Data Envelopment Analysis as a Complement to Marginal Analysis

A. M. Theodoridis, A. Psychoudakis and A. Christofi*

Abstract
The consideration in the present study is mainly conceptual. The objective is to show how Data Envelopment Analysis (DEA) can be used to reveal the true input-output relations in an industry. In the estimation of a production function it is assumed that all firms use the existing technology efficiently. However, in the real world the observed firms produce homogeneous outputs with differences in factor intensities and in managerial capacity. Hence, inefficiencies are hidden in the estimated production functions. In order to overcome this drawback of the parametric approach and to reveal the true nature of the input-output relations in production, given the available technology, the DEA approach is applied. In this study DEA is applied in order to select the farms that utilize efficiently the existing technology, allowing the estimation of a production function that reveals the true input-output relations in sheep-goat farming, using farm accounting data from a sample of 108 sheep-goat farms.

Introduction

Parametric approaches have been extensively used to estimate input-output relationships in a firm or in an industry in order to study the efficiency of resource allocation. The most celebrated of them is the Cobb-Douglas production function. The Cobb-Douglas function has been widely used in the early stages of empirical applications of production theory. However, this particular form has been unduly restrictive. To render a model operational and to limit the restrictive properties imposed on the production process, the translog production function is chosen very often and tested against the restricted Cobb-Douglas functional form. The estimation of translog functions has been extensively used for the flexibility it provides (Corbo and Meller, 1979; Berndt and Christensen, 1973).

In cross-sectional studies a sample of farms provides the required farm accounting data for the estimation of a specified function. In this case the results reflect the average farm, which fails to account for different endowments of fixed factors of production and managerial entrepreneurship (technical efficiency) across observations, since the production function is the wrong trap for capturing such differences. When a single equation model is estimated by using the Cobb-Douglas production function or a more flexible one like translog production function, one of the basic assumptions is that all farms are operating at technically efficient level. However, not all farms are technically efficient.

* Aristotle University of Thessaloniki, Department of Agricultural Economics, 541 24, Thessaloniki, Greece
Production is possible with a variety of factor proportions and production technologies. Often, several sizes of farms and techniques of production exist side by side in the same region. Where there are several production techniques, it is possible that the partial production elasticities (the estimated parameters of the function) will differ significantly among the different techniques. Consequently, valuable information has been lost. Many authors in the economic literature (Lau and Yotopoulos, 1971; Yotopoulos and Nugent, 1976; Doran, 1985; Singh and Patel, 1973; Bagi, 1981 and Sharma, 1983) have dealt with the aforementioned problem. A common method used to assess these differences is dividing the sample into groups on the basis of some predetermined criteria. It is the alternative of categorizing the sample of farms by different production techniques.

The same concept is also applied in this paper. However, in this study Relative Technical Efficiency is the classification criterion. Selection of the sub-samples is based on Data Envelopment Analysis (DEA). The sample has been divided according to the results from the application of DEA and sub-samples for estimating separate production functions are formed. This way the estimation procedures may lead to parameter estimates with clearer economic content.

DEA is a non-parametric approach that has been extensively used for determining efficiency frontiers and deals with the nature, existence, and departures from them. This approach defines a non-parametric frontier and measures the efficiency of each unit relative to that frontier. In other words, the DEA approach provides an analytical tool for determining effective and ineffective performance as the starting point for inducing theories about best-practice behavior (Charnes et al., 1994). Hence, two techniques have been used in this paper. The first technique is DEA that uses linear programming to construct a frontier that envelops all observations and computes the relative Technical Efficiency of each farm included in the sample. The second technique is an econometric approach that estimates a production function by fitting a regression plane to the data.

It has to be mentioned that the primary aim of the present study is not to compare the success of the two techniques- non-parametric and parametric analysis-, but to investigate whether the combination of these two approaches would lead to a result with clearer economic content. It is expected that future theoretical developments will draw DEA and econometric approaches even closer together. DEA has proved particularly adept at recovering relationships that remain hidden for other methodologies (Banker et al., 1986; Seiford and Thrall, 1990).

The remainder of this paper has been structured as follows. In section 2 we outline the two methodologies, and their advantages and disadvantages. In the next section, the data and the models for the application are described. The results from using the two methodologies and a discussion of the results are presented in section 4. Section 5 concludes the paper.

An Outline of the Methodologies

Data Envelopment Analysis (DEA) has its origins in the seminal work by Charnes et al. (1978) who reformulated Farrell’s (1957) approach. In this study, they described DEA as a “mathematical programming model applied to the observational data that provides a new way of obtaining empirical estimates of extremal relations – such as the production functions and / or efficient production possibility surfaces that are a cornerstone of modern economics”.

