One size does not fit all: an empirical investigation of the Romanian agriculture production function

Philip Kostov[1*], Sophia Davidova[2]

1 - Lancashire Business School, University of Central Lancashire, Greenbank Building, Preston, PR1 2HE, UK
2 - School of Economics, Keynes College, University of Kent, Canterbury, Kent, CT2 7NZ, UK

Contributed Paper prepared for presentation at the 90th Annual Conference of the Agricultural Economics Society, University of Warwick, England
4 - 6 April 2016

*Corresponding author: Lancashire Business School, University of Central Lancashire, Greenbank Building, Preston, PR1 2HE, UK. email:pkostov@uclan.ac.uk

Abstract
There are instances when one wants to consider homogeneous, with regard to some functional relationship, groups. For example in representative farm modelling one is interested in specifying groups of farms that have the same input/output relationship. This paper questions the logic behind the established approaches. It argues that one should use the underlying functional relationship to derive such groupings. Directly grouping farms with regard to their production function (or any other functional relationship of interest) not only asks the relevant question directly, but also makes the classification issue explicitly dependent on the choice of functional form. It provides a clear definition of what kind of homogeneity and/or representativeness were are looking at. If we want farms with similar production function either because this is the characteristic we are interested in or because we intend to model their production function in a follow–on simulation model, this is clearly the type of question we have to ask.
This paper proposes using finite regression mixture models to specify and estimate farm groups with regard to pre-specified functional relationship. We illustrate the proposed approach with regard to the production function. The properties and advantages of the proposed approach are discussed and explained.

Keywords: finite mixture models; production function
JEL code: C21; Q14

One size does not fit all: an empirical investigation of the Romanian agriculture production function

Motivation:

Economic theory has a long standing tradition of emphasising uniformity. After all the principle of the 'representative economic agent' is probably the best known theoretical abstraction in economics. Assuming such uniformity is very useful in deriving theoretical properties in that micro-economic models can be easily expressed into a common sense logic. Having the representative economic agent simply requires one to put herself/himself into the shoes of this agent and they deduce economic principles from generalising one's own response to external factors. This approach has been very fruitful in producing logical outcomes based on sound principles of rationality. Furthermore it also provides a basis for statistical investigation. To explain the latter consider the representative economic agent response to external factors, and let us assume for simplicity that this can be described by a known response function. Then since this concept is an abstraction and it is obtain just by averaging the reactions of the actual economic agents, the representative agent responses can be obtained just by averaging the responses of the real ones. Hence, although we cannot observe this, estimating a mean regression type of statistical model implicitly yields the response of the representative economic agent.

This uniformity principle, however useful has its limitations and has been questioned. From a theoretical point of view models of bounded rationality which combine two types of representative agents have been shown to be able to produce qualitatively different outcomes. For example De Long et al. (1990) present a model with rational agents and noise traders (who behave randomly and interact with the rational agents). One of the surprising outcomes of this model is the noise traders who non-intentionally (i.e. randomly) make very risky

investments may under certain conditions end up dominating the market. Kogan et al. (2006) further investigate this issue, which is now accepted in financial literature (see e.g. Cogley and Sargent, 2009; Le Baron, 2012; Luo, 2012).

Hereafter we will not be concerned with the theoretical challenges to this principle, but rather with empirical considerations that we outline below. A major problem in empirical research is the fact that theory rarely prescribes the form of the functional relationship between the variables in question. It is essentially not possible to know a forehead the functional form of this relationship. Hence the problem of 'representativeness' i.e. homogeneity in response becomes inter-wined with the issue of functional specification. There is clearly a trade-off here. Using more flexible functional representation reduces this problem, but also makes interpretation, and inference more difficult and in some cases impossible (as in the case of the curse of dimensionality problem). Using more restrictive functional representations results in more tractable models for which the representativeness assumption is more likely to be violated simply because the used functional representation is inadequate. Therefore the representativeness condition in empirical modelling is dependent of a given functional specification. In other words the question of whether the units of analysis exhibit the same relationship is only meaningful with regard to the given functional form of this relationship.

Modelling such heterogeneous responses has a long tradition in agricultural economics, particularly in the area of the so called 'representative farm modelling'. (may need some references here). This approach splits the units of interest (farms) into relatively homogenous groups and models these separately. Often the purpose of such modelling is to use the results for mathematical programing models for these different farm types. The way these groups are derived can however be problematic. Often some form of factor analysis or principal components analysis is applied with regard to farm characteristics to identify the

groups. The problem with this approach is that it yields group which a similar with read to the variables used in this analysis, not the functional relationship which is primary target of such approach.

The clustering approach is not entirely atheoretical. There is a long tradition in the hedonic literature dating back to at least Straszheim (1974, 1975) to identifying relatively homogeneous sub-markets. A point of departure for these early works is the premise that such groupings should be defined with regard to supply. The underlying logic of this approach is to define submarkets with regard to the concepts of substitutability and equilibrium. Note that in principle pairs of goods (or services) with similar characteristics are more likely to be close substitutes than pairs with dissimilar characteristics. The equilibrium process equates prices of characteristics across substitute goods. Therefore as long as submarkets are defined in such a way that they contain goods with similar characteristics, they should be close substitutes. This will ensure consistency of the prices within each submarket.

With regard to the farm classification issue as long as the characteristics used to perform clustering are elements of their cost of profit functions, the same argument applies and the resulting groups of farms should under a similar set of assumptions be homogeneous with regard to their cost (or profit) function, i.e. with regard to the functional relationship.

