

Technical efficiency of Greek olive growing farms: a robust approach with panel data

S. Kourtesi¹, P. Fousekis^{1*} and A. Polymeros^{2,3}

¹ Department of Economics. Aristotle University. Thessaloniki, Greece

² Department of Statistics. Athens University of Economics and Business. Athens, Greece

³ Farm Accountancy Data Network. Ministry of Rural Development and Food. Athens, Greece

Abstract

The assessment of technical efficiency in the agricultural sector and the influence of exogenous (environmental) variables on the production process has been a major topic of economic research especially for managers and policy makers. The methodological innovation of the present study involves the impact of environmental variables on efficiency and the utilization of panel data for the empirical analysis. This has been pursued using full nonparametric robust frontier techniques (the alpha-quantile estimator) and a panel data set of olive growing farms in Greece from the Farm Accountancy Data Network of the EU. According to the empirical results, the ratio of owned to total land, the ratio of family to total labor, the degree of specialization, and a farm's location have a statistically significant impact on performance, which is not constant but varies over the 2006 to 2009 period considered.

Additional key words: nonparametric estimation; conditional efficiency; environmental variables; agriculture.

Introduction

The assessment of technical efficiency (TE) provides information to managers and to policy makers about differences in performance among production units and the potential for improvements. Economic research on this important topic has evolved largely around two alternative approaches, namely, the parametric and the nonparametric. The first allows for random noise and, as a consequence, for some observations to lie outside the production set while the second assumes that all observations belong to the production set with probability equal to 1. The parametric models require restrictions on the shape of the production frontier (benchmark) and on the underlying data generation process (e.g. Stevenson, 1980; Battese & Coelli, 1988). Therefore, they lack robustness in cases where the functional forms of the frontier and/or the error structure are not correctly specified. The estimation of nonparametric frontier models has been, until recently, pursued through envelopment techniques such as the Data Envelopment

Analysis (DEA) (Charnes *et al.*, 1978) and the Free Disposal Hull (FDH) (Deprins *et al.*, 1984) that are quite appealing since they rely on very few assumptions. They are, however, by construction quite sensitive to outliers or to extreme values. This is certainly an important problem when one is interested in assessing TE of production units in economic activities where the amount of output is subject to random shocks. In farming, for example, the level of realized output can be quite different from the planned one because of weather conditions and pest attacks.

During the last decade considerable research effort has been devoted to the development of robust nonparametric efficiency estimators. Cazals *et al.* (2002) introduced the order- m estimator and Aragon *et al.* (2005) the α -quantile estimator. Both estimators are based on partial frontiers called order- m and α -quantile, respectively. The partial frontiers do not envelop all data points and, thus, they are more robust to extreme values and outliers compared to the FDH and DEA estimators, while keeping their asymptotic properties. Daouia &

* Corresponding author: fousekis@econ.auth.gr

Received: 27-04-13. Accepted: 14-11-13.

Abbreviations used: DEA (Data Envelopment Analysis); DGP (Data Generation Process); DMU (Decision Making Unit); EU (European Union); FADN (Farm Accountancy Data Network); FDH (Free Disposal Hull); LFA (Less Favoured Area); TE (Technical Efficiency).

Simar (2007) extended the α -quantile estimator to allow for production processes involving multiple inputs and outputs. Because of its continuous nature, the α -quantile estimator covers entirely the interior of the attainable set. It has, therefore, from an economic standpoint, an advantage over the discrete order- m estimator (that is, it gives a clearer indication of TE).

Against this background, the present work relies on the nonparametric α -quantile estimator to assess TE of olive growing farms in Greece. The European Union (EU) is by far the most important producer and consumer of olive oil; it offers 73% of the global olive oil supply and it accounts for 66% of global olive oil consumption¹. Greece is the third largest olive oil producer in the world behind Spain and Italy. Olive growing is a major feature of the heritage and socio-cultural life of Mediterranean regions. The cultivation of olive trees constitutes an agricultural activity in the country since the ancient times. Olive trees are well adapted to the soil and climatic conditions of Greece. Moreover, almost 80% of the total labor input required is applied during the winter period making, thus, olive farming a valuable complement to other activities, agricultural or non (Papageorgiou, 1987).

Olive oil prices in Greece have exhibited a clear downward trend over the most recent years. In particular, the wholesale price of olive oil in the country has decreased from about €290/100 kg in 2005 to about €170/100 kg in 2011. Over the same period the price index of intermediate inputs in Greek agriculture has risen from 100 in 2005 to 119 in 2011 (EC, 2012). Those developments have worked towards lower margins not only for olive growing farms but for the agricultural sector as a whole. The index of farm income in Greece has dropped from 100 in 2005 to 78 in 2011 (EC, 2012). If the recent trends prevail in the future as well, many olive growing farms in Greece are likely to face survival problems, unless they manage to improve substantially their productive performance (by taking more output with the same input levels or by reducing inputs levels required to produce a given level of output). It appears, therefore, that the empirical investigation of TE and its determinants for the olive growing farms in Greece is timely and potentially interesting.

It is widely recognized that efficiency estimates *per se* are of limited value unless they are somehow related to the operational environment of each individual production unit. Exogenous (environmental) factors may influence the performance of production units and, thus, they may be the cause of the observed efficiency differentials. To account for heterogeneity of the operational environment, the present work employs the nonparametric conditional efficiency approach which is based on the probabilistic formulation of the production process and it incorporates the operational environment by conditioning on exogenous characteristics (Aragon *et al.*, 2005; Daraio & Simar, 2005, 2006, 2007; Daouia & Simar, 2007; De Witte & Cortelainen, 2013). The conditional efficiency approach dispenses with the separability condition and it does not require any *a priori* specification regarding the impact of exogenous factors².

