

The Use of Yanai's Generalized Coefficient of Determination to Reduce the Number of Variables in DEA Models

Mauricio Benegas

School of Business, Accounting and Economics, Federal University of Ceará, Fortaleza, Brazil
Email: mauricio_benegas@caen.ufc.br

How to cite this paper: Benegas, M. (2017) The Use of Yanai's Generalized Coefficient of Determination to Reduce the Number of Variables in DEA Models. *American Journal of Operations Research*, 7, 187-200.
<https://doi.org/10.4236/ajor.2017.73013>

Received: February 22, 2017

Accepted: May 15, 2017

Published: May 18, 2017

Copyright © 2017 by author and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper proposes a new method to reduce the dimensionality of input and output spaces in DEA models. The method is based on Yanai's Generalized Coefficient of Determination and on the concept of pseudo-rank of a matrix. In addition, the paper suggests a rule to determine the cardinality of the subset of selected variables in a way to gain the maximal discretionary power and to suffer a minimal informational loss.

Keywords

DEA, Yanai's Generalized Coefficient of Determination, Pseudo-Rank, Dimension Reduction, Improving Discrimination

1. Introduction

The DEA (Data Envelopment Analysis) model is a nonparametric method for estimating production frontiers. The DEA involves the solution of a set of linear programming (LP) problems to determine a production frontier against which the technical efficiency of the Decision Making Units (DMU) will be calculated. The basic DEA model was originally proposed by [1], and that model is nowadays known as "the CCR model".

The basic CCR model proposal aims to maximize the ratio between a weighted sum of outputs and a weighted sum of inputs. The weights of these sums are chosen according to the feasibility conditions and assuming a hypothesis of constant returns to scale. Charnes, Cooper and Rhodes have previously transformed the fractional CCR model into a linear model whose dual is commonly referred to in the literature as the DEA (the details of this procedure can be found in [2] ch. 1 and 2) model.

The CCR model and its variants have been increasingly applied following the

original description by [1]. Since then, researchers worldwide have used the model as a tool to assess technical efficiency. The DEA model has become the method widely chosen for this type of study, especially in the absence of an explicit production function to define the relationship between inputs and outputs. It is worth to say that if there is evidence that the inputs and outputs can be linked through a function, then the stochastic frontier model can be used as an alternative to the DEA. Please refer to the work of [3] on models of stochastic frontier production models.

One of the most frequent problems associated with the CCR model is the lack of discrimination among DMUs when the number of inputs and outputs is rather large in relation to the number of DMUs. A large number of variables relative to the number of observations may entail a large number of efficient DMUs in the sample, thus reducing the model's ranking capability. This represents a characteristic of the CCR model: the lower the number of DMUs, the less active the restrictions imposed on the maximum efficiency multipliers. In fact, according to [4] DEA models are subject to the curse of dimensionality.

Several alternatives have been proposed to increase the ranking capacity of the CCR model (see, eg, [5] [6] [7] [8]), including the super-efficiency (e.g., [9]) and cross-efficiency evaluation methods ([10] [11] [12]). Other methods that use additional information (usually characterized by adding restrictions) include the cone-ratio and assurance-region approaches¹.

In the present work, we propose a simple and objective method to address a situation in which the original inputs and outputs have been correctly selected but the low number of observations has translated into low discrimination power. This approach is intended to reduce the dimensions of the inputs and outputs spaces and therefore does not require additional information. Additionally, this method does not require post-estimation procedures (as in the case of the super-efficiency and cross-efficiency approaches). Our proposed approach relies on multivariate statistical techniques, namely, Principle Components Analysis and the use of a correlation matrix. Again, it should be emphasized that this approach is not intended for the selection of inputs or outputs. Instead, it is a support tool used to increase the discriminatory power without a significant loss of information, *i.e.*, the discriminatory power is increased regardless of knowledge of which variables are essential for the model. The method is a tour de force of the linear algebra when applied to find similarity between subspaces by exploring the special structure of the space of positive semidefinite matrices.

Following this introduction, the remainder of the work is structured as follows: Section 2 briefly reviews multivariate variable selection in the DEA models, Section 3 introduces the proposed methodology, Section 4 discusses applications of the proposed method to the CCR model, and finally, Section 5 presents conclusions and suggestions for future research.

2. Selection/Summarization of Variables in the DEA Model

One of the issues relating to the CCR model is that of the dimensions of inputs

¹See [13] for a survey.

and outputs. A consequence of using a large number of variables relative to the number of observations is the loss of discrimination power due to the generation of a large number of efficient DMU's. There is no consensus on the optimal number of inputs and outputs to be used. However, [2] suggest the following rule: $J:J > \max\{M \times N, 3(M + N)\}$, where M and N correspond to the number of outputs and inputs respectively.

