Workshop: Modelos de Apoio à Decisão na Agricultura e Ambiente

EFFICIENT DECISIONS USING DEA

Armando B. Mendes
CEEAplA and Azores University
Mathematical Department
Portugal, amendes@uac.pt

Veska Noncheva
University of Plovdiv “Paisii Hilendarski”
Bulgaria, veska@uac.pt

Emilana Silva
CEEAplA and Azores University
9700-851 Angra do Heroísmo
Portugal, emiliana@uac.pt

Abstract

Data Envelopment Analysis (DEA) is becoming an increasingly popular management tool for decision support related to efficiency comparisons. The task of the DEA is to evaluate the relative performance of units of a system. Is not a problem solution technique but an important problem analysis method based in mathematical programming with some similarities with Multiple Criteria Decision Analysis (MCDA).

DEA makes it possible to identify efficient and inefficient units in a framework where results are considered in their particular context. The units to be assessed should be relatively homogeneous and were originally called Decision Making Units (DMUs). It is an extreme point method that compares each DMU with the "best" DMUs.

The “Productivity Analysis with R” (PAR) framework establishes a user-friendly data envelopment analysis environment with special emphasis on variable selection and aggregation, and summarization and interpretation of the results. PAR framework has been developed to distinguish between efficient and inefficient observations of performances and to advise explicitly for producers’ possibilities to optimize their production.

In this work we will apply PAR to farms in Terceira Island, with a small data set of 30 farms. This data set includes 14 input variables and 4 output variables. With PAR was possible to conclude that 4 farms are scale-efficient and the others are scale inefficiency due to decreasing returns to scale or increasing returns to scale. This implies that either the dairy farm is too big (the number of cows is too large) or to small and that the farmer can improve the productivity of inputs and hence reduce unit costs by reducing or increasing the dimension of the farm.

Keywords: Efficiency decisions, DEA, CCA, feature selection and extraction.

1. Introduction

Data Envelopment Analysis (DEA) is becoming an increasingly popular management tool. The task of the DEA is to evaluate the relative performance of units of a system. It has useful applications in many evaluation contexts and across several disciplines. Is not a problem solution technique but an important problem analysis method based on mathematical programming with some similarities with Multiple Criteria Analysis (MCA).

DEA makes it possible to identify efficient and inefficient units in a framework where results are considered in their particular context. The units to be assessed should be relatively homogeneous and were originally called Decision Making Units (DMUs). It is an extreme point method and compares each DMU with only the "best" DMU.

DEA can be a powerful tool when used wisely. A few of the characteristics that make it powerful are:

- DEA can handle multiple input and multiple output models.
- DMUs are directly compared against a peer or combination of peers.
- Inputs and outputs can have very different units. For example, one variable could be in units of lives saved and another could be in units of dollars without requiring an a priori tradeoff between the two.

The same characteristics that make DEA a powerful tool can also create problems. An analyst should keep these limitations in mind when choosing whether or not to use DEA.

- Since DEA is an extreme point technique, noise such as measurement error and outliers can cause significant problems.
- DEA is good at estimating "relative" efficiency of a DMU but it converges very slowly to "absolute" efficiency. In other words, it can tell you how well you are doing compared to your peers but not compared to a "theoretical maximum."
- The results of DEA are difficult to understand for many inputs and outputs.
- The method does not scale well with the number of variables (inputs and outputs) included in the model and the number of efficient DMUs are highly dependent on this number.

In this article we apply canonical correlation analysis as one possible mitigation of the problems identified before. This possible solution is applied to a data set of farms in Terceira Island. This is a small data set where the identified DEA problems are more acute.

2. DEA models

In this work, DEA is used to measure the technical efficiency and the scale efficiency of Azorean dairy and beef farms. Input-oriented technical efficiency measures satisfying three different types of scale behavior are specified and applied to the data on Terceira farms. These are constant returns to scale (CRS), non-increasing returns to scale (NIRS), and variable returns to scale (VRS).

Efficiency of a decision making unit (DMU) is defined as the ratio between a weighted sum of its outputs and a weighted sum of its inputs. We can find the DMU (or the DMUs) having the highest ratio. Then we can compare the performance of all other DMUs relative to the performance of these previous identified. We can calculate the relative efficiency of the DMUs.

