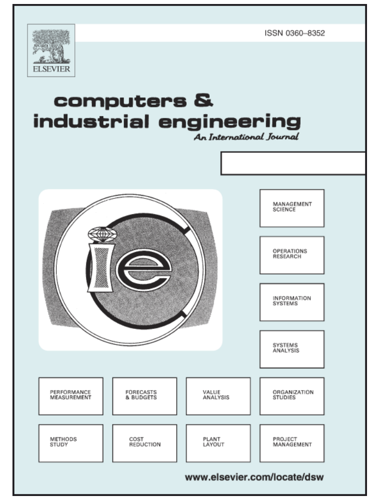
Journal Pre-proofs



A data envelopment analysis model for performance evaluation and ranking of DMUs with alternative scenarios

Mazyar Zahedi-Seresht, Shahrzad Khosravi, Josef Jablonsky, Petra Zykova

PII: S0360-8352(20)30672-0
DOI: https://doi.org/10.1016/j.cie.2020.107002
Reference: CAIE 107002

To appear in: *Computers & Industrial Engineering*

Received Date: 12 December 2018
Revised Date: 17 August 2020
Accepted Date: 24 November 2020

Please cite this article as: Zahedi-Seresht, M., Khosravi, S., Jablonsky, J., Zykova, P., A data envelopment analysis model for performance evaluation and ranking of DMUs with alternative scenarios, *Computers & Industrial Engineering* (2020), doi: https://doi.org/10.1016/j.cie.2020.107002

A data envelopment analysis model for performance evaluation and ranking of DMUs with alternative scenarios

Mazyar Zahedi-Seresht
Department of Econometrics, University of Economics, Prague, Czech Republic
email: zahedi_s@yahoo.com
Shahrzad Khosravi
Department of Econometrics, University of Economics, Prague, Czech Republic
email: khosravi_sh@ie.sharif.edu
Josef Jablonsky[*]
Department of Econometrics, University of Economics, Prague, Czech Republic
email: jablon@vse.cz
Petra Zykova
Department of Econometrics, University of Economics, Prague, Czech Republic
email: petra.zykova@vse.cz

[*] Corresponding author

A data envelopment analysis model for performance evaluation and ranking of DMUs with alternative scenarios

Abstract:

Data envelopment analysis (DEA) is a general tool for measuring the relative efficiency of homogeneous decision-making units (DMUs). DEA models usually deal with crisp data and do not consider the conditions in which the inputs and outputs are uncertain. Many researchers have focused their research on these types of conditions, in which they assumed fuzzy data, interval data, and probabilistic data, as well as other expressions of uncertainty in the dataset. Various models, such as mean value and variance, robust DEA, multiple criteria decision-making (MCDM) models, and several other models, have been proposed. This paper deals with instances in which uncertainty in the dataset is expressed by several alternative scenarios. The first presented model for problems with several alternative scenarios in their inputs and outputs is derived directly from the definition of the relative efficiency formula similar as those in traditional DEA models. This model is not linear and cannot be linearized. Due to this, we modify this model and derive a new model that is linear and can be solved easily. The proposed models have none of the common drawbacks attending other methods commonly applied to this set of issues. They are always feasible; moreover, they are able to generate a complete ranking of all DMUs using a computationally efficient procedure. Both models are illustrated using a numerical example with 10 DMUs and three scenarios for input and output values, and their results are compared and discussed.

Keywords: data envelopment analysis, efficiency, ranking, scenario-based data

1. Introduction

Data envelopment analysis (DEA) is a method for evaluating efficiency and performance of a set of homogenous decision-making units (DMUs) by solving linear programming problems. DEA models evaluate the relative efficiency of a set of DMUs that use multiple resources (inputs) for the production of multiple effects (outputs). The first DEA model was proposed by Charnes et al. (1978); since then, it has been developed by many other researchers. This model – often denoted as the Charnes, Cooper and Rhodes (CCR) model – is a generalisation of a single-input and single-output concept of efficiency, as introduced by Farell (1957). CCR models assume constant returns to scale in the definition of the production possibility set. Its generalization was proposed by Banker et al. (1984). The Banker, Charnes and Cooper (BCC) model considers variable returns to scale. Decision-makers must decide on the assumption of constant or variable returns to scale in the applications of this kind of DEA models and choose the most appropriate model. Both CCR and BCC models assign efficiency scores equal to 1 to the units expressing the efficient frontier, as defined by the model. The other inefficient units have efficiency scores lower than 1. The efficiency score allows ranking of inefficient DMUs, however, the efficient ones cannot be ranked due to their maximum identical score. As such, various methods and models have been proposed in the past in order to allow the ranking of originally efficient units. Many researchers have developed the DEA theory since the mentioned pioneering works. More information about progress in the theory and its practice can be found in Cooper et al. (1999), Seiford (1990), Seiford and Thrall (1990) and other publications.

Traditional DEA models, such as the CCR and BCC models, assume an expression of the dataset as crisp values, however, the observed values of inputs and outputs in real problems are often uncertain, fuzzy, interval-based, or scenario-based. To study this issue, various approaches have been proposed in the literature. O'Neal et al. (2002) suggested excluding those DMU(s) with missing or vague data to calculate efficiency. This approach affects the relative efficiency score of other DMUs of the set. That is why it is not acceptable for general applications. Several approaches to overcome the problem of missing or vague data use various imputation techniques, e.g. data averaging for the DMUs. Using this simplification in DEA models may cause misleading results since the efficiency scores of other DMUs of the set may be significantly over- or underestimated. Thus, it does not seem to be a reliable tool for the application of DEA models.

A stochastic approach has also been applied for uncertain data in DEA. Stochastic programming has undergone significant theoretical developments since the 1950s, starting with the pioneering works of Dantzig (1955) and Beale (1955). Nevertheless, this approach also suffers from its drawbacks. One of them concerns the determination of the probability distribution function in the absence of reliable empirical evidence. Another approach, proposed by Kuosmanen (2009), applies dummy variables instead of missing data. The missing output values are set to zero and the missing input data are considered as relatively large numbers. Two other approaches to vague data use the fuzzy methodology and interval data. They were developed by Sengupta (1992) and Cooper et al. (1999), respectively. Cooper et al. (1999) developed an interval method by converting the DEA model to the form of linear programming, allowing the combination of crisp and non-crisp data. Being able to assess the upper and lower bounds of the relative efficiency of DMUs is one of the difficulties that arises in applying the interval method. Despite this difficulty, some researchers (Despotis & Smirlis,2002; Entani et al., 2002; Kao, 2006; Kao & Liu, 2000; Smirlis et al., 2006; Wang & Yang, 2005) have developed numerous interval methods.