One of the most common ways of selecting variables in DEA is the use of correlation matrices for inputs and outputs. When two variables (inputs or outputs) are highly correlated, one is discarded, usually on the basis of ad hoc criteria. However, eliminating one or the other variable could have a dramatic impact on estimated efficiency².

In recent years, the application of multivariate statistical methods, especially Principal Components Analysis (PCA), has appeared as a satisfactory alternative for variable reduction³. The formulation of a DEA model in which inputs and outputs are summarized as principal components is the focus of the work of [14] and [15]. One limitation of using PCs instead of inputs and outputs is that these "new" variables may take a negative value, and therefore must be transformed for PCA. That transformation, however, may impact results. One way of overcoming the problem is to use the additive model of DEA (see [16]) that is invariant to translation of inputs and outputs. In [17] the author has also shown that an input-oriented BCC model (after [18]) is invariant to output translation.

There is also a practical difficulty. Even if the problem of negative PCs is resolved, the question remains of how to interpret the results in terms of the projection of inputs and outputs (that is, predicting quantities). The fact is that the only satisfactory way of accomplishing this is back transformation to original variables-which might require a considerable computational effort. Some authors select variables based upon their contribution to PCs. Specifically, the variables with the largest absolute linear combination coefficient are selected. Because it is common for the first few components to explain most data variance, the result is a considerably reduced subset of variables.

In [19] the authors employ the method of partial covariance analysis to identify the correlation between variables as well as the contribution of each variable to these correlations. With this technique, the authors demonstrate that the removal of variables with little contribution to the correlations does not significantly change the results.

The method we propose is based on the work of [20] and [21] and combines PCA with elements of the [19] method. It consists of generating a correlation matrix between two data sets-the orthogonal projection of data onto the subspace generated by a subset of PCs and the orthogonal projection of data onto the subspace generated by a subset of original variables. This matrix measure of the closeness of the two subspaces is known as Yanai's Generalized Coefficient of

²In [23] the authors presents an example that illustrate that point very well.

³In fact the use of PCA for summarizes a data set is broadly applied in several areas range from image processing ([24]) to meteorology ([25]).

Determination (GCD) (proposed in [22]). The resulting subset of PCs (which usually includes PCs that explain eighty to ninety percent of data variance) corresponds to the original variables that maximize the GCD.

The number of PCs is determined prior to the generation of the correlation matrix, estimating the pseudo-rank for the matrix, as proposed by [20] based on the specific geometric structure of the cone of positive semidefinite matrices.

In the following section, a brief discussion of the geometrical structure of the cone of positive semi-definite matrices and of PC analysis is presented and the GCD is subsequently defined.

3. The Pseudo-Rank of a Matrix and Yanai's Generalized Coefficient of Determination

This section briefly describes basic concepts described in detail in the work of [20] [21] [26].

Assuming C_p to be the cone of positive semidefinite matrices with dimension $p \times p$ provided with the Frobenius inner product $\langle \cdot, \cdot \rangle_F : C_p \times C_p \rightarrow \mathfrak{R}$ such that

$$\langle A, B \rangle_F = \text{tr}(A'B) \text{ for any } A, B \in C_p \quad (1)$$

The norm induced by (1) will be denoted by $\| \cdot \|_F$. For any matrix $V \in C_p$, the set

$$\text{Ray}(V) = \{ A \in C_p ; A = \lambda V, \lambda \geq 0 \}$$

is called the ray associated with V . The ray associated with the identity matrix of dimension p is called the central ray of C_p .

With the definitions above, it is possible to find the angle between the rays associated with any two matrices A and B . This angle is given by the arc whose cosine is

$$\cos(A, B) = \frac{\langle A, B \rangle_F}{\|A\|_F \|B\|_F} \quad (2)$$

In [19], the author demonstrates that C_p has a layered structure, with containing several cones fitted inside the other, just like showed in **Figure 1**.

Based on this observation, the author argues that the region close to the central ray contains only matrices of full rank, or at least with rank $p-1$ (which can occur at the boundary of such regions). The farther away from the central ray, the lower the rank of matrices.

