CEEP-BIT WORKING PAPER SERIES

CEEP-BIT

Robust data envelopment analysis based MCDM with the consideration of uncertain data

Ke Wang
Fajie Wei

Working Paper 87
http://ceep.bit.edu.cn/english/publications/wp/index.htm

Center for Energy and Environmental Policy Research
Beijing Institute of Technology
No.5 Zhongguancun South Street, Haidian District
Beijing 100081
February 2016

This paper can be cited as: *Wang K, Wei F. 2016. Robust data envelopment analysis based MCDM with the consideration of uncertain data. CEEP-BIT Working Paper.*

The authors thank the reviewers of the paper and are grateful to Professor Xin Chen and Dr. Yuhan Zhang for their valuable comments and suggestions on an earlier version of the paper. This paper was finished while Ke Wang was visiting University of Illinois at Urbana-Champaign. The views expressed herein are those of the authors and do not necessarily reflect the views of the Center for Energy and Environmental Policy Research.

The Center for Energy and Environmental Policy Research, Beijing Institute of Technology (CEEP-BIT), was established in 2009. CEEP-BIT conducts researches on energy economics, climate policy and environmental management to provide scientific basis for public and private decisions in strategy planning and management. CEEP-BIT serves as the platform for the international exchange in the area of energy and environmental policy.

Currently, CEEP-BIT Ranks 47, top 3% institutions in the field of Energy Economics at IDEAS（http://ideas.repec.org/top/top.ene.htm), and Ranks 52, top 3% institutions in the field of Environmental Economics at IDEAS (http://ideas.repec.org/ top/top.env.html).

Yi-Ming Wei
Director of Center for Energy and Environmental Policy Research, Beijing Institute of Technology

For more information, please contact the office:

Address:
Director of Center for Energy and Environmental Policy Research
Beijing Institute of Technology
No.5 Zhongguancun South Street
Haidian District, Beijing 100081, P.R. China

Access:
Tel: +86-10-6891-8551
Fax: +86-10-6891-8651
Email: ceeper@vip.163.com
Website: http://ceep.bit.edu.cn/english/index.htm

Robust data envelopment analysis based MCDM with the consideration of uncertain data

Ke Wang [1,2,3,1] , Fajie Wei [4]

1. Center for Energy and Environmental Policy Research, Beijing Institute of Technology, Beijing, China
2. School of Management and Economics, Beijing Institute of Technology, Beijing, China
3. Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, Urbana, USA
4. School of Economics and Management, Beihang University, Beijing, China

Abstract: The application of data envelopment analysis (DEA) as a multiple criteria decision making (MCDM) technique has been gaining more and more attention in recent research. In the practice of applying DEA approach, the appearance of uncertainties on input and output data of decision making unit (DMU) might make the nominal solution infeasible and lead to the efficiency scores meaningless from practical view. In this paper, we analyze the impact of data uncertainty on the evaluation results of DEA, and propose several robust DEA models based on the adaptation of recently developed robust optimization approaches, which would be immune against input and output data uncertainties. The robust DEA models we developed are based on input-oriented and output-oriented CCR model, respectively, when the uncertainties appear in output data and input data separately. Furthermore, our robust DEA models could deal with random symmetric uncertainty and unknown-but-bounded uncertainty, in both of which the distributions of the random data entries are permitted to be unknown. We implement the robust DEA models in a numerical example and the efficiency scores and rankings of these models are compared. The results indicate that the robust DEA approach could be a more reliable method for efficiency evaluation and ranking in MCDM problems.

Keywords: data envelopment analysis (DEA), multiple criteria decision making (MCDM), robust optimization, uncertain data, efficiency, ranking.

1. Introduction

Data envelopment analysis (DEA), initiated by Charnes et al. (CCR), is a mathematical programming methodology for measuring the relative efficiencies of a set of decision making unit (DMU) [1]. Cooper et al. pointed out that DEA has been used in evaluating the performances of many different kinds of entities engaged in many different kinds of activities in many different contexts [2]. Cook and Seiford summarized the major research in DEA over the last 30 years, which provides a very good research framework [3]. The application of DEA as an alternative multiple criteria decision making (MCDM) technique has been gaining more and more attention. Stewart contrasted DEA and MCDM, and pointed out that not only these two approaches are superficially similar problems, but the concepts of efficiency in DEA and Pareto optimality in MCDM are comparable [4]. He indicated that DEA could be seen as an alternative MCDM tool. Sarkis analytically compared DEA with the structure of MCDM and concluded that DEA could provide comparable results to traditional MCDM approaches, and DEA is advantageous to decision makers by requiring less information [5]. Therefore, the success of wide use of DEA in the area of performance evaluation together with the formal analogies between DEA and MCDM could make DEA as a good

---

[1] Corresponding author at Beijing Institute of Technology. Email: kewang2083@gmail.com. Tel/Fex: 0086-10-68918551.

alternative approach of MCDM.

In the conventional DEA models, all the data of the input and output are assumed to have the form of specific numerical values which are "know exactly". However this assumption may not be always true. In some situations, some data may not be known exactly with high accuracy but known as in the forms of interval data or ordinal data. Cook et al. extended the data type in DEA to ordinal form [6,7], and Cooper et al. discussed the case of interval data [8,9]. The mixtures of these data are referred to as imprecise data and the associated DEA models are proposed [8]. There are two different approaches in dealing with the imprecise data in DEA. One approach uses the scale transformations and variable alternations to convert the non-linear DEA model into a linear program, which were proposed by Cooper et al. [8], and then were simplified by Zhu [10] and further developed by Kao [11]. The other approach converts imprecise data into exact data first and then uses the standard linear DEA model to calculate the efficiency intervals. Kao developed such a method which applies a two-level mathematical programming [11].