In fuzzy approaches, several mathematical programming methods, including probabilistic planning and α-cut approaches, are applied to evaluate the relative efficiency of DMUs. Although there are circumstances where there is a sharp growth in the complexity of the fuzzy method, many researchers (Guo & Tanaka, 2001; Hatami-Marbini & Saati, 2009; Jahanshahloo et al. 2009; Le, 2003; Lertworasirikul et al., 2003; Liu, 2008; Liu & Chuang, 2009; Saati et al., 2002; Soleimani-Damaneh, 2009; Soleimani-Damaneh et al., 2006; and Wang et al., 2009) have studied fuzzy DEA and proposed their own models.

A new cost-efficiency data envelopment analysis (CE-DEA) approach with price uncertainty and an assumption that only the upper and lower bounds of input prices for all units are available has been developed by Toloo and Ertay (2014). Sadjadi and Omrani (2008) proposed a robust DEA model considering uncertainty in outputs and applying the model for performance evaluation of electricity distribution companies. Shokouhi et al. (2010) introduced an approach based on a robust optimisation model, in which input and output parameters are constrained to be within an uncertainty set under the assumption of a worst case efficiency (as defined by the uncertainty set), and supporting constraints. Hafezalkotob et al. (2015) proposed a robust DEA model to evaluate the Iranian electricity companies by using discrete robust optimisation approaches as introduced by Mulvey et al. (1995). Their model uses probable scenarios to capture the effect of ambiguous data in the presented case study. A possible infeasibility is the most significant problem attending this model. Arabmaldar et al. (2017) published a new framework based on the robust optimisation approach. According to their construct, evaluation of the efficiency of DMUs – in the presence of uncertainty – has been accomplished in fewer steps than when applying other models. They introduced two linear robust super-efficiency models based on the Andersen and Petersen (1993) model and verified the proposed approach on a real dataset. Wu et al. (2017) transformed the DEA optimisation model into a robust second-order cone – the effect was equivalent to immunising it against output perturbation in an uncertainty set. Zahedi-Seresht et al. (2016) formulated a robust DEA model that allows complete ranking of all DMUs with S alternative scenarios. The main disadvantage of their model is its computational complexity. It does not allow solving problems with higher numbers of DMUs and/or scenarios. Vaez-Ghasemi (2016) analysed the problem of supplier selection under the assumption of different scenarios. His study is one example of a real-world application of the DEA models with alternative scenarios. Zahedi-Seresht et al. (2017) formulated an original method to rank DMUs with alternative scenarios based on the Monte Carlo simulation and Mulvey's model.

From the overview above, only a few studies following dealt with the analysis of efficiency in cases where the DMUs are described by several alternative scenarios. This case is easiest for decision-makers to understand. The assumption of alternative scenarios can widely be used in practice, e.g. when the decision-maker is able to estimate and describe a future development of the system under evaluation by pessimistic, most likely, and optimistic values of variables (inputs and/or outputs). These three-point estimates are often used in practice. A good example is the PERT method (that works with three-time estimates for lengths of the activities in critical path analysis). Three-point estimates can be used in general, as approximations of continuous probabilistic distributions as discussed in detail in Keefer and Bodily (1984). In this paper, we assume a general case with S alternative scenarios. Performance analysis with the uncertainty defined by S alternative scenarios can find many real-world applications. They can assume several controllable crisp inputs and several uncertain input variables are defined using alternative scenarios. Future outputs of the DMUs are uncertain and described as discrete alternative scenarios that occur with certain probabilities. The following two examples illustrate the possible cases:

1. Performance evaluation of departments (DMUs) within a university. The number of employees may be considered as a crisp input. Alternative scenarios may describe

operating costs as inputs of the department, and the number of graduated and number of publications in journals as outputs.

2. Performance evaluation of a branch network of a bank (branches of the bank are DMUs). The number of employees of the branch and the number of accounts may be defined as crisp inputs. The number of transactions and the various financial indicators of the branch may be uncertain and described by alternative scenarios.

As discussed, many approaches were proposed for dealing with uncertainty in DEA models (interval data, fuzzy inputs/outputs, stochastic DEA, etc.) but almost none of them dealt with the assumption of alternative scenarios. The method introduced by Zahedi-Seresht et al. (2017) is quite complicated and computationally complex. It is a robust DEA model that combines optimization and the Monte Carlo approach to derive robust ranking of all DMUs under the assumption of alternative scenarios. The combination of these two approaches almost renders the model inapplicable for non-expert users. The aim of this paper is to fill a research gap in this area and propose a computationally efficient algorithm for evaluating DMUs under this assumption. The advantage of the approach proposed in this paper compared with other non-crisp models lies in its computational easiness (which allows solving instances with a high number of DMUs and variables). The computational easiness of our model is given by the necessity to solve just one linear program for each DMU to derive efficiency and/or super-efficiency scores.

The paper is organised as follows. The next section provides a brief formulation of traditional DEA models, as used further in this paper. Section 3 presents the mathematical details of the proposed model for ranking of DMUs in the case of alternative scenarios in inputs and/or outputs. Section 4 consists of a numerical illustration of the model's advantages, and the last section of our paper summarises the main points and findings of the study and discusses recommendations for future research.