Robust nonparametric efficiency estimators along with a probabilistic formulation of the production process have been applied to banking, mutual funds, post offices, and education (*e.g.* Daouia & Simar, 2007; Daraio & Simar, 2007; De Witte & Kortelainen, 2013). To the best of knowledge, there has been so far only one application to agricultural sector; that was by Kourtesi *et al.* (2012) on cereal growing farms in Greece. Relative to earlier works, the present one involves two innovations: (a) it offers both a qualitative and a quantitative characterization on the impact of environmental variables on efficiency (it determines the sign as well as the statistical significance of the impact of a given factor); (b) it utilizes panel data for the empirical analysis. It is known that nonparametric efficiency analysis typically relies on cross sectional data even when panel data are available (that means, it proceeds with the assessment of efficiency in each period separately without accounting for the fact that time may affect the performance of individual units either because of technological changes or because of changes in other important factors, *e.g.* weather conditions for the agricultural sector or resource stocks for resource-based industries). The innovations are based on recent developments concerning modeling and statistical inference in nonparametric frameworks (Racine *et al.*, 2006; De Witte & Kortelainen, 2013).

¹ http://ec.europa.eu/agriculture/olive-oil/index_en.htm, Accessed 26/03/2013.

² The separability condition holds when the environmental variables affect only the distribution of the efficiencies and not the range of achievable input-output combinations (shape of the production set). As shown by Simar & Wilson (2007), unless separability is validated, the typically used two-stage approach (in which the first stage estimates of the efficiency of production units are regressed in a second stage on a number of exogenous variables) is not meaningful.

Material and methods

The α -quantile output efficiency measure

Let $X \in R^p$ be the vector of inputs and $Y \in R^q$ be the vector of outputs from a given production process. Let also Ψ be the production set for the process, where Ψ satisfies the assumption of free disposability (e.g. Deprins *et al.*, 1984). As noted by Cazals *et al.* (2002), the production process can be described by the joint probability measure of (X, Y) on $R^p \times R^q$. This joint probability measure is completely characterized by the knowledge of the probability function:

$$H_{XY}(x, y) = \text{prob}(Y \geq y, X \leq x) \quad [1]$$

giving the probability that a decision making unit (DMU) that operates at level (x, y) is dominated; the support of H_{XY} is the production set Ψ . Relation [1] can be expressed as

$$H_{XY}(x, y) = \text{prob}(Y \geq y | X \leq x) \text{prob}(X \leq x) = S_{Y|X}(y|x)F_X(x) \quad [2]$$

where $S_{Y|X}(y|x)$ stands for the (non standard) conditional survival function and $F_X(x)$ for the distribution function of X .

Daouia & Simar (2007) define, for all x such that $F_X(x) > 0$ and for $\alpha \in (0, 1]$, the α -quantile output efficiency score for the unit operating at level (x, y) as:

$$\lambda_\alpha(x, y) = \sup \left\{ \lambda \mid S_{Y|X}(\lambda y | x) > 1 - \alpha \right\} \quad [3]$$

When $\lambda_\alpha(x, y) = 1$ the unit belongs to the α -quantile efficiency frontier. This means that only $(1 - \alpha)100\%$ of the units, using inputs less than (or equal to) x , can dominate the unit in question; a value of $\lambda_\alpha(x, y) > (<) 1$ indicates that a proportional increase (decrease) in outputs is necessary to bring the unit (x, y) on the α -quantile efficiency frontier. On the basis of [3], the α -quantile efficient frontier is the $p \times q$ vector $(x, \lambda_\alpha(x, y)y)$ where $(x, y) \in \Psi$. In the special case where $q = 1$, $y^E(x) = \lambda_\alpha(x, y)y = \phi(x)$ (where E is the expectation operator) stands for the α -quantile production function (Aragon *et al.*, 2005).

Note that for $\alpha = 1$ the α -quantile efficiency score reduces to the Farrell output efficiency score (the FDH deterministic efficiency estimator). The latter estimator, however, envelops all data points and it is very sensitive to extreme values and to outliers (e.g. Pastor *et al.*, 1999). With $\alpha < 1$, the α -quantile stochastic frontier is a partial one. Because of its continuous "trimming" nature, the α -quantile efficiency estimator does not envelope all data points and avoids the problems of

deterministic estimators (FDH and the DEA). As α approaches to 1 and the number of observations in the sample become very large the partial frontier estimator converges to the full frontier sharing the same properties as the FDH estimator (e.g. Aragon *et al.*, 2005; Daouia & Simar, 2007).

For the empirical implementation of [3] with n production units in the sample, Daouia & Simar (2007) propose the following procedure:

Define

$$\rho_i = \min_{l=1, \dots, q} \frac{y_i^l}{y_i^l}, \quad i = 1, \dots, n,$$

and let $N_x = n \hat{H}_{XY,n}(x, 0)$ where $\hat{H}_{XY,n}$ is the nonparametric estimator of the joint probability function

$$\hat{H}_{XY,n}(x, y) = \frac{\sum_{i=1}^n 1(x_i \leq x, y_i \geq y)}{n}$$

and $1(\bullet)$ is an indicator function which takes the value of one if the condition appearing in the parenthesis holds and takes the value of zero, otherwise. For $j = 1, \dots, N_x$, denote by $\rho_{(j)}^x$ the j th order statistic of ρ_i such that $x_i \leq x$: $\rho_{(1)}^x \leq \dots \leq \rho_{(N_x)}^x$. The nonparametric estimate of the α -quantile output efficiency score for the unit operating at level (x, y) is then

$$\hat{\lambda}_{\alpha,n}(x, y) = \begin{cases} \rho_{(\alpha N_x)}^x, & \alpha N_x \in N^* \\ \rho_{([\alpha N_x] + 1)}^x, & \alpha N_x \notin N^* \end{cases} \quad [4]$$

with N^* being the set of positive integers and $[\alpha N_x]$ being the integral part of αN_x .

