Kostov, P., Davidova, S., Bailey, A. (2017) “Comparative Efficiency of Family and Corporate Farms: Does Family Labour Matter”. Contributed paper presented at XV EAAE Congress, Parma, Italy.

Abstract

This paper examines the comparative efficiency of family vs corporate farms. It decomposes efficiency into two distinct sources - management capabilities and organisational differences. We find evidence for organisational efficiency gains from family farming, relative to corporate farming and these appear to increase with family involvement. With regard to the management capabilities however, family farms do not compare so favourably. Furthermore family involvement does not seem to have any systematic effect on the management capabilities derived efficiency. The findings indicate that further investigation of the way family farms employ and build management capabilities is needed to substantiate any ‘superiority’ claims.

JEL codes: C21; L25; Q12

Keywords: family farms; quantile regression; technical efficiency; efficiency density; managerial capabilities

1. Introduction

Family farming around the world faces multiple challenges. The main economic challenges surround two key issues of access, at the production and marketing stages respectively, to resources such as land, capital and labour, and to appropriate bargaining power in the food chain and other markets. These challenges vary in degree between large and small family farms, and by availability of family labour, and particularly between the different Member States of the EU having different farming structures.

Family farming has long been a concept that has attracted considerable attention from both researchers and policy makers. Indeed, from the very outset of the European Union’s Common Agricultural Policy (CAP), it has claimed family farmers as its main target group (Fennell, 1997). And yet there has never been an agreed EU definition on the concept (Hill, 1993). There are probably many reasons for this, some of which include the complex and multifaceted nature of the phenomenon. Since this paper focuses on the efficiency of family farms relative to corporate farms, it is nevertheless useful to briefly review some definitions of the concept in order to determine which aspects of such definitions should be of prioritised in the following analysis. In this regard we will start with the FAO definition proposed when announcing 2013 as an International year of family farming. It states that "a family farm is an agricultural holding which is managed and operated by a household and where farm labour is largely supplied by that household. The family and the farm are linked, co-*evolve and combine economic, environmental, social and cultural functions"* (FAO, 2013). There are two aspects of relevance to the study in this definition. One is the emphasis on the special mode of management and operation of the farm, exemplified by considerations of household control and use of family labour. The other aspect is the recognition that family farms may follow non-economic objectives dictated by family values. Still, in EU context family farming is more often viewed as a business. "Family farming is more than business but still a business" (EC, 2013). Such a view sits better within the traditions of economic analysis, in that the family values can be described in terms of ownership, control and inheritance of assets (Gasson and Errington, 1993)., and therefore are much more easily accounted for in contrast to more general and difficult to identify sociological notions.

Later attempts to define the concept such as the working document presented at the informal meeting of the Ministers of Agriculture state that "major share of capital is built up by the *manager and his or her family"* and that "*the major share of the family's inco*me is derived *from farming"* (Council of the EU, 26 July 2013). These appear to further lean towards the issue of control and (via the importance of income) to the overall involvement and intertwining of farming and family.

Therefore, for the purposes of this paper we will view a farm as a set of incentives and mechanisms that determine its operation. For family farming, this set of incentives and mechanisms would be different from those characterising non-family farms.

The latter group, called hereafter corporate farms, is also complex and heterogeneous. It includes several types of family-run and non-family run companies, production cooperatives, trusts and charities (Davidova and Thomson, 2014). One common characteristics is that they use predominantly hired labour.

The prospects for family farming are related to farm specialisation. Allen and Lueck (1998) suggest that the value of the family farming declines as farms specialise into the production of fewer outputs. By the same token, if a farm diversifies into processing stages requiring different equipment and skills, the typical family farm will be at a disadvantage. In summary, some economic and technological factors may lead to continued structural change towards non-family forms of organisation of agriculture.

On the other hand, family farming has survived and re-emerged from crises, wars and natural disasters, and adjusted to changing economic fortunes and in some countries to dramatic changes in political context. This has never been a smooth and painless process, and millions of small farmers have disappeared to give way often to larger, more efficient and more competitive farms not reliant on family labour and able to adopt new inputs and technologies (Davidova and Thomson, 2014).

Economic theory backs these survival prospects, suggesting arguments for the superiority of the family type of organisation in agriculture. Family farms are more flexible in responding to market constraints and more resilient to shocks than the more capital-intensive professionally managed farms relying on hired labou (Brookfield and Parsons, 2007). The theoretical arguments have been centred on the relative transaction costs associated with the employment of family and hired labour, and the different incentive structures which motivate individuals. Family labour has stronger incentives to be more productive and requires lower monitoring costs because the members of the farm family are residual claimants on farm profits (Allen and Lueck, 1998; Pollak 1985; Schmitt, 1991, 1993). The exposure of agriculture to weather, pests and disease risks makes it harder to link work effort to output, making it difficult to design the right incentives to motivate hired labour. Additionally, the spatial dispersion of work makes the monitoring costs in agriculture particularly large. Pollack (1985) emphasises families ability to provide incentives and monitor performance, ability that is higher than in corporate farms. This implies that the average productivity of labour on family farms may be higher when compared to hired workers on the corporate farms. Kostov et al. (2016) termed the above a 'motivation effect' of family labour. The concept is similar to what Valentinov (2007) calls 'efficiency of monitoring activities'.

However, Polack (1985) also underlines some disadvantages of a family business. These include the possible toleration of inefficiency and slack performance, and the possible lack of talent and skills required for a successful business performance. To these could be added the possible presence of non-economic objectives in family farming, e.g. to produce food for the household, provide employment to family members and preserve the farm for the next generation. Some of these disadvantages are due to the complex character of the family farm, which is at the same time a production unit, a consumption unit (household) and a focus of kinship and family (Djurfeldt, 1996). All this can limit farm growth, produce underemployment and reduce productivity. This might lead to a situation in which more family-oriented farms lose relative efficiency in comparison to the corporate farms. Kostov et al. (2016) refer to the latter as the ‘management capabilities deterioration effect’. Due to the existence of these two opposing effects (positive effect of more efficient monitoring and negative aspects of family farms discussed above), it cannot be said a priori that, from an economic point of view, family farming is a superior organisation of production.

As mentioned, in this paper the farm organisation is treated as a set of incentives and mechanisms that contribute to allowing a farm to achieve a certain level of efficiency. While the transaction costs literature is interested in what mechanisms and incentives are better suited to achieve this, the question in this paper is how the outcome would change if these mechanisms and incentives vary. The question has important policy relevance. Concerning the EU, for example, the European Parliament (EP) Committee on Agriculture and Rural Development (COMAGRI) published a motion for an EP Resolution on ‘How the CAP can improve job creation in rural areas’. It states that “support should be provided first and foremost to family farms run by one or more responsible, independent farmers who work on their farms in an effective manner and who are much better able to deal with any problems by adapting their production and/or their production methods and by diversifying their activities when necessary” (COMAGRI, 2016). What lies behind this statement is an implicit assumption that family farming is a superior form of organisation in agriculture. In this context, it is important to investigate to what extent in the 21^{st} century, with predominantly capital-intensive technology and high requirements to farm managers to run a business that is both economically viable and environmentally sustainable, family farming still can exhibit a superior economic performance.

