


Summer 7-8-2013

# Stochastic DEA with a Perfect Object and Its Application to Analysis of Environmental Efficiency

Alexander Vaninsky  
*CUNY Hostos Community College*

## How does access to this work benefit you? Let us know!

Follow this and additional works at: [http://academicworks.cuny.edu/ho\\_pubs](http://academicworks.cuny.edu/ho_pubs)

 Part of the [Environmental Indicators and Impact Assessment Commons](#), [Environmental Monitoring Commons](#), [Numerical Analysis and Computation Commons](#), [Other Applied Mathematics Commons](#), and the [Statistical Models Commons](#)

---

### Recommended Citation

Vaninsky, A. Y. (2013). Stochastic DEA with a Perfect Object and Its Application to Analysis of Environmental Efficiency. *American Journal of Applied Mathematics and Statistics*, 1(4), 57-63.

This Article is brought to you for free and open access by the Hostos Community College at CUNY Academic Works. It has been accepted for inclusion in Publications and Research by an authorized administrator of CUNY Academic Works. For more information, please contact [AcademicWorks@cuny.edu](mailto:AcademicWorks@cuny.edu).

# Stochastic DEA with a Perfect Object and Its Application to Analysis of Environmental Efficiency

Alexander Y. Vaninsky \*

Mathematics Department, City University of New York/Hostos Community College, New York, USA  
 \*Corresponding author: [avaninsky@hostos.cuny.edu](mailto:avaninsky@hostos.cuny.edu)

Received January 17, 2013; Revised July 03, 2013; Accepted July 04, 2013

**Abstract** The paper introduces stochastic DEA with a Perfect Object (SDEA PO). The Perfect Object (PO) is a virtual Decision Making Unit (DMU) that has the smallest inputs and greatest outputs. Including the PO in a collection of actual objects yields an explicit formula of the efficiency index. Given the distributions of DEA inputs and outputs, this formula allows us to derive the probability distribution of the efficiency score, to find its mathematical expectation, and to deliver common (group-related) and partial (object-related) efficiency components. We apply this approach to a prospective analysis of environmental efficiency of the major national and regional economies.

**Keywords:** Data Envelopment Analysis, DEA analytical solutions, stochastic DEA with a perfect object, efficiency decomposition, environmental efficiency

**Cite This Article:** Vaninsky, Alexander Y., "Stochastic DEA with a Perfect Object and Its Application to Analysis of Environmental Efficiency." *American Journal of Applied Mathematics and Statistics* 1, no. 4 (2013): 57-63. doi: 10.12691/ajams-1-4-2.

## 1. Introduction

Data Envelopment Analysis (DEA) estimates the relative efficiency of a group of objects referred to as Decision-Making Units (DMUs) that use inputs  $\mathbf{X} = (X_j, j = 1, \dots, r) > 0$  to produce outputs  $\mathbf{Y} = (Y_i, i = 1, \dots, s) > 0$ . It was developed in publications [3], [4], and [5]. DEA combines all of the indicators of each object into an efficiency index  $E$  scaled to an interval  $[0, 1]$ . An object is considered *efficient* if it receives a score equal to 1, and *inefficient* if it receives a score of less than 1. The DEA efficiency measure is based on the efficiency ratio suggested in [7]

$$E = \frac{\sum_{i=1}^s u_i Y_i}{\sum_{j=1}^r v_j X_j} \quad (1)$$

where  $\mathbf{u} = (u_1, \dots, u_s) \geq 0$  and  $\mathbf{v} = (v_1, \dots, v_r) \geq 0$  are the weights assigned to outputs and inputs, respectively. To estimate the weights, DEA sets up a series of optimization problems similar to the following one:

For each DMU $_k$ ,  $k = 1, \dots, N$ , find vectors  $\mathbf{u}_k = (u_{k1}, \dots, u_{ks}) \geq 0$  and  $\mathbf{v}_k = (v_{k1}, \dots, v_{kr}) \geq 0$  such that:

$$\text{Maximize } E_k = \frac{\sum_{i=1}^s u_{ki} Y_{ki}}{\sum_{j=1}^r v_{kj} X_{kj}}, \quad (2)$$

subject to  $\sum_{j=1}^r v_{kj} X_{kj} = 1$ , and  $E_m \leq 1$  for all DMU $_m$ ,  $m = 1, \dots, N$ , in the group with the same weight coefficients  $\mathbf{u}_k = (u_{k1}, \dots, u_{ks})$  and  $\mathbf{v}_k = (v_{k1}, \dots, v_{kr})$ .

Reference [4] showed that maximization of the efficiency ratio (2) is equivalent to solving a linear programming (LP) problem, one for each DMU in a group:

For each DMU $_k$ ,  $k = 1, \dots, N$ ,

Minimize  $\theta$

subject to

$$\sum_{m=1}^N \lambda_{km} X_{mj} \leq \theta_k X_{kj}, \quad j = 1, \dots, r; \quad (3)$$

$$\sum_{m=1}^N \lambda_{km} Y_{mi} \geq Y_{ki}, \quad i = 1, \dots, s;$$

$$\lambda_{km} \geq 0, \quad m = 1, \dots, N.$$

The LP algorithm forms a linear combination of DMUs that outperforms the currently selected DMU by both inputs and outputs. Publications [6,11] provide details and reviews of contemporary DEA.

