \sim N(x_{ij}, \sigma^2 a_{ij})$$
$$\tilde{y}_{rj} \sim N(y_{rj}, \sigma^2 b_{rj})$$

Also suppose that $\tilde{\xi}_i = \tilde{\xi}_{ij}, \tilde{\epsilon}_r = \tilde{\epsilon}_{rj}, j = 1, ..., n$.
Now consider the probable input constraint of the model (1).

$$p\{\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} \leq \theta \tilde{x}_{io}\} \geq 1 - \alpha \quad i = 1, ..., m$$

Now by placing $\tilde{h}_i = \sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} - \theta \tilde{x}_{io}$ and considering the normal distribution and error structure and assuming independence of inputs and outputs, the result is:

$$\tilde{h}_i = (\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} - \theta \tilde{a}_{io}) - \tilde{\xi}_i(\sum_{j=1}^{n} \lambda_j \tilde{a}_{ij} - \theta \tilde{x}_{io}) \quad i = 1, ..., m$$

So that:

$$\tilde{h}_i \sim ((\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} - \theta \tilde{a}_{io}), \sigma^2(\sum_{j=1}^{n} \lambda_j \tilde{a}_{ij} - \theta \tilde{x}_{io})^2) \quad i = 1, ..., m$$

According to (3) the constraint (3) turns into the following deterministic form:

$$\sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} - \phi^{-1}(\alpha)\sigma|\sum_{j=1}^{n} \lambda_j \tilde{a}_{ij} - \theta \tilde{a}_{io}| \leq \theta \tilde{x}_{io} \quad i = 1, ..., m$$

Also, the probable constraint output of the model (1) becomes the following deterministic form:

$$\sum_{j=1}^{n} \lambda_j \tilde{y}_{rj} + \phi^{-1}(\alpha)\sigma|\sum_{j=1}^{n} \lambda_j \tilde{b}_{rj} - \theta\tilde{b}_{ro}| \geq \tilde{y}_{ro} \quad r = 1, ..., s$$

As a result, model (1) becomes the following deterministic form:

$$\begin{aligned}
& min\theta && (3)\\
& s.t. \quad \sum_{j=1}^{n} \lambda_j \tilde{x}_{ij} - \phi^{-1}(\alpha)\sigma|\sum_{j=1}^{n} \lambda_j \tilde{a}_{ij} - \theta\tilde{a}_{io}| \leq \theta\tilde{x}_{io} \quad i = 1, ..., m \\
& \sum_{j=1}^{n} \lambda_j \tilde{y}_{rj} + \phi^{-1}(\alpha)\sigma|\sum_{j=1}^{n} \lambda_j \tilde{b}_{rj} - \theta\tilde{b}_{ro}| \geq \tilde{y}_{ro} \quad r = 1, ..., s \\
& \lambda_j \geq 0 \quad j = 1, ..., n
\end{aligned}$$

The above model is a non-linear model due to the existence of the absolute value. To remove the absolute term, we use the ideal programming presented by Charnes and Cooper (1961,1977). For this purpose, consider the following transformations:

$$\begin{aligned}
& |\sum_{j=1}^{n} \lambda_j \tilde{a}_{ij} - \theta\tilde{a}_{io}| = p_i^+ + p_i^- \quad i = 1, ..., m \\
& \sum_{j=1}^{n} \lambda_j \tilde{a}_{ij} - \theta\tilde{a}_{io} = p_i^+ + p_i^- \quad i = 1, ..., m \\
& p_i^+ p_i^- = 0 \\
& |\sum_{j=1}^{n} \lambda_j \tilde{b}_{rj} - \theta\tilde{b}_{ro}| = q_r^+ + q_r^- \quad r = 1, ..., s \\
& \sum_{j=1}^{n} \lambda_j \tilde{b}_{rj} - \theta\tilde{b}_{ro} = q_r^+ + q_r^- \quad r = 1, ..., s \\
& q_r^+ q_r^- = 0
\end{aligned}$$

By placing the above transformations in problem (3), the resulting model is a non-linear problem due to the existence of the $p_i^+ p_i^- = 0$ and

$q_r^+ q_r^- = 0$ constraints. Since when the linear programming problem has an optimal solution, this solution occurs at the vertex point. In this case, for each $i = 1, ..., m$, at least one of the values of p_i^+ or p_i^- and for each $r = 1, ..., s$ at least one of the values of q_r^+ or q_r^- will be zero.because their vector coefficients are linearly dependent and cannot be present simultaneously in the basic matrix of the optimal solution. As a result, by using the simplex algorithm, this optimal point can be calculated and the constraints of $p_i^+ p_i^- = 0$ and $q_r^+ q_r^- = 0$ can be removed from the problem.

$$
\begin{aligned}
& min\theta \\
& s.t. \quad \sum_{j=1}^{n} \lambda_j x_{ij} - \phi^{-1}(\alpha)\sigma(p_i^+ + p_i^-) \leq \theta x_{io} \quad i = 1, ..., m \\
& \sum_{j=1}^{n} \lambda_j a_{ij} - \theta a_{io} = p_i^+ - p_i^- \quad i = 1, ..., m \\
& \sum_{j=1}^{n} \lambda_j b_{rj} + \phi^{-1}(\alpha)\sigma(q_r^+ + q_r^-) \geq y_{ro} \quad r = 1, ..., s \\
& \sum_{j=1}^{n} \lambda_j b_{rj} - \theta b_{ro} = q_r^+ - q_r^- \quad r = 1, ..., s \\
& \lambda_j \geq 0 \quad j = 1, ..., n
\end{aligned}
$$

4.2 P-model

Here is a more general class that Charns and Cooper refer to as "P-Models."These types of models are formulated based on the concept of probability.

Definition 4.3. *Probability measures how certain we are a particular event will happen in a specific instance.*

In the following, the general and basic form of the P-Model model is introduced to adapt the usual definitions of "DEA efficiency" to the field of limited chance programming. In this model, the Chance Constrained framework is used for the multiple CCR model[21].

$$\begin{aligned} & Maxp(\frac{\sum_{j=1}^{n} u_r \tilde{y}_{ro}}{\sum_{j=1}^{n} v_i \tilde{x}_{io}} \geq 1) \\ & s.t. \;\; p(\frac{\sum_{j=1}^{n} u_r \tilde{y}_{rj}}{\sum_{j=1}^{n} v_i \tilde{x}_{ij}} \geq 1) \geq 1-\alpha_j \quad j=1,...,n \\ & u_r, v_i \geq 0 \quad r=1,...,s, \quad i=1,...,m \end{aligned} \tag{4}$$

where P means the probability and v_i, u_r are the respective weights of inputs and outputs which are obtained by solving the problem. Model (4) for $u_r = 0$ and $v_i > 0$ is feasible for all i and r[16]. Therefore, for any continuous probability distribution we can write:

$$p(\frac{\sum_{j=1}^{n} u_r^* \tilde{y}_{ro}}{\sum_{j=1}^{n} v_i^* \tilde{x}_{io}} \geq 1) + p(\frac{\sum_{j=1}^{n} u_r^* \tilde{y}_{ro}}{\sum_{j=1}^{n} v_i^* \tilde{x}_{io}} \leq 1) = 1$$

or

$$p(\frac{\sum_{j=1}^{n} u_r^* \tilde{y}_{ro}}{\sum_{j=1}^{n} v_i^* \tilde{x}_{io}} \leq 1) = 1-\alpha^* \geq 1-\alpha_o$$