This study employs as a criterion for a family farm the use of family labour in the broadest possible sense, i.e. any farm which employs family (unpaid) labour. The source of farm-level data is the EU Farm Accountancy Data Network (FADN), since FADN allows the classification of farms according to the use of family or non-family labour.

A non-parametric non-separable farm production function is estimated, and the efficiency scores are derived for both family and corporate farms. The distribution of these efficiency scores for the two farm types is then compared by using their empirical probability density functions. The relative difference in the probability density functions for the efficiency scores of family and corporate farms is used to represent the efficiency gains, or losses, attributable to family farming. One of the contributions of the paper to the debate about the survivability of family farming under the conditions of technical change and weak power in the globalised food chain is that by re-specifying the production function used to derive efficiency scores, the total efficiency effects are decomposed into two distinct components, namely M-efficiency which measures the individual farm's managerial capabilities to realise any efficiency gains and F-efficiency, associated with the potential gains/losses derivable from the family form of organisation of farming activities.

The analysis considers separately four EU Member States, each characterised by a different mix of family and corporate farms - the Czech Republic, Hungary, Romania and Spain.

The rest of the paper is structured as follows. The next section presents the analytical approach and section 3 details the estimation strategy. Section 4 describes briefly the data and section 5 presents the results. Section 6 formulates the conclusions and policy implications.

2. Analytical Approach

In order to rank farms by their technical efficiency, first, a measure of technical efficiency for all farms is calculated. If family farms are more efficient than the corporate ones, then the expectations are that there will be more family farms at the higher levels of efficiency and vice versa. By comparing the distribution of such an efficiency measure for family and corporate farms, inference about the relative efficiency of family farms can be carried out. Logically such a comparison could be based on their probability distribution functions. In order to do this we consider the differential between the empirical probability density functions of the efficiency measure for family versus corporate farms. Hereafter we will refer to the above as (efficiency) density differential.

If indeed family farms are more efficient than their corporate counterparts, then there will be more family farms (relative to the corresponding number of corporate farms) at higher efficiency ranges. This means that there will be higher probability of finding family (vs corporate) farms at larger values of the efficiency measure. Therefore this should translate into a positive efficiency density differential at the higher efficiency ranges. Similarly at lower efficiency ranges, the expectation would be for a negative density differential. Putting the above two expected outcomes together means that, under the hypothesis of more efficient family farms, upward-sloping density differentials will be obtained - from negative values at the lower efficiency ranges to positive values at the higher efficiency ranges. Similarly, a downward-sloping density differential would indicate that family farms are less efficient. The density differential can be viewed in the same way as a probability density function (PDF). The main difference is that since the PDF function shows the probability of occurrence of a certain outcome, it is only defined for non-negative values. The density differential on the other hand is essentially a difference between two distinct PDFs. As such, it will produce negative, as well as positive values. So the interpretation for the density differential values is in terms of excess probability. Positive values for the efficiency density differential mean that family farms are more likely to obtain the corresponding values of the efficiency measure and vice versa. Consider the following example. Let us assume that for some level of efficiency (i.e. value of the efficiency measure) the family farms PDF value is 30% and that for corporate farms is 20%. Then the density differential value at that point would be 0.3-0.2=0.1. This would mean that there will be, in relative terms, more family farms than corporate farms at that particular point, i.e. achieving this particular efficiency level. The result from such a simple differential is however tricky to interpret. The relative density differential is much easier to interpret. In the above example the relative (of family vs corporate) density differential at the same point will be (0.3-0.2)/0.2=0.5. In this case, we could say that a family farm is 50% more likely to achieve this particular level of efficiency than a corporate one. Therefore, hereafter we will use the relative efficiency density differential.

The efficiency density differential between the two groups of farms is decomposed into two distinct components. Using a standard approach to estimate an efficiency frontier and thus allocate efficiency scores to each farm essentially amounts to imposing the same efficiency frontier to both farm types. In this case, since we do not explicitly distinguish between family and corporate farms, any effects that are due to such differences, are not accounted for in the estimation of the production frontier. Such differences (between family and corporate farms) will then be fully absorbed by the efficiency scores. Consequently, the density differentials will account for these differences, but will also include efficiency effects that do not depend on the family/corporate farm distinction. We want to distinguish between these two types of effects and for this reason we propose to decompose the efficiency density differential as described below.

Let us assume that family involvement in farming can potentially have transformative effect on farm productivity (although we do not assume whether any such effect is positive or negative). What we mean under transformative effect is that family involvement could modify the nature of the production relationship (i.e. lead to essentially different production function). Furthermore, let us also approximate the extent of family involvement by the amount of family labour employed. Then a proper account of the farm production potential, that takes into account the possibility for such transformative effect, could be derived by considering a production function that treats family and hired labour as different inputs. As a result, the efficiency frontier derived from this approach will vary with the type of farm and, hence, the maximum output possible for a given farm will depend on the amount of family labour it employs. Since we hypothesise that the level of family involvement can effectively modify the production function, this means that only non-additive specifications for the production function are admissible. Let us consider what this alternative approach estimates. Since the differences between family and corporate farms are accounted for by the extent of family involvement (which is zero for corporate farms), the production possibilities frontier estimated under this approach will account for any such differences. Hence in a way each farm under this approach will have its own specific maximum achievable output that depends on the level of family involvement. Therefore by accounting for the difference between family and corporate farms in the frontier, we effectively exclude such effects from the efficiency scores and therefore from the resulting efficiency density differentials.

The efficiency measure obtained from differentiating between family and hired labour input is referred to hereafter as M-efficiency. If one assumes that the mechanisms through which this potential is to be realised is the management of farm production, then M-efficiency is in fact a measure of the management capability of a farm.

Therefore, the efficiency density differential can be viewed as consisting of two components, namely the management capability of a farm (which we measure by M-efficiency), and a remainder, which would in fact measure the differences between family and corporate farms. Hereafter, we will refer to this difference as F-efficiency density differential and we construct it as simply taking the difference between the aggregate efficiency differential and the M-efficiency differential.