If the "not known exactly data" in DEA are considered to be more general as, say, "uncertain data", which have uncertain entries and the "true" values of these data are unknown. For example, to consider a measure of the data violation, assume that the true values of the uncertain data are obtained from the "nominal values" by random perturbations $a_j \rightarrow \tilde{a}_j = (1+\xi_j)a_j$ where ξ_j is a random variable distributed on $[-0.001, 0.001]$. We have several reasons to believe that some of the input and output data of DEA, which characterize certain evaluating systems or processes, could hardly be known to exact or with high accuracy. Therefore, as Ben-Tal and Nemirovski mentioned in their research, it is quite nature to assume that these input and output data are indeed uncertain [12]. In fact, in the real-world evaluation problem, there is no guarantee that all the input and output data could be observed accurately, especially under the situation that some of the evaluation information are came from the expert's subjectively judgment or approximately estimate, which are provided as imprecise data. When such kinds of data are used in DEA, a certain DMU is reluctant to admit that it performs worse than another. Furthermore, in the survey study of some benchmark problems, Ben-Tal and Nemirovski pointed out that in linear programming problems, even "quiet small perturbations of 'obvious uncertain' data coefficients can make the 'nominal' optimal solution heavily infeasible and thus practically meaningless" [12]. As a linear programming based approach, DEA will never be able to escape from the impact of data uncertainty, i.e., a small perturbation on input and output data of DMU could make a big change on the efficiencies, so the results of the ranking could be unreliable in many cases especially when the efficiency of particular DMU is close to another. We will take a further analysis of the impact of data uncertainty on DEA in Section 2. Therefore, in the application of DEA, we need to develop some "robust" DEA models which are capable of generating more stable efficiency scores and more reliable related ranking, so that a small change in input and output data cannot change the evaluation results.

Based on the robust optimization approach of Ben-Tal and Nemirovski [12,13,14], and Bertsimas and Sim [15,16,17], Sadjadi and Omrani proposed a new DEA method with the consideration of uncertain data, which could be seen as a robust DEA [18,19]. Since they proposed a good idea for formulating the DEA model which could deal with symmetric and bounded random variables as the entries of the data, their model should be further improved. Because though they pointed out that the uncertainty should be considered in different parts of input and output data, they only proposed the model which considers the output uncertainty. The reason that they did not give the model which considers the input uncertainty might be that the output uncertainty model is based on input-oriented CCR linear model, but the input uncertainty model could hardly be formulated based on the same oriented CCR linear model with clear meanings of the entries of uncertainties.

Recent developed DEA sensitivity analysis could be considered as a method which may deal with DEA uncertainty problem. Charnes et al. utilized the concept of distance to determine radii of stability within which the occurrence of data variations will not alter a DMU's classification status [20]. Another type of DEA sensitivity analysis is based on super efficiency DEA approach, which is initiated by Andersen and Petersen [21]. This technique developed by Charnes et al. [22] could be used in the situation where simultaneous proportional variance is occurred in all inputs and outputs for a specific DMU under evaluation. Then Zhu proposed a new approach, where the DEA sensitivity analysis could be done in a general situation that the input and output data of a testing DMU and the remaining DMU are allowed to vary simultaneously and unequally [23,24]. For a summary discussion of DEA sensitivity analysis, see Cooper et al. [25]. However, we have to emphasize that our approach in this paper is quite different from sensitivity analysis. In DEA sensitivity analysis, people are interested in how much the efficiency score of DMU to a perturbed problem can differ from the nominal problem. In contrast, we want to know by how much the optimal efficiency to the nominal problem can violate the constraints of the perturbed problem. Furthermore, DEA sensitivity analysis can only quantify the stability of the nominal efficiency with respect to input and output data perturbations, but it does not show the way on how to improve the stability of DMU efficiency. The latter issue is exactly what is addressed by robust DEA method.

In this paper, we develop several robust DEA models based on Ben-Tal and Nemirovski's robust optimization approaches [12,13,14], which could deal with uncertainty data and provide stable and reliable evaluation results. And the models proposed in this paper, which could consider the input uncertainties as well as output uncertainties are based on output-oriented and input-oriented CCR models, respectively, so as to make the meanings of the uncertainties in these models clear. Furthermore, our robust DEA models also could deal with the data uncertainty in different forms: random symmetric uncertainty and unknown-but-bounded uncertainty.