Here $\alpha^*$is the optimal value of the model (4), so $1-\alpha^*$ is the probability of achieving a value of at least unity choosing these optimal weights, and therefore $1-\alpha^*$ is the probability of not achieving this value[58].

$$\begin{aligned} & Max \;\; p(\frac{\sum_{j=1}^{n} u_r^* \tilde{y}_{ro}}{\sum_{j=1}^{n} v_i^* \tilde{x}_{io}} \geq 1) \\ & s.t. \;\; p(\frac{\sum_{j=1}^{n} u_r^* \tilde{y}_{rj}}{\sum_{j=1}^{n} v_i^* \tilde{x}_{ij}} \geq 1) + p(\frac{\sum_{j=1}^{n} u_r^* \tilde{y}_{rj}}{\sum_{j=1}^{n} v_i^* \tilde{x}_{ij}} \leq 1) \geq 1 \quad j=1,...,n \end{aligned}$$

It should be noted that $\alpha_o \geq \alpha^*$ because $1-\alpha_o$ is always $DMU_o$for the constraint $j = o$ as the probability of inefficiency. It is predetermined. Therefore, we call DMU_ostochasticly efficient if and only if $\alpha_o = \alpha^*$.
In programming models with probability constraints, it is assumed that the stochastic variables are in a multivariate distribution. Therefore, it is possible to examine the value of α^* even before generating the data. Charnes et al. (1958) related the programming with probability constraints to the policies related to programming black oil production for an oil company.This led to forming a risk assessment committee to select the appropriate options for α^*.

4.3 Satisfaction-model

The satisfaction model is a generalization of model (5), which is analyzed further[21]:

$$
\begin{aligned}
&Max\ \ p(\frac{\sum_{j=1}^{n} u_r^* \tilde{y}_{ro}}{\sum_{j=1}^{n} v_i^* \tilde{x}_{io}} \geq \beta_0) \\
&s.t.\ \ p(\frac{\sum_{j=1}^{n} u_r^* \tilde{y}_{rj}}{\sum_{j=1}^{n} v_i^* \tilde{x}_{ij}} \leq \beta_j) \geq 1-\alpha_j \ \ j=1,...,n \\
&\qquad p(\frac{\sum_{j=1}^{n} u_r^* \tilde{y}_{rj}}{\sum_{j=1}^{n} v_i^* \tilde{x}_{ij}} \geq \beta_j) \geq 1-\alpha_j \ \ j=n,...,n+k
\end{aligned}
$$

In model (5), β_o is called the Aspiration level, whose value is imposed by an external factor, such as the budget model of Estdari (1960), or adopted by the manager for some activities, such as Simon's concept of satisfaction (1957). Aspirational level is the value of a target variable that should be achieved or replaced by a satisfactory decision.
The above models are very general and are generally considered for conceptual interpretation. In order to perform the calculations, one should use their deterministic equivalent form. For this purpose, we assume that all inputs and outputs are random variables with multivariate normal distribution with specific parameters. Choosing a multivariate normal distribution may be a bit limiting. But by using transformations, other types of distributions can be converted almost into normal distribution forms.
In the following, using Cooper et al.'s (1996) transformation, we transform model (5) into a deterministic form.
According to the constraint of model (5), we can write:

$$
\begin{aligned}
&p(\frac{u^T \tilde{y}_j}{v^T x_j \leq \beta_j}) = p(u^T \tilde{y}_j \leq v^T x_j \beta_j) \\
&= p(\frac{u^T \tilde{y}_j - u^T \overline{y}_j}{\sqrt{u^T \Sigma_j u}} \leq -\frac{u^T \overline{y} - \beta_j v^T x_j}{\sqrt{u^T \Sigma_j u}})
\end{aligned} \quad (5)
$$

where $y_j = E[\tilde{y}_j]$, $x_j = E[\tilde{x}_j]$ and $\Sigma_j = cov(\tilde{y}_{ij}, \tilde{y}_{kj})$: The stochastic variable z_j is defined as follows:

$$z_j = \frac{u^T\tilde{y}_j - u^T\overline{y}_j}{\sqrt{u^T\Sigma_j u}}, \quad j = 1, ..., n.$$

so that z_j has a standard normal distribution($z_j \sim (0, 1)$). By inserting (6) in (5), the relation (6) can be rewritten as follows:

$$p(z_j \leq -\frac{u^T\overline{y}_j - \beta_j v^T x_j}{\sqrt{u^T\Sigma_j u}}) \geq 1 - \alpha_j$$

According to the standard normal distribution function and its invertibility, (6) is written as follows:

$$-\frac{u^T\overline{y}_j - \beta_j v^T x_j}{\sqrt{u^T\Sigma_j u}} \geq \phi^{-1}(1 - \alpha_j)$$

Where$\phi$is the standard normal distribution function and ϕ^{-1} is its inverse.

Now, by introducing non-negative spacer variable η_j, which Charnes and Cooper introduced (1963), relation (6) is replaced by the following two inequalities:

$$\begin{aligned} &u^T\overline{y}_j - \beta_j v^T x_j - \eta_j\phi^{-1}(\alpha_j) \leq 0 \\ &C_{\alpha_j}[\eta_j^2 - u^T\Sigma_j u] \geq 0 \end{aligned} \tag{6}$$

So that:

$$C_{\alpha_j} = \begin{cases} 1 & \alpha_j < 0.5 \\ 0 & \alpha_j = 0.5 \\ -1 & \alpha_j > 0.5 \end{cases}$$

By placing the deterministic constraints (6), instead of the probabilistic constraint of model (5), model (7) is obtained.

$$\begin{aligned} &Max \;\; p(\frac{u^T\tilde{y}_o}{v^T x_o} \geq \beta_o) \\ &s.t. \;\; u^T\overline{y}_j - \beta_j v^T x_j - \eta_j\phi^{-1}(\alpha_j) \leq 0 \quad j = 1, ..., n \\ &C_{\alpha_j}[\eta_j^2 - u^T\Sigma_j u] \geq 0 \quad j = 1, ..., n \\ &u, v, \eta \geq 0 \end{aligned} \tag{7}$$

Due to random variables $\tilde{x}_o$ and $\tilde{y}_o$ in the objective function, problem (7) has not yet turned into a deterministic form. For this purpose, consider problem (8), which is equivalent to problem (7).

$$\begin{aligned}
&Max \;\; \gamma \\
&s.t. \;\; p(\frac{u^T\tilde{y}_o}{v^T x_o} \geq \beta_o) \geq \gamma \\
&u^T\overline{y}_j - \beta_j v^T x_j - \eta_j \phi^{-1}(\alpha_j) \leq 0 \quad j = 1, ..., n \\
&C_{\alpha_j}[\eta_j^2 - u^T \Sigma_j u] \geq 0 \quad j = 1, ..., n \\
&u, v, \eta \geq 0
\end{aligned} \tag{8}$$

Now, the first constraint of the model (8), like the previous constraints, turns into a deterministic form.

$$\begin{aligned}
&p(\frac{-u^T\tilde{y}_o}{v^T x_o} \geq \beta_o) = p(-u^T\tilde{y}_o \leq v^T x_o \beta_o) \\
&= p(\frac{-u^T\tilde{y}_o + u^T\overline{y}_o}{\sqrt{u^T\Sigma_o u}} \leq -\frac{v^T x_o\beta_o - u^T\overline{y}_o}{\sqrt{u^T\Sigma_o u}}) \\
&= p(\tilde{z}_o \leq -\frac{v^T x_o\beta_o - u^T\overline{y}_o}{\sqrt{u^T\Sigma_o u}})
\end{aligned}$$

