Applying regression quantiles to farm efficiency estimation

Eleni A. Kaditi and Elisavet I. Nitsi

Centre of Planning and Economic Research (KEPE)

Amerikis 11, 106 72 Athens, Greece

kaditi@kepe.gr ; nitsi@kepe.gr

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association 2010 AAEA,CAES, & WAEA Joint Annual Meeting, Denver, Colorado, July 25-27, 2010

Applying regression quantiles to farm efficiency estimation

Eleni A. Kaditi and Elisavet I. Nitsi

Centre of Planning and Economic Research (KEPE)

Amerikis 11, 106 72 Athens, Greece

kaditi@kepe.gr ; nitsi@kepe.gr

Abstract

This article is concerned with the methodological question of frontier production functions estimation for agriculture, and the appropriateness of regression quantiles, as a useful semi-parametric approach. Better insights are reached using the proposed methodology that provides robust farm efficiency scores estimates. Using the 2007 Farm Accountancy Data Network (FADN) data for Greece, analysis shows that the distribution of efficiency scores is closer to normality when employing regression quantiles, while underestimation of efficiency obtained by other parametric or deterministic methods based on the conditional mean can be avoided. The results further suggest that government support aimed at enhancing farms viability should be directed towards payments decoupled from output or prices, as well as rural development payments that affect productivity in a uniform way.

Keywords: Efficiency, Quantile Regression, Agriculture

JEL codes: C14, D24, Q18

Applying regression quantiles to farm efficiency estimation

Eleni A. Kaditi and Elisavet I. Nitsi

Efficiency measurement is a topic of continuing interest to agricultural researchers and policy-makers, who aim to allocate effectively decreasing agricultural funds across heterogeneous farmers and maintain an adequate standard of living in rural communities. This article is concerned with the methodological question of frontier production functions estimation for agriculture, and the appropriateness of regression quantiles, as a useful semi-parametric approach that provides robust farm efficiency scores estimates.

In the economics literature, two approaches have been widely used to estimate efficiency, the non-parametric data envelopment analysis (DEA) and the stochastic frontier analysis (SFA). DEA has been developed since Charnes *et al.* (1978) and Färe *et al.* (1985) provided measures of efficiency in production, based on the work of Debreu (1951) and Farell (1957) that makes no assumptions about the functional form of the frontier model and the errors distribution. In contrast, Aigner *et al.* (1977) and Meeusen and van den Broeck (1977) proposed the SFA approach that uses maximum likelihood to estimate the production frontier and two random terms; inefficiency and the standard normal error. Both methodologies have been criticized. DEA for the hull that it maps out, as it could be affected to a significant degree by the presence of random disturbances in the data, while SFA makes assumptions for the functional form of the inefficiency distribution and is sensitive towards outliers, raising the possibility of misspecification. Nevertheless, these approaches have been extensively used to estimate farm efficiency (e.g. Coelli and Prasada Rao, 2005; Wadud and White, 2000).

In the current analysis, a first attempt is made to employ regression quantiles as a potential alternative approach to estimate efficiency scores in agriculture. Quantile regression was developed by Koenker and Bassett (1978) and it provides a description of a response variable as a conditional function of a set of covariates broader than the methods based on conditional means (i.e. ordinary least squares or maximum likelihood). This approach requires an assumption about the functional form of the frontier, while it does not require the imposition of a particular form on the distribution of the inefficiency term as in SFA. It also avoids the criticism aimed at DEA, a pure deterministic approach that does not allow for random error in the observed values of the dependent variable, as despite the recently developed bootstrap techniques employed to analyze the sensitivity of DEA efficiency estimates and obtain confidence intervals (Wilson, 1995; Simar and Wilson, 2000), it allows observations to lie above the fitted curve as a result of pure chance, requiring that a functional form is fitted. In addition, the proposed approach is very robust compared to conditional mean regression against outliers. Quantile regression functions are also especially useful in the case of heteroskedasticity. As farm level data typically display considerable heterogeneity (Kaditi and Nitsi, 2009), quantile regression is especially suited for empirical efficiency analysis.