2. Data envelopment analysis and the impact of uncertain data on it

The original fractional DEA model which is an input-oriented CCR model is presented in model (1):

$$
\begin{aligned}
&\max \frac{\sum_{r=1}^{s} v_r y_{rj_0}}{\sum_{i=1}^{m} u_i x_{ij_0}} \\
&s.t. \frac{\sum_{r=1}^{s} v_r y_{rj}}{\sum_{i=1}^{m} u_i x_{ij}} \le 1,\ j = 1,\dots,n, \\
&\qquad v_r, u_i \ge 0, r = 1,\dots s, i = 1,\dots,m.
\end{aligned} \tag{1}
$$

which evaluates the relative efficiencies of n DMUs, each with m inputs and s outputs denoted by x_{ij} and y_{rj}, respectively, by maximizing the ratio of weighted summation of outputs to weighted summation of inputs. u_i and v_r are weights associated with inputs and outputs, respectively. In addition, x_{ij_0} and y_{rj_0} are the ith input and rth output for the DMU under evaluation. Model (1) is a non-linear programming problem, and it is equivalent to the following linear programming problem (2) which is more computational convenient:

$$
\begin{aligned}
&\max \sum_{r=1}^{s} \mu_r y_{rj_0} \\
&s.t.\ \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \omega_i x_{ij} \le 0,\ j = 1,\dots,n, \\
&\qquad \sum_{i=1}^{m} \omega_i x_{ij_0} = 1, \\
&\qquad \mu_r, \omega_i \ge 0, r = 1,\dots s, i = 1,\dots,m.
\end{aligned} \tag{2}
$$

Now we analyze how the uncertain data affect the DEA results. We intend to take the analysis based on the simplest assumption: The "true" value $\tilde{x}_{ij}$ and $\tilde{y}_{rj}$ of the uncertain input and output data are obtained from the nominal value x_{ij} and y_{rj} by random perturbation: $\tilde{x}_{ij} = (1+\varepsilon_x \xi_{ij}) x_{ij}$ and $\tilde{y}_{rj} = (1+\varepsilon_y \xi_{rj}) y_{rj}$, where ε_x and ε_y are given uncertainty levels and ξ_{ij} and ξ_{rj} are random variables distributed symmetrically in the interval $[-1,1]$. The random perturbations affecting the uncertain data of a particular inequality constraint are independent identical distribution. Now we look at a DEA example which includes ten DMUs, each has three inputs and two outputs. For the inputs and outputs data, see Table 1.

Table 1 The input and output data for each DMU

DMU	Input			Output	
	x_1	x_2	x_3	y_1	y_2
A	11	15	115	102	64
B	10	24	127	109	64
C	30	20	109	90	69
D	7	20	82	124	70
E	11	16	123	88	77
F	8	11	135	109	89
G	14	25	118	120	72
H	35	18	126	80	69
I	9	18	86	117	60
J	12	19	75	94	88

First we solve the problem in standard CCR model (1) so as to get the nominal optimal solution of weights to compute the efficiency scores and ranking, which are shown in the second and third columns of Table 2. Then using the nominal optimal solution of weights reported by the CCR model (1), we compute the efficiency scores and rankings when the input or output data are randomly perturbed for the uncertainty level $\varepsilon = 0.1$. The results of the analysis are shown in the fourth to eleventh columns of Table 2. The impact of perturbations appearing on single input x_2 or single output y_2, are shown in the forth to fifth or sixth to seventh columns; and the impact of perturbations appearing on all input x or all outputs y are shown in the eighth to ninth or tenth to eleventh columns. From the analysis results we could find that: in 10 DMUs, the nominal efficiency scores of 3 DMUs (DMU D, F and I) turned out to be greater than 1 when input x_2 have uncertain entries; the nominal efficiency scores of 3 DMUs (DMU D, F and J) turned out to be greater than 1 when output y_2 have uncertain entries; the nominal efficiency scores of 3 DMUs (DMU D, F and I) turned out to be greater than 1 when all input x have uncertain entries; and the nominal efficiency scores of 3 DMUs (DMU D, F and J) turned out to be greater than 1 when all input y have uncertain entries. The efficiency scores greater than 1 violate the inequality constraint of model (1). The reason of the efficiency scores becoming greater than 1 is that the optimal solution of weights from the nominal problem becomes (partially) infeasible when the data uncertainties appear. In another word, it is because that the perturbed efficiency scores are computed by using the nominal optimal solution of weights obtained from the standard CCR model in which the input and output data are assumed exact. Furthermore, the ranking for nominal efficiency scores and the rankings for all other efficiency scores affected by uncertain data entries are inconsistent with each other, and the largest gap between the rankings for the same DMU is 3

(DMU J). Therefore, the analysis of the impact of data uncertainty on DEA leads to the following conclusion: in real-world evaluation process, when applying DEA approaches, one cannot ignore the possibility that a small perturbation on input or output data of DMU can make the nominal optimal solution of the weights (partially) infeasible, which leads to (some of) the efficiency scores meaningless from practical view and the ranking unreliable. Consequently, in the application of DEA, there exists a real need of robust DEA method which is immune against input and output data uncertainty that can heavily affect the quality of efficiency scores, and is capable of generating more reliable ranking for decision making units.

Table 2 The results from nominal and uncertainty problem

DMU	CCR		Perturbation on x_2		Perturbation on y_2		Perturbations on all x		Perturbations on all y	
	Efficiency score	Ranking	Efficiency score	Ranking	Efficiency score	Ranking	Efficiency score	Ranking	Efficiency score	Ranking
A	0.859 1	5	0.818 8	5	0.859 2	5	0.824 3	5	0.828 1	5
B	0.670 1	9	0.659 5	9	0.672 5	9	0.661 5	9	0.636 7	10
C	0.703 4	8	0.735 6	7	0.706 9	8	0.701 4	8	0.680 6	8
D	1.000 0	1	1.041 5*	1	1.002 1*	3	1.065 0*	2	1.001 2*	3
E	0.776 5	6	0.764 9	6	0.757 6	6	0.783 3	6	0.749 4	6
F	1.000 0	1	1.029 9*	2	1.006 1*	2	1.095 0*	1	1.020 8*	2
G	0.740 2	7	0.731 4	8	0.714 4	7	0.708 1	7	0.695 9	7
H	0.638 4	10	0.657 8	10	0.651 2	10	0.656 8	10	0.673 4	9
I	0.986 9	4	1.028 8*	3	0.982 5	4	1.036 3*	3	0.956 7	4
J	1.000 0	1	0.949 0	4	1.035 4*	1	0.967 2	4	1.035 3*	1

* denotes the efficiency scores which are greater than 1 and violate the model constraint.

