Determination of Weights for the Ultimate Cross Efficiency: A Use of Principal Component Analysis Technique

Qifa Qu
School of Management, University of Science and Technology of China,
Hefei, Anhui Province, P. R. China, 230026

***Abstract*—Data Envelopment Analysis (DEA) has becoming more and more important in evaluating the performance of homogenous Decision Making Units (DMUs). Cross efficiency evaluation method, a DEA extension technique, can be utilized to identify efficient DMUs and to rank DMUs in a peer appraisal mode, instead of a pure self-evaluation of traditional DEA models. Traditionally, the ultimate cross efficiency is determined based on the average assumption. However it cannot ensure this result contains the most information of the cross-efficiency matrix (CEM). In the current paper, we use principal component analysis (PCA) to determine the ultimate cross-efficiency of each DMU and then rank them. Compared with the tradition average cross efficiency evaluation method, the method proposed in this paper can contain the most of the information of CEM. Finally, an empirical example is illustrated to examine the validity of the proposed method.**

***Index Terms*—DEA; Cross-efficiency; Principal component analysis (PCA); Weights**

I. INTRODUCTION

Data envelopment analysis (DEA), as a non-parametric programming techniue, can provide a relative efficiency measure for peer decision making units (DMUs) with multiple inputs and multiple outputs. DEA has become a very effective approach in identifying the best practice frontiers and evaluating the performance of each DMU. DEA has been extensively applied in performance evaluation and benchmarking of schools, hospitals, bank branches, production plants, and so on (Charnes et al, 1994; Charnes and Zlobec, 1978; Wang, Parkan and Luo, 2007; Saen, 2010). However, traditional DEA models are not very appropriate for ranking DMUs since they simply classify the units into two groups: efficient and inefficient as the work of Charnes et al. (1978). Moreover, it is often possible in DEA that some inefficient DMUs are in fact better overall performers than some efficient ones. This is because of the unrestricted weight flexibility problem in DEA by being involved in an unreasonable self-rated scheme (Dyson and Thannassoulis, 1988; Wong and Beasley, 1990). In order to maximize its own DEA efficiency, the DMU under evaluation always ignores other inputs and outputs, thus it heavily weighs few favorable measures from its own view.

Sexton et al. (1986) first proposed Cross efficiency evaluation method which is developed as a DEA extension technique, and can be utilized to identify efficient DMUs and to completely rank DMUs. The main idea of cross evaluation is to use DEA in a peer evaluation instead of a pure self-evaluation for providing an efficiency ordering among all the DMUs. Cross efficiency evaluation has been used in various applications, e.g., efficiency evaluations of nursing homes (Sexton et al. 1986), R&D project selection (Oral et al., 1991), preference voting (Green et al., 1996), and others.

In the traditional cross efficiency evaluation method, average cross efficiency has been widely used, however, there are still several disadvantages for utilizing the ultimate average cross efficiency to evaluate and rank DMUs (Jahanshahloo, Lotfi, Jafari and Maddahi, 2011), like the losing association with the weights by averaging among the cross efficiencies (Despotis, 2002), which means that this method can not clearly provide the weights to help decision maker improve his performance. Especially, the average cross efficiency measure is not good enough since it is not a Pareto solution. Considering the shortcomings above, Wu et al. (2009) eliminate the average assumption for determining the ultimate cross efficiency scores, and DMUs are considered as the players in a cooperative game, in which the characteristic function values of coalitions are defined to compute the Shapley value of each DMU, and the common weights associated with the imputation of the Shapley values are used to determine the ultimate cross efficiency scores. But all these methods cannot ensure its results contain the most information of the cross-efficiency matrix.

