Comparative efficiency of family and corporate farms: does family labour matter?

Philip Kostov, Sophia Davidova and Alastair Bailey[1]

Abstract

We examine the comparative efficiency of family relative to corporate farms, using FADN data for the Czech Republic, Hungary, Romania and Spain. We estimate a non-parametric non-separable farm production function, and derive efficiency scores for both family and corporate farms. We decompose efficiency into two distinct sources - organisational differences and management capabilities. We find evidence for organisational efficiency gains from family farming, relative to corporate farming, and these appear to increase with family involvement. However, with regard to the management capabilities, family farms do not compare so favourably. Furthermore, family involvement does not seem to have any systematic effect on the efficiency derived from management capabilities. Our findings suggest that further investigation of the way family farms employ and build management capabilities is needed to substantiate any 'superiority' claims.

1. Introduction

Family farming around the world faces many challenges, the two critical ones being: access to resources such as land, capital and labour; bargaining power in the food chain and input 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.

---

[1]   Philip Kostov is the contact author (PKostov@uclan.ac.uk) and is in the Lancashire Business School, University of Central Lancashire, Preston, UK. Sophia Davidova and Alastair Bailey are both in the School of Economics, University of Kent, Canterbury, UK. They wish to thank three anonymous reviewers for their comments and suggestions. Any remaining errors are the responsibility of the authors.

Family farming has long attracted considerable attention from both researchers and policy makers. From its inception, the European Union's Common Agricultural Policy (CAP) has claimed family farmers as its main target group (Fennell, 1997). Yet there has never been an agreed EU definition on the concept (Hill, 1993), which reflects the multifaceted nature of the phenomenon. However, to analyse their relative efficiency, we need to be clear about what we mean by a family farm. The FAO, announcing 2013 as an International Year of Family Farming, 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 relevant aspects in this definition for us: a) the particular mode of management and operation of the farm, exemplified by considerations of household control and use of family labour; b) recognition that family farms may follow non-economic objectives. Nevertheless, family farming is often viewed as a business. "*Family farming is more than business but still a business*" (EC, 2013), conforming to the traditions of economic analysis, where family values can be described in terms of ownership, control and inheritance of assets (Gasson and Errington, 1993). This perspective is also evident in statements that: "a *major share of capital is built up by the manager and his or her family*" and that "*the major share of the family's income is derived from farming*" (Council of the EU, 26 July 2013).

Theoretical arguments for the apparent persistence of family farms have focused on the relative transaction costs associated with the employment of family and hired labour, and on the different incentive structures which motivate individuals. It is argued that 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. This implies that the average productivity of labour on family farms may be higher than for hired workers on corporate farms. (Pollak, 1985; Kostov *et al*., 2016; Valentinov, 2007).

However, Pollak (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 successful business performance. To these could be added the

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 the two opposing effects - the positive effect of more efficient monitoring and negative aspects of family farms discussed above - it cannot be said *a priori* that family farming is a more efficient form of organisation.

Transaction costs literature focuses on mechanisms and incentives that are better suited to achieve efficiency. In this study we ask how the outcome would change if these mechanisms and incentives vary. The extent to which the family farming exhibits a superior economic performance has important policy relevance since there is a claim that "support should be provided first asnd foremost to family farms" (COMAGRI, 2016). ?

We define a family farm on the basis of the use of family labour: any farm which employs family (unpaid) labour is defined here as a family farm. Our farm-level data is the EU Farm Accountancy Data Network (FADN), which allows the classification of farms according to the use of family or non-family labour.

We estimate a non-parametric non-separable farm production function, and derive the efficiency scores for both family and corporate farms. The distribution of these efficiency scores for the two types of farm organisation 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. We decompose the total efficiency effects into two distinct components: M-efficiency which measures the individual farm's managerial capabilities to realise any efficiency gains; 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. Section 2 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 provides the conclusions and policy implications.

2. Analytical Approach

We calculate a measure of technical efficiency for all farms and we compare the probability distribution of this technical efficiency measure for family and corporate farms, which we refer to as the (efficiency) density differential. 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. This density differential is essentially a difference between two distinct probability density functions (PDFs), generating negative as well as positive values. 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*. Furthermore, for ease of interpretation, we use a *relative* efficiency density differential. Suppose that, for a given value of the efficiency measure, the PDF value is 30% for family farms and 20% for corporate farms. The relative (of family vs corporate) density differential at this point is (0.3-0.2)/0.2=0.5, meaning that a family farm is 50% more likely to achieve this level of efficiency than a corporate one.

