Central Lancashire Online Knowledge (CLoK)

Title	A Quantile Regression Analysis of the Effect of Farmers' Attitudes and Perceptions on Market Participation
Type	Article
URL	https://clok.uclan.ac.uk/id/eprint/5917/
DOI	https://doi.org/10.1111/j.1477-9552.2012.00366.x
Date	2013
Citation	Kostov, Phillip and Davidova, Sophia (2013) A Quantile Regression Analysis of the Effect of Farmers' Attitudes and Perceptions on Market Participation. Journal of Agricultural Economics, 64 (1). pp. 112-132. ISSN 0021857X
Creators	Kostov, Phillip and Davidova, Sophia

It is advisable to refer to the publisher's version if you intend to cite from the work.
https://doi.org/10.1111/j.1477-9552.2012.00366.x

For information about Research at UCLan please go to http://www.uclan.ac.uk/research/

A quantile regression analysis of the effect of farmers' attitudes and perceptions on market participation

Philip Kostov and Sophia Davidova[1]

[1] Philip Kostov is Reader in Quantitative Economics, Lancashire Business School, University of Central Lancashire, Preston, Lancashire, PR1 2HE, e-mail: PKostov@uclan.ac.uk. Sophia Davidova is Professor of European Agricultural Policy, School of Economics, University of Kent, Canterbury, Kent, CT2 7NZ, e-mail: S.M.Davidova@kent.ac.uk. Research for this paper benefitted from data collected within the EU FP6 Programme SCARLED project (SSPE-CT-2006-044201). The authors are grateful to Alastair Bailey, Lena Fredriksson and Cornelia Suta for their contribution to the design of the survey instrument and the dataset.

A quantile regression analysis of the effect of farmers' attitudes and perceptions on market participation

Abstract

The objective of this study is to investigate the 'subjective' determinants of farmers' participation in output markets in five EU New Member States (NMS) characterised by a large semi-subsistence sector. It employs more rigorous methods than previous studies in this area to model the heterogeneity amongst farmers according to their market integration, i.e. the quantile regression models. The study also uses the Bayesian adaptive lasso to simultaneously select important covariates and estimate the corresponding quantile regression models. It does not fix any arbitrary cut-off points to differentiate semi-subsistence from commercial farming since the impact of different motivations and perceptions is investigated on the extent to which the farms are integrated into the output markets. The analytical results indicate that only three variables affect all quantiles. Apart from this, the actual structure of variables varies across different quantiles. Some variables affect the share of output sold at the lower quantiles (i.e. for semi-subsistence and less commercial farmers) only, while some other variables are only significant at the upper quantiles (i.e. for more commercial farms). Advisory services, and particularly agricultural business advice and advice on food safety and product quality standards, can facilitate the further market integration of the 'intermediate' farms – semi-subsistence and semi-commercial.

Keywords: *quantile regression, Bayesian adaptive lasso, semi-subsistence farmers, commercial farmers, NMS*

JEL classifications: *C11, C21, D12, Q12*

1. Introduction

The analysis of the characteristics of subsistence and semi-subsistence farming in Europe has become more prominent due to the two 'Eastern' enlargements of the European Union (EU) in 2004 and 2007. Several theoretical and empirical studies have been carried out in relation to semi-subsistence farming in different European countries (Kostov and Lingard, 2002;

Kostov and Lingard, 2004; Mathijs and Noev, 2004; Petrovici and Gorton, 2005; Latruffe *et al*., 2008; Davidova *et al*., 2009; Davidova *et al*., 2012 forthcoming; Fritzsch *et al*., 2011). Davidova (2011) estimates that in 2007 in the current 27 EU Member States there were 5.9 million farmers who used more than 50% of the output for household consumption.

One of the main characteristics of semi-subsistence farmers is their partial engagement in market activity. Although this is a common feature, semi-subsistence farmers are heterogeneous in terms of their objectives in farming, farm assets, human capital, income sources and strategies. Miracle (1968) criticised the concept of subsistence farming because it obscures the heterogeneity in farmers' situations and the diversity of their decision-making process. Davidova *et al*. (2009), Fritzsch *et al*. (2011) and Davidova *et al*. (2012, forthcoming) acknowledged this heterogeneity and employed cluster analysis to produce a typology of semi-subsistence farmers. However, cluster analysis does not allow for a formal procedure to select the relevant variables and the best model. Davidova *et al*. (2009) attempted to mitigate this shortcoming by using the clusters in a stepwise linear regression together with other variables in order to investigate the determinants of the share of output sold. However, the variables for the inclusion in the regression were selected by the researchers based on previous studies and not by a formal procedure.

The objective of this article is to investigate the 'subjective' determinants of farmers' participation in output markets in five EU New Member States (NMS) – Bulgaria, Hungary, Poland, Romania and Slovenia. These countries account for 94% of all farms that use more than 50% of the output for household consumption in the NMS and 84% of those in the whole EU (Davidova, 2011). One of the contributions of this study is that it employs a more rigorous method than previous studies in this area to model the heterogeneity amongst farmers according to their market integration, i.e. a quantile regression. This not only allows one to model the heterogeneous effects of covariates on the response variable but also takes into account the unobserved heterogeneity and allows for heteroscedasticity among the disturbances (Koenker, 2005). The study also implements the Bayesian adaptive lasso to simultaneously select important covariates and estimate the corresponding quantile regression models.

Another contribution of this study is that it does not fix any arbitrary cut-off points to differentiate semi-subsistence from commercial farming. Where definitions are concerned,

3

the threshold distinguishing subsistence from commercial farming becomes a point of contention. The most widely accepted pragmatic solution is the one proposed by Wharton (1969) who conceptualised market participation as a continuum from zero to 100% of output sold with a cut-off point differentiating subsistence and semi-subsistence, on the one hand, from semi-commercial and commercial farming, on the other, defined at 50%. Various other approaches have been empirically employed in order to decrease the arbitrariness of the thresholds. Petrick and Tyran (2003) split their sample of Polish farms into two equally sized groups – one above and one under the median value of own consumed goods in the total value of output. Cadot *et al*. (2006) employed a two-step procedure for switching regression models with unknown switch points. In the case of their sample of Madagascan farmers, the threshold of the share of sales in the total output turned out to be zero, thus subsistence farming meant a total isolation from output markets. Pingali and Rosegrant (1995) adopted a qualitative approach to characterise food production systems with increasing commercialisation, i.e. subsistence, semi-commercial and commercial. In the present study, no cut-off point is pre-defined since the impact of different motivations and perceptions is investigated on the extent to which the farms are integrated into the output markets. This is carried out for a range of quantiles which cover the continuum from zero to 100% share of sales in output.

The study also fills a gap in the literature on market participation which mainly investigates the persistence of subsistence and semi-subsistence farming in relation to transaction costs or, separately, entry and/or exit costs (e.g. Cadot *et al*., 2006; Petrick and Tyran, 2003; De Janvry *et al*., 1991; Key *et al*., 2000). Several studies have explored empirically the association of farmers' marketing behaviour with household assets, location and household characteristics (for a review of this type of empirical literature on Africa, see Barrett, 2008). Much less attention has been paid to the way farmers' objectives, values and attitudes shape their marketing behaviour. One attempt in this direction is the above-mentioned Davidova *et al*. (2009) study which used attitudinal statements to cluster farm households and in the second step employed these clusters in a stepwise regression together with other variables characterising farm assets, location and technology. The present study allows both the structure of farmers' attitudes and the way these are related to market participation to be explored.

4

The analytical results indicate that only three variables affect all quantiles. Apart from this, the actual structure of variables included in the empirical modelling varies across different quantiles. Some variables affect the share of output sold at the lower quantiles (i.e. for semi-subsistence and less commercial farmers) only, while some other variables are only significant at the upper quantiles (i.e. for more commercial farms). These more detailed insights that can allow policy makers to target better those semi-subsistence farmers in NMS who are on their way to commercialisation could not be achieved by the mean regression methods prevailing in most previous research.

This paper consists of six sections. The next section presents the conceptual framework and the third section describes the data used. Section 4 presents the empirical modelling and Section 5 discusses the results. Section 6 concludes.