Suppose there are *n* DMUs. Suppose *m* input items and *s* output items are selected. Let the input and output data for DMUs be, respectively:

$$\underset{m\times n}{X} = (x_{ij})_{i=1,\dots,m;\, j=1,\dots,n}$$
$$\underset{s\times n}{Y} = (y_{kj})_{k=1,\dots,s;\, j=1,\dots,n} \qquad (1)$$

Given the data, we measure the efficiency of each DMU. Hence we need *n* optimizations (one for each DMU to be evaluated). Let the DMU, which are being evaluated, be designated as DMU_o (o=1,2,…,*n*).

Following Norman and Stoker (1991), the input oriented CRS model aims to minimise inputs while satisfying at least the given output levels. The CRS input-oriented measure of technical efficiency for the farms θ^*_{CRS} is calculated as the solution to the following mathematical programming problem:

$$\min_{\theta,\lambda} \theta$$

subject to:

$$\theta x_0 - X\lambda \geq 0$$
$$Y\lambda \geq y_0$$
$$\lambda \geq 0 \qquad (2)$$

where, for any DMU_o there are the following arrays:

$$\underset{m\times 1}{x_o} = (x_{1o}, x_{2o}, \dots, x_{mo})^T$$
$$\underset{n\times 1}{\lambda} = (\lambda_1, \dots, \lambda_n)^T \qquad (3)$$

and θ is a real variable. The scalar value θ represents a proportional reduction in all inputs such that $0 \leq \theta \leq 1$, and θ^*_{CRS} is the minimum value of θ so that $\theta^*_{CRS}x_0$ represents the vector of technically efficient inputs for the farm we are evaluating. This model is also known as CCR model, which are the initials of the 3 authors: Charnes, Cooper & Rhoades (1978).

For an inefficient DMU_o, we define its reference set E_o by $E_o = \{j \mid \lambda_j^* > 0\}$, $j=1,\ldots,n$.

Maximum technical efficiency is achieved when θ^*_{CRS} is equal to unity. In other words, according to the DEA results, when θ^*_{CRS} is equal to unity, a farm is operating at best-practice and cannot, given the existing set of observations, improve on this performance. When θ^*_{CRS} is less than unity, the DEA results imply that a farm is operating at below best-practice and can, given the existing set of observations. To improve the productivity of its inputs benchmarking partnerships are recommended, emulating the best practices of its best-practice reference set of farms.

The non-increasing returns to scale technical efficiency θ^*_{CRS} is calculated as the solution to the following mathematical programming problem:

$$\min_{\theta,\lambda} \theta$$

subject to:

$$\begin{aligned} &\theta x_0 - X\lambda \geq 0 \\ &Y\lambda \geq y_0 \\ &\lambda \geq 0 \\ &\ell\lambda \leq 1 \end{aligned} \tag{4}$$

where ℓ is a $(1\text{x}n)$ vector of ones.

The variable returns to scale technical efficiency $\theta^*_{VRS}$is calculated by:

$$\min_{\theta,\lambda} \theta$$

subject to:

$$\begin{aligned} &\theta x_0 - X\lambda \geq 0 \\ &Y\lambda \geq y_0 \\ &\lambda \geq 0 \\ &\ell\lambda = 1 \end{aligned} \tag{5}$$

The input-oriented scale efficiency measure is defined as the ratio of overall technical efficiency to variable returns to scale technical efficiency:

$$SE = \frac{\theta^*_{CRS}}{\theta^*_{VRS}} \tag{6}$$

.

If the value of the ratio (7) is equal to unity (i.e. SE=1), then the dairy farm is scale-efficient. This means that the farm is operating at its optimum size and hence that the productivity of inputs cannot be improved by increasing or decreasing the size of the dairy farm. If the value of the ratio is less than unity (i.e. SE < 1), then the DEA results indicate that the farm is not operating at its optimum size. If SE < 1 and $\theta^*_{NIRS} = \theta^*_{CRS}$, then the DEA results suggest that scale inefficiency is due to increasing returns to scale. This means that by increasing the size of the dairy farm, the farmer can improve the productivity of inputs and thereby reduce unit costs. If SE < 1 and $\theta^*_{NIRS} > \theta^*_{CRS}$, then the DEA results suggest that scale inefficiency is due to decreasing returns to scale. This implies that the dairy farm is too big and that the farmer can improve the productivity of inputs and hence reduces unit costs by reducing its size.