The conditional α -quantile output measure and the influence of exogenous variables on the production process

Let $Z \in R^r$ a vector of environmental variables which, although exogenous, may influence the probabilistic production process. To account for the operational environment in efficiency estimation with partial stochastic frontiers Cazals *et al.* (2002) and Daraio & Simar (2005) considered the DGP (data generation process) of the random variable (X, Y, Z) and focused on the conditional distribution (X, Y) of for a given value of Z :

$$H_{XY|Z}(x, y | z) = \text{prob}(Y \geq y, X \leq x | Z = z) = S_{Y|X,Z}(y|x, z)F_{X|Z}(x|z) \quad [5]$$

giving the probability that the unit (x, y, z) will be dominated by other units facing exactly the same ope-

rational environment (that means, having $Z = z$). The support of $H_{XY|Z}$ is denoted by Ψ^Z (a set possibly different from the production set Ψ). Daouia & Simar (2007) defined the conditional α -quantile efficiency estimator as:

$$\lambda_\alpha(x, y|z) = \sup \left\{ \lambda \mid S_{Y|X,Z}(\lambda y|x, z) > 1 - \alpha \right\} \quad [6]$$

$\lambda_\alpha(x, y|z)$ has an interpretation analogous to that of $\lambda_\alpha(x, y)$; that is, $(1 - \lambda_\alpha(x, y|z))100\%$ stands for the radial feasible change in all outputs a unit operating at (x, y, z) should perform to reach the efficient boundary of the set Ψ^z . For the empirical implementation of [6] with n production units in the sample, Daouia & Simar (2007) propose the following procedure:

For $j = 1, \dots, N_x$, denote by $Z_{[j]}^x$ the observation Z_i corresponding to the order statistic $\rho_{(j)}^x$ and let

$$R_{X,Z} = \sum_{i=1}^n 1(X_i \leq x) K\left(\frac{z - Z_i}{h_n}\right) > 0, \text{ where } K \text{ is a kernel}$$

with compact support and h_n is the bandwidth of appropriate size. The nonparametric estimate of the conditional α -quantile output efficiency score is then

$$\hat{\lambda}_{\alpha,n}(x, y|z) = \begin{cases} \rho_{(k)}^x, & L_{k+1} \leq 1 - \alpha < L_k \quad (k = 1, \dots, N_x - 1) \\ \rho_{(N_x)}^x, & 0 \leq 1 - \alpha < L_{N_x} \end{cases} \quad [7]$$

with $L_{k+1} = \left(\frac{1}{R_{XZ}}\right) \left(\sum_{j=k+1}^{N_x} K\left(\frac{z - Z_{[j]}^x}{h_n}\right)\right)$.

The influence of exogenous (environmental) variables on the production process can be evaluated by comparing the conditional efficiency measure with the unconditional one (that means the ratio of the radial distances from the conditional and the unconditional frontiers, respectively). Specifically, Daraio & Simar (2005 and 2007) propose the estimation of the following smooth nonparametric regression model

$$\hat{\varphi}_{\alpha,n,i} = g(z_i) + e_i, \quad [8]$$

where:

$$\hat{\varphi}_{\alpha,n,i}(x_i, y_i|z_i) = \frac{\hat{\lambda}_{\alpha,n,i}(x_i, y_i|z_i)}{\hat{\lambda}_{\alpha,n,i}(x_i, y_i)}, \quad i = 1, 2, \dots, n, \quad [9]$$

g is a conditional smooth mean function and e_i is the error term (with $E(e_i|z_i) = 0$). In the output-oriented efficiency and for a univariate and a continuous Z , a horizontal smoothed regression curve, implies that the exogenous factor has no influence whatsoever on the TE; an increasing (decreasing) regression curve

implies that TE rises (falls) with the amount of Z . When an exogenous factor has a favorable impact, it can be viewed as substitute input which augments the productivity of the X inputs. In the opposite case, the presence of Z reduces productivity by entailing more of the X inputs per unit of output. It should be noted that the impact is not necessarily monotonic for all values of Z . An increasing part of the regression may be followed by a decreasing one and the opposite. Therefore, the approach allows for the existence of different impacts locally.

De Witte & Kortelainen (2013) extended the ideas of Daraio & Simar (2005 and 2007) to general Z vectors involving multiple continuous and discrete (both ordered and unordered) environmental factors. Moreover, they developed appropriate tools of statistical inference. With multiple Z factors, the visualization of individual impacts can be achieved through the so-called *partial smooth regression plots* where only one such factor at a time is allowed to change and the rest are kept at fixed values. For instance, the continuous factors are set at the first, the second or the third quartile, while at the same time each discrete factor is set on one of its specific values.

The statistical inference relies on the Local Linear model which allows one to estimate simultaneously both the conditional smooth mean function in [8] as well as the gradient vector associated with the multiple environmental factors. The approach involves the minimization problem

$$\min_{\{\gamma, \delta\}} \sum_{i=1}^n (\hat{\varphi}_i - \gamma - \delta(z_i^c - z^c))^2 K_h(z, z_i) \quad [10]$$

where $z_i = (z_i^c, z_i^d, z_i^u)$ includes the values of the continuous and the discrete (ordered and unordered) environmental variables, γ and δ are local coefficients, and K_h is the generalized product kernel function with appropriate bandwidth; the gradient vector of interest is $\delta(z^c)$. Let $Z_s \in Z$ be the s th component of vector Z and Z_o be the vector of all other environmental variables. Then, the null hypothesis (no influence) is

$$\begin{aligned} E(\varphi|Z_s, Z_o) = E(\varphi|Z_o) &\Rightarrow \frac{\partial E(\varphi|Z_s, Z_o)}{\partial Z_s} = \\ &= 0 \Rightarrow \delta(Z_s) = 0 \quad \text{almost everywhere} \end{aligned} \quad [11]$$

and the alternative is $\delta(Z_s) \neq 0$ (De Witte & Kortelainen, 2013). Details on the estimation of the gradient $\delta(z^c)$ and the associated with it vector of p -values are presented in the De Witte & Kortelainen (2013). The p -values together with the *partial smooth regression*

plots allows a researcher to characterize the sign of the impact of each individual environmental factor on efficiency as well as its statistical significance.

