J. Ind. Eng. Int., 7(15), 43-50, Fall 2011
ISSN: 1735-5702

Considering undesirable variables in PCA-DEA method: A case of road safety evaluation in Iran

***Alimohammad Ahmadvand*[1]*; Zeinab Abtahy*[2*]*; Mahdi Bashiri*[3]**

[1] *Professor, Dept. of Industrial Engineering, Emam Hossein University, Tehran, Iran*

[2] *M.Sc., Dept. of Industrial Engineering, Khoramabad Branch, Islamic Azad University, Khoramabad, Iran*

[3]*Associate Professor, Dept. of Industrial Engineering, Shahed University, Tehran, Iran*

Received: 11 August 2009; Revised:14 June 2010; Accepted: 3 August 2010

Abstract: This paper presents an approach composed from Data Envelopment Analysis (DEA) model and a multivariate statistical method, Principle Component Analysis (PCA), considering undesirable input and output variables. PCA is used to improve discrimination power of DEA and making variables as independent as possible to avoid overlapping of Decision Making Units (DMU's) information. The advantage of the proposed approach is considering undesirable output and input variables simultaneously in PCA-DEA method; furthermore it was applied for performance assessment of different province's road safety level in Iran.

Keywords: *DEA; PCA; PCA-DEA; Performance evaluation; Road safety; Undesirable variables*

1. Introduction

Standard DEA models rely on some restrictive assumptions, e.g. variables need to be strictly positive and as independent as possible, increasing inputs and decreasing outputs are not allowed (treating undesirable variables), excessive number of variables toward DMUs often arising discrimination problems (considering efficient DMUs as inefficient and vice versa) and so on. A unique feature of this paper is proposing a new method considering aforementioned restrictions simultaneously. Several approaches have been proposed to improve discrimination and full ranking; deriving common weights (e.g. Hermans *et al.*, 2009), setting a range of weights corresponding experts' opinions to restrict input and output's weight (e.g. (Hermans et al., 2009)), and reducing data. Data reduction is a way to improve discrimination power of DEA performing by the use of multivariate statistical analysis methods such as Variable Reduction (Jenkins and Anderson, 2003) and PCA. Originally, PCA is a data reduction method. PCA was used to evaluate comparable DMUs prior to DEA (Jenkins and Anderson, 2003). The idea of working with the ratio of every output to inputs proposed by Zhu (1998) and slightly modified by Permchandra (2001). Shanmugam and Johnson (2007) mentioned a new method to evaluate DMUs by PCA.

PCA-DEA was persuaded by Adler and Yazhemsky (2010), Adler and Golany (2001), Shanmugam and Johnson (2007) and rest to attain "the best" of the DEA and PCA approaches and to improve discrimination power of DEA.

To encounter possibility of negative PCA new variables called Principle Components (PCs), translation invariance property guarantees that (an envelopment form of) DEA models using the original-negative-data and the same model using the translated-positive data are equivalent, i.e. both have the same optimal solution (Agha and Lawrence, 1990).

In the case of undesirable DEA outputs, ignoring undesirable variables or including them like inputs does not reflect the true production process (Lawrence *et al*., 2002). Under the context of BCC model, Zhu (1998) introduced a linear transformation approach preserving convexity and linearity of BCC model and it can be used only in BCC models. A non-radial DEA model was used by Hadi *et al.* (2005) to treat both undesirable input and output factors.

Liang *et al.,* (2009) proposed a PCA-DEA model based on Zhu (1998) in the presence of undesirable outputs. In this paper a composite PCA-DEA method is proposed following Liang *et al.* (2003) and Shanmugam and Johnson (2007) while considering undesirable output and input variables simultaneously for the first time.

The materials are organized as follows in the article: In Section 2, the DEA, PCA and PCA-DEA and proposed methodology are explained briefly. Section 3 describes a case study of Iran's

*Corresponding Author Email: zeinab.abtahy@yahoo.com
Tel.: + 98- 9163675398

provinces road safety level evaluation solved by proposed method and corresponding results. Finally, the researchers summarize and conclude the paper in Section 4.

