The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/0025-1747.htm

Received 5 November 2014
Revised 18 July 2015
Accepted 14 August 2015

Data envelopment analysis with interactive variables

Aibing Ji, Hui Liu and Hong-jie Qiu
College of Public Health, Hebei University, Baoding, China, and
Haobo Lin
Computer Centre, Hebei University, Baoding, China

Abstract

Purpose – The purpose of this paper is to build a novel data envelopment analysis (DEA) model to evaluate the efficiencies of decision making units (DMUs).

Design/methodology/approach – Using the Choquet integrals as aggregating tool, the authors give a novel DEA model to evaluate the efficiencies of DMUs.

Findings – It extends DEA model to evaluate the DMU with interactive variables (inputs or outputs), the classical DEA model is a special form. At last, the authors use the numerical examples to illustrate the performance of the proposed model.

Practical implications – The proposed DEA model can be used to evaluate the efficiency of the DMUs with multiple interactive inputs and outputs.

Originality/value – This paper introduce a new DEA model to evaluate the DMU with interactive variables (inputs or outputs), the classical DEA model is a special form.

Keywords Data envelopment analysis, Choquet integrals, Efficiency evaluation, Interactive variables

Paper type Research paper

1. Introduction

Data envelopment analysis (DEA) is a common methods for evaluating the performance of organizational units. DEA is a non-parametric optimizing mathematical method that was first introduced by Charnes and Cooper Rhodes (1978), then was developed by Banker *et al.* (1984). According to DEA, it combines and transforms multiple inputs and outputs into a single efficiency index, efficiency will be calculated relatively among a homogeneous collection of decision-making units (DMUs) based on some equal inputs and outputs. This approach first establishes an "efficient frontier" formed by a set of DMUs that exhibit best practices and then assigns the efficiency level to other non-frontier units according to their distances to the efficient frontier. The basic idea has since generated a wide range of variations in measuring efficiency. Nowadays, various DEA efficiency models are available for different types of measuring requirement, such as the constant returns to scale model (Charnes and Cooper Rhodes, 1978), the variable-returns-to-scale (Banker *et al.*, 1984) model, the additive model (Charnes *et al.*, 1985) the slacks-based measures and the free disposal hull model (Bardhan *et al.*, 1996), etc. It also has been applied to various industrial and non-industrial contexts, such as banking

© Aibing Ji, Hui Liu, Hong-jie Qiu and Haobo Lin. Published by Emerald Group Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 3.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial & non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/3.0/legalcode

National Social Science Fund (14BJY010), science and technology research and development programs of Hebei Province (12276104D-3) and Research Foundation of College of Medicine in Hebei University (2012A3003).

Management Decision
Vol. 53 No. 10, 2015
pp. 2390-2406
Emerald Group Publishing Limited
0025-1747
DOI 10.1108/MD-11-2014-0631

(Cooper *et al.*, 2008; Kao and Liu, 2009), education (Ray and Jeon, 2008), hospital (Ancarani *et al.*, 2009), and so on.

In all the classical DEA models, the combination form of the multiple inputs (or outputs) is a linear weighted sum, it needs an assumption that there is no interaction among the contributions from individual information sources so that the joint contribution is just the simple sum of contributions from individual information sources. But in practice, the variables (inputs or outputs) are usually strongly correlated, there are interactions among variables. In general, the inherent interaction cannot be ignored in the efficiency evaluation. For example, in the efficiency evaluation of Greek commercial banks (Varias and Sofianopoulou, 2012), the inputs (such as fixed assets and total assets) are closely related, if the weighted sum of multiple inputs (or outputs) is used in the DEA model, then the evaluation results may have deviations. The selection of variables (inputs or output) has significant impacts on DEA efficiency estimates (Nataraja and Johnson, 2011), Dyson *et al.* (2001) showed that the omission of a highly correlated variable or adding of uncorrelated variable can have a significant impact on the efficiency estimates of some production units.

The collinearity of variables is a special interaction, the collinearity of variables can affect the evaluation result of DEA (Jukti, 1995; Fengxia Dong and Mitchell, 2015).

Some researchers have proposed several models to solve these problems. Adler and Golany (2001, 2002) suggested using the principal component analysis (PCA) to produce uncorrelated linear combinations of original inputs and outputs, and construct a PCA-based DEA model. Adler and Yazhemsky (2010) concluded that PCA-DEA outperforms classical DEA by comparing their discrimination performance in a simulation exercise. Independent component analysis (ICA) is an another information aggregation tool to resolve the problem of variables correlation, ICA is a novel statistical technique used to extract independent variables from observed multivariate statistical data where no relevant data mixture mechanisms are available (Hyvarinen and Oja, 2000). Using ICA to extract independent variables (inputs or outputs) in DEA efficiency measurement (Kao *et al.*, 2011) can partly overcome the effect of variables correlation.

Although the above two DEA models can avoid efficiency misjudgment for the DMUs with interactive variables (inputs or outputs), but they divide efficiency measurement into two stages, first use the variables (inputs, outputs) of the DMUs to generate independent variables (inputs, outputs) and then construct DEA model for the generated independent variables (inputs, outputs), which lead to high computation complexities. Especially, in the two DEA models, the variables (inputs, outputs) of DEA model are not practical variables of the DMUs, but are the variables produced by the original variables (in PCA-based DEA model, they are uncorrelated combinations of original variables; in ICA-based DEA model, they are the basis vector of original variables), the results of efficiency evaluations have fewer guiding for production activities.

The weighted sum is the Lebesgue-like integrals on discrete spaces, in which the global contribution toward the objective is regarded as the sum of contributions from each attribute. It needs a basic supposition that there is no interaction among attributes. None of the classical aggregation tools are related to interactions among attributes. Choquet integral (Choquet, 1954) is a generalization of the Lebesgue integral, when the fuzzy measure is additive, the Choquet integrals coincide with Lebesgue integrals. Choquet integral with respect to efficiency measures (also called fuzzy measures or non-additive measures) is a non-linear integral (Wang *et al.*, 2010), it can provide the attribute aggregation tool which considers interactions among attributes.