The first constraint of model(8)can be written as follows:

$$p(\tilde{z}_o \leq -\frac{v^T x_o\beta_o - u^T\overline{y}_o}{\sqrt{u^T\Sigma_o u}}) \geq \gamma.$$

Which is equivalent to:

$$\frac{u^T\overline{y}_o - v^T x_o\beta_o}{\sqrt{u^T\Sigma_o u}} \geq \phi^{-1}(\gamma)$$

Now the model (8) can be rewritten as below:

$$\begin{aligned}
&Max \;\; \gamma \\
&s.t. \;\; \frac{u^T\overline{y}_o - v^T x_o\beta_o}{\sqrt{u^T\Sigma_o u}} \geq \phi^{-1}(\gamma) \\
&u^T\overline{y}_j - \beta_j v^T x_j - \eta_j \phi^{-1}(\alpha_j) \leq 0 \quad j = 1, ..., n \\
&C_{\alpha_j}[\eta_j^2 - u^T \Sigma_j u] \geq 0 \quad j = 1, ..., n \\
&u, v, \eta \geq 0
\end{aligned} \tag{9}$$

Although problem (9) is deterministic, it is non-convex because of the fractional constraint.To solve this problem, consider the following:

$$
\begin{aligned}
& Max \;\; \xi \\
& s.t. \;\; \frac{u^T\overline{y}_o - v^T x_o \beta_o}{\sqrt{u^T\Sigma_o u}} \geq \xi \\
& u^T\overline{y}_j - \beta_j v^T x_j - \eta_i \phi^{-1}(\alpha_j) \leq 0 \;\; j=1,...,n \\
& C_{\alpha_j}[\eta_j^2 - u^T\Sigma_j u] \geq 0 \;\; j=1,...,n \\
& u, v, \eta \geq 0
\end{aligned}
\tag{10}
$$

As shown in (9), $\phi^{-1}(\gamma)$ is a strictly increasing function of γ. (9) and (10) have the same solution structure and in each pair of optimal solutions for these two problems, we have:

$$\xi^* = \phi^{-1}(\gamma^*)$$

Here * is to represent the optimal value. Now it is easy to see that (10) is equivalent to the following problem:

$$
\begin{aligned}
& Max \;\; \frac{u^T\overline{y}_o - v^T x_o \beta_o}{\sqrt{u^T\Sigma_o u}} \\
& s.t. \;\; u^T\overline{y}_j - \beta_j v^T x_j - \eta_j \phi^{-1}(\alpha_j) \leq 0 \;\; j=1,...,n \\
& C_{\alpha_j}[\eta_j^2 - u^T\Sigma_j u] \geq 0 \;\; j=1,...,n \\
& u, v, \eta \geq 0
\end{aligned}
\tag{11}
$$

Consider the positive variable ω as follows:

$$\omega = \sqrt{u^T\Sigma_o u}$$

(11) can be rewritten as follows:

$$
\begin{aligned}
& Max \;\; \frac{u^T\overline{y}_o - v^T x_o \beta_o}{\omega} \\
& s.t. \;\; u^T\Sigma_o u - \omega^2 \geq 0 \\
& u^T\overline{y}_j - \beta_j v^T x_j - \eta_j \phi^{-1}(\alpha_j) \leq 0 \;\; j=1,...,n \\
& C_{\alpha_j}[\eta_j^2 - u^T\Sigma_j u] \geq 0 \;\; j=1,...,n \\
& u, v, \eta \geq 0
\end{aligned}
\tag{12}
$$

This problem (12) includes a fractional function in the objective function. Hence, the Charnes-Cooper transformation($CharnesandCooper$(1962))can be used:
Considering $t := \frac{1}{\omega}$, $\mu := tu$, $\nu := tv$ and$\zeta := t\eta$ problem (12) can be written as the following quadratic programming problem.

$$\begin{aligned} & Max \;\; u^T\overline{y}_o - v^T x_o\beta_o \\ & s.t. \;\; u^T\Sigma_o u \geq 1 \\ & u^T\overline{y}_j - \beta_j v^T x_j - \zeta_j\phi^{-1}(\alpha_j) \leq 0 \;\; j = 1, ..., n \\ & C_{\alpha_j}[\zeta_j^2 - u^T\Sigma_j u] \geq 0 \;\; j = 1, ..., n \\ & u, v, \zeta \geq 0 \end{aligned} \tag{13}$$

Theorem 4.4. *Let* (μ_*, ν_*, ζ_*) *and* (u_*, v_*) *be optimal solutions of (13) and (5), respectively; then*

$$\phi(\mu^{*T}\overline{y}_o - \beta_o\nu^{*T}x_o) = p(\frac{u^{*T}\overline{y}_o}{v^{*T}x_o} \geq \beta_o)$$

Furthermore, DMUo is stochastically efficient if and only if $\phi(\mu^{*T}\overline{y}_o - \beta_o\nu^{*T}x_o) = \alpha_o$.

Proof. in $Cooper$(1996)[21] □ Now assume more specifically that the output components are related only through some basic underlying factors. In the sense that this factor solely determines all components of each output. More precisely,

$$\begin{aligned} & Max \;\; \mu^T\tilde{y}_o - \beta_o\nu^T x_o \\ & s.t. \;\; \mu^T b_o = 1 \\ & B_j\nu^T x_j - \mu^T\overline{y}_j \geq \phi^{-1}(1 - \alpha_j)\mu^T b_j \quad j = 1, ..., n \\ & \mu \geq 0, \nu \geq 0 \end{aligned} \tag{14}$$

The dual of (14) is as follows:

$$\begin{aligned} & Min \quad -\theta \\ & s.t. \quad \sum_{j=1}^{n} \lambda_j[\overline{y}_j + \phi^{-1}(1-\alpha_j b_j)] \geq \overline{y}_o + \theta b_o \\ & \sum_{j=1}^{n} \lambda_j \beta_j x_j \leq \beta_o x_o \\ & \lambda \geq 0 \end{aligned} \tag{15}$$

Deterministic equivalents under the assumption that outputs and inputs are both stochastic
The dual of (15) is as follows:

$$\begin{aligned} & Max \quad \mu^T \overline{y}_o - \beta_o \nu^T \overline{x}_o \\ & s.t. \;\; \mu^T \Sigma_o^{oo} \mu - 2\beta_o \mu^T \Sigma_o^{oI} \nu + \beta_o^2 \nu^T \Sigma_o^{II} \nu \geq 1 \\ & \mu^T \overline{y}_j - \beta_j \mu^T \overline{x}_j - \phi^{-1}(\alpha_j \zeta_j) \leq 0 \quad j = 1, ..., n \\ & C_{\alpha_j}[mu^T \Sigma_j^{oo} \mu - 2\beta_j \mu^T \Sigma_j^{oI} \nu + \beta_j^2 \nu^T \Sigma_j^{II} \nu - \zeta_j^2] \leq 0 \quad j = 1, ..., n \\ & \mu \geq 0, \nu \geq 0, \zeta \geq 0, \quad j = 1, ..., n. \end{aligned} \tag{16}$$

$$\begin{aligned} & \Sigma_j^{oo} = (cov(\tilde{y}_{ij}, \tilde{y}_{kj}))_{s\times s} \\ & \Sigma_j^{oI} = (cov(\tilde{y}_{ij}, \tilde{x}_{kj}))_{s\times m} \\ & \Sigma_j^{II} = (cov(\tilde{x}_{ij}, \tilde{x}_{kj}))_{m\times m} \\ & j = 1, ..., n. \end{aligned}$$

Now the previous theorems and definitions for stochastic efficiency can be expanded as follows:

Theorem 4.5. *Let* (μ_*, ν_*, ζ_*) *and* (u_*, v_*) *be optimal solutions of (16) and (5), respectively; then:*

$$\phi(\mu^{*T} \overline{y}_o - \beta_o \nu^{*T} \overline{x}_o) = p(\frac{u^{*T} \tilde{y}_o}{v^{*T} \tilde{x}_o} \geq \beta_o)$$

Furthermore, DMU_o *is called stochastic efficient if and only if :* $\phi(\mu^{*T} \overline{y}_o - \beta_o \nu^{*T} \overline{x}_o) = \alpha_o$.

