

Efficiency measurement using Data Envelopment Analysis combined with Principal Component Analysis

Agnese Rapposelli

Dipartimento di Metodi Quantitativi e Teoria Economica

Università “G.D’Annunzio” Chieti-Pescara

Viale Pindaro 42, Pescara

e-mail:agnese.rapposelli@virgilio.it

Abstract: Frontier analysis methods, such as Data Envelopment Analysis (DEA), seek to investigate the technical efficiency of productive systems which employ input factors to deliver outcomes. In economic literature one can find extreme opinions about the role of input/output systems in assessing performance. For instance, it has been argued that if all inputs and outputs are included in assessing the efficiency of units under analysis, then they will all be fully efficient. Discrimination can be increased, therefore, by being parsimonious in the number of factors. To deal with this drawback, we suggest to employ Principal Component Analysis (PCA) in order to aggregate input and output data. In this context, the aim of the present paper is to evaluate the performance of an Italian airline for 2004 by developing a model based upon PCA and DEA techniques.

Keywords: technical efficiency, Data Envelopment Analysis (DEA), Principal Component Analysis (PCA), airline.

1. Introduction

Modern efficiency measurement begins with Farrell (1957), who introduced the seminal concept of technical efficiency. He drew upon the theory of Pareto optimality stating that a productive system is technically efficient if it either maximises output for a given amount of input or minimises input to achieve a given level of output. Moreover, he introduced the concept of the best practice frontier, also called efficiency frontier: according to him, the measure of technical efficiency is given by the relative distance between the observed production and the nearest benchmark production lying on the frontier. It was the seed for later exploitation, following its rediscovery by Charnes et al. (1978) and subsequent relabelling as CCR-efficiency under the broader heading of Data Envelopment Analysis (DEA) (Stone, 2002). DEA method measures technical efficiency relative to a deterministic best practice frontier, which is built empirically from observed inputs and outputs using linear programming techniques. Its main advantage is that it allows several inputs and several outputs to be considered at the same time.

Apart of the nature of the inputs and outputs used in assessing efficiency, it must be remembered that questions can also be raised concerning the appropriate number of inputs and outputs for describing an activity process. Also in DEA context the first problem in the selection of inputs and outputs is to include factors indiscriminately. As DEA allows flexibility in the choice of inputs and outputs weights, the greater the number of factors included the less discriminatory the method appears to be. In order to

individuate a significant number of inefficient organisations, the literature suggests, in the case of static analysis that the number of units is greater than $3(m + s)$, where $m + s$ is the sum of the number of inputs and number of outputs. Another suggested rule (Dyson et al., 2001) is that, to achieve a reasonable level of discrimination, the number of units has to be at least $2m \times s$. Thus the number of inputs and outputs included in a DEA assessment should be as small as possible in relation to the number of units being assessed. This can be achieved by using Principal Component Analysis (PCA), which is able to reduce the data to a few principal components whilst minimising the loss of information. This process provides therefore a more parsimonious description of a relatively large multivariate data set.

Following on from the above discussion, our objective is to adapt the techniques of efficiency measurement, such as DEA, to the problem at hand. This study proposes a modified DEA model that includes PCA results and apply it to measure the technical efficiency of the Italian airline Air One, by comparing its domestic routes for 2004.

This paper is organised as follows. Section 2 provides the technical framework for the empirical analysis, Section 3 describes inputs and outputs used and lists the results obtained.

2. Methods

2.1 Data Envelopment Analysis (DEA)

DEA is a linear-programming technique for measuring the relative efficiency of a set of organisational units, also termed Decision Making Units (DMUs). The basic DEA models measure the technical efficiency of a DMU in terms of the maximal radial contraction to its input levels (input orientation) or expansion to its output levels feasible under efficient operation (output orientation).

In this study we use the following output-oriented BCC formulation of DEA method (Banker, Charnes and Cooper, 1984), which assumes variable returns to scale of activities (VRS):

$$\begin{aligned} e_0 &= \max \phi_0 \\ \text{subject to} \\ \sum_{j=1}^n \lambda_j x_{ij} &\leq x_{ij0}, \quad i = 1, \dots, m \end{aligned} \quad (1)$$

$$\phi_0 y_{rj0} - \sum_{j=1}^n \lambda_j y_{rj} \leq 0, \quad r = 1, \dots, s \quad (2)$$

$$\sum_{j=1}^n \lambda_j = 1, \quad (3)$$

$$\lambda_j \geq 0, \quad \forall j \quad (4)$$

where y_{rj} is the amount of the r -th output to unit j , x_{ij} is the amount of the i -th input to unit j , λ_j are the weights of unit j and ϕ_0 is the scalar expansion factor for DMU j_0 currently assessed.

DMU j_0 is said to be efficient (has a score of unity) if no other unit or combination of units can produce more than DMU j_0 on at least one output without producing less in some other output or requiring more of at least one input.