In this paper, we shall discuss the efficiency evaluation model with interaction variables (inputs or outputs), in which Choquet integrals are used to aggregate the multiple inputs and outputs into a single efficiency index. In Section 2, we present some preliminaries. In Section 3, we propose the DEA model with interaction variables. In Section 4, we apply the proposed DEA model to the efficiency evaluation.

2. Background

In this section, some preliminaries for DEA and Choquet integral are presented.

2.1 The DEA model and cross-efficiency evaluation

Charnes, Cooper, and Rhodes introduced the method of DEA to address the problem of efficiency measurement for DMUs with multiple inputs and multiple outputs. They have applied in the efficiency evaluation for the non-market agencies like schools, hospitals, which produce identifiable and measurable outputs from measurable inputs but generally lack market prices of outputs (and often of some inputs as well). Suppose that there are n DMUs, each producing m outputs from s inputs. DMU_0, the DMU to be evaluated. DMU_k use the input bundle $x^k = (x_{1k}, x_{2k}, \ldots, x_{sk})$ to produce the output bundle $y^k = (y_{1k}, y_{2k}, \ldots, y_{mk})$. The weighted sums are used to aggregate of inputs and outputs, the measure of efficiency of any DMU is obtained as the maximum of ratio of total weighted outputs to the total weighted inputs subject to the condition that the similar ratios for every DMU be less than or equal to unity. Its mathematical model is as follows:

$$\max h_0 = \frac{\sum_{j=1}^{m} u_j y_{j0}}{\sum_{i=1}^{s} v_i x_{i0}} \tag{1}$$

Subject to:

$$\frac{\sum_{j=1}^{m} u_j y_{jk}}{\sum_{i=1}^{s} v_i x_{ik}} \leqslant 1, \ k = 1, 2, \ldots, n$$

$$u_j \geqslant 0; \ (j = 1, 2, \ldots, m);$$
$$v_i \geqslant 0; \ (i = 1, 2, \ldots, s).$$

This is a linear fractional functional programming problem, by the C2 transformation (Charnes and Cooper, 1962), the linear fractional functional programming problem (1) can be transformed to the following equivalent linear programming:

$$\max \sum_{j=1}^{m} u_j y_{j0} \tag{2}$$

Subject to:

$$\sum_{j=1}^{m} u_j y_{ik} - \sum_{i=1}^{s} v_i x_{ik} \leqslant 0; \quad (k = 1, 2, \ldots, n);$$

$$\sum_{i=1}^{s} v_i x_{i0} = 1;$$

$$u_j \geqslant 0; v_i \geqslant 0 \quad (j = 1, 2, \ldots, m; i = 1, 2, \ldots, s)$$

The above efficiency evaluation models are called CCR models which are treated in input-oriented forms. The efficient DMUs in CCR model are called CCR efficient.

The dual form of the programming (2) is:

$$\min \theta$$

Subject to:

$$\begin{aligned}
&\sum_{k=1}^{n}\lambda_k x_{ik} \leqslant \theta x_0, \quad (i = 1, 2, \ldots, s);\\
&\sum_{k=1}^{n}\lambda_k y_{jk} \geqslant y_0, \quad (j = 1, 2, \ldots, m)\\
&\lambda_k \geqslant 0, \quad (k = 1, 2, \ldots, n)
\end{aligned} \tag{3}$$

Banker, Charnes and Cooper (BCC) model (Banker *et al.*, 1984) adds an additional constant variable, w_0, in order to permit variable returns-to-scale:

$$\max z = \sum_{j=1}^{m} u_j y_{j0} - w_0$$

Subject to:

$$\begin{aligned}
&\sum_{j=1}^{m} u_j y_{jk} - \sum_{i=1}^{s} v_i x_{ik} - w_0 \leqslant 0, \; k = 1, 2, \ldots, n\\
&\sum_{i=1}^{S} v_i \, x_{i0} = 1\\
&u_j \geqslant 0; \; (j = 1, 2, \ldots, m);\\
&v_i \geqslant 0; \; (i = 1, 2, \ldots, s),
\end{aligned} \tag{4}$$

The dual form of the programming (4) is:

$$\min \theta$$

Subject to:

$$\begin{aligned}
&\sum_{k=1}^{m} x_{ik}\lambda_k \leqslant \theta x_0, \; i = 1, \ldots, s\\
&\sum_{k=1}^{n} y_{jk}\lambda_k \geqslant y_o, \; j = 1, 2, \ldots, m\\
&\sum_{k=1}^{n} \lambda_k = 1\\
&\lambda_k \geqslant 0, \; k = 1, 2, \ldots, n
\end{aligned} \tag{5}$$

It can be seen from DEA models (2) (or (4)) that the essence of CCR (BCC) models is that the DMU evaluated tries to find out its weight vector to maximizing its weighted output with the constraints that its weighted input is fixed as unity and the weighted output is not larger than the weighted input for all DMUs. In other words, each DMU seeks its favorite weight vector to its own advantage.

To overcome the inability of DEA in discriminating between DEA efficient units, cross-efficiency evaluation (Sexton *et al.*, 1986; Doyle and Green, 1994, 1995) has been suggested in the DEA literature. In the cross-efficiency evaluation, each DMU determines a set of input and output weights individually, leading to n sets of weights for n DMUs. The n sets of weights are then used to assess the efficiencies of the n DMUs, resulting in n efficiency values for every DMU. The n efficiency values for each DMU are finally averaged as an overall efficiency value of the DMU. It is believed that the cross-efficiency evaluation can guarantee a unique ordering for the DMUs. Due to its good discrimination power, the cross-efficiency evaluation has been used in a variety of applications, including preference voting and project ranking (Green *et al.*, 1996), industrial robot selection (Baker and Talluri, 1997). The cross-efficiency evaluation has also been deeply studied theoretically (Ramón and Ruiz, 2010; Cook and Zhu, 2014).