As in the previous section, assume that the input and output components are related only through a basic factor:

$$\tilde{x}_{ij} = \overline{x}_{ij} + a_{ij}\xi$$
$$\tilde{y}_{rj} = \overline{y}_{rj} + b_{rj}\xi$$

Using the same analysis as before, we have:

$$\begin{aligned} &Max \quad \mu^T \overline{y}_o - \beta_o \nu^T \overline{x}_o \\ &s.t. \ \ |\beta_o \nu^T a_o - \mu^T b_o| \geq 1 \\ &\beta_j \nu^T \overline{x}_j - \mu^T \overline{y}_j \geq \phi^{-1}(1-\alpha_j)|\beta_j \nu^T a_j - \mu^T b_j| \quad j = 1, ..., n \\ &\mu \geq 0, \nu \geq 0 \end{aligned} \tag{17}$$

The presence of absolute values in the constraints makes problem (17) no longer an ordinary linear programming problem.The goal programming theory developed by Charnes and Cooper (1961, 1977) can be used to transform problem (17) into a quadratic programming problem.
Consider the expression$|\beta_j \nu^T a_j - \mu^T b_j|$. if $\beta_j \nu^T a_j - \mu^T b_j \geq 0$,let$\eta_j^+ = \beta_j \nu^T a_j - \mu^T b_j$otherwise $\eta_j^- = -(\beta_j \nu^T a_j - \mu^T b_j)$ Hence:

$$\begin{aligned} &(\beta_j \nu^T a_j - \mu^T b_j) = \eta_j^+ - \eta_j^-, \\ &\eta_j^+ \eta_j^- = 0, \eta_j^+ \geq 0, \eta_j^- \geq 0 \end{aligned}$$

Due to the nature of goal programming, inequalities must be satisfied in any solution. So:

$$\begin{aligned} &\eta_o^+ + \eta_o^- \geq 1 \\ &(\beta_o \nu^T a_o - \mu^T b_o) = \eta_o^+ - \eta_o^- \\ &\eta_o^+ \eta_o^- = 0, \eta_o^+ \geq 0, \eta_o^- \geq 0 \end{aligned}$$

to replace the first constraint in (17), and use

$$\begin{aligned} &(\beta_o \nu^T a_o - \mu^T b_o) = \eta_o^+ - \eta_o^- \\ &(\beta_j \nu^T a_j - \mu^T b_j) = \eta_j^+ - \eta_j^-, \\ &\eta_j^+ \eta_j^- = 0 \\ &\eta_j^+ \geq 0, \eta_j^- \geq 0 \end{aligned}$$

to replace the second constraint in (17) for each$j$.[21]

$$
\begin{aligned}
&Max \quad \mu^T\overline{y}_o - \beta_o\nu^T\overline{x}_o \\
&s.t. \ \ \eta_o^+\eta_o^- \geq 1 \\
&\beta_j\nu^T\overline{x}_j - \mu^T\overline{y}_j \geq \phi^{-1}(1-\alpha_j)(\eta_j^+\eta_j^-) \quad j=1,...,n \\
&\beta_j\nu^T a_j - \mu^T b_j = \eta_j^+ - \eta_j^-, \quad j=1,...,n \\
&\eta_j^+\eta_j^- = 0 \quad j=1,...,n \\
&\mu \geq 0, \nu \geq 0, \eta_j^+ \geq 0, \eta_j^- \geq 0 \quad j=1,...,n.
\end{aligned} \tag{18}
$$

Problem (18) is referred to as in efficiency analysis form with the maximization directed to the choices of z and v, which yield the largest value of a satisfactory probability of achieving an aspiration ratio level of weighted outputs to weighted inputs allowed by the constraints. In this way, we retain contact with the earlier discussion of both "satisficing" concepts and DEA frontiers.

5 Detailed Analyses of The Literature

In this section, the stochastic models presented in the literature have been analyzed in detail. Some structural features of the stochastic models presented so far are listed in Tables (1) and Table (2). In table (1), in the CRS and VRS columns, the type of return to scale of the preceding models is determined.In 18 articles, variable return to scale is considered; in 34 articles, constant return to scale efficiency is considered. In the black box and network columns, the type and structure of the problem are determined in previous articles.Stochastic network models are used in 8 articles, and black box structure is used in 44 articles to solve the problem. In the model type column, the stochastic structure of the preceding models is specified. The Undesirable Data column applies to those problems for which undesirable data is used.In four articles, modeling in the presence of undesirable outputs has been done. The envelope or multiple types of the presented models are specified in the last two columns. Table (2) shows the application of stochastic DEA models in the past. Also, the inputs, outputs, and nature of the data used (deterministic, stochastic, and descriptive) have been determined in previous research. In some research, a numerical example has been used to validate the model.In the surveys conducted (Table 2), we have categorized the types of data into three categories: deterministic, stochastic, and descriptive. In some articles, the data used are not mentioned, which are not mentioned in table (2). By using tables (1) and (2), comprehensive information can be obtained about the research done in the past (in the field of stochastic data Envelopment analysis).

Stochastic data envelopment analysis models are divided into four categories as follows.

Chance Constraint Programming: *Chance Constraint Programming accepts stochastic changes in data and allows constraints to be violated up to a given probability.*

Fuzzy Stochastic: *It is similar to Chance Constraint Programming except that the data set is treated as fuzzy.*

Statistical Analysis: *In this method, by designing an algorithm and using the relationships governing statistics and probabilities, and using statistical software, after choosing the appropriate distribution, they produce reliable random data. In the end, they analyze the sensitivity of the presented model.*

DEA and SFA: *The DEA method is more flexible in the case of multiple inputs and outputs but cannot effectively deal with measurement error in the data, while stochastic frontier analysis(SFA) is more effective in the presence of noise. SFA is also a method to determine the efficiency frontier in the presence of noise.*

Therefore, the round combination of methods for solving problems shows a suitable performance. The strengths of one approach can cover the weaknesses of another approach. The purpose of using DEA is to improve the accuracy of SFA efficiency estimates.

