

## Research Article

# Efficiency in the Worst Production Situation Using Data Envelopment Analysis

**Md. Kamrul Hossain,<sup>1</sup> Anton Abdulbasah Kamil,<sup>1</sup> Adli Mustafa,<sup>2</sup> and Md. Azizul Baten<sup>3</sup>**

<sup>1</sup> School of Distance Education, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia

<sup>2</sup> School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia

<sup>3</sup> Department of Decision Science, School of Quantitative Sciences, Universiti Utara Malaysia, 01610 Sintok, Kedah, Malaysia

Correspondence should be addressed to Md. Kamrul Hossain; [kamrulstat@gmail.com](mailto:kamrulstat@gmail.com)

Received 28 June 2012; Accepted 10 January 2013

Academic Editor: Graham Wood

Copyright © 2013 Md. Kamrul Hossain et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Data envelopment analysis (DEA) measures relative efficiency among the decision making units (DMU) without considering noise in data. The least efficient DMU indicates that it is in the worst situation. In this paper, we measure efficiency of individual DMU whenever it losses the maximum output, and the efficiency of other DMUs is measured in the observed situation. This efficiency is the minimum efficiency of a DMU. The concept of stochastic data envelopment analysis (SDEA) is a DEA method which considers the noise in data which is proposed in this study. Using bounded Pareto distribution, we estimate the DEA efficiency from efficiency interval. Small value of shape parameter can estimate the efficiency more accurately using the Pareto distribution. Rank correlations were estimated between observed efficiencies and minimum efficiency as well as between observed and estimated efficiency. The correlations are indicating the effectiveness of this SDEA model.

## 1. Introduction

Producer performance is influenced by three phenomena: efficiency with which management organizes production, the effect of environmental factors, and random error [1]. If all of these phenomena are influenced positively to production, then it is the best situation for production of the DMU, and if all of these phenomena are influenced the production negatively, then it is the worst situation. Several models of single stage, or multistage have been proposed to incorporate environmental factors in DEA. Banker and Morey [2] proposed a single-stage DEA method for environmental variables. An obstacle of this method is that the direction of impact on production of the environmental factors must be known in advance. In some research, a two-stage approach is used to describe the effect of environmental factors. First stage calculates efficiency from DEA based on inputs and outputs and a second-stage regression analysis tries to explain variation in first-stage efficiency score. A step further, development of the two-stage method was done by McCarty and Yaisawarng

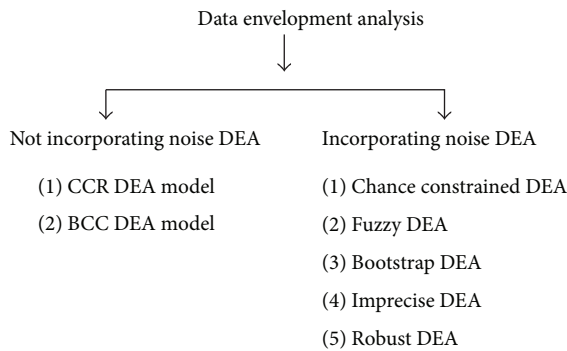
[3] and Bhattacharyya et al. [4] by using the second-stage regression residuals to adjust the first-stage efficiency scores. Input oriented DEA to inputs and environmental factors or output oriented DEA to outputs and the environmental factors applied in the first stage. Then, the inputs or outputs were replaced by their residual projections. In the second stage of this method again applied DEA to expanded data set consisting of the originally efficient observations, the originally inefficient observations, and the radial projections of the originally inefficient observations. Camanho et al. [5] suggested a method of DEA considering internal and external nondiscretionary factors. In this method, they generalized models of Banker and Morey [2] and their extension to both nondiscretionary inputs and outputs by Golany and Roll [6], the model of Ruggiero [7] and its extension described by Ruggiero [8]. Lotfi et al. [9] measured the relative efficiency of decision making units with nondiscretionary inputs and interval discretionary data. They showed upper bound by assuming the best performance of a DMU against the worst performance of the rest of other DMU and similarly lower

bound by assuming the worst performance of a DMU against the best performance of the rest of other DMU. Sadjadi and Omrani [10] presented DEA model with uncertain data. Simar and Wilson [11–13] proposed bootstrap algorithm to study statistical properties of DEA models. In bootstrap method, it is difficult to find an appropriate value of a smoothing parameter. Another problem of this method is the large number of iterations. Entani et al. [14] introduce a method to calculate interval efficiency. Kao and Liu [15] used the interval data to measure the interval efficiency. Our method will provide an interval efficiency, the novelty we will be taking into consideration is that the study approach lies with an assumption of noise data in the existing DEA method and bounded Pareto distribution is used to estimate efficiency scores at a time.

The main focus of this paper is to develop a method for DEA including noise in data. Toward the end, we formulate a method of SDEA using BCC model. The objective of the paper is to measure minimum efficiency of DMUs. The rest of the study is organized as follows. Section 2 deals with the background of this study. Section 3 develops a methodology for SDEA and characteristics of bounded Pareto distribution. Section 4 presents and discusses an empirical example. And conclusions are included in the final section.