In the first stage, the optimal weights of inputs and outputs are calculated for each DMU using the classical DEA formulation (CCR model or BBC model). Given the results of the first stage, the weights used by the DMU can be utilized for calculating the peer rated efficiency for each of the other DMUs. The peer evaluation score, ϑ_{pk}, indicates the efficiency score for DMU_k using the weights obtained by DMU_p:

$$\vartheta_{pk} = \frac{\sum_{j=1}^{m} u_{jp} y_{jk}}{\sum_{i=1}^{s} v_{ip} x_{ik}}$$

The cross-efficiency scores can be summarized in a cross-efficiency matrix. Note that $0 \leqslant \vartheta_{pk} \leqslant 1$ and the elements in the diagonal, ϑ_{pp}, represent the standard DEA efficiency score. Then the overall efficiency value of the DMU_k is calculated by:

$$\vartheta_p = \frac{1}{n} \sum_{k=1}^{n} \vartheta_{pk}$$

2.2 Efficiency measures and the Choquet integral

The concept of efficiency measure was initiated in the 1950s (Choquet, 1954) and has been well developed since 1970s (Denneberg, 1994; Wang and Klir, 2008; Wang *et al.*, 2010).

Considering the non-linear relationships, particularly interactions among attributes, the non-linear integrals can be used as data aggregation tools. Studies of non-linear integrals can be found in literature, such as (Wang and Klir, 2008; Wang *et al.*, 2010; Denneberg, 1994). The weighted sum is also referred to as the Lebesgue-like integral on a finite space (Wang *et al.*, 2010) The Choquet integral (Choquet, 1954; Wang *et al.*, 2010), one of non-linear integrals, is more appropriate for information fusion and data mining applications because it contains very important information regarding the interactions among attributes in the database. It can be utilized to aggregate the values of feature attributes according to the values of the set function μ. Thus, the Choquet integral with respect to an efficiency measure is chosen as the non-linear data aggregation tool.

Let finite set $X = \{x_1, x_2, \ldots, x_n\}$ be the attributes in a multi-dimensional data set. Several important types of efficiency measure are defined as the follows (Denneberg, 1994; Wang and Klir, 2008; Wang *et al.*, 2010).

Definition 1. An efficiency measure μ defined on X is a set function μ: $P(X) \to [0,+\infty)$ satisfying $\mu(\phi) = 0$ (where $P(X)$ denotes the power set of X); μ is called a monotone measure if it satisfies $\mu(\phi) = 0$ and $\mu(E) \leqslant \mu(F)$ for $E \subset F$, where E, F are any sets in $P(X)$.

The fuzzy measure (Sugeno, 1974) is a monotone efficiency measure on X.

Definition 2. An efficiency measure is said to be regular if $\mu(X) = 1$.

Definition 3. A signed efficiency measure μ defined on X is a set function μ: $P(X) \to (-\infty, +\infty)$ satisfying $\mu(\phi) = 0$.

For efficiency measure μ, $A \in P(X)$, $B \in P(X)$, there are three cases:

Case 1: If $\mu(A \cup B) > \mu(A) + \mu(B)$, implying that A and B have a multiplicative effect.

Case 2: If $\mu(A \cup B) = \mu(A) + \mu(B)$, implying that A and B have an additive effect.

Case 3: If $\mu(A \cup B) < \mu(A) + \mu(B)$, implying that A and B have a substitutive effect.

The efficiency measure is often used with the non-linear integral for aggregating information evaluation by considering the influence of the substitutive and multiplication effect among all attributes (see the example 1 below, the worker x_1 and x_2 have a multiplicative effect; the worker x_2 and x_3 have a substitutive effect).

Definition 4. Let $(X, P(X), \mu)$ be an efficiency measure space, f: $X \to (-\infty, +\infty)$ be a measurable real-valued function, then the Choquet integral of $f(x)$with respect to the efficiency measure μ is defined as follows:

$$(c)\int f\, d\mu = \int_{-\infty}^{0} [\mu(\mathrm{F}_\alpha) - \mu(\mathrm{X})]\mathrm{d}\alpha + \int_{0}^{\infty} \mu(\mathrm{F}_\alpha) d\alpha$$

Where (c) indicates the type of integral being the Choquet integral and $\mathrm{F}_\alpha = \{x | f(x) \geqslant \alpha,\, x \in X\}$, $\alpha \in [0, \infty)$, $\int_0^\infty \mu(\mathrm{F}_\alpha) d\alpha$ are the Riemann integral. When f are non-negative functions, $(c)\int f\, d\mu = \int_0^\infty \mu(\mathrm{F}_\alpha) d\alpha$

The Choquet integral is a non-linear integral, it is generalization of the Lebesgue integral and coincide with the Lebesgue integral for additive efficiency measure.

When $X = \{x_1, x_2, \cdots, x_n\}$ is a finite set, the values of f, i.e., $f(x_1), f(x_2), \cdots, f(x_n)$, can be sorted in a non-decreasing order so that $f(x'_1) \leqslant f(x'_2) \leqslant \cdots \leqslant f(x'_n)$, where $\{x'_1, x'_2, \ldots, x'_n\}$ is a certain permutation of $\{x_1, x_2, \cdots, x_n\}$, then the Choquet integral is obtained by:

$$(C)\int f d\mu = \sum_{i=1}^{n} \left[f(x'_i) - f(x'_{i-1})\right] \mu(A_i) \tag{6}$$

Where $f(x'_0) = 0$ and $A_i = \{x'_i, x'_{i+1}, \ldots, x'_n\}$.

We can easily verify the formula (6) is equivalent to:

$$(C)\int f d\mu = \sum_{i=1}^{n} f(x'_i)[\mu(A_i) - \mu(A_{i+1})] \tag{7}$$

where $A_{n+1} = \phi$.