3. Terceira farms efficiency

Terceira is the second biggest island, on area but also on economic and agricultural context, in the Azorean archipelago. The Azores islands are a Portuguese territory with a population of about 250 000 inhabitants. The majority (about 75%) of this population lives in S. Miguel and Terceira islands. The main economic activity is dairy and meat farming. In S. Miguel, Terceira and S. Jorge islands, about 24% of the farms produce only milk, about 13% of farms produce only meat and 24% produce both and some cultures as well. The remaining farms produce other agricultural production. Dairy policy depends on Common Agricultural Policy of the European Union and is limited by quotas. In this context, decision makers need knowledge for deciding the best policies in promoting quality and best practices. One of the goals of our work is to provide Azorean Government with a reliable tool for measurement of productive efficiency of the farms.

In Azores there are about 15.107 farmers. Azorean farms are small - about 8 hectares per farm, what is about the half of the average European farm dimension (15.8 in 2003). The production system is primarily based on grazing (about 95% of the area). In the last years, the most representative expenses – based in data of FADN (Farm Account Database Network) are on concentrates, annual depreciation, rents and fertilizers. The subsidies are important for the dairy farms, and in 2004 they were about 61.6% of all profit. Some of these subsidies are compensations for low selling prices received by farmers, and so they are due after the production of meat and milk, others are incentives to investment and compensation for high prices of production factors. There are also subventions to improve ecological production.

The aim of this study is the best-practice farms to be identified and help the less efficient farms to identify their relevant benchmark partners. The former should then be able to identify and emulate the better practices of the latter and thereby eliminate the controllable sources of inefficiency.

Some research work on the dairy sector in Azores has been already done (see Marote and Silva, 2002; Silva, *et al.,* 2001). The beef sector in Azores has been investigated by means of Stochastic Frontier Analysis (see Silva, 2004).

Dairy and beef farms produce a wide range of outputs and also use a large number of resources. Any resource used by an Azorean dairy farm is treated as an input variable and because of it the list of variables that provide an accurate description of the milk and meat production process is large. The names of all input variables used in the analysis are the following: EquipmentRepair, Oil, Lubricant, EquipmentAmortization, AnimalConcentrate, VeterinaryAndMedicine, OtherAnimalCosts, PlantsSeeds, Fertilizers, Herbicides, LandRent, Insurance, MilkSubsidy, MaizeSubsidy, SubsidyPOSEIMA, AreaDimension, and DairyCows. All these variables are metrical numerical quantities. Almost all are expressed in monetary values except the AreaDimension, stated in square meters and DairyCows is a head count. The main output variables are Milk and Cattle also expressed in monetary values and representing how much the farm earned by selling these two main products. All subsidies are reflected in the two output variables used in the study (ProductionSubsidy and FactorsSubsidy) and they are evaluated in currency values. In Terceira there are 30 farms with milk and cattle production.

DEA allows the measurement of technical efficiencies of individual farms, taking account of all quantitative variables.

The analysis of the Terceira's farms efficiency is implemented in R statistical software version 2.8.1 using the DEA, FEAR and CCA packages and routines developed by the authors (see R Development Core Team, 2007).