2. Methods

There are some difficulties in performance evaluation of DMUs in DEA method:

(1) Overlapping of DMU's information because of variables dependency

(2) Necessity of full ranking achievement and

(3) Existing undesirable variables.

So the researchers proposed a new composite PCA-DEA model dealing with aforementioned restriction.

2.1. Data envelopment analysis

Twenty years after Farrell's seminal work, DEA was proposed by Charnes to evaluate the relative efficiency of the similar DMUs (Khodabakhshi, 2003). To overcome DEA positivity restriction, Agha and Lawrence, (1990) established conditions under which translation invariance hold in DEA. Indeed, Agha and Lawrence (1990) provide the classification invariance concerning to additive or BCC models and generalize the results on the matter of translation invariance by allowing inputs and outputs to take not only zero but also negative values in DEA. They prove linear translation could not change the differentiation among efficient, weak efficient and non-efficient solutions (FU Bo-xin *et al.*, 2007). In this way, there is no change to the rank of DMUs. Lovell and Pastor (1995) classified basic DEA models in translation invariance property and mentioned that no currently known (basic) DEA models satisfies both desirable translation invariance and unit invariance properties. They proposed a normalized BCC model satisfying unit invariance and partial translation invariance, and normalized additive model satisfying both properties simultaneously. Robert (1996) mentioned that the multiplier form of DEA models did not have the same translation invariance properties as the envelopment form Table1.

Input and output orientations are combined in additive models. Efficiency score is not measured explicitly but is implicitly present in output and input slack respectively. More inefficiencies in both inputs and outputs. DMU_0 is Additive efficient if and only if output and input slack variables were equal to zero. Moreover, whereas objective function of output or input oriented DEA model reflects only weak efficiency, the objective in additive model reflects:

2.1.1. Weighted additive DEA model

From Table 1, lack of unit invariance property is observed for additive models. Lovell and Pastor (1995) proposed a normalized additive model possessing both translation invariance and unit invariance properties.

Aassume n DMUs to be evaluated; DMU_j , $j=1,\dots,n$, consume $X_j=(x_{1j},\dots,x_{mj})$, a column vector of m inputs and produce $Y_j=(y_{1j},\dots,y_{rj})$, a column vector of r outputs.

Definition 1. The normalized weighted additive DEA model has the same constraints as the additive DEA model. It replaces the objective function in additive DEA model with

$$\min \sum_{j=1}^{n}\sum_{k=1}^{r} w_{jk}^{+} s_{k}^{+} - \sum_{j=1}^{n}\sum_{i=1}^{m} w_{ji}^{-} s_{i}^{-} \tag{1}$$

Where;

$$w_k^{+}=(\frac{1}{\delta_k^{+}}),\ w_i^{-}=(\frac{1}{\delta_i^{-}}),\ k=1,\dots,r,$$

$$i=1,\dots,m,\ \delta_k^{+}=(\delta_1^{+},\dots,\delta_r^{+}), \tag{2}$$

$$\delta_i^{-}=(\delta_1^{-},\dots,\delta_m^{+})\delta_k^{+},\delta_i^{-}$$

Are sampling standard deviations of output and input variables respectively. So normalized weighted additive DEA model to evaluate DMU_0 with data (X_0, Y_0) is presented as follows:

Table 1: Summary of basic DEA model's characteristics.

Model	CCR-I	CCR-O	BCC-I	BCC-O	ADD
inputs	Semi- positive	Semi- positive	Semi- positive	Free	free
outputs	free	free	Free	Semi- positive	free
Translation invariance		-	With respect only to outputs	with respect only to inputs	With convexity constraint only
Unit invariance					
Return to scale	CRS	CRS	VRS	VRS	CRS or VRS
Objective function	[0, 1]	[0, 1]	[0, 1]	[0, 1]	