However, it is possible that quite different farms could have similar cost/profit/production functions yet their inputs may not be substitutes. To simplify the discussion we will focus from now on the production function, but our argument will be equally applicable to other functional relationships. It is therefore more logical to ask the question directly: which are the farms that have the same production function? This question could avoid some of the pitfalls of the clustering approach. For example let us assume that all farms have the same production

function, but there are two very different groups in terms of appearance: capital and labour intensive ones. Since the first groups will have high stock of capital and low labour usage while the other one will have lots of labour but little capital, clustering methods which are based on relative differences in some distance measure are likely to yield two separate farm groups. One may say this should not be a big issue because modelling them separately should yield the same production functions anyway. However in the presence of noise in the data, outliers and measurement errors, this may not be as straightforward. Even if none of these empirical issues exist, using the different functional specifications for the production functions of the already defined clusters can produce very different outcomes (to whether their production functions are similar or not).

Directly grouping farms with regard to their production function (or any other functional relationship of interest) not only asks the relevant question directly, but also makes the classification issue explicitly dependent on the choice of functional form. It provides a clear definition of what kind of representativeness were are looking at. If we want farms with similar production function either because this is the characteristic we are interested in or because we intend to model their production function in a follow–on simulation model, this is clearly the type of question we have to ask. A clustering type of approach in contrast asks a very different question. It asks how similar the farms appear with regard to some predefined characteristics. Such a question leaves the issue of 'representativeness' very vague. It also implicitly claims a kind of logically inconsistent universality. For example one may use some set of 'relevant variables' to cluster farms and them assume that the functional relationship is homogeneous within each cluster. But the same approach could be applied to a wide range of relationships such as e.g. cost, profit and production functions. So the farms in the same cluster are assumed to have the same type of functional relationship for all of the above (and more). Surely this is a very heroic assumption.

This paper proposes using finite regression mixture models to specify and estimate farm groups with regard to pre-specified functional relationship. We illustrate the proposed approach with regard to the production function. The properties and advantages of the proposed approached are discussed and explained.

Background

In this paper we look at the issue of the production function that underlies the Romanian agriculture.

There is the inheritance of transition that could potentially create quite heterogenous farming structure that incorporates the successors to the old cooperative farms and emerging modern farms. Strictly speaking adapting pre-existing farming structures carries forward a set of constraints that can restrict the possible production responses that such legacy farms can exhibit. Starting anew on the other hand does not imply such restrictions and then potentially could result in drastically different technological relationships, which will result in production responses that are not alike those of the pre-existing farms. Furthermore a rather turbulent transition period could be characterised by a series of shocks and then establishing a new farm could be quite different depending on when exactly one establishes such resulting in potentially even more diversity in underlying technologies.

Finally there is another more practical consideration. Economic analysis is often based)as we will do here) on aggregate relationships, which undoubtedly contain unobserved heterogeneity. For example when we look at the issue of production function, since technologies are very different for different farm typologies, it is reasonable to consider different production functions for different types of farming typologies, e.g. livestock, crop etc farms. Yet doing so results in large number of underlying models without actually solving the problem of unobserved heterogeneity since even within a certain typology,

different technologies could co-exists, based on characteristics that are not directly observable. Note that clustering approaches have no chance when trying to deal with such unobserved heterogeneity due to the very fact that since it depends on unobservable characteristics, it is impossible to account them during a classification stage. Therefore from a purely practical point of view we are looking at the following trade-off: we would want a small number of functional relationships (ideally a single one) but on the other hand would want these relationships to encompass both the similarities and differences amongst the units of interest. In other words subject to the constraints defined by the choice of functional relationship (in this context production function in terms of functional form and components) we want the best combination of (possibly more than one) such functional relationships that describe the data. Hence the question becomes: how many distinct production functions can describe the output response of Romanian agriculture and what are their characteristics. In this way we not only provide a characterisation of the agricultural economy, but also simultaneously determine their behaviour.

Methodology

The finite mixture of regressions can be considered a generalisation of the mixture of distributions model (MDM). To illustrate the general structure of a mixture of distributions model (MDM), let us denote the set of n d-dimensional vectors comprising the available data by $y = \{y_1, \ ,y_n\}$ (i.e. the sample contains n observations and d variables). In this application we will use a single y variable (*i.e.* d=1), but it will be useful to present the multidimensional case since its treatment is essentially the same. Additionally, it allows one to demonstrate the generality of the used approach. It is assumed that each y_i arises from a d-dimensional probability distribution with the following density:

$$f\left(y_i|\theta\right)=\sum_{k=1}^{K}p_k g\left(y_i|\lambda_k\right) \tag{1}$$

where p_k are the mixing proportions ($0<p_k<1$ for all k and $\sum_{k=1}^{K}p_k=1$), and $g\left(y_i|\lambda_k\right)$ is some (d-dimensional) probability distribution, parameterised by λ_k. This means that y can be viewed as drawn from K different underlying probability distributions. The 'standard' application of the MDM uses the unconditional distribution of y_i. When the conditional, with regard to some set of (explanatory) variables $\boldsymbol{X}$, distribution is used instead we essentially obtain a mixture of regressions:

$$f\left(y_i|\theta,X\right)=\sum_{k=1}^{K}p_k g\left(y_i|\lambda_k,X\right) \tag{2}$$

In this representation the parameters λ_k include a fully parameterised regression model, i.e. they include regression coefficients, as well as the distribution parameters. In this study we will use a linear regression specification (as introduced in De Sarbo and Cron, 1988; see also Wedel and Kamakura, 2001) but in principle any other parametric specification, i.e. a mixed model could be used instead. As we will demonstrate later the nature of the estimation algorithm is very general and allows for a wide range of specifications which may also be non-parametric. Equation (2) states that the data-generating process for y, conditional on X, is a mixture of regressions. Thus if y is the output and X are the production factors, this expression provides us with a basis for using the production function to define separate types of farms, which are characterised by distinct production functions.