2. Conceptual framework

The partial market integration of small scale farmers has been conceptualised by theoretical models. The most frequently used is the transaction cost model (e.g. De Janvry *et al*., 1991; Löfgren and Robinson, 1999; Key *et al*., 2000). This is a static perfect foresight equilibrium model which demonstrates that the presence of transaction costs leads to a band between a lower selling price and a higher buying price for an identical commodity. When the equilibrium solution falls within that band neither sale nor purchase is desired resulting in a subsistence state. Within this framework a farmer decides to be subsistence or commercial according to externally determined prices and transaction costs. The latter, however, are household specific.

Another model is the two-stage decision process model of Kostov and Lingard (2004) (henceforth KL). Unlike the transaction cost model, which has been extensively employed in empirical studies (e.g. Mathijs and Noev, 2004 for Central and Eastern Europe; Key *et al*., 2000 for Mexico), the latter, to the best of our knowledge, has not been tested empirically. There are two main reasons for this. First, it is a dynamic model and proper testing would require a longitudinal dataset with a sufficiently long time dimension, which is not available. The other reason lies in the nature of the model. Unlike the transaction cost model which is firmly based on assumptions such as perfect foresight, rational expectations and static equilibrium, the KL model is based on the concepts of dynamic transaction costs and farmers' orientation. The dynamic transaction costs (Langlois, 1992) are defined with regard

5

to a change. As such, they are an intrinsically subjective and procedural rationality concept.[2] In contrast to rational choice, individuals are not assumed to know all future situations and options, so their choice of outcome is generated through a deliberation process (Chaserant, 2003). In addition, change can only be evaluated against the existing *status quo* and as such the model assumes, similarly to all behavioural economic or finance models, a reference point. Decisions are made locally with regard to the reference point (the existing *status quo*), rather that globally with reference to a global equilibrium.

The other important feature of the KL model is that the dynamic transaction costs are defined with regard to an uncertain future following any changes. If the future is uncertain, i.e. the consequences of any actions leading to a change cannot be fully anticipated, then rational expectations are not necessarily assumed in this model. The use of the concept of dynamic transaction costs suggests that subjective views can affect economic actions. This possibility is made even more explicit through the use of the concept of orientation. KL defines two types of farmers' orientation, namely subsistence and commercial, depending on the primary objective in farming. Subsistence oriented farmers have household consumption as a primary objective, while commercially oriented ones view the marketing of the output as a primary objective. It has to be noted that orientation in the sense of the KL model is different from the observed outcome (i.e. the actual level of subsistence or commercialisation measured by the share of output sold). The concept of orientation is subjective since it reflects subjective views/attitudes. The subjective attitude influences an individual's decisions (see Kostov and Lingard, 2004 for details).

Therefore, in essence, the fundamental difference between the two models is that while within the transaction cost framework the degree of market integration of a given farm/farm household is externally determined by 'objective' factors (i.e. prices and transaction costs), in the KL approach the subjective evaluation of the *status quo*, the nature of the alternatives and the subjective utility contributions of the attitudes towards agricultural concept, measured by orientation, affect the degree of market participation. This does not indicate that the KL model is entirely subjective. Objective factors still exert influence, but this influence is not direct, it occurs via the subjective evaluation of these objective factors, and the degree of

---

[2] Procedural rationality includes the cognitive processes that are involved in a choice.

6

correspondence between the objective facts and their subjective interpretation is not restricted *a priori*.

The policy implications of the KL approach depend on the structure of orientation since the model allows for a differentiated approach to different types of farmers distinguished as being subsistence or commercial. Since the main objective of the present study is to measure the effect of farmers' objectives and perceptions on the extent of their market integration, the KL approach appears to present a more adequate conceptual framework. As indicated previously, due to the lack of longitudinal data it is not possible to study the dynamic changes in farmers' orientation. Therefore, this paper focuses on a static investigation of the effects of the subjective evaluations on the extent of market integration.

3. Data

The data for the empirical study were generated through a primary survey carried out within the EU FP6 project 'Structural Change in Agriculture and Rural Livelihoods' (SCARLED). The survey was focused on agricultural households in five EU NMS (Bulgaria, Hungary, Poland, Romania and Slovenia) characterised by a large semi-subsistence sector. Only households that reported being engaged in agricultural production in 2006 and/or 2003 (including production from house gardens) were included in the sample.

The survey instrument was designed in such a way that both quantitative and qualitative information was collected. It required quantitative data on: household members, time allocation and income sources; inputs and outputs, including information on purchased inputs and self-consumed or marketed output product by product; land and non-land assets, and labour use. The largest part of the questionnaire consisted of qualitative statements measured on a 5-point Likert scale concerning motivation in farming, attitudes to commercialisation, barriers to and drivers for income diversification, and market integration. These are the variables used in the empirical analysis in the present study.

The survey used geographical cluster sampling. Regions and villages were selected through a two-stage clustered sampling process. In the first stage, three regions in each of the five

7

surveyed countries were selected using EUROSTAT data at the NUTS3[3] level according to their degree of economic development – poor, average or prosperous – corresponding to a gross domestic product (GDP) per capita relative to the national average. Since the emphasis was on rural areas, the regions of the capital city and other large cities were excluded from the selection. In the second stage, three villages per NUTS3 region were selected – a prosperous, an average and a poor one in comparison to the regional average – and agricultural households in these villages were surveyed (for more detail, see Davidova *et al*., 2009). The survey was implemented in 2007-2008 by face-to-face interviews using local enumerators.

Based on the survey data, the dependent variable for the empirical analysis was constructed, i.e. the share of output sold in the total agricultural output. This measures the degree of (output) market integration. In the literature, this is the measure used most often in defining subsistence, although it has sometimes been criticised as reflecting farmers' behaviour in output markets only (Miracle, 1968). The subsistence-commercial continuum could also be defined with regard to the integration in input markets. However, the latter is more difficult to measure and does not provide any information about output use and the output supply response which is of interest to policy makers from the point of view of food security and farm revenues. In this study the focus is on output market integration only.

The construction of the dependent variable required several calculations. First, the total value of sales was established product by product. For livestock products this value was available in the questionnaire, while for crops it was derived by multiplying the sale price by the quantity sold. Second, similarly, the total value of output was derived on a product by product basis by multiplying the quantities produced by the price (directly available for crops and derived from the sales value for livestock). Third, the share of output sold was calculated as a ratio of sales value in the value of the total output.

One issue in the above calculations was the use of different units for quantities and prices/values within the questionnaire. Since the values were aggregated, this was not a problem as long as the values for all products were expressed in the same units (national

---

[3] NUTS stands for Nomenclature for Territorial Units for Statistics. In the countries analysed NUTS 3 level corresponds to districts in Bulgaria, counties in Hungary and Romania, regions in Slovenia and sub-regions in Poland.

currency units). Whenever this was not the case, the aggregation was not possible. For this reason, all households for which the units of measurement were different and could not be easily reconciled were removed. Furthermore, households for which there were missing data for any product (meaning that a household reported producing a particular product, but some data were missing and it was not possible to calculate the corresponding values) were excluded as well.

The above procedure led to 766 observations for which the share of output sold could be calculated. The number of useable independent variables was 128. Out of these 17 variables were excluded due to too many missing values. This resulted in the 280 full observations for 111 independent variables which constituted the dataset analysed.

4. Empirical Methodology

This study employs a quantile regression approach. Quantile regression is an important method for modelling heterogeneous effects of variables on a response and at the same time it takes into account unobserved heterogeneity and allows for heteroscedasticity among the disturbances (Koenker, 2005). The quantile regression can be written as:

$$y_i = X_i^T \beta_\tau + u_{\tau i} \quad u_{\tau i} \sim H_{\tau i} \ subject \ to \ H_{\tau i}(0) = \tau \tag{1}$$

where the index i denotes the individual agent (household/farm), y_i is the dependent (response) variable and X_i is the vector of covariates for individual i, β_τ denotes the quantile specific linear effects and $0 < \tau < 1$ is a given (i.e. fixed and known) quantile. One can easily note the similarity in notation to the linear regression model. In a sense, the quantile regression is a generalisation of the latter. However, instead of being fixed, the coefficients are allowed to vary with the quantile. For this reason they are represented in (1) as an unknown function of the quantile τ. The unknown error term $u_{\tau i}$ is characterised by an unspecified cumulative distribution function $H_{\tau i}$. No specific distributional assumptions are made about this distribution function except from the restriction in (1), which implies that the distribution function at 0 is τ. The latter is known as a linear quantile restriction and it can be relaxed if non-parametric versions of the quantile regression models are considered.