Example 1 Let $\{x_1, x_2, x_3\}$ be the set of three workers, function f: $X \to [0, \infty)$ be the numbers of hours of their work in a certain day with:

$$f(x) = \begin{cases} 6, & \text{if } x = x_1 \\ 3, & \text{if } x = x_2 \\ 4, & \text{if } x = x_3 \end{cases}$$

Individual and joint efficiency measure μ: $P(X) \to [0, +\infty)$ are as following:

Set	Value of μ
ϕ	0
$\{x_1\}$	5
$\{x_2\}$	6
$\{x_3\}$	7
$\{x_1, x_2\}$	14
$\{x_1, x_3\}$	13
$\{x_2, x_3\}$	9
$\{x_1, x_2, x_3\}$	17

$(C)\int f d\mu = 3 \times 17 + 1 \times 13 + 2 \times 5 = 74$

Because of the interactions among attributes, it is not equal to the weighted sum: $6\times5+3\times6+4\times7 = 76$.

3. DEA with interactive variables

According to the actual production activities, the efficiency measures are positive and monotone, therefore, in the following, let us suppose that the efficiency measures are always positive and monotone. For a collection of DMUs, the attribute set of inputs is $X = \{x_1, x_2, \ldots, x_s\}$, the attribute set of outputs is $Y = \{y_1, y_2, \ldots, y_m\}$. Taking into account the interaction among the multiple inputs (outputs), the efficiency measure $\mu(\{x_i\})$ represent the efficiency of the input index x_i, the efficiency measure $\mu(A)$ (A is not a single point set) represent the joint efficiency of attribute set $A \subset X$; and it is similar to the efficiency measure ν on $P(Y)$. Let $f_k(x_i)$ denote the numerical information that kth DMU obtained from input attribute x_i, $g_k(y_j)$ denote the numerical information that kth DMU obtained from output attribute y_j. The kth DMU use the input bundle $f_k(x) = (f_k(x_1), f_k(x_2), \ldots, f_k(x_s))$ to produce the output bundle $g_k(y) = (g_k(y_1), g_k(y_2), \ldots, g_k(y_m))$. The proposed measure of the efficiency of any DMU is obtained as the maximum of the ratio of the aggregate outputs to aggregate inputs, where the aggregate inputs (or outputs) are calculated by the Choquet integral. The efficiency measure μ and ν can be determined by the following optimization problem ((CH-CCR)^I Model):

$$\max \left\{ h_0 = \frac{(c)\int g_0 dv}{(c)\int f_0 d\mu} \right\} \tag{8}$$

Subject to:

$$\begin{cases} \frac{(c)\int g_k dv}{(c)\int f_k d\mu} \leqslant 1, \ k = 1, 2, \ldots, n \\ 0 \leqslant \mu(A) \leqslant \mu(B), \ for\ A \subseteq B, \quad A, B \in P(X); \\ 0 \leqslant v(C) \leqslant v(D), \ for\ C \subseteq D, \quad C, D \in P(Y); \end{cases}$$

The efficiency ratio ranges from zero to one, the k-DMU is considered relatively efficient if it receives a score of one. Thus, each unit will choose weights so as to maximize self-efficiency, given the constraints.

The result of the DEA is the determination of the hyper-planes that define an envelope surface. DMUs that lie on the surface determine the envelope and are deemed efficient, whilst those that do not are deemed inefficient.

We utilize the following transformations:

$$t = \frac{1}{(c)\int f_0 d\mu}, (t > 0),\ \omega = t\upsilon,\ \lambda = t\mu \tag{9}$$

The set functions λ: $X \rightarrow [0, +\infty)$, ω: $Y \rightarrow [0, +\infty)$ are all also non-negative monotone efficiency measures.

By using the transformations (9), the $(\text{CH-CCR})^{\text{I}}$ model (8) can be changed into the following equivalent CH-CCR model:

$$\max \left\{ h_0 = (c)\int g_0 d\omega \right\} \tag{10}$$

Subject to:

$$\begin{cases} (c)\int g_k d\omega - (c)\int f_k d\lambda \leqslant 0, k = 1, 2, \ldots, n \\ (c)\int f_0 d\lambda = 1 \\ 0 \leqslant \omega(A) \leqslant \omega(B),\ for\ A \subseteq B, \quad A, B \in P(Y); \\ 0 \leqslant \lambda(C) \leqslant \lambda(D),\ for\ C \subseteq D, \quad C, D \in P(X); \end{cases}$$

When the efficiency measures λ on X, ω on Y are all additive, the CH-CCR model (10) is coincide with the classical CCR model, therefore CH-CCR model is a generalization of the classical CCR.

The CH-CCR model (10) are in essence a linear programming. For the Choquet integral $(c)\int f_k d\lambda$ (or $(c)\int g_k d\omega$), once the integrand f_k (or g_k) is given, the calculation of its Choquet integral only involves the value of λ(or ω) at the sets in a chain from the universal set X (or Y) to the empty set. In order to express the $(c)\int f_k d\lambda$ (or $(c)\int g_k d\omega$) in an explicit linear form of the unknown parameters that are the value of λ(or ω), we introduce an alternate calculation formula for the Choquet integral as follows (Wang *et al.*, 2010):

$$c)\int f_k d\lambda = \sum_{j=1}^{2^s - 1} z_{jk}\lambda_j \tag{11}$$

Where $\lambda_j = \lambda\left(\cup_{j_i = 1}\{x_i\}\right)$ if j is expressed in terms of binary digits $j_s j_{s-1} \ldots j_2 j_1$ for every $j = 1, 2, \ldots, 2^s - 1$ and:

$$z_{jk} = \begin{cases} \min\limits_{i | frc(j/2^i) \in [\frac{1}{2}, 1)} f_k(x_i) - \max\limits_{i | frc(j/2^i) \in [0, \frac{1}{2})} f_k(x_i) & \text{if it is} > 0 \text{ or } j = 2^s - 1 \\ 0 & \text{otherwise} \end{cases}$$

Similarly,

$$(c)\int g_k d\omega = \sum_{j=1}^{2^m - 1} \eta_{jk}\omega_j,$$

where $\omega_j = \omega\left(\cup_{j_i=1}\{y_i\}\right)$ if j is expressed in terms of binary digits $j_s j_{s-1}\ldots j_2 j_1$ for every $j=1, 2,\ldots, 2^m-1$ and:

$$\eta_{jk} = \begin{cases} \min\limits_{i|frc(j/2^i)\in[\frac{1}{2},1)} g_k(x_i) - \max\limits_{i|frc(j/2^i)\in[0,\frac{1}{2})} g_k(x_i) & \text{if it is} > 0 \text{ or } j = 2^m-1 \\ 0 & \text{otherwise} \end{cases}$$