A two-stage approach is used in this framework, employing quantile regression in both stages. In the 1st-stage, the estimated efficiency scores are computed, while in the 2nd-stage, these scores are regressed over a set of covariates, including policy measures and farm characteristics at different points of the conditional efficiency distribution. For reasons of comparison, stochastic frontier techniques, data envelopment analysis and least squares are applied in the respective stages. Farm level data is retrieved from the Farm Accountancy Data Network (FADN) dataset for Greece for 2007.[1]

Quantile Production Function

Quantile regression estimators are robust to deviations from distributional hypotheses, which is an appealing characteristic in the production function context because of the asymmetric distribution of the stochastic error. The efficient production frontier is estimated by a quantile regression of high percentile, which essentially describes the production process as the obtained regression parameters display the 'optimal' technique used by the most efficient farms, i.e. farms representing the efficient production frontier. Efficiency estimates for all farms are actually derived by using the obtained coefficients and comparing each farm's factual output with its potential output using the 'optimal' technique.

To estimate the production function, cross sectional data for n farms is assumed indexed by i ($i = 1,...,n$) using k different inputs contained in the input vector x_i' to produce a single output y_i. The conditional τ^{th} quantile of y ($\tau \in [0,1]$), given a covariate vector x', can be computed employing the conditional quantile function denoted linearly in logarithms by:

$$Q_{\ln y}(\tau|x) = \beta(\tau)\ln x' \qquad (1)$$

whereas the estimator $\hat{\beta}(\tau)$ can be obtained as the solution of the minimization problem:

$$\min_{\beta \in \Re^p} \sum_{i=1}^{n} \rho_\tau \left(\ln y_i - \beta(\tau)\ln x_i'\right) \qquad (2)$$

Assuming a linear relationship between $\ln x$ and $\ln y$:

$$\ln y' = \beta_0 + \beta(\tau)\ln x' + u_i \qquad (3)$$

the conditional quantile becomes:

$$Q_{\ln y}(\tau|x) = \beta_0 + \beta(\tau)\ln x' + F_u^{-1}(\tau) = \left[\beta_0 + F_u^{-1}(\tau)\right] + \beta(\tau)\ln x' \qquad (4)$$

where $F_u^{-1}(\tau)$ is the quantile of the error term distribution.

Some arbitrariness remains in terms of the choice of τ for the estimation of the production frontier, as quantiles differentiation depends on the size of the sample and the

[1] Source: "EU-FADN – DG AGRI L-3".

amount of information it contains about the upper tail of the conditional distribution (Koenker, 2005). One might conjecture that the higher the number of observations, the higher the quantile τ can be chosen. As further explained below, it seems evident that the analysis should focus on the top quantiles, as these percentiles represent the production frontier in the upper tail of the conditional distribution where 'best-practice' farms are operating.

To estimate the production function in agriculture, a multi-input-one-output model is further employed, signifying the appropriateness of the quantile regression approach. The inputs included are *capital*, measured as the value of total assets, *labor*, denoted by the number of working hours, *land* expressed in hectares, and *intermediates* measured as the value of various expenses per farm. Data for 2007 were retrieved from the FADN dataset for Greece, which includes physical, structural, economic and financial data for 4 014 farms.

Summary statistics are presented in Table 1. On average, Greek farms' output values about €30 000. The average size is about 12 Ha, whereas the operator, family-members and hired-staff work for about 3 200 hours a year. The second column provides the mean obtained from the FADN standard results database. The extrapolated data from the sample to all farms in Greece covered by the survey have been obtained by a special weighting system where each farm in the sample has a weight corresponding to the number of agricultural holdings it represents. As a result, the FADN mean shows high deviations from the sample mean for both the output and all inputs, though the figures are close to the sample median. This characteristic of the sample provides an additional argument in favor of the use of regression quantiles, which is more indicative, as the effect of the covariates on the conditional median is estimated rather than the mean of output.

Table 1. Descriptive statistics, 2007

	Mean	***Mean****	***Median***	***SD***	***Min***	***Max***
Production, €	29 687	19 176	22 183	29 424	582	469 159
Capital, €	104 463	78 576	81 735	85 213	730	875 508
Labor, hours	3 206	2 693	2 810	2 014	506	22 560
Land, Ha	12.14	7.04	7.20	14.95	0.1	180
Intermediates, €	12 537	7 691	8 068	14 313	226	212 730

*: FADN Public Database.

In this framework, a simple Cobb-Douglas production function is estimated in logs with the use of quantile regression:

$$\ln y_i = \beta_0 + \beta_1 \ln x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i \tag{5}$$

where u is the iid error term.