However, matrices of full rank are also found outside of the core. Because they have may have eigenvalues close to zero, they behave as low-rank matrices. The question then is how far (into the core of the cone) does one have to move to avoid such matrices? The answer to this question lies in the concept of the pseudo-rank of a matrix. According to [20], the pseudo-rank of a $V \in C_p$ matrix is the smallest integer $k^* \leq p$ such that

$$\cos(V, I_p) \leq \sqrt{\frac{k^*}{p}} \quad (3)$$

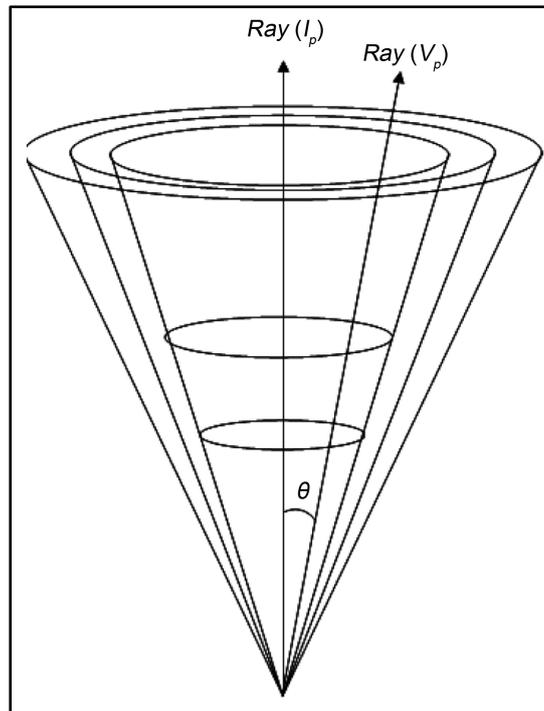


Figure 1. Cone of positive semidefinite matrices.

The author shows that this value is given by

$$k^* = \left\lceil \frac{\text{tr}(V)^2}{\text{tr}(V^2)} \right\rceil \quad (4)$$

where $\lceil z \rceil$ is the nearest to z greater integer. Letting V the covariance/correlation matrix of a data matrix A , with p variables and n observations, or $V = n^{-1}A'A$. In this case the pseudo-rank of V corresponds to the number of components to be used as representative of all accumulated variance associated with A .

Yanai's Generalized Coefficient of Determination

Let A represent a data matrix with dimension $(n \times p)$, where p indicates the number of variables and n the number of observations for each variable. In this context, A may refer to a matrix of discretionary/nondiscretionary outputs/inputs. It is important to keep in mind that in DEA models, the number of observations indicates the number of columns in the pertinent matrices. Thus, for the implementation of the proposed method, matrix A should be considered the transposed output/input matrix.

Given the covariance matrix/correlation of data $S = n^{-1}A'A$, let Λ and P be the diagonal matrix of eigenvalues (arranged in decreasing order) and the matrix of normalized eigenvectors of S respectively. The PCs are the columns of the matrix $(n \times p)$ given by $C = AP$. Using the spectral decomposition of S , it is easy to show that the covariance matrix of C is exactly Λ , so that the variables in C are uncorrelated. For this reason the AP transformation is sometimes called data "decorrelation".

Consider K to be a subset of indices associated with $k \leq p$ PCs arranged in decreasing order of eigenvalues (in general the first k 's). Similarly, let Q be defined as a subset of indexes associated with the $q \leq p$ original variables. The sets K and Q are the subspaces generated by vectors with indices K and Q respectively. The following matrices are then defined:

- A_K is the submatrix of A in which columns with indexes in K are maintained;
- $S_K = n^{-1}A'_K A_K$ is the covariance matrix associated with A_K ;
- Λ_K is the matrix of the eigenvalues associated with S_K ;
- P_K is the matrix of eigenvectors associated with eigenvalues in Λ_K ;

Let us assume P_K to be the matrix of orthogonal projection on the subspace K such that

$$P_K = n^{-1}AS_K^{-1}A'$$

where S_K^{-1} is the Moore-Penrose generalized inverse of S_K . Similarly, P_Q is the matrix of orthogonal projection on the subspace Q defined as

$$P_Q = n^{-1}AI_Q S_Q^{-1}I_Q A'$$

where I_Q is the identity matrix of the submatrix obtained by selecting the q columns with indices in Q and $S_Q = n^{-1}I_Q A' I_Q$.

Given the definitions above, Yanai's GCD between subspaces Q and K is defined as:

$$GCD(Q, K) = \frac{\langle P_Q, P_K \rangle_F}{\|P_Q\|_F \|P_K\|_F} \tag{5}$$

Supposing that $K = \{1, 2, \dots, k^*\}$ remains fixed, where k^* is the pseudo-rank of the covariance matrix, in this case the selection of variables exhibiting the greatest contribution to the principal components selected in K is the set of indices \tilde{Q} such that

$$\tilde{Q} = \arg \max_Q GCD(Q, K)$$

4. Reduced CCR Model⁴

A practical example of the proposed method is provided in this section. For that, the CCR model will be presented with the original variables (hereafter called the "general model" and denoted by CCRg), and with the subset of selected variables ("reduced model," denoted by CCRr). For the sake of simplicity, only product-oriented models will be discussed below.