The comparative efficiency of family and corporate farms is thus decomposed into two components: management capabilities (i.e. M-efficiency) and difference due to organisational form, (i.e. F-efficiency), in this case difference between family and corporate farms. The basis for this decomposition is the family involvement in the business operations. However one needs to acknowledge that family farms are also able to trade off returns to family owned land and capital in order to maintain family relevant values (such as preserving the farm for future generations). Under perfectly competitive markets this would not be an issue. In the presence of a range of possible market imperfections, e.g. if capital or land markets are more constrained than labour markets, such non-labour related difference will accumulate in the M-efficiency differentials.

When asking the question of whether family farms are indeed superior form of organisation of agricultural production, it is the F-efficiency differential that is of primary interest. If for whatever reason family farms lack the management capabilities to utilise their potential, they will be characterised by lower M-efficiency and this may (partially of fully) offset the positive effect of F-efficiency. It is however possible that family farms are better in achieving their potential than their non-family counterparts in which case they will exhibit a higher M-efficiency than the corporate farms. Therefore, the efficiency decomposition, presented above provides a useful framework to evaluate the relative efficiency of family and corporate farms.

3. Implementation details

3.1. Efficiency estimation

In order to construct efficiency density differential and their decomposition as discussed previously, one just needs efficiency scores for each farm in the sample. The exact method used to estimate such efficiency scores is inconsequential and a variety of alternatives could be used. The main requirement is that the individual efficiency scores for each farm in the dataset need to be estimated based on two different production functions: one that includes labour as a single input, and another that differentiates family and non-family labour. There are different methods available to calculate such efficiency scores. The conventional approach to efficiency relies upon some form of frontier estimation and conceptualises efficiency in terms of distance between actual realisations and the estimated frontier. There are several estimation measures. The most basic and easy to interpret one is the distance function which is simply $y/\hat{y}$ (y being actual production, while $\hat{y}$ being the corresponding maximum output, defined by the efficient frontier). The efficient frontier can be recovered non-parametrically via e.g. Data Envelopment Analysis (DEA), or estimated parametrically by employing some additional assumptions.

This paper uses a different way to calculate an efficiency frontier, based on quantile regression methods. The approach taken and the rationale for it are outlined below.

The conditional quantile regression estimated a pre-specified quantile of the conditional distribution of the response variable, given a set of covariates. When the functional relationship that is modelled is a production function, this conditional quantile can be interpreted as a level of technical efficiency. For example, the 0.9th quantile regression models a hypothetical farm that is more efficient than 90% of the other farms in the sample and one can define this as a 90% efficient farm. Consider the quantile regression for $\tau = \left(\mathrm{n}-1\right)/\mathrm{n}$, where n is the sample size. This will result in estimating the production function for the most efficient farm in the dataset. The latter is effectively an estimate of the efficiency frontier.

Hendricks and Koenker (1992, p. 58) noted that the frontier models "correspond closely to models for extreme quantiles of a stochastic production surface". So using extreme conditional quantile functions was adopted in Bernini et al (2004) and Behr (2010). Detailed analysis of such quantile based efficiency measures is presented in Aragon et al. (2005). The above papers adopt a restrictive linear quantile model formulation, which requires parametric production function. Nonparametric extensions have been considered in Martins-Filho and Yao (2008), Wang and Wang (2013) and Wang et al. (2014).

However, estimating extreme quantiles results in loss of estimation efficiency (see Chernozhukov, 2005). In simple terms the issue is as follows. In the mean based frontier estimation the overall shape of the frontier is based on an average estimated relationship (modified by a parametric assumption about the distribution of the efficiency scores). The overall shape of the estimated relationship for conditional quantiles however changes with the quantile. So using two different quantile functions as 'frontiers' can change the relative distance from each observation to the frontier and hence can affect the sorting of efficiency scores amongst farms. It is the efficiency sorting, not so much the values of the efficiency scores, that affect the construction of probability densities and their differentials, which are a centrepiece of our approach. Such a problem would not arise if the corresponding quantile functions are parallel to each other. Hence, more extreme quantiles estimates are conceptually closer to the actual frontier, but their estimation is less reliable in statistical sense. Less extreme quantiles are further from the actual frontier, but their shape estimation is more reliable.

The estimation efficiency loss can be reduced by estimating a less extreme tail quantile. There is a trade-off between the complexity of the production function model (in that more complex specification involve higher number of parameters and therefore more estimation efficiency loss) and the choice of such tail quantile. For example for parametric specifications of the production frontier Bernini et al (2004) and Behr (2010) proposed using the 0.95^{th} quantile regression. However non-parametric estimation involves greater degree of complexity. Hence in order to ensure reliable estimation of the upper conditional quantile functions (see Chernozhukov, 2005; Chernozhukov and Fernandez-Val, 2011 for details), which will be used as 'frontiers', we use the 0.9th quantile for non-parametric estimation. Hereafter, these upper quantile regressions are referred to as τ -quantile envelopes, e.g. the 0.9th quantile regression is referred to as a 0.9th quantile envelope. Quantile based frontier models do not envelope all the data, and as such, similarly to the m-output frontier models of Cazals et al. (2002), are robust to outliers.

The paper employs quantile regression residuals to construct efficiency measures. In a τ -quantile regression, approximately τ of the residuals will be negative and 1- τ of the residuals will be positive. The distance function derived from such a quantile regression model would then produce a measure of efficiency that we will hereafter refer to as an efficiency score. The efficiency scores are a linear function of the residuals i.e. $y / \hat{y} = \frac{y - \hat{y}}{\hat{y}} + 1 = 1 + \frac{\varepsilon}{\hat{y}}$ where ε are the residuals. This distance function efficiency scores are a measure of technical efficiency thus larger numbers denote better technical efficiency.

The above means that the efficiency scores can be interpreted as conditional (on τ) efficiency. When τ =0.9, 10 per cent of the farms will have positive residuals since by the actual definition of the 'efficient' farms, 10 per cent of the actual farms are more efficient. This efficiency measure will be under 1 for 90 per cent of the farms and 10 per cent of the obtained values from the data sample will actually exceed 1.

Therefore, unlike the conventional efficiency scores, the ones we obtain here are not limited from above. It would be tempting to 'standardise' these residuals to a [0,1] interval to resemble the standard efficiency measures. If we were to do so, however, we would lose the natural conditional efficiency interpretation and mask their dependence on the level of efficiency represented by the quantile envelope. Instead, in this paper the density function of the conditional efficiency scores is used to compare the efficiency distribution of different farms. Using the upper quantile envelope (i.e. high τ) with non-parametric estimation allows a natural equivalence to non-parametric DEA estimation. If we wanted results to resemble more conventional approach, we could have trimmed away the observation lying outside the quantile envelope (i.e. effectively treating them as outliers). Such an approach would however exclude the most efficient farms and is undesirable.