Thirty-nine distinct quantile regression estimates, that is a whole spectrum of production functions corresponding to different quantiles of the conditional distributions of output given inputs, are presented for a (horizontal) quantile scale ranging from 0.025 to 0.975 as the solid curve with filled dots (Figure 1). The shaded grey area depicts a 90 percent point-wise confidence band for the quantile regression estimates that were obtained by bootstrapping with 2 000 sample replications. The dotted line in each figure shows the least squares estimate of the conditional mean effect, whereas the two dashed lines represent conventional 90 percent confidence intervals for the latter estimate. The coefficients describing the impact of labor and capital on production have an upward trend along the output distribution, with some exceptions. A considerable dispersion is observed for the intermediates at different quantiles of the distribution, as the estimate at the 0.025 quantile is around 0.651, whereas it reaches 0.263 when evaluated at quantile 0.975 indicating a negative relationship. Quantile regression estimates suggest also a positive relationship between land and output, although this relationship becomes statistically significant only for point estimates above the 0.80 quantile. Finally, it is obvious that in all cases results from OLS estimates would lead to simplistic and false conclusions.

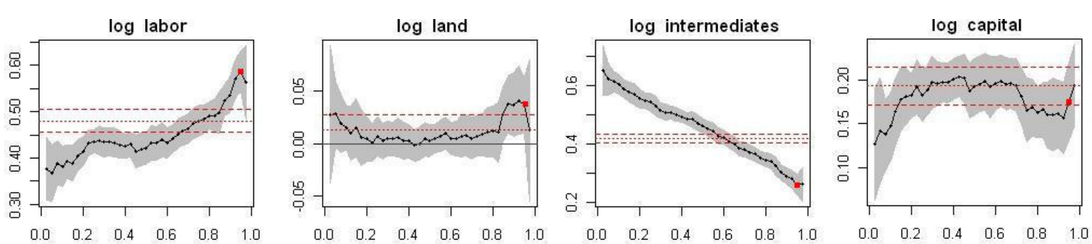

Figure 1. OLS and Quantile regression estimates

The importance of the differences in the quantile parameter estimates was formally examined with the relevant hypotheses testing. The corresponding test statistics for the pure location shift hypothesis and the location-scale shift hypothesis proposed by Khmaladze (1981) and Koenker and Xiao (2002) were performed. Two tests were computed for each hypothesis; a joint test that all covariates effects satisfy the null hypothesis that all the conditional quantile production functions have the same slope parameters, and a coefficient-by-coefficient version of the test. Both tests were decisively rejected (with values 21.97 and 16.26, respectively). The effects of the coefficient-by-coefficient tests are also highly significant.

Having produced a family of production functions, the attention should now be drawn on the particular segment of the conditional distribution that can reflect the production frontier. The choice of the appropriate τ for the estimation of the production frontier focuses on the top quantiles, i.e. $\tau \geq 0.95$. Figure 2 illustrates the estimated efficiency frontier for such quantiles. Using equation (5), it is examined whether farm i belongs to the quantile curve of order τ_i. In particular, the order of the quantile frontier indicates that farm i produces more than (100τ)% of all farms using inputs smaller or equal to x_i and produces less than the 100(1- τ)% remaining farms (Aragon *et al.*, 2005; Daouia and Simar, 2007). If τ_i is close to one, then the farm (x_i, y_i) can be seen to be performing relatively efficiently. As the order of the quantile frontier increases, the number of outliers reduces, whereas farm i denoted by a filled-square becomes relatively inefficient. That is, the number of observations above the quantile estimates $\widehat{q}_{\tau,n}$ decreases with τ. However, given the large sample of farms, the number of observations above the quantile frontier $\widehat{q}_{\tau=0.95,n}$ remains large, while it is very small at $\widehat{q}_{\tau=0.99,n}$. An illustration is given by farm i, which lies above the $\widehat{q}_{\tau=0.95,n}$ frontier, but below the $\widehat{q}_{\tau=0.99,n}$. This indicates that the empirical quantile frontier $\widehat{q}_{\tau=0.975,n}$ defines a reasonable benchmark value, so that $\tau = 0.975$ is chosen for the present analysis.