A sample of indicator variables $z = \{z_1, \ldots, z_n\}$, sometimes referred to as *labels*, can be assigned to the observed data. These are defined as: $z_i = \{z_{i1}, \ldots, z_{iK}\}$, where each z_{ik} assumes the value of 1 if y_i arises from the k-th mixture component and the value 0, otherwise. When the

sample of indicator variables is known the problem is one of discriminant analysis, where the vector of parameters to be estimated is $\theta = (p_1, .. p_K, \lambda_{11}, .., \lambda_{1J}, .., \lambda_{K1}, .., \lambda_{KJ})$, where λ_{ij} denotes the *j*-th parameter of the *i*-th regression model. When the indicator variables are unknown the problem is one of cluster analysis or conditional density estimation. In the cluster analysis case, in contrast to the density estimation, we are also explicitly interested in estimating the indicator variables.

One can obtain the maximum likelihood estimate for the parameters, θ, by using the Expectation Maximisation (EM) algorithm of Dempster et al. (1977) and then applying the 'maximum a-posteriori' (MAP) principle to assign a value to the indicator variables, z_i. The EM is standard algorithm for estimating MDM.

The EM algorithm used in the analysis consists of the following two steps, namely, the E(xpectation) step and the M(aximisation) step. In the E step the conditional probability of z_{ik} being equal to one, estimated during the *m*-th iteration for all *i* and *k* is given by:

$$t_{ik}^{(m)} = t_k^{(m)}\left(y_i \middle| \theta^{(m-1)}, X\right) = \frac{p_k^{(m-1)} g\left(y_i \middle| \lambda_k^{(m-1)}, X\right)}{\sum_{l=1}^{K} p_l^{(m-1)} g\left(y_i \middle| \lambda_l^{(m-1)}, X\right)} \tag{3}$$

where the (bracketed) superscripts denote estimates for the parameters during the corresponding iteration.

In the M step the ML estimate, $\theta^{(m)}$ of θ, is updated using the conditional probabilities, $t_{ik}^{(m)}$, as conditional mixing weights. This leads to maximizing:

$$F\left(\theta \middle| y, t^{(m)}\right) = \sum_{i=1}^{n}\sum_{k=1}^{K} t_{ik}^{(m)} \ln\left(p_k g\left(y_i \middle| \lambda_k, X\right)\right) \tag{4}$$

The updated expressions for the mixing proportions are given by:

$$p_k^{(m)} = \frac{\sum_{i=1}^{n} t_{ik}^{(m)}}{n} \tag{5}$$

The updating of λ_k depends on the parametric specification and therefore, no general formula can be given. Note however that while the expectation step (3) is pretty standard, the maximisation step (4) depends on the particular statistical model assumed, i.e. type of conditional distributions and type of conditioning (e.g. linear regression). The maximisation step is essentially the standard maximisation routine used to estimate the conditional model given some fixed (determined in the expectation step) mixing proportions. The generic expression (4) expresses calculating the log-likelihoods for each separate component and maximising the weighed likelihood with weights given by the posterior probabilities $p_k^{(m)}$. Thus by adapting the maximisation step a wide range of models could be fitted. I.e. if we have an estimation routine for any type of statistical model by plugging this estimation routine in the M-step of the EM algorithm we can obtain (and estimate) a mixture of such models. In particular we could be interested in a general mixture of regressions sometimes referred to as latent class regression or cluster-wise regression. Using a generic term, where under 'regression' we understand any parametric (or even nonparametric) model that provides us with a parameterised conditional distribution, we can consider our data as coming from a mixture of such models. In particular in this application we are interested whether a single parameterised in a specified way production function can be applied to the whole land price dataset, or several submarkets exhibiting different pricing functions could be identified.

So far we have considered estimating a mixture model for the purposes of classifying observations into a pre-defined number of distributions (sub-samples or clusters). In other

words the above estimation procedure only works with predefined fixed number of clusters. However, the number of clusters is typically unknown. Choosing the appropriate number of mixing distributions (clusters) is essentially a model selection problem. One can estimate the regression mixture models for different number of clusters and then selects amongst these.

A popular criterion in model selection problems is the Bayesian Information Criterion (BIC) (Schwarz, 1978).

$$\mathrm{BIC}_{mK} = -2\, L_{mk} + \nu_{mK} \ln(n) \qquad (6)$$

where m is any model (thus m denotes the choice of the parametric (conditional) distributions $g(.)$ or any combination thereof, K is the number of components, L is the (maximised) complete log-likelihood and ν is the number of free parameters in the model. If the choice of $g(.)$ is taken for granted, then (5) suggests a strategy of consecutive estimation of (m, K) models for K=1,2, until BIC increases. It is clear that if (m, K) and $(m, K+1)$ provide essentially the same fit then the BIC for (m, K) will be smaller, since it has less free parameters. The consecutive estimation strategy also ensures against the danger of over-fitting the statistical model (2).