Using a standard approach to estimate an efficiency frontier and thus allocate efficiency scores to each farm imposes the same efficiency frontier for both farm types. We call this an A production function. Any and all differences (between family and corporate farms) are then fully absorbed in 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 decompose the efficiency density differential.

Family involvement (amount of family labour employed) in farming might have a transformative effect on farm productivity, where family involvement could modify the nature of the production relationship (i.e. lead to essentially different production function). Hence, we consider a production function that treats family and hired labour as different inputs (we denote this as a B production function). 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.

Since the A production function implicitly assumes that family labour has the same impact as hired labour and the B production function accounts for any transformative effects of family labour, the difference between A and B will measure this effect (if such exists).

These two different function formulations (distinguishing or ignoring family involvement) are used to estimate production possibilities frontiers and calculate the efficiency score for each farm. Let's consider the B frontier. For any given farm it already reflects the extent of family involvement (which is zero for corporate farms). The efficiency score measures the individual farm capability to achieve its typology dependent potential. We denote this hereafter as M-efficiency[2] which is a measure of the management and organisational capability of a farm. The efficiency score for the same farm evaluated against the A frontier ignores the farm type and this is the only source of difference between these two efficiency scores. All other potential contributors to farm efficiency are treated in the same way. Therefore the difference between the aggregate efficiency differential and the M-efficiency differential will account for the difference in output capacity between family and corporate farms. We term the latter the F-efficiency density differential.[3]

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 or 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 higher M-efficiency than 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

As explained above, our individual efficiency scores for each farm need to be estimated based on two different production functions: one that includes labour as a single input (A), and the other that differentiates family and non-family labour (B). The conventional approach to efficiency relies upon some form of frontier estimation and conceptualises efficiency in

---

[2] The notation M indicates Management capabilities.
[3] The notation F indicates organisational Form.

terms of distance between actual realisations and the estimated frontier. There are several estimation measures. The most basic and easy to interpret is the distance function, which is simply $y/\hat{y}$ (y being actual production, and $\hat{y}$ 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.

The estimated frontier in the stochastic frontier approach is a fitted production function plus disturbance (usually Gaussian and iid) term and plus an efficiency score shifter (term reflecting the distribution of the efficiency scores usually with a half-normal distribution). In effect, a mean production function is estimated, which is then 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, which is undesirable.

An alternative is to use a quantile regression approach, which does not require any distributional assumptions and hence does not restrict the distribution of the derived efficiency measures. 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, those 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 has the advantage that it does not impose the same input/output relationship on both efficient and inefficient farms. Non-parametric alternatives such as DEA are less susceptible to this problem, but lead to the other extreme, i.e. that inefficient farms have no contribution to the frontier estimation while outliers can have a major effect. In the limit, the quantile regression approach converges to a DEA in the case of non-parametric estimation of an extreme quantile.

The conditional quantile regression estimates 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 $\tau = (n-1)/n$, 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) note 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). These papers adopt a restrictive linear quantile model formulation, which requires a 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). The overall shape of the estimated relationship for conditional quantiles changes with the quantile. As a result, using two different quantile functions as 'frontiers', as in our case, 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. In principle, the finer the quantile, the more likely are the quantile functions to approximate the actual frontier, but the less reliable are the estimates of the functions. Less extreme quantiles are further from the actual frontier, but their shape estimation is more reliable.

There is a trade-off between the complexity of the production function model (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 0.95th 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. 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 are robust to outliers, similar to the *m*-output frontier models of Cazals et al. (2002).

We employ 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, referred 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.

These distance function efficiency scores are a measure of technical efficiency, where larger numbers denote better technical efficiency.

This 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 definition of the 'efficient' farms, 10 per cent of the actual farms are more efficient. The 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, ours are not limited from above. It would be tempting to 'standardise' these residuals to a [0,1] interval to resemble the standard efficiency measures. But it is more informative in our case to use the density function of the conditional efficiency scores 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 observations 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.