Results

The empirical analysis in this study relies on information from the Farm Accountancy Data Network (FADN) of the EU which is an important tool for agricultural policy analysis and simulation. The panel data for the years 2006-2009 include 519 observations from specialist olive farms located in all regions of Greece³. Farm output (Y) is the total revenue from olive production (measured in euros). The production inputs (X) include: (a) total labor (comprising all family and non family labor and measured in working hours); (b) total land under olive trees (measured in 100 m²); (c) fertilizers and pesticides (measured in euros); and (d) other costs (including electric power, fuel, depreciation, interest and miscellaneous expenses, all measured in euros)⁴. We note that the vector of inputs X employed here is in line with those used in earlier empirical studies on TE of olive growing farms in Greece (e.g. Giannakas *et al.*, 2000; Tzouvelekas, *et al.*, 2001, 2002a; Karagiannis & Tzouvelekas, 2009)⁵.

The environmental/exogenous factors include: (a) the ratio of family to total labor; (b) the ratio of owned to total land; (c) the ratio of irrigated to total land; (d) the degree of specialization; (e) the region at which the farm is located; (f) the type of a farm's location, and (g) the year of observation⁶. The first three variables are continuous while the last four are discrete. The degree of specialization is obtained from the ratio of land under olives trees to total land under crops; in particular this categorical variable takes the value of

2 if the ratio is higher than 0.9 and the value of 1, otherwise⁷. The region involves two categories (1: the farm is located in Macedonia, Thrace or Thessaly (Northern-Central Greece), 2: the farm is located in Epirus, Peloponnese, Ionian Islands, Continental Greece, Aegean Islands or Crete (Southern Greece or Greek Islands). The type of location includes two categories [1: the farm is not located in a less favoured area (LFA); 2: the farm is located in a LFA]⁸. The year of observation involves four categories, one for each of the four years considered (see Table 1). The presence of the year of observation (time dummy) among the environmental variables allows for different frontiers (benchmarks) over time. This makes it possible to utilize fully the information contained in the panel of observations and, at the same time, to avoid a potential contamination of the resulting efficiency estimates due to frontier shifts from one year to the other. We note that because the period considered here involves only four years the discrete time variable is likely to capture frontier shifts because of variations in weather conditions rather than the effect of technical change (which typically takes much more time to have an impact on the production possibilities set). The choice of environmental variables is to a certain extent constrained by data availability. Nevertheless, the same or similar environmental variables have been considered as relevant in almost all earlier empirical studies on the topic (e.g. Tzouvelekas *et al.*, 1997, 2001, 2002a; Zhu *et al.*, 2008).

Table 1 presents descriptive statistics for the variables used in the empirical analysis. The sample includes very small as well as very large olive growing farms (in terms of the land under olive trees as well as the output). Considerable variability also appears to exist with respect to the use of the production inputs. As far

³ FADN considers a farm as a "specialist olive" farm if it obtains more than 2/3 of its revenues from the production of olives.

⁴ All the nominal variables were deflated using price indices (base year 2005) from the Hellenic Statistical Authority.

⁵ As noted Banker *et al.* (2007), the use of aggregate revenue or aggregate cost data in TE analysis may result in estimates reflecting a mix of technical and allocative efficiency. Here, we deal with single output and allocative efficiency from the production side is not an issue. Fertilizers and pesticides as well as "other" inputs are expressed in monetary units to facilitate aggregation. This is a very common practice in empirical analysis of technical efficiency (e.g. Odeck, 2007; Latruffe *et al.*, 2008; Larsen, 2010).

⁶ The variables family to total labor and owned to total land may not be truly exogenous. In the relevant literature the Z variables represent factors which are not input levels but they may impact on the performance of productive units. Here, following the relevant studies by Latruffe *et al.* (2008) and Larsen (2010) those two variables have been included among the Z ones to capture potential effects of a farm's integration in agricultural input markets on efficiency.

⁷ It is known that the convergence rate of nonparametric efficiency estimators decreases with the number of continuous environmental variables (e.g. Li & Racine, 2007; Jeong *et al.*, 2010). The degree of specialization has been treated as a categorical variable in order to: (1) to improve the accuracy of the estimators and (2) allow for a sharp contrast between very high rates of specialization in olive production and lower ones.

⁸ For the definition of LFA in the EU see Articles 18 and 20 of the Regulation EC No 1257/1999.

Table 1. Descriptive statistics

	Minimum	Maximum	Mean	Standard deviation
Output (euros)	516	75,566	14,837.95	11,888
Labor (hours)	213	7,720	2,793.42	1,406
Land (100 m ²)	50	2,840	590.43	396
Fertilizers and pesticides (euros)	5	12,509	1,153.83	1,543
Other costs (euros)	154	30,641	4,414.34	4,395
Ratio of family to total labor	0.28	1	0.87	0.16
Ratio of owned to total land	0	1	0.87	0.25
Ratio of irrigated to total land	0	1	0.29	0.40
Year (1=2006, 2=2007, 3=2008, 4=2009)	1	4	2.64	1.10
Specialization (1 < 0.9, 2 > 0.9)	1	2	1.64	0.48
Region (1 = Northern-Central Greece, 2 = Southern Greece or Islands)	1	2	1.95	0.23
Type of location (1 = not LFA, 2 = LFA)	1	2	1.74	0.44

LFA: less favoured area.

as the environmental/exogenous (Z) variables are concerned, it appears that the cultivation relies mostly on family labor and on family owned land, while the irrigated land represents a small percentage of the total land. Also the majority of the olive growing farms are highly specialized, located in the Southern Greece or in Islands and in LFA. Regarding the year variable, 19.9%, 25.4%, 25.6%, and 29.1% of the total observations come from 2006, 2007, 2008 and 2009 respectively.

To estimate the α -quantile output efficiency scores (both the unconditional and the conditional ones) α has been set equal to 0.991 so that to achieve a level of robustness at about 10%, as suggested by Daraio & Simar (2006 and 2007)⁹. For the conditional efficiency estimation we use the Uniform kernel function for the continuous variables and the Aitchison & Aitken (1976)'s discrete univariate kernel for the categorical ones¹⁰. Following Hall *et al.* (2004), Li & Racine (2007) and Jeong *et al.* (2010), we rely on least squares cross-validation for the bandwidth choice (conditional bandwidth estimation).