Table 1: Structural characteristics of previous non-deterministic models in data envelopment analysis.
Legend:(BB: Black Box), (N: Network), (VRS: Variable Returns to Scale), (CRS: Constant Returns to Scale), (MT: Model Type), (UD: Undesirable Output), (MF: Multiple Form), (EF; Envelopment Form)

Author	MT	BB	N	VRS	CRS	UD	MF	EF
Amirteimoori et al. (2023)	P-Model, linear	*		*				*
Zhou et al. (2021)	P-Model, linear		*		*			*
Tavassoli et al.(2020)	fractional, Non-linear	*		*				*
Wanke et al.(2020)	dynamic network DEA,SFA		*	*				*
Al-Khasawneh et al. (2020)	non-parametric bootstrapped analysis	*		*				*
Beraldi and Bruni. (2020)	E-Model, Nonlinear	*			*		*	
Mehdizadeh et al. (2020)	P-Model, linear		*	*			*	
Davtalab et al. (2019)	mean and median ordering in SDEA, partial and linear	*			*			*
Jradi and Ruggiero. (2019)	Combination of a quantileregression and DEA	*		*				*
Hosseini et al. (2019)	P-Model, Nonlinear		*		*			*
Chen and Zhu. (2019)	P-Model, Nonlinear	*			*			*
Kao and Liu. (2019)	Mean-Var, linear	*		*			*	
Park et al. (2018)	Discrete event simulation (PS-DE) and SDEA		*					
Wen et al. (2018)	P-Model, based on Additive model,linear	*						*
Nasseri et al. (2018)	Fuzzy Stochastic Data Envelopment Analysis	*			*	*	*	
Aleskerov et al. (2017)	combination of DEA and SFA	*			*		*	

Chen et al. (2017)	A combination including non-linear, linear and statistical simulation		*	*		*	*	
Zhou et al. (2017)	P-Model, linear		*		*			*
Aleskerov et al. (2017)	combination of DEA and SFA	*						
Branda and Kopa. (2016)	P-model, linear	*			*			*
Sinuany-Stern and Friedman. (2016, February)	Statistical Analysis in the DEA Context							
Mitropoulos et al. (2015)	P-Model and Bayesian analysis, Nonlinear	*			*			*
Jin et al. (2014)	E-Model, linear	*			*	*	*	
Branda and Kopa. (2014)	Conditional Value at Risk(CVaR)	*		*	*			*
Wei et al. (2014)	P-Model, linear	*			*		*	
Tavana et al. (2014)	P-Model, Nonlinear	*			*			*
Wu et al. (2013)	P-model	*			*	*		*
Assaf, A. G. (2012)	A Bayesian combination of DEA and SFA	*						
Hosseinzadeh Lotfi et al. (2012)	P-Model, Nonlinear, Stochastic centralized resource allocation (SCRA)	*		*				*
Beraldi and Bruni. (2012)	Conditional Value at Risk(CVaR)	*			*		*	
Simar and Zelenyuk. (2011)	Stochastic FDH/DEA (SFA)	*		*				*
Tsionas and Papadakis. (2010)	P-Model and Bayesian analysis, Nonlinear	*			*			*
Wu and Olson. (2010)	P-Model (DEA VaR), Nonlinear	*			*			*
Khodabakhshi. (2010)	P-Model, Nonlinear	*		*				*

Barnum et al. (2010)	DEA and statistical Panel Data Analysis (PDA)	*			*			*
Bruni et al. (2009)	P-Model (LLT-model (Land etal.,1993)), Nonlinear	*		*				*
Simar, L. (2007)	Stochastic FDH/DEA (SFA)	*		*				*
Margari et al. (2007)	combination of DEA and SFA	*		*				*
Olesen, O. B. (2006)	P-Model, Nonlinear	*			*			*
Cooper et al. (2004)	P-Model, Nonlinear	*		*				*
Tsionas, E. G. (2003)	A Bayesian combination of DEA and SFA	*						
Cooper et al. (2002)	P-Model, Nonlinear	*		*				*
Chen. (2002)	P-Model, Nonlinear	*			*			*
Post. (2001)	Mean-Var, linear	*			*			*
Huang and Li. (2001)	P-Model, linear	*		*				*
Sueyoshi, T. (2000)	E-Model, Nonlinear	*			*		*	
Cooper et al. (1996)	P-Model, Nonlinear	*						
Cooper et al. (1995)	combination of DEA and SFA	*			*			*
Olesen and Petersen. (1995)	P-Model, Nonlinear	*			*			*
Land et al. (1993)	E-Model, Nonlinear	*			*			*
Sengupta, J. K. (1987)	P-Model, Nonlinear	*			*		*	
Charnes and Cooper. (1963)	P-Model and E-Model, Nonlinear	*			*		*	

Table 2: Input, output and type of data used in previous studies

Author	case study (application)	data type	output	Input
Amirteimoori et al. (2023)	A comparison of inferences about the scale characteristics of a manufacturing firm in 48 states in the United States	deterministi Stochastic Descriptive	1- the gross value of production	1-production labor measured in terms of the number of hours worked, 2-non-production labor 3-capital 4- energy 5- and materials
Zhou et al. (2021)	In this application, 16 commercial banks in China are included in evaluation.	deterministic, Stochastic	1-loan, 2-profit	1-employee, 2-fixed assets, 3-and expenses
Tavassoli et al.(2020)	Choosing the most suitable supplier in the supply chain of Sapco Company, a subsidiary of Iran Khodro (the data is related to 2017)	Deterministic Descriptive Stochastic	1-Efficiency of energy consumption 2-Supplier experience 3-Product quality	1-Number of delayed days 2-Offered price from suppliers 3-Shipping cost 4-Total annual cost of electricity 5-Cost of work safety and labor health
Wanke et al.(2020)	An efficiency comparison in OECD banking	Stochastic	1-Net Interest Margin (NIM) – Shared 2-Total Equity 3-Net Income	1-Net Loans 2-Personnel Expenses 3-Total Earning Assets 4-Fixed Assets 5-Loan Loss Reserve 6-Costs
Beraldi and Bruni. (2020)	Numerical examples	deterministic Stochastic	1-earnings before interest, 2- taxes, 3- depreciation and amortization, 4- cash flow	total liabilities
Mehdizadeh et al. (2020)	Performance evaluation of 16 commercial banks in China has been conducted to verify the applicability of the proposed approaches at different levels.	deterministic Stochastic	Stage 1 : 1-product Deposits 2- Interbank Deposits Stage 2: Loan and Prot as - nal outputs	Stage 1 : 1-consume Employee, 2-Fixed assets, 3- Expenses Stage 2: the productions of stage 1 as inputs
Davtalab et al. (2019)	Performance evaluation and ratings of 10 companies from Grundfeld-green	Stochastic	1- gross investment	1-market value of the firm at the end of the previous year, 2- value of the stock of plant and equipment at the end of the previous year
Jradi and Ruggiero. (2019)	Using simulated data, we compare the model to the econometric stochastic frontier model under different distributional assumptions.	Not determined.	Not determined.	.
Hosseini et al. (2019)	productivity evaluation of branches of a university system (An empirical application on education institutes for MPI evaluation with stochastic data)	Stochastic deterministic	1-Number of publications 2- Number of graduate students 3- Earnings (in billion Rials)	1-Number of academic staff 2-Number of non-academic staff for teaching office 3- other costs (in billion Rials) 4- Number of laboratories, studios, and libraries 5- Number of non-academic personnel for graduate office