3.2. Quantile regression and density ratios estimation
In this paper 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:

$$y = f_\tau(X) + u_\tau \tag{0}$$

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

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 (0) 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).

Our estimates follow the indirect method of Li and Racine (2008), which 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: $pdf_{fam}(z)/pdf_{corp}(z)-1$, where $pdf_{fam}(z)$ and $pdf_{corp}(z)$ are the empirical probability density functions for the efficiency scores for the family and corporate farms. The efficiency scores themselves are calculated as $z = y/\hat{y} = 1 + \dfrac{u_\tau}{\hat{y}}$, 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 (0).

Therefore we need to estimate the density ratio $pdf_{fam}(z)/pdf_{corp}(z)$. 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. 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

Our data come from the EU's Farm Accountancy Data Network (FADN) for 2008 – near the year of accession of Central and Eastern European countries to the EU. 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 relatively balanced mix between corporate and family farms. Romania is the EU Member State with the largest number of 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, and as such they are not included in FADN (Davidova et. al., 2013). However, semi-subsistence farms affect the nature of the commercial farms which are the FADN field of observation since they present an important source of 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 (89.8 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 (both 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 being used as separate inputs in the production function B for the M-efficiency estimation are added together as a single input in the production function A for the 'total' efficiency estimation.

Capital is calculated as the total fixed assets, net of the value of land, permanent crops and quotas (all in monetary terms). 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.

The family farms are split into four groups, according to family to hired labour ratio, where higher values of this ratio signify farms with greater family involvement. Table 1 presents some summary statistics for corporate and family farms, with the latter grouped according to the degree of family involvement. The 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. 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 farms are larger in all countries. 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 quartile is actually lower than in the third, while for the other two countries the values in the third and fourth quartiles are almost the same. The decrease in total labour input from lower to higher quartiles explains this apparently anomalous situation, but raises the question of the most appropriate measure of family involvement: 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 farm 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			
		(<25%]	(25-50%]	(51-75%]	(>75%]
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

Our focus is investigation of the theoretical conjecture that the potential benefits of family farming are related to lower monitoring costs. Such benefits are more likely to accrue due to the amount of family labour rather than its share. The absolute amount of family labour represents the family commitment to and dependence on the farm operation. Higher level of commitment is hypothesised to lead to better organisation and 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 four EU Member States. 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 a negative F-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 corporate farms are superior, the opposite would be true.

A mostly downward-sloping total efficiency effect (ignoring the difference between family and hired labour) 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, in general, not as efficient as corporate ones. However, for Romania and the Czech Republic 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 farms.

The efficiency decomposition results are more subtle and varied amongst countries. For Romania and Hungary the M-efficiency differentials are relatively small, so the total efficiency differential follows the F-efficiency differential (which is derived as Total − M). In other words, the total production effect of family farms in these two countries is mostly characterised by F-efficiency. Hence, in these two countries there is not much apparent difference in the managerial capability between family and corporate farms, but family farms are generally not as efficient as the corporate ones, implying a lower capacity of family farms to achieve efficiency.

In Romania, at the top end of the efficiency range, the F-efficiency effect roughly offsets the negative impact of M-efficiency and as a result the total efficiency differential is close to zero. This provides some limited support for the superiority of the most efficient family farms

in this case. However, these family farms do not 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 poorer management capabilities.