9

The linear quantile restriction leads to the following interpretation: the model describes the quantile function $Q_{y_i}\left(\tau \mid X_i\right)$ of the response variable y_i conditional on a vector of covariates X_i at a given quantile τ. More specifically:

$$Q_{y_i}\left(\tau \mid X_i\right) = H_{y_i}^{-1}\left(\tau \mid X_i\right) = X_i^T \beta_\tau \tag{2}$$

Therefore, in other words, the quantile regression represents the τth quantile of the response variable as a linear function of the covariates. In contrast, the linear regression model describes the mean of the dependent variable. The fundamental difference is that the mean models assume that the response variable is conditionally Gaussian, which means that the mean equation applies to all parts of the distribution. The quantile regression makes no such distributional assumptions and, hence, the conditional quantile function that is estimated can vary across quantiles. It would also be useful to clarify that in estimating any quantile (including the most extreme ones) the (linear) quantile regression uses all available observations.

The conditional quantile can be alternatively expressed as the following optimisation problem (see Koenker and Bassett, 1978):

$$\arg\min_{\beta_\tau} \sum_{i=1}^n \rho_\tau\left(y_i - X_i^T \beta_\tau\right) \tag{3}$$

where $\rho_\tau(.)$ is the so called 'check function', i.e. $\rho_\tau(u) = u\left(\tau - I(u < 0)\right)$, with $I(.)$ denoting the indicator function. Solving (3) leads to the most popular linear quantile regression estimator, namely the linear programming estimator of Koenker and Bassett (1978).

Koenker and Machado (1999) noted that the minimisation problem (3) can be recast as an equivalent maximum likelihood theory problem where the distribution of the response variable is the skewed asymmetric Laplace distribution. This has been exploited to propose Bayesian versions of the quantile regression (see Yu and Moyeed, 2001). Although distributional assumptions are necessary for the Bayesian approach, these are pretty innocuous since the Bayesian estimation of a quantile regression is simply established to be equivalent to the frequentist estimation that does not employ any.

Since quantile regression models are estimated in this study, it is necessary to consider the distribution of the dependent variable (the proxy for market integration). This is a ratio taking

values from zero (no sales) to 1 (all produce is sold). Using this variable directly in a regression model ignores the fact that the dependent variable is measured over an interval and can introduce an estimation error. In particular, a model should not predict values below or above one and running an unrestricted regression model could result in parameter estimates that predict such inadmissible values. For this reason, it is preferable to use a fractional response model. The most widely applied approach in modelling a fractional response variable is to transform the original variable in such a way that the interval restriction no longer holds. The latter can be expressed by applying the logit transform $y^* = log(y/1-y)$, where y is the original (interval valued) fractional response variable, and build a model for the transformed variable y^*. This can be more easily seen if one considers the opposite transform, i.e. that $y = exp(y^*)/[1+exp(y^*)]$, showing that for any value of y^*, y is guaranteed to be in the (0,1) interval. The problem arises when the fractional variable is measured at the boundary of the unit interval (i.e. when it takes the value 0 or 1), because then the logit transform is undefined. The latter can be overcome by a preliminary 'scaling' of the fractional variable to map it from the [0,1] to the (0,1) interval. This can be achieved -by replacing y by $(y+e_1)/(1+e_2)$, where e_1 and e_2 are some small numbers, such that $e_1 < e_2$. Adding e_1 moves y away from zero, while dividing by $(1+e_2)$ scales back its values and as long as $e_1 < e_2$ the scaled values will be lower than 1. Here $e_1 = 10^{-32}$ and $e_2 = 10^{-8}$ are used.

To allow for an unrestricted (conceptually) dependent variable, the application of the above logit transform also preserves the ranking of the dependent variable, which is an important property particularly in view of employing a quantile regression. Furthermore, the coefficients in the transformed model can be interpreted in the usual way with regard to their signs. Similarly, larger coefficients indicate a larger effect. Their magnitude, however, would not have a direct interpretation, although one can use estimated coefficients to calculate 'odds ratios' in the same way as in a logistic regression. Since the magnitude of the effects is not a primary focus of this study and owing to the large volume of results this is not applied here.

In addition to estimating a quantile regression for a range of quantiles, the interest in this study is also in determining which variables affect the corresponding conditional quantiles, thus there is a variable selection problem. The standard tools for analysis of regression results, such as t statistics and F tests, are based on the implicit assumption that the set of predictors is fixed in advance. In practice, however, in most empirical problems this set is not fixed but rather chosen adaptively, using some formal procedure (e.g. stepwise regression and

all-subsets regression) or an informal one (researcher selected variables that provide a good fit). Under adaptive selection of regressors these classical tests are biased. This bias ultimately affects the variable selection process since the above variable and model selection procedures are essentially based either directly on the F statistics or on some other related statistics. Two possible strategies can be employed to circumvent the above problem. One is to apply bias reducing adjustments to sequential F tests in an adaptive variable selection algorithm. The other option, which is followed here, is to use a penalised regression.

Penalised (also called regularised) regression methods have emerged as important techniques for variable selection. Since Tibshirani (1996) introduced the least absolute shrinkage and selection operator (lasso) as a method that can simultaneously achieve both parameter estimation and variable selection, a vast literature has developed which analyses and extends this method, and/or suggests other regularisation approaches based on alternative penalties. Although consistent in terms of variable selection since it retains the important variables, the original lasso estimator applies fixed amount of shrinkages to all coefficients, which can be a problem when the so-called oracle property is desired. In simple terms, an estimator possessing the oracle property will have the same asymptotic distribution of the coefficient estimates as the 'oracle' estimator, i.e. the estimator implemented knowing which coefficients are zero. This allows an oracle estimator to be used not only for variable selection but also for inference. In particular, the original lasso estimator can be an oracle one only under some special circumstances subject to non-trivial conditions (see Zou, 2006 for details). However, when an adaptive amount of shrinkage for each regression coefficient is implemented this leads to estimators which possess the oracle property (Zou, 2006).

Two of the most popular regularisation approaches, namely the least absolute shrinkage and selection operator (lasso) of Tibshirani (1996), and the smoothed clipped absolute deviations (SCAD) method of Fan and Li (2001), have already been considered in a quantile regression setting (see Li and Zhu, 2008; Wu and Liu, 2009; Belloni and Chernozhukov, 2009). In general, these papers have established the consistency of such regularised estimators for quantile regression problems subject to appropriately chosen 'optimal' penalty parameter(s).

A regularised (penalised) linear quantile regression estimator can be formally defined as:

$$\min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau \left(y_i - X^T \beta_\tau \right) + \lambda J \left(X^T \beta_\tau \right) \tag{4}$$

where $J(.)$ is a given penalty function.

The shrinkage effect is determined by the positive penalty parameter λ that needs to be chosen according to some criterion (typically information criterion or cross-validation).

The adaptive lasso estimator for the linear quantile regression can be defined as a weighted lasso problem in the following way:

$$\min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau \left(y_i - X^T \beta_\tau \right) + \lambda \sum_{j=1}^{d} \tilde{w}_j \left| \beta_{j\tau} \right| \tag{5}$$

where $\left| . \right|$ denotes the *L1* norm, while the weights are given by $\tilde{w}_j = \dfrac{1}{\left| \tilde{\beta}_{j\tau} \right|^\gamma}$ for some $\gamma > 0$,

where $\tilde{\beta}_{j\tau}$ are initial estimates for the parameters. Usually $\tilde{\beta}_{j\tau}$ are obtained by an unpenalised quantile regression. The conventional lasso estimator is a particular case when all weights are equal rather than adaptively chosen.