Therefore, through the above transformations, the CH-CCR model (10) can transform a linear programming with decision variables λ_j ($j=1,2,\ldots,2^s-1$) and ω_j ($j=1,2,\ldots,2^{\mathrm{m}}-1$):

$$\max\left\{h_0 == \sum_{j=1}^{2^m-1}\eta_{j0}\omega_j\right\}$$

Subject to:

$$\begin{cases} \sum\limits_{j=1}^{2^m-1}\eta_{jk}\omega_j - \sum\limits_{j=1}^{2^s-1} z_{jk}\lambda_j \leqslant 0,\ k=1,2,\ldots,n \\ \sum\limits_{j=1}^{2^s-1} z_{j0}\lambda_j = 1 \\ 0\leqslant\lambda(A)<\lambda(B);\ \text{for very } A,B\in P(X),\ A\subset B; \\ 0\leqslant\omega(C)<\omega(D),\ \text{for every } C,D\in P(Y),\ \mathrm{C}\subset D; \end{cases} \tag{12}$$

For the simple efficiency evaluation problem (less DMUs, less inputs, less outputs), we can solve the optimization problem (10) directly; for a reasonably large efficiency evaluation problem (large DMUs or large inputs or large outputs), we can use the genetic algorithm (Wang *et al.*, 2010) to solve the optimization problem (12).

Definition 5. The performance of DMU_0 is efficient iff there exist the optimal solution μ on $P(X)$, ν on P(Y) in $(CH-CCR)^I$ Model, such that:

$$h_0 = \frac{(c)\int g_0 dv}{(c)\int f_0 d\mu} = 1$$

The Definition 5 is equivalent to the following definition.

Definition 5'. The performance of DMU_0 is CH-CCR efficient if there exist the optimal solution λ on $P(X)$, ω on $P(Y)$ in (10), such that: $h_0=(c)=\int g_0 d\omega = 1, (c)\int f_0 d\lambda = 1$.

In order to obtain the overall efficiency value for every DMU, the optimal efficiency measures of inputs and outputs are calculated for each DMU using the CH-CCR model. Given the results of the first stage, the efficiency measures used by the DMU_p are denoted by λ_p on $P(X)$, ω_p on $P(Y)$, they can be utilized for calculating the peer rated efficiency for each of the other DMUs. The peer evaluation score, ϑ_{pk}, indicates the efficiency score for DMU_k using the efficiency measure obtained by DMU_p:

$$\vartheta_{pk} = \frac{(c)\int g_k d\omega_p}{(c)\int f_k d\lambda_p}$$

The cross-efficiency scores can be summarized in the cross-efficiency matrix. Note that $0\leqslant\vartheta_{pk}\leqslant 1$ and the elements in the diagonal, ϑ_{pp}, represent the standard

DEA efficiency score. Then the overall efficiency value of the DMU_k is calculated by:

$$\vartheta_p = \frac{1}{n}\sum_{k=1}^{n} \vartheta_{pk}$$

4. Illustration with numerial examples

In this section, we provide three numerical examples to show the performance of the proposed models. One simple example is given to illustrate the proposed models is consistent with the classical DEA when the interaction (correlations) between the input (output) variables are very low, and then we provide two numerical examples that have frequently appeared in the related literature to show the performance of the proposed models is different from the classical DEA for highly correlated input (or output) variables.

Example 1: in this example, we are to evaluate the community health center (DMU) of Hebei Province in China, evaluating input indexes are confirmed as the public expenditure (10,000 Yuan) X_1, the number of medical staffs X_2 and the fixed assets (10,000 Yuan) X_3; the output index is considered as the numbers of medical service (thousands) (including inpatient service and childhood immunization) Y_1, the numbers of management of chronic diseases (thousands) Y_2. The data are shown in Table I.

The correlation matrix of the inputs is:

$$\begin{pmatrix} 1 & -0.10663 & -0.09349 \\ -0.10663 & 1 & 0.024742 \\ -0.09349 & 0.024742 & 1 \end{pmatrix}$$

DMU	X_1	X_2	X_3	Y_1	Y_2
1	236	266	302	181	231
2	254	229	269	164	239
3	379	213	268	179	176
4	308	306	366	221	222
5	312	260	332	188	221
6	298	398	279	211	311
7	286	329	368	231	267
8	279	306	399	198	243
9	305	332	297	238	275
10	288	309	308	243	292
11	246	336	332	190	242
12	214	320	309	188	283
13	269	303	298	209	204
14	288	296	336	194	268
15	332	380	312	203	235
16	268	288	359	239	206
17	256	269	378	216	173
18	299	271	319	228	188
19	245	332	277	231	219
20	298	269	338	232	191

Table I. Data of inputs and outputs

The correlation matrix of the outputs is:

$$\begin{pmatrix} 1 & 0.07484 \\ 0.07484 & 1 \end{pmatrix}$$

Applying the CCR and CH-CCR models given before to the data in Table I, we obtain the results shown in Table II.

As shown in Table II, the efficiency evaluation of CCR and CH-CCR models are same (although the efficiency value are different), and the rank of the overall efficiency is also the same. The reason may be that the correlations between the input (output) variables are very low, so we may consider the input (output) variables are mutually independent and there little interactions between variables (inputs and outputs).

Example 2: the data set is taken from a previous study (Guo, Jia and Qiu) for comparing the efficiency of eight retail enterprises of Chengdu city in China, evaluating input indexes are confirmed as production flexibility (day) X_1, order advance (day) X_2, and guaranteeing cost or the returned merchandise cost (10,000 Yuan) X_3; the output index is considered as complete execution rate of orders (per cent) Y_1, satisfaction rate of customer (per cent) Y_2, and intelligence capital rate (per cent) Y_3. DMU_1 ($j = 1, 2, \ldots, 8$) means each retail enterprise in Chengdu city. Data are shown in Table III.