One of the disadvantages of DEA is that it measures the relative efficiency only. To work around this, some DEA publications suggest a best practice comparison. But all of them remain within the framework of conventional DEA and use the LP algorithm. In publications [17,18,19] explicit formulas for the DEA efficiency scores are obtained, assuming that a Perfect Object (PO) is added to the group of actual DMUs. The PO is a virtual DMU having the smallest inputs and greatest outputs; it serves as a benchmark for efficiency comparisons. The obtained

formulas obviate the need to use the LP algorithm and require only a moderate number of simple operations with ratios of inputs and outputs.

Another disadvantage of conventional DEA is its deterministic nature. Some authors claim that the DEA, unable to deal with stochastic noise in inputs and outputs, cannot distinguish between actual inefficiency and short-lived negative effects or plain bad luck. To address these concerns, a stochastic version of DEA (SDEA) was developed in publications [2,10,15], to name just a few. SDEA is a rapidly developing dimension of DEA, and new approaches, procedures, and algorithms continue to appear.

Conventional SDEA is aiming to find a region of stochastic dominance or to eliminate the stochastic noise. Since DEA with a Perfect Object (DEA PO) furnishes an explicit formula for the efficiency index, it provides new opportunities for stochastic extension of DEA. Firstly, the explicit formula allows for the computation of the probability density function (pdf) of the efficiency index directly. Secondly, it introduces stochasticity unrelated to imprecision in data measurement. Thirdly, it introduces an alternative breakdown of the efficiency index into two components. The first component is related to the efficiency of a group as a whole, the second - to the relative efficiency of an individual DMU in the group. This decomposition stems from the fact that in the framework of stochastic DEA with a perfect object (SDEA PO), an individual DMU may be considered as an occurrence of a random normalized DMU, having all inputs and outputs scaled to the interval of [0,1]. The mathematical expectation of the normalized DMU's efficiency, referred to below as common efficiency, is due to fundamental, systemic factors underlying the group. The difference between the observed and common efficiency is called partial efficiency, which is determined by the factors specific to a particular DMU. This paper develops the above-mentioned SDEA PO approach and demonstrates its application.

The paper is organized as follows. In Section 2, stochastic DEA with a perfect object is developed, and Section 3 presents its application to the analysis of prospective environmental performance of the major economies and regions. The appendix provides a computer program written in the R language, meant to compute the common efficiency.

## 2. Stochastic DEA with a Perfect Object

Consider a group of Decision Making Units  $DMU_k$ ,  $k = 1, \dots, N$  using inputs  $\mathbf{X}_k = (X_{jk}, j = 1, \dots, r) > 0$  to produce outputs  $\mathbf{Y}_k = (Y_{ik}, i = 1, \dots, s) > 0$ . Add to the group a Perfect Object with smallest inputs and greatest outputs, and denoted as  $DMU_0$ . Then, for each  $k = 1, \dots, N$ , we get:

$$\begin{aligned} X_{j0} &\leq X_{jk}, j = 1, \dots, r, \\ Y_{i0} &\geq Y_{ik}, i = 1, \dots, s. \end{aligned} \tag{4}$$

When it is unambiguous, we drop the lower index  $k$  for readability. The values of  $i$  and  $j$  providing maximum values to the ratios  $Y_i/Y_{0i}$  and  $X_{0j}/X_j$  for  $DMU_k$  are labeled as  $i^*$  and  $j^*$ , respectively:

$$\begin{aligned} i^* &= \{i \mid Y_i^*/Y_{0i} = \max_{1 \leq l \leq s} Y_l/Y_{0l}\} \\ j^* &= \{j \mid X_{0j}/X_{j^*} = \max_{1 \leq l \leq r} X_{0l}/X_l\} \end{aligned} \tag{5}$$

The following theorem was proved in [18]:

**Theorem.** Let  $i^*$  and  $j^*$  be unique for each DMU in a group. Then the IM CRS PO and OM CRS PO efficiency scores are equal to each other and may be found as:

$$E_{CRS PO} = \max_{0 \leq j \leq r} \frac{X_{0j}}{X_j} \times \max_{0 \leq i \leq s} \frac{Y_i}{Y_{0i}} = \left( \frac{X_{0j^*}}{X_{j^*}} \right) \left( \frac{Y_{i^*}}{Y_{0i^*}} \right) \tag{6}$$

where IM, OM, and CRS stand for input minimization, output maximization, and constant returns to scale, respectively.

In case of non-unique  $i^*$  or  $j^*$ , any of them may be used in the formula (6). For ease of notation, we will refer to  $E_{CRS PO}$  simply as  $E$ . Figure 1 provides a geometric interpretation of the theorem for one input and one output. In this case, the IM efficiency score equals to the ratio of the distances from the y-axis to the frontier and to the DMU, while the OM efficiency, to the ratio of the distances from the x-axis to the DMU and to the frontier. The theorem follows from elementary geometric considerations.