The adaptive lasso, when implemented in a quantile regression setting, retains the oracle property (Zou and Yuan, 2008) similarly to the mean regression case. *L1* norm estimators are by far the most widely studied regularisation estimators for quantile regressions (see for variable selection applications e.g. Belloni and Chernozhukov, 2009; Wu and Liu, 2009; Zou and Yuan, 2008).

The main computational cost in implementing the penalised quantile regression stems from the fact that an optimal value for the penalty needs to be chosen. This is typically accomplished by some form of cross-validation. A direct application would, however, require repeatedly computing optimisation problems similar to (5). The latter could be computationally expensive. An alternative is to estimate the whole regularisation path for lasso type problem (i.e. the lasso solution for a range of penalty values) directly using efficient path solution algorithms. Li and Zhu (2008) proposed such an algorithm for the lasso type of quantile regression problem. The advantage of such algorithms is that, since they exploit the piecewise linear property of the regularisation path, they can be obtained at a fraction of the computational cost of the overall regularised estimator. This facilitates the implementation of cross-validation.

It is well recognised that in linear regressions such regularisation methods have equivalent Bayesian formulations obtainable by adopting suitable prior distributions on the regression

coefficients (see Park and Casella, 2008; Hans, 2009 for details and discussion on Bayesian lassos). Li *et al*. (2010) employ a Laplace prior on the quantile regression coefficients to obtain a Bayesian version of the lasso (and some other regularisation approaches). Alhamzawi *et al.* (2011) combine Laplace priors on the coefficients with inverse Gamma priors on the individual shrinkage for each parameter to obtain a Bayesian adaptive lasso regression. The advantage of the latter is that the individual amounts of shrinkage are no longer given but are treated as unknown and therefore estimated from the data jointly with the parameters. A disadvantage of the Bayesian lassos is that since continuous priors are imposed on the regression parameters, unlike the case of the frequentist counterparts, draws from the posterior distributions are never exactly zero. This means that some *ad hoc* (typically thresholding) methods must be applied to implement variable selection. In this paper a Bayesian approach is followed and we apply a hard thresholding at the 95% confidence limit. From the Bayesian point of view discrete priors (such as e.g. the spike and slab prior) can be used to automatically perform a variable selection. For example, in a linear quantile regression the stochastic search variable selection (SSVS) method, proposed in Reed *et al*. (2009), can be applied. This, however, is much more computationally demanding and in a high dimensional model setting, such as the one adopted here, the sampler will visit the 'true model' only a handful of times which would impede the efficient inference. Finally, although we have to implement hard thresholding to arrive at the final model, the Bayesian adaptive lasso, by virtue of its (asymptotic) equivalency to the frequentist version, shares its oracle property and hence we effectively combine this with the small sample advantages of the Bayesian approach. Furthermore, the Bayesian estimation provides confidence intervals, unlike the frequentist versions in which the final model needs to be re-estimated to obtain these. The implementation in this study follows Alhamzawi *et al.* (2011) and the reader is referred to Appendix 1 for brief technical details.

5. Discussion of results

As mentioned previously, one of the contributions of this research is that it provides more detailed insights into the effect of the motivations and perceptions of farm households on their market integration within the continuum from subsistence to fully commercial. The results are meaningful in two aspects. First, the actual structure of the variables included in the model varies across different quantiles. This means that some variables affect the share of

output sold at the lower quantiles (i.e. for more semi-subsistence farmers) only, while some other variables are only significant at the upper quantiles (i.e. for more commercial farmers). Second, the modelling strategy applied allows for heterogeneous impact, i.e. even when some variables affect the outcome at different quantiles, their impact is allowed to vary and be different for different types of farmers.

In the discussion that follows the focus is on the differences in the estimated coefficients for different quantiles. It must be noted that when discussing the changes in the magnitude of the estimated effects only tentative conclusions can be drawn. For a proper assessment of whether such effects are significantly different across quantiles formal statistical tests are required, something that is not done in this paper in order to simplify the discussion and limit the volume of results. Nevertheless, a discussion of the way the effects change across the quantiles is informative with regard to the underlying processes.

Seventeen different quantile regressions have been estimated for quantiles from the 0.05^{th} up to the 0.85^{th} in steps of 0.05. Since there are a number of fully commercial farms (100% sales) in the sample, at the 0.85^{th} quantile the commercial state is reached and there is little to gain in terms of further insights in estimating a model at higher quantiles. However, for policy implications it is worth stating that farmers in the 0.85^{th}, 0.90^{th} and 0.95^{th} quantiles sell all of their output and do not use any part of it for household consumption.

For interpretation purposes it is useful to characterise the quantiles for the untransformed dependent variable. As explained in the methodology section, the actual dependent variable in the model is a transformation of these shares, but since the transformation is rank preserving it is more informative to look at the corresponding quantiles for the untransformed sale shares in output. Table 1 presents the coefficient estimates with the actual sales shares the corresponding empirical quantiles of the dependent variable refer to. A more detailed representation, including the 95% confidence intervals for the estimates is provided in Appendix 2. This table suggests that even the least market integrated farmers in the estimation sample are not subsistence but can be defined as semi-subsistence. Farmers in the lowest quantile (0.05) sell on average 22% of their output; the 0.15^{th} quantile reaches 50% of sales. Starting from the 0.45^{th} quantile farmers sell more than 80% of their output and can be treated as typical commercial farms. A large number of the more subsistence farms were excluded from the estimation sample because of missing or unreliable data. Therefore, the

15

estimation results have to be interpreted with regard of the actual degree of market integration, denoted by the corresponding empirical quantiles.

Table 1 about here

As mentioned previously, 111 variables were used in the selection process and 34 important covariates were retained. These fall into six groups: household off-farm occupation (E variables); incomes (F variables); land assets (G variables); agricultural production, use and sales (H variables); contribution of own food production to household welfare (I variable) and future farming activities (K variables). A description of the retained variables, together with the way they are measured, is presented in Table 2.

Table 2 about here

A first important point that emerges from Table 1 is that the estimated model does not show any sign reversals, i.e. there are no variables that have a negative effect at some quantiles while exerting a positive impact at other quantiles. This simplifies significantly the description of the results.

There are three variables that influence all quantiles. The first is the agreement that the current aim in farming is to provide work for the household members (H2b). The estimated impact of this variable is negative, implying that a higher level of agreement with this statement is associated with a lower level of commercialisation. Higher scores for this variable imply that farmers may pursue this objective at the price of underemployment and low labour productivity. In this case, farmers may not be competitive in the market and they are forced to consume a great deal of their output. Therefore, it is not surprising that the coefficients of the variable are negative across all quantiles. The other point worth noting is that the estimated effect generally weakens with the quantile (i.e. its magnitude is lower at the upper quantiles). Hence, this impact is stronger for the less commercial farmers.

The second variable that affects all the quantiles is the agreement with the importance of the access to credit to invest in non-farm businesses with a view to exiting from agriculture (K9e). The relationship is positive across the board. In terms of magnitude, broadly speaking, it remains at a constant level before decreasing at the upper quantiles. If access to credit is deemed important for exit from agriculture (as the higher evaluation of this variable indicates), this shows that the households feel they need loans to start up non-farm businesses and are interested in doing so. The credit market, however, can be characterised by credit

rationing. It has been widely documented that due to credit market imperfections there is credit rationing in the NMS. Small farmers have difficulty in accessing loans and they face liquidity problems (e.g. Latruffe *et al*., 2010; Latruffe, 2005; Petrick, 2004; Davis *et al*., 2003; Bezemer, 2002; Swinnen and Gow, 1999). Furthermore, the above studies have identified the credit rationing faced by farmers in these countries as 'type 2' rationing (see Stiglitz and Weiss, 1981), where some applicants are refused loans altogether despite their willingness to pay the market rate of interest, as opposed to 'type 1' rationing (Jaffee and Russell, 1976) where borrowers prefer larger loans than those on offer and where everyone willing to pay the market rate of interest obtains some funding, although a lower amount than they would prefer. In the presence of type 2 credit rationing, one of the mechanisms to overcome it is to provide collateral and savings to reduce the credit contribution to the project (see e.g. Bester, 1985 for details). One possible explanation for the positive relationship is that the stronger the farmers agree they need credit to start up non-farm activities, which they cannot access, the stronger their motivation is to integrate into the market in order to fund their investment through internal accumulation. Hence, increasing sales could be a way to build-up cash funds. The decrease of the effect at the highest quantiles is probably a reflection of the decreased possibility of further increasing the share of sales when a farm gets closer to a fully commercial state.