Model (1): weighted additive DEA model

$$\min \sum_{j=1}^{n}\sum_{k=1}^{r} w_{jk}^{+} s_{k}^{+} - \sum_{j=1}^{n}\sum_{i=1}^{m} w_{ji}^{-} s_{i}^{-}$$

$$s.t.\ \sum_{j=1}^{n} Y_{kj}\lambda_j - s_k^{+} = y_{ko} \qquad (k = 1,...,r)$$

$$-\sum_{i=1}^{m} X_{ij}\lambda_j - s_i^{-} = y_{ko} \qquad (i = 1,...,m)$$

$$\sum_{j=1}^{n} \lambda_j = 1 \qquad (j = 1,...,n)$$

$$\lambda_j, s_k^{+}, s_i^{-} \geq 0$$

Where s^- is a column m-vector of input slack variables, s^+ is a column r-vector of output slack variables; λ is a column vector of n non-negative variables satisfying the convexity constraint $\left[\sum_{j=1}^{n}\lambda_j = 1\right]$.

2.2. PCA-DEA method

To improve discrimination and variables dependency, PCA was applied to deal with the data before implementation of DEA model. PCA transforms a set of correlated variables into an uncorrelated (and smaller number of) new variables, called principle component (PCs), which are linear combination of original variables with minimum loss of information. The first principle, accounts for the maximum variance in the sample data, the second new variable, accounts for the maximum variance which is not considered by first component and so on. To implement PCA Cauchy distributed and highly correlated variables are necessary.

For comprehensive domain envelopment one may choose as many indices as possible, but in a DEA model, the selected indices must be as independent as possible and multiple inputs and outputs lead to multiple correlations making the information of DMUs overlap. Also an excessive number of input and output variables toward number of DMUs result in a large number of efficient units. Thus it is preferable to keep this ratio low. PCA can be used for these purposes with minimum loss of information whilst ensuring similar results to those achieved by the original DEA model.

Adler *et al.*0 0 proposed a PCA-DEA method to improve discrimination power of DEA and proved it is a more appropriate method than variable reduction (Jenkins and Anderson, 2003). Shanmugam and Johnson (2007) mentioned that: “The PCA does not rate the DMU in the same way as DEA does” and only when the DMU’s inputs and outputs are stochastic (preferably by multivariate Gaussian distribution), PCA could alternatively be used to rank DMUs and because ratio of every output to inputs proposed by Zhu (Khodabakhshi, 2003) are Cauchy distributed and not Gaussian, the method of Zhu (Khodabakhshi, 2003) and Permchandra (2001) and those based on them are flawed. Shanmugam and Johnson (2007) Pursued applying PCA on inputs and outputs data separately and using these separately acquired variables as new inputs and outputs of DEA to overcome the flaw of trapping into a Cauchy distribution.

2.3. The proposed method

To improve the discrimination power of the DEA model, the approach of Shanmugam and Johnson (2007) is followed. However, the most published papers in PCA-DEA followed the method of Zhu (Khodabakhshi, 2003) whose defect was mentioned before (division of input to output variables and trapping into Cauchy distribution trapping). The possibility of negative PC values (for inputs and outputs) makes the exploit of thorough (not partial) translation invariance property necessary, and in this case a normalized additive DEA model (Knox Lovell and Jesfis, 1995) seems to be a good option.

Positivity restriction of DEA does not allow increasing inputs or decreasing outputs. So following Liang *et al.* (2009) undesirable variables are prepared. Suppose n similar DMUs which DMU_j, j=1,2,…,n consumes X'_J ; a column vector of k desirable and m − k undesirable inputs and produces Y'_j ; a column vector of t desirable and r − t undesirable outputs.

$$X' = [\, X'^{D} \quad X'^{U} \,] = [\, X'_1 \dots X'_n \,]_{m\times n} \qquad i = 1,2,...,m \tag{3}$$

$$Y' = [\, Y'^{D} \quad Y'^{U} \,] = [\, Y'_1 \dots Y'_n \,]_{r\times n} \qquad k = 1,2,...,r \tag{4}$$

In order to take DEA on the original data, several steps are carried out as follows:

Step 1: Examining if principle components analysis is an appropriate technique; examining Variable's Gaussian distribution by the normality test of variables and examining if there is "any statistical significant correlation between original variables" (by SPSS software).