In general, DEA methodology uses a set of production units of a sample to construct an efficiency frontier consisting of all possible linear combinations of efficient production units. The frontier technology consists of convex input and output sets enveloping the data points with linear facets. Consequently, the efficient units lie by definition on that frontier while the inefficiency of units that are not on the frontier is indicated in direct proportion to their distance from the frontier. Individual units are considered as Decision Making Units (DMUs) and efficiency can be measured relative to the highest observed performance rather than against some average. The proposed measure of efficiency of any DMU is obtained as the maximum of a ratio of weighted outputs to weighted inputs subject to the condition that the similar ratios for every DMU be less than or equal to unity.

Since DEA is now a well-established method it is not necessary to go into details about the theoretical background of the approach. The basic version of the DEA model, which is also known as the CCR model (it was initially proposed by Charnes, Cooper and Rhodes) can be found in Charnes et al. (1978, 1979, 1981). The extensions that have been proposed can be found in Färe et al. (1985) and in Seiford and Thrall (1990).

For the purpose of this study out of the six measures of relative efficiency (overall cost-minimising efficiency, technical efficiency, allocative efficiency, pure technical efficiency, scale efficiency and efficiency due to input congestion) only technical efficiency is computed applying the input oriented model.

The main advantage of DEA is that it does not require specification of the functional form of the production function. DEA calculations focus on individual observations in contrast to population averages. It can simultaneously utilize multiple outputs and multiple inputs with each being stated in different units of measurement. DEA also focus on revealed best-practice frontiers rather than on central-tendency properties or frontier and it generates the set of “peer” units with which a unit is compared.

However, several properties that represent strengths in one capacity may act as limitations in another. One of the main criticisms of DEA is that the method does not at first sight have any statistical foundation, i.e. that it is not possible to make inference about estimated DEA parameters, sensitivity, asymptotic properties etc. This poses a problem, seeing that uncertainty and measurement errors will often be present in observed data. Sometimes Stochastic Production Frontier (SPF) analysis may be preferred as a method that includes measurement errors and uncertainty (Aigner et al., 1977; Kumbhakar and Lovell, 2000).

The Cobb-Douglas model has been widely used in the agricultural economics. The use of single equation models for agricultural production functions has been justified by Griliches (1957), Mundlak and Hoch (1965), Hopper (1965) and Zellner et al. (1966) who argue that because inputs in agriculture are largely predetermined because of a considerable lag in production and due to the fact that error is weather determined, simultaneous equation bias will be small for well specified production functions. The production environment in the present study does not seem to differ from the specification requirements postulated by the authors mentioned above.

The primary purpose of the estimation of a production function is to obtain estimates of regression coefficients and marginal factor productivities, which can be useful for the study of efficiency when they are compared with marginal factor costs.

Parametric approach requires more assumptions about the production function and also about the distribution of the errors, although it is possible to test for the validity of
the assumptions and to determine whether particular variables are relevant. The main weaknesses of the regression approach is that it fits a function on the basis of average behavior; it requires pre-specification of the functional form; it does not take efficiency into consideration; it only gives residuals.

Data, Models and Methods

The farm accounting data for this empirical application were collected through a farm management survey, of a sample of 108 sheep-goat farms, carried out during the 2001-2002 period. The farms included in the sample are located in West Macedonia, Greece, which is an area where, traditionally, sheep-goat farming is an important sector of the livestock production. Furthermore, it has to be mentioned that West Macedonia is a region where several sizes of farms and techniques of production exist side by side. The initial sample included 180 farms but in the end, those, which reared either goats or sheep, were excluded.

All these farms have the required characteristics for the empirical application of DEA. Each DMU consumes varying amounts of inputs to produce different outputs. The application of DEA involves the identification and measurement of relevant inputs and outputs, which are common in all units. The relevant inputs used in this empirical application are: (1) The number of sheep in the herd, (2) The number of goats in the herd (3) The acreage on irrigated land, in 1000m$^2$ (stremma) (4) The acreage on non-irrigated land, in 1000m$^2$ (stremma) (5) Labour used, in hours (6) Machinery (annual expenses in Euros) (7) Buildings (annual expenses in Euros) (8) Variable Cost (in Euros) (9) Feed Purchased, in tonnes. The only output that has been used is Gross output, in Euros. It is of importance to state here that the relative efficiency score associated to a DMU is not affected by the choice of a different unit of measure. The measure of efficiency is independent of the units of measurement used. This property is referred to as “units invariance” (Cooper et al., 2000).