The third variable that affects all quantiles relates to the functioning of the land market (G9j). The respondents were asked about their agreement with the statement that they would like to sell land. Similarly to the previous one, the relationship is positive and the strength of the relationship is more pronounced for farmers at the semi-subsistence end. At first glance, this variable seems to have a wrong sign. If land is sold, this would be expected to reduce output which, everything else being equal (e.g. the self-consumed quantity remaining unchanged), would decrease the marketed share. So, following this logic the willingness to sell land could be expected to have a negative impact on the dependent variable. The above argument, however, assumes implicitly that all available land is cultivated to its full potential. If, however, this is not the case and the land under consideration for sale is underutilised (which is the reason for wanting to sell it) then a positive effect can materialise. Since farmers may not be able to cultivate all their land, in order to improve the factor mix that will help them to market more of their output they need to sell some land. The relationship suggests that there are land market imperfections in NMS which corroborates the work of Ciaian and Swinnen

(2006). There is also an important difference between the two land transactions, namely selling land or renting it out. The preference to rent land out may indicate a greater emotional attachment to land (i.e. the traditional attitude), while selling shows a greater degree of calculative commercial-like attitude. In other words, the preference for selling land over renting it out demonstrates a commercial orientation (in the sense of Kostov and Lingard, 2004) and could be expected to increase the share of marketed production.

The rest of the selected variables only affect some quantiles and not others. Furthermore, a particular clustering of such effects is observed since most of these variables tend to affect mainly either the lower or the upper quantiles, i.e. either only the semi-subsistence and semi-commercial farmers or the commercial ones, but not both groups. Furthermore, the frontiers of impact for such variables are not fixed but change with the variable, so that the effect of some variables may be concentrated at the more extreme quantiles while some other variables impact the centre of the distribution (i.e. the intermediate quantiles). The interpretation below is structured according to the cluster of effects either in the lower or upper quantiles. In this way it could be clarified which motivations and attitudes affect low market integration and which the high one.

Two variables impact only on the lowest quantile (0.05). The first one refers to the aim of transferring to the next generation (H2c). It has a positive impact on market integration. Such motivation would be impeded if the value and productive capacity of the farm were decreased which might be the case without a generation of cash from sales. Therefore, at the most semi-subsistence end of the spectrum (these farmers sell on average 22% of their output) sales of output are required to maintain the productive capacity and the value of the family farm for the next generation.

The second variable conveys the idea that off-farm work is a stepping stone out of agriculture (k9b). The least market integrated farmers appear to see off-farm work as a way out of agriculture and a substitute for the other potential source of cash income, i.e. further market integration. Therefore, the effect of this variable on the response is negative.

A wide range of variables only impacts on several quantiles at the lower end. In fact the results in this paper suggest that the subjective factors are more important in shaping the market behaviour of less market integrated farms. According to the direction of the relationship these variables can be grouped into those with a positive effect on market

integration and those with a negative one9b. The impact of these variables is discussed in turn.

Following the KL framework, the variable F7 contains information on the subjective assessment of respondents of their incomes and consumption in kind in 2006 in comparison to the situation in 2003. Therefore, they subjectively assess the change with respect to a local reference point. The increasing values indicate a perception of improvement. The perceived improvement exerts a positive impact, but it is only observed at the lower ($q<0.35$) quantiles. Furthermore, the magnitude of the impact decreases with the quantile. The higher evaluation of the change in the respondents' income and consumption situation creates a perception of improving circumstances and optimism. This should reduce their risk aversion and hence induce further sales, since the perceived risks are lower. Such an effect could be expected to be greater when risk aversion is higher and the initial wealth stock is lower. Both these conditions point towards the lower quantiles. This is further reinforced by the decrease in this effect at higher quantiles.

H3c refers to the use of agricultural business advice and information. It shows a positive impact only present at the lower ($q<0.3$) quantiles. This impact decreases in magnitude with the quantile. In general the sign is to be expected. Since farmers for which this variable is important sell between 20% and 60% of their production, they cannot be described as fully commercial. Using such advice shows a desire to expand and, therefore, should lead to increased sales. As farmers become more commercial, they are likely to gain both experience and knowledge obtained through advisory services, meaning that the marginal effect of additional advice should decrease until it disappears.

The next variable from the same group is H6f 'We lack information and advice on markets and prices' also has a positive impact at the lower ($q<0.4$) quantiles and the effect decreases with the quantile. On the one hand, this variable emphasises once again the perceived need for information and advice by farmers at the less market integrated end. On the other, it reflects a perception of unachieved potential with regard to prices and sales. Therefore, it accounts for a desire to sell (a commercial orientation in the KL sense) which explains its positive impact. As the sales share increases (up to 70%) this perception diminishes until its effects on market behaviour disappear.

A more optimistic evaluation of the mid-term (next 5 years) economic prospects of the farm, as measured by K5, exerts a positive impact at the lower quantiles (q<0.25) and this impact reduces with the quantile. The effect is to be expected in terms of direction. It is manifested for semi-subsistence and partially for intermediate type of farms (semi-commercial) only as they have to decide whether to further their commercialisation. The selection of this variable suggests, in agreement with the conceptual framework, that views (subjective evaluations) about future prospects affect market behaviour.

G9a ('Would like to buy more land') has a positive impact at the lower quantiles (q<0.3) only and this effect decreases with the quantile. Clearly a desire to buy land reflects expansion plans and commercial orientation. So farms with such an orientation should be expected to sell more relative to other farms (with a subsistence orientation). At first glance, it appears counter-intuitive that such an effect is only present at the lower quantiles. However, for less commercialised farms the difference made by an expansion in the land area would be larger because the change at the margin on the marketed surplus would be larger as well.

Several variables negatively affect the market integration of lower quantiles. The first variable – E15h – points towards the insufficient availability of low cost credit as a constraint to engaging in off-farm business and impacts on (q<0.25) quantiles (i.e. households who sell roughly less than half of their production). This is a logical result. If farmers perceive this to be an important constraint chaining them to agriculture, they may be less interested in expanding their agricultural activities and sell more.

K8b measures the perceived importance of contracts with buyers for the probability of investing in farming. Its effect is pronounced at the 0.1^{th}- 0.4^{th} quantile range and has a negative impact on market integration. This effect could be expected under the assumption that such contracts are generally missing for smaller and less market integrated farms. Despite various innovations by private capital and foreign investors (see Gow and Swinnen, 2001; Dries and Swinnen, 2004) aimed at promoting vertical integration in the agriculture of the NMS, these have not always incorporated smaller less market integrated farms. Csaki *et al*. (2008) suggest that these farms have had difficulties in meeting the conditions and accessing the Special Accession Programme for Agriculture and Rural Development (SAPARD) funds. Consequently, they have had problems in upgrading and expanding their supply, and as a result the 'anti-small farm bias' has increased. If there is no interest by

20

buyers in entering into formal contractual relations with less market integrated farmers and these relations are important for realising sales this would lead to both less investment and a lower share of sales.

The last variable that exerts a negative effect on market integration for the lowest quartiles (q<0.45), G9f, indicates that respondents wish to rent-in more land. It seems that due to the thin land market and high competition from the large private producers and corporate farms (companies and co-operatives), the smaller farmers feel that they cannot rent-in more land to expand and have a larger marketable surplus. Swinnen and Vranken (2008) argue that corporate farms in the NMS have led to imperfect competition in the land markets influencing the rent rates and rental contract conditions.

The other aspect of interest in this study concerns the variables that mainly influence the upper quantiles, thus the market integrated farmers. Of larger policy relevance are those subjective factors that may positively affect farmers' market integration. For this reason, the following discussion focuses mainly on the variables which have a positive relationship with market integration.

The first such variable is E14 which measures the intention to earn income through starting up an own off-farm business within the next 5 years. It exerts a positive impact and this impact is only present in the two highest quantiles. Business orientation (entrepreneurship), as indicated by a higher score for this variable, even when it is non-agricultural, increases commercialisation. Given that these quantiles denote a virtually full commercialisation, the above result also hints that such farmers may have reached the limits of what they can extract as income from agriculture and, hence, could be inclined to look off-farm.