For input data:

Step 2: Preparing PCA matrix by reversing undesirable inputs:

$$X = [\, X^D \quad X^U \,] = [\, X_1 \ldots X_m \,]_{n\times m} \qquad (5)$$

Step 3: Performing PCA on X.

Step 4: Confirm PCs; selecting $\{PC_l\}$ where $l=1\ldots P \leq m$ from scree plot or eigenvalue-greater-than-one rule.

Step 5: Performing translation invariance on PCs (new DEA inputs), if necessary, sake of insuring strictly positive input data. Input PCs data having been increased by the most negative value in the vector plus one (Adler and Yazhemsky, 2010).

$$X_l = PC_l + b > 0 \qquad l = 1 \ldots p \leq m \qquad (6)$$

$$b = -min\ PC_l + 1 \qquad (7)$$

In order to take DEA on the output data, the aforementioned steps are carried out on outputs too.

Step 6: Preparing PCA matrix by reversing undesirable outputs:

$$Y = [\, Y^D - Y^U \,] = [\, Y_1 \ldots Y_r \,]_{r\times n} \qquad (8)$$

Step 7: Performing PCA on Y.

Step 8: Confirming PCs; $\{PC_l\}$ where $l = 1\ldots P \leq r$ (scree plot or Eigen value greater than one rule).

Step 9: Performing translation invariance on PCs (new DEA outputs), if necessary, sake of insuring strictly positive output data.

$$Y_l = PC_l + b > 0 \qquad l = 1 \ldots p \leq r \qquad (9)$$

$$b = -min\ PC_l + 1 \qquad (10)$$

Step 10: Performing Model (1) (normalized additive DEA model) using translated input-oriented PCs as inputs of the model and translated output-oriented PCs as outputs.

3. Case study

In the early development literature, crash rate and crash fatality was traditionally used for measuring road safety. Although the extent use of only the crash rate and crash fatality, in the recent years safety ranking of roads, it is fully recognized that the multidimensionality of road safety problem could not be achieved substantially (Hermans and Van den Bossche, 2008). In order to consider different aspects of road safety, European safety net proposed the use of performance indicators which comprehensively consider different aspects of road safety (Hermans *et al.,* 2009). Alhaji (2005) proposed composite road safety development index calculated on the basis of three most important risk areas; behavioral, infrastructural and vehicle aspect.

There are several methods to evaluate road safety performance and one may classify them into two main groups: socioeconomic road safety methods consist of Multi Criteria Analysis (MCA) and Cost Benefit Analysis (CBA) and structural road safety methods consist of statistical analysis and data mining, but only those methods accepting several road safety indices, e.g. DEA and data mining, could comprehensively evaluate it. DEA is usually performed to compare similar decision making units' efficiency, in this case the road safety of Iran's state, through the use of weighted averages and to improve the efficiency of those units that are not efficient. When assessing the performance of roads, DEA combines performance in terms of several desirable and undesirable attributes into a single measure, the efficiency score. Desirable and undesirable indices are explained in Table 2. The advantages of applying DEA in road safety evaluation have been discussed in (Hermans and Van den Bossche, 2008). Odeck (2008; 2006) used a DEA based Malmquist index to evaluate progress in target achievement in road safety context. DEA was applied to evaluate safety level of 21 European countries (Hermans *et al.*, 2009). In this section, the performance of 30 Iran's states was analyzed using DEA. Analyzing road safety of different region enables the government to make just policies to improve the level of road safety at a country. Several human and structural attributes and those related to car performance are considered. The choice of attributes is influenced by data availability. Most of the data are collected from TRANSPORTTAION AND ROAD SAFETY ORGANIZATION- 2006 (1386 in solar) annals, for more consistency. Table 3 displays output and input data.