Principal Component Analysis (PCA) is widely used in multivariate statistics such as factor analysis. Azadeh and Ebrahimipour (2002, 2004) pointed out it had been used to reduce the number of variables under study and consequently by ranking and analysis of decision-making units, such as industries, universities, hospitals, cities, and so on. PCA is performed by identifying Eigen structure of the covariance or singular value decomposition of the original data. It has been studied with data envelopment analysis in many studies, but it was only used to validate

Corresponding author 's. E-mail: qqf@ustc.edu.cn.

doi:10.4304/jsw.7.10.2177-2181

the results of DEA or decrease the number of indicators under study. For example, Zhao et al. (2005) applied PCA to identify the input and output indicators of distibutors in, Azadeh and Ghaderi (2006, 2008) applied it to validate the results of DEA, and Liang, et al. (2009) applied PCA to deal with undesirable outputs and simultaneously reduce the dimensionality of data set, consequently increasing the discriminatory power of DEA. In the current paper, we use principal component analysis to determine the ultimate cross efficiency scores instead of average cross efficiency scores, which has some theoretical and practical advantages, for example, the proposed approach can contain most of the information of cross efficiency matrix (CEM) in cross evaluation. The rest of the paper unfolds as follows: Section 2 briefly reviews the original cross efficiency concept in DEA. Section 3 revisits the method of principal component analysis and introduces its application in determining the ultimate cross-efficiency. In Section 4, a numerical example is used to illustrate the proposed method. Finally, conclusions are made in Section 5.

II. CROSS-EFFICIENCY EVALUATION

Using the traditional denotations in DEA, we assume that there are a set of n DMUs, and each $DMU_j (j=1,2,\cdots,n)$ produces s different outputs using m different inputs which are denoted as $x_{ij}(i=1,2,\cdots,m)$ and $y_{rj}(r=1,2,\cdots,s)$, respectively. For any evaluated $DMU_d(d=1,2,\cdots,n)$, the efficiency score E_{dd} can be calculated by using the following CCR model [3].

$$\begin{aligned} &\max \sum_{r=1}^{s} \mu_r y_{rd} = E_{dd} \\ &s.t.\ \sum_{i=1}^{m} \omega_i x_{ij} - \sum_{r=1}^{s} \mu_r y_{rj} \ge 0,\ j=1,2,\cdots,n \\ &\quad \sum_{j=1}^{m} \omega_i x_{id} = 1 \\ &\quad \omega_i \ge 0, i=1,2,\cdots,m; \mu_r \ge 0, r=1,2,\cdots,s \end{aligned} \tag{1}$$

Cross efficiency is often calculated as a two-phase process. The first phase is calculated by using the DEA model (1). For each DMU under evaluation by model (1), we can obtain a set of optimal weights $w_{1d}^*, w_{2d}^*, ..., w_{md}^*, \mu_{1d}^*, \mu_{2d}^*, ..., \mu_{sd}^*$. Then the cross-efficiency of each DMU_j using the weights of DMU_d, namely E_{dj}, can be calculated as follows.

$$E_{dj} = \frac{\sum_{r=1}^{s} \mu_{rd}^* y_{rj}}{\sum_{i=1}^{m} \omega_{id}^* x_{ij}}, d, j=1,2,\cdots,n \tag{2}$$

where $\omega_{id}^*, i=1,...,m$ represent the optimal weights of the ith input and $\mu_{rd}^*, r=1,...,s$ represent the optimal weights of the rth output for DMU_d.

As shown in Table 1, E_{dj} is the efficiency score of DMU_j using the weights that $DMU_d (j=1,2,\cdots,n)$ has chosen. We can see that the efficiency scores of E_{dd} are calculated by the CCR model (1), so they can be seen as self-evaluation. Each of the columns of the cross efficiency matrix (CEM) in Table 1 is then averaged to get a mean cross efficiency measure for each DMU.

TABLE I
A GENERALIZED CROSS EFFICIENCY MATRIX

Rating DMU_j	Rated DMU_d				
	1	2	3	…	n
1	E_{11}	E_{12}	E_{13}	…	E_{1n}
3	E_{21}	E_{22}	E_{23}	…	E_{2n}
…	…	…	…	…	…
…	…	…	…	…	…
n	E_{n1}	E_{n2}	E_{n3}	…	E_{nn}
Mean	$\bar{E}_1$	$\bar{E}_2$	$\bar{E}_3$	…	$\bar{E}_n$

In Table 1, for each $DMU_j (j=1,2,...,n)$, the average of all $E_{dj}(d=1,2,...,n)$, namely $\bar{E}_j = \frac{1}{n}\sum_{d=1}^{n} \bar{E}_{dj}$, can be used as a new efficiency measure for DMU_j, and will be referred to as the cross efficiency score for DMU_j.