Figure 1. Efficiency differentials by country
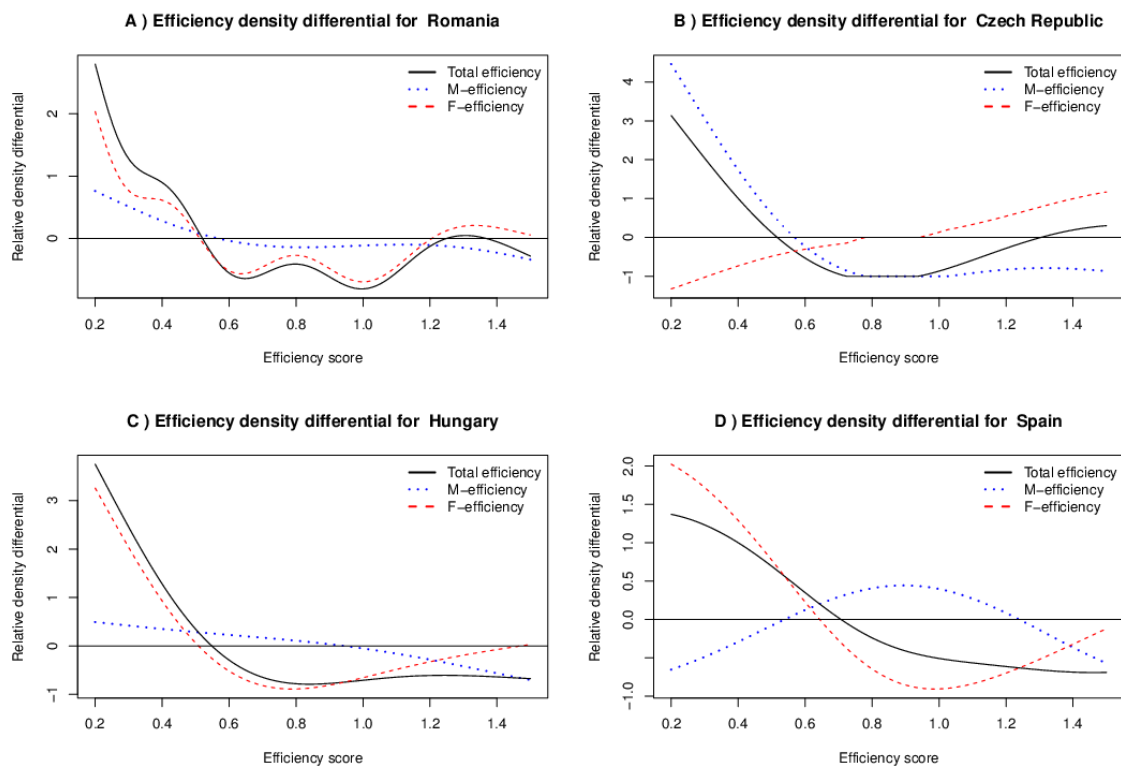
For Hungary, at the top end of the efficiency density range the M-efficiency effect appears to dominate and to account for virtually the entire efficiency differential.

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, the potential efficiency gains from family farming are not realised, being more than offset by M-efficiency losses (i.e. the lower capability of the family labour).

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 efficiency gains appear to be realised only above an efficiency threshold, 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 that these only materialise at a significantly high threshold of 1.75 AWU of family labour input. To re-examine these findings, we consider 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. 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 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 an upward slope 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.