The approach applied consists of three steps. In the first step the input oriented DEA model is applied in a sample of 108 farms of the sheep-goat sector. Only 67 of these farms are technically efficient. DEA is applied again using this time as initial sample these 67 farms. The results indicate that only 49 of the farms are relatively technically efficient. The same procedure is followed using the sample of 49 farms, and the results indicate that all farms lie on the efficiency frontier. Thus, a sub-sample has been formulated where all DMU’s are relative technical efficient.

The aim of this procedure is not to estimate the efficiency score of the DMUs, but to end up with a sample where each of the farms is laid on the efficiency frontier. Through this non-parametric analysis three sub-samples have been formulated; the first one contains 108 farms, the second 67 farms, the third 49 farms.

Based on these results a parametric analysis using Least Squares has been applied in order to estimate the production function parameters. In view of the research problem to be explored and the aims of the study, two analytical models- namely translog and Cobb-Douglas production function- are specified. Translog production function has been estimated from a cross section sample of farms and tested against Cobb-Douglas (by using the likelihood ratio test) in order to see if Cobb-Douglas production function forms an adequate representation of the data. The translog production function is often written in its logarithmic form and the mathematical model to be estimated is:
\begin{equation}
\ln y = \beta_0 + \sum_{i=1}^{3} \beta_i \ln x_i + \sum_{j=1}^{3} \sum_{i=1}^{j} \beta_{ij} \ln x_i \ln x_j
\end{equation}

where the dependent variable \( y \) is some measure of output (in this case gross output in euros), \( X_1, X_2, X_3 \) are “independent” variables representing some measure of the inputs, and the \( \beta_j (j = 1, 2, 3) \) are unobserved population parameters. In this case

\( X_1 \) is the Labor Cost expressed in Euros,
\( X_2 \) is the Fixed Capital (annual expenses of machinery and buildings and value of live capital) expressed in Euros,
\( X_3 \) is the Variable Cost, Purchased Feed and Rent expressed in Euros.

Based on the results of the DEA three models (function of the same sample) are specified in order to obtain the estimates of the regressions coefficients. Hence, in total OLS is applied three times and three different elasticities are calculated for each input. As it has already mentioned model I contains 108 observations, model II contains 67 observations and model III contains 49 observations.

**Results**

The initial sample consists of 108 farms. DEA was applied on this sample and the average technical efficiency for this group was 94.35 percent. 41 of the farms included in the initial sample resulted to be technically inefficient (this means a percent below 100). By excluding these 41 farms from the sample a new sub-sample was constructed, which consisted of 67 farms. The same input oriented DEA model was applied to this sub-sample and the average technical efficiency for this group was 97.93 percent. This time 18 farms were technically inefficient. By excluding these 18 farms from the sample a second sub-sample was formulated, which included 49 farms. Again the same procedure was applied and all farms resulted to be technically efficient. This means that all 49 farms lie on the frontier. The results from the application of DEA in each case are presented in Table 1.

**Table 1. DEA Results**

<table>
<thead>
<tr>
<th>Sample (Number of farms)</th>
<th>Number of inefficient farms</th>
<th>Technical Efficiency (Mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model I</td>
<td>108</td>
<td>41</td>
</tr>
<tr>
<td>Model II</td>
<td>67</td>
<td>18</td>
</tr>
<tr>
<td>Model III</td>
<td>49</td>
<td>49</td>
</tr>
</tbody>
</table>

An implicit assumption of production functions is that they assume that there are no different endowments of fixed factors of production and no management bias; in other words all farms are technically efficient. Nevertheless, the production frontier indicates the maximum potential output for a given set of inputs. From the production frontier it is possible to measure the relative efficiency of certain groups or set of practices from the relationship between observed production and some ideal or potential production.
(Greene, 1993). This ascertainment was the elementary guide for the study. The basic concept was to investigate through DEA how the production function estimators are affected by the aforementioned drawback of the parametric method.

In order to examine this case thoroughly and to reveal this particular aspect of the problem, the specified production function was estimated for each sample formulated with the assistance of DEA. Furthermore, a generalized Likelihood-Ratio test (LR) was performed to test whether or not translog could be an appropriate functional form of the production function estimated in this study. The result of the LR test suggests that the Cobb-Douglas gives a more appropriate model and an adequate representation of the data under examination. The null hypothesis that all parameters are zero is accepted by the test at the 5 percent level. In all three cases LR test provides a statistic which does not exceed the critical value $x^2_6 = 12.6$. Thus translog is rejected confidently in favour of the Cobb-Douglas model and all results presented here from now on refer solely to the Cobb-Douglas production function.