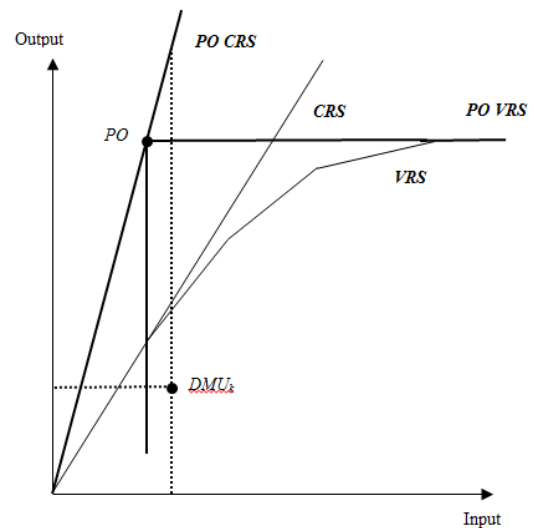


Figure 1. DEA frontiers for one input and one output

A Perfect Object ( $DMU_0$ ) is located at the point  $X_0 = \min X_k, Y_0 = \max Y_k, k = 1, \dots, N$ .

As follows from formula (6), given the joint distribution of relative inputs and outputs, the probability distribution of the efficiency score can be found explicitly. The underlying theory and corresponding formulas can be found in [16]. From formula (6), it follows that the efficiency scores for each  $DMU_k, k = 1, \dots, N$  are determined by the product of maximum normalized inputs and outputs. This observation allows for the introduction of stochasticity into DEA PO that is not related to the inaccuracy in the measurement of inputs or outputs. In contrast, stochastic DEA PO considers normalized actual DMUs as occurrences of identically distributed random DMUs. By doing so, it converts a group of DMUs under consideration into a statistical sample for which the probability distribution of the efficiency score can be estimated. The mathematical expectation of the group's

efficiency score is referred to below as common efficiency  $E_c$ :

$$E_c = M(E), 0 \leq E_c \leq 1 \tag{7}$$

where symbol  $M$  stands for the mathematical expectation. The common efficiency is the efficiency of the whole group and it serves as a benchmark for the evaluations of the *partial* efficiencies of the group's individual members.

Common efficiency shows how well the group is performing overall. Its use in the group-oriented approach to efficiency estimation is practical in some applications, such as a comparative study of national and regional environmental efficiency. In this case it is useful to evaluate how well a group of economies meets environmental standards, as well as efficiency variation among its members. For every particular DMU<sub>k</sub>, the difference between individual efficiency and common efficiency is referred to as partial efficiency  $E_{pk}$ :

$$E_{pk} = E_k - E_c, -1 \leq E_{pk} \leq 1. \tag{8}$$

As follows from formula (8), partial efficiency  $E_{pk}$  is scaled to the interval [-1,1]. It shows how well an individual DMU<sub>k</sub> performs with regard to the group as a whole. Negative values of partial efficiency  $E_{pk}$  mean that a corresponding DMU performs worse than the group at large and requires special attention of the regulatory body.

It should be noted that while conventional DEA efficiency scores are bound to the interval [0,1], publications [1] and [13] introduced “super efficiency” that allows it to be greater than 1. Similarly, negative values of the partial efficiency introduced in this paper extend the interval of feasible values of efficiency scores further to values below zero.

Introduction of common and partial efficiency leads to additive decomposition of the total efficiency of the DMU<sub>k</sub> as follows:

$$E_k = E_c + E_{pk} \tag{9}$$

where  $E_k$ ,  $E_c$ , and  $E_{pk}$  stand for the total efficiency score and its common and partial components for DMU<sub>k</sub>, respectively.

With inputs and outputs normalized to the interval [0,1] and considered as occurrences of identically distributed random objects, the Beta distribution is a convenient choice for fitting their distributions, see [8]. The support of the Beta distribution is the interval [0,1] - same as that of the DEA PO normalized inputs and outputs. It has two parameters:  $\alpha$  and  $\beta$  that control the shape of the probability density function (pdf). Depending on the values of  $\alpha$  and  $\beta$ , the Beta-pdf may be increasing, decreasing, bell-shaped, U-shaped, or horizontal. A statistical computer language R provides a procedure to estimate parameters of the Beta distribution. It is available on the internet for free download [22].

Recall that the pdf of the Beta distribution is:

$$f_B(x) = B(\alpha, \beta)^{-1} x^{\alpha-1} x^{\beta-1} \tag{10}$$

where  $B(\alpha, \beta)$  is the Beta-function

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt \tag{11}$$

The mathematical expectation and variance of the Beta distribution are as follows:

$$M_B = \frac{\alpha}{\alpha + \beta}, V_B = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \tag{12}$$