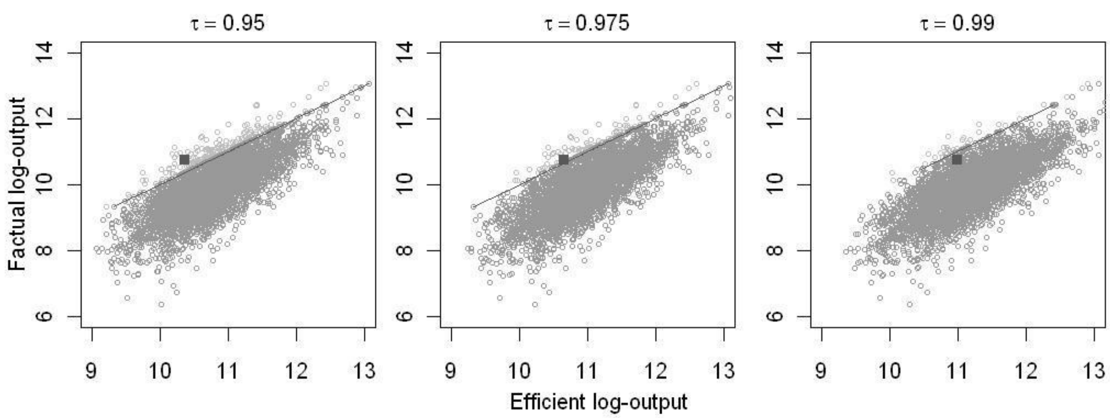

Figure 2. Estimated efficiency frontiers for different τ

Quantile Frontier Model and Efficiency Scores

As $\tau = 0.975$ has been chosen for defining the benchmark farms, the estimated elasticities for the quantile regression model appear in Table 2. For reasons of comparison, a maximum likelihood estimator (MLE) is also performed using equation (5) for the SFA, presuming that *u* is composed of a two-sided stochastic term that accounts for statistical noise and a nonnegative term representing the inefficiency component.[2]

[2] That is: $u_i = \varepsilon_i + v_i$, where $\varepsilon_i \overset{iid}{\sim} N\left(0, \sigma_\varepsilon^2\right)$ and $v_i \overset{iid}{\sim} N^+\left(0, \sigma_v^2\right)$.

Using quantile regression, the statistical significance of input coefficients are consistent with the results found using the stochastic frontier approach. The estimations for capital and land are very similar, though only the former appears to be statistically significant. Labor elasticity exceeds the remaining in both cases, whereas the estimated coefficient for intermediates is much lower in the quantile regression. The elasticities add up to 1.03 and 1.1 for the quantile regression and SFA. That is, the returns to scale for agriculture in Greece are just greater than constant.

Table 2. Estimates of production frontier models

	Quantile regression ($\tau = 0.975$)			***SFA***		
	Estimate	***Std. error***	***p-value***	***Estimate***	***Std. error***	***p-value***
Capital (x_{1i})	0.194	0.033	0.000	0.193	0.013	0.000
Labor (x_{2i})	0.563	0.047	0.000	0.482	0.015	0.000
Land (x_{3i})	0.013	0.042	0.762	0.013	0.009	0.133
Intermediates (x_{4i})	0.263	0.032	0.000	0.413	0.011	0.000
Intercept	1.984	0.461	0.000	0.480	0.171	0.004

To demonstrate the quantile regression and SFA frontier estimation, the relations between efficient and factual outputs obtained by both methods are illustrated in Figure 3.[3] The estimated efficiency frontier of the 0.95 quantile regression is also plotted[4]. As the data contains outliers, the quantile regression appears less sensitive to extreme values. On the contrary, the SFA approach is sensitive to large observations in the output direction. The efficient output produced by SFA is more spread out leading to an underestimation of efficiency, given that the maximum likelihood estimation is based on the conditional mean and as such it does not take into account the possible difference in the production technology of the most efficient farms in the upper tail of the output distribution, being possibly identified even as outliers by the SFA estimation.

---

[3] DEA is not included as it is a pure deterministic approach that does not allow for random error in the observed values and as a result the efficient output cannot be calculated.
[4] The corresponding SFA frontier is not shown given that the estimated efficiency scores does not produce fully efficient farms, i.e. on the frontier.