Chen and Zhu. (2019)	Efficiency measurement in nine airline companies namely Alaska Airlines, Air Canada, Delta, Hawaiian Airlines, Jet Blue, Southwest Airlines, United Continental, Spirit Airlines, and Allegiant.	deterministic Stochastic	1-passenger revenue 2-return on equity	1-operating cost per available seat mile (exclud- ing fuel costs), 2-employees 3- fuel cost per available seat mile,.
Kao and Liu. (2019)	Efficiencies of Taiwanese commercial banks	deterministic Stochastic	1-deposits 2-short-term loans 3- medium-and-long-term loans	1-labor 2-physical capita, 3-and purchased funds
Park et al. (2018)	Evaluating vendor performance in the supply chain	Not determined	Not determined	Not determined
Nasseri et al. (2018)	evaluate the performance some of commercial bank branches in Iran	deterministic Stochastic Descriptive	1-received interest, 2-fee, 3- nonperforming loans (NPAs) (delay in delivering loans and other facilities) as undesirable output.	1-personnel rate (weighted combination of personal qualifications, quantity, education and others), 2-total of deposits (TDs) (of current, short duration and long duration accounts).
Aleskerov et al. (2017)	evaluate the performance the universities	deterministic Stochastic	1- Number of undergraduates 2-Number of taught postgraduates 3-Number of postgraduates 4-Research income 5-Four dummy variables that assess department's performance	1-General expenditures 2-Equipment expenditures 3-Research income
Mitropoulos et al. (2015)	The paper discusses the statistical advantages of this method using cross-sectional data from asample of 117 Greek public hospitals. (An application to Greek public chospitals)	Stochastic deterministic	1-Inpatient admissions, 2-Outpatient visits	1-Doctors, 2-Other personnel, 3-Beds, 4-Operating cost
Jin et al. (2014)	The proposed model has been applied to evaluate the environmental performance of Asia- Pacific Economic Cooperation (APEC) economies in 2010.	Stochastic deterministic	1-GDP (billion 2005 US in PPP) as the desirable output 2-CO2 emissions (Mt) as the undesirable output	1-total energy consumption (Mtoe) 2-labor force (thousand)
Wei et al. (2014)	evaluate the performance of a subset of the selected gas stations in Tokyo	deterministic Stochastic	1-sales of gasoline 2-sales of petrol	1-number of employees, 2- the space size of a gas station, 3-and the monthly operational cost
Wu et al. (2013)	evaluate the Chinese provincial environment efficiency	Stochastic	1-GDP as the desirable output, 2-total industrial emission of waste water, waste gas, as well as waste solids as undesirable output	1-total energy assumption, 2total population

Assaf, A. G. (2012)	measures and compares the efficiency of leading tour operator and hotel companies across several Asia Pacific countries	deterministic Stochastic	total revenues	1-Number of rooms" (proxy for fixed capital), 2-Number of FTE" (full time equivalent employees),3-Other operational costs(administrative costs, costs of utilities and rent)
Hosseinzadeh Lotfi et al. (2012)	car factory which wants to allocate a portion of its producible parts to its ten subject firms	deterministic Stochastic	1-cost of parts, 2-amount of produced parts	1-marginal prices, 2-machinery maintenance expenses
Beraldi and Bruni. (2012)	A sample of 20 Italian firms	deterministic Stochastic	1-earnings before interest, 2-taxes, depreciation and amortization, 3-cash flow	1-liabilities, 2-average duration of accounts receivable
Tsionas and Papadakis. (2010)	efficiency analysis of the Greek banking system for the period 1993–1999.	Stochastic	1-Loans, 2-Investments, 3-liquid assets	1-Labor, 2-Capital, 3-deposits
Khodabakhshi. (2010)	As an empirical example, the proposed method is applied using some actual data of year 2000 to Iranian electricity distribution units.	deterministic Stochastic	1-units of energy delivere 2- number of customers 3- size service area	1-Operating costs 2-number of employees transformer capacity 3-network length
Bruni et al. (2009)	exhaustive and systematic efficiency evaluation of screening units in Italy	deterministic Stochastic	1-number of screening performed (Scr), 2-true positive (TP), 3-true negative (TN)	1-average monthly cost for administrative and medical staff (AvSC), 2-average monthly costs for materials and equipments (AvPC)
Margari et al. (2007)	evaluate the performance of 42 Italian public-owned LPT companies from 1993 to 1999	deterministic Stochastic	1-kilometers supplied, 2-total number of, 3-workers rolling stock size	1-fuel consumption, 2-total operating costs
Tsionas, E. G. (2003)	efficiency measurement in US airlines	deterministic Stochastic		1-Labor, 2-Fuel, 3-Flight equipment, 4-ground property materials
Chen. (2002)	measure the technical efficiency of 39 banks in Taiwan	deterministic Stochastic	1-Loans, 2-Investments, 3-non-interest revenue, 4-interest revenue	1-Labour, 2-Assets, 3-Deposits, 4-Branches
Post. (2001)	Efficiency measurement of 49 large Indonesian banks	deterministic Stochastic	1-total net loans (TNL), 2-interest income, 3-Other income OI.	1-customer deposits CD, 2-interest expenses (IE), 3-other expenses (OE)
Sueyoshi, T. (2000)	This research applies the proposed approach to plan the restructure strategy of a Japanese petroleum company	Stochastic deterministic	1-Gasoline, 2-Petrol	1-No. of employees, 2-Size of gas station, 3-Operation cost

Cooper et al. (1995)	Using data from Chinese sources, this paper reports results from a study of the impact of the 1978 economic reforms for the period 1966-88 on the Textiles, Chemicals and Metallurgical Industries.	deterministic Stochastic	Income(Yuan)	1-Labor, 2-Capital
Land et al. (1993)	analyzed the results from 49 school sites enrolled in the Program Follow Through experiment, and 21 'Non-Follow Through' sites.	Stochastic Descriptive	1-total reading scores, 2-total math scores, 3- total Coopersmith scores (an index of a child self-esteem)	1-education level of mother, 2- parent occupation index, 3-parental visit index, 4- counseling index 5,-number of teachers
Sengupta, J. K. (1987)	Efficiency measurement of public elementary schools in California for the year 1977-1978.	Stochastic Descriptive	1- achievement scores of sixth grade elementary school pupils	1-teacher salary, 2- average class size, 3- proportion of minority students, 4- proxy variable indicating student quality

6 Discovering the Research Gap

In all the Stochastic network DEA models that have been presented so far, the distribution of data in all stages has been considered the same [63][60][62][17][39][26].This is while the distribution of data can be different in each stage. In stochastic models based on statistical simulation for stochastic data generation, the distribution of data is considered the same in each period[7][48][39][17][2].

In terms of the nature of the data type, in some articles, the data are not homogeneous but are considered homogeneous. In other words, there are different types of problem data. In this category of articles, a suitable model is not used to solve the problems. Also, a method for homogenizing the type of uncertainty in the data is not provided.[51][56][33][43][36][3].

In some other articles, several models have been used to cover all types of data[51].

One of the appropriate methods for modeling problems with descriptive data, which have inherent uncertainty, is the theory of possibility. This method has not been used in the reviewed articles.

In none of the articles the type of uncertainty space is not specified. According to the definitions of probability and expected value, it can be said that in problems with a strong uncertainty space, the concept

of probability (P-Model) should be used, and in problems with a weak uncertainty space, the concept of expected value (E-Model) should be used.
In some articles, due to the existence of undesirable outputs, an inappropriate model has been used for the problem. Because by writing a dual model, undesirable outputs appear as model inputs [17][28][36] .
The empty place of the mining process is quite evident in the research done to determine the type of data uncertainty.

7 Conclusion

In this article, stochastic data envelopment analysis models have been investigated.The main purpose of this article is to answer the main questions of the research. In order to answer the research questions, the articles were examined in two functional and structural dimensions.Due to the large volume of articles on the non-deterministic DEA, only articles including stochastic models were reviewed. Stochastic DEA models were divided into four categories:1-Chance Constraint Programming, 2-Fuzzy Stochastic, 3-Statistical Analysis, 4-DEA and SFA
The importance of understanding the real-world problem-solving environment is clear to everyone. Process mining methods can help to understand the environment and produce data appropriate to the problem.In order to compare deterministic and non-deterministic models for a problem in the presence of uncertainty and non-deterministic data, process mining must be performed because by performing the mining process, data corresponding to the uncertainty space of the problem will be produced.After generating data and knowing the type of uncertainty space, the suitable model is selected for the problem. After generating suitable data and solving the problem, the non-deterministic model can be compared with the deterministic model.
After a careful review of the literature, some research gaps were identified. Following research on stochastic data envelopment analysis models, suggestions have been made for future research, including:
In the following, suggestions are provided to solve the research gap discovered in previous studies:

1- In future research in voice, it is possible to use different distributions in the network DEA model.
2- The mining process should be used to better understand the problem space and the type of data.
3- Found an answer to this question: Is there a way to homogenize the uncertainty in the data?
4- It is suggested to use the concepts of possibility theory in problems with descriptive data.