The preference for agricultural work to other self-employment activity (measured by E15b) positively affects the sales of more commercial farmers (marketed share in excess of 90%). This is due to the fact that when household members have a stronger preference towards agriculture, they view it as a preferable choice of livelihood and in order to make a living out of their choice they need to sell more. For more subsistence farmers, a similar preference would have meant a more traditional attitude and therefore might have had a negative effect. However, this variable does not affect the marketing behaviour of the lower quantiles for the sample analysed.

21

The variable E33 contains information on whether the household plans to earn an income through taking up wage employment within the next 5 years. It is selected at the higher quantiles (q>0.55). It should be noted that there is a slight difference in that this variable indicates specific plans, rather than measuring general attitudes as was the case with some other variables from the E15 category. Although the quantiles at which this variable appears are exactly the same as the quantiles affected by the previous E15b ('Household members prefer agricultural work'), here the estimated effects are negative, thus such plans decrease the share of output sold. Therefore, there is a difference in the direction of the effect compared to similar self-employment plans. Taking a wage job does not require start-up funds and hence does not reveal entrepreneurial orientation. Furthermore, if realised, such plans would most likely reduce the households' on-farm labour input, and under imperfect rural labour markets, largely documented for the NMS (see e.g. Rizov and Swinnen, 2004; Macours and Swinnen, 2005), also agricultural output and hence sales. What is interesting is that this particular variable is not selected at the lower quantiles. This confirms the conjecture of Kostov and Lingard (2002, 2004) that semi-subsistence farmers may be locked into agriculture because of the lack of alternative income and employment opportunities. If they perceive a general lack of such opportunities, and the fact that taking up wage employment cannot materialise, then they treat wage employment as an aspiration rather than a specific plan for the mid-term future and therefore this variable, suggesting specific plans, does not affect the behaviour of these farmers.

One factual production-related variable is H3d measuring the extent to which the farm fully employs household members. This variable has a positive impact, only present at the highest quantile. The alternative (not being able to fully employ household members) would mean under-utilisation of resources and therefore insufficient production. For fully commercial households this would tend to reduce the share of sales. For other households the net effect would depend on the motivation behind such employment. Since there could be two different such motivations – to increase production or to provide livelihood – and it is difficult to reconcile these two motivations, such an effect is not pronounced at the lower quantiles.

One variable affecting the upper quartiles relates to the issue of farm succession – K3 'A successor for the farm has already been designated by the household'. This variable impacts positively on the marketed share, but the effect is expressed at the higher (q>0.65) quantiles. Since this variable eliminates a fundamental uncertainty about the farm succession and

uncertainty would be expected to impact negatively on farm operation and therefore sales, it is not surprising that the effect is positive.

At the end, due to their policy relevance, two different sets of evaluations about household's abilities to adapt to the EU regulations that have been retained by the model are discussed. K10a (veterinary and phytosanitary standards) exerts negative impact at relatively higher (q>0.25) quantiles. The higher values of the variable indicate that households perceive that adapting to the standards is easy. One possible explanation of the negative impact is if these standards are easy to be met by a particular farm, it is likely to be easy to be met by other farms. This would lead to an increased competition (because it would lead to a higher level of EU compliant production) and hence restrict the marketing abilities of a specific farm. It has to be noted that this effect appears for mostly commercial (sale share above 61%) farms and is totally missing for the semi-subsistence ones, which conforms to the above interpretation.

The effect of the other variable K10b (food safety and quality standards), on the other hand, is positive and is present at the 0.1^{th} to 0.35^{th} quantiles (i.e. 45-71% share of sales). The standards in K10b refer to the product and hence could be expected to have a direct influence on the ability to market the particular product. Therefore, K10b has the expected sign since the easiness to meet such standards enhances the marketing ability of the farm. This impact is manifested mainly for intermediate type of farms (semi-subsistence with a relatively large share of sales and semi-commercial), since more commercial farms in order to achieve their high share of sales would have by definition met these standards. The positive relationship suggests that another way in which policy can facilitate commercialisation is training and advice on product standards and the necessary technology adjustments.

6. Conclusions

This study focuses on the impact of motivations, perceptions and attitudes of agricultural households in five EU NMS – Bulgaria, Hungary, Poland, Romania and Slovenia – on their market participation. The conceptual framework is based on Kostov and Lingard (2004), suggesting that objective factors influence market behaviour through their subjective evaluations. Within this approach the subjective evaluation of the *status quo*, the nature of the alternatives, and the subjective utility contributions of the attitudes towards agricultural concept, measured by orientation (subsistence or commercial), affect the degree of market participation.

The methodology applied is a quantile regression method coupled with the Bayesian adaptive lasso to simultaneously select important covariates and estimate the corresponding quantile regression models. It provides more detailed insights that can allow policy makers to better target those semi-subsistence and semi-commercial farmers who would like to integrate further into the output markets, insights that could not be achieved by the mean regression methods prevailing in most previous research.

The analytical results are meaningful in two aspects. First, the actual structure of the variables included in the model varies across different quantiles. Second, the modelling strategy applied allows for heterogeneous impact, i.e. even when some variables affect the outcome at different quantiles, their impact is allowed to vary and be different for different types of farmers.

The results support the assumption about the heterogeneity of farm households, and the impact of their subjective motivations, attitudes and evaluations of future prospects on their marketing behaviour. Out of 34 variables retained by the model only three are important to all quantiles (17 quantiles altogether). The remaining variables affect some quantiles but not others. What is particularly interesting is that some clustering of effects has emerged since most of the variables tend to affect mainly either the lower or the upper quantiles, i.e. either only the semi-subsistence and semi-commercial farmers or the commercial ones, but not both groups. This indicates that the marginal effects of some subjective factors on farmers' orientation, subsistence or commercial in the Kostov and Lingard (2004) sense, are large for farmers with a specific orientation but not for farmers with a different one. Thus, some subjective factors affect the market integration of semi-subsistence and semi-commercial farmers but not that of the commercialised ones and *vice versa*. This suggests that policy measures have to be more targeted and that blanket measures, if effective at all, may not be effective for the overall policy target group.

The results indicate that what may help some semi-subsistence farmers to integrate into the market is not the traditional agricultural support and/or support similar to measure 141 in the EU Rural Development Regulations for 2007-2013 for semi-subsistence farmers in the NMS undergoing restructuring: rather, it is rural development and job creation in rural areas. The latter will not only increase the living standards of these farmers but will release resources for

more efficient uses in the commercial sector since several of the farmers feel locked in agriculture due to the perception that other options are unavailable .

The results also suggest an important role for the advisory services in facilitating the market integration of farmers selling between 20 and 60% of their output. In the current (2007-2013) EU rural development policy there is support for advisory services amounting to 80% of the eligible cost per service and capped at 1,500 Euro. However, the main focus is on advice on how to keep the land in good agri-environmental conditions (GAEC) and meet the occupational and safety standards, thus the emphasis is on the conditions necessary to receive direct payments from CAP Pillar 1, which is not very beneficial for small and not well market integrated farmers (Council Regulation 1698/2005). In the proposal for the new regulations post-2013, the focus is widened with actions necessary to mitigate climate change and maintain biodiversity – but again nothing specific for the millions of semi-subsistence farmers. There is a general provision that "advice may also cover issues linked to the economic, agricultural and environmental performance of the holding or enterprise" (COM(2011) 627 final/2:13), thus advice on market integration is not precluded but at the same time is not targeted either. The results also show the need for more targeted training and advice to meet EU food safety and quality standards. In many cases these standards require investment. This links to the other policy area that featured prominently in the results – access to credit. It has been felt as a constraint by the lower end of market integrated farmers who want to start up an off-farm business, thus to diversify or to add value to their agricultural products. There are new financial engineering instruments; for example, a rural credit guarantee scheme. However, experience in the use of such instruments is still rather limited (CSIL, Centre for Industrial Studies, 2010).