III. DETERMINATION OF ULTIMATE CROSS EFFICIENCY USING PCA

A. Principal Component Analysis

Principal Component Analysis (PCA) is widely used in multivariate statistics such as factor analysis. The objective of PCA is to identify a new set of variables such that each new variable, called a principal component, is a linear combination of original variables and also the first new variable, accounts for the maximum variance in the sample data and so on, and The new variables (principal components) are uncorrelated.

Assume there are n samples, each sample has p indicators (variables) $X_1, X_2, \cdots, X_p$, and the original data matrix is

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \underline{\underline{\Delta}}(X_1, X_2, \cdots, X_p) \tag{3}$$

where

$$X_i=\begin{bmatrix} X_{1i} \\ X_{2i} \\ \vdots \\ X_{ni} \end{bmatrix} \qquad i=1,\cdots,p.$$

The linear combination (i.e., composite indicator vector) of p vectors $X_1,...,X_p$ in data matrix X can be written as follows:

$$\begin{cases} F_1 = a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p \\ F_2 = a_{12}X_1 + a_{22}X_2 + \cdots + a_{p2}X_p \\ \quad\cdots\cdots \\ F_p = a_{1p}X_1 + a_{2p}X_2 + \cdots + a_{pp}X_p \end{cases} \tag{4}$$

which can also abbreviated as $F_i = a_{1i}X_1 + a_{2i}X_2 + ... + a_{pi}X_p, i=1,..., p.$

The above equations requires $a_{1i}^2 + a_{2i}^2 + \cdots + a_{pi}^2 = 1, i=1,\cdots,p$, and other two constraints are as follows:

(i) F_i and F_j $(i \neq j, i, j=1,\cdots,p)$ are not relevant.

(ii) F_1 is a linear combination of all $X_1,\cdots,X_p$, which has the largest variance. F_2 is a linear combination that has largest variance of all of $X_1,\cdots,X_p$ which are not related to F_1. Analogy, F_p is a linear combination that has largest variance of all of $X_1,\cdots,X_p$, which is not related to $F_1, F_2, \cdots, F_{p-1}$.

B. The Proposed Method

In this section, we will introduce our proposed method, its main objective is to identify a few principal components that can cover most of information in CEM and to obtain their feature vectors and contribution rates. Then, the weight of each indicator can be obtained by multiplying the corresponding contribution rates with their own feature vectors. Finally, we can get the ultimate cross efficiency scores of each DMU using the weights obtained.

Firstly, we transform the traditonal cross efficiency matrix into its transpose matrix, and denoted as follows:

$$E'=\begin{bmatrix} E_{11} & E_{21} & \cdots & E_{n1} \\ E_{12} & E_{22} & \cdots & E_{n2} \\ \vdots & \vdots & & \vdots \\ E_{1n} & E_{2n} & \cdots & E_{nn} \end{bmatrix}$$

where $E_i^{'}=\begin{bmatrix} E_{i1} \\ E_{i2} \\ . \\ . \\ . \\ E_{in} \end{bmatrix}, i=1,...,n.$

Next, we introduce the seven steps for determining the weights of each indicator based on principal component analysis technique.

- Step 1. Standardization

We denote $E_i'=\begin{bmatrix} E_{i1} \\ E_{i2} \\ \vdots \\ E_{in} \end{bmatrix}$, i=1,2,…,n . For each E_i', we use the Standardized formula $T_i = \dfrac{E_i' - E(E_i')}{\sqrt{Var(E_i')}}, i=1,...,n$ to get the standardized matrix $T=[T_1, T_2, ..., T_n]$.

- Step 2. Establishing the correlation coefficient matrix between variables

$$R=(r_{ij})_{n\times n} \qquad \mathrm{R}=T'T$$

- Step3. Solving the Eigen value $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ of R and the corresponding unit eigenvector

$$a_1=\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix}, a_2=\begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix}, \cdots, a_p=\begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{nn} \end{bmatrix}$$

- Step 4. Writing down the principal component

$$F_i = a_{1i}T_1 + a_{2i}T_2 + \cdots + a_{ni}T_n \qquad \mathrm{i}=1,\cdots,\mathrm{n}$$

If the cumulative contribution rate of the first k principal components reaches 85%, it shows that the k principal components contain the most of the indicators' information for measuring. In this way, we choose the first k principal components to determine the indicators' weights in Step 5.