We will use the BIC as a main model choice criterion, although details on some alternatives will also be provided. The reasons for our choice are outlined below. The BIC is based on an asymptotic approximation of the integrated log-likelihood, valid under some regularity conditions. In spite of the fact that these regularity conditions are usually violated in mixture models, it has been proven that the BIC is consistent under some conditions and efficient on a practical ground (e.g. Fraley and Raftery, 1998). Moreover the whole class of penalised likelihood estimators (of which the BIC is a special class) are consistent (Keribin, 2000). Using the BIC can be viewed as an approximate Bayes factors inference. The BIC is

furthermore approximately equivalent to the popular in information theory Minimum Description Length (MDL) criterion.

Alternatively one may use cross-validation to select the appropriate model. This would however substantially increase the computational burden.

If one needs to select of model where in addition to the model fit the ability to define well separated clusters is taken into account, the integrated complete likelihood (ICL) criterion can be used. The ICL can be expressed (Biernacki *et al*., 2002) as BIC with an additional entropy penalty term as follows:

$$\mathrm{ICL}_{mK} = -2\,BIC_{mk} - 2\sum_{i=1}^{n}\sum_{k=1}^{K} z_{ik} \ln t_{ik} \qquad (7)$$

where the conditional probabilities t_{ik} and the cluster membership indicators z_{ik} are defined as in (3). Depending on the purposes of applying a finite mixture model, often one may not be explicitly interested in the degree of separation of our clusters. Nevertheless, applying the ICL can be used as an illustration to how clustering criteria can lead to different results.

One can note that the mixture models with increasing number of components can be analysed in a nested models framework. Therefore the Likelihood Ratio (LR) test can be readily applied to consecutively test for number of components. In order to provide a valid small sample inference the distribution of the LR tests statistic can be simulated via bootstrap. In order to do this we need to simulate under the null, estimate the model under the alternative and calculate the LR test statistic. The estimation step needs a number of different initialisation (starting) points to avoid local optima. This means that such a bootstrap approach will be very expensive in computational terms. For this reason instead of using it in

a consecutive manner, we will only implement it to additionally test the model selected by the information criteria.

The finite regression mixture approach describes the functional relationship as am hierarchical mixture model, where the data generation process generates each observation from a finite set of underlying sub-models, which define separate clusters. By definition these clusters represent different functional relationships, in this case different production functions. Hence we define the representativeness condition directly with regard to the production function, conditional on its functional form. Note that technically the data generation process assumed in Bayesian hierarchical models, in particular model averaging (bot Bayesian and frequentist) as well Bayesian variable selection models follows the same hierarchical mixture representation. The advantage of the finite mixture approach is the ease by which data observations can be attached to the different underlying production functions.

Data and choice of functional form

We use data from the 2008 FADN for Romania. The implementation of CAP creates methodological issues about how to treat the CAP subsidies and by choosing the year immediately after accession we hopefully avoid some of these issues. The key question is whether the Romanian farms can be described by the same production function. As already discussed this question requires us to specify the inputs and the functional form for the specific production function. Here we will use the translog functional specification.

There is extensive literature on the issue of the productions functions and their theoretical and empirical properties (Griliches and Ringstad, 1971; Berndt and Christensen, 1973; Christensen and Lau, 1973). With regard to the problem in hand, it is advisable to employ a

production function specification that is sufficiently flexible, since in a finite regression modelling framework we advocate there is a clear trade-off between flexibility and the potential number of farm groups in that more flexible functional forms will reduce the number of farm clusters. In the production function literature the term ‘flexible’ has a specific meaning. According to Diewert (1974) a functional form can be denoted as ‘flexible’ if its shape is only restricted by theoretical consistency. The unrestricted Translog non-homothetic and imposes no restrictions of the production technology. It can be restricted to satisfy the homotheticity, homogeneity or separability, but in this application we will not apply any such restrictions. The main reason for this is that by avoiding such restriction we can maintain its generality. Furthermore, as our previous argument demonstrates there is a clear trade-off between flexibility and the potential number of clusters in that more flexible specifications would result in smaller number of clusters. Therefore when question is whether a single production function specification is sufficient to describe the data, it makes sense to avoid imposing restrictions that could inflate the potential number of clusters.

Although in more recent research the translog appears to have somewhat fallen out in favour with empirical researchers, it is still the most extensively investigated second order flexible functional form and surely the one with the most empirical applications as its empirical applicability in terms of statistical significance is outstanding (Feger, 2000). Furthermore the fact that the translog function can be considered as a second order (Taylor series) approximation of a more general production function provides a sound justification in applying it here, since the uncertainty about the production function is major justification for the present proposal.

An important reason for the popularity of the translog specification is that it is linear with regard to the parameters, which means that standard linear regression techniques can be used

for estimation and testing purposes. In principle estimating a finite regression model simply requires plugging in the M step an estimation routine for the underlying model, which creates tremendous flexibility since this means that the underlying model can be fully non-parametric. Linear specifications offer considerable savings in terms of computational costs.

Results

Both the BIC and the ICL point to two clusters (see Table 1). Bearing in mind that the BIC is a model fitting criterion expressing the fit of the statistical model, a single common translog production function is not sufficient to describe the Romanian farms. Furthermore taking into account that the ICL accounts for both model fit and cluster separation, the fact that the ICL also points to two cluster model demonstrates that these two clusters are well separated. In practical terms this means that (at least some of) the corresponding coefficients are significantly different resulting in two quite different production functions (subject to the functional restriction of a translog functional form)

In order to confirm the above conclusion we implemented LR bootstrap tests for 2 mixtures (clusters). Such tests are based on model fitting like the BIC and do not take into account the cluster separation (as ICL does) and as such are only comparable to the BIC results. However since BIC and ICL agree on the number of clusters, this technical difference is not crucial in this particular implementation. The probability levels for these are shown in Table 2. These bootstrap tests are very costly in terms of computational time and in most applied work one could expect that information criteria would be preferable to determine the number of mixture components. The LR bootstrap tests agree with the information criteria in that the Romanian farms can be split into two distinct clusters with regard to their underlying production function.