## 2. Background

Data envelopment analysis is a nonparametric method in operations research for measure production efficiency of DMUs. It is a popular method to measure performance of DMUs. Farrell [16] was motivated to develop better methods and models for evaluating productivity performance:



The initial DEA model, as presented by Charnes et al. [17] and built on the earlier work of Farrell [16], is known as CCR model. It is a constant returns to scale model. CCR model is as follow:

$$\begin{aligned}
 & \text{Max}_{\theta, \lambda} \quad \theta_i \\
 & \text{subject to} \quad x_i \geq \sum_{i=1}^n \lambda_i x_i, \\
 & \quad \quad \quad \theta y_i \leq \sum_{i=1}^n \lambda_i y_i,
 \end{aligned} \tag{1}$$

where  $x_i \geq 0$  is the input and  $y_i \geq 0$  is the output for the  $i$ th firm.  $\lambda_i \geq 0$  represents intensity of variable.  $\theta_i$  is the efficiency of the  $i$ th firm.

In the CCR model, if a constraint  $\sum_{i=1}^n \lambda_i = 1$  is adjoined, then it is known as BCC models [18].

Chance Constraint DEA (CCDEA) [19] considers two sets of constraints. There is a set of chance constraints that probability of efficient of all DMUs is only 0.05 or less. The CCDEA model is

$$\begin{aligned}
 & \text{Min} \quad \theta \\
 & \text{subject to} \quad \text{Prob} \left( \sum \lambda_i y_i \leq y_i \right) \leq 0.05, \\
 & \quad \quad \quad \theta x_i \geq \sum \lambda_i x_i, \quad \theta \text{ unrestricted in sign.}
 \end{aligned} \tag{2}$$

Fuzzy DEA is an extension of the CCR DEA model incorporating fuzzy numbers [20]. The Fuzzy DEA model is

$$\begin{aligned}
 & \text{Max} \quad \mu^t Y_0 \\
 & \text{subject to} \quad v^t X_0 \approx \bar{I}, \\
 & \quad \quad \quad \mu^t Y_j \leq v^t X_j \quad (j = 1, 2, \dots, n), \quad \mu \geq 0, \quad v \geq 0,
 \end{aligned} \tag{3}$$

where  $X_j$  and  $Y_j$  are  $s$ -dimensional  $L$ - $L$  fuzzy input vector and  $m$  dimensional fuzzy output vector of  $j$ th DMU.  $v$  and  $\mu$  are coefficient vectors of  $X_j$  and  $Y_j$ , respectively. The index  $0$  indicates the evaluated DMU.

The bootstrap method introduced by Efron [21] is used in Bootstrap DEA [11–13]. Algorithm of Bootstrap DEA is as follows [11].

*Step 1.* Transform the input-output vectors using the original efficiency estimates  $\{\hat{\theta}_i, i = 1, 2, \dots, n\}$  as  $(\hat{x}_i, y_i) = (x_i, \hat{\theta}_i, y_i)$ .

*Step 2.* Given a set of estimated efficiencies  $\{\hat{\theta}_i\}$ , use  $h = 0.90n^{-1} \min\{\hat{\sigma}, R/1.34\}$  to obtain bandwidth parameter  $h$ .  $\hat{\sigma}$  denotes the standard deviation estimate of efficiency estimates and  $R$  denotes the interquartile range of empirical distribution, respectively.

*Step 3.* Generate  $\{\delta^-\}$  by resampling, with replacement, from the empirical distribution of estimated efficiencies.  $\delta^-$  is the nonsmoothed resample of original efficiencies.

*Step 4.* Generate sequence  $\{\tilde{\delta}^-\}$  using

$$\tilde{\delta}^- = \begin{cases} \delta^- + h\epsilon^-, & \text{if } \delta^- + h\epsilon^- \leq 1, \\ 2 - (\delta^- + h\epsilon^-), & \text{otherwise.} \end{cases} \tag{4}$$

*Step 5.* Generate the smoothed pseudo efficiencies  $\{\gamma^-\}$  using  $\gamma^- = \bar{\delta} + (\delta^- - \bar{\delta})/\sqrt{1 + h^2/\hat{\sigma}^2}$ .  $\gamma^-$  is the smoothed resample efficiencies,  $\bar{\delta} = (\sum \delta^-)/n$ .

*Step 6.* Let the bootstrap pseudo data be given by  $(\bar{x}_i, \bar{y}_i) = (\hat{x}_i/\gamma^-, y_i)$ .

*Step 7.* Estimate the bootstrap efficiencies using the pseudo data and the linear program  $\hat{\theta} = \min\{\theta : y \leq Yz, \theta x \geq Xz, \sum z = 1, z \in R_+\}$ .

*Step 8.* Repeat Steps 2 to 7  $B$  times to create a set of  $B$  firm specific bootstrapped efficiency estimates.

Imprecise DEA (IDEA) model is an extension of the CCR model incorporating imprecise data information and this DEA is nonlinear and nonconvex program [22]. Sometimes, outputs and inputs are imprecise data in the forms of bounded data, ratio bounded data, weak ordinal data, and strong ordinal data [23, 24]. If data follow any of the above forms, then DEA is IDEA as follows:

$$\begin{aligned} \text{Max} \quad & \pi_o = \sum_r \mu_r y_{r0} \\ \text{subject to} \quad & \sum_r \mu_r y_{rj} - \sum_i w_i x_{ij} \leq 0, \\ & \sum_i w_i x_{ij} = 1, \\ & (x_{ij}) \in \Theta_i^-, \quad (y_{rj}) \in \Theta_i^+, \quad \mu_r, w_i \geq 0, \end{aligned} \tag{5}$$

where  $(x_{ij}) \in \Theta_i^-$  and  $(y_{rj}) \in \Theta_i^+$  represent any form of imprecise data.  $\mu_r$  and  $w_i$  are the weights of output and input, respectively.