Variable's Gaussian distribution is one of the elements of performing PCA (Shanmugam and Johnson, 2007). The normality test of some variables by SPSS software brings in Figures 1 and 2. For input data, PCA offers the following results. The first three eigenvalues are greater than one, capturing 64% of the variation in the input variables: 34% of variance is covered by PC_1, 16% of remained variance is covered by PC_2 and 14% by PC_3.The principal component scores can be computed using:

*PC_1 = 0.528 * fleet age percent + 0.641* way light percent + 0.749 * highway percent + 0.016 * removed black spot no + 0.678 * police station no + 0.615 * road red arc base no + 0.765 * trespass percent + 0.157 * public instruction.*

*PC_2 = -0.115 * fleet age percent - 0.4* way light percent – 0.12 * highway percent + 0.881 * removed black spot no + 0.409 * police station no – 0.125 * road red arc base no + 0.156 * trespass percent – 0.31 * public instruction.*

Table 2: Desirable and undesirable indices explanation.

Desirable /Undesirable	Definition	indices
Desirable	Average proportion of less than 6 years fleet age in each province by the end of 2006.	Fleet age
Desirable	The proportion of km of province's roads having standard light to the total Km of each province's roads by the end of 2006.	Way light
Desirable	The proportion of km of province's highway to the total Km of each province's roads by the end of 2006.	highway
Desirable	The proportion of the number of removed black spot points to the total recognized black spots of each province by the end of 2006.	Remove blacksp
Desirable	The number of active road police station in each province by the end of 2006.	police station
Desirable	The number of road red arc in each province by the end of 2006.	road red
Undesirable	The percent of detected road trespass in each province in 2006.	Trespass
Desirable	The percent of public road safety instruction (banners, animations…) by the end of 2006.	public instr
Desirable	The percent of driver road safety instruction by the end of 2006.	Driver ins
Undesirable	The number of each province's road crashes in 2006.	crash
Undesirable	The number of each province's road casualties in 2006.	casualty

Table 3: Inputs and outputs.

2006	Inputs									outputs	
INDICES PROVINCE	fleet age	Way light	High way	Remove Blacksp	police station	road red	trespass	public instr	Driver ins	crash	casualty
AZARBAIJAN SHARQI	0.62	0.9	3.8	1	10	3	2.3	0.024	135	8948	633
AZARBAIJAN GHARBI	0	1.7	4.14	0.7	7	6	7.7	0.187	150	4940	535
ARDEBIL	1.71	10.1	4.8	0.18	5	10	0.41	0.081	75	1755	227
ESFAHAN	0.44	4.4	43.6	0.18	13	11	6.34	0.032	150	11963	926
ILAM	9.53	0.16	0.83	1	4	3	0.9	0.008	95	1219	140
BUSHEHR	2.05	1.22	11.5	0.43	5	3	2.6	0.024	120	1499	307
TEHRAN	2.26	17.6	21.6	1.33	14	13	8.2	0.024	400	2806	1071
CHARMAHAL	1.74	3.5	2.12	0.4	4	2	0.91	0.024	81	1602	260
KHORASAN JONUBI	0.89	1.81	0	0.1	4	8	0.47	0.040	94	2432	272
KHORASAN RAZAVI	0.73	2.38	4.4	0.07	5	7	0.37	0.073	826	1016	160
KHORASAN SHOMALI	1.25	2.9	3	0.74	3	11	6.84	0.024	60	5752	1209
KHUZESTAN	4.28	4.25	14.6	0.1	10	4	5.92	0.016	142	7229	1047
ZANJAN	3.41	2.72	1.14	0.59	6	2	1.81	0.008	0	2882	385
SEMNAN	2.36	6.32	16.04	0.21	6	5	1.33	0.016	189	3882	391
SISTAN	2.78	0.57	0.38	2.8	9	5	2.98	0.024	169	2386	769
FARS	1.26	2.1	6.34	0.28	13	4	3.04	0.016	290	8382	1328
QAZVIN	1.45	9.02	6.85	0.59	6	9	2.4	0.024	27	5031	606
QOM	1.83	14.6	21.9	0.28	4	4	1.8	0.024	86	4955	366
KORDESTAN	0	2.23	1.1	0.43	4	11	2.5	0	307	2509	462
KERMAN	0	2.3	9.9	1	9	7	2.4	0.040	165	4743	1115
KERMANSHAH	2.24	5.9	8.6	0.16	9	6	3.9	0.024	185	3731	387
KOHKILUIE	29.01	1.73	0.3	0.24	3	2	0.4	0.024	85	1641	171
GOLESTAN	0.73	11.6	11.8	0.36	4	5	5.73	0.081	191	4240	421
GILAN	0	8.4	13.4	1	8	10	6.26	0.008	285	9220	959
LORESTAN	1.36	3.43	4.65	0.28	6	13	2.25	0.016	83	5086	666
MAZANDARAN	0	12.3	10.9	0.064	10	10	5.03	0.040	0	1645	731
MARKAZI	1.17	3.04	9.2	1	8	9	2.12	0.016	0	4622	550
HORMOZGAN	1.35	0.87	3.4	0.69	5	8	1.16	0.008	120	2912	503
HAMEDAN	1.05	2.45	15.4	1	5	4	6.7	0.032	140	3627	605
YAZD	0.56	1.02	6.7	1	6	3	1.23	0.024	66	2270	268