The estimated frontier in the stochastic frontier approach is a fitted production function plus disturbance (usually Gaussian and iid) term plus an efficiency score shifter (term reflecting the distribution of the efficiency scores usually with a half-normal distribution). So one can think of the stochastic frontier approach along the following lines. A mean production function is estimated. Then this function is shifted outwards by the efficiency scores using an assumption about their distribution. If the efficiency scores were uniformly distributed the frontier would have exactly the same shape as the mean production function. In this estimation process this shape is determined by all observations in equal degree and also depends on an arbitrary assumption on the distribution of the efficiency measures. The latter is an undesirable feature. The strength of the quantile regression approach is that it does not employ any distributional assumptions and hence does not restrict the distribution of the efficiency measures that are being calculated from it. The other advantage is that in quantile regression the shape of the estimated function changes with the quantiles and although all observations contribute to it, the ones that are associated with higher efficiency are weighted differently to those that are related to lower efficiency farms. Since the shape of the frontier is essential in calculating efficiency scores, the quantile approach is preferable in that it does not effectively impose the same input/output relationship on both efficient and inefficient farms. Non-parametric alternatives such as DEA are less susceptible to the above mentioned problem. The issue with such alternatives is that they lead to the other extreme, i.e. that inefficient farms have no contribution to the frontier estimation and outliers drastically affect the latter. In effect, the quantile regression approach would essentially converge to DEA in the case of non-parametric estimation of an extreme quantile.

3.2. Quantile regression and density ratios estimation

Here the unknown production function is estimated non-parametrically and thus avoids the necessity to specify any pre-defined functional form. The non-parametric quantile regression applied here can be expressed as:

$$y = f_{\tau}\left(X\right) + u_{\tau} \tag{1}$$

$$\text{st } q_{\tau}\left(u_{\tau} \mid X\right) = 0 \tag{2}$$

where y is the dependent variable vector, X is a matrix containing the covariates, $f_\tau(.)$ is a quantile dependent arbitrary (i.e. unspecified) function, τ is the quantile being modelled and u_τ are the residuals. In contrast to the more widely known linear quantile regression specification, the effect of the covariates is given by a non-parametrically specified function, which itself is quantile dependent, and the conditional quantile restriction in (2) is specified with regard to this non-parametric function. This quantile restriction states that the τ quantile of the residuals is zero. In the standard mean regression the corresponding assumption is that the expected value of the residuals is zero and that they are characterised by some assumed distribution (usually Gaussian with a fixed variance).

Estimations in this paper follow the indirect method of Li and Racine (2008). It consists of two distinct steps. First, the conditional distribution function of the dependent variable with regard to the covariates is estimated. This is a standard (conditional) density estimation problem. Then conditional quantile estimates are recovered by inverting the conditional density function via locally constant kernel estimation approach. This is also a standard non-parametric problem, since it relates the conditional quantiles of the dependent variables (obtained from the conditional density) to the covariates using a kernel weighting with bandwidths specified in the first step.

Following the quantile envelope function estimation we obtain efficiency score measures for each farm in the sample. Then we split the farms into family and corporate farms and construct the efficiency probability density differentials which form the basis of the analysis. The relative density differentials used in the paper are simply: $\mathrm{pdf}_{\mathrm{fam}}(z)/\mathrm{pdf}_{\mathrm{corp}}(z)-1$, where $\mathrm{pdf}_{\mathrm{fam}}(z)$ and $\mathrm{pdf}_{\mathrm{corp}}(z)$ are the empirical probability density functions for the efficiency scores for the family and corporate farms. The efficiency scores themselves are calculated as $z=y/\hat{y}=1+\frac{u_\tau}{\hat{y}}$, where $\hat{y}$ are the fitted values from the estimated quantile envelope (i.e. $\hat{y}=f_\tau(X)=y-u_\tau$) and u_τ are the residuals from the estimated quantile regression model in (1).

Therefore we need to estimate the density ratio $\mathrm{pdf}_{\mathrm{fam}}(z)/\mathrm{pdf}_{\mathrm{corp}}(z)$. A straightforward but naïve approach to density-ratio estimation would be to separately estimate the corresponding probability densities (corresponding to the numerator and the denominator of the ratio), and then take the ratio of the estimated densities. However, unless we have simple parametric density model, density estimation could be problematic, particularly in high-dimensional cases (Vapnik 1998). Therefore, for reliable statistical inference it is preferable to directly estimate the density ratio without going through separate density estimation for the numerator and the denominator. There are many methods that have been used for that purpose. One of the most popular methods, due to its ease of implementation and computational properties, is the unconstrained Least-Squares Importance Fitting approach. This approach is a squared-loss version of the M-estimator for the linear density-ratio model. It has a closed-form solution and the 'leave-one-out' cross-validation score associated with this approach can be analytically computed. Here a kernel version of the latter is employed (Kanamori et al, 2012) together with analytical 'leave-one-out' cross-validation. The method provides us with density ratio and we convert this to a relative density differential by simply subtracting 1 from it.

4. Data

The data used in this study comes from the EU's Farm Accountancy Data Network (FADN). The FADN samples include only commercial holdings defined in terms of their economic size, so that very small and semi-subsistence family farms are excluded. The analysis focuses on four EU Member States which differ substantially according to their farm structure. These states are the Czech Republic, Hungary, Romania and Spain. The Czech Republic has a farm structure with a particularly high share of corporate farms which do not rely heavily on family labour. The farm structure in Hungary presents a mix between corporate and family farms. Romania is the EU Member State with the largest number of small family farms and a high proportion of these farms, 93 per cent, are considered to be semi-subsistence holdings which consume more than half of the farm output within the household (Davidova et. al., 2013). Semi-subsistence farms affect the nature of the commercial farms which are the FADN field of observation since they present an inelastic supply of commercial farm labour. Finally, Spain's agriculture is dominated by family farms.

Altogether, these four countries account for 12,929 observations in the EU FADN dataset, out of which 11,606 are family farms (89.7 per cent of all observations). In order to apply the analytical approach described above, the following variables were extracted from the FADN dataset. The dependent variable is total output measured in value expressed in the currency of the respective country (cu). The production function is specified with regard to labour, land (in hectares of utilised agricultural area (UAA)), capital and intermediate consumption, the last two components are also expressed in value terms. Total labour input is measured in annual work units (AWU) and is split into two variables: family labour and non-family labour. As explained previously, these two labour variables are used as separate inputs in the production functions in the M- efficiency estimation and are added together as a single input for the 'total' efficiency estimation.

Capital is calculated as the difference between the total fixed assets, on the one hand, and land, permanent crops and quotas (all in monetary terms), on the other. In this way a proxy for a long-term capital is obtained. The value of the land is excluded from the capital measure in order to avoid double counting since it is used as a separate input to the production function. FADN bundles together the value of land with permanent crops and policy quotas, and does not provide information for the quality of land. Finally total intermediate consumption is extracted as calculated in FADN.