Robust structure proposed by Ben-Tal and Nemirovski [25] and Bertsimas and Sim [26] is used in DEA by Sadjadi and Omrani [10]. They mentioned it as robust DEA. The robust DEA model is expressed as follows [10]:

$$\begin{aligned} \text{Max} \quad & \hat{z}_0 \\ & \sum_i v_i x_{i0} = 1, \\ & \sum_r u_r y_{r0} - \hat{z}_0 - \lambda_0 p_0 - \sum_{j \in J_i} q_{r0} \geq 0, \\ & \sum_i v_i x_{ij} - \sum_r u_r y_{rj} - \lambda_j p_j - \sum_{j \in J_i} q_{rj} \geq 0, \\ & j = 1, 2, \dots, n \\ & p_j + q_{rj} \geq e y_{rj} t_r, \quad \forall r, j \\ & -t_r \leq u_r \leq t_r, \quad \forall r \\ & p_j, q_{rj} \geq 0, \quad u_r, v_i \geq 0, \end{aligned} \tag{6}$$

where  $z$  is the efficiency,  $y_{rj}$  and  $x_{ij}$  are the  $r$ th output and  $i$ th input of  $j$ th DMU.  $\lambda_j$  is the budget of uncertainty for constraint  $j$ .  $u_r$  and  $v_i$  are the weights of output and input, respectively.  $e$  is the utmost probability to violate constraint.  $p_j$  and  $q_{rj}$  are the dual variables.

In this paper we describe a methodology based on BCC model incorporating noise in data.

### 3. Method

We propose a method of DEA to know the performance of a DMU in the worst situation. In the first stage, we apply DEA with input and output data to know the efficiency level of DMU. Thereafter, in the second stage, we maximize the gap between frontier output and observed output to assess what will be the maximum loss if the DMU is in the worst situation. In the final stage, we will again use DEA for the worst condition of individual DMU output with observed output of other DMUs and will determine the efficiency level.

#### 3.1. DEA Incorporating Error

*Stage 1.* DEA is familiar with input-oriented method and output-oriented method. We choose the output oriented DEA method of Banker et al. [18] which can be expressed as the linear programming problem as follows:

$$\begin{aligned} \text{Max}_{\theta, \lambda} \quad & \theta_i^a \\ \text{subject to} \quad & x_i \geq \sum_{i=1}^n \lambda_i x_i, \\ & \theta_i^a y_i \leq \sum_{i=1}^n \lambda_i y_i, \\ & \sum_{i=1}^n \lambda_i = 1, \end{aligned} \tag{7}$$

where  $\theta_i^a$  is a function of output  $y$ . This LP problem is solved  $n$  times and  $n$  is the number of firms.

*Stage 2.* Model can be classified into two groups as neo-classical model and frontier model [27]. This classification is depending on the interpretation of the deviation terms  $\varepsilon_i \in \mathfrak{R}$ . Considering the assumption of neoclassical model, all firms are efficient, and deviations  $\varepsilon_i \in \mathfrak{R}$  are seen as random, uncorrelated noise terms that satisfy the Gauss-Markov assumptions. But in the frontier models, all deviations from the frontier are attributed to inefficiency which implies that,  $\varepsilon_i \leq 0$ , for all  $i = 1, 2, \dots, n$  [28]. A function

$$y_i = g(\cdot) + \varepsilon_i \tag{8}$$

Is a frontier model if  $\varepsilon_i \in \mathfrak{R}$  are interpreted as composite error terms that include both inefficiency and noise components where  $g(\cdot)$  is the production frontier Ainger et al. [29]. Then, output due to environmental factor and inefficiency  $\varepsilon_i$  is  $\varepsilon_i = g(\cdot) - y_i \geq 0$  since frontier output is always greater than or equal to the observed output. Now, to maximize the output  $\varepsilon_i$ , we use output-oriented DEA method. The method

can be expressed as the linear programming problem as follows:

$$\begin{aligned}
 & \text{(b)} \\
 & \text{Max}_{\lambda} \quad \theta_i^b \\
 & \text{subject to} \quad x_i \geq \sum_{i=1}^n \lambda_i x_i, \\
 & \quad \quad \quad \varepsilon_i \theta_i^b \leq \sum_{i=1}^n \lambda_i \varepsilon_i, \\
 & \quad \quad \quad \sum_{i=1}^n \lambda_i = 1,
 \end{aligned} \tag{9}$$

where  $\theta_i^b$  is a function of  $\varepsilon$ .

*Stage 3.* Let us suppose that the maximized quantity of  $\varepsilon_i$  is  $\varepsilon_{i\max}$ . To calculate efficiency of  $i$ th firm in worst production situation take into account the output of the  $i$ th firm is  $g(\cdot) - \varepsilon_{i\max}$  whereas inputs of all forms and output of other firms are observed data.