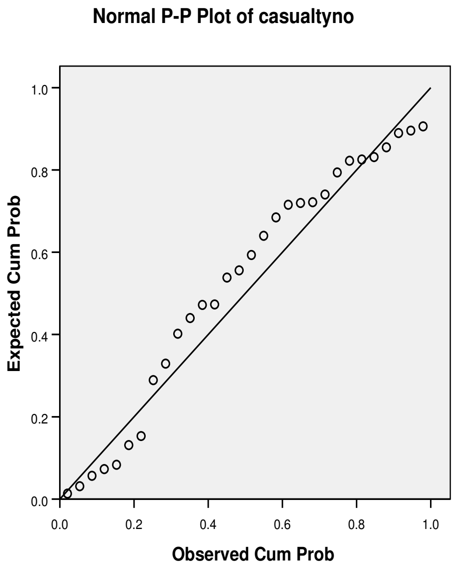

Figure 1: The normality tests of casualty number.

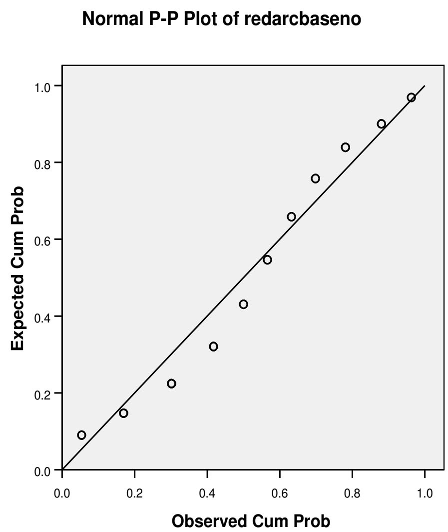

Figure 2: The normality test of road red arc base number.

*PC_3 = -0.359 * fleet age percent - 0.198* way light percent – 0.355 * highway percent + 0.167 * removed black spot no - 0.202 * police station no + 0.033 * road red arc base no + 0.243 * trespass percent + 0.843 * public instruction*

These results reveal the following points: (i) the input variables fleet age percent, Way light percent, high way percent, road red arc base no, police station number. and trespass percent are an important cluster of variables and they are identified in the first principal component; (ii) the input variable removed black spot number is also important and it is according to the second principal component; (iii) the input variable public instruction is also important and mentioned in the third principle component. The interpretation of input principle component can be as follows: (i) PC_1 is interpreted as road safety policies, services and infrastructures; (ii) PC_2 is interpreted as road reformation; (iii) PC_3 is interpreted as safety instruction.

For output variables, PCA results as follows: the first eigenvalue is greater than one, capturing 80% of variation in the output variables and the principle component score can be computed using: PC_4 = 0.892* suburban crash no + 0.892* suburban casualty number.

The output principle component is interpreted as road safety products.

Table 4: Additive DEA results using transformed PC1, PC2, PC3 and PC4.