Table 2 presents the frequency distributions of the unconditional and the conditional efficiency estimates. The average value of the unconditional efficiency estimates is 1.24 suggesting that output could be increased by 24%, provided that farms in the sample will follow the same rules of input use as the best-practice farms do. The majority (more than 60%) of the efficiency estimates lie in the interval [1, 1.5). Nevertheless, there has been a considerable proportion of farms which can

be classified as “super-efficient” (10.79% with an efficiency estimate below 1) and also a sizable proportion of farms (25.62% with an estimate between 1.5 and 2) which appear to be highly inefficient. Also, 164 farms (or 31.60%) of the total lie on the respective unconditional α -quantile frontiers.

The average value for the conditional efficiency estimates is 1.03. The overwhelming majority of the estimates (98%) lie in the interval [1-1.5); there are no “super-efficient” farms, while the proportion of highly inefficient farms has fallen to about 2%. Also, 474 farms (or 91.33%) of the total lie on the respective conditional α -quantile frontiers. Overall, accounting for the operational environment leads to a much more concentrated distribution of the estimated efficiency scores suggesting that the operating environment does affect the productive performance of the olive growing farms

Table 2. Frequency distribution of the unconditional and the conditional estimates

Efficiency score	Unconditional estimates		Conditional estimates	
	No. of farms	% of farms	No. of farms	% of farms
[0.5-1)	56	10.79	0	0
1	164	31.60	474	91.33
(1-1.25)	96	18.50	17	3.27
[1.25-1.5)	70	13.49	18	3.47
[1.5-2)	133	25.62	10	1.93

⁹ The level of robustness refers to the percentage of the sample points to be left outside of the partial frontier (that means the percentage of “super-efficient” units in the sample).

¹⁰ As noted by Daraio & Simar (2005) only kernels with compact support can be employed for continuous variables.

in Greece¹¹. There is a strong indication, therefore, that comparing the performance of olive growing farms in Greece on the basis of the unconditional efficiency scores is certainly unfair. The conditional efficiency scores derived when the environmental variables are included in the analysis provide a much more objective picture of the existing differentials and the potential for improvements. It is very likely that the same result will hold for other countries and other agricultural activities where productive units operate under different environments.

For the nonparametric estimation of the Local Linear model we employ again the Uniform kernel function for the continuous variables and the Aitchison & Aitken (1976)'s discrete univariate kernel function for the categorical ones¹². For bandwidth choice we also use the least-squares cross-validation method. Figs. 1 to 2 present partial smooth regression plots visualizing the impact of each individual environmental/exogenous factor on the performance of farms in the sample. The associated with them Local Linear models have been estimated setting the rest or the continuous factors at their 50 quantile value (a choice

typically made in earlier applications of robust non-parametric efficiency estimators, e.g. De Witte & Kortelainen, 2013).

Fig. 1a indicates that the ratio of family to total labor has an unfavorable impact on a farm's productive performance. The same is true for the ratio of owned to total land (Fig. 1b). However, for the ratio of irrigated to total land it doesn't come out any particular pattern regarding the impact (Fig. 1c). A positive impact of a continuous exogenous factor means that the larger the value of that factor, the more the unconditional efficiency score will benefit from it. For unordered discrete variables (like the variable region) one cannot give a similar interpretation as categories have no natural ordering. Nevertheless, partial regression plots can still be used to obtain an indication with regard to which category is better for productive performance. Next for the discrete environmental factors, we observe (Fig. 2a) that the value of ϕ corresponding to 2 (year 2007) is above all other values, followed by the value of ϕ corresponding to 4 (year 2009); the value of ϕ corresponding to 1 (year 2006) is only above value of ϕ corresponding to 3 (year 2008). From Fig. 2b follows

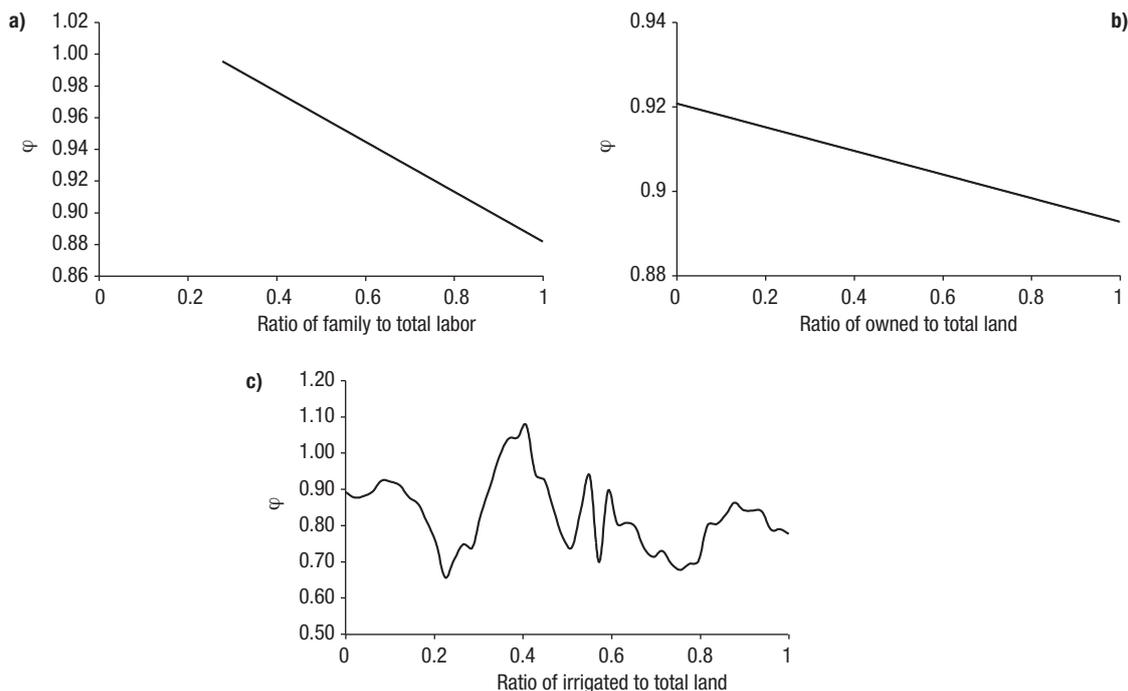


Figure 1. Partial regression plots of the impact of continuous environmental variables: (a) impact of family to total labor, (b) impact of owned to total land and (c) impact of irrigated to total land.