References

[1] F. T. Aleskerov, V. Y. Belousova, and V. V. Petrushchenko, "Models of data envelopment analysis and stochastic frontier analysis in the efficiency assessment of universities,"*Automation and Remote Contro*l, vol. 78, no. 5, pp. 902–923, 2017.

[2] J. A. Al-Khasawneh, N. Essaddam, and T. Hussain, "Total productivity and cost efficiency dynam- ics of us merging banks: A non-parametric bootstrapped analysis of the fifth merger wave," *The Quarterly Review of Economics and Finance*, vol. 78, pp. 199 211, 2020.

[3] A. Amirteimoori, B. K. Sahoo, V. Charles, and S. Mehdizadeh, Stochastic Benchmarking: Theory and Applications, vol. 317. Springer Nature, 2022.

[4] A. G. Assaf, "Benchmarking the asia pacific tourism industry: A bayesian combination of dea and stochastic frontier," *Tourism Management*, vol. 33, no. 5, pp. 1122–1127, 2012.

[5] Banker, R.D., "Stochastic data envelopment analysis," *Working Paper, Carnegie-Mellon University School of Urban and Public Affairs, Pittsburgh*,PA, 1986.

[6] Banker, R.D., " Maximum likelihood, consistency and DEA: Statistical foundations," *Management Science*,39, 1993, 1265-1273.

[7] D. T. Barnum, J. M. Gleason, M. G. Karlaftis, G. T. Schumock, K. L. Shields, S. Tandon, and S. M. Walton, "Estimating dea confidence intervals with statistical panel data analysis," Journal of Applied Statistics, vol. 39, no. 4, pp. 815–828, 2012.

[8] M. Branda and M. Kopa, "On relations between dea-risk models and stochastic dominance efficiency tests," *Central European Journal of Operations Research*, vol. 22, no. 1, pp. 13–35, 2014.

[9] M. Branda, "Stochastic programming problems with generalized integrated chance constraints," *Optimization*, vol. 61, no. 8, pp. 949–968, 2012.

[10] M. Branda, "Sample approximation technique for mixed-integer stochastic programming problems with several chance constraints," *Operations Research Letters*, vol. 40, no. 3, pp. 207–211, 2012.

[11] M. Branda, "Diversification-consistent data envelopment analysis with general deviation measures," *European Journal of Operational Research*, vol. 226, no. 3, pp. 626–635, 2013.

[12] M. Branda and M. Kopa, "Dea models equivalent to general nth order stochastic dominance effi- ciency tests," *Operations Research Letters*, vol. 44, no. 2, pp. 285–289, 2016.

[13] P. Beraldi and M. Bruni, "Efficiency evaluation under uncertainty: a stochastic dea approach," *Decisions in Economics and Finance*, vol. 43, no. 2, pp. 519–538, 2020.

[14] M. Bruni, D. Conforti, P. Beraldi, and E. Tundis, "Probabilistically constrained models for efficiency and dominance in dea," *International Journal of Production Economics*, vol. 117, no. 1, pp. 219–228, 2009.

[15] M. E. Bruni, P. Beraldi, and G. Iazzolino, "Lending decisions under uncertainty: a dea approach," *International Journal of Production Research*, vol. 52, no. 3, pp. 766–775, 2014.

[16] A. Charnes and W. W. Cooper, "Deterministic equivalents for optimizing and satisficing under chance constraints," *Operations research*, vol. 11, no. 1, pp. 18–39, 1963.

[17] Z. Chen, P. Wanke, J. J. M. Antunes, and N. Zhang, "Chinese airline efficiency under co2 emissions and flight delays: A stochastic network dea model," *Energy Economics*, vol. 68, pp. 89–108, 2017.

[18] K. Chen and J. Zhu, "Computational tractability of chance constrained data envelopment analysis," *European Journal of Operational Research*, vol. 274, no. 3, pp. 1037–1046, 2019.

[19] T.y. Chen, "A comparison of chance-constrained dea and stochastic frontier analysis: bank effi- ciency in taiwan," *Journal of the Operational Research Society*, vol. 53, no. 5, pp. 492–500, 2002.

[20] W. W. Cooper, S. Kumbhakar, R. M. Thrall, and X. Yu, "Dea and stochastic frontier analyses of the 1978 chinese economic reforms," *Socio-Economic Planning Sciences*, vol. 29, no. 2, pp. 85–112, 1995.

[21] W. W. Cooper, Z. Huang, and S. X. Li, "Satisficing dea models under chance constraints," *Annals of operations research*, vol. 66, no. 4, pp. 279–295, 1996.

[22] W. W. Cooper, H. Deng, Z. Huang, and S. X. Li, "Chance constrained programming approaches to congestion in stochastic data envelopment analysis," *European Journal of Operational Research*, vol. 155, no. 2, pp. 487 501, 2004.

[23] W. W. Cooper, H. Deng, Z. Huang, and S. X. Li, "Chance constrained programming approaches to technical efficiencies and inefficiencies in stochastic data envelopment analysis," *Journal of the Operational Research Society*, vol. 53, no. 12, pp. 1347–1356, 2002.

[24] M. Davtalab-Olyaie, M. Asgharian, and V. P. Nia, "Stochastic ranking and dominance in dea," *International Journal of Production Economics*, vol. 214, pp. 125–138, 2019.

[25] F. Hosseinzadeh Lotfi, N. Nematollahi, M. Behzadi, M. Mirbolouki, and Z. Moghaddas, "Central- ized resource allocation with stochastic data," *Journal of Computational and Applied Mathematics*, vol. 236, no. 7, pp. 1783–1788, 2012.

[26] S. S. Hosseini, R. KAZEMI MATIN, M. Khunsiavash, and Z. Moghadas, "Measurement of pro- ductivity changes for general network production systems with stochastic data," *S¯adhan¯a*, vol. 44, no. 3, pp. 1–17, 2019.

[27] Z. Huang and S. X. Li, "Stochastic dea models with different types of input-output disturbances," *Journal of Productivity Analysis*, vol. 15, no. 2, pp. 95–113, 2001.

[28] J. Jin, D. Zhou, and P. Zhou, "Measuring environmental performance with stochastic environmental dea: The case of apec economies,"*Economic Modelling*, vol. 38, pp. 80–86, 2014.

[29] S. Jradi and J. Ruggiero, "Stochastic data envelopment analysis: A quantile regression approach to estimate the production frontier," *European Journal of Operational Research*, vol. 278, no. 2, pp. 385–393, 2019.

[30] C. Kao and S.-T. Liu, "Stochastic efficiency measures for production units with correlated data," *European Journal of Operational Research*, vol. 273, no. 1, pp. 278–287, 2019.

[31] C. Kao and S.-T. Liu, "Stochastic data envelopment analysis in measuring the efficiency of taiwan commercial banks,"*European journal of operational Research*, vol. 196, no. 1, pp. 312–322, 2009.