Let us suppose that there are J DMUs under study, each using a vector $\mathbf{x} \in \mathfrak{R}_+^N$ of inputs to produce a vector $\mathbf{y} \in \mathfrak{R}_+^M$ of outputs with a technology defined by

$$T_{CCR_g} = \{(\mathbf{x}, \mathbf{y}); \mathbf{x} \geq \mathbf{X}\lambda, \mathbf{y} \leq \mathbf{Y}\lambda, \lambda \geq 0\}$$

where $\mathbf{X}_{(N \times J)}$, $\mathbf{Y}_{(M \times J)}$ and $\lambda_{(J \times 1)}$ are input and output matrices and the vector

⁴The use of the CCR model is an example. The method can actually be applied to any DEA model.

of intensities respectively.

Given a DMU j , its technical efficiency in the model CCRg, denoted by ET_j^g , is estimated by solving the following linear programming problem (the minimization of slacks is omitted for simplicity)

$$ET_j^g = \begin{cases} \max_{\theta, \lambda} \theta \\ \text{subject to} \\ (\mathbf{x}_j, \theta \mathbf{y}_j) \in T_{CCR_g} \end{cases} \quad (6)$$

Let us now suppose that a set of variables was selected following the procedure presented in Section 3. For simplicity's sake, let us assume that only outputs were selected. Let us denote by \mathbf{y}_j^q the vector product of a DMU $j = 1, 2, \dots, J$, with selection of $q < N$ outputs, and let us denote by \mathbf{Y}^q the respective reduced output matrix. The technology in the CCRr model is defined as

$$T_{CCR_g} = \{(\mathbf{x}, \mathbf{y}^q); \mathbf{x} \geq \mathbf{X}\lambda, \mathbf{y}^q \leq \mathbf{Y}^q\lambda, \lambda \geq 0\}.$$

Given a DMU j , its technical efficiency in the model CCRr, denoted by ET_j^r , is estimated by solving the following linear programming problem

$$ET_j^r = \begin{cases} \max_{\theta, \lambda} \theta \\ \text{subject to} \\ (\mathbf{x}_j, \theta \mathbf{y}_j^q) \in T_{CCR_r} \end{cases} \quad (7)$$

The reduced model is obtained in three stages, as summarized in the **Figure 2** below:

One issue not covered by the procedure presented above is that of defining the cardinality of the subset of selected variables. One suggestion would be to combine the gain in discrimination with some measure that would reveal loss of information. The gain in discrimination would be obtained by the difference between the percentage of efficient DMUs in the general model and in the reduced model. The complexity of this issue relates to what measure of informational loss

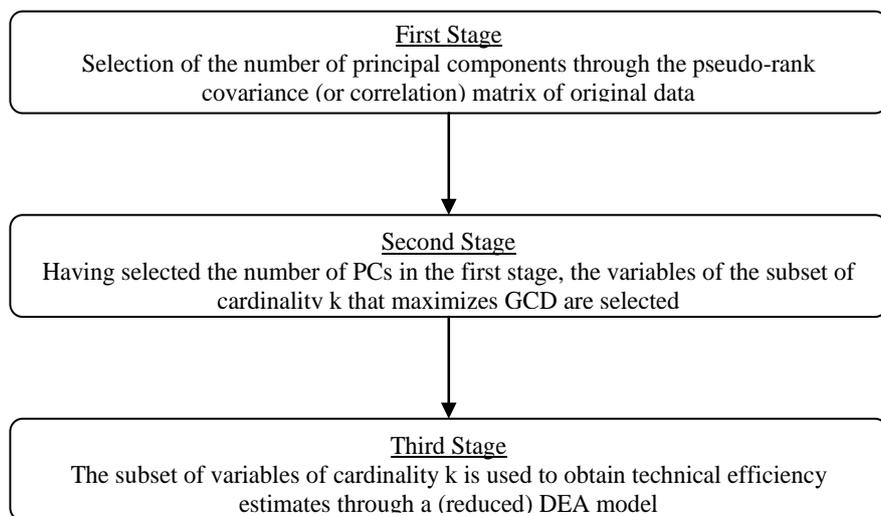


Figure 2. Steps for obtaining the reduced model.

should be used. One possibility is to use the Kolmogorov-Smirnov statistic, which quantifies the difference between distributions. In this context, this statistic is given by

$$KS(q) = \sup_x |F(x) - F_q(x)|$$

where F and F_q are the empirical cumulative distribution functions of technical efficiency estimated by the general and reduced models (the latter with a subset q of selected variables) respectively.