¹¹ From the two-sided Kolmogorov-Smirnov test as well as the Wilcoxon test with p -values $< 2.2 \cdot 10^{-16}$, we reject the null hypothesis that the two distributions (unconditional and conditional) coincide.

¹² All computations have been carried out in *R*. The code utilizes *np* package by Hayfield & Racine (2008).

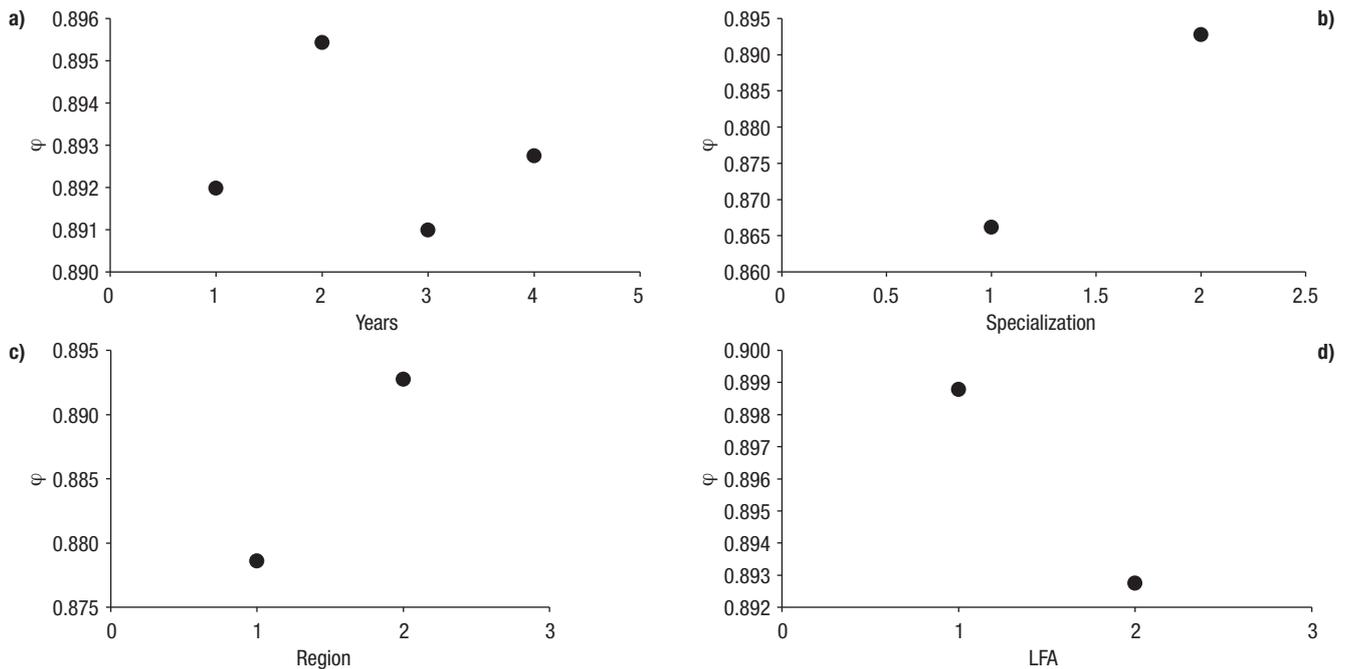


Figure 2. Partial regression plots of the impact of discrete environmental variables: (a) impact of years, (b) impact of specialization, (c) impact of region, and (d) impact of type of location, where LFA is a less favoured area.

that a very high (above 0.9) degree of specialization has a favorable impact on productive performance. We also observe (Fig. 2c) that the olive growing farms located in the Southern Greece or in the Islands outperform the farms located in the Northern-Central Greece. Finally, Fig. 2d suggests that farms that are not located in LFA outperform, *ceteris paribus*, the farms located in LFA.

As mentioned above, the information from the partial regression plots together with the *p*-values from the testing of the null hypotheses of no influence can be used to characterize the impact of environmental/exogenous variables and whether it is significantly different from zero. Table 3 presents the results. From the three continuous environmental variables considered, two (the ratio of family to total labor and the ratio of owned to total land) are statistically significant at 5% and 1% level respectively, while all the four categorical environmental variables are statistically significant at 5% level (year, region and type of location) or 1% level (specialization).

Discussion

The measurement and the explanation of efficiency differentials among decision making units has been an important topic of economic research over the last 40

years and it has been pursued using alternative methodologies. This is not accidental, since the efficiency analysis provides valuable information to managers and policy makers regarding the productive performance across a sample of units and the potential for improvements. In this context, the present work investigates the performance of olive growing farms in Greece over 2006 to 2009 using recently developed fully nonparametric robust partial frontier techniques (the α -quantile estimator) which allow for the inclusion of mixed (both continuous and discrete) environmental/exogenous variables.

According to our results the unconditional estimates indicate considerable efficiency differentiation among the farms in the sample. However, much of this differentiation disappears once the operational environment is accounted for. Out of the seven environmental/exogenous factors considered in the present study, six are statistically significant at the conventional levels. The ratio of family to total labor and the ratio of owned to total land appear to have a negative impact on efficiency. The performance of farms located in LFAs is inferior and the same holds true for farms located in the Northern-Central part of the country. Specialization on olive growing results in higher efficiency. Also the time variable included in the empirical analysis turned out to be statistically significant.