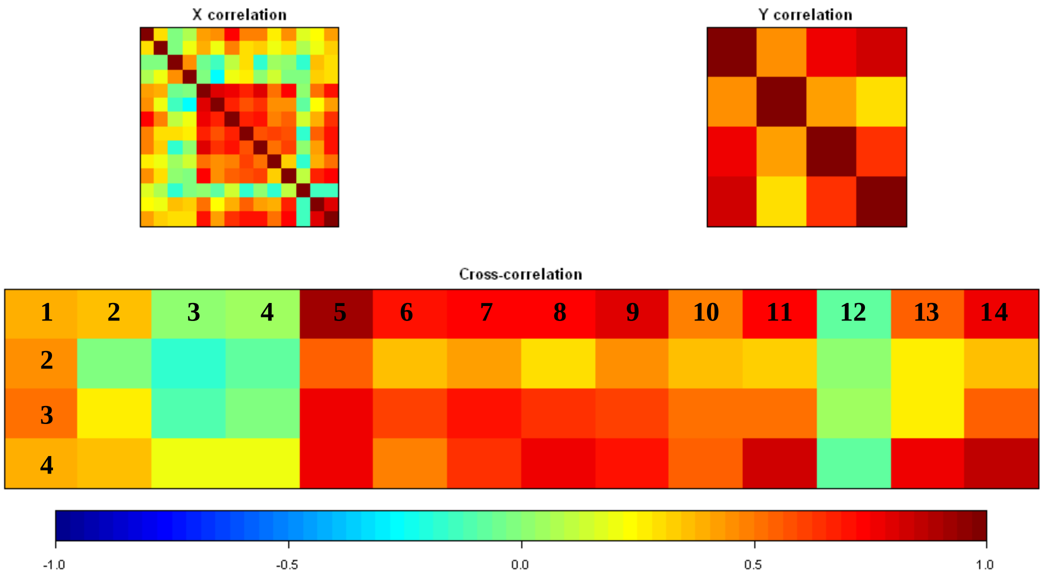
At first we applied DEA with all variables and we received that maximum technical efficiency is achieved by 28 farms. According to the DEA result 28 farms are operating at best practice and cannot improve on this performance. However, the farmers are mainly old, more than 55 years old and low educated, mainly with basic education with only about four years. This characterization of the Azorean farms is based on national agricultural institutional data and previous works (see, for example, Marote and Silva, 2002; Silva, 2004; Silva, *et al.,* 2001). It is also well-known that when one has few DMUs and many inputs or outputs many of the DMUs appear on the DEA frontier. We made Azorean agricultural industry look good by restricting the sample size to farms on Terceira only, and by increasing the number of inputs and outputs taking account too many variables.

In order to build an adequate model of dairy and beef productivity an appropriate variable selection method is needed.

The application of canonical correlation analysis highlights the correlations between input and output data sets, called X and Y, respectively. The correlation matrixes are visualized in Figure 1. In this figure we can see high correlations between the milk production and the two subsidy variables. This correlation is much higher than with the beef production. This translates de fact that subsidies have more impact in milk production than in beef production. In cross-correlation matrix we can also see the high correlation between milk production and money spent in concentrate food, which was anticipated as intensive farms produce much more milk and consumes much more concentrate than no intensive production. This is not so visible in beef production. The factors subsidies are directly related with the number of caws (DairyCows) and dimension of the farm (AreaDimension, LandRent) and so the high correlation is easily explained. As expected, production subsidies are more correlated with food concentrate, veterinary and other animal costs. What was unexpected was the low correlation of those variables with the production subsidies.

Figure 1. Visualization of sample correlation coefficients.

(the numbers in lower picture are the same used to identify variables in table 1)



The structure coefficients are given in Table 1. If we consider that the two Canonical Correlation variates extracted represent reasonably well all the data, then it makes sense to choose both inputs and outputs with the biggest structure coefficients to be included in the DEA model. The chosen input variables are AnimalConcentrate and DairyCows. Let the output data for DMUs be Milk and FactorsSubsidy.

Table 1. Input and output variables and their structure weights.

	Input variables (X)	structure weights	structure weights
1	EquipmentRepair	-0.463883	-0.447400
2	Oil	-0.367265	-0.354215
3	Lubricant	-0.084217	-0.081225
4	EquipmentAmortization	-0.122065	-0.117728
5	AnimalConcentrate	**-0.923175**	-0.890372
6	VeterinaryAndMedicine	-0.659249	-0.635824
7	OtherAnimalCosts	-0.756439	-0.729560
8	PlantasSeeds	-0.807528	-0.778834
9	Fertilizers	-0.819276	-0.790165
10	Herbicides	-0.575547	-0.555096
11	LandRent	-0.831971	-0.802408
12	Insurance	0.071831	0.069279
13	AreaDimension	-0.706559	-0.681453
14	DairyCows	-0.884397	**-0.852972**

	Output variables (Y)	structure weights	structure weights
1	Milk	**-0.923248**	-0.957263
2	Cattle	-0.486988	-0.504929
3	ProductionSubsidy	-0.728532	-0.755373
4	FactorsSubsidy	-0.908093	**-0.941549**

In the absence of environmental differences (*i.e.* differences in soil quality, animal genetics, climate and other unspecified variables) and errors in the measurement of inputs and outputs, pure technical inefficiency would reflect departures from best-practice farm management. The way to eliminate this latter source of inefficiency would be to form a benchmarking partnership with relevant best-practice farms with a view to identifying and then emulating their farm management practices.