Let K^* and K_q^* be the number of efficient DMUs in the general and reduced models respectively; let $\delta_q = (K^* - K_q^*)/K$ such that δ_q represents, in proportional terms, the gain in discrimination power of the reduced model in relation to the general model. Then the optimal cardinality would be given by q^* such that

$$q^* \in \arg \max_q \left\{ \frac{\delta_q}{KS(q)}; KS(q) \leq 1.36\sqrt{\frac{2}{K}} \right\} \quad (8)$$

in which $KS(q) \leq 1.36\sqrt{2/K}$ is included so that optimal cardinality will depend on acceptance of the null hypothesis of the Kolmogorov-Smirnov test. The amount $1.36\sqrt{2/K}$ represents the nullity condition of the Kolmogorov-Smirnov test with a significance level of 0.05, such that if $KS(q) > 1.36\sqrt{2/K}$ the null hypothesis of equality between the distributions is rejected. It should be noted that $1.36\sqrt{2/K}$ is generally valid for $K \geq 8$, otherwise it is necessary to consult tabulated values.⁵

If there are multiple solutions to (8) the lowest maximizer q^* is selected, so that

$$q^* = \min_{\hat{q}} \left\{ \hat{q} \in \arg \max_q \left\{ \frac{\delta_q}{KS(q)}; KS(q) \leq 1.36\sqrt{\frac{2}{K}} \right\} \right\} \quad (9)$$

An Application to Real Data

This example employs real-world data previously described by [28], who examined the technical efficiency of the public health care system in Brazil. The database uses one input (x), represented by the annual per capita expenditure on health of the three levels of government, and 12 outputs (y_p , $i=1, \dots, 12$), representing health indicators available in the Ministry of Health's Information System DATASUS.⁶ For the present example, to reduce the power of discrimination, only 12 of the 27 states studied by [28] were selected. The data and some of the descriptive statistics are shown in **Table 1** (Appendix **Table A1** provides a description of the variables used in the example).

In this example, the Kolmogorov-Smirnov null hypothesis is accepted with a significance level of 0.05 for a value up to 0.5552, such that the condition to select the cardinality of the subset of selected outputs becomes

⁵See [27] for details.

⁶See site www2.datasus.gov.br.

Table 1. Data of example.

State	x	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}	Y_{11}	Y_{12}
RO	387.5	68.2	73.8	70.9	80	0.8	1.7	111.2	106.9	107.5	108.3	47.6	70
AC	556.6	68.6	73.8	71.1	71	0.8	2	86.5	83.64	110.5	90.14	40.3	67.4
AM	513.4	68.4	74.4	71.3	78	0.9	1.5	106.3	93.88	131.1	92.36	57.1	72.8
RR	674.5	67.2	72.1	69.6	83	1.1	1.6	93.13	88.62	114.1	89.85	71.9	79.3
PA	263.8	68.8	74.7	71.7	76	0.8	1.6	116.6	112.3	144.5	119.8	55	76.9
AP	512.5	66.2	74.1	70.1	79	0.8	1.5	94.12	96.3	126.7	98.17	28.2	90.5
TO	503.8	68.8	73.3	71	78	1.1	1.8	99.57	107.2	110.1	107.4	21	69.9
MA	285.3	63.4	71.3	67.2	69	0.6	2.4	108.2	106.6	135.6	111.9	50.1	59
PI	315.8	65.6	71.7	68.6	73	0.8	2.5	101.9	102.1	106	104.6	61.7	49.6
CE	291.4	65.7	74.4	69.9	74	0.9	1.9	108.3	108.3	113.3	112.3	40.8	71.3
RN	405	66.3	74.1	70.1	69	1.2	2.3	98.87	95.45	108.2	94.61	45.1	82.3
PB	335.1	65.2	72.2	68.6	68	1.1	2.7	105.7	103.6	117.8	104	48.5	74.7
Max	674.5	68.8	74.7	71.7	83	1.2	2.7	116.6	112.3	144.5	119.8	71.9	90.5
Min	263.8	63.4	71.3	67.2	68	0.6	1.5	86.5	83.64	106	89.85	21	49.6
Mean	420.4	66.9	73.3	70	75	0.9	2	102.5	100.4	118.8	102.8	47.3	72
SE	130.1	1.73	1.18	1.33	4.8	0.2	0.4	8.529	8.772	12.64	9.732	13.9	10.6

Source: [28].

$$q^* = \min_{\hat{q}} \left\{ \hat{q} \in \arg \max_q \left\{ \frac{\delta_q}{KS(q)}; KS(q) \leq 0.5552 \right\} \right\}$$

Table 2 shows the results obtained by applying the proposed variable selection method to the output matrix. Variables were selected into subsets with cardinality from 1 to 10. Using the pseudo-rank of the output covariance matrix (Equation (4)), the first four PCs were used. These for PCs accounted for approximately 84% of the total variability of the sample.