3. Robust optimization

The robust optimization technique has recently been introduced into the mathematical programming problem to deal with the entering of perturbation. When modeling the optimization problem with data uncertainty, robust optimization technique could provide a solution that is guaranteed to be good for all or most possible realizations of the uncertainty in the parameters. Soyster investigated the robust optimization approaches in which the column vectors of the constraint matrix were assumed to belong in a convex uncertainty sets [26], and Ben-Tal and Nemirovski proposed a new method to model the uncertainty data based on interval and ellipsoidal uncertainty sets [12,13,14]. To present the robust optimization method, consider a linear programming problem (3):

$$\begin{aligned} &\min c^T x \\ &s.t. \quad Ax \le b, \\ &\qquad l \le x \le u. \end{aligned} \qquad (3)$$

where A is the matrix of coefficients which is assumed to be affected by uncertainty and x is the vector of decision variables. To model the uncertainty in coefficients under robust optimization technique, consider a particular row, the ith row, of the matrix A and let J_i represent the subscript set of coefficients in row i that are subject to uncertainty. Ben-Tal and Nemirovski proposed that there are two ways to implement the robust optimization technique, depending on whether to treat the uncertainty affecting the coefficients as "unknown-but-bounded", or as random [12].

Under the "unknown-but-bounded" coefficient assumption, a solution which immune against entry-wise uncertainty of given magnitude ε affecting uncertain coefficients of linear programming is needed, i.e., the solution should have such characteristics: it is feasible for the nominal problem (3); suppose that in the *i*th inequality constraint of (3), each entry of the coefficient $\tilde{a}_{ij}, j \in J_i$, of uncertain data is assumed to range in the interval $[a_{ij} - \varepsilon|a_{ij}|, a_{ij} + \varepsilon|a_{ij}|]$, where $\varepsilon > 0$ is a given "uncertainty level", and "| |" is the absolute value sign. Whatever are the true values of uncertain coefficient from these intervals, the solution must satisfy the *i*th constraint. Such a solution is called (ε)-reliable solution. Then the robust counterpart of (3) is:

$$\begin{aligned} &\min c^T x \\ &s.t. \quad Ax \le b, \\ &\qquad \sum_j a_{ij} x_j + \varepsilon \sum_{j \in J_i} |a_{ij}||x_j| \le b_i \quad \forall i, \\ &\qquad l \le x \le u. \end{aligned} \tag{4}$$

It is easy to see that (4) is equivalent to the linear programming (5):

$$\begin{aligned} &\min c^T x \\ &s.t. \quad Ax \le b, \\ &\qquad \sum_j a_{ij} x_j + \varepsilon \sum_{j \in J_i} |a_{ij}| y_j \le b_i \quad \forall i, \\ &\qquad -y_j \le x_j \le y_j \quad \forall j \in J_i, \\ &\qquad l \le x \le u. \end{aligned} \tag{5}$$

where x_j and y_j are decision variables extended by x_j. Then, the way to get a robust optimal solution to (3) is to solve the ε-interval robust counterpart (5) (*IRC*[ε]) of the uncertainty problem.

Under the random coefficient assumption, the second characteristic of the solution in the above assumption will be translated to its probabilistic version: in the *i*th inequality constraint of (3), each entry of the coefficient $\tilde{a}_{ij}, j \in J_i$, of uncertain data is obtained from the nominal values a_{ij} by random perturbations: $\tilde{a}_{ij} = (1 + \varepsilon \xi_{ij}) a_{ij}$, where $\varepsilon > 0$ is a given uncertainty level (percentage of perturbations), $\xi_{ij} = 0$ for $j \notin J_i$, and the perturbations ξ_{ij} are independent random variables symmetrically distributed in the interval [-1,1]. The solution of this problem is called (ε, Ω)-almost reliable solution. Then the robust counterpart of (3) is:

$$\begin{aligned} &\min c^T x \\ &s.t. \quad Ax \le b, \\ &\qquad \sum_j a_{ij} x_j + \varepsilon \left[\sum_{j \in J_i} |a_{ij}| y_{ij} + \Omega \sqrt{\sum_{j \in J_i} a_{ij}^2 z_{ij}^2} \right] \le b_i \ \forall i, \\ &\qquad -y_{ij} \le x_j - z_{ij} \le y_{ij} \quad \forall j \in J_i, \forall i, \\ &\qquad l \le x \le u. \end{aligned} \tag{6}$$

where x_j, y_{ij} and z_{ij} are x_j extended decision variables, and Ω is a preference parameter or "safety parameter". Parameter Ω is controlled by the given "reliability level" $\kappa > 0$, which we will explain in the proposition below. In (6), under the coefficients uncertainty, we have a proposition as:

Proposition. Assume (x_j, y_{ij}, z_{ij}) are the feasible solution of the optimization problem (6), where Ω is a positive parameter. Then for every *i*, the probability of the event $\sum_j \tilde{a}_{ij} x_j > b_i$ is at most $\kappa = \exp(-\Omega^2/2)$, i.e., the probability of the violation of the *i*th constraint in (6) is at most κ.