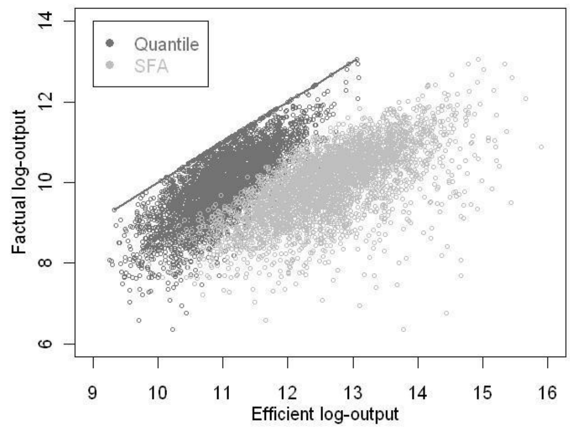

Figure 3. Estimated efficiency frontiers for quantile regression and SFA

Comparing efficiency estimates in Table 3, the average efficiency score in the quantile regression model is 90.4%, that is higher than the one obtained in the stochastic frontier model and the data envelopment analysis. In the former case, efficiency score is 78.9%, whereas in the latter it is 71.4%. The correlation of efficiency scores obtained from the three different approaches is also examined. The Spearman's Rho nonparametric rank statistic show high correlation coefficient between the efficiency scores obtained from the quantile regression and the SFA model, i.e. 0.94 ($p = 0.000$). The two regression methods are therefore in accord when scoring inefficiency of individual farms in the sample. The correlation between the efficiency scores produced by DEA and both quantile regression and SFA is also high but negative (-0.92 and -0.88, respectively).

Table 3. Efficiency scores

	Mean	***Median***	***SD***	***Min***	***Max***
SFA	0.789	0.795	0.052	0.462	0.902
DEA	0.714	0.748	0.161	0.016	0.983
Quantile regression	0.904	0.908	0.051	0.623	1.000

The D'Agostino *et al.* (1990) normality test is, finally, used to show statistically (at the 1% level of significance) that the distribution of the efficiency scores obtained by DEA and SFA methods is negatively skewed and kurtic (i.e. the skewness is -23.795 and -21.721, while the kurtosis is 10.850 and 14.828, respectively). These results suggest that the distribution of the dependent variable significantly departs from normality implying considerable heterogeneity in farm level data and justifying the use of quantile regression. This also becomes apparent by the results of the normality test on the efficiency scores obtained by the estimation of the production frontier via quantile regression. Both skewness and kustosis were found much lower (skewness = -15.363, kurtosis = 7.661), though there still exists some deviation from normality, allowing the use of quantile regression approach in the 2nd stage of the analysis (Figure 4).

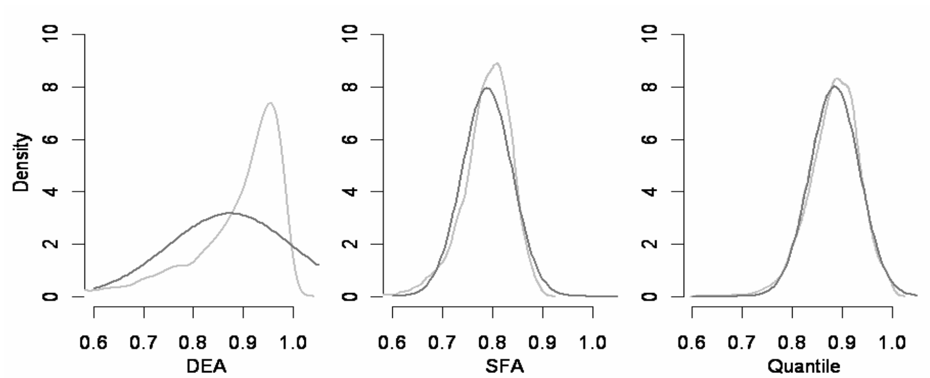

Figure 4. Efficiency scores distributions

Quantile Regression Estimates

The efficiency scores computed in the 1st-stage are now regressed using a number of covariates suggested in the literature. Government policies are distinguished between *Decoupled payments*, *Rural development payments* and *Other payments*, and they are expressed as the share of each category in the total farm revenue. The *Farm size* is measured by a dummy derived from each farmer's European Size Unit (ESU). Nine different economic size classes are essentially used based on the classification provided by FADN. Two variables are included regarding the technology employed. The capital to labor ratio is used as a first proxy of farm *Technology*, whereas the ratio of *Unpaid labor* hours to total farm labor hours indicates the workforce composition. Financial information concerning each farm is also included using the share of *Owned land* in the total land operated. To capture differences in farming practices among farms producing different types of output, a binary variable that equals one is introduced, if a farm is producing mainly livestock and zero otherwise (*Specialization*). The *Age* of the farm's operator, as well as regional dummies are also included.