Note that in this example, $\arg \max \left\{ \delta_q / KS(q); KS(q) \leq 0.5552 \right\} = \{4, 5, 6, 7, 8, 9, 10\}$ and therefore $q^* = 4$. The last part of the example appears in **Figure 2**, which shows the estimated densities for the general and reduced models with cardinality from 1 to 8. The procedure to estimate the densities uses the Gaussian kernel. The bandwidth was selected minimizing the mean integrated square error (see [29] for details).

The results shown in **Figure 3** suggest that in the presence of four selected variables, the inclusion of an additional variable does not have a significant impact on the comparison between the general model and the reduced model. This observation confirms the conclusions obtained by applying the method of subset cardinality selection suggested by equation (9).

5. Conclusions and Further Research

This paper proposes a method for reducing the dimension of input/output matrices used to estimate production frontiers through the CCR model (or its va-

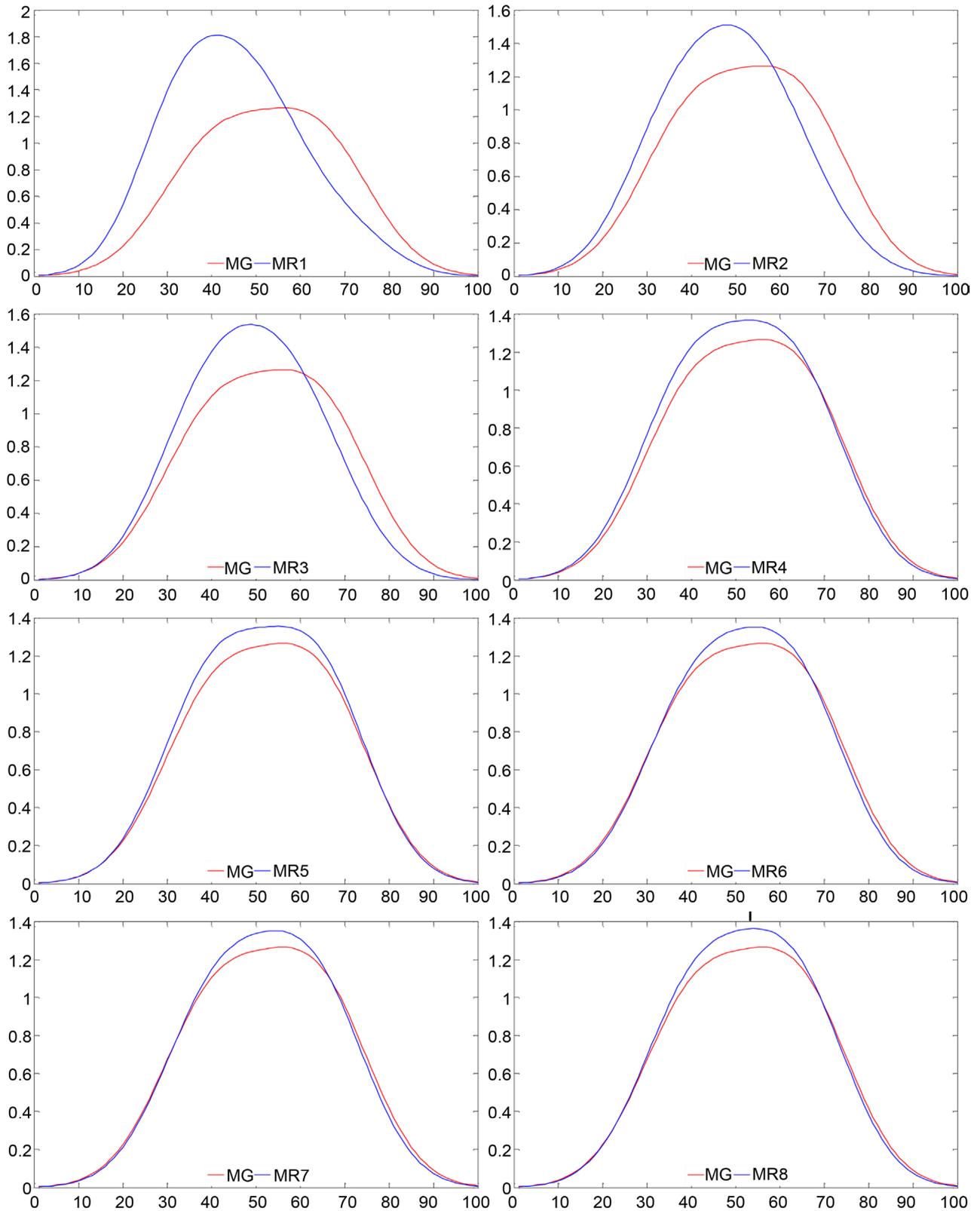


Figure 3. Estimated densities for TE in general and reduced models.

riants). The method is based on Yanai’s Generalized Coefficient of Determination (GCD) and on the concept of pseudo-rank of a matrix. Additionally, a rule

Table 2. Results for general and reduced models^a.