This can be mathematically shown as the following:

$$\begin{aligned}
 & \text{(c)} \\
 & \text{Max}_{\theta, \lambda} \quad \theta_i^c \\
 & \text{subject to} \quad x_i \geq \sum_{i=1}^n \lambda_i x_i, \\
 & \quad \quad \quad \left( g(\cdot) - \varepsilon_{i\max} \right) \theta_i^c \leq \lambda_i \left( g(\cdot) - \varepsilon_{i\max} \right) \\
 & \quad \quad \quad + \sum_{j=1, j \neq i}^n \lambda_j y_j, \\
 & \quad \quad \quad \sum_{i=1}^n \lambda_i = 1,
 \end{aligned} \tag{10}$$

where  $\theta_i^c$  is a function of frontier output and output due to environmental factor and shows the minimum level of efficiency of the  $i$ th firm.

*3.2. Estimation of DEA Efficiency.* Using the above methodology, we get minimum efficiency of DMU and the highest known efficiency is 1. Pareto distribution is usually used to describe the allocation of wealth among individuals, but in this paper we use bounded Pareto distribution because of its properties. Pareto distribution is applicable if the range of a variable is a certain value to infinity. The range of efficiency of BCC model to maximize the output is a certain value to 1. This situation justifies the use of bounded Pareto distribution.

The probability density function of bounded Pareto distribution is

$$f(x) = \left( \frac{\alpha L^\alpha x^{-\alpha-1}}{1 - (L/H)^\alpha} \right), \tag{11}$$

where  $L \leq x \leq H$ ,  $\alpha > 0$  with.

Mean

$$\frac{L^\alpha}{1 - (L/H)^\alpha} * \left( \frac{\alpha}{\alpha - 1} \right) * \left( \frac{1}{L^{\alpha-1}} - \frac{1}{H^{\alpha-1}} \right), \quad \alpha \neq 1. \tag{12}$$

Variance

$$\frac{L^\alpha}{1 - (L/H)^\alpha} * \left( \frac{\alpha}{\alpha - 2} \right) * \left( \frac{1}{L^{\alpha-2}} - \frac{1}{H^{\alpha-2}} \right), \quad \alpha \neq 2. \tag{13}$$

Then, the probability density function of bounded Pareto distribution for efficiency is  $f(\theta) = (\alpha(\theta_i^c)^\alpha \theta^{-\alpha-1} / (1 - (\theta_i^c)^\alpha))$  as upper bound for efficiency is 1, where  $\theta_i^c$  is the minimum efficiency. Using pdf  $f(\theta)$ , we generate random number of  $\theta$ .

### 4. Empirical Example

Table 1 shows the efficiency from observed outputs and inputs mentioned as observed efficiency. Firm 1, Firm 2, Firm 4, Firm 12, Firm 13, and Firm 20 are showing to be perfectly efficient among the 20 firms. Firm 16 has the less efficiency among the firms. Frontier output is the projected output obtained from the software DEAP version 2.1. Gap between observed and frontier outputs is the amount of output lost at DMU because of inefficiency. Inefficiency is influenced by some environmental factors. If a DMU can control these factors, then it will be efficient DMU. Maximum loss of output is calculated from Stage 2 of methodology. We calculated minimum output as

$$\begin{aligned}
 & \text{minimum output} \\
 & = (\text{frontier output} - \text{maximum loss of output}).
 \end{aligned} \tag{14}$$

Minimum output and observed output are found the same for Firm 7 and Firm 16 since from the beginning these firms are in the worst position among firms. Efficiency in the worst situation is the efficiency of individual firm when it will follow the situation of Firm 7 and Firm 16. Maximum loss is the loss of output due to completely negative effect of unexplained factors on production. Maximum loss of firm 1 will be 273 units if all unexplained factors effect as they effect in firm 7 or firm 16 because firm 7 and firm 16 are in worst production situation.

To calculate the efficiency in the worst situation, we consider the minimum output of the firm with observed output of other firms. For efficiency in the worst situation of Firm 1, use the minimum output 2202 unit with observed output of Firm 2 to Firm 20 as output and observed inputs of all firms. Observed efficiency of Firm 16 (0.823) is the lowest. And in the worst situation, Firm 18 and Firm 19 are also showing the lowest efficiency level (0.823) with Firm 16.

Table 2 shows results of estimated efficiency from bounded Pareto distribution. We arbitrarily choose 0.5, 1.5, 2.5, 3.5, and 4.5 as shape value, efficiency in the worst situation as lower bound, and 1 as highest value to draw 10 random numbers. Mean of random numbers is the estimated efficiency. Bias is the difference between observed efficiency and estimated efficiency. In Robust Bootstrap DEA, original

TABLE 1: Efficiency level and projected output using observed inputs and outputs.