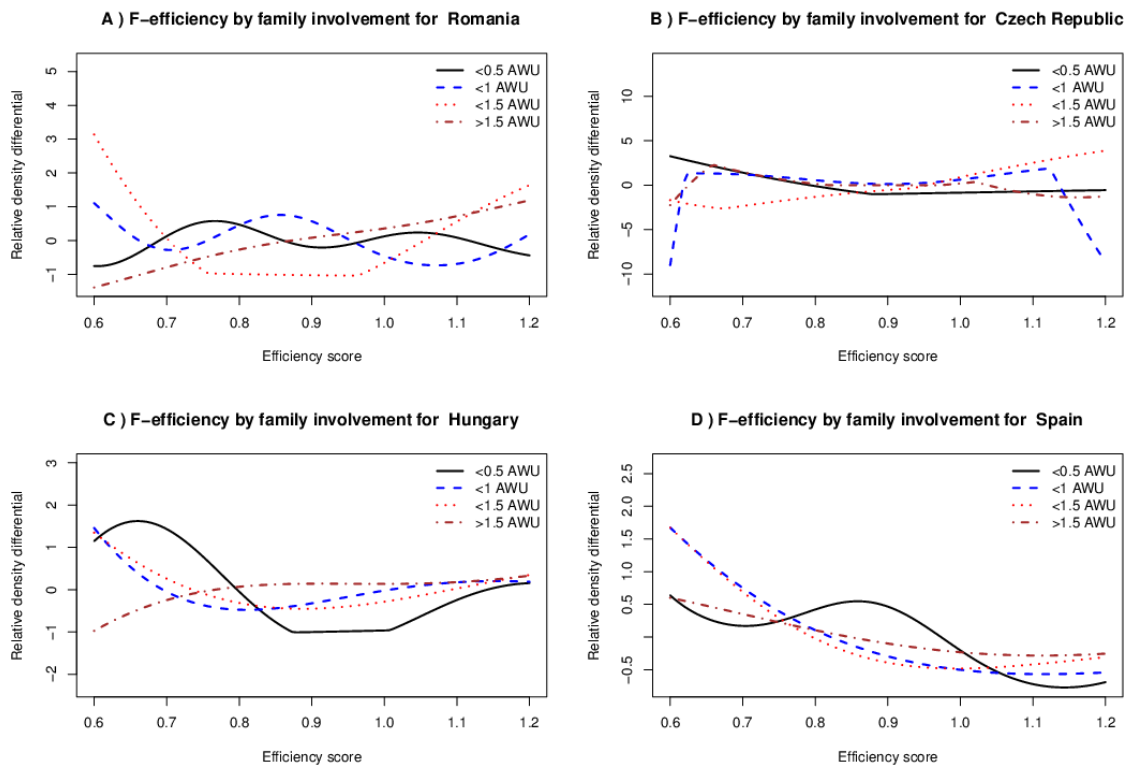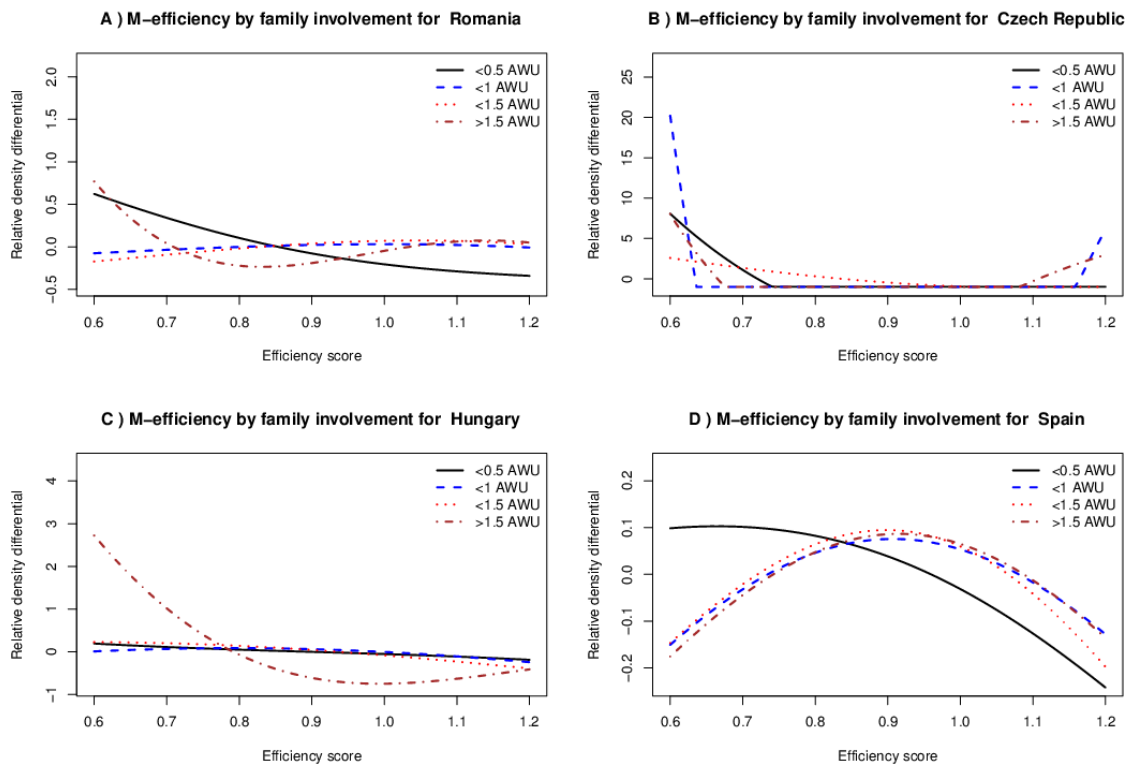## Figure 2. Changes in F-efficiency effects by family involvement

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

There are two features of the F-efficiency differential graphs (Figure 2) that are of particular interest: first, the relative ordering of these differentials according to the extent of family involvement, and second, the shape of these differentials.