State	GM	RM1	RM2	RM3	RM4	RM5	RM6	RM7	RM8	RM9	RM10
RO	0.71	0.62	0.65	0.67	0.69	0.69	0.71	0.71	0.71	0.71	0.71
AC	0.51	0.42	0.42	0.47	0.49	0.49	0.49	0.49	0.49	0.49	0.49
AM	0.59	0.49	0.49	0.51	0.57	0.59	0.59	0.59	0.59	0.59	0.59
RR	0.53	0.40	0.40	0.40	0.48	0.53	0.53	0.53	0.53	0.53	0.53
PA	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AP	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61	0.61
TO	0.64	0.48	0.48	0.52	0.63	0.63	0.64	0.64	0.64	0.64	0.64
MA	1.00	0.71	0.86	0.87	0.87	0.87	0.88	0.88	0.87	0.87	0.87
PI	1.00	0.54	0.73	0.80	0.85	0.94	0.94	0.94	0.94	0.94	0.94
CE	1.00	0.84	0.84	0.88	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RN	0.88	0.70	0.70	0.70	0.88	0.88	0.88	0.88	0.88	0.88	0.88
PB	1.00	0.76	0.76	0.76	1.00	1.00	1.00	1.00	1.00	1.00	1.00
δq	-	0.33	0.33	0.33	0.17	0.17	0.17	0.17	0.17	0.17	0.17
$KS(q)$	-	0.42	0.42	0.42	0.17	0.17	0.17	0.17	0.17	0.17	0.17
$\delta q/KS(q)$	-	0.80	0.80	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Source: Author's estimates. ^aThe initials GM and RM_q refer to general and reduced models with cardinality q , respectively.

is suggested to support the choice of cardinality of the subset of selected variables. This rule seeks to combine maximum gain in discriminatory power with minimal loss of information.

Through an example that employs real-world data, it was found that the pseudo-rank of the output correlation matrix indicates that the first four PCs should be maintained. The GCD for output subsets with cardinality from 1 to 10 was then calculated. The cardinality rule indicated the subset with four of the 12 original outputs. Finally, the estimation of densities of the general and reduced models suggested that the cardinality decision rule can support decisions concerning the number of variables required in the model to obtain maximum discrimination with minimal loss of information.

Further research is suggested on the cardinality decision rule taking into consideration various measures of loss of information. Also warranted are studies comparing the proposed method with other methods of summarization and selection.

References

- [1] Charnes, A., Cooper, W.W. and Rhodes, E. (1978) Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*, **2**, 429-444.
- [2] Cooper, W.W., Seiford, L.M. and Tone, K. (2006) *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, Springer, New York.
- [3] Kumbhakar, S.C. and Lovell, C.A.K. (2000) *Stochastic Frontier Analysis*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/cbo9781139174411>