In order to provide an insight into the data for each country, the family farms are split into four groups, according to family to hired labour ratio. The logic is that higher values of this ratio signifies family farms with greater family involvement. Table 1 presents some summary statistics for corporate farms and family farms with the latter grouped according to the degree of family involvement. The degree of family involvement is measured by the family labour input and is grouped according to the quartiles of the latter. The fourth group (i.e. fourth quartile according to the family labour input) contains by far the largest number of farms, with almost half of the total number of farms in the three EU New Member States and 73 per cent of the Spanish farms falling in this category. Considerable differences between countries are noticeable. In Romania, Hungary and the Czech Republic, corporate farms account for a large proportion of total output, UAA and labour utilisation, while family farms dominate Spain's agriculture in these respects. As expected, corporate ones are larger in all countries. It is also informative to compare the average values of the family labour input. One would have expected that family labour input should increase with the corresponding quartile, but this is not uniformly the case. In the Czech Republic and Hungary the average family labour in the fourth group is actually lower than in the third one, while for the other two countries the values in the third and fourth group are almost the same. The extent of decrease of total labour input raises when moving from lower to higher quartiles, which may explain the above counter-intuitive situation. This raises the question of what is a more appropriate measure of family involvement, namely the amount of family labour or its share in total labour. The share approach, as reflected in Table 1, can lead to defining a 'typical' family farms as one with modest size employing very little hired labour.

Table 1. Summary statistics for corporate farms and family farms by quartile of family labour usage

	Corporate farms	Family farms by quartile of family involvement			
		(0,0.25]	(0.25,0.5]	(0.5,0.75]	(0.75,1]
Romania					
Share of farms (%)	0.30	0.05	0.05	0.10	0.51
Share of UAA (%)	0.91	0.05	0.01	0.01	0.03
Share of Output (%)	0.94	0.04	0.00	0.00	0.01
Share of Labour (%)	0.81	0.08	0.01	0.02	0.08
Average UAA (ha)	823	330	36	23	14
Average Output (000 cu)[1]	2,234	620	76	29	18
Average Labour (AWU)	24.68	16.17	2.74	2.13	1.38
Average Family Labour (AWU)	0.00	0.75	1.02	1.33	1.32
Czech Republic					
Share of farms (%)	0.35	0.04	0.08	0.05	0.48
Share of UAA (%)	0.82	0.04	0.05	0.02	0.07
Share of Output (%)	0.88	0.03	0.03	0.01	0.05
Share of Labour (%)	0.89	0.03	0.02	0.01	0.05
Average UAA (ha)	1,280	632	313	166	80
Average Output (000 cu)[1]	1,884	553	296	182	75
Average Labour (AWU)	43.67	12.03	5.04	3.42	1.85
Average Family Labour (AWU)	0.00	1.55	1.91	2.15	1.83
Hungary					
Share of farms (%)	0.21	0.08	0.14	0.13	0.45
Share of UAA (%)	0.66	0.06	0.08	0.06	0.15
Share of Output (%)	0.77	0.05	0.06	0.04	0.08
Share of Labour (%)	0.72	0.07	0.07	0.04	0.09
Average UAA (ha)	673	164	119	96	70
Average Output (000 cu)[1]	1,198	205	147	107	60
Average Labour (AWU)	20.26	5.67	2.96	2.02	1.15
Average Family Labour (AWU)	0.00	0.80	1.06	1.22	1.10
Spain					
Share of farms (%)	0.01	0.03	0.08	0.15	0.73
Share of UAA (%)	0.02	0.04	0.09	0.14	0.71
Share of Output (%)	0.02	0.12	0.15	0.15	0.56
Share of Labour (%)	0.02	0.14	0.14	0.16	0.54
Average UAA (ha)	133	65	58	51	50
Average Output (000 cu)[1]	273	376	168	100	73
Average Labour (AWU)	6.85	8.13	3.00	2.00	1.33
Average Family Labour (AWU)	0.00	0.92	1.13	1.27	1.28

[1] cu means currency unit

However, in this paper we focus on whether we can find support for the theoretical conjecture that suggests that the potential benefits of family farming are related to lower monitoring costs. Any such benefits are more likely to accrue due to the amount of family labour rather than its share. The share of family labour implicitly captures the differential effort involved in family and hired labour input. In other words, the potential benefit this measure captures is due to the expectation that family members work harder. The absolute amount of family labour on the other hand represents the commitment of the family to the farm operation. Higher level of commitment is expected to lead to more organisational improvements that could result in a more productive farm. Therefore, we measure family involvement by the amount of family labour, rather than its share.

5. Results

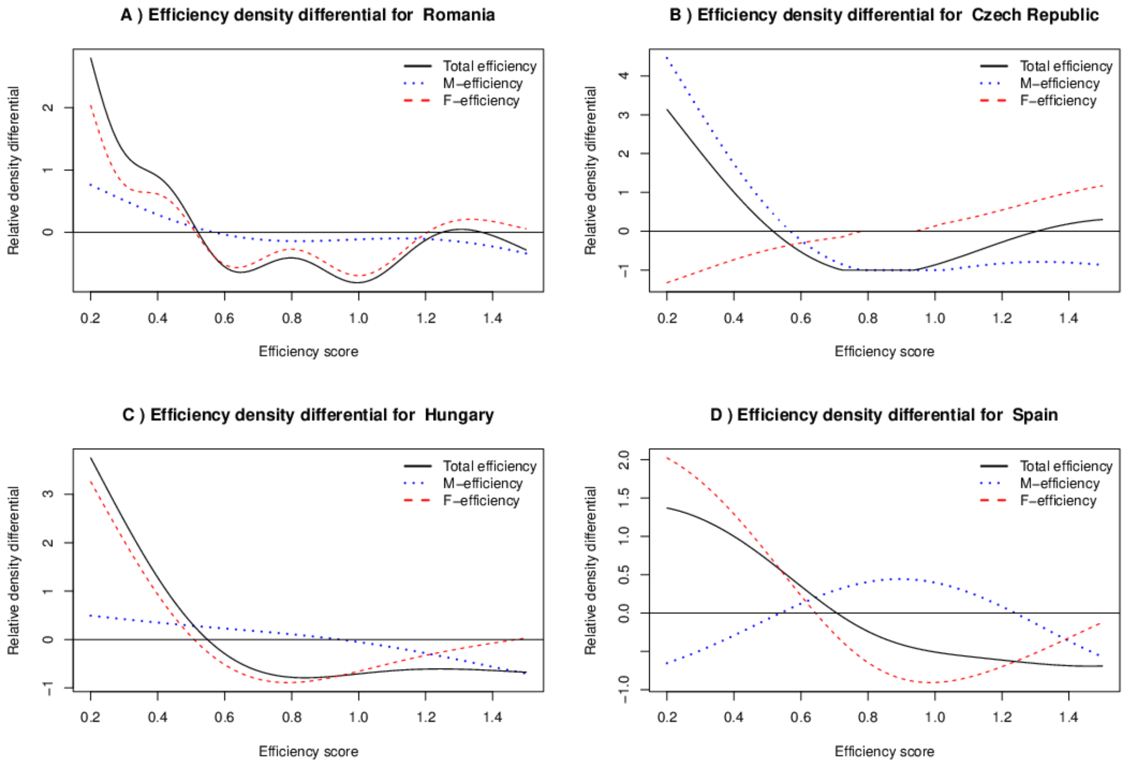
Figure 1 presents the relative density differentials associated with the efficiency decomposition, i.e. total, M and F-efficiency differentials, for each of the EU Member States analysed. They are computed as density differentials between family and corporate farms. If the theoretical conjecture of the superiority of family farms relative to corporate ones is true, one would expect more probability mass in the family farms distribution towards the upper end of the density range. This would show as a negative differential at the lower end of the density range and positive one at the upper end, resulting in generally upward-sloping efficiency differential. If the family farms are inferior, the opposite would be true.