	Input1	Input2	Input3	Output	S3 -	S2 -	S1 -	S+	λ	Z
AZARBAIJAN S	6.35	4.7	2.78	5.27	1.78	2.87	5.35	2.42	λ22=1	-7.33
AZARBAIJAN G	8.85	2.57	7.35	6.18	6.35	0.74	7.85	1.5	λ22=1	-10.03
ARDEBIL	7.41	1	3.6	7.52	0	0	0	0	λ3=1	0
ESFAHAN	13.57	2.86	1.3	4.00	0.3	1.03	12.57	3.66	λ22=1	-8.05
ILAM	3.28	3.86	2.2	7.84	0	0	0	0	λ5=1	0
BUSHEHR	5.87	2.98	2.55	7.35	1.55	1.15	4.87	0.34	λ22=1	-4.29
TEHRAN	14.58	4.03	1.77	1	0.76	2.2	13.58	6.69	λ22=1	-11.64
CHARMAHAL	4.49	2.66	2.73	7.46	1.73	0.83	3.49	0.23	λ22=1	-3.62
KHORASAN J	5.17	1.98	3.25	7.29	2.26	0.15	4.17	0.39	λ22=1	-3.9
KHORASAN R	5.79	1.7	3.75	7.82	0	0	0	0	λ10=1	0
KHORASAN S	7.81	3.17	3.6	4.26	2.6	1.34	6.81	3.43	λ22=1	-8.02
KHUZESTAN	8.62	3.02	1.9	4.45	0.87	1.19	7.6	3.24	λ22=1	-6.55
ZANJAN	4.81	3.48	2.32	6.92	1.32	1.65	3.8	.769	λ22=1	-4.35
SEMNAN	7.07	2.2	1.7	6.74	0.7	0.34	6.07	.949	λ22=1	-3.73
SISTAN	6.21	7.43	3.48	5.98	2.48	5.6	5.21	1.709	λ22=1	-9.61
FARS	7.68	3.88	2.06	3.52	1.06	2.05	6.68	4.169	λ22=1	-7.65
QAZVIN	7.88	2.53	2.46	5.98	1.46	0.69	6.88	1.709	λ22=1	-5.45
QOM	8.34	1.21	1.53	6.63	0	0	0	0	λ18=1	0
KORDESTAN	6.42	2.86	2.61	6.77	1.6	1.03	5.42	.914	λ22=1	-4.80
KERMAN	7.72	4.09	3.03	4.68	3.09	2.26	6.72	3.012	λ22=1	-8.89
KERMANSHAH	8.09	2.68	2.27	6.77	1.27	0.85	7.09	.913	λ22=1	-4.99
KOHKILUIE	1	1.83	1	7.69	0	0	0	0	λ22=1	0
GOLESTAN	8.87	1.47	3.92	6.60	1.79	0	5.089	1.014	λ22=0.57 λ 3=0.43	-4.10
GILAN	10.26	3.8	2.33	4.34	1.33	1.97	9.26	3.326	λ22=1	-8.25
LORESTAN	7.54	2.47	2.53	5.81	1.53	0.64	6.54	1.875	λ22=1	-5.45
MAZANDARAN	10.78	1.90	2.5	3.79	1.5	0.07	9.78	3.897	λ22=1	-7.45
MARKAZI	7.58	4.00	2.42	6.2	1.42	2.17	6.584	1.491	λ22=1	-6.34
HORMOZGAN	5.59	3.42	2.53	6.601	1.53	1.59	4.59	1.089	λ22=1	-4.99
HAMEDAN	7.99	3.91	3.20	6.21	2.2	2.08	6.99	1.476	λ22=1	-7.1
YAZD	5.39	4.05	2.83	7.33	1.83	2.22	4.39	0.36	λ22=1	-5.22

The province group with perfect rating score are; ARDEBIL, ILAM, KHORASAN- RAZAVI, QOM and KOHKILUIE with Z= 0. For safety improvement, GOLESTAN could emulate ARDEBIL and KOHKILUIE and other province could emulate KOHKILUIE. Table 4 displays model solution.