A cumulative distribution function (cdf) of the maximum normalized inputs and outputs that appear in the formula (6) can be found using the formulas for the probability distribution of order statistics, see [16]. Let  $\zeta = \langle \zeta_1, \zeta_2, \dots, \zeta_p \rangle$  be a random vector with cdf

$$F_\zeta(x_1, x_2, \dots, x_p) = P(\zeta_1 < x_1, \zeta_2 < x_2, \dots, \zeta_p < x_p) \tag{13}$$

and  $\eta = \max(\zeta_q, q=1, \dots, p)$ . Then the cdf of the random variable  $\eta$  is:

$$F_\eta(x) = P(\eta < x) = P(\zeta_1 < x, \zeta_2 < x, \dots, \zeta_p < x) = F_\zeta(x, x, \dots, x). \tag{14}$$

A pdf of the random variable  $\eta$ , denoted as  $f_\eta(x)$ , can be found by partial differentiation of its cdf with respect to  $x_1, x_2, \dots, x_p$  and consequent substituting  $x$  for every  $x_q, q = 1, \dots, p$ :

$$f_\eta(x) = dF_\eta(x) / dx = F_{1\zeta}(x, x, \dots, x) + F_{2\zeta}(x, x, \dots, x) + \dots + F_{p\zeta}(x, x, \dots, x), \tag{15}$$

where  $F_{q\zeta}(x, x, \dots, x)$  stands for the partial derivative of  $F_\zeta(x, x, \dots, x)$  with respect to  $q$ -th argument,  $q = 1, \dots, p$ .

In case of independent random variables  $\zeta_1, \zeta_2, \dots, \zeta_p$ ,

$$F_\eta(x) = F_{\zeta_1}(x) F_{\zeta_2}(x) \dots F_{\zeta_p}(x), \tag{16}$$

so that:

$$f_\eta(x) = \sum_{q=1}^p f_{\zeta_q}(x) \prod_{\substack{t=1 \\ t \neq q}}^p F_{\zeta_t}(x) \tag{17}$$

where  $f_{\zeta_q}(x)$  is a pdf of the random variable  $\zeta_q$ . In the case of two inputs and two outputs considered in the next section, this formula becomes:

$$f_\eta(x) = f_{\zeta_1}(x) F_{\zeta_2}(x) + f_{\zeta_2}(x) F_{\zeta_1}(x). \tag{18}$$

For three inputs or outputs, the formula is:

$$f_\eta(x) = f_{\zeta_1}(x) F_{\zeta_2}(x) F_{\zeta_3}(x) + f_{\zeta_2}(x) F_{\zeta_1}(x) F_{\zeta_3}(x) + f_{\zeta_3}(x) F_{\zeta_1}(x) F_{\zeta_2}(x). \tag{19}$$

It may be noted that the number of additive terms in formulas (16) through (19) grows linearly with the number of inputs or outputs, and their computation requires just minor changes in the computer program.

In addition to the pdf of the maximum value, a formula (6) also requires computation of the pdf of the product of two random variables having support [0,1]. Designating the pdf of the joint distribution of the factors as  $g(x, y)$ , we get [16],

$$f(x) = \int_0^1 \frac{1}{t} g\left(t, \frac{x}{t}\right) dt \tag{20}$$

that simplifies to

$$f(x) = \int_0^1 \frac{1}{t} g_1(t) g_2\left(\frac{x}{t}\right) dt \quad (21)$$

for two independent random variables.

Publication [9] emphasizes computational problems that may arise when formulas (20) or (21) are used in practice. However, these problems may be avoided if inputs and outputs are independent. In this case, the mathematical expectation of the product of the maxima of normalized inputs and outputs equals to the product of their mathematical expectations:

$$E_c = E_{output} \times E_{input}, \quad (22)$$

where

$$E_{output} = \int_0^1 x \cdot f_{output}(x) dx,$$

$$E_{input} = \int_0^1 x \cdot f_{input}(x) dx \quad (23)$$

where  $E_{output}$  and  $E_{input}$  are calculated using a formula (18) for two elements, formula (19) for three, and formula (17), for a greater number of the elements. Formula (22) holds for any number of inputs and outputs.

Formulas (17), (18), (19), (22) and (23) shed light on how SDEA PO operates, and on its ability to fix a problem related to conventional DEA revealed in the literature. As mentioned in literature - [1], for example - DEA may assign an efficiency score of 1 (fully efficient) to an object with a single very large output or a very small input. Similarly, DEA PO may assign a high efficiency score to an object having just one very large output *and* one very small input. These observations may undermine the ability of DEA to assign ranks objectively. Contrary to the deterministic versions, stochastic DEA PO, when producing a common efficiency score  $E_c$ , blends all of the indicators of all objects. By doing so, it obviates an opportunity for any single input or output to have a decisive impact.

The steps performed by the SDEA PO on the way to finding the common efficiency  $E_c$  are as follows. First, it blends inputs and outputs into two random vectors, with each component corresponding to one input or output, respectively. Next, it fits the Beta - distribution to the sample, and uses the distributions of inputs and outputs to obtain the pdf's of maximum values of relative inputs and outputs -  $f_{input}$  and  $f_{output}$  - as shown by formulas (17), (18) and (19). Then the mathematical expectations  $E_{output}$  and  $E_{input}$ , as given by formulas (23), are produced. Finally, the common efficiency score  $E_c$  is obtained as a product of two mathematical expectations, see formula (22) - a stochastic equivalent of formula (6). As a result, all inputs and outputs, rather than the maxima only, are included in the common efficiency score  $E_c$ .