Let us first consider the total efficiency effects, derived from the production function that does not separate family and non-family labour. A mostly downward-sloping total efficiency effect is observed for all countries. This means that there are fewer family farms at the higher efficiency levels than corporate ones, suggesting that family farms are, on aggregate, not as efficient as corporate ones. For Romania and the Czech Republic however if only the top of the density range (i.e. the top 10 per cent efficient farms) is considered, family farms compare favourably to corporate ones.

The efficiency decomposition results are more subtle and varied amongst countries. For Romania and Hungary the M-efficiency differentials are relatively small and this results from the fact that the total efficiency differential follows closely the F-efficiency differential. In other words, the total production effect of family farms in these two countries is mostly defined by the F-efficiency. Hence, in these two countries there is not much difference in the managerial capabilities contribution to efficiency in family and corporate farms, but family farms are generally not as efficient as the corporate ones.

In Romania, at the top end of the efficiency spectrum, the F-efficiency effect roughly offsets the negative impact of M-efficiency and as a result the total efficiency differential is close to zero. So some, although very limited, support for the superiority of the most efficient family farms can be found in this case. But then these family farms do not manage to take advantage of this since they are unable to translate this potential into actual (i.e. total) efficiency gains, due to worse M-efficiency, i.e. due to worse management capabilities.

Figure 1. Efficiency differentials by country



For Hungary, at the top end of the efficiency density range the M-efficiency effect appears to dominate and to account for virtually the entire efficiency differential.

For the Czech Republic the M-efficiency differential dominates the efficiency differentials, but the F-efficiency is clearly expressed translating into an overall positive total efficiency differential at the highest efficiency scores. The F-efficiency differential has the expected upward slope demonstrating the superiority of family farming. However, similarly to

Romania, this potential efficiency gains from family farming fail to realise due to the fact that they are more than offset by M-efficiency losses.

Finally, for Spain the hat-like shape of the M-efficiency differential, combined with the weak recovery in F-efficiency effect, tentatively suggests that both these effects may be present (which is easier to notice in the right-hand part of the graph) with an overall small negative total efficiency differentials.

In summary, Figure 1 shows that the F-efficiency density differential is upward sloping (although only becoming positive for the Czech Republic) as efficiency increases. This confirms the results of Kostov et al (2016) that family farming gains are realised after an efficiency threshold is reached, i.e. only for the most efficient farms.

Using a counterfactual simulation approach, Kostov et al. (2016) claim that family involvement generates additional output effects, but these only materialise at a significantly high threshold of 1.75 AWU of family labour input. To re-examine such findings, the analysis considers the extent of family involvement, measured by the amount of family labour input, assuming that family farms that employ more family labour are more deeply engaged in the farming business. Such farms are denoted here as ‘more involved’. Since family labour is a continuous variable, it is difficult to assess the impact of family involvement. For this reason we discretise the variable by synthetically splitting family farms into groups defined by 0.5 AWU increments of family labour. The corresponding efficiency differentials (compared to corporate farms) for each of these groups are calculated. These differentials are used to investigate the way family involvement affects the efficiency decomposition.

The corresponding F and M-efficiency differentials are presented in Figures 2 and 3 respectively. To facilitate interpretation, only the efficiency score levels of 0.6 to 1.2 are presented. The reason for this is as follows. The highest efficiency scores could represent outliers. Similarly we remove the most inefficient levels to eliminate the impact of any potential outliers. In general, the argument for superiority of family farms should translate into downward slope for the M-efficiency differential and an upward one for the F-efficiency differential. Based on Kostov et al (2016), we would expect such results to gradually appear with a move from less to more involved family farms.