For the proof of this proposition, see [12]. Then, the way to get a robust optimal solution to (3) is to solve the (ε, Ω)-robust counterpart (6) ($RC[\varepsilon, \Omega]$) of the uncertainty problem.

As Ben-Tal and Nemirovski indicated, $RC[\varepsilon, \Omega]$ is less conservative than $IRC[\varepsilon]$, because the feasible solution (x_j, y_j) of $IRC[\varepsilon]$ is a feasible solution of $RC[\varepsilon, \Omega]$ when z_{ij} is set to be 0. Infect, in the case of large set J_i, $IRC[\varepsilon]$ can be much more restrictive than $RC[\varepsilon, \Omega]$ [12]. Here, we highlight the essential difference between "unknown-but-bounded" coefficient assumption and random coefficient assumption: $IRC[\varepsilon]$ assumes that uncertainty is represented by a range of potential values, and no probability distribution is assigned to this range of potential values. To satisfy a constraint in the $IRC[\varepsilon]$ means to satisfy the constraint for any uncertainty variable in the designated uncertainty set, with no exception. On the other hand, $RC[\varepsilon, \Omega]$ represents uncertainty as a random variable, which takes certain probability distribution. In $RC[\varepsilon, \Omega]$, constraints can only be satisfied in the probability sense, that is, the constraint should be satisfied with a probability greater than a threshold value of $1-\kappa$. Furthermore, $RC[\varepsilon, \Omega]$ has a practical drawback compared to $IRC[\varepsilon]$, which is that the former non-linear programming problem, although convex and "well-structured", is more demanding computationally than the later linear programming program.

In next section, when adopting the robust optimization technique in DEA, we will use both $RC[\varepsilon, \Omega]$ and $IRC[\varepsilon]$ for different DEA models.

4. Robust data envelopment analysis

In order to generate an "uncertainty-immune" DEA approach, we adopt the robust optimization technique in DEA. Since the uncertainties may appear in different part of the input or output data of DMU, and the uncertainties may appear in different form as the random symmetric or unknown-but-bounded uncertainty, we propose four different robust DEA models with the consideration of these different uncertainties.

When the uncertainty appears in the output data, we adopt the $RC[\varepsilon, \Omega]$ in DEA based on the input-oriented CCR model, so as to avoid the appearance of uncertainty in the input related equality constraint. And in order to avoid the appearance of uncertainty in objective function, we express the objective function as max*z*, and add the constraint $z_{j_0} - \sum_{r=1}^{s} \mu_r y_{rj_0} \le 0$ into the constraints. Therefore, the input-oriented robust DEA model with output uncertainty is expressed in model (7):

$$
\begin{aligned}
& \max z_{j_0} \\
& s.t.\ \ z_{j_0} - \sum_{r=1}^{s} \mu_r y_{rj_0} + \varepsilon_y \left(\sum_{r \in R} \left| y_{rj_0} \right| \alpha_{rj} + \Omega_y \sqrt{\sum_{r \in R} y_{rj_0}^2 \beta_{rj}^2} \right) \le 0, \\
& \sum_{i=1}^{m} \omega_i x_{ij_0} = 1, \\
& \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \omega_i x_{ij} \le 0, \ \ \forall j, \\
& \sum_{r=1}^{s} \mu_r y_{rj} + \varepsilon_y \left(\sum_{r \in R} \left| y_{rj} \right| \alpha_{rj} + \Omega_y \sqrt{\sum_{r \in R} y_{rj}^2 \beta_{rj}^2} \right) - \sum_{i=1}^{m} \omega_i x_{ij} \le 0, \ \ \forall j, \\
& -\alpha_{rj} \le \mu_r - \beta_{rj} \le \alpha_{rj}, \ \ \forall r \in R, \forall j, \\
& \mu_r, \omega_i \ge 0, r = 1,\dots,s, i = 1,\dots,m.
\end{aligned} \tag{7}
$$

where $\omega_i, \mu_r, \alpha_{rj}$ and β_{rj} are decision variables, and x_{ij} and y_{rj} are the ith input and rth output for DMUj, respectively. $y_{rj}, r \in R,$ have random uncertainties ranged in $[(1-\varepsilon_y)\left|y_{rj}\right|, (1+\varepsilon_y)\left|y_{rj}\right|]$, and R is the subscript set of outputs which are subject to uncertainty. In addition, the efficiency of DMU under evaluation is z_{j_0}, and x_{ij_0} and y_{rj_0} are the ith input and rth output for the DMU under evaluation. This robust DEA model can guarantee on the probability of $1-\kappa_y = 1-\exp(-\Omega_y^2/2)$ that the robust solution is feasible.