Recently, there has been a more extensive policy debate in Europe concerning the special policy needs of semi-subsistence and small farmers (ENRD, 2010). In the Commission proposals for the new regulations there is substantial detail concerning a simplified scheme for direct payments to small farmers. However, the results here as well as several previous studies (see e.g. Davidova, 2011) argue for support for rural development (CAP Pillar 2). The measures in Pillar 2 are multi-annual and can be tuned to specific regional needs. In the Commission proposal for a Regulation for Rural Development post-2013 'small farmers' constitute a 'thematic sub-programme' (COM(2011) 627 final/2). In principle, the Member States are allowed to give higher support rates for some activities covered by the sub-

programmes but the probability that low market integrated farmers will be the beneficiaries of such national decisions is low since they have to compete with the remaining sub-programmes, including setting up young farmers, support to mountainous areas and support for short-supply chains. It is unlikely that the least market integrated farmers, who usually have small farms, will have a very strong and organised voice and be capable of influencing decision-making in the direction of their interests.

References

Alhamzawi, R., Yu, K. and Benoit, D.F. *Bayesian adaptive LASSO quantile regression*, Working paper 2011/728 (University of Gent, Faculty of Economics and Business Administration, July 2011).

Barrett, C. 'Smallholder market participation: Concepts and evidence from Eastern and Southern Africa', *Food Policy* Vol. 33, (2008) pp. 299-317.

Belloni, A. and Chernozhukov, V. *L1-penalized quantile regression in high-dimensional sparse models*, Working Paper 10/09 (Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 2009).

Bester, H. 'Screening versus Rationing in Credit Markets with Imperfect Information', *American Economic Review* Vol. 75, (1985) pp. 850-855.

Bezemer, D.J. 'Credit Markets for Agriculture in the Czech Republic', *Europe-Asia Studies*, Vol. 54, (2002) pp. 1301-1317.

Cadot, O., Dutoit, L. and Olarreaga, M. *How costly is it for poor farmers to lift themselves out of subsistence?* World Bank Policy Research Working Paper No 3881 (The World Bank, April 2006).

Chaserant, C.. 'Cooperation, contracts and social networks: From a bounded to a procedural rationality approach', *Journal of Management and Governance*, Vol. 7, (2003) pp 163-186.

Ciaian, P. and Swinnen, J.F.M. 'Land market imperfections and agricultural policy impacts in the new EU Member States: a partial equilibrium analysis', *American Journal of Agricultural Economics,* Vol. 88, (2006) pp. 799-815.

COM(2011) 627 final/2 Proposal for a Regulation of the European Parliament and of the Council on support for rural development by the European Agricultural Fund for Rural Development (EAFRD), Brussels 19.10.2011.

Council Regulation (EC) No 1698/2005 on support for rural development by the European Agricultural Fund for Rural Development (EAFRD), Official Journal of the European Union, L277 21.10.2005.

Csaki, C., Forgacs, C., Milczarek-Andrzejewska, D. and Wilkin, J. (eds.) *Restructuring Market Relations in Food and Agriculture in Central and Eastern Europe: Impact upon Small Farmers* (Budapest: Agroinform Publisher Co. Ltd, 2008).

CSIL, Centre for Industrial Studies. Lessons from shared management in cohesion, rural development and fisheries policies (Final report, Milan 2010) available at

http://ec.europa.eu/regional_policy/sources/docgener/studies/pdf/shared_management/shared_management_final_report.pdf (last accessed 8 January 2012).

Davidova, S. 'Semi-subsistence farming: An elusive concept posing thorny policy issues', Presidential address, *Journal of Agricultural Economics* Vol. 62, (2011) pp. 503-524.

Davidova, S., Fredriksson, L. and Bailey, A. 'Subsistence and semi-subsistence farming in selected EU new Member States', *Agricultural Economics,* Vol. 40, (2009) pp. 733-744.

Davidova, S., Fredriksson, L., Gorton, M., Mishev, P. and Petrovici, D. 'Subsistence farming, incomes and agricultural livelihoods in the new Member States of the European Union, *Environment and Planning C Government and Policy* (forthcoming).

Davis, J., Buchenrieder. G. and Thomson, K. 'Rural Credit for Individual Farmers: A Survey Analysis of Rural Financial Markets', in Davidova S. and Thomson K. (eds) (2003) *Romanian Agriculture and Transition toward the EU* (Lanham/Boulder/New York/Oxford, Lexington Books).

De Janvry, A., Fafchamps, M. and Sadoulet, E. (1991) 'Peasant household behaviour with missing markets: Some paradoxes explained', *The Economic Journal,* Vol. 101, (1991) pp. 1400-1417.

Dries, L. and Swinnen, J.F.M. 'Foreign direct investment, vertical integration and local suppliers: Evidence from the Polish dairy sector', *World Development*, Vol. 32, (2004) pp. 1525-1544.

European Network for Rural Development (ENRD). *Semi-subsistence farming in Europe: Concepts and key issues* (Background paper prepared for the seminar "Semi-subsistence farming in the EU: Current situation and future prospects", Sibiu Romania 13-15 October 2010) available at: http://enrd.ec.europa.eu/app_templates/filedownload.cfm?id=FB3C4513-AED5-E24F-E70A-F7EA236BBB5A (last accessed 1 November 2011).

Fan, J. and Li, R. (2001), 'Variable selection via nonconcave penalized likelihood and its oracle properties', *Journal of the American Statistical Association*, Vol. 96, (2001) pp. 1348-1360.

Fritzsch, J., Wegener, S., Buchenrieder, G., Curtiss, J. and Gomez y Paloma, S. 'Is there a future for semi-subsistence farm households in Central and Southeastern Europe? A multiobjective linear programming approach', *Journal of Policy Modeling*, Vol. 33, (2011) pp. 70-91.

Gilks, W. R. Derivative-free adaptive rejection sampling for Gibbs sampling, in Bernardo, J., Berger, J., Dawid, A. P. and Smith A. F. M (eds), *Bayesian Statistics 4,* (1992), pp. 169-94, Clarendon Press, Oxford.

Gow, H. and Swinnen, J. F. M. 'Private enforcement capital and contract enforcement in transition economies', *American Journal of Agricultural Economics*, Vol. 83, (2001) pp. 686-690.

Hans, C. 'Bayesian lasso regression', *Biometrika*, Vol. 96, (2009) pp. 835–845.

Jaffee, D. M. and Russell, T. 'Imperfect Information, Uncertainty, and Credit Rationing', *The Quarterly Journal of Economics*, Vol. 90, (1976) pp. 651-666.

Key, N., Sadoulet, E. and De Janvry, A. 'Transactions costs and agricultural household supply response, *American Journal of Agricultural Economics,* Vol, 82**,** (2000) pp. 245-259.

Koenker, R. and Bassett, G. 'Regression quantiles', *Econometrica*, Vol. 46, (1978) pp. 33-50.

Koenker, R. and Machado, J. 'Goodness of fit and related inference processes for quantile regression', *Journal of the American Statistical Association*, Vol. 94, (1999) pp. 1296-1310.

Koenker, R. *Quantile Regression* (Econometric Society Monograph Series, Cambridge: Cambridge University Press, 2005).

Kostov, P. and J. Lingard 'Subsistence agriculture in transition economies: its roles and determinants', *Journal of Agricultural Economics*, Vol. 55, (2004) pp. 565-579.

Kostov, P. and J. Lingard 'Subsistence farming in transitional economies: lessons from Bulgaria', *Journal of Rural Studies*, Vol. 18, (2002) pp. 83-94.

Latruffe, L. 'The impact of credit market imperfections on farm investment in Poland', *Post-Communist Economies*, Vol. 17, (2005) pp. 349-362.

Latruffe, L., Davidova, S. and Desjeux, Y. Perpetuation of subsistence farming in Western Balkans: The role of factor market imperfections. *82nd Annual Conference of the UK Agricultural Economics Society (AES)* (Royal Agricultural College, 31 March-2 April 2008).

Latruffe, L., Davidova, S., Douarin, E. and Gorton, M. 'Farm expansion in Lithuania after accession to the EU: The role of CAP payments in alleviating potential credit constraints', *Europe-Asia Studies*, Vol. 62, (2010) pp. 351-365.