- Step 5. Determination of PCA-based weights

The weight of the ith indicator is

$$w_i = \sum_{j=1}^{k}[(\lambda_j \Big/ \sum_{m=1}^{n}\lambda_m)/(\sum_{p=1}^{k}(\lambda_p \Big/ \sum_{m=1}^{n}\lambda_m))*a_{ij}] \quad i=1,\cdots,n \tag{5}$$

- Step 6. Determination of efficiency value

The efficiency value using the weights in Step 5 is shown as follows:

$$E_i = w_1E_{1i} + w_2E_{2i} + \cdots + w_nE_{ni} \quad i=1,\cdots,n \tag{6}$$

- Step 7. Determination of the ultimate cross efficiency

In order to limit the efficiency E_i into the range of [0, 1], we use the following formula

$$E_i^* = E_i / \max\{E_1,...,E_n\} \tag{7}$$

IV. ILLUSTRATION

In order to illustrate the method which has been proposed above, we consider a simple numerical example shown in Table 2. There are five DMUs, each DMU has three inputs X_1, X_2, X_3 and two outputs Y_1, Y_2. After solving the CCR model，we can obtain the cross efficiency matrix listed in Table 3.

TABLE II
DATA OF THE EMPIRICAL EXAMPLE

	X_1	X_2	X_3	Y_1	Y_2
DMU_1	7	7	7	4	4
DMU_2	5	9	7	7	7
DMU_3	4	6	5	5	7
DMU_4	5	9	8	6	2
DMU_5	6	8	5	3	6

TABLE III
CROSS EFFICIENCY MATRIX

	DMU_1	DMU_2	DMU_3	DMU_4	DMU_5
DMU_1	0.6578	0.9333	1.0000	0.8	0.4500
DMU_2	0.4478	1.000	0.9965	0.7323	0.4643
DMU_3	0.3710	0.7489	1.0000	0.2092	0.6402
DMU_4	0.4587	1.0000	0.9313	0.8571	0.3817
DMU_5	0.4082	0.7143	1.000	0.1786	0.8571

Then we transform the cross efficiency matrix of Table 3 into its transpose matrix shown in Table 4. All data in Table 4 are during 0 and 1, so we should not standardize them. Then using the PCA technique, we can get the several principal components that can cover most of the information of the matrix. Total variance explained from SPSS is shown in Table 5.

TABLE IV
THE TRANSPOSE OF CROSS EFFICIENCY MATRIX

	DMU_1	DMU_2	DMU_3	DMU_4	DMU_5
DMU_1	0.6578	0.4478	0.371	0.4587	0.4082
DMU_2	0.9333	1.0000	0.7489	1.0000	0.7143
DMU_3	1.0000	0.9965	1.0000	0.9313	1.0000
DMU_4	0.8000	0.7323	0.2092	0.8571	0.1786
DMU_5	0.4500	0.4643	0.6402	0.3817	0.8571

TABLE V
TOTAL VARIANCE EXPLAINED

Com ponent	Initial Eigen values			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.342	66.848	66.848	3.342	66.848	66.848
2	1.578	31.563	98.411	1.578	31.563	98.411
3	0.078	1.57	99.981			
4	0.001	0.019	100			
5	-2.30E-016	-4.60E-015	100			

From Table 5, we can see that the first two principal components cover the 98.411% information of the cross efficiency matrix, so we choose them. The first two principal components and their corresponding weights in every indicator are shown in Table 6.

TABLE Ⅵ
COMPONENT MATRIX

	Component Matrix	
	1	2
a	0.899	-0.379
b	0.975	-0.192
c	0.742	0.668
d	0.877	-0.464
e	0.511	0.859

Then we get the weights of all indicators by formula (5) as follows

$$w=[w_1,...,w_5]=[0.2373, 0.3133, 0.4463, 0.2074, 0.4092]$$

and the corresponding efficiency values of all DMUs by formula (6) are

$$E=[E_1,...,E_5]=[0.7259, 1.3666, 1.5989, 0.7642, 0.9672]$$

Finally, by formula (7), we can obtain the ultimate cross efficiency as follows

$$E^*=[E_1^*,...,E_5^*]=[0.454, 0.8547, 1, 0.478, 0.6049]$$

V. CONCLUSIONS

Aiming at the flaws when the ultimate average cross efficiency scores are used to evaluate and to rank DMUs, we eliminate the assumption of average and utilize the PCA technique to determine the ultimate cross efficiency scores for each DMU. Finally, a numerical example is illustrated to prove the effectiveness of the proposed approach. We should point out that the numerical example in this paper is chosen only for illustrative purposes and for better understanding of the main principles of the proposed approach, so how the proposed approach can be used in the real-world application case is obvious an interesting research in the future.