On the contrary, when the uncertainty appears in the input data, we adopt the RC[ε, Ω] in DEA based on the output-oriented CCR model, so as to avoid the appearance of uncertainty in the output related equality constraint. And in order to avoid the appearance of uncertainty in objective function, we express the objective function as min$z$, and add the constraint $\sum_{i=1}^{m} \omega_i x_{ij_0} - z_{j_0} \le 0$ into the constraints. Therefore, the output-oriented robust DEA model with input uncertainty is expressed in model (8):

$$
\begin{aligned}
& \min z_{j_0} \\
& s.t.\ \ \sum_{i=1}^{m} \omega_i x_{ij_0} + \varepsilon_x \left(\sum_{i \in I} \left| x_{ij_0} \right| \rho_{ij} + \Omega_x \sqrt{\sum_{i \in I} x_{ij_0}^2 \sigma_{ij}^2} \right) - z_{j_0} \le 0, \\
& \sum_{r=1}^{s} \mu_i y_{rj_0} = 1, \\
& \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \omega_i x_{ij} \le 0, \ \ \forall j, \\
& \sum_{r=1}^{s} \mu_i y_{rj} - \sum_{i=1}^{m} \omega_i x_{ij} + \varepsilon_x \left(\sum_{i \in I} \left| x_{ij} \right| \rho_{ij} + \Omega_x \sqrt{\sum_{i \in I} x_{ij}^2 \sigma_{ij}^2} \right) \le 0, \ \ \forall j, \\
& -\rho_{ij} \le \omega_i - \sigma_{ij} \le \rho_{ij}, \ \ \forall i \in I, \forall j, \\
& \mu_r, \omega_i \ge 0, r = 1,\dots,s, i = 1,\dots,m.
\end{aligned} \tag{8}
$$

where $\omega_i, \mu_r, \rho_{ij}$ and σ_{ij} are decision variables, and x_{ij} and y_{rj} are the ith input and rth output for DMUj, respectively. $x_{ij}, r \in I,$ have random uncertainties ranged in $[(1-\varepsilon_x)\left|x_{ij}\right|, (1+\varepsilon_x)\left|x_{ij}\right|]$, and I is the subscript set of inputs which are subject to uncertainty. In addition, the efficiency of DMU under evaluation is z_{j_0}, and x_{ij_0} and y_{rj_0} are the *i*th input and *r*th output for the DMU under evaluation. This robust DEA model can guarantee on the probability of $1-\kappa_x = 1-\exp(-\Omega_x^2/2)$ that the robust solution is feasible.

Similarly, we could adopt the *IRC*[ε] in DEA based on the input-oriented and output-oriented CCR models, respectively, and give the related robust DEA models with output uncertainty and input uncertainty in the form of model (9) and (10):

$$\begin{aligned}
&\max z_{j_0} \\
&s.t.\ \ z_{j_0} - \sum_{r=1}^{s} \mu_r y_{rj_0} + \varepsilon_y \sum_{r\in R} \left|y_{rj_0}\right| \alpha_r \le 0, \\
&\sum_{i=1}^{m} \omega_i x_{ij_0} = 1, \\
&\sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \omega_i x_{ij} \le 0, \quad \forall j, \\
&\sum_{r=1}^{s} \mu_r y_{rj} + \varepsilon_y \sum_{r\in R} \left|y_{rj}\right| \alpha_r - \sum_{i=1}^{m} \omega_i x_{ij} \le 0, \quad \forall j, \\
&-\alpha_r \le \mu_r \le \alpha_r, \quad \forall r \in R, \\
&\mu_r, \omega_i \ge 0, r = 1,\dots,s, i = 1,\dots,m.
\end{aligned} \tag{9}$$

$$\begin{aligned}
&\min z_{j_0} \\
&s.t.\ \ \sum_{i=1}^{m} \omega_i x_{ij_0} + \varepsilon_x \sum_{i\in I} \left|x_{ij_0}\right| \rho_i - z_{j_0} \le 0, \\
&\sum_{r=1}^{s} \mu_i y_{rj_0} = 1, \\
&\sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} \omega_i x_{ij} \le 0, \quad \forall j, \\
&\sum_{r=1}^{s} \mu_i y_{rj} - \sum_{i=1}^{m} \omega_i x_{ij} + \varepsilon_x \sum_{i\in I} \left|x_{ij}\right| \rho_i \le 0, \quad \forall j, \\
&-\rho_i \le \omega_i \le \rho_i, \quad \forall i \in I, \\
&\mu_r, \omega_i \ge 0, r = 1,\dots,s, i = 1,\dots,m.
\end{aligned} \tag{10}$$

In model (9) and (10), $\omega_i, \mu_r, \alpha_r$ and ρ_i are decision variables, and x_{ij} and y_{rj} are the *i*th input and *r*th output for DMU*j*, respectively. $y_{rj}, r \in R,$ and $x_{ij}, r \in I,$ have bounded uncertainties ranged in the interval $[x_{ij} - \varepsilon_x \left|x_{ij}\right|, x_{ij} + \varepsilon_x \left|x_{ij}\right|]$ and $[y_{rj} - \varepsilon_y \left|y_{ij}\right|, y_{rj} + \varepsilon_y \left|y_{rj}\right|]$, respectively, and R and I is the subscript set of outputs and inputs which are subject to uncertainty, respectively. In addition, the efficiency of DMU under evaluation is z_{j_0}, and x_{ij_0} and y_{rj_0} are the *i*th input and *r*th output for the DMU under evaluation.