Table 3 presents the estimation results, while Table 4 shows the summary statistics for the used variables, both for the overall sample and by cluster. In order to facilitate the discussion the summary statistics in Table 4 are for the raw variables (rather than their logarithmic transformation which is used in specification and estimation). Cluster 1 is smaller with 296, while Cluster 2 consists of 574 observations. Cluster 1 contains bigger farms. Comparing the mean values for the two clusters, the only input for which Cluster 2 has larger values is land. Hence in general we can say the first cluster is characterised by larger farms. The larger average value of land in the input mix of the farms in the second cluster suggests that these might use a production technology that is much more land intensive, something that the estimation results might throw a more light on.

It is difficult to ascertain the differences between the clusterwise production functions given in Table 2, due to their non-linear form. A reliable way to compare two nonlinear functions is by comparing their partial correlation plots. It simple terms this amounts to using the estimated models to predict the dependent variable and plotting the predicted values against the values for a given factor by keeping the other factors values fixed at 'typical' values. In this way one can visualise the effect of a given production factor when the rest of the inputs are kept fixed. The main issue here is what would be reasonable values to fix the other inputs at. This would depend on the purpose of the above plot. Often one is interested in average effects and in such cases using simply the averaged over the estimation sample values would be an easy way to achieve this. Sometimes averaging would not be a reasonable strategy. See e.g. Kostov et al (2008) who discuss the different options and in particular the pitfalls of averaging discrete values and suggest alternatives for creating reference points to use for such comparisons. In this case all the inputs are continuous variables and we want to compare the two clusters production functions. Therefore averaging over the estimation sample is a viable option.

Note that although the translog is a non-linear functional form, from estimation point of view it is still linear in parameters and technically it is a linear regression model. Therefore predicting from it is very easy. We simply need to create a prediction sample containing a range of values for input variable the effect of which we want to investigate and the 'typical' values for the other inputs, create the relevant (transformed) variables needed in the translog specification and predict from the estimated linear model. The only choice we need to make is the range of values for the analysed input. We use a regular grid of 100 points defined over the range over which the input in question is observed. Since the two clusters are quite different in their input mixes (see table 4) it is reasonable to produce separate such ranges for each cluster. In this way the values for the variable of interest are actually observable within the estimation sample. The resulting plots show the range of values for each input by cluster and this facilitates the interpretation of the results. It also avoids the danger of predicting outside the range each of the two clusters is defined over. In particular, since Cluster 2 consists of smaller farms, we would not want to plot predictions of what its production function would yield for very large farms, since it simply would not be applicable to them. As for the variables over which any such plot is conditioned upon (i.e. the other inputs) averaging over the whole sample is applied in order to ensure that the effects plotted for the two clusters are comparable (since all the rest is being equal. Since the summary statistics for both clusters exhibit considerable dispersion, it is easy to verify that such common 'typical' values lie comfortable within the range of observable values for each of the two clusters and therefore the synthetic observations that we create in order to produce the effects of interest are entirely feasible.

Simply plotting the effects for each input can provide a useful overview of the differences between the corresponding production functions. However the usefulness of such a comparison would be limited if we cannot say how different are these statistically. The latter

can be accomplished if we can provide confidence intervals for such effects. These can be obtained by bootstrapping the corresponding models. Here we will follow Kostov et al. (2008) in using nonparametric case bootstrap.

The partial correlation plots for the inputs are presented in Figures 1-4. In producing these we have transformed both output and the input back into the original units, since results for the logarithms are not amenable to meaningful interpretation.

Due to the non-linear nature of the model we obtain asymmetric confidence intervals. The first noteworthy feature of these figures is that cluster 2 is considerably more homogenous in terms of the underlying production function in that the confidence intervals for the effect of all inputs are narrower than those for cluster 1. Although this on its own is not that surpising given the larger dispersion in the underlying inputs (see standard deviations in Table 4), the latter by no means guarantees a higher dispersion of the estimated effects. This difference in the homogeneity means that the farms in cluster two are much better characterised by their underlying production function than those in cluster 1. Taking into account that there are actually considerably more farms in cluster 2 and that cluster 1 farms are larger, it looks like the growth in farm size could be responsible for farms moving away from a common production function. The other important results is that these difference in the form of different width of the corresponding confidence intervals, but also in terms of underlying mean effects for each separate input, are quite unevenly distributed amongst the different inputs. We comment on these difference below. Furthermore in order to better explain any such difference we have calculated the own elasticities derivable form the estimated translog specifications for both clusters and these are presented in Table 5. For comparison purposes it also includes overall elasticities are calculated from a common (single cluster) translog specifically applied to the full sample. We calculated the own elasticities for each farm and

the mean values and their standard deviations are summarised in table 5. Since the elasticities are in fact properties of the underlying production functions, they can be used to complement the partial correlation plots effects.