Given the fact that the distribution of the efficiency scores departs from normality, quantile regression is also employed in the 2nd-stage. The empirical results are shown in Table 4, where the 0.10, 0.25, 0.50, 0.75 and 0.90 quantiles are reported. In addition, OLS estimates showing the mean effects of all covariates are presented. To ensure an adequate coverage of the confidence intervals, 2 000 replications were performed for the regression quantiles. The numbers in parentheses are therefore the bootstrapped standard errors computed to improve statistical efficiency.

Significant differences are observed among the selected quantiles. In particular, the negative impact of government support on farm efficiency indicates that the motivation for improving farms' performance is lower when they are supported by government policies. For the farms that have higher efficiency scores, the marginal effect of subsidies is lower. This means that the farms that perform well are less sensitive to government support and tend to reduce their efficiency at a lower level when receiving agricultural payments.

More specifically, as shown in Figure 5, where each of the plots gives information about the relevant covariate for government support, at any chosen quantile, the question that can be answered is how different is the impact of the corresponding variable on farm efficiency, given a specification of all other conditioning factors. For the variable for decoupled payments, the OLS estimate shows that efficiency declines by 3.6 percent. That is, an increase of 1 percent of subsidies contribution related to the 1st pillar of the Common Agricultural Policy (CAP) to farmers' income leads to a decrease of 3.6 percent in efficiency. However, the quantile regression estimates show higher losses in efficiency for the lower tail of the distribution, where farms are less productive, while in the upper tail, where farmers are more efficient, the reduction in efficiency is relatively smaller. That is, a reduction in efficiency by 2.1 percent at the 0.95 quantile up to 6.8 percent at the 0.05 quantile. The conventional least squares confidence interval does then a poor job of representing this range of disparity.

The opposite effect is observed when considering other government payments. The mean estimate is negative and equal to the coefficient obtained at the 0.50 quantile, remaining statistically significant. The impact of this scheme of government support though varies considerably among the selected quantiles, while its magnitude doubles when comparing the lower and upper tails of the distribution. In terms of the rural development payments, it appears that government support related to the 2nd pillar of the CAP affects in a rather similar manner farms' performance independently of their efficiency level. In particular, the negative impact on farm efficiency is about 3 – 3.5 percent at all quantiles, with the exception of the estimations obtained at the lower quantiles.

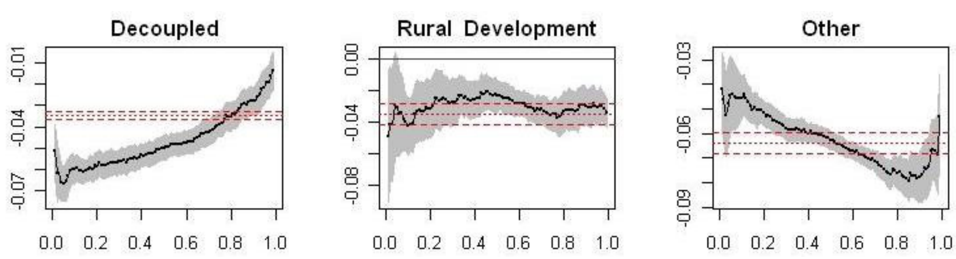

Figure 5. OLS and Quantile regression estimates for government support

Farm size has a positive impact on farm efficiency since it increases efficiency, though different quantiles show a disparity from 1.5 percent at the 0.05 quantile to 1.1 percent at the 0.95 quantile, implying that as a farm becomes larger, it looses efficiency. The OLS estimates show an increase in mean efficiency by 1.4 percent. Moreover, the technology variable appears to affect farm efficiency, though at a rather small rate, remaining statistically significant for all quantiles. It also appears that there is a negative relationship between efficiency and a farm's workforce composition. The relevant coefficient is -2.1 percent for the OLS estimates and it varies along quantiles (from -0.9 percent at the 0.75 quantile up to -1.7 percent at the 0.25 quantile). Its negative sign indicates that farms with a lower proportion of unpaid labor are more efficient. Unpaid laborers appear to have fewer incentives than hired labor to act efficiently, whereas hired labor may be more qualified and more able to perform specialized tasks than unpaid (family) labor.