It may be noted that computationally SDEA PO is robust with regard to the numbers of inputs and outputs. Each additional input or output just adds one more additive term in the structure of formula (17) and one multiplicative term to each addend.

### 3. Example of Applications

This section presents an example of application of the SDEA PO to the prospective comparative analysis of environmental efficiency of major national and regional economies. While conventional DEA is widely used in environmental performance research, see [20], for example, SDEA PO extends the opportunities in this arena. In this section, we analyze how well a group of major national and regional economies is expected to perform from the environmental point of view, and which of them need improvement. In calculations, we use prospective data for 2030, available on the website of the Energy Information Administration of the United States [21]. The website provides prospective data for GDP, population, total primary energy consumption, and carbon dioxide emissions. The data on the area were collected from the website of the United Nations Statistics Division [23].

Efficiency scores obtained by using a DEA PO or an SDEA PO are independent of the units of measurement. For convenience, we transformed the quantitative indicators into the percentage of the world total, as shown in Table 1. A select group of objects includes major national and regional economies that cover 80% to 94% of the total for each quantitative indicator. At the next step, the data in Table 1 were further transformed into the ratios presented in Table 2, columns (2) to (5), similar to that suggested in [14].

In general, given five quantitative indicators: Gross Domestic Product ( $G$ ), Population ( $P$ ), Area ( $A$ ), Primary Energy Consumption ( $R$ ), and Equivalent  $\text{CO}_2$  Emissions ( $C$ ), there are 20 ratios available for inclusion in the DEA model, as shown in Table 3. However, most of these ratios are functionally dependent. Thus, the ratios that are symmetrical about the main diagonal of Table 3 are inverses of each other. For example,  $R/C = 1/(C/R)$ . Also, any three ratios that form an  $L$ -shape in Table 3 are interconnected as well. For instance,  $(P/R)(A/P) = (A/R)$ . It may be shown that with five quantitative indicators, there are only four functionally independent ratios.

In this paper, we used functionally independent ratios with clear economic interpretations referred to below as environmental ratios. Among them are Energy intensity of GDP ( $R/G$ ) and Emissions intensity of energy ( $C/R$ ) - the ratios typical for environmental studies. They were used as DEA inputs. For outputs, we used GDP per capita ( $G/P$ ), a commonly used economic indicator, and the ratio of the Area to  $\text{CO}_2$ -equivalent emissions ( $A/C$ ). The last ratio is atmospheric clearness of a country or region. By choosing it, we stress the responsibility of each country or region for the atmospheric quality within its borders.

Functional independence of the selected ratios can be proven by using a functional determinant of the Jacobian matrix [12]. To prove it, we consider these ratios as functions of the quantitative indicators  $C$ ,  $A$ ,  $R$ ,  $P$ , and  $G$ , that is  $h_1 = G/P$ ,  $h_2 = A/C$ ,  $h_3 = R/G$ , and  $h_4 = C/R$ . With straightforward calculations, it can be shown that the

Jacobian matrix  $\frac{\partial(h_1, h_2, h_3, h_4)}{\partial(C, A, R, P, G)}$  has rank 4, so that the

functions  $h_i$ ,  $i=1, \dots, 4$ , and, therefore corresponding environmental ratios, are functionally independent.

Columns (6) to (9) in Table 2 present normalized environmental ratios obtained as the ratios of  $i$ -output to maximum output, and of minimum input to  $j$ -input, respectively. They were used further as DEA PO inputs

and outputs. For example, for the United States, the normalized values of inputs and outputs are as follows:  $output_1 = 3.6723/3.6723 = 1.0000$ ,  $output_2 = 0.4141/5.8420 = 0.0709$ ,  $input_1 = 0.6643/1.0174 = 0.6529$ , and  $input_2 = 0.6634/0.9278 = 0.7150$ . Maxima of the two normalized outputs and inputs are given in columns (10) and (11), while column (12) presents their product

expressed in percent. For instance, for the U.S.,  $max(1.0000, 0.0709) = 1.000$ ,  $max(0.6529, 0.7150) = 0.7150$ . The product is  $1.0 \times 0.7150 = 0.7150$  or 71.50%, which is the total efficiency score calculated by the formula (6) and shown in column (12) of Table 2 in the row corresponding to the U.S.

**Table 1. Projections for 2030, % of the World Total**

Country/ Region	Abbrev.	CO2	Energy	Population	Area	GDP
(1)	(2)	(3)	(4)	(5)	(6)	(7)
United States	USA	15.6117	16.8257	4.5034	6.4651	16.5379
Canada	Can	1.8568	2.6725	0.4684	6.7038	1.4802
Mexico	Mex	1.3852	1.4726	1.5372	1.3189	2.2039
OECD Europe	OECD	11.1136	13.3488	6.8212	3.3716	16.8056
Australia/New Zealand	ANZ	1.3014	1.2272	0.3603	5.3442	1.1965
Russia	Rus	4.8381	5.4950	1.4891	11.4800	3.0767
China	Chi	29.2555	23.0979	17.5093	6.4722	19.2757
India	Ind	5.2507	4.7723	18.0857	2.2071	7.1841
Middle East	MidE	5.6430	5.6040	3.5307	4.7408	3.7110
Africa	Afr	3.4733	3.2179	18.2299	20.2911	4.3336
Brazil	Bra	1.7639	2.6588	2.8342	5.7170	2.8527
Other Central and South America	CSAnB	2.4463	2.9179	4.1792	6.6127	3.6535
	Max	29.2555	23.0979	18.2299	20.2911	19.2757
	Min	1.3014	1.2272	0.3603	1.3189	1.1965
	Average	6.9950	6.9426	6.6290	6.7270	6.8593