Let us first consider the capital input. Cluster 1 employs more capital (Table 4) than cluster 2 uses wider range of capital inputs (Table 4 and Figure 1). Furthermore cluster 1 is also more capital intensive in that in manages to extract considerably more output from the capital it employs. This can be ascertained from the fact that the average contribution of capital to output is higher for cluster 1 over the whole range of capital values. Taking into account the associated confidence intervals, which do not overlap, the difference in these effects is statistically significant. One can also note that the confidence intervals for the effect of capital in cluster 2 are quitter narrow showing that cluster 2 consists of farms which are very homogeneous with regard to the contribution of capital to their output. In contrast to this the corresponding confidence intervals for cluster 1 are considerably wider. If we compare the (own) elasticity of capital, it is higher in cluster 1 (Table 5), which also shows from the graph in that the slope of its production curve is steeper for cluster 1. Yet quite interestingly both the mean values and standard deviation for the capital elasticity in cluster 1 coincide (subject to of course rounding error) with those derived from a single cluster full sample estimation. Taking into account that there are smaller number, although much larger farms in cluster 1, this shows that this cluster defines the role of capital in Romanian agriculture.

With regard to labour, again cluster 1 is characterised by larger farms employing both more labour and having a wider range of labour inputs (Table 4). Note however that on average the labour/capital mix is not that different between the two clusters (which can be inferred by dividing the average values for labour and capital in Table 4 and comparing the ratios). However, contrary to the case for capital, the farms in cluster 2 make much better use of

labour in that they manage (except for the very small farms) to extract considerably more output per unit of labour employed (Figure 2). Hence we can view the sector 2 farms as more labour intensive. Intriguingly the average labour output elasticities for the two sectors are rather similar (Table 5). Once again the dispersion of the labour effects looks larger in cluster 1, but if we look at the width of the confidence interval at similar values for the labour input and in particular for values observed over the larger farms in cluster 2 these are actually of similar magnitude. So unlike any of the other inputs contibutions, we cannot claim that labour effects are more homogenous in cluster 2.

Although cluster 1 in general uses less own land (Table 4), the output from the two clusters with regard to own land is not statistically different (Figure 3). While cluster 2 appears to be more land intensive in terms of both the slope of its partial effect, as well as its own elasticity (Table 5), this effect does not appear to be statistically significant, mainly due to the large dispersion of the land effect in sector 1. We used here the monetary value for land and therefore in this way accounted for the potentially different productive capacity of land. Owing to the latter one should expect that the partial effect of such value measured input would be similar across clusters.

The difference between the land effects raises an interesting possibility. It might be that farms manage better own land and hence they could obtain higher output form it. If this is the case cluster 1 could be heterogeneous with regard to this how well the corresponding farms manage rented in land. It is not possible in the present model to capture such heterogeneity. Furthermore the lack of separation between the partial land effects could be also due to the flexible functional form that leads to a smaller number of clusters.

Finally consider the effects of IC. These mirror the case of capital. Cluster 1 comprises of larger farms, which are relatively more productive, both in terms of the average output they

can extract from IC , but also that this output effect is statistically larger than this attributable to farm sin cluster 2. Similarly to the case of capital cluster 2 shows considerably homogeneity with regard to this effect. Furthermore the examination of the clusterwise and overall own elasticities mirror the case of the capital input in that cluster 1 dominates in defining the total contribution of IC in Romanian agriculture.

Conclusions

This paper uses unrestricted translog production function specification to cluster Romanian farms. Our results suggest that there are (at least) two very different farm types with distinct production functions. The larger cluster contains relatively smaller farms, which however cultivate larger land areas. In addition to this they are more labour intensive in that they extract more output form their labour input. The other cluster consist of smaller number of relatively larger farms whose production function is more capital intensive in that they manage to make better use of their capital and intermediate consumption. This split alongside the capital-labour trade-off and in particular the much greater heterogeneity that we observe with regard to the smaller capital-intensive cluster suggests a traditional vs new farming technologies explanation of our results. In particular this mean that more traditional farming structures, most likely inheriting the technological constraints of the pre-transition era are identifiable with the labour intensive sector. There is however also a new emerging capital-based agriculture. The latter is considerably more heterogenous in terms of its production technology. It is therefore clear that the differences between these two types of farms are consistent with the legacy of central planning and emergence of new commercial farms during transition. There are two important implications of the above farming structure. First, the rise in global food prices would intensify the process of structural transformation exemplified the emergence of capital intensive farms. Since the aggregate production

function of Romanian agriculture can be viewed as weighted average of the tow underlying 'technologies' this essentially means a transition from the more labour intensive into the more capital intensive cluster.. Such a transformation could also perhaps surprisingly avoid the detrimental effects of such transition on overall employment due to the fact that it does not entail the classical 'substitution of capital for labour'. Note that in terms of their input mix (i.e. the ratio of capital to labour) the two types of farms are very similar which means that transformation of traditional into new farms will not replace labour with capital, but essentially 'upgrade' capital with more productive one.

References:

De Long, J. B. , Shleifer, A., Summers L. H., and R. J. Waldmann (1990) Noise Trader Risk in Financial Markets, *Journal of Political Economy* 98, 703-738.

Kogan, L., S. Ross, J. Wang, and M. Westerfield *(*2006*),* The Price Impact and Survival of Irrational Traders, *Journal of Finance,* 61, 195–229.

Luo, G. Y. (2012) Conservative traders, natural selection and market efficiency, *Journal of Economic Theory*, 147 (1), 310-335.

LeBaron, B., (2012) Heterogeneous gain learning and the dynamics of asset prices, *Journal of Economic Behavior & Organization,* 83 (3), 424-445.

Cogley, T., T. J. Sargent (2009) Diverse Beliefs, Survival and the Market Price of Risk, *The Economic Journal*, 119, 354-376.