The correlation matrix of the inputs is:

$$\begin{pmatrix} 1 & 0.35381 & 0.50931 \\ 0.35381 & 1 & 0.45547 \\ 0.50931 & 0.45547 & 1 \end{pmatrix}$$

DMU	CCR efficiency	Overall efficiency value for CCR	Overall efficiency rank for CCR	CH-CCR efficiency	Overall efficiency value for CH-CCR	Overall efficiency rank for CH-CCR
DMU1	0.92449	0.79411	8	0.92491	0.84765	8
DMU2	1	0.81906	5	1	0.87617	5
DMU3	1	0.82673	4	1	0.88293	4
DMU4	0.87827	0.72672	16	0.91799	0.80108	16
DMU5	0.87934	0.72156	17	0.91372	0.79985	17
DMU6	1	0.76411	13	1	0.82717	13
DMU7	0.93242	0.79292	9	0.942698	0.84666	9
DMU8	0.84351	0.71915	18	0.842796	0.78986	18
DMU9	1	0.82856	3	1	0.88653	3
DMU10	1	0.90077	1	1	0.96356	1
DMU11	0.86797	0.71865	19	0.87946	0.75164	19
DMU12	1	0.83276	2	1	0.91485	2
DMU13	0.89589	0.74091	14	0.90078	0.81457	14
DMU14	0.91996	0.76867	11	0.83342	0.83461	11
DMU15	0.81219	0.63864	20	0.81226	0.71519	20
DMU16	1	0.80526	7	1	0.86374	7
DMU17	0.96309	0.73989	15	0.9676	0.80638	15
DMU18	1	0.76761	12	1	0.8305	12
DMU19	1	0.8085	6	1	0.86891	6
DMU20	1	0.77637	10	1	0.8385	10

Table II.
Results of the CCR model

The correlation matrix of the outputs is:

$$\begin{pmatrix} 1 & 0.15762 & 0.00489 \\ 0.15762 & 1 & 0.60974 \\ 0.00489 & 0.60974 & 1 \end{pmatrix}$$

Applying the CCR and CH-CCR models given before to the data in Table III, we obtain the results shown in Tables IV-VI.

From the correlation matrix of inputs and outputs, there are interaction between some variables (inputs or outputs). As shown in Tables IV and V, the efficiency evaluation of CCR and CH-CCR models are slightly different, both the CCR and CH-CCR models define DMU_1, DMU_3 and DMU_8 as efficient, the DMU_5 is efficient in CCR model, but inefficient in CH-CCR model, the differences are probably caused by the interaction of variables (inputs or outputs). On the other hand, it can be seen from Table V that the input x_2, x_3 and the output y_3 have no contribution to the efficiency evaluation independently, but their joint contribution with other indexes cannot be ignored. And from Table VI, it can be seen that because the interaction of variables are little and the efficiency differences among the DMUs are not big enough, so in the cross-evaluation, although the overall efficiency values are different, but the rank of the overall efficiency is the same.

	Inputs			Outputs		
DMU	x_1	x_2	x_3	y_1	y_2	y_3
DMU1	15	15	0.05	0.80	0.800	0.42
DMU2	70	25	0.10	0.90	0.900	0.53
DMU3	45	16	0.07	0.96	0.885	0.47
DMU4	40	30	0.12	0.85	0.750	0.32
DMU5	35	25	0.11	0.75	0.845	0.44
DMU6	60	18	0.15	0.85	0.755	0.25
DMU7	55	20	0.08	0.70	0.850	0.51
DMU8	30	12	0.09	0.95	0.700	0.46

Table III. Inputs and outputs data table

		Input weights			Output weights		
DMU	CCR efficiency	x_1	x_2	x_3	y_1	y_2	y_3
DMU1	1	0.004167	0.0625	0	0	1.25	0
DMU2	0.7894	0	0.026619	3.345324	0	0	1.348961
DMU3	1	0.000573	0.055859	1.149595	0	1.129944	0
DMU4	0.7968	0.000084	0.018961	3.565185	0.5799099	0	0
DMU5	1	0.000371	0.036202	0.74504	0	0.732313	0
DMU6	0.719	0	0.055556	0	0	0.952381	0
DMU7	0.9495	0	0.033273	4.181655	0	0	1.686151
DMU8	1	0.018129	0.038012	0	1.052632	0	0

Table IV. Results of the CCR model

Table V.
Results of the CH-CCR model

DMU	CH-CCR efficiency	Efficiency measures of input set							Efficiency measures of output set						
		{1}	{2}	{3}	{1,2}	{1,3}	{2,3}	{1,2,3}	{1}	{2}	{3}	{1,2}	{1,3}	{2,3}	{1,2,3}
DMU1	1	0.066667	0	0	0	0.016402	0	0.066667	0	0	0	0	0	0	2.380952
DMU2	0.7149	0	0	0	0.026619	0	0	3.371942	0	0	0	0	0	0	1.348921
DMU3	1	0	0	0	0.034549	0	0	6.423438	1.041667	0	0	1.041667	1.041667	0	1.041667
DMU4	0.4929	0.000084	0	0	0.019045	0.000084	0	3.58423	0.57991	0	0	0.57991	0.57991	0	0.57991
DMU5	0.6188	0.000371	0	0	0.03657	0.000371	0	0.781622	0	0.732313	0	0.732313	0	0.732313	0.732313
DMU6	0.719	0	0	0	0.055556	0	0	0.055556	0	0	0	0.952381	0	0	0.952381
DMU7	0.8599	0	0	0	0.03327	0	0	4.214928	0	0	0	0	0	0	1.686151
DMU8	1	0.000153	0	0	0.03457	0.000153	0	6.505965	1.052632	0	0	1.052632	1.052632	0	1.052632

Note: For simplicity, input set $\{x_i, x_j, x_k\}$ is denoted by $\{i, j, k\}$

Example 3: a real data set taken from a previous study by Sherman and Gold (1985) for comparing the efficiency of 14 bank branches is given in Table VII. The comparison is based on three input and four output as follows:

Input 1: rent (thousands of dollars).
Input 2: full time equivalent personnel.
Input 3: supplies (thousands of dollars).
Output 1: loan applications, new pass-book loans, life insurance sales.
Output 2: new accounts, closed accounts.
Output 3: travelers checks sold, bonds sold, bonds redeemed.
Output 4: deposits, withdrawals, checks sold, treasury checks issued, B per cent checks, loan payments, pass-book loan payments, life insurance payments, mortgage payments.