Firm number	Observed output	Frontier output	*Gap between observed and frontier outputs	Observed efficiency	Maximum loss of output	Minimum output	Efficiency in worst situation
1	2475	2475	1	1	273	2202.00	0.968
2	2261	2261	1	1	271.40	1989.60	0.928
3	2179	2235.66	56.66	0.975	293.84	1941.82	0.869
4	2075	2075	1	1	288.56	1786.44	0.905
5	1850.2	1925.35	75.15	0.961	267.71	1657.65	0.861
6	1790.7	1993.6	202.9	0.898	276.31	1717.29	0.861
7	1676	1994.42	318.42	0.840	318.42	1676.00	0.840
8	1870	1903.32	33.32	0.982	251.11	1652.21	0.868
9	1874.6	1999.12	124.52	0.938	285.71	1713.41	0.857
10	1616.9	1892.05	275.15	0.855	299.99	1592.06	0.841
11	1734	1939.21	205.21	0.894	249.59	1689.62	0.871
12	1916	1916	1	1	167.91	1748.09	0.983
13	1808	1808	1	1	156.56	1651.44	0.924
14	1850.7	1917.41	66.71	0.965	229.17	1688.24	0.880
15	1831.8	1921.06	89.26	0.954	259.28	1661.78	0.865
16	1500	1821.97	321.97	0.823	321.97	1500.00	0.823
17	1745	1839.02	94.02	0.949	318.91	1520.11	0.827
18	1512.325	1610.94	98.62	0.939	284.68	1326.26	0.823
19	1506.852	1633.42	126.57	0.923	288.65	1344.77	0.823
20	1894.557	1894.56	1	1	298	1596.56	0.948

\*We use one instead of zero in the gap between observed and frontier outputs which will not influence the result.

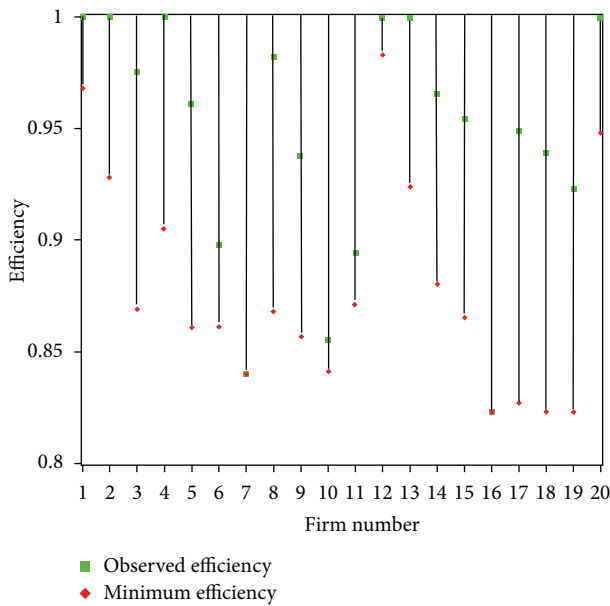


FIGURE 1: Efficiency interval of firms.

efficiency is greater than the estimated efficiency [30], but in our case the bias is both positive and negative. Mean of bias is the lowest for shape 0.5. Mean of bias is increasing with increased the value of shape.

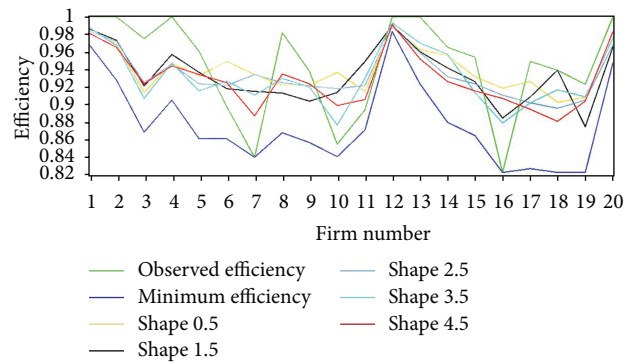


FIGURE 2: Observed, minimum and estimated efficiency comparison.

Figure 1 presents interval efficiency of firms. Among the perfectly efficient firm, firm 12 is highly consistent and firm 4 is most inconsistent. That is Firm 12 will be less affected in worst situation. Range of efficiency is almost the same for Firm 7, Firm 10, Firm 16, Firm 17, Firm 18, and Firm 19.

Figure 2 indicates the comparison among the observed, minimum, and estimated efficiencies. Estimated efficiency is following almost the same pattern for different shape parametric values. Estimated efficiency curve is found smooth and it might be due to a parametric estimation. Observed and minimum efficiency lines are showing the same point for

TABLE 2: Bias in estimated efficiency using bounded Pareto distribution.