It is not surprising that the results from the estimation of the C-D did yield significant results. All of the variables had the expected sign. The results of the cross-sectional analysis are presented in the following table 2.

### Table 2. Results of cross-section analysis, standard error in parentheses, *** indicates significance at the 1% level, ** at the 5% level, * at the 10% level

<table>
<thead>
<tr>
<th>Variable</th>
<th>108 farms</th>
<th>67 farms</th>
<th>49 farms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.6775**</td>
<td>0.7686*</td>
<td>0.8314*</td>
</tr>
<tr>
<td>Labour Cost</td>
<td>0.3050***</td>
<td>0.2745***</td>
<td>0.2807***</td>
</tr>
<tr>
<td>Fixed Capital</td>
<td>0.2106***</td>
<td>0.2332***</td>
<td>0.2311***</td>
</tr>
<tr>
<td>Variable Capital</td>
<td>0.4727***</td>
<td>0.4684***</td>
<td>0.4591***</td>
</tr>
<tr>
<td>Sum of elasticities</td>
<td>0.988</td>
<td>0.976</td>
<td>0.971</td>
</tr>
<tr>
<td>$R^2$ adjusted</td>
<td>0.897</td>
<td>0.873</td>
<td>0.885</td>
</tr>
</tbody>
</table>

The size of the adjusted coefficients of multiple determination suggests that the major part of the interfarm variation in output is explained by the observed inputs (0.90 in the first case, 0.87 in the second case and 0.89 in the third case). All of the input coefficients are significant at the 1 percent level, using a two-tailed test.

Under perfect competition, the sum of regression coefficients measures returns to scale. In our example, in all three cases the sum of regression coefficients is equal to one. This means that the farms operate under constant returns to scale. This is an expected result since there are a priori theoretical reasons to believe that Constant Returns to Scale will prevail (Heady and Dillon, 1961; Yotopoulos, 1968).

The input coefficients are interpreted as elasticities of production. Variable Capital (containing the value of the purchased feed and the rent of the land as already has been mentioned), 0.47 in the first model and 0.46 in the third model has the highest coefficient in all three models. Next in importance is the coefficient of Labour Cost, 0.31 in the first model and 0.28 in the third model. Last in importance is the coefficient of
Fixed Capital, 0.21 in the first and 0.23 in the third model. The shares of the factors of production are consistent with a priori expectations.

The Chow-test has been applied to the three samples to test if the differences between the parameter estimates were statistically significant (Maddala, 2001). When testing Model I (N=108) against Model II (N=67) the F-statistic has a value of 1.25 with d.f. 67 and 104. Hence, at the 15 percent level of significance, the hypothesis of equality between the parameter estimates is rejected. The same results occurs when Model I is tested against Model III (N=49). The F-statistic has a value of 1.29 with d.f. 49 and 104, which lead us to reject the hypothesis of stability. On the contrary, when Model II is tested against Model III the difference between the parameter estimates is proved to be no statistically significant.

The Chow-test is inaccurate if the error variances of the samples are unequal (Schmidt and Sickles, 1977). For this reason it was desirable to test the equality of the variances. At the 5 percent level of significance the hypothesis of equality of the error variances was not rejected. Thus, in all three cases the Chow-test is quite accurate.

Through our analysis we try to reveal the true nature of the production function. We make an attempt to obtain regression coefficients of production function that are free of technical inefficiency. This could become clear from the discussion of marginal productivities below.

The marginal product of a factor can be computed as the product of the factor’s elasticity times its average product. Given the relevant elasticities, marginal productivities can be computed at any point of the production function. It is convenient, however, to present the discussion in terms of the “average farm”, i.e., at the geometric means of output and inputs. And estimation at the geometric means is the most relevant in the context of a Cobb-Douglas function (Heady and Dillon, 1961).

The geometric means of the variables, the marginal productivities of the inputs and their opportunity cost are presented in Table 3. The dimensions of these values are also presented in the table.