Table 3. Nonparametric significance tests

	<i>p</i> -value	Impact as revealed from the partial regression plot	Conclusion (using the <i>p</i> -value and the evidence from the partial regression plot)
Ratio of family to total labor	0.038 **	Unfavorable	Negative and statistically significant effect of the ratio of family to total labor on productive performance
Ratio of owned to total land	0.009 *	Unfavorable	Negative and statistically significant effect of the ratio of owned to total land on productive performance
Ratio of irrigated to total land	0.881	Not a specific pattern	Not statistically significant effect
Year	0.012 **	Year 2007 is favorable	Positive and statistically significant effect of the year 2007
Specialization	0.006 *	Degree of specialization > 0.9 is favorable	Positive and statistically significant effect of a very high degree of specialization on productive performance
Region	0.037 **	Location in Southern Greece or islands is favorable	Positive and statistically significant effect of a Southern or Island region on productive performance
Type of location	0.011 **	Location not in LFA is favorable	Positive and statistically significant effect of a not LFA region on productive performance

***: statistically significant at the 1% and 5% level, respectively. LFA: less favoured area.

We note that there is a number of earlier works on TE regarding olive growing farms in Greece, all based on the parametric stochastic frontier approach. It is, therefore, interesting to compare their results to those from the robust nonparametric α -quantile estimator (especially with respect to the influence of certain environmental factors on TE).

Tzouvelekas *et al.* (2002a) applied a parametric stochastic frontier model to a sample of olive growing farms in Greece. They found that the ratio of family to total labor has a negative effect on efficiency. Exactly the same result has been obtained by Zhu *et al.* (2008) who also applied a parametric model with non monotonic inefficiency effects to FADN data of Greek olive farms in the period 1995-2004. The study of Tzouvelekas *et al.* (2001) inferred a similar conclusion analyzing the TE of organic and conventional Greek olive growing farms using a stochastic production frontier methodology and a translog functional specification. Tzouvelekas *et al.* (1997), using again parametric methods, reported a negative but not statistically significant impact of the ratio of family to total labor on productive performance. The above results are certainly in agreement with those of the present study. The negative and statistically significant impact of this particular variable provides an indication that traditional, family-farming practices

in olive growing are less efficient than practices depending more on hired labor. It appears, therefore, that contractual agreements give adequate incentives to hired labor and/or that family labor tends to be less experienced/efficient than the hired one.

The impact of the share of owned to total land was found to be negative in Zhu *et al.* (2008) something which is in line with the results here. Tzouvelekas *et al.* (2002b), however, reported a positive impact. A positive relationship between the attained efficiency and the share of owned to total land can be attributed to agency problems between land owners and land lenders (Tzouvelekas *et al.*, 2002b; Latruffe *et al.*, 2008). Specifically, short-run contracts accompanied with an upfront payment may induce renters to “mine” the soil degrading its quality and reducing, thus, its productivity. A negative relationship, on the other hand, suggests the land lenders use inputs more efficiency to cover all operating expenses (including land rent). As noted by Gavian & Ehui (1999) agency problems may be mitigated through long-run contracts, monitoring, collateral pledges by lenders as well as by reputation effects.

With regard to the impact of irrigation, note that there are two main processes of olive growing in Greece: (1) the traditional, found in mountainous or in hilly areas which typically involves no irrigation and (2) the

modern one, which relies intensively on irrigation and on mechanization. There are few olive tree varieties which grow in marginal regions (mountainous or hilly areas; poor and stony soil) and cannot be replaced by other crops. The finding that irrigation has not a statistically significant impact is probably related to the fact that the FADN data includes a varying mix of traditional and modern process.

High degree of specialization has a positive and statistically significant impact on productive performance (Fig. 2b). Tzouvelekas *et al.* (1997) and Zhu *et al.* (2008), report a negative impact (although in the latter study the impact was not significant). Our result appears to be reasonable since specialization implies greater focus on a given agricultural activity. Nevertheless, the benefits of specialization in terms of higher performance should be weighed against those of production diversification. The agricultural economics literature offers substantial empirical evidence that farmers are risk averse (*e.g.* Sckokai & Moro, 2006). Given that agriculture is a risky business, production diversification may be a more appropriate strategy for Greek olive growers.

The performance of farms located in Southern Greece and in Greek islands is superior to those located in the Northern part of the country. This again appears to be reasonable given that in Crete and in Peloponnese lie the largest and the most suited areas for the olive tree growing. Farms located in a LFA tend to be less efficient compared to those that are not. The same result has been reported by Tzouvelekas *et al.* (1997) and by Zhu *et al.* (2008). This is an interesting finding if one takes into account that the large majority of olive growing farms in Greece (74%) is located in LFA (with the percentages of LFAs in Central-North and Southern Greece being about the same). Figs. 2c and 2d, taken together, appear to provide evidence that olive growing in Greece would benefit from further concentration into regions and locations offering natural comparative advantage. We note, however, that there is another interpretation of our empirical result which leads to a different policy recommendation. That is, farmers in the LFAs should be compensated for their natural disadvantage. The implementation of such a policy requires of course that the society and the policy makers feel that continuation of olive farming in the LFAs of Greece is worthwhile.

Finally, the statistical significance of the time variable justifies the use of panel data instead of four separate cross sections in the empirical analysis of effi-

ciency. It turns out that efficiency in olive growing farms in Greece is not constant but it may change from one period to another. The number of years is definitely too short to draw any general conclusions about trends in efficiency and to relate them with changes in the Common Agricultural Policy. Such an analysis would be beyond the scope of this work.

This research implies a theoretically consistent way to incorporate time effects in modeling technical efficiency with nonparametric methods (robust or traditional). Given that the standard nonparametric approaches typically rely on the analysis of otherwise unrelated cross-section samples, the approach utilized here appears to have a very distinct advantage; it brings a number of time periods together in a theoretically consistent way and, provided that an adequate number of periods is available, it allows one to determine policy relevant trends in performance over time. Nevertheless, the extension of the time period considered and the consequent results may provide guidelines for the policy planning and/or the evaluation of applied programmes.