The above robust DEA models have a same goal which is to maximize the worst-case efficiency score of the under evaluating DMUj_0 (model (8) and (10) are to minimize the reciprocal efficiency scores), i.e., in these models, DMUj_0 is seeking its best efficiency under the worst situation that the output data of DMUj_0

are negatively perturbed (output decreased) and its input data are positively perturbed (inputs increased) by uncertainties, while the output data of all the other DMUs are positively perturbed and their input data are negatively perturbed by uncertainties. Therefore the solving process of robust DEA approach could be seen as a worst-case optimization process, which is the "spirit" and core idea of robust optimization technique.

Model (7) and (8) are non-linear programming problems, and model (9) and (10) are linear programming problems, and all of them could be solved by using ordinary non-linear programming packages. We have to point out that, when adopting the RC[ε, Ω] in DEA, the original DEA linear programming models are transferred into the robust DEA non-linear programming model (7) and (8), which might be harder to solve. In addition, as the size of the problem increases in terms of the variables and the uncertainties, the robust DEA models becomes more complicated. But when adopting the *IRC*[ε] in DEA, the robust DEA model (9) and (10) are still linear. The complexity of the non-linear robust DEA model may be one weakness of the robust DEA approach compared with the standard DEA approach. However, as mentioned above in Section 2 and 3, the most advantage of robust DEA is its uncertainty immune property, which will lead to a more stable efficiency score and a more reliable ranking for each decision making units.

5. Numerical example

In order to give a better understanding of the performance of the proposed robust DEA approach, we implement our models in the same numerical example we used above when analyzing the impact of data uncertainty on DEA results. There are ten DMUs each has three inputs and two outputs. We assume that all of the input and output data entries could be perturbed by random variables ξ_{ij} distributed symmetrically in the interval $[-1,1]$, the uncertainty level is considered to be $\varepsilon = 0.1$ (both for input *x* and output *y*), and the reliability level is set to $\kappa = 0.05$ (Ω=2.45). The results of the robust DEA models are shown in Table 3. The fourth, fifth and sixth columns show the directly computed efficiency (perturbed efficiency), robust efficiency and related ranking of model (7) which deals with the uncertainty appearing in all outputs, respectively. Then, the seventh, eighth and ninth columns show the directly computed efficiency (perturbed efficiency), robust efficiency and related ranking of model (8) which deals with the uncertainty appearing in all inputs, respectively. We have to indicate that since model (8) is output-oriented and the efficiency scores reported by it are greater than or equal to 1. For compare convenience, we use the reciprocal of these scores which are between 0 and 1.

Table 3 The results from different approaches

DMU	CCR		Perturbations on all y			Perturbations on all x		
	Efficiency	Ranking	Perturbed efficiency	Robust efficiency	Ranking	Perturbed efficiency	Robust efficiency	Ranking
A	0.8591	5	0.7524	0.7022	5	0.7431	0.7043	5
B	0.6701	9	0.5791	0.5480	9	0.5960	0.5493	9
C	0.7034	8	0.6187	0.5755	8	0.6358	0.5796	8
D	1.0000	1	0.9104	0.8182	1	0.9454	0.8182	1
E	0.7765	6	0.6806	0.6352	6	0.7052	0.6356	6
F	1.0000	1	0.9661	0.8182	1	0.9981	0.8182	1
G	0.7402	7	0.6326	0.6056	7	0.6380	0.6065	7
H	0.6384	10	0.6122	0.5224	10	0.5971	0.5275	10
I	0.9869	4	0.8686	0.8067	4	0.9332	0.8078	4
J	1.0000	1	0.9683	0.8182	1	0.9433	0.8182	1

Mean	0.8375	-	-	0.6850	-	-	0.6865	-

Note that Table 3 only reports a single result for each DMU with uncertain input and output data, in spite of the fact that we also implement model (9) and (10) to obtain the efficiency scores and related rankings for each DMU. The reason is that with our setup of κ, both $IRC[\varepsilon]$ and $RC[\varepsilon,\Omega]$ have essentially the same optimal values. Similar phenomenon appeared in Ben-Tal and Nemirovski case study within relative inaccuracy 10^{-7} for each result of both the robust counterparts [12].

A possible explanation of such phenomenon is that the reliability level was set high and thus leaded to a large Ω. With this Ω, $RC[\varepsilon,\Omega]$ will be less conservative than $IRC[\varepsilon]$ only when there were a large number of uncertain data entries per constraint. Actually, we only assume two or three uncertain data entries for each constraint of our robust DEA models. In our numerical example, the robust efficiency of each DMU from $IRC[\varepsilon]$ satisfies the constraints with no exception and the robust efficiency from $RC[\varepsilon,\Omega]$ satisfies the constraints with very high probability. Therefore, both of the two approaches have essentially the same solutions.

Comparing the results from different robust DEA models and nominal CCR model, we could find that, since there is no perturbation and perturbations appear in input and output data, the average efficiency scores decrease from 0.837 to 0.685. The rankings from CCR model and robust DEA models have ties between DMU D, F, and J, which are evaluated as efficient DMUs under nominal CCR model and as the most efficient DMUs which having highest sufficiency scores under robust DEA models. The rankings from these robust DEA models are coincident with that of the nominal CCR model. Another important thing we find is that, by testing the optimal solution of weights reported by these robust DEA approaches and computing the efficiency scores under the same random perturbations on input and output data we used in the analysis of section 2, all of these weights are feasible and the related perturbed efficiency scores satisfy the constraint that the efficiency scores should be no more than one.