Firm	Lower bound	Observe efficiency	Shape 0.5		Shape 1.5		Shape 2.5		Shape 3.5		Shape 4.5	
			Estimate efficiency	Bias	Estimate efficiency	Bias	Estimate efficiency	Bias	Estimate efficiency	Bias	Estimate efficiency	Bias
1	0.968	1	0.987	0.013	0.986	0.014	0.984	0.016	0.988	0.012	0.981	0.019
2	0.928	1	0.969	0.031	0.973	0.027	0.972	0.028	0.967	0.033	0.965	0.035
3	0.869	0.975	0.914	0.061	0.921	0.054	0.925	0.05	0.907	0.068	0.924	0.051
4	0.905	1	0.947	0.053	0.957	0.043	0.946	0.054	0.947	0.053	0.944	0.056
5	0.861	0.961	0.933	0.028	0.937	0.024	0.924	0.037	0.916	0.045	0.934	0.027
6	0.861	0.898	0.949	-0.051	0.918	-0.02	0.922	-0.024	0.927	-0.029	0.925	-0.027
7	0.84	0.84	0.934	-0.094	0.915	-0.075	0.934	-0.094	0.911	-0.071	0.887	-0.047
8	0.868	0.982	0.922	0.06	0.913	0.069	0.925	0.057	0.93	0.052	0.935	0.047
9	0.857	0.938	0.922	0.016	0.904	0.034	0.921	0.017	0.92	0.018	0.924	0.014
10	0.841	0.855	0.937	-0.082	0.914	-0.059	0.918	-0.063	0.877	-0.022	0.899	-0.044
11	0.871	0.894	0.915	-0.021	0.949	-0.055	0.922	-0.028	0.93	-0.036	0.906	-0.012
12	0.983	1	0.989	0.011	0.99	0.01	0.989	0.011	0.993	0.007	0.991	0.009
13	0.924	1	0.963	0.037	0.96	0.04	0.959	0.041	0.97	0.03	0.952	0.048
14	0.88	0.965	0.956	0.009	0.942	0.023	0.932	0.033	0.957	0.008	0.926	0.039
15	0.865	0.954	0.932	0.022	0.926	0.028	0.924	0.03	0.914	0.04	0.916	0.038
16	0.823	0.823	0.919	-0.096	0.885	-0.062	0.911	-0.088	0.879	-0.056	0.907	-0.084
17	0.827	0.949	0.926	0.023	0.909	0.04	0.902	0.047	0.902	0.047	0.895	0.054
18	0.823	0.939	0.903	0.036	0.939	0	0.896	0.043	0.917	0.022	0.881	0.058
19	0.823	0.923	0.908	0.015	0.875	0.048	0.905	0.018	0.909	0.014	0.904	0.019
20	0.948	1	0.976	0.024	0.966	0.034	0.968	0.032	0.976	0.024	0.983	0.017
Mean bias			0.005		0.011		0.011		0.013		0.016	

TABLE 3: Rank correlation matrix of minimum efficiency, observed efficiency, and estimated efficiency for different values of shape parameter.

		Minimum efficiency	Observed efficiency	Estimated efficiency				
				Shape 0.5	Shape 1.5	Shape 2.5	Shape 3.5	Shape 4.5
Estimated efficiency	Minimum efficiency	1						
	Observed efficiency	0.827**	1					
	Shape 0.5	0.760**	0.601**	1				
	Shape 1.5	0.867**	0.740**	0.701**	1			
	Shape 2.5	0.903**	0.761**	0.805**	0.785**	1		
	Shape 3.5	0.864**	0.764**	0.703**	0.833**	0.787**	1	
	Shape 4.5	0.882**	0.842**	0.753**	0.697**	0.836**	0.828**	1

\*\*Significant at 1 percent level of significance.

Firm 7 and Firm 16. Again minimum efficiency in the worst situation occurred in cases of Firm 16, Firm 18 and Firm 19.

Correlation (0.842) is high for shape parameter 4.5 between observed and estimated efficiencies. Table 3 presents rank correlations matrix among observed efficiency, minimum efficiency, and estimated efficiency with different values of shape parameter. All the correlations are recorded significantly high at 1 percent level of significance, which indicates the effectiveness of the developed SDEA method. This result supports the result of Yuichiro [31].

Correlation between minimum efficiency and observed efficiency is 0.827. Minimum efficiency is found relatively more correlated with estimated efficiency (0.903) when shape

parameter is 2.5. Though estimated efficiency with shape parameter 0.5 has minimum bias, correlation with minimum efficiency as well as observed efficiency is relatively low.

Figure 3(a) shows the scatter diagram between minimum efficiency and observed efficiency. The efficiency correlation coefficient is calculated as 0.707. Scatter diagram of observed efficiency and estimated efficiency with different values of shape parameter is shown in Figures 3(b)–3(f), and correlation coefficients of efficiency are 0.521, 0.667, 0.592, 0.759, and 0.731 for shape parameters 0.5, 1.5, 2.5, 3.5, and 4.5, respectively. The correlation between observed efficiency and estimated efficiency is increased as the value of shape parameter is increased. All the correlations are significant at 5