In addition, farms renting land may be more efficient relative to farms that own the operated land, as the relevant coefficient is statistically significant and negative for all farms. Direct costs of land rentals create then stronger incentives to work the land in a more efficient manner, relative to the opportunity costs borne by owned land. The variable for specialization has an inconsistent effect on farm efficiency, as its impact is positive and significant at the lower quantiles, it becomes though negative and significant above the 0.80 quantile, whereas it remains insignificant in the other cases. The opposite marginal effects in these quantiles indicate that the degree of specialization affects efficiency non-monotonically in the sample. Interpreting the results, livestock producers are increasing their efficiency relative to crop producers by 0.4 percent at the mean estimate, as in the 0.50 quantile.

In terms of farmers' age, it appears that older farmers might be less efficient in comparison to younger ones, though the coefficient is statistically significant in the upper tail of the distribution. Finally, the estimated coefficients for the regional dummies indicate that efficiency is higher in all three regions in comparison with the reference region, which is Sterea Ellada-Nissoi Egaiou-Kriti. However, in the higher quantiles, that is the farms that are more efficient, the coefficients are negative and statistically non-significant.

The pure location shift and the location-scale shift hypothesis were, finally, performed in the 2^{nd}-stage to test the null hypothesis that all the conditional quantile functions have the same slope parameters. Both tests were rejected (with values 63.60 and 37.42, respectively). The effects of the coefficient-by-coefficient tests are also tested and show high significance except of the *Age* and the *Regions*.

Table 4. Empirical results

	OLS estimates	*Quantile regression estimates*				
		0.10	***0.25***	***0.50***	***0.75***	***0.90***
Decoupled payments	-0.036 (0.001)***	-0.062 (0.004)***	-0.060 (0.004)***	-0.051 (0.004)***	-0.040 (0.004)***	-0.029 (0.004)***
Rural development payments	-0.038 (0.004)***	-0.041 (0.011)***	-0.028 (0.011)***	-0.028 (0.007)***	-0.038 (0.005)***	-0.033 (0.007)***
Other payments	-0.066 (0.003)***	-0.047 (0.006)***	-0.056 (0.004)***	-0.066 (0.002)***	-0.078 (0.003)***	-0.082 (0.006)***
Farm size	0.014 (0.001)***	0.015 (0.001)***	0.014 (0.001)***	0.013 (0.001)***	0.014 (0.001)***	0.012 (0.001)***
Technology	0.0002 (0.000)***	0.00015 (0.000)***	0.00022 (0.000)***	0.00021 (0.000)***	0.00018 (0.000)***	0.00014 (0.000)***
Unpaid labor	-0.021 (0.004)***	-0.014 (0.006)**	-0.017 (0.006)***	-0.015 (0.004)***	-0.009 (0.004)**	-0.014 (0.007)**
Owned land	-0.022 (0.002)***	-0.023 (0.003)***	-0.026 (0.003)***	-0.025 (0.003)***	-0.022 (0.003)***	-0.022 (0.003)***
Specialization	0.0044 (0.002)***	0.0089 (0.003)***	0.0076 (0.002)***	0.0038 (0.002)**	-0.0029 (0.002)	-0.0085 (0.003)***
Age	-0.0001 (0.000)	0.0001 (0.000)	-0.000 (0.000)	-0.0001 (0.000)	-0.0001 (0.000)*	-0.0002 (0.000)**
Region 1	0.008 (0.002)***	0.017 (0.003)***	0.009 (0.002)***	0.009 (0.002)***	0.006 (0.002)***	-0.002 (0.003)
Region 2	0.011 (0.002)***	0.005 (0.003)	0.007 (0.003)***	0.011 (0.002)***	0.018 (0.003)***	0.019 (0.003)***
Region 3	0.007 (0.002)***	0.016 (0.003)***	0.006 (0.003)**	0.006 (0.003)**	0.003 (0.003)	-0.005 (0.004)
Intercept	0.901 (0.006)***	0.841 (0.010)***	0.879 (0.010)***	0.903 (0.007)***	0.919 (0.007)***	0.959 (0.011)***

Region 1 refers to *Macedonia–Thrace*; Region 2 is *Ipiros–Peloponnisos–Nissoi Ioniou*; Region 3 represents *Thessalia*, and Region 4 denotes *Sterea Ellada–Nissoi Egaiou–Kriti*.

Values in the parentheses are Standard Errors. Significance levels: 0.01***, 0.05**, 0.1*.