4. Conclusion

In conclusion, it is pointed out that before applying PCA, multivariate Gaussian distribution assumption of data variables (whether input or output) should be tested. Undesirable input variable; trespass percent and undesirable output variables; crash number and casualty number are considered by reversing (multiplying by minus). To avoid finding the ratio of input to the output data; PCA is performed on input data: fleet age percent, way light percent, highway percent, removed black spot number, police station number, road red arc number, reversed trespass percent, public instruction percent and driver instruction number and on reversed output data; crash number and crash fatality separately. Because of un-positivity possibility of input and output principle components, the researchers consider a DEA model possessing general translation invariance property (normalized additive DEA model). The advantages of the proposed method are:

(i) Avoidance of finding the ratio of input to output and falling into Cauchy distribution trap.

(ii) Considering undesirable input and output variables simultaneously in composite PCA-DEA method.

References

Adler, N.; Golany, B., (2001), Evaluation of deregulated airline networks using data envelopment analysis combined with principle component analysis with an application to Western Europe. *European Journal of Operational Research*, 132, 260-273.

Adler, N.; Yazhemsky, E., (2010), Improving discrimination in data envelopment analysis: PCA-DEA or variable reduction. *European Journal of Operational Research,* 202(1), 273-284.

Agha, I.; Lawrence, M. S., (1990), Translation Invariance in Data Envelopment Analysis. *Operations Research Letters*, 9(6), 403-405.

AlHaji, G., (2005), *Towards a road safety development index*. Ph.D. thesis, Linko pings University, Sweden.

FU Bo-xin, SONG Xiang-qun, GUO Zi-jian, FANG Zhuo, (2007), *Study on improved DEA-Based MPI and its application in dynamic performance evaluation*. Proceeding of 14th International Conference on Management Science & Engineering, Harbin, P. R. China.

Hadi, A.; Vencheh, R.; Kazemi, M.; Tavassoli Kajani, M., (2005), Undesirable factors in efficiency measurement. *Applied Mathematics and Computation*, 163, 547-552.

Hermans, E.; Brijs, T.,; Wets, G.; Vanhoof, K., (2009), Benchmarking road safety: Lessons to learn from a data envelopment analysis. *Accident Analysis and Prevention*, 41,174-182.

Hermans, E.; Van den Bossche, F., (2008), combining road safety infor-mation in a performance index. *Accident Analysis and Prevention*, 40, 1337-1344.

Jenkins, L.; Anderson, M., (2003), A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *European Journal of Operational Research*, 147, 51-61.

Jesfis T., (1996), Translation invariance in data envelopment analysis: A generalization. *Annals of Operations Research*, 66, 93-102.

Khodabakhshi, M., (2003), *New input and output orentation models in analysis and applications*. Ph. D. thesis.

Knox Lovell, C. A.; Jesfis, T., (1995), Units invariant and translation invariant DEA models. *Operations Research Letters*, 18, 147-151.

Lawrence, M. S.; Zhu, J.; Jesfis, T., (2002), Pastor, Modeling undesirable factors in efficiency evaluation. *European Journal of Operational Research*, 142, 16-20.

Liang, L.; Li, Y.; Li, S., (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.

Odeck, J., (2006), Identifying traffic safety best practice: An application of DEA and Malmquist. *Omega*, 34, 28-40.

Odeck, J., (2008), The effect of mergers on efficiency and productivity of public transport services. *Transportation Research*, 42, 696-708.

Permachandra, I. M., (2001), A note on DEA vs. principle component analysis: An improve-ment to Joe Zhu's approach. *European Journal of Operation Research,* 132, 553-560.

Robert, M., (1996), Thrall, The lack of invariance of optimal dual solutions under translation. *Annals of Operations Research*, 66, 103-108.

Shanmugam, R.; Johnson, Ch., (2007), At a crossroad of data envelopment and principal component analyses. *Omega*, 35, 351-364.

Zhu, J.; (1998), Data envelopment analysis vs. principal component analysis. *European Journal of Operational Research*, 111(1), 50-61.