The output of DEA therefore includes measures of each farm's scale efficiency (SE), pure technical efficiency, overall technical efficiency and identification of its best-practice benchmark (see Table 2). The latter identifies potential benchmark partners along with their respective contributions to the best-practice benchmark.

The result of DEA is given in Table 2. On this basis the following preliminary conclusions can be made.

Table 2. Efficiency measures for Terceira farms.

DMU	Constant returns to scale	Non-increasing returns to scale	Scale efficiency (SE)	Reference set	DMU	Constant returns to scale	Variable returns to scale	Scale efficiency (SE)	Reference set
1	0.8407	0.8407	0.9765	12; 20	16	0.7148	0.8310	0.9390	12; 14; 13
2	0.8630	0.8769	0.9841	12; 13; 14	17	0.4305	0.5901	0.7297	14; 12; 20
3	0.4497	0.4497	0.6110	13; 14	18	0.6939	0.7136	0.9724	12; 20
4	0.5842	0.5909	0.9887	12; 14; 20	19	0.5292	0.9361	0.5654	14; 13; 12
5	0.7894	1.000	0.7893	14; 13; 12	20	1.000	1.000	1.000	
6	0.6560	1.000	0.6560	14; 13; 12	21	0.7155	0.7206	0.9930	14; 12; 20
7	0.7793	0.7793	0.9940	12; 20	22	0.8722	0.8815	0.9894	12; 20
8	0.5451	0.8296	0.6571	14; 12; 13	23	0.9687	0.9855	0.9829	14; 20; 12
9	0.6626	0.6991	0.9477	13; 14; 12	24	0.8770	0.8867	0.9890	12; 14; 20
10	0.6714	0.6714	1.000	14; 20	25	0.8177	0.8273	0.9884	12; 14; 20
11	0.8098	0.8098	0.9891	20; 12	26	0.6811	0.8114	0.8394	14; 12; 20
12	1.000	1.000	1.000		27	0.5935	0.6665	0.8904	12; 4; 13
13	1.000	1.000	1.000		28	0.5002	0.5226	0.9572	12; 20
14	1.000	1.000	1.000		29	0.8349	0.8489	0.9835	12; 20; 14
15	0.8127	0.8127	0.9780	12; 14; 20	30	0.9221	0.9602	0.9603	12; 20

The farms 12, 13, 14, and 20 are scale-efficient. This means that the farms are operating at its optimum size and hence that the productivity of inputs cannot be improved by increasing or decreasing this kind of production factors. The others 26 farms in Terceira are not operating at its maximum efficiency. The farms 1, 3, 7, 10, 11, 15, 18, 22, 23, 25, 27, 28 and 29 can improve the productivity of inputs and thereby reduce unit costs. We can conclude this analysing the DEA results, in Table 2, which suggest that scale inefficiency is due to increasing returns to scale. The others 13 farms are too big and so, the farmer can improve the productivity of inputs and hence reduce unit costs by reducing the size of the farm (the number of cows, the pasture, *etc.*). The reference set for each inefficient farm is depicted in Table 2. It identifies potential benchmark partners.

We can summarize that thirteen Agricultural farms in Terceira have SE<1 and $\theta_{CRS}^{j} = \theta_{NIRS}^{j}, j = 1, 2, \dots, 13$. Because of it the DEA results suggest that scale inefficiency is due to increasing returns to scale. This means that by increasing the size of the dairy farm (number of cows, pasture, *etc.*), the farmer can improve the productivity of inputs and thereby reduce unit cost. Other thirteen Agricultural farms in Terceira have SE<1 and $\theta_{CRS}^{j} < \theta_{NIRS}^{j}, j = 1, 2, \dots, 13$. Because of it the DEA results suggest that scale inefficiency is due to decreasing returns to scale. This implies that the dairy farm is too big (the number of cows is too large) and that the farmer can improve the productivity of inputs and hence reduce unit costs by reducing the number of cows in the farm.