The results of the numerical example could be summarized as follows: the uncertainty immune efficiency scores and related rankings for decision making units, i.e., the reliable solutions of robust DEA models, do exist, and the “price” of immune to uncertainty in terms of the efficiency scores is not high (the largest reduction of the average efficiency score is 0.15). Moreover, there is no “price” of immune to uncertainty in terms of the rankings, i.e., although the efficiency scores decreasing, the rankings of DMUs are not changed by adopting robust DEA approach. Therefore, by applying the robust DEA approach in this example, we gain a lot in the ability of getting evaluation results withstanding data uncertainty, while lose a little in the optimality of the efficiency scores, as well as lose nothing in the consistency of the ranking of DMUs compared with the nominal problem.

6. Conclusion

In this paper, we analyze the impact of input and output data uncertainty on the result of DEA and propose several robust DEA models based on the newly developed robust optimization approaches. These models could deal with uncertain input and output entries, and provide more reliable evaluation results. The robust DEA model presented in this paper where the output data are subject to uncertainties is based on input-oriented CCR model. On the contrary, we present an output-oriented CCR based robust DEA model where the input data are subject to uncertainties. Furthermore, two different forms of uncertainties, random symmetric and unknown-but-bounded uncertainty, are considered in our robust DEA models. We implement these models in a numerical example and the efficiency scores and rankings of the robust DEA models and standard CCR model are compared. The results from our analysis and the numerical example indicate that considering the input and output data uncertainties when applying DEA approach is very important, and using robust DEA approach could be more reliable for efficiency evaluation and ranking in MCDM problems. Since we consider the uncertainties affecting output and input data separately in this study, further research should focus on how to deal with uncertainties appearing in output and input data simultaneously.

Acknowledgment

The authors thank the reviewers of the paper and are grateful to Professor Xin Chen and Dr. Yuhan Zhang for their valuable comments and suggestions on an earlier version of the paper. This paper was finished while Ke Wang was visiting University of Illinois at Urbana-Champaign.

Reference

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

[2] W. W. Cooper, L. M. Seiford, J. Zhu. *Handbook on Data Envelopment Analysis*. Boston: Kluwer Academic Publishers, 2004.

[3] W. D. Cook, L. M. Seiford. Data envelopment analysis (DEA) - Thirty years on. *European Journal of Operational Research*, 2009, 192(1): 1-17.

[4] T. J. Stewart. Relationship between Data Envelopment Analysis and Multicriteria Decision Analysis. *Journal of the Operational Research Society*, 1996, 47(5): 654-665.

[5] J. Sarkis. A comparative analysis of DEA as a discrete alternative multiple criteria decision tool. *European Journal of Operational Research*, 2000, 123(3): 543-557.

[6] W. D. Cook, M. Kress, L. Seiford. On the use of ordinal data in data envelopment analysis. *Journal of Operational Research Society*, 1993, 44(2): 133-140.

[7] W. D. Cook, M. Kress, L. Seiford. Data envelopment analysis in the presence of both quantitative and qualitative factors. *Journal of Operational Research Society*, 1996, 47(7): 945-953.

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

[9] W. W. Cooper, K. S. Park, G. Yu. An illustrative application of IDEA (Imprecise Data Envelopment Analysis) to a Korean mobile telecommunication company. *Operations Research*, 2001, 49(6): 807-820.

[10] J. Zhu. Imprecise data envelopment analysis (IDEA): A review and improvement with an application. *European Journal of Operational Research*, 2003, 144(3): 513-529.

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

[12] A. Ben-Tal, A. Nemirovski. Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical Programming*, 2000, 88(3): 411-421.

[13] A. Ben-Tal, A. Nemirovski. Robust convex optimization. *Mathematics of Operations Research*, 1998, 23(4): 769-805.

[14] A. Ben-Tal, A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letter*, 1999, 25(1): 1-13.

[15] D. Bertsimas, M. Sim. Robust discrete optimization and network flows. *Mathematical Programming Series B*, 2003, 98(1-3): 49-71.

[16] D. Bertsimas, M. Sim. The price of robustness. *Operations Research*, 2004, 52(1): 35-53.

[17] D. Bertsimas, M. Sim. Tractable approximations to robust conic optimization problems. *Mathematical*

Programming Series B, 2006, 107(1-2): 5-36.

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

[19] S. J. Sadjadi, H. Omrani. A bootstrapped robust data envelopment analysis model for efficiency estimating of telecommunication companies in Iran. *Telecommunications Policy*, 2010, 34(4): 221-232.

[20] A. Charnes, L. Neralic. Sensitivity analysis of the additive model in data envelopment analysis. *European Journal of Operational Research*, 1990, 48(3): 332-341.

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

[22] A. Charnes, J. Rousseau, J. Semple. Sensitivity and stability of efficiency classifications in data envelopment analysis. *Journal of Productivity Analysis*, 1996, 7(1): 5-18.

[23] J. Zhu. Super-efficiency and DEA sensitivity analysis. *European Journal of Operational Research*, 2001, 129(2): 443-455.

[24] J. Zhu. Erratum to "Super-efficiency and DEA sensitivity analysis". *European Journal of Operational Research*, 2007, 179(1): 277.

[25] W. W. Cooper, S. Li, L. M. Seiford, Kaoru Tone, R. M. Thrall and J. Zhu. Sensitivity and Stability Analysis in DEA: Some Recent Developments. *Journal of Productivity Analysis*, 2001, 15(3): 217-246.

[26] A. L. Soyster. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 1973, 21(5): 1154-1157.