2. Traditional DEA models

DEA is a mathematical programming method for evaluating and measuring the performance of DMUs. Let us consider n DMUs with nonnegative row vectors $x^j = \left(x_{1j}, x_{2j}, \ldots, x_{sj}\right)^{\mathsf{T}}$ and $y^j = \left(y_{1j}, y_{2j}, \ldots, y_{rj}\right)^{\mathsf{T}}$, $j = 1, \ldots, n$, of inputs and outputs respectively, where x_{ij} represents the value of the i-th input and y_{kj} is the value of the k-th output for DMU$_j$, $j=1,\dots,n$. The efficiency score of the DMU$_p$ can be expressed as the ratio

$$e_p(u,v)=\frac{u^{\mathbf{T}}y^p}{v^{\mathbf{T}}x^p},$$

i.e. the weighted sum of outputs divided by the weighted sum of inputs with positive column vectors of weights u and v, respectively. Podinovski and Athanassopoulos (1998) and Podinovski (2001) formulated the relative measure of the efficiency of the DMU$_p$ $f_p(u,v)$ as follows:

$$\begin{aligned}&f_p(u,v)=\frac{e_p(u,v)}{\max\limits_j\{e_j(u,v)\}},\\&u,v\geq\varepsilon,\ j=1,\dots,n,\end{aligned}\tag{1}$$

where ε is an infinitesimal constant. Obviously $f_p(u,v)$ does not exceed one for any DMU. It is possible to show that the optimal solution for the model (1) is identical to the CCR DEA model, introduced by Charnes et al. (1978). The fractional form of this model is formulated as follows:

$$\begin{aligned}\text{Maximize}\quad & f_p(u,v)=\frac{u^{\mathbf{T}}y^p}{v^{\mathbf{T}}x^p}\\ \text{subject to}\quad & \frac{u^{\mathbf{T}}y^j}{v^{\mathbf{T}}x^j}\leq 1,\ j=1,\dots,n,\\ & u\geq\varepsilon, v\geq\varepsilon.\end{aligned}\tag{2}$$

Model (2) is not linear in its objective function but can be easily linearized using Charnes-Cooper transformation. The input-oriented formulation of this model is as follows:

$$\begin{aligned}\text{Maximize}\quad & f_p(u,v)=u^{\mathbf{T}}y^p\\ \text{subject to}\quad & u^{\mathbf{T}}y^j-v^{\mathbf{T}}x^j\leq 0,\ j=1,\dots,n,\\ & v^{\mathbf{T}}x^p=1,\\ & u\geq\varepsilon, v\geq\varepsilon.\end{aligned}\tag{3}$$

Models (2) and (3) assume the availability of a crisp input and output dataset. The following section generalises these models for the case of several scenarios.

3. A DEA model with alternative scenarios

Let us consider the set of n DMUs with s inputs and r outputs. Instead of crisp data, which is not available, the decision-maker has several alternative scenarios for the inputs and outputs, and is able to estimate their probabilities of occurrence. Table 1 shows the complete dataset, with n DMUs and S scenarios. Each scenario has the probability of its occurrence p_i, $i = 1,\ldots, S$, $\Sigma_i\, p_i = 1$. Let us denote vectors of inputs and outputs for the DMU$_j$ and the i-th scenario as x_i^j and y_i^j, $j = 1,\ldots,n$, $i = 1,\ldots,S$, respectively.

Table 1: *A set of n DMUs with s inputs, r outputs and S scenarios*

	Scenarios	I_1	...	I_s	O_1	...	O_r
DMU$_1$	1	x_{11}^1		x_{s1}^1	y_{11}^1		y_{r1}^1
	…						
	S	x_{1S}^1		x_{sS}^1	y_{1S}^1		y_{rS}^1
DMU$_2$	1	x_{11}^2		x_{s1}^2	y_{11}^2		y_{r1}^2
	…						
	S	x_{1S}^2		x_{sS}^2	y_{1S}^2		y_{rS}^2
...							
DMU$_n$	1	x_{11}^n		x_{s1}^n	y_{11}^n		y_{r1}^n
	…						
	S	x_{1S}^n		x_{sS}^n	y_{1S}^n		y_{rS}^n

Model (1) for evaluation of DMU$_p$ with S alternative scenarios can be re-written as follows:

$$\text{Maximize} \quad f_p(u_i, v_i) = \frac{p_1 \dfrac{u_1 y_1^p}{v_1 x_1^p} + \ldots + p_S \dfrac{u_S y_S^p}{v_S x_S^p}}{\max\limits_j \left\{ p_1 \dfrac{u_1 y_1^j}{v_1 x_1^j} + \ldots + p_S \dfrac{u_S y_S^j}{v_S x_S^j} \right\}}, \tag{4}$$
$$u_i \geq \varepsilon, v_i \geq \varepsilon,\ i = 1,\ldots,S.$$

where u_i, $i = 1,\ldots, S$, is the row weight vector attached to the outputs in the i-th scenario, and v_i, $i = 1,\ldots, S$, is the row weight vector attached to the inputs in the i-th scenario. Model (4) can be reformulated based on model (2) as follows:

$$
\begin{aligned}
&\text{Maximize} \quad && f_p\left(u_i, v_i\right) = p_1 \frac{u_1 y_1^p}{v_1 x_1^p} + \ldots + p_S \frac{u_S y_S^p}{v_S x_S^p}, \\
&\text{subject to} \quad && p_1 \frac{u_1 y_1^j}{v_1 x_1^j} + \ldots + p_S \frac{u_S y_S^j}{v_S x_S^j} \le 1, \qquad j = 1, \ldots, n, \\
& && u_i \ge \varepsilon, v_i \ge \varepsilon, \; i = 1, \ldots, S.
\end{aligned} \tag{5}
$$

We will move model (5) using Charnes-Cooper transformation to a model with a linear objective function. Let us put $q_i = \dfrac{1}{v_i x_i^p} > 0$, $i = 1, \ldots, S$, and multiply the nominator and denominator of all ratios in the constraints by this variable:

$$
\begin{aligned}
&\text{Maximize} \quad && f_p\left(u_i, v_i\right) = p_1 q_1 u_1 y_1^p + \ldots + p_S q_S u_S y_S^p, \\
&\text{subject to} \quad && \\
& && p_1 \frac{q_1 u_1 y_1^j}{q_1 v_1 x_1^j} + \ldots + p_S \frac{q_S u_S y_S^j}{q_S v_S x_S^j} \le 1, \; j = 1, \ldots, n, \\
& && q_i v_i x_i^p = 1, \; i = 1, \ldots, S, \\
& && u_i', v_i \ge \varepsilon.
\end{aligned} \tag{6}
$$