**Table 2. Prospected ratios and efficiency**

2030 Projected ratios					Normalized ratios a				Efficiency			
Output-1	Output-2	Input-1	Input-2		Output-1	Output-2	Input-1	Input-2	Max Output	Max Input	DEA PO,% <sup>b</sup>	Partial Eff,% <sup>c</sup>
GDP/Pop	Area/CO2	Energy/GDP	CO2/ Energy		GDP/ Pop	Area/CO2	Energy/ GDP	CO2/ Energy				
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
USA	3.6723	0.4141	1.0174	0.9278	1.0000	0.0709	0.6529	0.7150	1.0000	0.7150	71.50	25.07
Can	3.1604	3.6104	1.8055	0.6948	0.8606	0.6180	0.3679	0.9548	0.8606	0.9548	82.17	35.74
Mex	1.4337	0.9521	0.6682	0.9407	0.3904	0.1630	0.9942	0.7052	0.3904	0.9942	38.81	-7.62
OECD	2.4637	0.3034	0.7943	0.8326	0.6709	0.0519	0.8363	0.7968	0.6709	0.8363	56.11	9.68
ANZ	3.3211	4.1067	1.0256	1.0605	0.9044	0.7030	0.6477	0.6256	0.9044	0.6477	58.57	12.14
Rus	2.0661	2.3728	1.7860	0.8805	0.5626	0.4062	0.3719	0.7535	0.5626	0.7535	42.39	-4.04
Chi	1.1009	0.2212	1.1983	1.2666	0.2998	0.0379	0.5544	0.5238	0.2998	0.5544	16.62	-29.81
Ind	0.3972	0.4203	0.6643	1.1003	0.1082	0.0720	1.0000	0.6029	0.1082	1.0000	10.82	-35.61
MidE	1.0511	0.8401	1.5101	1.0069	0.2862	0.1438	0.4399	0.6588	0.2862	0.6588	18.86	-27.57
Afr	0.2377	5.8420	0.7425	1.0794	0.0647	1.0000	0.8946	0.6146	1.0000	0.8946	89.46	43.03
Bra	1.0065	3.2412	0.9320	0.6634	0.2741	0.5548	0.7127	1.0000	0.5548	1.0000	55.48	9.05
CSAnB	0.8742	2.7032	0.7987	0.8384	0.2381	0.4627	0.8317	0.7913	0.4627	0.8317	38.49	-7.94
Max	3.6723	5.8420	1.8055	1.2666								
Min	0.2377	0.2212	0.6643	0.6634								

Notes:

- a. Normalization is done by taking the ratios of i-output to the maximum output and of the minimum input to j-input, respectively.
- b. DEA PO efficiency is a product of normalized Max Output by Max Input ratios. The value is given in %.
- c. The difference between DEA PO efficiency and the common efficiency of 46.43%.

**Table 3. Ratio matrix**

	C	R	P	A	G
C	-	C/R	C/P	C/A	C/G
R	R/C	-	R/P	R/A	R/G
P	P/C	P/R	-	P/A	P/G
A	A/C	A/R	A/P	-	A/G
G	G/C	G/R	G/P	G/A	-

For analysis, we used SDEA PO to separate total efficiency into additive components: common and partial efficiency, as given by the formula (9). The common component  $E_c$  is a part of the total efficiency; it was calculated using SDEA PO and a computer program provided in the Appendix section. Methodologically,  $E_c$

corresponds to the worldwide environmental efficiency level obtained by weighing CO<sub>2</sub>-equivalent emissions against economic development, technological progress, energy consumption, area, and population. It was calculated by fitting the Beta distribution to the normalized environmental ratios and then by using formula (18) for finding the probability distribution of the corresponding maxima, and formulas (23) and (22) for calculations of the mathematical expectations and common efficiency, respectively. For the calculations, we used a program written by the author in R - language, version R 2.13.0 given in the Appendix.

While the R language procedure "fitdist" with default parameters (maximum likelihood method) gave the best results in simulations, it failed to fit to actual data.

Because of that, we used the program parameters providing maximum goodness of fit estimation (MGE) using the Cramér–von Mises (CvM) distance. Table 4 presents the values of the parameters  $\alpha$  and  $\beta$  for the Beta distribution. These values were used in the pdf functions  $g_1$  and  $g_2$  in formula (10). By doing so, we obtained the common efficiency score  $E_c = 46.43\%$ . This value was used in column 13 of Table 2 to calculate partial efficiency scores as the difference between total and common efficiency, see formula (8).