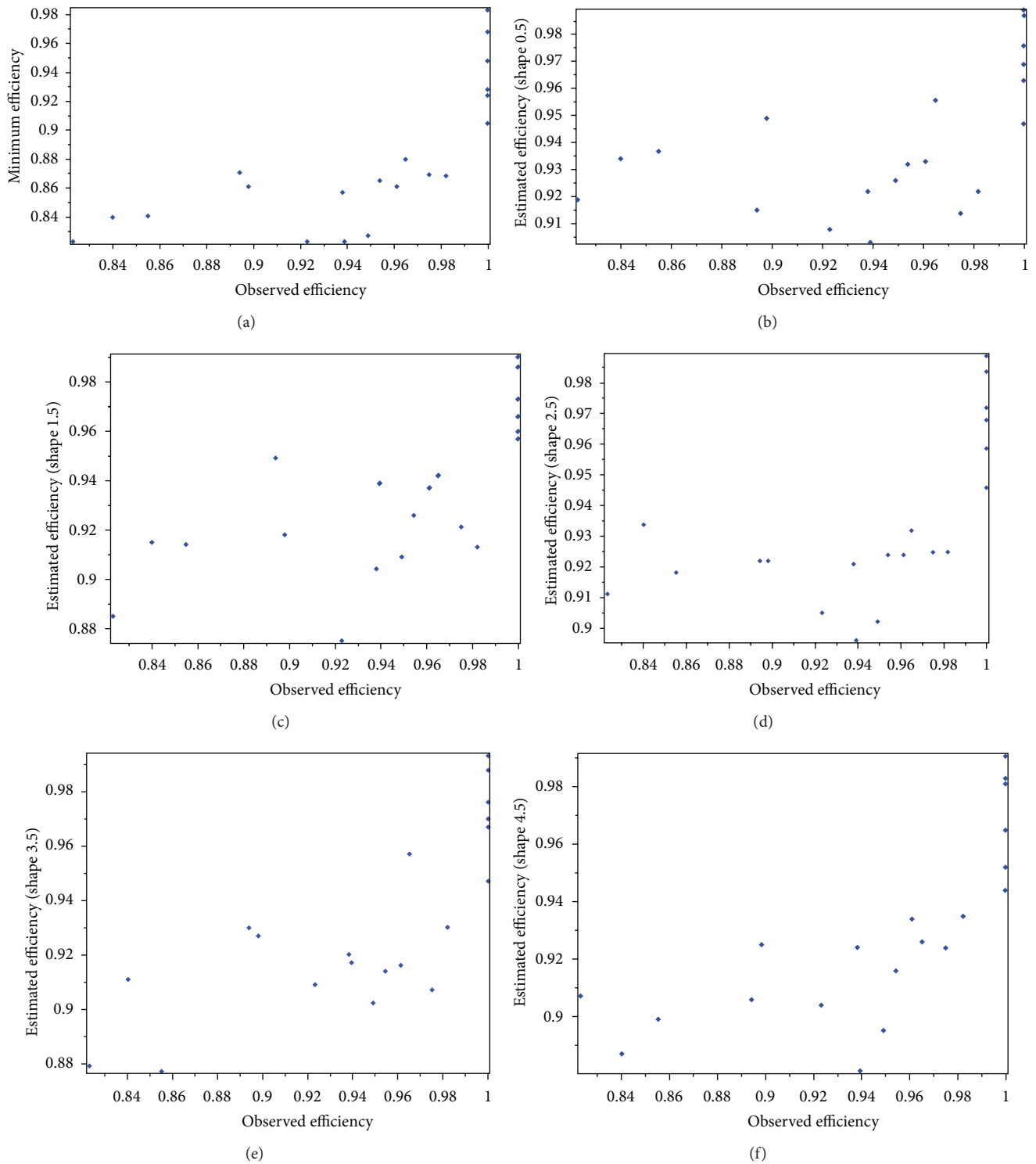


FIGURE 3: (a) Observed and minimum efficiencies. (b), (c), (d), (e), and (f) are observed and estimated efficiencies for different values of shape parameter of farms.

percent level of significance. Estimated efficiency with shape parameter 2.5 has the highest correlation with minimum efficiency. And the lowest correlation between minimum efficiency and estimated efficiency is found when shape parameter is 0.5.

To select the appropriate value of shape parameter, we can set up the following hypotheses.

(Ha) There is significant difference between observed efficiency and estimated efficiency.

TABLE 4: Appropriate shape parameter for estimate efficiency.

Comparison of efficiency score	Paired Differences		P value
	Mean	Standard deviation	
Observe versus shape 0.5	0.00475	0.04831	0.665
Observe versus shape 1.5	0.01085	0.04247	0.267
Observe versus shape 2.5	-0.01085	0.04595	0.304
Observe versus shape 3.5	-0.01295	0.03786	0.143
Observe versus shape 4.5	0.01585	0.03959	0.089

Table 4 shows that there is no significant difference between observed efficiency and estimated efficiencies for shape parameters (0.5, 1.5, 2.5, and 3.5) but there is difference between observed and estimated efficiencies for shape parameter 4.5 at 10% level of significance. In Table 4, the  $P$  value is decreasing for higher value of shape parameter. To estimate efficiency, the shape parameters 0.5, 1.5, 2.5, and 3.5 can be used but the small value of shape parameter is more appropriate.

## 5. Conclusion

This paper presents a method of the stochastic data envelopment analysis which handles data with noise in DEA. The method consists of three stages. In the first stage, we use BCC DEA model to see the gap between observed and frontier production. From the second stage, we can assess minimum production level. In policy implication, it will help to determine the size of inventory. The third stage measures lower bound of efficiency for each DMU. It provides an interval efficiency to use statistical properties. For the small value of shape parameter, mean of bias is also small. Small value of shape parameter is more appropriate to estimate efficiency using the Pareto distribution. The result shows an almost similar pattern of observed efficiency, minimum efficiencies and estimated efficiencies. Significantly high correlations in between observed and minimum efficiency as well as between observed and estimated efficiency are recorded. Rank correlations are showing the effectiveness of the developed SDEA method.