Now we put $u_i' = q_i u_i$ and $v_i' = q_i v_i$, $i = 1, \ldots, S$. The final version of the model for efficiency evaluation of the DMUs with S alternative scenarios is as follows:

$$
\begin{aligned}
&\text{Maximize} \quad && f_p\left(u_i', v_i'\right) = p_1 u_1' y_1^p + \ldots + p_S u_S' y_S^p, \\
&\text{subject to} \quad && \\
& && p_1 \frac{u_1' y_1^j}{v_1' x_1^j} + \ldots + p_S \frac{u_S' y_S^j}{v_S' x_S^j} \le 1, \qquad j = 1, \ldots, n, \\
& && v_i' x_i^p = 1, \; i = 1, \ldots, S, \\
& && u_i' \ge \varepsilon, v_i' \ge \varepsilon.
\end{aligned} \tag{7}
$$

Model (7) can be re-written in a more compact way as model (8):

$$
\begin{aligned}
&\text{Maximize} \quad && f_p\left(u_i', v_i'\right) = \sum_{i=1}^{S} p_i u_i' y_i^p, \\
&\text{subject to} \quad && \\
& && \sum_{i=1}^{S} p_i \frac{u_i' y_i^j}{v_i' x_i^j} \le 1, \qquad j = 1, \ldots, n, \\
& && v_i' x_i^p = 1, \qquad i = 1, \ldots, S, \\
& && u_i' \ge \varepsilon, v_i' \ge \varepsilon.
\end{aligned} \tag{8}
$$

The following properties of model (8) are obvious and need not be proved:

a) Model (8) is always feasible.

b) The optimal objective function of model (8) is lower than or equal to 1.

c) At least one of the DMUs under evaluation is efficient, i.e. at least for one DMU, the optimal objective function of model (8) equals 1.

d) Model (8) has the same optimal solution as model (4).

An important property of model (8) that express the relation between the results of traditional CCR models and the results of model (8) is formulated in the following theorem.

Theorem 1. If a unit under evaluation is efficient in one of the scenarios, then it is always efficient in model (8).

Proof.

Let us consider DMU$_p$ and (u_q^*, v_q^*) the optimal solution for model (3) for its q-th scenario. Assume that DMU$_p$ is efficient in this scenario, i.e.

$$\frac{u_q^* y^p}{v_q^* x^p} = 1, \qquad \text{and} \qquad \frac{u_q^* y^j}{v_q^* x^j} \le 1, \qquad j = 1, \ldots, n.$$

Let $u_q'^* = \dfrac{u_q^*}{p_q}, v_q'^* = v_q^*$, where p_q is the probability of the q-th scenario. Then the following relations hold

$$f_p = \left(\sum_{i=1, i \ne q}^{S} p_i u_i' y_i^p \right) + p_q u_q'^* y_q^p = \left(\sum_{i=1, i \ne q}^{S} p_i u_i' y_i^p \right) + 1,$$

$$p_q \frac{u_q'^* y_q^j}{v_q'^* x_q^j} \le 1, \quad j = 1, \ldots, n,$$

and at least one of these relations is binding. For $\varepsilon = 0$, all weights $u_i' = 0, i \ne q$, and the optimal objective function $f_p = 1$.

For $\varepsilon > 0$, assume that $u_q'^*$ variables are reduced to their new values $u_q'^{**}$ and this change leads to a reduction of $p_q u_q'^* y_q^p$ by Δ, i.e. $p_q u_q'^* y_q^p = p_q u_q'^{**} y_q^p + \Delta$. To keep feasibility and maximum efficiency, the other weights of model (8) must fulfill the following relations:

$$\sum_{i=1, i \ne q}^{S} p_i u_i'^* y_i^p = \Delta,$$

$$\sum_{i=1, i \ne q}^{S} p_i \frac{u_i'^* y_i^j}{v_i'^* x_i^j} \le \Delta, \qquad j = 1, \ldots, n.$$

It is clear this set of constraints has a feasible solution considering an appropriate setting of the Δ value and ε values in model (8).

By this, it is proven that the maximum efficiency way can be reached in model (8) for the case where DMU_p is CCR efficient in at least one of the scenarios independently on the probabilities of the scenarios.

Model (8) is linear in its objective function but has non-linear constraints and is rather hard to solve. Unfortunately, it is not possible to find a linearized problem for model (8). However, we can formulate a new linear model with the constraints that are derived from model (8). We suggest re-writing the first set of constraints in model (8) as follows:

$$p_i \frac{u_i' y_i^j}{v_i' x_i^j} \le 1, \qquad j = 1,\dots,n, i = 1,\dots,S, \tag{9}$$

Relation (9) can be expressed as (10):

$$p_i \frac{u_i' y_i^j}{v_i' x_i^j} + \delta_i^j = 1, \qquad j = 1,\dots,n, i = 1,\dots,S, \tag{10}$$

where $1 - p_i \le \delta_i^j \le 1, i = 1,\dots,S.$ In order to transform it into a linear form, we can write:

$$p_i \frac{u_i' y_i^j}{v_i' x_i^j} + \frac{\Delta_i^j}{v_i' x_i^j} = 1, \qquad j = 1,\dots,n, i = 1,\dots,S, \tag{11}$$

and then receive the following linearized version:

$$p_i u_i' y_i^j - v_i' x_i^j + \Delta_i^j = 0, \qquad j = 1,\dots,n, i = 1,\dots,S, \tag{12}$$

To summarise, the complete linear model is as follows:

$$\begin{aligned}
&\text{Maximize} \quad F_p\left(u_i', v_i'\right) = \sum_{i=1}^{S} p_i u_i' y_i^p, \\
&\text{subject to} \\
&\qquad p_i u_i' y_i^j - v_i' x_i^j + \Delta_i^j = 0, \qquad j = 1,\dots,n, i = 1,\dots,S, \\
&\qquad v_i' x_i^p = 1, \qquad i = 1,\dots,S, \\
&\qquad v_i' x_i^j - p_i v_i' x_i^j \le \Delta_i^j \le v_i' x_i^j, \qquad j = 1,\dots,n, i = 1,\dots,S, \\
&\qquad u_i' \ge \varepsilon, v_i' \ge \varepsilon.
\end{aligned} \tag{13}$$

Theorem 2. The optimal solution for model (13) is a feasible solution for model (8).