Table 4. Beta-distribution parameters, 2030

	GDP/Pop	Area/CO2	Energy/GDP	CO2/Energy
$\alpha$	0.807672	0.598249	2.063552	10.20527
$\beta$	0.939740	1.092585	0.988334	4.295321

Figure 2 shows total and partial efficiency graphically. As follows from Table 2 and Figure 2, the U.S. and Canada are expected to have high values of total efficiency (71.50% and 82.17%, respectively), and positive values of partial efficiency (25.07% and 35.74%, respectively). For the OECD countries of Europe and Australia/New Zealand region, the expected total efficiency is sufficiently lower: 56.11% and 58.57%, respectively, though partial efficiency is still positive - 9.68% and 12.14%, respectively. When looking at the BRIC countries (Brazil, Russia, India, and China) that are conventionally considered as the leading force of future economic development, it may be noted that only Brazil is expected to have positive partial efficiency of 9.05%. The other three bear negative values: Russia -4.04%, China -29.81, and India -35.61%. The Middle East is another region with a large negative partial efficiency score of -27.57%.

Obtained results may be used to develop recommendations on economic restructuring leading to better environmental performance worldwide.

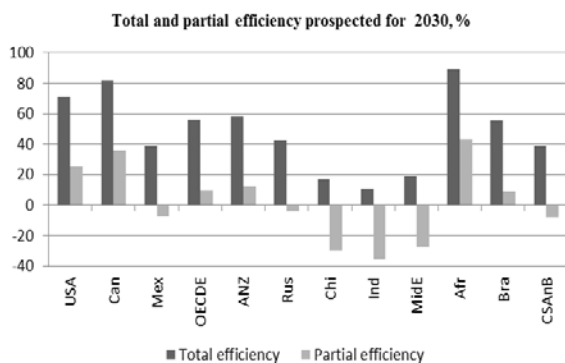


Figure 2. Total and partial efficiency. Common efficiency = 46.43%

## 4. Conclusions

The paper introduces Stochastic DEA with a Perfect Object (SDEA PO) and demonstrates its application to prospective analysis of environmental efficiency.

SDEA PO considers a group of DMUs as a sample of identically-distributed normalized DMU occurrences. An explicit formula derived for the DEA PO efficiency index allows us to obtain the probability distribution of the group's efficiency score. The mathematical expectation of this score characterizes the efficiency of the group at large and is called common efficiency. The deviation of the

observed or prospective efficiency of a particular DMU from the group level is called partial efficiency. It reflects the relative performance of a specific DMU in the group. The SDEA PO approach is applied to the analysis of environmental efficiency of major national and regional economies. It reveals how different national and regional economies perform and can aid in the development of clean air policies.

## Acknowledgement

Partial support for this research was provided by a PSC-CUNY Award # 65135-00 43, jointly funded by The Professional Staff Congress and The City University of New York. Author is thankful to the anonymous reviewer for useful comments.

## References

- Andersen, P., Petersen, N., "A procedure for ranking efficient units in data envelopment analysis," *Management Science*, 39(10), 1261-1264, 1993.
- Banker, R., *Stochastic data envelopment analysis*. Carnegie-Mellon University, Pittsburgh, 1986.
- Banker, R. D., Charnes, A., Cooper, W., "Some models for estimating technical and scale efficiency in Data Envelopment Analysis," *Management Science*, 30(9), 1078-1092, 1984.
- Charnes, A., Cooper, W., "Programming with linear fractional functionals," *Naval Research Logistics Quarterly*, 9, 181-186, 1962.
- Charnes, A., Cooper, W., Rhodes, E., "Measuring the efficiency of decision-making units," *European Journal of Operational Research*, 2, 429-444, 1978.
- Cooper, W., Seiford L., and Tone, K., *Data Envelopment Analysis: A comprehensive text with models, applications, references and DEA-Solver software*. (2nd Ed). New York: Springer, 2006.
- Farrell, M. "The measurement of production efficiency," *Journal of the Royal Statistical Society, Series A*, 120 (3), 253-282, 1957.
- Evans, M., Hastings, N., & Peacock, B., "Beta Distribution", In: *Statistical Distributions*, 3rd Ed. New York: Wiley, p. 34-42, (Chapter 5), 2000.
- Glenn, A., Evans, D., and Leemis, L., "APPL: A Probability Programming Language," *The American Statistician*, 55 (2), 156-166, 2001.
- Gstach, D., "Another approach to data envelopment analysis in noisy environments: DEA+," *Journal of Productivity Analysis*, 9, 161-176, 1998.
- ICON Group International, *Data Envelopment Analysis: Webster's Timeline History, 1981-2007*. ICON Group International, Inc., 2009.
- Kaplan W., *Advanced calculus*. (4th Ed.) New York: Addison-Wesley, 1993.
- Khodabakhshi, M., Asgharian, M., Gregoriou, G., "An input-oriented super-efficiency measure in stochastic data envelopment analysis: Evaluating chief executive officers of US public banks and thrifts. *Expert Systems with Applications*, 37, 2092-2097, 2010.
- Mereste U., "On the matrix method of analysis of economic efficiency of social production," ("O matrichnom metode analiza ekonomicheskoi effektivnosti obshestvennogo proizvodstva"), *Ekonomika i matematicheskie metody*, XVIII (1), 138-49, 1982. (in Russian).
- Sengupta, J., "Efficiency measurement in stochastic input-output systems," *International Journal of Systems Science*, 13, 273-287, 1982.
- Spiegel, M., Schiller, J., and Srinivasan, R., *Probability and Statistics*. (2<sup>nd</sup> Ed.), Schaum's Outlines. New York: McGraw Hill, 2000.
- Vaninsky, A., "Environmental Performance of the United States Energy Sector: A DEA Model with Non-Discretionary Factors and Perfect Object," *Publications of the World Academy of Science. Engineering and Technology*, 30, 139-144, 2009.