Figure 2. Changes in F-efficiency effects by family involvement

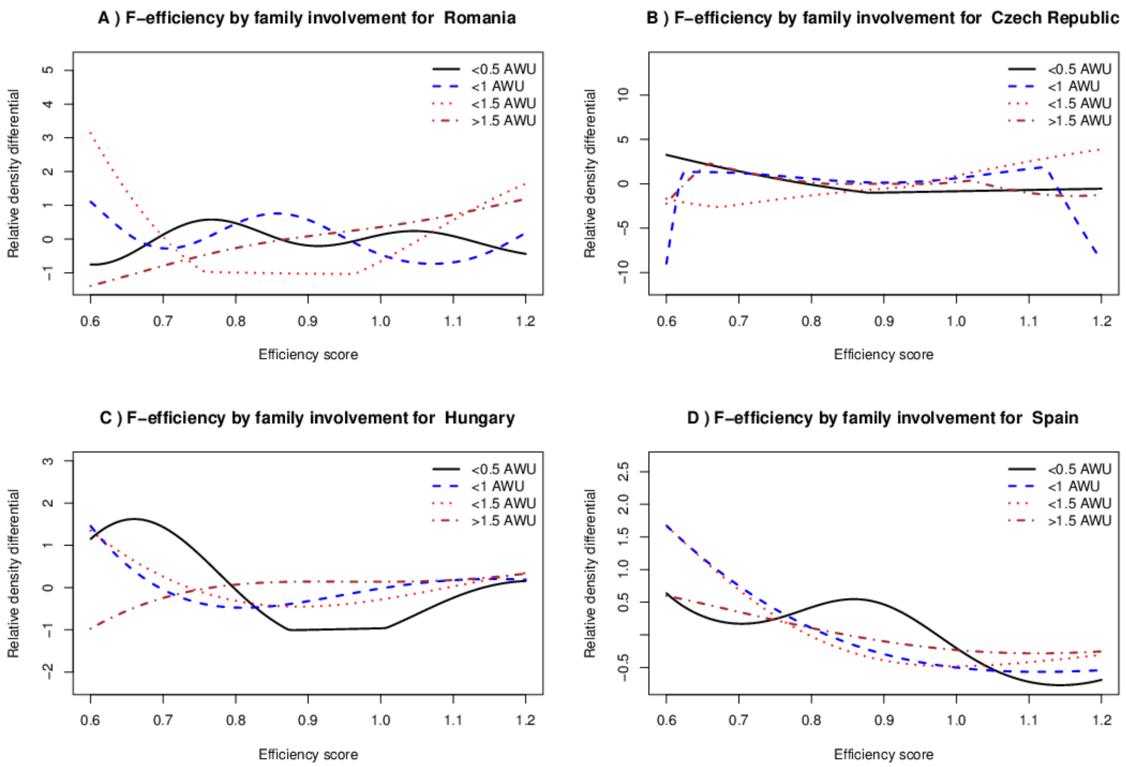
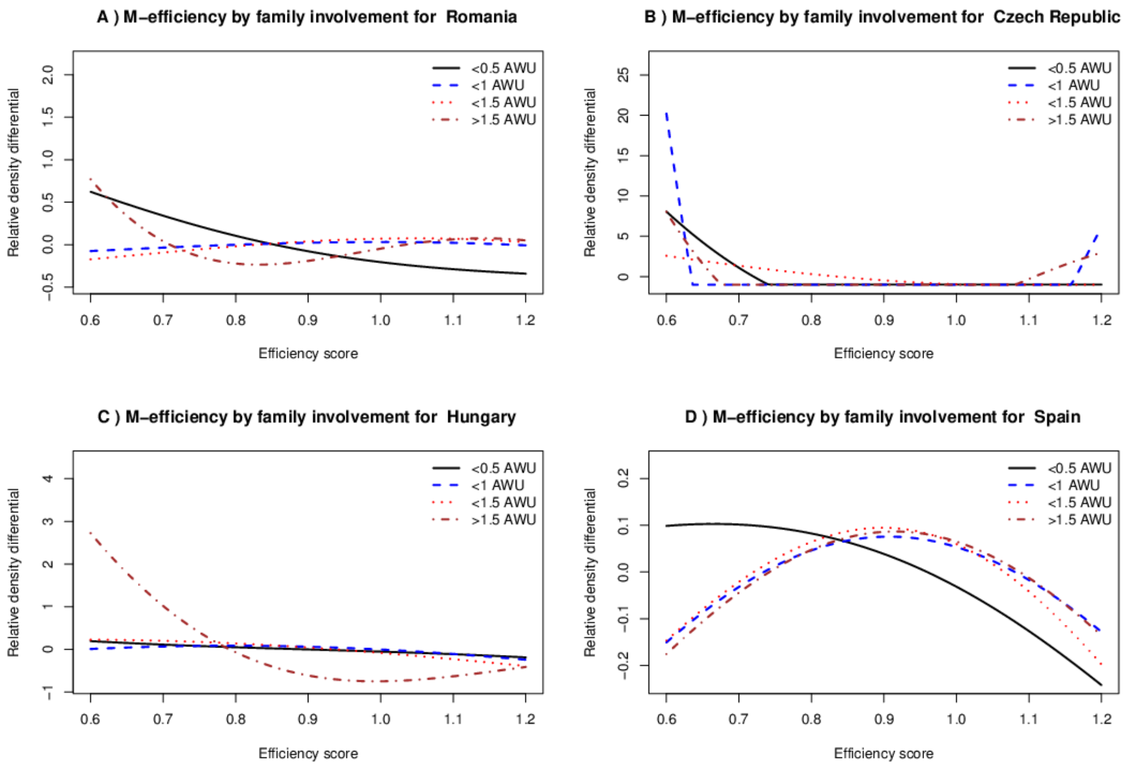


Figure 3. Changes in M-efficiency effects by family involvement



Considering F-efficiency, in general, it is enhanced by greater family involvement. This is consistent with the hypothesis that greater family involvement will lead to positive efficiency gains due to better monitoring. Let us look at the shape of the relevant differentials for adjacent categories. In some of these adjacent categories (in particular in the case of the <0.5 and the <1 AWU category pairing, as well as the <1.5 and >1.5 AWU pairing) the corresponding differentials may slightly deviate from this ordering since the next category (of greater family involvement) may not always lead to greater F-efficiency. This could be a result of actual thresholds which are slightly different from the ones used here to create the underlying family involvement groups. The most important conclusion from Figure 2 is that there is evidence of monitoring effects for all countries with the possible exception of Spain. Another point is that in the case of the Czech Republic it appears that too much family involvement has a negative impact, since the <1.5 AWU category shows the best potential efficiency gains (i.e. it is more efficient than the >1.5 AWU category). However, such a result needs to be treated with caution due to the smaller number of observations in the >1.5 AWU category. In general, these results corroborate the assumption that greater family involvement improves efficiency.

With regard to M-efficiency, Figure 3 presents how this changes with the level of family involvement. Unlike F-efficiency, the M-efficiency differential effects (in the sense of regular pattern across categories) are only observed for some of the groups of family farms considered here. For example, in Romania only the low (<0.5 AWU) and high (>1.5 AWU) levels of family involvement demonstrate such effects. In the Czech Republic, there is no evidence for family involvement affecting the M-efficiency effect. However, the results have to be treated with caution due to the small number of family farms in some categories. In Hungary the M-efficiency effect occurs at all levels of family involvement but only for the highest one (>1.5 AWU) it is visibly stronger. In Spain, however, we observe exactly the opposite case with the strongest effect for the least involved (<0.5 AWU) family farms and very little difference in the other categories. Hence unlike the case of F-efficiency, there does not appear to be any definite link between family involvement and M-efficiency (which represents management capabilities).

6. Conclusions and policy implications

This paper examined quantitatively the claims of the 'superiority' in terms of economic efficiency of the family form of organisation of agricultural production compared to non-family 'corporate' farming. The theoretical arguments related to the superiority of family farming are centred on family labour, which could be more productive and involving lower monitoring costs than hired labour as it is more motivated in its role as a residual claimant on farm profits (Allen and Lueck, 1998; Pollak 1985). Recently, Kostov et al. (2016) have established limited empirical support for this conjecture.