Proof.

Let us consider that $(u_i'^*, v_i'^*)$ is the optimal solution for model (13). Then the following relations hold:

$$v_i'^* x_i^j - p_i v_i'^* x_i^j \leq \Delta_i^{j*} \leq v_i'^* x_i^j \Rightarrow$$
$$v_i'^* x_i^j - p_i v_i'^* x_i^j \leq -p_i u_i'^* y_i^j + v_i'^* x_i^j \leq v_i'^* x_i^j \Rightarrow$$
$$-p_i v_i'^* x_i^j \leq -p_i u_i'^* y_i^j \leq 0 \Rightarrow$$
$$v_i'^* x_i^j \geq u_i'^* y_i^j \geq 0 \Rightarrow$$
$$\frac{u_i'^* y_i^j}{v_i'^* x_i^j} \leq 1 \Rightarrow$$
$$p_i \frac{u_i'^* y_i^j}{v_i'^* x_i^j} \leq p_i \Rightarrow$$
$$\sum_{i=1}^{S} p_i \frac{u_i'^* y_i^j}{v_i'^* x_i^j} \leq 1, \qquad j = 1, \ldots, n.$$

By this, the proof is completed.

It is clear, that the set of constraints in model (13) is stronger than the set of constraints in model (8). It means that any feasible solution for model (13) is also a feasible solution for model (8) but not vice versa, i.e. a feasible solution within model (8) need not be feasible in model (13). This relation was proved above. The properties of model (13) are as follows:

a) The optimal objective function value (efficiency score) $f_p^*(u_i', v_i')$ of model (8) is always greater than or equal to the optimal value of model (13) $F_p^*(u_i', v_i')$.
b) The efficiency score of the unit under evaluation derived by model (13) is greater than or equal to the worse efficiency score of this unit over all scenarios.
c) The efficiency score of the unit under evaluation derived by model (13) is lower than or equal to the best efficiency score of this unit over all scenarios.
d) If a unit under evaluation is efficient in model (3) in all scenarios, then it is efficient in model (13) also and vice versa.

In order to find an optimal efficiency score $f_p^*(u_i', v_i')$, i.e. the solution of model (8), it is necessary to solve this non-linear model using a global non-linear solver, and even if it is available, it is hard to obtain an optimal solution for a higher number of scenarios. The optimal solution of model (13) can be explained better to decision-makers because, depending on the probabilities of occurrences of the scenarios, the final efficiency score $F_p^*(u_i', v_i')$ is always between the worst and the best efficiency scores of all the alternative scenarios. This efficiency score can be regarded as a good representation of the performance of the DMU under evaluation. Decision-makers consider both the worst scenario (with the worse efficiency score) and the best scenario (with the best efficiency score) that can occur with certain probabilities.

It is clear, the decision-maker will expect that the efficiency score of the aggregated DMU to be between these two extreme values. Our model (13) fulfils this expectation.

An important conclusion of our study consists in the fact that for the identification of efficient units in model (8), it must be possible to solve linear programs (13) which is computationally efficient. It is expressed by the following corollary of Theorem 2.

Corollary 1. If a DMU under evaluation is efficient in model (13) then it is efficient in model (8) also.

Proof.

The optimal objective function $f_p^*\left(u_i', v_i'\right)$ of model (8) is greater than or equal to the optimal objective function $F_p^*\left(u_i', v_i'\right)$ of model (13). i.e. if $F_p^*\left(u_i', v_i'\right) = 1$ then $f_p^*\left(u_i', v_i'\right) = 1$, and the unit DMU_p is efficient in both models.

The number of efficient DMUs identified by model (13), i.e. the DMUs with $F_p^*\left(u_i', v_i'\right)$ = 1, may be higher than 1. To rank the efficient units, we propose a super-efficiency model derived from that suggested by Andersen and Petersen (1993). This model removes the unit under evaluation from the set of DMUs, then its super-efficiency score is greater than 1 (which allows the complete ranking of all units). The mathematical formulation of this super-efficiency model is as follows:

$$\begin{aligned} &\text{Maximize} && F_p\left(u_i', v_i'\right) = \sum_{i=1}^{S} p_i u_i' y_i^p, \\ &\text{subject to} && \\ & && p_i u_i' y_i^j - v_i' x_i^j + \Delta_i^j = 0, && j = 1,\ldots,n, i = 1,\ldots,S, j \neq p, \\ & && v_i' x_i^p = 1, && i = 1,\ldots,S, \\ & && (1 - p_i) v_i' x_i^j \le \Delta_i^j \le v_i' x_i^j, && j = 1,\ldots,n, i = 1,\ldots,S, \\ & && u_i' \ge \varepsilon, v_i' \ge \varepsilon. && \end{aligned} \tag{14}$$

There have been a few studies published that deal with efficiency analysis in the case of alternative scenarios. The proposed approaches in these studies have various drawbacks that do not allow their broader applications. In this study, a new model, formulated as an alternative to solution of nonlinear model (8), is used as a computationally friendly way of deriving efficiency and/or super-efficiency scores of DMUs.