- [18] Vaninsky, A., "DEA with a Perfect Object: Analytical Solutions," *Communications in Mathematics and Applications*, 2(1), 1-13, 2011.
- [19] Vaninsky, A. "Explicit formulas for efficiency scores and weight coefficients in DEA problems with a perfect object," *International Journal of Mathematical Modeling, Simulation and Applications* 4(3): 217-233, 2011.
- [20] Zhou P, Ang B., Poh KL., "Measuring environmental performance under different environmental DEA technologies," *Energy Economics*, 30, 1-14, 2008.
- [21] [www.eia.gov](http://www.eia.gov) - the website of the U.S. Energy Information Administration (EIA).
- [22] <http://www.r-project.org/> - the website concerning R -language.
- [23] <http://unstats.un.org/unsd/default.htm> - the website of the United Nations Statistics Division.

## Appendix

This section presents a program in the R language, version R 2.13.0, for fitting parameters of the Beta distribution to statistical data for two inputs and two outputs. In case of a greater number of inputs or outputs, the functions *maxoutpdf* and *maxinppdf* need be changed appropriately, as given in the text by formulas (17) or (19). In this version of the program, the statistical data were included in the R script file. In general, they should be imported from a spreadsheet or database using R language tools. Symbol “#” stands for a comment line or a part of it.

```
library(fitdistrplus) #MML
#----- DEA Outputs -----
#GDP/Pop 2030
d1<-c(1.000000, 0.860593, 0.390419, 0.670897,
0.904363, 0.562622, 0.299780, 0.108168, 0.286214,
0.064733, 0.274090, 0.238058)
#vv1<-fitdist(d1,"beta")
vv1<-fitdist(d1,"beta", method="mge", gof="CvM")
a1=vv1$estimate[1][[1]]
b1=vv1$estimate[2][[1]]
a1
b1
#-----
#Area/CO2 2030
d2<-c(0.070886, 0.618005, 0.162977, 0.051930,
0.702954, 0.406162, 0.037869, 0.071951, 0.143808,
1.000000, 0.554806, 0.462710)
#vv2<-fitdist(d2,"beta")
vv2<-fitdist(d2,"beta", method="mge", gof="CvM")
a2=vv2$estimate[1][[1]]
b2=vv2$estimate[2][[1]]
```

```
a2
b2
#-----
#----- DEA Inputs -----
#Energy/GDP 2030
d3<-c(0.652922, 0.367918, 0.994173, 0.836308,
0.647689, 0.371945, 0.554360, 1.000000, 0.439888,
0.894605, 0.712715, 0.831749)
#vv3<-fitdist(d3,"beta")
vv3<-fitdist(d3,"beta", method="mge", gof="CvM")
a3=vv3$estimate[1][[1]]
b3=vv3$estimate[2][[1]]
a3
b3
#-----
#CO2/Energy 2030
d4<-c(0.714979, 0.954810, 0.705225, 0.796809,
0.625570, 0.753452, 0.523762, 0.602941, 0.658813,
0.614608, 1.000000, 0.791293)
#vv4<-fitdist(d4,"beta")
vv4<-fitdist(d4,"beta", method="mge", gof="CvM")
a4=vv4$estimate[1][[1]]
b4=vv4$estimate[2][[1]]
a4
b4
#-----
#----- pdf's for DEA max ratios for outputs and
inputs-----
maxoutpdf<-function(x,aa1,bb1,aa2,bb2)
{dbeta(x,aa1,bb1)*pbeta(x,aa2,bb2)+dbeta(x,aa2,bb2)*
pbeta(x,aa1,bb1)}
maxinppdf<-function(x,aa3,bb3,aa4,bb4)
{dbeta(x,aa3,bb3)*pbeta(x,aa4,bb4)+dbeta(x,aa4,bb4)*
pbeta(x,aa3,bb3)}
maxout<-function(x) {x*maxoutpdf(x,a1,b1,a2,b2)}
maxinp<-function(x) {x*maxinppdf(x,a3,b3,a4,b4)}
#----- Mathematical expectations of max ratios of
inputs and outputs -----
effout<-integrate(maxout,0,1)
effout$value
effinp<-integrate(maxinp,0,1)
effinp$value
#-----
#----- Common efficiency -----
effcomm=effout$value*effinp$value
effcomm
```