The paper aims at contributing more detail to the debate on whether family farm is superior under modern conditions when technology, globalised food chains, and even policy requirements favour larger, often corporate farms. To test the empirical predictions of theoretical models, the analysis compared the efficiency distributions of family and corporate farms by decomposing total efficiency into two distinct components, namely F-efficiency, associated with the potential gains/losses derivable from the family form of organisation of farming activities, and M-efficiency, which measures the individual farm's managerial capabilities to realise any such efficiency gains. The results suggest that both effects are present to a certain degree. However, while greater family involvement (measured by family labour input) appears to always enhance F-efficiency, this is not so for M-efficiency. The M-efficiency does not show a regular pattern of variation with the level of family involvement. Furthermore, the relationship between the M-efficiency effect and the degree of family involvement is not common for all analysed EU Member States. For some countries (e.g. Hungary) a greater family involvement increases this effect while for others (e.g. Spain) it reduces it. For Romania the effect is larger for both small and high level of family involvement. These results suggest that the management capabilities that determine M-efficiency are most likely dependent on the nature of industry/country and as such can interact with the family involvement in a variety of ways, unlike the F-efficiency which is strictly increasing with greater family involvement. Therefore, it appears that the existing emphasis on the family effects, i.e. F-efficiency, that seems to underlie much of the literature on the superiority of family farming, is misplaced. It is only by decreasing, or in the best case removing the negative M-efficiency effects, i.e. by improving the management capabilities of family farms, that total efficiency gains could be achieved. These results imply that policy makers in Europe should focus more on enhancing management capabilities of family farmers if the aim is to strengthen family farming for economic and most of all for non-economic reasons – e.g. social and environmental.

This paper only considered four EU Member States – it would be interesting to see the similarities and differences across all EU Member States in order to discern a more consistent pattern across groups of European countries possessing a similar farm structure. In order to test the robustness of the results another approach, e.g. a parametric one, could be used.

References:

Allen, D and Lueck, D. (1998) The nature of the farm, Journal of Law and Economics, 41(2):343-386.

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

Behr, A. (2010) Quantile regression for robust bank efficiency score estimation, In European Journal of Operational Research, 200(2), 568-581.

Bernini, C., Freo, M. & Gardini, A. (2004) Quantile estimation of frontier production function, Empirical Economics, 29: 373-381.

Brookfield, H. and Parsons, H. (2007) Family farms: Survival and prospects. A World-Wide Analysis, Routledge: Abingdon, Oxon.

Cazals, C., Florens, J.P., Simar, L., 2002. Nonparametric frontier estimation: a robust approach. Journal of Econometrics 106, 1–25.

Chernozhukov, V. (2005) Extremal Quantile Regression, Annals of Statistics, 33 (2), 806-839.

Chernozhukov, V. and I. Fernandez-Val (2011) Inference for Extremal Conditional Quantile Models, with an Application to Market and Birthweight Risks, The Review of Economic Studies, 78 (2), 559-589.

Council of the European Union (26 July 2013). Family farming prospects in the context of globalization, Discussion paper. 12786/13, AGRI516. Available at: http://static.eu2013.lt/uploads/documents/Programos/Discussion%20documents/Informal_AGRI_DP.PDF.

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

Davidova, S. and Thomson, K. (2014) "Family Farming in Europe: Challenges and Prospects". In-depth Analysis of the AGRI Committee, European Parliament, Agriculture and

Rural Development. http://www.europarl.europa.eu/RegData/etudes/note/join/2014/529047/IPOL-AGRI_NT(2014)529047_EN.pdf

Davidova, S., Bailey, A., Dwyer, J., Erjavec, E., Gorton, M., Thomson, K. (2013) Semi-Subsistence Farming – Value and Directions of Development, study prepared for the European Parliament Committee on Agriculture and Rural Development. Available at: http://www.europarl.europa.eu/RegData/etudes/etudes/join/2013/495861/IPOL-AGRI_ET(2013)495861_EN.pdf

Djurfeldt, G. (1996) Defining and operationalizing family farming from a sociological perspective, Sociologia Ruralis, 36(3):340-351.

European Commission (2013). Agriculture and Rural Development Family Farming. Available at: http://ec.europa.eu/agriculture/family-farming/index_en.htm

European Parliament (EP) Committee on Agriculture and Rural Development (2016), Draft Report on how the CAP can improve job creation in rural areas, (2015/2226(INI)) Committee on Agriculture and Rural Development (6.04.2016), Available at http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+COMPARL+PE-580.547+02+DOC+PDF+V0//EN&language=EN

FAO (2013) 2014 IYFF FAO Concept Note (Modified May 9, 2013). Available at: http://www.fao.org/fileadmin/templates/nr/sustainability_pathways/docs/2014_IYFF_FAO_Concept_Note.pdf.

Fennell, R. (1997) The Common Agricultural Policy. Oxford: Clarendon.

Gasson, R. and Errington, E. (1993). The Farm Family Business. CAB International, Wallingford, UK.

Hendricks, W. & R. Koenker (1992) Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity. Journal of the American Statistical Association 87, pp.58-68.

Hill, B. (1993). The 'Myth' of the family farm: Defining the family farm and assessing its importance in the European Community. Journal of Rural Studies, 9(4), 359-370.

Kanamori, T. Suzuki, M. Sugiyama (2012) Statistical analysis of kernel-based least-squares density-ratio estimation, Machine Learning, 86 (3), 335-367.

Kostov, P., S. Davidova and A. Bailey (2016) Effect of family labour on output of farms in selected EU Member States: A non-parametric quantile regression approach, paper presented at 90th Annual Conference of the UK Agricultural Economics Society 4th-6th April 2016, Warwick available at: http://ageconsearch.umn.edu//handle/236358

Li Q, Racine J (2008) Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data, Journal of Business and Economic Statistics, 26(4), 423–434.

Martins-Filho, C. and F. Yao, (2008) A smooth nonparametric conditional quantile frontier estimator, Journal of Econometrics, 143(2), 317-333.

Pollak, R. (1985) A transaction cost approach to families and households, J. Econ. Lit, 23, 581-608.

Schmitt, G. (1989) Farms, farm households, and productivity of resource use in agriculture, European Review of Agricultural Economics,16 (2): 257-284.

Schmitt, G. (1991) Why is the agriculture of advanced Western economies still organized bb family farms? Will this continues to be so in the future? European Review of Agricultural Economics, 18(3-4): 443-458.

Valentinov, V. (2007) Why are cooperatives important in agriculture? An organizational economics perspective, Journal of Institutional Economics, 3(1): 55-69.

Vapnik, V. N. (1998) Statistical learning theory. New York: Wiley.

Wang, Y. and S Wang, (2013) Estimating α-frontier technical efficiency with shape-restricted kernel quantile regression, Neurocomputing, 101, 243-251.

Wang, Y. , S. Wang, C. Dang, and W. Ge, (2014) Nonparametric quantile frontier estimation under shape restriction, European Journal of Operational Research, 232(3), 671-678.