Let us consider the relative ordering of the different efficiency differentials. F-efficiency, in general, appears to be enhanced by greater family involvement. This can be inferred from the fact that the corresponding efficiency differentials for farms with greater family involvement increase for higher levels of technical efficiency. This is consistent with the hypothesis that greater family involvement will lead to positive efficiency gains due to lower monitoring costs. However, in this study this particular effect can only be observed at a relatively higher level of technical efficiency - 0.9 and above. In other words, the positive effect of family involvement is stronger for more efficient farms.

The other characteristic feature is the actual shape of these differentials. Although in general the expected upward sloping efficiency differentials, at least at the higher efficiency levels, can be observed, this is not the case for Spain, where the F-efficiency differentials appear relatively flat at the higher efficiency levels. Due to the prevalence of family farms in Spanish agriculture, the estimated differentials are heavily dependent on characteristics of a small number of commercial farms, which could explain why any such differences are relatively small.

Overall, Figure 2 provides evidence of the presence of efficiency enhancing monitoring effects, associated with higher degree of family involvement.

Figure 3 presents how M-efficiency changes with the level of family involvement. Let us first consider the relative ordering of the M-efficiency differentials. Unlike F-efficiency, we cannot observe any regularity in the way these change with the level of family involvement In particular for Romania and Spain family involvement over 0.5 AWU improves M-efficiency, but this effect flattens quickly and the effects of all the other groups (0.5-1.0, 1.00-1.5 and over 1.5 AWU) are practically indistinguishable. In Hungary, on the other hand, too much family involvement (>1.5 AWU) reduces M-efficiency. Such finding should not be surprising, since M-efficiency measures managerial capabilities and the extent of family involvement may not be an appropriate measure for this. Greater family involvement could improve managerial capabilities (as we observe for Romania and Spain), but too much family involvement could seriously deteriorate them (see the result for Hungary and to some degree Spain).

With regard to the relative slope of the M-efficiency differential, again the findings are inconclusive. In the case of Romania, greater degree of family involvement results in upward sloping M-efficiency differential, suggesting a beneficial overall efficiency effect. However, in the case of Spain we observe downward sloping differential which suggest insufficient managerial capabilities, meaning that family farms fail to realise their potential.

6. Conclusions and policy implications

We examine 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 for family farming are centred on family labour, which could be more productive and involve lower monitoring costs than hired labour because of its motivation as a residual claimant on farm profits (Allen and Lueck, 1998; Pollak 1985). Recently, Kostov et al. (2016) provided limited empirical support for this conjecture.

To test the predictions of theoretical models empirically, we compare the efficiency distributions of family and corporate farms. We decompose total efficiency into two distinct components, namely M-efficiency, which measures the farm's managerial and organisational capabilities to realise efficiency gains, and F-efficiency as a residual from total efficiency, which measures the difference in production capacity between family and corporate farms

In other words, family farms might be expected to have the capacity to perform well (achieve higher levels of efficiency) because of better motivation and lower management (organisational) costs. But family farms might be limited in their capability to deliver outputs (and revenues) because of insufficient managerial capabilities. By employing a non-parametrically estimated non-separable production frontier to derive individual farm efficiency scores we allow family involvement (measured by family labour input) to have a potentially transformative effect on the production possibilities and this effect is captured by the F-efficiency differential.

The use of conditional quantile techniques produces robust frontier estimation thus accumulating any outlier effects in the efficiency scores. Summarising the efficiency scores via probability measure (i.e. their probability density), reduces the effect of outliers by simultaneously benefiting from the correct ordering of the efficiency scores due to the robustly estimated frontier.

The results suggest that greater family involvement improves F-efficiency, hence effectively increasing the production capacity of the family farm. However, the extent to which family farms can realise this potential, as measured by M-efficiency, varies from one country to another. We can tentatively suggest that such managerial capabilities initially increase with family involvement before the increase in family labour starting to have a detrimental effect. The presented decomposition of efficiency effects demonstrates that contribution of managerial capabilities (i.e. M-efficiency) to the aggregate efficiency is more important that this of the potential capacity benefits of family farming (i.e. F-efficiency).

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 the total efficiency gains could be achieved. These results imply that policy should focus on enhancing management capabilities of family farmers in order to allow them to benefit from the advantages of family labour input.

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.