The correlation matrix of the inputs is:

$$\begin{pmatrix} 1 & 0.31997 & 0.93579 \\ 0.31997 & 1 & 0.3679 \\ 0.93579 & 0.3679 & 1 \end{pmatrix}$$

DMU	Overall efficiency value for CCR	Overall efficiency rank for CCR	Overall efficiency value for CH-CCR	Overall efficiency rank for CH-CCR
DMU1	0.989286	2	0.988638	2
DMU2	0.626938	5	0.637388	5
DMU3	0.938803	3	0.949344	3
DMU4	0.437821	8	0.429798	8
DMU5	0.572817	6	0.544193	6
DMU6	0.570601	7	0.513081	7
DMU7	0.717636	4	0.693837	4
DMU8	1	1	1	1

Table VI. Overall efficiency value for CCR model and CH-CCR model

DMU	Input 1	Input 2	Input 3	Output 1	Output 2	Output 3	Output 4
1	140,000	42,900	87,500	484,000	4,139,100	59,860	2,951,430
2	48,800	17,400	37,900	384,000	1,685,500	139,780	3,336,860
3	36,600	14,200	29,800	209,000	1058,900	65,720	3,570,050
4	47,100	9,300	26,800	157,000	879,400	27,340	2,081,350
5	32,600	4,600	19,600	46,000	370,900	18,920	1,069,100
6	50,800	8,300	18,900	272,000	667,400	34,750	2,660,040
7	40,800	7,500	20,400	53,000	465,700	20,240	1,800,250
8	31,900	9,200	21,400	250,000	642,700	43,280	2,296,740
9	36,400	76,000	21,000	407,000	647,700	32,360	1,981,930
10	25,700	7,900	19,000	72,000	402,500	19,930	2,284,910
11	44,500	8,700	21,700	105,000	482,400	49,320	2,245,160
12	42,300	8,900	25,800	94,000	511,000	26,950	2,303,000
13	40,600	5,500	19,400	84,000	287,400	34,940	1,141,750
14	76,100	11,900	32,800	199,000	694,600	67,160	3,338,390

Table VII. Sherman and Gold data set on 14 bank branches and their efficiency score

The correlation matrix of the outputs is:

$$\begin{pmatrix} 1 & 0.73518 & 0.55795 & 0.53992 \\ 0.73518 & 1 & 0.43056 & 0.45917 \\ 0.55795 & 0.43056 & 1 & 0.6922 \\ 0.53992 & 0.45917 & 0.6922 & 1 \end{pmatrix}$$

Applying the CCR and CH-CCR models given before to the data in Table VII, we obtain the results shown in Table VI.

The results in Table VIII show that the efficiency evaluation of CCR and CH-CCR models are different, the DMU3 is efficient in CCR model, but inefficient in CH-CCR model. And in the overall efficiency evaluation, because the interaction of variables are big enough, the ranks of the DMU3, DMU8 and DMU5, DMU7, DMU13 are different.

From the above numerial examples, we conclude that the interaction between variables (inputs or outputs) can affect the efficiency evaluation. When there are no any interaction between the variables (inputs and outputs), the proposed models is consistent with the classical DEA, and the efficiency evaluation of the proposed models is different from the classical DEA for highly correlated input (or output) variables.

The numerial examples also show that the proposed model is a generalization of the classical CCR.

5. Conclusion

In this paper, a novel DEA model is proposed for evaluating the efficiencies of DMUs, and the Choquet integrals are used to aggregate the multiple inputs and outputs into a single efficiency index. This model extends DEA model to evaluate the DMU with interactive variables (inputs or outputs), the classical DEA model is a special form. Besides, the proposed model is used to analysis some numerical examples and the results show good performance. In the future, the other DEA model with interactive variables would be studied.

DMU	CCR efficiency	CH-CCR efficiency	Overall efficiency value for CCR	Overall efficiency rank for CCR	Overall efficiency value for CH-CCR	Overall efficiency rank for CH-CCR
1	1	1	0.483847	13	0.460054	13
2	1	1	0.923078	1	0.923069	1
3	1	0.9688	0.847371	3	0.864444	4
4	1	1	0.667029	8	0.690591	8
5	0.9041	0.9041	0.567036	11	0.587961	10
6	1	1	0.868264	2	0.911811	2
7	0.782	0.782	0.566842	10	0.577323	12
8	1	1	0.817737	4	0.875426	3
9	1	1	0.366268	14	0.325619	14
10	1	1	0.694751	5	0.727538	5
11	0.9668	0.9668	0.679443	7	0.70027	7
12	0.8522	0.8522	0.610857	9	0.636658	9
13	0.9049	0.9049	0.54811	12	0.582011	11
14	1	1	0.689768	6	0.721158	6

Table VIII. Efficiency and overall efficiency value for CCR model and CH-CCR model

References

Adler, N. and Golany, B. (2001), "Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe", *European Journal of Operational Research*, Vol. 132 No. 2, pp. 18-31.

Adler, N. and Golany, B. (2002), "Including principal component weights to improve discrimination in data envelopment analysis", *Journal of the Operational Research Society*, Vol. 53 No. 9, pp. 985-991.

Adler, N. and Yazhemsky, E. (2010), "Improving discrimination in data envelopment analysis: PCA-DEA or variable reduction", *European Journal of Operational Research*, Vol. 202 No. 1, pp. 273-284.

Ancarani, A., Di Mauro, C. and Giammanco, M.D. (2009), "The impact of managerial and organizational aspects on hospital wards' efficiency: evidence from a case study", *European Journal of Operational Research*, Vol. 194 No. 1, pp. 280-293.