2.2 Principal Component Analysis (PCA)

The technique of PCA explains the variance structure of a matrix of data by using a few linear combinations of this original variables. The object is to take p variables $X_1, X_2, \dots, X_p$ and find linear combinations of these to produce principal components $Z_1, Z_2, \dots, Z_p$ that are uncorrelated. The principal components are also ordered so that Z_1 displays the largest amount of variation, Z_2 displays the second largest amount of variation, and so on: that is, $\text{var}(Z_1) \geq \text{var}(Z_2) \geq \dots \geq \text{var}(Z_p)$. Often much of the total system variability can be accounted for by a small number k of the principal components, which can then replace the initial p variables.

Principal components used here are computed based on the correlation matrix rather than on covariance, as the variables are quantified in different units of measure.

3. Case study

As mentioned before, this study evaluates the comparative performance of thirty Air One domestic routes for the year 2004. The airline production process defined in this empirical study is characterised by a higher number of inputs than those considered in Nissi and Rapposelli (2008), where only two categories of costs have been included in the DEA model. The input selected here are the number of seat available for sale (aircraft capacity), block time hours and several airline costs categories such as total variable direct operating costs (DOCs), total fixed direct operating costs (FOCs), commercial expenses, overhead costs and financial costs. With regard to the output side of the model, the output variables are the number of passengers carried, passenger scheduled revenue, cargo revenue and other revenues. All data were selected from Financial Statements as at 31st December 2004 and from various internal reports.

Unlike the previous paper, in this study all the original input variables are very highly correlated. This is therefore good material for using PCA to produce a reduced number of inputs. The first principal component Z_1 explains 98.22% of the variance of the data vector, so inputs will be included in the DEA model in the form of principal component rather than original data. The general DEA formulation has to be modified to incorporate Z_1 directly into the linear programming problem. In particular, the constraint (1) is replaced with the following:

$$\sum_{j=1}^n \lambda_j z_{1j} \leq z_{1j_0} \quad (5)$$

Generally inputs and outputs of DEA models need to be strictly positive, but the results of a PCA can be negative. It has been argued (Pastor, 1996) that the BCC output-oriented model is input translation invariant. Hence, the efficiency classification of DMUs is preserved if the values of principal component Z_1 are translated by adding a

sufficiently large scalar β (1 in this case) such that the resulting values are positive for each DMU j .

The DEA analysis has been performed by using DEA-Solver software. The efficiency scores of Air One routes, in descending order of efficiency, are shown in Table 1.

Table 1: *Efficiency ratings of Air One domestic routes*

DMU	Score	DMU	Score	DMU	Score	DMU	Score	DMU	Score	DMU	Score
TT	1	Z	1	EE	0.9462	F	0.7912	H	0.7370	JJ	0.6800
A	1	V	1	L	0.8418	U	0.7808	N	0.7159	OO	0.6608
CC	1	J	1	T	0.8275	W	0.7650	QQ	0.7080	SS	0.6431
VV	1	E	0.9866	FF	0.8232	Q	0.7547	G	0.6997	PP	0.6296
DD	1	S	0.9674	X	0.8027	AA	0.7541	B	0.6833	MM	0.6198

Two different remarks can be made. First, the result consists of a large number of inefficient units, nearly one third of the 30 DMUs and the average level of technical efficiency is 0.8273. Second, eight routes are fully efficient and three more routes are quite close to the best practice frontier. This would suggest that Air One routes are not operating at a very high level of efficiency, so there is room for improvement in several routes. However, these results could be improved. Further research in this field could include in the DEA-PCA model the presence of undesirable outputs, such as the number of delayed flights, which may also reflect the service quality of the network.

References

- Adler N., 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.
- Banker R.D., Charnes A., Cooper W.W. (1984) Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis, *Management Science*, 30, 1078-1092.
- Charnes A., Cooper W.W., Rhodes E. (1978) Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2, 429-444.
- Cooper W.W., Seiford L.M., Tone K. (2000) *Data Envelopment Analysis, a comprehensive test with models, applications, references and DEA-Solver software*, Kluwer Academic Publishers, Boston.
- Dyson R.G., Allen R., Camanho A.S., Podinovski V.V., Sarrico C.S., Shale E.A. (2001) Pitfalls and protocols in DEA, *European Journal of Operational Research*, 132, 245-259.
- Farrell M.J. (1957) The measurement of productive efficiency, *Journal of the Royal Statistical Society*, 120, 253-281.
- Nissi E., Rapposelli A. (2008) A Data Envelopment Analysis study of airline efficiency, in: *Research methodology on Data Envelopment Analysis*, Mantri J.K. (Eds.), Brown Walker Press, Boca Raton, FL, USA, 269-280.
- Pastor J. T. (1996) Translation invariance in data envelopment analysis: a generalization, *Annals of Operations Research*, 66, 93-102.
- Stone M. (2002) How not to measure the efficiency of public services (and how one might), *Journal of Royal Statistical Society A*, 165, 3, 405-434.