## References

- [1] H. O. Fried, C. A. K. Lovell, S. S. Schmidt, and S. Yaisawarng, "Accounting for environmental effects and statistical noise in data envelopment analysis," *Journal of Productivity Analysis*, vol. 17, no. 1-2, pp. 157-174, 2002.
- [2] R. D. Banker and R. C. Morey, "Efficiency analysis for exogenously fixed inputs and outputs," *Operations Research*, vol. 34, no. 4, pp. 513-521, 1986.
- [3] T. McCarty and S. Yaisawarng, "Technical efficiency in New Jersey school districts," in *The Measurement of Productive Efficiency*, H. Fried, C. Lovell, and S. Schmidt, Eds., pp. 271-287, Oxford University Press, 1993.
- [4] A. Bhattacharyya, A. K. Lovell, and P. Sahay, "The impact of liberalization on the productive efficiency of Indian commercial banks," *European Journal of Operational Research*, vol. 98, no. 2, pp. 332-345, 1997.
- [5] A. S. Camanho, M. C. Portela, and C. B. Vaz, "Efficiency analysis accounting for internal and external non-discretionary factors," *Computers and Operations Research*, vol. 36, no. 5, pp. 1591-1601, 2009.
- [6] B. Golany and Y. Roll, "Some extensions of techniques to handle non-discretionary factors in data envelopment analysis," *Journal of Productivity Analysis*, vol. 4, no. 4, pp. 419-432, 1993.
- [7] J. Ruggiero, "On the measurement of technical efficiency in the public sector," *European Journal of Operational Research*, vol. 90, no. 3, pp. 553-565, 1996.
- [8] J. Ruggiero, "Performance evaluation when non-discretionary factors correlate with technical efficiency," *European Journal of Operational Research*, vol. 159, no. 1, pp. 250-257, 2004.
- [9] F. H. Lotfi, G. R. Jahanshahloo, and M. Esmaili, "Non-discretionary factors and imprecise data in DEA," *International Journal of Math Analysis*, vol. 1, no. 5, pp. 237-246, 2007.
- [10] S. J. Sadjadi and H. Omrani, "Data envelopment analysis with uncertain data: an application for Iranian electricity distribution companies," *Energy Policy*, vol. 36, no. 11, pp. 4247-4254, 2008.
- [11] L. Simar and P. W. Wilson, "Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models," *Management Science*, vol. 44, no. 1, pp. 49-61, 1998.
- [12] L. Simar and P. W. Wilson, "Statistical inference in nonparametric frontier models: the state of the art," *Journal of Productivity Analysis*, vol. 13, no. 1, pp. 49-78, 2000.
- [13] L. Simar and P. Wilson, "A general methodology for bootstrapping non parametric frontier models," *Journal of Applied Statistics*, vol. 27, no. 6, pp. 779-802, 2000.
- [14] T. Entani, Y. Maeda, and H. Tanaka, "Dual models of Interval DEA and its extension to interval data," *European Journal of Operational Research*, vol. 136, no. 1, pp. 32-45, 2002.
- [15] C. Kao and S. 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.
- [16] M. J. Farrell, "The measurement of productive efficiency," *Journal of the Royal Statistical Society A*, vol. 120, pp. 253-289, 1957.
- [17] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429-444, 1978.
- [18] R. D. Banker, A. Charnes, and W. W. Cooper, "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management Science*, vol. 30, no. 9, pp. 1078-1092, 1984.
- [19] C. K. Land, C. A. K. Lovell, and S. Thore, "Chance-constrained data envelopment analysis," *Managerial and Decision Economics*, vol. 14, pp. 541-554, 1993.



- [20] P. Guo and H. Tanaka, "Fuzzy DEA: a perceptual evaluation method," *Fuzzy Sets and Systems*, vol. 119, no. 1, pp. 149–160, 2001.
- [21] B. Efron, "Bootstrap methods: another look at the jackknife," *The Annals of Statistics*, vol. 7, no. 1, pp. 1–26, 1979.
- [22] J. Zhu, "Imprecise data envelopment analysis (IDEA): a review and improvement with an application," *European Journal of Operational Research*, vol. 144, no. 3, pp. 513–529, 2003.
- [23] W. W. Cooper, K. S. Park, and G. Yu, "Idea and AR-IDEA: models for dealing with imprecise data in DEA," *Management Science*, vol. 45, no. 4, pp. 597–607, 1999.
- [24] S. Kim, C. Park, and K. Park, "An application of data envelopment analysis in telephone offices evaluation with partial data," *Computers and Operations Research*, vol. 26, no. 1, pp. 59–72, 1999.
- [25] A. Ben-Tal and A. Nemirovski, "Robust solutions of Linear Programming problems contaminated with uncertain data," *Mathematical Programming B*, vol. 88, no. 3, pp. 411–424, 2000.
- [26] D. Bertsimas and M. Sim, "The price of robustness," *Operations Research*, vol. 52, no. 1, pp. 35–53, 2004.
- [27] T. Kuosmanen and M. Fosgerau, "Neoclassical versus frontier production models? Testing for the skewness of regression residuals," *Scandinavian Journal of Economics*, vol. 111, no. 2, pp. 351–367, 2009.
- [28] T. Kuosmanen and A. L. Johnson, "Data envelopment analysis as nonparametric least-squares regression," *Operations Research*, vol. 58, no. 1, pp. 149–160, 2010.
- [29] D. J. Aigner, C. A. K. Lovell, and P. Schmidt, "Formulation and estimation of stochastic frontier production function models," *Journal of Econometrics*, vol. 17, no. 1, pp. 21–37, 1977.
- [30] S. J. Sadjadi and H. Omrani, "A bootstrapped robust data envelopment analysis model for efficiency estimating of telecommunication companies in Iran," *Telecommunications Policy*, vol. 34, no. 4, pp. 221–232, 2010.
- [31] Y. Yoshida, "Endogenous-weight TFP measurement: methodology and its application to Japanese-airport benchmarking," *Transportation Research E*, vol. 40, no. 2, pp. 151–182, 2004.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