[32] M. Khodabakhshi, "An output oriented super-efficiency measure in stochastic data envelopment analysis: Considering iranian electricity distribution companies," *Computers and Industrial Engineering*, vol. 58, no. 4, pp. 663–671, 2010.

[33] K. C. Land, C. K. Lovell, and S. Thore, "Chance-constrained data envelopment analysis,"*Manage- rial and decision economics*, vol. 14, no. 6, pp. 541–554, 1993.

[34] B. B. Margari, F. Erbetta, C. Petraglia, and M. Piacenza, "Regulatory and environmental effects on public transit efficiency: a mixed dea-sfa approach," *Journal of Regulatory Economics*, vol. 32, no. 2, pp. 131–151, 2007.

[35] S. Mehdizadeh, A. Amirteimoori, M. H. Behzadi, and S. Kordrostami, " Stochastic two-stage network-structures under p-models: A dea based approach," *International Journal of Industrial Mathematics*, vol. 12, no. 3, 2020.

[36] S. H. Nasseri, A. Ebrahimnejad, and O. Gholami, "Fuzzy stochastic data envelopment analysis with undesirable outputs and its application to banking industry," *International journal of fuzzy systems*, vol. 20, no. 2, pp. 534–548, 2018.

[37] O. B. Olesen, "Comparing and combining two approaches for chance constrained dea," *Journal of Productivity Analysis*, vol. 26, no. 2, pp. 103–119, 2006.

[38] O. B. Olesen and N. Petersen, "Chance constrained efficiency evaluation," *Management science*, vol. 41, no. 3, pp. 442–457, 1995.

[39] S. Park, C. Ok, and C. Ha, "A stochastic simulation-based holistic evaluation approach with dea for vendor selection," *Computers and Operations Research*, vol. 100, pp. 368–378, 2018.

[40] Pegoraro, M., Uysal, M. S.,and van der Aalst, W. M., "Conformance checking over uncertain event data," *Information Systems*, 102, 101810, 2021.

[41] Pegoraro, M., Uysal, M. S.,and van der Aalst, W. M., "PROVED: A tool for graph representation and analysis of uncertain event data," *In International Conference on Applications and Theory of Petri Nets and Concurrency* , (pp. 476-486). Cham: Springer International Publishing, (2021, June).

[42] T. Post, "Performance evaluation in stochastic environments using mean-variance data envelopment analysis," *Operations Research*, vol. 49, no. 2, pp. 281–292, 2001.

[43] J. K. Sengupta, "Data envelopment analysis for efficiency measurement in the stochastic case," *Computers and operations research*, vol. 14, no. 2, pp. 117–129, 1987.

[44] J. K. Sengupta, "Efficiency measurement in stochastic input-output systems," *International Journal of Systems Science*, 13(3), 273-287.

[45] A. Schaefer, M. Cassidy, K. Marshall, and J. Rossi, "Internal audits and executive education: A holy alliance to reduce theft and misreporting," *Employee Relations Law Journal*, vol. 32, no. 1, pp. 61–84, 2006.

[46] L. Simar and V. Zelenyuk, "Stochastic fdh/dea estimators for frontier analysis,"*Journal of Produc- tivity Analysis*, vol. 36, no. 1, pp. 1–20, 2011.

[47] L. Simar, "How to improve the performances of dea/fdh estimators in the presence of noise?," *Journal of Productivity Analysis*, vol. 28, no. 3, pp. 183–201, 2007.

[48] Z. Sinuany-Stern and L. Friedman, "Statistical analysis in the dea context," in 2016 Second Inter- national Symposium on Stochastic Models in Reliability Engineering,*Life Science and Operations Management (SMRLO)*, pp. 469–474, IEEE, 2016.

[49] T. Sueyoshi, "Stochastic dea for restructure strategy: an application to a japanese petroleum com- pany," *Omega*, vol. 28, no. 4, pp. 385–398, 2000.

[50] M. Tavana, R. Khanjani Shiraz, and A. Hatami-Marbini, "A new chance-constrained dea model with birandom input and output data," *Journal of the operational research society*, vol. 65, no. 12, pp. 1824–1839, 2014.

[51] M. Tavassoli, R. F. Saen, and D. M. Zanjirani, "Assessing sustainability of suppliers: A novel stochastic-fuzzy dea model," *Sustainable production and consumption*, vol. 21, pp. 78–91, 2020.

[52] E. G. Tsionas and E. N. Papadakis, "A bayesian approach to statistical inference in stochastic dea," *Omega*, vol. 38, no. 5, pp. 309–314, 2010.

[53] E. G. Tsionas, "Combining dea and stochastic frontier models: An empirical bayes approach," *European Journal of Operational Research*, vol. 147, no. 3, pp. 499–510, 2003.

[54] J. Von Neumann and O. Morgenstern, "Theory of games and economic behavior.(1967 edition) john wiley and sons," *Inc., New York, New York*, vol. 641, 1944.

[55] C. Wu, Y. Li, Q. Liu, and K. Wang, "A stochastic dea model considering undesirable outputs with weak disposability," *Mathematical and computer modelling*, vol. 58, no. 5-6, pp. 980–989, 2013.

[56] D. D. Wu and D. Olson, "Enterprise risk management: a dea var approach in vendor selection," *International Journal of Production Research*, vol. 48, no. 16, pp. 4919–4932, 2010.

[57] D. D. Wu and C.-G. Lee, "Stochastic dea with ordinal data applied to a multi-attribute pricing problem,"*European Journal of Operational Research*, vol. 207, no. 3, pp. 1679–1688, 2010.

[58] G. Wei, J. Chen, and J. Wang, "Stochastic efficiency analysis with a reliability consideration," *Omega*, vol. 48, pp. 1–9, 2014.

[59] M. Wen, T. Zu, M. Guo, R. Kang, and Y. Yang, "Optimization of spare parts varieties based on stochastic dea model," *IEEE Access*, vol. 6, pp. 22174–22183, 2018.

[60] P. Wanke, M. G. Tsionas, Z. Chen, and J. J. M. Antunes, "Dynamic network dea and sfa models for accounting and financial indicators with an analysis of super-efficiency in stochastic frontiers: An efficiency comparison in oecd banking," *International Review of Economics and Finance*, vol. 69, pp. 456–468, 2020.

[61] H. Yang, D. Shao, B. Liu, J. Huang, and X. Ye, "Multi-point source identification of sudden water pollution accidents in surface waters based on differential evolution and metropolis–hastings– markov chain monte carlo," *Stochastic Environmental Research and Risk Assessment*, vol. 30, no. 2, pp. 507–522, 2016.

[62] Z. Zhou, L. Lin, H. Xiao, C. Ma, and S. Wu, "Stochastic network dea models for two-stage sys- tems under the centralized control organization mechanism," *Computers and Industrial Engineering*, vol. 110, pp. 404–412, 2017.

[63] Z. Zhou, W. Sun, H. Xiao, Q. Jin, and W. Liu, "Stochastic leader–follower dea models for two-stage systems," *Journal of Management Science and Engineering*, vol. 6, no. 4, pp. 413–434, 2021.

Sadegh Farzam
PhD candidate of applied mathematics
Department of Mathematics
Shahid Rajaee Teacher Training University
Tehran, Iran
s.farzam@sru.ac.ir

Farhad Hosseinzadeh Lotfi
Professor of Mathematics
Department of Mathematics
Science and Research Branch, Islamic Azad University
Tehran, Iran
farhad@hosseinzadeh.ir

Hamid Mesgarani
Associate Professor of Mathematics
Department of Mathematics
Shahid Rajaee Teacher Training University
Tehran, Iran
hmesgarani@sru.ac.ir