REFERENCES

[1] Azadeh, A., Ebrahimipour, A. (2002). An integrated approach for assessment of manufacturing sectors based on machine performance: the cases of automotive and food and beverages industries, *Proceeding of the Second International Conference on Manufacturing on Complexity University of Cambridge*, UK.

[2] Azadeh, A., Ebrahimipour, A. (2004). An integrated approach for assessment and ranking of manufacturing systems based on machine performance. *International Journal of Industrial Engineering*, 11, 349-363.

[3] Azadeh, A., Ghaderi, S. F., Maghsoudi, A. (2006). Location optimization of solar plants by an integrated multivariable DEA-PCA method. *IEEE International Conference on Industrial Technology*, 11, 2748-2754.

[4] Azadeh, A., Ghaderi, S. F., Maghsoudi, A. (2008). Location optimization of solar plants by an integrated hierarchical DEA PCA approach. *Energy Policy*, 36, 3993-4004.

[5] Charnes, A., Cooper, W. W., Lewin, A. Y., Seiford, L. M. (Eds.)(1994). *Data envelopment analysis: Theory, methodology, and applications*. Boston: Kluwer.

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

[7] Charnes, A., Zlobec, S. (1989). Stability of efficiency evaluations in data envelopment analysis. *Mathematical Methods of Operations Research*, 3, 167-179.

[8] Despotis, D. K. (2002). Improving the discriminating power of DEA: Focus on globally efficient units. *Journal of the Operational Research Society*, 53, 314–323.

[9] Dyson, R. G., Thannassoulis, E. (1988). Reducing weight flexibility in data envelopment analysis. *Journal of Operational Research Society*, 39, 563–576.

[10] Green, R., Doyle, J., Cook, W. D. (1996). Preference voting and project ranking using DEA and cross-evaluation. *European Journal of Operational Research*, 90, 461-472.

[11] Jahanshahloo, G. R., Lotfi, F. H., Jafari, Y., Maddahi, R. (2011). Selecting symmetric weights as a secondary goal in DEA cross-efficiency evaluation. *Applied Mathematical Modelling*, 35, 544-549.

[12] Liang, L., Li, Y. J., Li, S. B. (2009). Increasing the discriminatory power of DEA in the presence of the undesirable outputs and large dimensionality of data sets with PCA. *Expert systems with applications*, 36, 5895-5899.

[13] Oral, M., Kettani, O. Lang, P. (1991). A methodology for collective evaluation and selection of industrial R&D projects. *Management Science*, 37, 871–883.

[14] Saen, R. F. (2010). Restricting weights in supplier selection decisions in the presence of dual-role factors. *Applied Mathematical Modelling*, 34, 2820-2830.

[15] Sexton, T. R., Silkman, R. H., Hogan, A. J. (1986). *Data envelopment analysis: Critique and extensions*. In Silkman, R. H. (Ed.). Measuring efficiency: An assessment of data envelopment analysis (vol. 32, pp. 73–105). San Francisco: Jossey-Bass.

[16] Wang, Y. M., Parkan, C., Luo, Y. (2007). Priority estimation in the AHP through maximization of correlation coefficient. *Applied Mathematical Modelling*, 31, 2711-2718.

[17] Wong, Y. H. B., Beasley, J. E. (1990). Restricting weight flexibility in data envelopment analysis. *Journal of the Operational Research Society*, 41, 829–835.

[18] Wu, J., Liang, L., Yang, F. (2009). Determination of the weights for the ultimate cross efficiency using Shapley value in cooperative game. *Expert Systems with Applications*, 36, 872-876.

[19] Zhao, Y. Q., Chen, J., Han, J. X. (2005). Evaluation of distributors using PCA-DEA algorithmic model. *Industrial Engineering Journal*, 6, 95-98.