Conclusions

The article examines efficiency in Greek agriculture using farm level data for 2007. In the 1st-stage, production frontiers are estimated by the methods of quantile regression, SFA and DEA, while in the 2nd-stage, these scores are regressed over a set of covariates at different points of the conditional efficiency distribution. Empirical results suggest that the sector is characterized by increasing returns to scale, while the average efficiency obtained using SFA and DEA is about 79 and 71 percent, respectively. The efficiency scores obtained from the quantile regression frontier estimation are though higher (90 percent). The SFA leads to an overestimation of inefficiency, since the employed MLE-estimation is based on the conditional mean, which does not take into account differences in production technology used in different segments of the output distribution.

Furthermore, the distribution of efficiency scores is closer to normality when employing regression quantiles.

Factors that affect efficiency are examined using quantile regressions to capture the remaining deviance from normality. The results suggest that government support aimed at enhancing farms viability should be directed towards payments decoupled from output or prices, as well as rural development payments that affect productivity in a uniform way. It further appears that small farms are relatively more efficient than their counterparts, due to their flexibility to adjust easier in a continuous changed environment. Farms location, specialization and labor composition are also statistically significant determinants of efficiency. Less successful is the variable measuring farms' age.

Overall, a semi-parametric estimator of the efficient frontier is employed, based on conditional quantiles of an appropriate distribution associated with the production process. This line of research generates further discussion on the issue of the appropriate methodology for the estimation of efficiency, as well as on the effect of various covariates that should be estimated at different points of the conditional efficiency distribution rather than just only the mean. The proposed methodology essentially provides better estimates of the production frontier function, leading to robust farm efficiency scores that can be used as more accurate regressors in the 2^{nd}-stage to examine the relevant (policy) questions.

References

Aigner, D.J., C.A.K. Lovell and P. Schmidt 1977. Formulation and estimation of stochastic frontier function models. *Journal of Econometrics* 6: 21-37.

Aragon, Y., A. Daouia and C. Thomas-Agnan 2005. Nonparametric frontier estimation: A conditional quantile-based approach. *Econometric Theory* 21: 358-389.

Charnes, A., W. W. Cooper and E. Rhodes 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research* 2: 429-444.

Coelli, T. and D.S. Prasada Rao 2005. Total factor productivity growth in agriculture: A Malmquist index analysis of 93 countries, 1980-2000. *Agricultural Economics* 32(s1): 115-134.

D'Agostino, R.B., A. Balanger and Jr. R.B. D'Agostino 1990. A suggestion for using powerful and informative tests of normality. *Amer Statistician* 44(4): 316-321.

Daouia, A. and L. Simar 2007. Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics* 140: 375-400.

Debreu, G. 1951. The coefficient of resource utilization. *Econometrica* 19: 273-292.

Färe, R., S. Grosskopf and C. A. K. Lovell. 1985. *The Measurement of Efficiency of Production* Boston: Kluwer-Nijhoff.

Farell, M.J 1957. The measurement of productive efficiency. *Journal of the Royal Statistical Society* 120(3): 253-281.

Kaditi, E.A. and E.I. Nitsi 2009. A two-stage productivity analysis using bootstrapped Malmquist indices and quantile regression. No 52845, Selected Paper presented at 111th EAAE-IAAE Seminar, Canterbury, UK, June 26-27.

Khmaladze E.V. (1981). Martingale approach to the theory of goodness-of-fit tests. *Theory of Probability and its Applications*. 26: 240-257.

Koenker, R. 2005. *Quantile Regression,* New York, NY: Cambridge University Press.

Koenker R. and Z. Xiao 2002 Inference on the quantile regression process. *Econometrica* 70(4): 1583-1612.

Koenker, R. and G. Bassett 1978. Quantile Regression. *Econometrica* 46(1): 33-50.

Meeusen, W. and J. van den Broeck 1977. Efficiency estimation from Cobb-Douglas production functions with composite errors. *International Economic Review* 18: 435-444.

Simar L. and P.W. Wilson 2000 A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics* 27(6): 779-802.

Wadud, A. and B. White 2000. Farm household efficiency in Bangladesh: A comparison of stochastic frontier and DEA methods. *Applied Economics* 32: 1665-1673.

Wilson, P.W. 1995. Detecting influential observations in data envelopment analysis. *Journal of Productivity Analysis* 6: 27-45.