On the basis of this study, senior management can only make some preliminary conclusions each must be confirmed with domain knowledge. The assessment questions were analysed through the type of mathematical and statistical analysis described above, but extensive and detailed subsequent analysis of pointed farms is required before any sound decision can be made.

4. CONCLUSION

The primary objective of our work has been to measure the scale efficiency of Azorean dairy and beef industry using data envelopment analysis (DEA). As the study proceeded another objective emerged. This was to find a variable selection algorithm appropriate for the efficiency analysis. In our approach to efficiency measurement CCA provides selection of both input and output units and then DEA provides the overall technical efficiency, scale efficiency and pure technical efficiency. This approach is applied to all Terceira farms and it will be also applied to all Azorean farms. It could form the basis for performance improvement by identifying potential benchmark partners for the less efficient Azorean farms.

Note that this method, in contrast to other published works, (see Chilingerian, 1995, and Salinas-Jimenez and Smith, 1996) tends to choose highly correlated variables. The reasoning expressed for eliminating high correlated variables with existing model variables are that those variables are merely redundant and so not useful. In fact, this is generally accepted in several areas but we think this reasoning cannot be applied to DEA models as we found very different models choosing only one variable for input and output and choosing two variables as presented in this article. In the limit, when correlation coefficient is one, both variables have the same information and produce the same model using only one or the two. When correlation is high but not one, the two variables represent two different perspectives for efficiency analysis and can produce different models, and so should be included. In this way is preferable to eliminate variables that influence less the efficient DMUs chosen.

ACKNOWLEDGEMENT

This work has been partially supported by Direcção Regional da Ciência e Tecnologia of Azores Government through the project M.2.1.2/l/009/2008.

REFERENCES

Charnes, A., W. Cooper, & E., Rhodes (1978) "Measuring the efficiency of decision-making units," *European Journal of Operational Research* vol. 2, pp. 429–444.

Chilingerian, J.A. 1995. Evaluating physician efficiency in hospitals: A multivariate analysis of best practices, *European Journal of Operational Research* 80 pp. 548–574

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.

Lewin A.Y., R.C. Morey and T.J. Cook, 1982. Evaluating the administrative efficiency of courts, *OMEGA* **10** (4), pp. 401–411

Marote, Eusébio and Silva, Emiliana, 2002. Análise dinâmica da eficiência das explorações leiteiras da ilha Terceira. *Actas do Congresso de Zootecnia*, 12ª ed.

Norman M. and B. Stoker, 1991. *Data Envelopment Analysis: The assessment of performance*, John Wiley and Sons, Chichester, England.

Nunamaker, Thomas R 1985. Using data envelopment analysis to measure the efficiency of non-profit organizations: A critical evaluation, *Managerial and decision economics*, vol. 6(1), pp.50 -58.

R Development Core Team. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, http://www.R-project.org, 2008

Salinas-Jimenez, J. and Smith, P. 1996. Data envelopment analysis applied to quality in primary health care, *Annals of Operations Research* 67, pp. 141–161.

Sigala, M., Airey, D., Jones, P. and Lockwood, A. 2004. ICT paradox lost? A stepwise DEA methodology to evaluate technology investments in tourism settings, *Journal of Travel Research* 43, pp. 180–192.

Silva, E., Berbel, J. and Arzubi, A. 2001. Tipología y análisis no paramétrico de eficiencia de explotaciones lecheras en Azores (Portugal) a partir de datos de RICA-A. *Economía agraria y recursos naturales: Nuevos enfoques y perspectivas – Actas do Congreso de la Asociación Española de Economía Agraria,* 4ª ed., Universidad Pública de Navarra

Silva, Emiliana and Venâncio, Fátima 2004. A competitividade das explorações pecuárias no Faial: Recurso a metodologias alternativas. Actas do Congresso de Economistas Agrícolas, 4ª ed.

Wagner, Janet M. and Shimshak, Daniel G. 2007. Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives, *European journal of operational research*, vol. 180, pp. 57 -67.