Dempster, A.P., N.M. Laird, and D.B. Rubin (1977) Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society (B)*, 39, 1-38.

Keribin, C. (2000) Consistent estimation of the order of mixture models, *Sankhya,* SeriesA 62(1), 49-66.

McLachlan, G.J. and D. Peel (2000) *Finite Mixture Models*, New York: Wiley.

Schwarz, G., (1978) Estimating the dimension of a model, *Annals of Statistics*, 6(2), 461-464.

Straszheim, M.R. (1974) Hedonic estimation of housing market prices: A further comment. *The Review of Economics and Statistics* 56 (3): 404-406.

Straszheim, M.R. (1975) *An Econometric Analysis of the Urban Housing Market.* New York: National Bureau of Economic Research.

De Sarbo W.S. and W.L. Cron (1988) A maximum likelihood methodology for clusterwise linear regression, *Journal of Classification*, 5, 249–282.

Wedel, M. and W.A. Kamakura (2001). *Market Segmentation - Conceptual and Methodological Foundations*, International Series in Quantitative Marketing, Kluwer Academic Publishers, Boston, MA, 2nd edition.

Table 1 Information Criteria Results for number of clusters

Number of clusters	BIC	ICL
1	2931.021	NA
2	2831.105	2931.105
3	2849.514	3783.277
4	2864.836	3931.209
5	2908.899	3906.923
6	2956.215	4279.229

Table 2. Bootstrapped LR test (5000 replications)

Test	P value
2 (NULL) vs 1 clusters	0.72
2 (NULL) vs 3 clusters	0.17

Table 3. Estimated translog for overall sample and clusters

	All data		Cluster 1		Cluster 2	
	Coefficient	P-value	Coefficient	P-value	Coefficient	P-value
(Intercept)	9.68	0.00	10.64	0.00	6.77	0.00
capital	-0.21	0.01	-0.33	0.06	0.08	0.20
labour	1.01	0.00	1.07	0.06	0.49	0.00
land	0.57	0.00	0.72	0.00	0.25	0.00
ic	-0.37	0.00	-0.45	0.10	0.08	0.18
I(0.5 * capital^2)	0.03	0.00	0.03	0.00	0.01	0.00
I(0.5 * labour^2)	0.05	0.16	0.00	0.26	0.31	0.00
I(0.5 * land^2)	0.13	0.00	0.14	0.00	0.16	0.00
I(0.5 * ic^2)	0.06	0.00	0.06	0.14	0.02	0.00
capital*labour	-0.03	0.05	-0.06	0.10	0.04	0.00
capital*land	-0.02	0.00	-0.03	0.01	-0.01	0.00
capital*ic	0.01	0.07	0.02	0.16	-0.01	0.12
labour*land	-0.04	0.00	0.02	0.03	-0.20	0.00
labour*ic	-0.02	0.20	-0.01	0.42	-0.02	0.01
land*ic	-0.05	0.00	-0.07	0.00	-0.01	0.06

Note: variable labels refer to variables in natural logarithms (i.e. capital is the natural logarithm of the capital variable)

Table 4 Summary statistics for the clusters

	Cluster 1			
	mean	sd	min	max
Output, 000s	1,530	22,563	0	506,143
capital, 000s	472	2,219	0	37,216
labour	14	54	0	680
land	219	757	0	11,196
ic, 000s	763	9,502	0	212,143

	Cluster 2			
	mean	sd	min	max
Output 000s	225	660	1	9,978
capital, 000s	197	732	0	15,334
labour	6	14	0	142
land	306	1,211	0	21,565
ic, 000s	182	980	0	23,479

Table 5 Elasticities (own)

	capital	labour	land	ic
All data				
Average	0.13	0.44	0.26	0.18
SD	0.06	0.12	0.26	0.13
Cluster1				
Average	0.13	0.44	0.23	0.19
SD	0.06	0.12	0.25	0.13
Cluster2				
Average	0.10	0.44	0.40	0.10
SD	0.05	0.39	0.30	0.03