By looking at the geometric means of the independent variables, which are computed for 10000 units of output so that a comparison between the different cases can take place, it is absolutely clear that farms that compose the sample in the third model are utilizing inputs in a more productive sense. All inputs, apart from Variable Capital, are decreasing in order to produce the same level of output (the increase in Variable capital

<table>
<thead>
<tr>
<th>Table 3. Marginal value products of production factors used</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of farms:</strong></td>
</tr>
<tr>
<td><strong>Sample means:</strong></td>
</tr>
<tr>
<td>Output (€)</td>
</tr>
<tr>
<td>Labor (€)</td>
</tr>
<tr>
<td>Fixed Capital (value in €)</td>
</tr>
<tr>
<td>Variable Capital (€)</td>
</tr>
<tr>
<td><strong>Marginal Products:</strong></td>
</tr>
<tr>
<td>Labor (€/€)</td>
</tr>
<tr>
<td>Fixed Capital (%)</td>
</tr>
<tr>
<td>Variable Capital (€/€)</td>
</tr>
</tbody>
</table>
can be considered as negligible). Hence, as the inefficiency, which is present in the sample under examination is obliterated the productivity is improved. Fewer inputs are demanded for the production of the same output, thereby releasing resources for other economic activities. The computation of the resource cost of the average farm for each case leads us also to similar deduction. Resource cost to achieve the output of 10,000 Euros in the first model is 11.5 percent higher than in the third model where all farms operate relatively efficiently. The difference in the resource cost between the second and the third model is 4 percent. This result suggests that as our sample under examination becomes more efficient the resource cost is decreasing.

The marginal products of considered inputs change in the expected way i.e. for decreasing inputs the marginal products increase and vice versa. The marginal product of Variable Capital, computed at the geometric mean of input and output is 1.16 €/€ in the first, 1.19 €/€ in the second and 1.11 €/€ in the third model. It is decreasing but at a very slow rate. The marginal product of the second in importance variable, Labour Cost is decreasing from 0.94 €/€ to 0.89 €/€. An increase is observed at the marginal product of Fixed Capital. The marginal product of Fixed Capital is increasing from 6.76 percent in the first model to 8.64 percent in the third model.

One can immediately notice that the results indicate that it is crucial to take the existence of technical inefficiency into account. From the point of view of econometric research, if one ignores differences in technical efficiency among firms, one bias the parameter values obtained in the estimation of the production function. Our suppositions were fully confirmed and this is clear from the aforementioned findings.

**Conclusion**

The combination of the assumptions of the same production function, the same prices, and perfect profit maximisation for all firms invalidates the concept of the production function per se. All firms would produce the same quantity of output and use the same quantity of inputs. However, not all firms have the same entrepreneurial capacity. The usual interpretation of the production function is that, although individual firms attempt to maximise profits, they are not uniformly successful in doing so due to differences in their managerial abilities.

The study of technical efficiency per se has been an important aspect of the study of development, because it quantifies the productive contribution of factors that are not easily amenable to measurement, such as technology and management. The concept of technical efficiency was developed to introduce systematic deviations in the quantities of inputs that firms use and in the quantity of output they produce, while retaining the assumptions of maximising behaviour by firms that face the same product and factor prices.

A non-parametric approach, known as Data Envelopment Analysis (DEA), defines an efficiency frontier and measures the efficiency of each unit relative to that frontier. A useful tool in revealing relationships that remains hidden for parametric approaches and in quantifying differences in efficiency.

Be motivated by the fact that estimates of the parameters of production functions are subject to bias as a result of the presence of technical inefficiency we applied DEA in a sample of 108 farms with the view to reveal the true nature of the production function. In order to do this we fit a translog production function, which is rejected in favor of the
Cobb-Douglas production function for three samples, which were composed according to the results of DEA. The fact that there is significant difference in the allocation of the input resources and its marginal products between the first sample, which is characterized by inefficiency, and the third sample where all farms are relatively technically efficient confirms our initial stand that DEA can be complement to parametric analysis.

Notes
1. Because of the volume of published work on the Cobb-Douglas function, we cannot begin to site all relevant literature on each topic discussed; therefore this article should not be regarded as a review article on Cobb-Douglas production functions.
2. This is contrast to parametric methods. The concept of these methods is to define the function explicitly in order to determine the frontier for an industry.
3. Recently the statistical properties of the DEA estimators have been investigated by several authors (Banker, 1993, Kneip et al., 1996, Korostelev et al., 1995). The approach to remedy this shortcoming of the DEA method is to apply bootstrap techniques in order to obtain measures of statistical precision in the estimates (Simar and Wilson, 2000a, 2000b, Lothgren and Tambour, 1999). Bootstrap (Efron, 1979, Efron and Tibshirani, 1993) is a general method for estimating statistical properties of deterministic parameters.

References


Maddala G.S., Introduction to Econometrics, 2001, West Sussex, John Wiley &Sons Ltd.