The presented models eliminate most drawbacks of the previous models and methods dealing with alternative scenarios in efficiency analyses – Zahedi-Seresht et al. (2016, 2017). The most significant among them can be summarized as follows:

1. *Infeasibility*. One of the major problems of previous approaches with multi-scenario inputs and outputs is the possible infeasibility of the models (see Zahedi-Seresht et al., 2016). Model (13) is always feasible.
2. *Computational complexity*. The presented linear model (13) has only $S(2n+1)$ constraints, where n is the number of DMUs and S the number of scenarios. The number of variables $(S \cdot n+s+r)$ is quite low also (s and r are the number of inputs and outputs, respectively). In contrast, the time complexity of the algorithm introduced in Zahedi-Seresht et al. (2016) is $O\left(S\left(s+r\right)^3 n\right)$, which is much worse.
3. *Complete ranking of the DMUs*. The models introduced in previous studies can offer a complete ranking of the DMUs, but the efficiency score of the best DMUs is not necessarily equal to 1 (Zahedi- Seresht et al., 2017). That is why they cannot distinguish among the efficient and inefficient DMUs. In the proposed models (8) and (13), at least one of the DMUs is efficient, i.e. its efficiency score is equal to 1. Moreover, a super-efficiency model (14) has been proposed to rank the originally efficient units.
4. Podinovski et al. (2017) proposed nonparametric production technology with multiple technologies as different scenarios. They worked on multi-hybrid returns to scale (MHRS) as a new model, based on returns-to-scale technology. However, our non-linear and linear models can be modified for any type of technology.

4. A numerical example

Zahedi-Seresht et al. (2017) studied robust DEA with alternative scenarios and introduced a case study from a large engineering company. We use the same dataset for illustrative purposes and comparison of results for the method presented in this paper. The company has 10 branches (DMUs); two inputs and one output are considered:

- Input 1 - Engineering manpower measured by the number of employees.
- Input 2 - Tangible and intangible assets in millions of LCY (local currency).
- Output - The volume of invoices in local or foreign currencies (Payments) received in millions of LCY.

Table 2: *Dataset for all scenarios*

	Manpower	Assets	Output - Payments		
	Input 1	Input 2	Scenario 1	Scenario 2	Scenario 3
DMU_1	606	293	3054	2974	2455
DMU_2	797	569	897	948	862
DMU_3	247	614	777	836	760
DMU_4	376	126	987	829	860
DMU_5	876	553	644	670	3852
DMU_6	2766	365	2814	2999	2360
DMU_7	245	715	2305	2009	2196
DMU_8	145	147	128	109	127
DMU_9	136	14054	5070	5370	5828
DMU_{10}	141	1559	3000	2900	2756

The company authorities decided that two real scenarios for output variable (Payments) must be considered for the future analysis. These two scenarios are complemented by the third, rather hypothetical scenario. This last scenario was added for the illustrative purposes of the presented algorithms only. The probabilities of the scenarios are 0.57, 0.42, and 0.01, respectively. The probability of the last scenario is very low as it is for experimental purposes only. The inputs of all scenarios are identical. They differ only in the values of the output. The complete dataset is presented in Table 2. A detailed description of the example is not necessary for our illustrative purposes; it can be found in Zahedi-Seresht et al. (2017).

Table 3 contains efficiency and super-efficiency scores and rankings of all DMUs obtained by traditional DEA models. Efficiency scores for all three scenarios are computed by CCR model (3) for all scenarios independently. Super-efficiency scores are derived by the Andersen and Petersen model. Inefficient units have efficiency scores lower than 1. Efficient units have super-efficiency scores higher than 1. The results show that the rankings of the units obtained for the scenarios are not identical.

Table 3: *Scenario analysis – efficiency (super-efficiency) scores and ranking of DMUs*

	CCR model scen. 1	Rank	CCR model scen. 2	Rank	CCR model scen. 3	Rank
DMU_1	1.672	3	1.925	1	1.203	3
DMU_2	0.206	8	0.227	8	0.237	9
DMU_3	0.364	7	0.430	7	0.380	8
DMU_4	0.752	5	0.648	6	0.815	6
DMU_5	0.139	10	0.149	9	1.011	5
DMU_6	0.740	6	0.809	5	0.772	7
DMU_7	1.072	4	0.963	4	1.161	4
DMU_8	0.147	9	0.132	10	0.170	10
DMU_9	1.752	2	1.920	2	2.192	1
DMU_{10}	1.824	1	1.916	3	1.669	2

The results of the models proposed in this paper are included in Table 4. The first two columns of this table contain efficiency scores and ranking of DMUs obtained by non-linear model (8). The results were computed using the global non-linear solver included in the LINGO modelling system. The last two pairs of columns show the same information for linear model (13) and super-efficiency model (14). As expected, the efficiency scores of all DMUs computed using linear model (13) are always between the worst and the best CCR efficiency scores over all scenarios. An extreme situation occurs for DMU_5 that is very far from the efficient frontier in evaluation by considering the first two scenarios, but it is efficient in the last scenario. Even though this scenario has a very low probability, DMU_5 is identified as efficient in model (8). This example demonstrates one of the properties of this model that was proved in the paper. Model (13) and its super-efficiency version (14) returns for this unit the efficiency score corresponding to expectations of decision-makers. Due to this property, it is hardly estimating deviations between efficiency scores of both models (8) and (13). Model (8)'s return results may be quite far from the results from model (13), as demonstrated on the case of DMU_5 in our example. Model (13) reflects the probabilities of the scenarios in a more suitable way.

Table 4: *Scenario analysis – results of the proposed models*

	Eff. score Model (8)	Rank	Eff. score Model (13)	Rank	Eff. score Model (14)	Rank
DMU_1	1.000	1	1.000	1	1.773	3
DMU_2	0.239	9	0.215	8	0.215	8
DMU_3	0.431	8	0.392	7	0.392	7
DMU_4	0.815	6	0.709	6	0.709	6
DMU_5	1.000	1	0.152	9	0.152	9
DMU_6	0.809	7	0.769	5	0.769	5
DMU_7	1.000	1	1.000	1	1.027	4
DMU_8	0.171	10	0.141	10	0.141	10
DMU_9	1.000	1	1.000	1	1.825	2
DMU_{10}	1.000	1	1.000	1	1.861	1

In order to obtain efficiency scores of all DMUs for all scenarios using the traditional CCR model, it is necessary to solve model (3) $n \cdot S$ times, where n is the number of DMUs and S is the number of scenarios. Model (13) is linear and has the same computational complexity as the traditional DEA model (3). To obtain the results for all n DMUs, this model must be solved n times only. This means that its computational complexity is independent of the number of scenarios involved. The efficiency scores of all units obtained by this model are always lower than or equal to 1. Super-efficiency scores obtained by model (14) allow ranking of originally efficient DMUs.