Baker, P.C. and Talluri, S. (1997), "A closer look at the use of data envelopment analysis for technology selection", *Computers & Industrial Engineering*, Vol. 32 No. 1, pp. 101-108.

Banker, R.A., Charnes, A. and Cooper, W.W. (1984), "Some models for estimating technical and scale inefficiencies in data envelopment analysis", *Management Science*, Vol. 30 No. 9, pp. 1078-1092.

Bardhan, I., Bowlin, W.F., Cooper, W.W. and Sueyoshi, T. (1996), "Models and measures for efficiency dominance in DEA", *Journal of the Operations Research*, Vol. 39 No. 3, pp. 333-344.

Charnes, A. and Cooper, W.W. (1962), "Programming with linear fractional functional", *Naval Research Logistics*, Vol. 9, pp. 181-185.

Charnes, A. and Cooper Rhodes, W.W.E. (1978), "Measuring the efficiency of decision making units", *European Journal of Operational Research*, Vol. 2 No. 6, pp. 429-44.

Charnes, A., Cooper, W.W., Golany, B., Seiford, L. and Stutz, J. (1985), "Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions", *Journal of Econometrics*, Vol. 30, pp. 91-107.

Choquet, G. (1954), "Theory of capacities", *Annales de l'Institut Fourier*, No. 5, pp. 131-295.

Cook, W.D. and Zhu, J. (2014), "DEA Cobb-Douglas frontier and cross-efficiency", *Journal of the Operational Research Society*, Vol. 65 No. 2, pp. 265-268.

Cooper, W.W., Timothy, W.R., Deng, H., Wu, J. and Zhang, Z. (2008), "Are state-owned banks less efficient? A long- vs short-run data envelopment analysis of Chinese banks", *International Journal of Operational Research*, Vol. 3 No. 5, pp. 533-556.

Denneberg, D. (1994), *Fuzzy Measure and Integral*, Kluwer Academic Publishers, Dordrecht, Boston, MA and London.

Doyle, J. and Green, R. (1994), "Efficiency and cross-efficiency in DEA: derivations, meanings and uses", *Journal of the Operations Research Society*, Vol. 45 No. 5, pp. 567-578.

Doyle, J.R. and Green, R.H. (1995), "Cross-evaluation in DEA: improving discrimination among DMUs", *INFOR*, Vol. 33 No. 3, pp. 205-222.

Dyson, R.G., Allen, R., Camanho, A.S., Podinovski, V.V., Sarrico, C.S. and Shale, E.A. (2001), "Pitfalls and protocols in DEA", *European Journal of Operational Research*, Vol. 132 No. 2, pp. 245-259.

Fengxia Dong, P.D. and Mitchell, J.C. (2015), "Measuring farm sustainability using data envelope analysis with principal components: the case of Wisconsin cranberry", *Journal of Environmental Management*, Vol. 147, pp. 175-183.

Green, R.H., Doyle, J.R. and Cook, W.D. (1996), "Preference voting and project ranking using DEA and cross-evaluation", *European Journal of Operational Research*, Vol. 90 No. 3, pp. 461-472.

Hyvarinen, A. and Oja, E. (2000), "Independent component analysis: algorithms and applications", *Neural Networks*, Vol. 13 Nos 4-5, pp. 411-430.

Jukti, K.K (1995), "Effects of multicollinearity in data envelopment analysis (DEA) models of product market efficiency", *Proceedings of the Academy of Marketing Science (AMS) Annual Conference*.

Kao, C. and Liu, S.-T. (2009), "Stochastic data envelopment analysis in measuring the efficiency of Taiwan commercial banks", *European Journal of Operational Research*, Vol. 196 No. 1, pp. 312-322.

Kao, L.-j., Lu, C.-j. and Chiu, C.-C. (2011), "Efficiency measurement using independent component analysis and data envelopment analysis", *European Journal of Operational Research*, Vol. 210 No. 2, pp. 310-317.

Nataraja, N.R. and Johnson, A.L. (2011), "Guidelines for using variable selection techniques in data envelopment analysis", *European Journal of Operational Research*, Vol. 215 No. 3, pp. 662-669.

Ramón, N. and Ruiz, J.L. (2010), "On the choice of weights profiles in cross-efficiency evaluations", *European Journal of Operational Research*, Vol. 207 No. 3, pp. 1564-1572.

Ray, S.C. and Jeon, Y. (2008), "Reputation and efficiency: a non-parametric assessment of America's top-rated MBA programs", *European Journal of Operational Research*, Vol. 189 No. 1, pp. 245-268.

Sexton, T.R., Silkman, R.H. and Hogan, A.J. (1986), "Data envelopment analysis: critique and extensions", in Silkman, R.H. (Ed.), *Measuring Efficiency: An Assessment of Data Envelopment Analysis*, Jossey-Bass, San Francisco. CA, pp. 73-105.

Sherman, H.D. and Gold, F. (1985), "Bank branch operating efficiency: evaluation with DEA", *J. Bank Finance*, Vol. 9 No. 2, pp. 297-315.

Sugeno, M. (1974), "Theory of fuzzy integrals and its applications", PhD thesis, Tokyo Institute of Technology, Tokyo.

Varias, A.D. and Sofianopoulou, S. (2012), "Efficiency evaluation of Greek commercial banks using", *DEA Journal of Applied Operational Research*, Vol. 4 No. 4, pp. 183-193.

Wang, Z. and Klir, G.J. (2008), *Generalized Measure Theory*, Springer, New York, NY.

Wang, Z., Yang, R. and Lung, K. (2010), *Nonlinear Integrals and their Applications in Data Mining*, World Scientific Publishing Company, New Jersey, London.

Further reading

Guo, J., Jia, L. and Qiu, L. (2006), "Research on supply chain performance evaluation based on DEA/AHP model", *Proceedings of the 2006 IEEE Asia-Pacific Conference on Services Computing (APSCC'06), Guangzhou, 12-15 December, 0-7695-2751-5/06*.

Corresponding author
Dr Hui Liu can be contacted at: jab@hbu.edu.cn

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**