Figure 1. Effect of capital

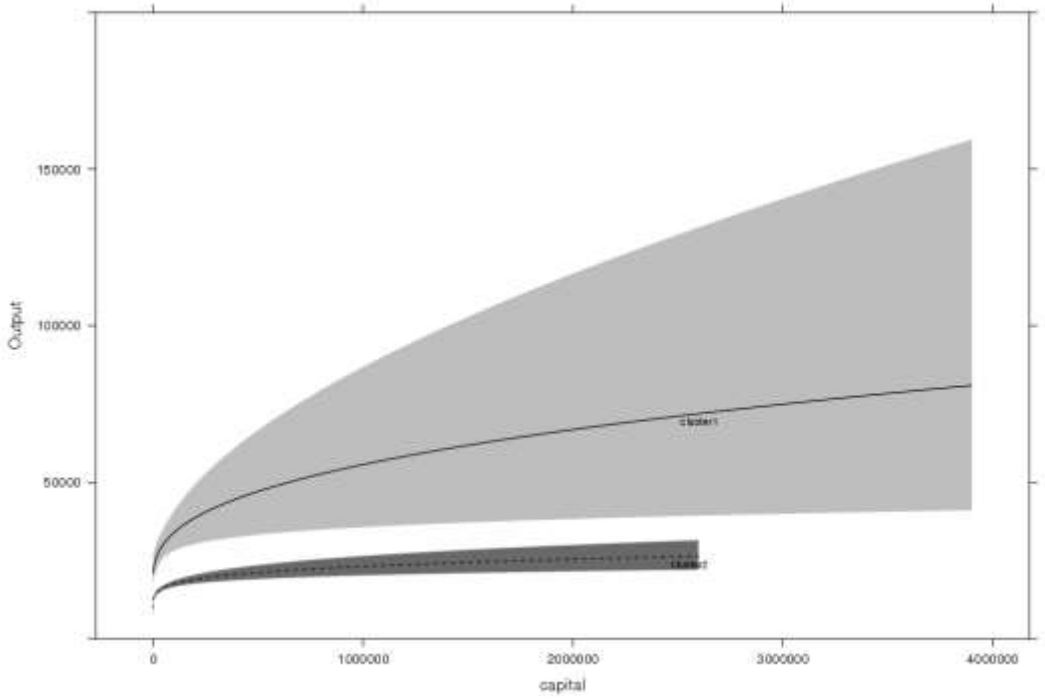

Figure 2. Effect of labour

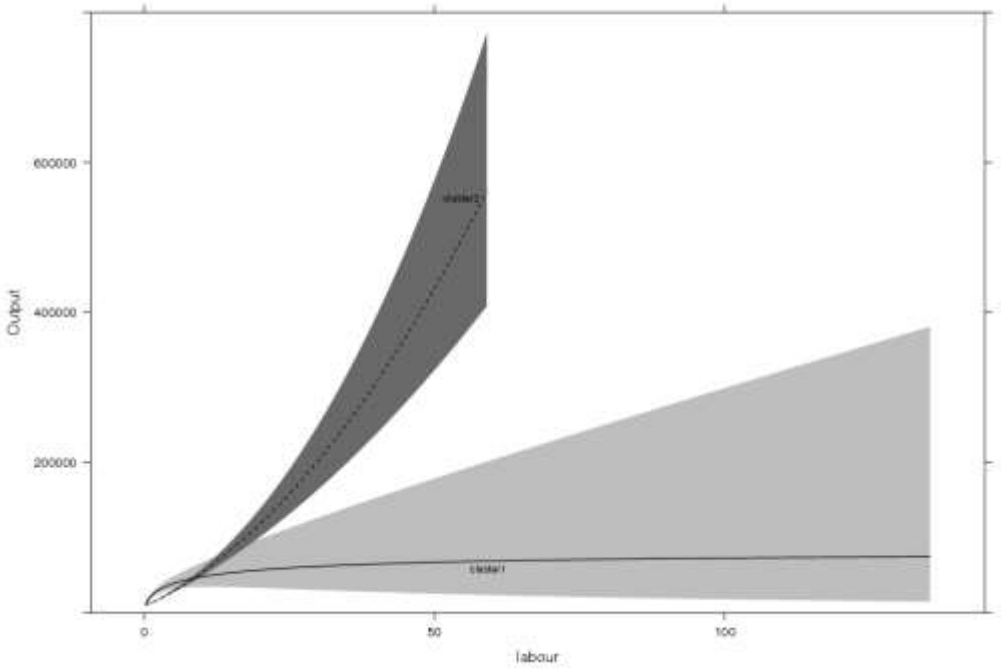

Figure 3 Effect of land

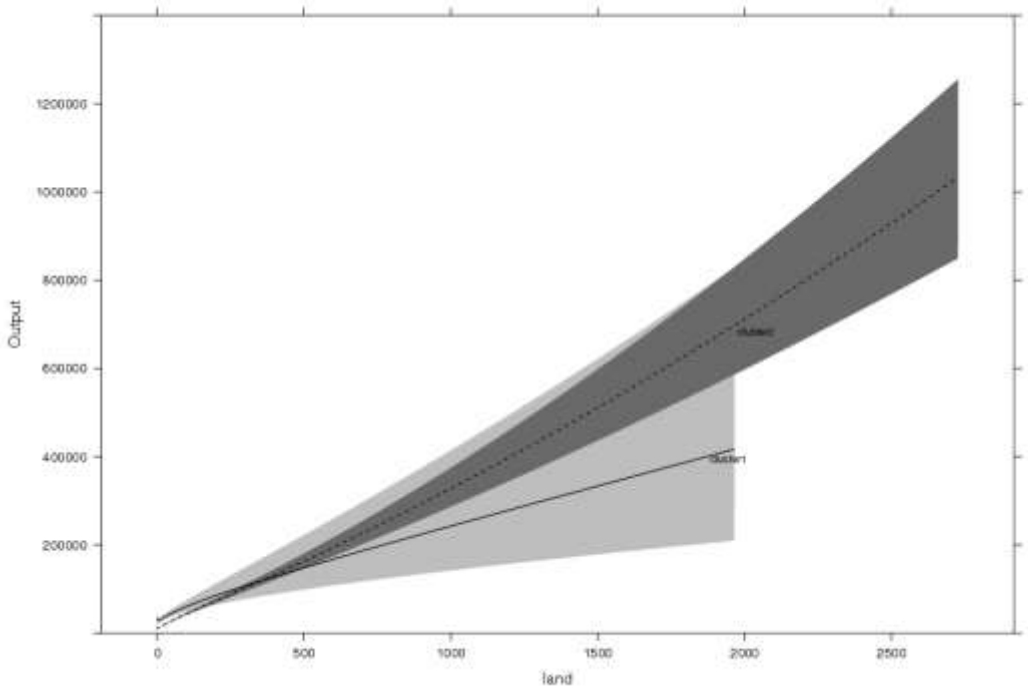

Figure 4. Effect of IC