An interesting task in the analysis of efficiency of DMUs with alternative scenarios is sensitivity analysis with respect to the probabilities of scenarios. It is clear that, in case of maximum probability for the i-th scenario ($p_i = 1$), the efficiency (super-efficiency) scores are identical to those obtained by the traditional CCR model (3) for this scenario. Table 5 contains results for $p_1 = 0.2$, 0.4, 0.6, and 0.8, including the rankings of all DMUs. The probability for the third scenario remains unchanged ($p_3 = 0.01$). In our example, this analysis shows that the final ranking is identical for all values of probabilities. Of course, it is given by a high level of similarity of all scenarios, and it is not a general conclusion.

Table 5: *Sensitivity analysis for model (14)*

	$p_1 = 0.2$	Rank	$p_1 = 0.4$	Rank	$p_1 = 0.6$	Rank	$p_1 = 0.8$	Rank
DMU_1	1.867	3	1.817	3	1.958	3	1.715	3
DMU_2	0.223	8	0.219	8	0.237	8	0.211	8
DMU_3	0.416	7	0.403	7	0.433	7	0.377	7
DMU_4	0.671	6	0.691	6	0.777	6	0.733	6
DMU_5	0.156	9	0.154	9	0.166	9	0.149	9
DMU_6	0.795	5	0.781	5	0.848	5	0.753	5
DMU_7	0.987	4	1.009	4	1.127	4	1.052	4
DMU_8	0.135	10	0.139	10	0.155	10	0.145	10
DMU_9	1.887	2	1.853	2	2.011	2	1.786	2
DMU_{10}	1.895	1	1.877	1	2.050	1	1.840	1

The example presented in this section was quite simple and the efficiency scores obtained by traditional CCR model for all scenarios independently are close in number. That is the reason why the experiments presented in Table 5 are very similar and rankings of the DMUs are identical. Of course, in general, the number of scenarios may be relatively high and the results (efficiency scores) may differ significantly. In this case, the sensitivity analysis of results with respect to the probabilities is of a higher importance. One of the ways in which to perform a deeper analysis is by using the Monte Carlo simulation.

5. Conclusions

This paper contains a contribution to the DEA theory for the case in which uncertainty in the dataset is expressed using multiple alternative scenarios. This situation may occur frequently in cases where the decision-maker is focused on an analysis of future development of a system under control, i.e. how the efficiency of the units under evaluation will change under the assumption of several directions of a future development.

The main contribution of this paper is the formulation of models for efficiency evaluation under the assumption of multiple alternative scenarios with a given probability of their occurrence. At first, a model that is derived from traditional CCR is introduced. This model assumes an aggregated DMUs defined as the weighted sum of all alternative scenarios. Unfortunately, this first model is not linear and cannot be linearized. Its solution is not easy to obtain and has some other negative properties. The second model for dealing with alternative scenarios is a modification of the first one. This model is linear and is computationally efficient. Like other traditional DEA models, this model can result in a conclusion where more than one DMU is efficient (but at least one of them is always efficient). To distinguish among several efficient DMUs, a super-efficiency modification of this model was introduced. The properties of both efficiency and super-efficiency models have been derived and discussed. They have been illustrated on a small example with ten DMUs and three scenarios.

Future research directions in this field can consist of a verification of properties of the model on large-scale examples or more extensive case studies in a real economic context. For future studies, we are going to analyze a relation between the efficiency scores of linear and non-linear models proposed in this paper for non-efficient DMUs. Also, properties of non-linear model (8) and its relation to the results of the analysis for independent scenarios may be interesting to investigate. Further, the research can be focused on deriving feasibility conditions for super-efficiency model (14) and the sensitivity analysis of results for all models presented.

References

Andersen, P. & Petersen, N.C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management Science*, 39(10): 1261–1264.

Arabmaldar, A., Jablonsky, J., & Saljooghi, F.H. (2017). A new robust DEA model and super-efficiency measure. *Optimization*, 66(5): 723–736.

Banker, R.D., Charnes, A., & Cooper, W.W. (1984). Some models for estimating technical and scale efficiencies in data envelopment analysis. *Management Science*, 30(9): 1078-1092.

Beale, E.M.L. (1955). On minimizing a convex function subject to linear inequalities. *Journal of Royal Statistics Society*, 17(2): 173-184.

Charnes, A, Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6): 429–444.

Cooper. W. W., Park, K. S., & Yu, G. (1999). IDEA and AR-IDEA: Models for dealing with imprecise data in DEA. *Management Science*, 45(4): 597–607.

Cooper, W. W., Seiford, L., & Tone, K. (1999). *Data envelopment analysis: A comprehensive text with models, applications, references and DEA-Solver software*. London: Kluwer Academic.

Dantzig, G. B. (1955). Linear programming under uncertainty. *Management Science*, 1(3-4): 197–206.

Despotis, D. K., & Smirlis, Y.G. (2002). Data envelopment analysis with imprecise data. *European Journal of Operational Research*, 140(1): 24–36.

Entani, T., Maeda, Y., & Tanaka, H. (2002). Dual models of interval DEA and its extension to interval data. *European Journal of Operational Research,* 136(1): 32–45.

Guo, P., & Tanaka, H. (2001). Fuzzy DEA: A perceptual evaluation method. *Fuzzy Sets and Systems*, 119(1): 149–160.

Farell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120(3): 253–290.

Hafezalkotob, A., Haji-Sami, E., & Omrani, H. (2015). Robust DEA under discrete uncertain data: A case study of Iranian electricity distribution companies, *Journal of Industrial Engineering International,* 11(2): 199–208.

Hatami-Marbini, A., & Saati, S. (2009). Stability of RTS of efficient DMUs in DEA with fuzzy under fuzzy data. *Applied Mathematical Sciences*, 3(44): 2157–2166.

Jahanshahloo, G.R., Sanei, M., Rostamy-Malkhalifeh, M., & Saleh, H. (2009). A comment on "A fuzzy DEA/AR approach to the selection of flexible manufacturing systems". *Computers and Industrial Engineering*, 56(4): 1713–1714.

Kao, C. (2006). Interval efficiency measures in data envelopment analysis with imprecise data. *European Journal of Operational Research*, 174(2): 1087–1099.