Acknowledgements

We are grateful to the Greek FADN for providing the data set.

References

- Aitchison J, Aitken CBB, 1976. Multivariate binary discrimination by kernel method. *Biom* 63: 413-420.
- Aragon Y, Daouia A, Thomas-Agnan C, 2005. Nonparametric frontier estimation: a conditional quantile-based approach. *Econom Theory* 21: 358-389.
- Banker R, Chang H, Natarajan R, 2007. Estimating DEA technical and allocative inefficiency using aggregate cost or revenue data. *J Prod Anal* 27: 115-121.
- Battese GE, Coelli TJ, 1988. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *J Econom* 3: 387-399.
- Cazals C, Florens JP, Simar L, 2002. Nonparametric frontier estimation: a robust approach. *J Econom* 1: 1-25.
- Charnes A, Cooper WW, Rhodes E, 1978. Measuring the efficiency of decision making units. *Eur J Oper Res* 2: 429-444.
- Daouia D, Simar L, 2007. Nonparametric efficiency analysis: a multivariate conditional quantile approach. *J Econom* 140: 375-400.
- Daraio C, Simar L, 2005. Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *J Product Anal* 24(1): 93-121.

- Daraio C, Simar L, 2006. A robust nonparametric approach to evaluate and explain the performance of mutual funds. *Eur J Oper Res* 175: 516-542.
- Daraio C, Simar L, 2007. Advanced robust and nonparametric methods in efficiency analysis, Methodology and applications, Series: Studies in Productivity and Efficiency. Springer.
- De Witte K, Kortelainen M, 2013. What explains performance of students in a heterogeneous environment? Conditional efficiency estimation with continuous and discrete environmental variables. *Appl Econ* 45(17): 2401-2412.
- Deprins D, Simar L, Tulkens H, 1984. Measuring labor efficiency in post offices. In: The performance of public enterprises: concepts and measurements (Marchand M, Pestieau P, Tulkens H, eds). North-Holland. pp: 243-267.
- EC, 2012. Member states factsheets-Greece. European Commission, Agricultural Policy Perspectives, May 2012.
- Gavian S, Ehui S, 1999. Measuring the production efficiency of alternative land tenure contracts in a mixed crop-livestock system in Ethiopia. *Agric Econ* 20: 37-49.
- Giannakas K, Tran KC, Tzouvelekas V, 2000. Efficiency, technological change and output growth in Greek olive growing farms: a box-cox approach. *Appl Econ* 32: 909-916.
- Hall P, Racine JS, Li Q, 2004. Cross-validation and the estimation of conditional probability densities. *J Am Stat Assoc* 99(486): 1015-1026.
- Hayfield T, Racine JS, 2008. Nonparametric econometrics: the np package. *J Stat Soft* 27(5): 1-32.
- Jeong S, Park B, Simar L, 2010. Nonparametric conditional efficiency measures: asymptotic properties. *Ann Oper Res* 173: 105-122.
- Karagiannis G, Tzouvelekas V, 2009. Measuring technical efficiency in the stochastic varying coefficient frontier model. *Agric Econ* 40: 389-396.
- Kourtesi S, Fousekis P, Polymeros A, 2012. Conditional efficiency estimation with environmental variables: evidence from Greek cereal farms. *Sci Bull-Econ Sci* 11: 43-52.
- Larsen K, 2010. Effects of machinery-sharing arrangements on farm efficiency: evidence from Sweden. *Agric Econ* 41: 497-506.
- Latruffe L, Davidova S, Balcombe K, 2008. Application of a double bootstrap to investigation of determinants of technical efficiency of farms in Central Europe. *J Product Anal* 29: 183-191.
- Li Q, Racine JS, 2007. Nonparametric econometrics: theory and practice. Princeton Univ. Press. Princeton, NJ, USA.
- Odeck J, 2007. Measuring technical efficiency and productivity growth: a comparison of SFA and DEA on Norwegian grain production data. *Appl Econ* 39: 2617-2630.
- Papageorgiou LC, 1987. The role of the olive tree in Greece. *Olivae* 19: 7-11.
- Pastor J, Ruiz J, Sirvent I, 1999. A statistical test for detecting influential observations in DEA. *Eur J Oper Res* 115: 542-554.
- Racine JS, Hart J, Li Q, 2006. Testing the significance of categorical predictor variables in nonparametric regression models. *Econom Rev* 25(4): 523-544.
- Sckokai P, Moro D, 2006. Modelling the reforms of the common agricultural policy for arable crops under uncertainty. *Am J Agric Econ* 88: 43-56.
- Simar L, Wilson P, 2007. Estimation and inference in two-stage, semi-parametric models of production processes. *J Econom* 136(1): 31-64.
- Stevenson RE, 1980. Likelihood functions for generalized stochastic frontier estimation. *J Econom* 13(1): 57-66.
- Tzouvelekas V, Giannakas K, Midmore P, Mattas K, 1997. Technical efficiency measures for olive-growing farms in Crete, Greece. *Int Adv Econ Res* 3(2): 154-170.
- Tzouvelekas V, Pantzios CJ, Fotopoulos C, 2001. Technical efficiency of alternative farming systems: the case Greek organic and conventional olive-growing farms. *Food Policy* 29: 546-569.
- Tzouvelekas V, Pantzios CJ, Fotopoulos C 2002a. Empirical evidence of technical efficiency levels in Greek organic and conventional farms. *Agric Econ Rev* 3: 49-60.
- Tzouvelekas V, Pantzios CJ, Fotopoulos C, 2002b. Measuring multiple and single factor technical efficiency in organic farming. The case of Greek wheat farms. *Br Food J* 104(8): 591-609.
- Zhu X, Karagiannis G, Lansink AO, 2008. Analysing the impact of direct subsidies on the performance of the Greek olive farms with a non-monotonic efficiency effects model. Proc 12th Cong Eur Assoc Agric Econ, EAAE, Ghent, Belgium, Aug 26-29, Paper No 43612.