- [4] Zimek, A., Schubert, E. and Kriegel, H.-P. (2012) A Survey on Unsupervised Outlier Detection in High-Dimensional Numerical Data. *Statistical Analysis and Data Mining*, **5**, 363-387. <https://doi.org/10.1002/sam.11161>
- [5] Adler, N. and Golany, B. (2002) Including Principal Component Weights to Improve Discrimination in Data Envelopment Analysis. *Journal of the Operational Research Society*, **53**, 985-991. <https://doi.org/10.1057/palgrave.jors.2601400>
- [6] Angulo-Meza, L. and Lins, M.P.E. (2002) Review of Methods for Increasing Discrimination in Data Envelopment Analysis. *Annals of Operations Research*, **116**, 225-242. <https://doi.org/10.1023/A:1021340616758>
- [7] Podinovski, V.V. and Thanassoulis, E. (2007) Improving Discrimination in Data Envelopment Analysis: Some Practical Suggestions. *Journal of Productivity Analysis*, **28**, 117-126. <https://doi.org/10.1007/s11123-007-0042-x>
- [8] Senra, L.F.A.C., Naci, L.C., Soares de Melo, J.C.B. and Angulo-Meza, L. (2007) Estudo Sobre Métodos de Seleção de Variáveis em DEA. *Pesquisa Operacional*, **27**, 191-207. <https://doi.org/10.1590/s0101-74382007000200001>
- [9] Andersen, P. and Petersen, N.C. (1993) A Procedure for Ranking Efficient Units in Data Envelopment Analysis. *Management Science*, **39**, 1261-1264. <https://doi.org/10.1287/mnsc.39.10.1261>
- [10] 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, 73-105. <https://doi.org/10.1002/ev.1441>
- [11] Doyle, J.R. and Green, R. (1994) Efficiency and Cross-Efficiency in Data Envelopment Analysis: Derivatives, Meanings and Uses. *Journal of the Operational Research Society*, **45**, 567-578. <https://doi.org/10.1057/jors.1994.84>
- [12] Green, R.H., Doyle, J.R. and Cook, W.D. (1996) Preference Voting and Project Ranking Using Data Envelopment Analysis and Cross-Evaluation. *European Journal of Operational Research*, **90**, 461-472.
- [13] Athanassopoulos, A.D. (2012) Discriminating among Relatively Efficient Units in Data Envelopment Analysis: A Comparison of Alternative Methods and Some Extensions. *American Journal of Operations Research*, **2**, 1-9. <https://doi.org/10.4236/ajor.2012.21001>
- [14] Ueda, T. and Hoshiai, Y. (1997) Application of Principal Component Analysis for Parsimonious Summarization of DEA Inputs and/or Outputs. *Journal of Operational Research Society of Japan*, **40**, 466-478.
- [15] 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*, **132**, 260-273.
- [16] Ali, A.I. and Seiford, L.M. (1990) Translation Invariance in Data Envelopment Analysis. *Operations Research Letters*, **9**, 403-405.
- [17] Pastor, J. (1996) Translation Invariance in Data Envelopment Analysis: A Generalization. *Annals of Operations Research*, **66**, 91-102. <https://doi.org/10.1007/BF02187295>
- [18] Banker, R.D., Charnes, A. and Cooper, W.W. (1984) Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, **30**, 1078-1092. <https://doi.org/10.1287/mnsc.30.9.1078>
- [19] Jenkins, L. and 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.

- [20] Cadima, J.F.L. (2001) Redução de Dimensionalidade Através duma Análise em Componentes Principais: um critério para o número de Componentes Principais a reter. *Revista de Estatística (INE)*, 1o. quadrimestre, 37-49.
- [21] Cadima, J.F.L. and Jolliffe, I.T. (2001) Variable Selection and the Interpretation of Principal Subspaces. *Journal of Agricultural, Biological and Environmental Statistics*, **6**, 62-79. <https://doi.org/10.1198/108571101300325256>
- [22] Yanai, H. (1974) Unification of Various Techniques of Multivariate Analysis by Means of Generalized Coefficient of Determination (G.C.D.). *Journal of Behavior-metrics*, **1**, 45-54. <https://doi.org/10.2333/jbhmk.1.46>
- [23] 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*, **132**, 245-259.
- [24] Karamizadeh, S., Abdullah, S.M., Manaf, A.A., Zamani, M. and Hooman, A. (2013) An Overview of Principal Component Analysis. *Journal of Signal and Information Processing*, **4**, 173-175. <https://doi.org/10.4236/jsip.2013.43B031>
- [25] Richman, M.B, Mercer, A.E., Leslie, L.M., Doswell III, C.A. and Shafer, C.M. (2013) High Dimensional Dataset Compression Using Principal Components. *Open Journal of Statistics*, **3**, 356-366. <https://doi.org/10.4236/ojs.2013.35041>
- [26] Jolliffe, I.T. (1986) *Principal Component Analysis*. Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4757-1904-8>
- [27] Gibbons, J.D. and Chakraborti, S. (2003) *Nonparametric Statistical Inference*. 4th Edition, CRC Press, London.
- [28] Benegas, M. and Silva, F.G. (2010) *Estimação da Eficiência Técnica do SUS nos Estados Brasileiros na Presença de Variáveis Contextuais*. Texto para Discussão, CAEN-UFC.
- [29] Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Appendix

Table A1. Results for general and reduced models.

Output	Description
Y_1	Male life expectancy
Y_2	Female life expectancy
Y_3	Combined life expectancy
Y_4	Survival rate (%)
Y_5	Physicians per 1000 Inhabitants
Y_6	Hospital beds per 1000 inhabitants
Y_7	Coverage MMR (%)
Y_8	Tetavalent vaccine coverage (%)
Y_9	BCG vaccination coverage (%)
Y_{10}	Polio vaccine coverage (%)
Y_{11}	Sanitation coverage (%)
Y_{12}	Garbage collection coverage (%)

Source: [28].



Submit or recommend next manuscript to SCIRP and we will provide best service for you:

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact ajor@scirp.org