Kao, C. & Liu, S. T. (2000). Fuzzy efficiency measures in data envelopment analysis. *Fuzzy Sets and Systems*, 113(3): 427–437.

Keefer, D. L., & Bodily, S. E. (1983). Three-point approximations for continuous random variables. *Management Science*, 29(5): 595–609.

Kuosmanen, T. (2009). Data envelopment analysis with missing data. *Journal of the Operational Research Society*, 60(12): 1767–1774.

Lertworasirikul, S., Fang, S. C., Joines, J. A., & Nuttle, H. L. W. (2003). Fuzzy data envelopment analysis (DEA): A possibility approach. *Fuzzy Sets and Systems,* 139(2): 379–394.

Liu, S. T. (2008). A fuzzy DEA/AR approach to the selection of flexible manufacturing system. *Computers and Industrial Engineering*, 54(1): 66–76.

Liu, S. T., & Chuang, M. (2009). Fuzzy efficiency measures in fuzzy DEA/AR with application to university libraries. *Expert Systems with Applications*, 36(2): 1105–1113.

Mulvey, J.M., & Ruszczynski, A. (1995). A new scenario decomposition method for large-scale stochastic optimization. *Operations Research*, 43(3): 477–490.

Mulvey, J. M., Vanderbei J. V., & Zenios, S. A. (1995). Robust optimization of large-scale systems. *Operations Research*, 43(2): 264–281.

O'Neal, P. V., Ozcan, Y. A., & Yanqiang, M. (2002). Benchmarking mechanical ventilation services in teaching hospitals. *Journal of Medical Systems*, 26(3): 227–240.

Podinovski, V. V., Olesen, O. B., Sarrico, C. S. (2017). Nonparametric production technologies with multiple component processes. *Operations Research*, 66(1): 282-300.

Podinovski, V. V. (2001). DEA models for the explicit maximization of relative efficiency. *European Journal of Operational Research*, 131(3): 572–586.

Podinovski, V. V., & Athanassopoulos, A. D. (1998). Assessing the relative efficiency of decision making units using DEA models with weight restrictions. *Journal of the Operational Research Society*, 49(5): 500–508.

Saati, S., Memariani, A., & Jahanshahloo, G. R. (2002). Efficiency analysis and ranking of DMUs with fuzzy data. *Fuzzy Optimization and Decision Making*, 1(3): 255–267.

Sadjadi, S. J., & Omrani, H. (2008). Data envelopment analysis with uncertain data: An application for Iranian electricity distribution companies. *Energy Policy*, 36(11): 4247-4254.

Seiford, L. M. (1996). Data envelopment analysis: The evolution of the state of the art (1978–1995). *Journal of Productivity Analysis*, 7(1): 99–137.

Seiford, L. M., & Thrall, R. M. (1990). Recent developments in DEA: The mathematical programming approach to frontier analysis. *Journal of Econometrics*, 46(1): 7-38.

Sengupta, J. K. (1992). A fuzzy systems approach in data envelopment analysis. *Computers and Mathematics with Applications*, 24(8–9): 259–266.

Shokouhi, A. H., Hatami-Marbini, A., Tavana, M., & Saati, S. (2010). A robust optimization approach for imprecise data envelopment analysis. *Computers & Industrial Engineering*, 59(3): 387-397.

Smirlis, Y. G., Maragos, E. K., & Despotis, D. K. (2006). Data envelopment analysis with missing values: An interval DEA approach. *Applied Mathematics and Computation*, 177(1): 1–11.

Soleimani-Damaneh M. (2009). Establishing the existence of a distance-based upper bound for a fuzzy DEA model using duality. *Chaos, Solitions and Fractals*, 41(1): 485–490.

Soleimani-Damaneh, M., Jahanshahloo, G. R., & Abbasbandy, S. (2006). Computational and theoretical pitfalls in some current performance measurement techniques and a new approach. *Applied Mathematics and Computation*, 181(2): 1199–1207.

Toloo, M., & Ertay, T. (2014). The most cost efficient automotive vendor with price uncertainty: A new DEA approach. *Measurement*, 52(1): 135–144.

Vaez-Ghasemi, M. (2015). Supplier selection among alternative scenarios by data envelopment analysis. *Iranian Journal of Optimization*, 7(2): 815–821.

Wang, Y. M., Greatbanks, R., & Yang, J.B. (2005). Interval efficiency assessment using data envelopment analysis. *Fuzzy Sets and Systems,* 153(3): 347–370.

Wang, Y. M., Luo, Y., & Liang, L. (2009). Fuzzy data envelopment analysis based upon fuzzy arithmetic with an application to performance assessment of manufacturing enterprises. *Expert Systems with Applications*, 36(3): 5205–5211.

Wu, D., Ding, W., Koubaa, A., Chaala, A., & Luo, C. (2017). Robust DEA to assess the reliability of methyl methacrylate-hardened hybrid poplar wood. *Annals of Operations Research,* 248(1–2): 515–529.

Zahedi-Seresht, M., Mehrabian, S., & Jahanshahloo, G.R. (2016). A new method for ranking distribution companies with several scenarios data by using DEA/MADM. *International Journal of Applied Operational Research*, 6(3): 11–24.

Zahedi-Seresht, M., Jahanshahloo, G.R., & Jablonsky, J. (2017). A robust data envelopment analysis model with different scenarios. *Applied Mathematical Modelling*, 52: 306–319.

Acknowledgements

The research is supported by the Grant Agency of the Czech Republic, project no. 19-08985S.

Credit Author Statement

Zahedi-Seresht, Mazyar: Conceptualization, Methodology

Khosravi, Shahrzad: Writing-Original Draft Preparation

Jablonsky, Josef: Supervising, Reviewing, Software

Zykova, Petra: Writing-Reviewing, Editing

Highlights

- An original model for ranking of DMUs in data envelopment analysis models with alternative scenarios in the dataset is introduced.
- The properties of the new model are derived and mathematically proved.
- Comparing to the conventional models, the proposed approach ensures feasibility and is computationally efficient.
- The model is illustrated on an example with 10 DMUs and 3 scenarios.