Li, Q., Xi, R. and Lin, N. 'Bayesian regularized quantile regression', *Bayesian Analysis*, Vol. 5, (2010) pp. 1-24.

Li, Y. and Zhu, J. 'L1-Norm quantile regression', *Journal of Computational and Graphical Statistics,* Vol. 17, (2008) pp. 1-23.

Langlois, R. (1992). 'Transaction cost economics in real time', *Industrial and Corporate Change*, Vol1, (1992) pp. 99-127.

Löfgren, H. and Robinson, S. 'Nonseparable farm household decisions in a computable general equilibrium model', *American Journal of Agricultural Economics,* Vol. 81**,** (1999) pp. 663-670.

Macours, K. and Swinnen, J. F. M. 'Agricultural Labor Adjustments in Transition Countries: The Role of Migration and Impact on Poverty', *Review of Agricultural Economics,* Vol. 27, (2005) pp. 405-411.

Mathijs, E. and Noev, N. 'Subsistence farming in Central and Eastern Europe: Empirical evidence from Albania, Bulgaria, Hungary, and Romania', *Eastern European Economics,* Vol. 42**,** (2004) pp. 72-89.

Miracle, M. 'Subsistence agriculture: Analytical problems and alternative concepts', *American Journal of Agricultural Economics*, Vol. 50, (1968) pp. 292-310.

Park, T. and Casella, G. 'The Bayesian LASSO', *Journal of the American Statistical Association,* Vol. 103, (2008) pp. 681-686.

Petrick, M. 'A microeconometric analysis of credit rationing in the Polish farm sector', *European Review of Agricultural Economics*, Vol. 31, (2004) pp. 77-101.

Petrick, M. and Tyran, E. 'Development perspective of subsistence farmers in Southeastern Poland: Social buffer stock or commercial agriculture' in Abele, S. and Frohberg, K. (eds.), *Subsistence agriculture in Central and Eastern Europe: How to break the vicious circle* (conference proceedings, IAMO: Halle, 2003).

Petrovici, D. and Gorton, M. 'An evaluation of the importance of subsistence food production for assessments of poverty and policy targeting: Evidence from Romania', *Food Policy,* Vol. 30**,** (2005) pp. 205-223.

Pingali, P. And Rosegrant, M. 'Agricultural commercialization and diversification: processes and policies', *Food Policy*, Vol. 20, (1995) pp. 171-185.

Reed, C., Dunson, D. and Yu, K. *Bayesian variable selection in quantile regression*, Technical report (Department of Mathematical Sciences, Brunel University, 2009).

Rizov, M. and Swinnen, J. F. M. 'Human capital, market imperfections, and labor reallocation in transition', *Journal of Comparative Economics,* Vol. 32, (2004) pp. 745-774.

Stiglitz, J. E. and Weiss, A. 'Credit rationing in markets with imperfect information', *American Economic Review*, Vol. 71, (1981) pp. 393-410.

Swinnen, J. and Gow, H. 'Agricultural credit problems and policies during the transition to a market economy in Central and Eastern Europe', *Food Policy*, Vol. 24, (1999) pp. 21-47.

Swinnen, J. and Vranken L. *Review of transitional restrictions on acquisition of agricultural land in NMS* (Brussels: CEPS, 2008) available at: http://ec.europa.eu/internal_market/capital/docs/study_en.pdf (last accessed 6 January 2012).

Tibshirani, R. J. 'Regression shrinkage and selection via the LASSO', *Journal of the Royal Statistical Society, Series B*, Vol. 58, (1996) pp. 267-288.

Tsionas, E. G. 'Bayesian quantile inference', *Journal of Statistical Computation and Simulation*, Vol. 73, (2003) pp. 659-674.

Wharton, C. R. Jr. 'Subsistence agriculture: Concepts and scope, in Wharton, C. R. Jr (ed), *Subsistence Agriculture and Economic Development* (New Brunswick, NJ: Transaction Publishers, 1969, pp.12-20).

Wu, Y. and Liu, Y. 'Variable selection in quantile regression', *Statistica Sinica*, Vol. 19, (2009) pp. 801–817.

Yu, K. and Moyeed, R. A. 'Bayesian quantile regression', *Statistics and Probability Letters*, Vol. 54, (2001) pp. 437-447.

Zou, H. 'The adaptive lasso and its oracle properties', *Journal of the American Statistical Association*, Vol. 101, (2006) pp. 1418-1429.

Zou, H. and Yuan, M. 'Regularized simultaneous model selection in multiple quantiles regression', *Computational Statistics & Data Analysis*, Vol. 52, (2008) pp. 5296–5304.

Table 1 Coefficient estimates

Quantile	q=0.05	q=0.10	q=0.15	q=0.20	q=0.25	q=0.30	q=0.35	q=0.40	q=0.45	q=0.50	q=0.55	q=0.60	q=0.65	q=0.70	q=0.75	q=0.80	q=0.85
Share	0.2201	0.4547	0.5000	0.5275	0.6091	0.6673	0.7128	0.7500	0.8157	0.8571	0.8824	0.9023	0.9267	0.9571	0.9822	0.9998	1.0000
E14																0.02018	0.02073
E15b												0.01948	0.01917	0.01885	0.01847	0.01732	0.01665
E15g	0.03922	0.03580	0.032	0.02676	0.02519					0.01774	0.01868						
E15h	-0.04040	-0.03721	-0.0311	-0.0253													
E33												-0.01840	-0.01936	-0.01974	-0.02009	-0.02090	-0.02069
F7	0.09872	0.08125	0.0605	0.0516	0.0450	0.0389											
H2b	-0.04573	-0.05020	-0.0512	-0.0507	-0.0490	-0.0461	-0.0432	-0.0400	-0.0372	-0.03367	-0.03011	-0.02778	-0.02519	-0.02340	-0.02148	-0.02050	-0.01903
H2c	0.02441																
H2e	0.02946		0.0288	0.0310	0.0364	0.0365	0.0307									0.02189	
H3c	0.04956	0.04452	0.0398	0.0341	0.0286												
H3d																	0.01499
H6d						0.0298		0.0249				0.02097					
H6e																0.01816	0.01833
H6f	0.05750	0.05160	0.0457	0.0383	0.0326	0.0279	0.0233										
H6g		-0.02769		-0.0270	-0.0260	-0.0280	-0.0284	-0.0284	-0.0294	-0.03002	-0.03002	-0.03116	-0.03097	-0.03252	-0.03192	-0.03241	-0.03368
K3														0.02774	0.02955	0.02596	0.02870
K4				0.0246	0.0271	0.0259											
K5	0.05597	0.05334	0.0480	0.0359													
K8b		-0.02446	-0.0277	-0.0279	-0.0295	-0.0274	-0.0256	-0.0227									
K8c																	-0.01596
K9b	-0.02134																
K9e	0.02953	0.02992	0.0301	0.0288	0.0307	0.0330	0.0340	0.0331	0.0318	0.03197	0.03041	0.02867	0.02824	0.02673	0.02493	0.02285	0.02197
K10a						-0.0271	-0.0267	-0.0250		-0.02391		-0.02182	-0.02046		-0.01797	-0.01668	-0.01634
K10b		0.03256	0.0334	0.0359	0.0354	0.0335	0.0298										
I6_2006	-0.09692																

G9a	0.02476	0.02389	0.0219	0.0207	0.0198												
G9d															-0.01717	-0.01695	-0.01669
G9e	0.03195	0.02675													0.01958	0.01983	0.02010
G9f	-0.03918	-0.03624	-0.0323	-0.0274	-0.0253	-0.0239	-0.0218	-0.0200									
G9j	0.03228	0.02983	0.0263	0.0261	0.0254	0.0243	0.0225	0.0217	0.0204	0.01934	0.01867	0.01822	0.01809	0.01793	0.01766	0.01745	0.01763
G9n			0.0287	0.0313	0.0286	0.0246											
G9o									-0.0216	-0.02158	-0.02156	-0.02002	-0.01914	-0.01779	-0.01624	-0.01601	-0.01534
G9q	-0.03669	-0.03087	-0.0281			-0.0203	-0.0213	-0.0221									
G10						0.05189	0.0644	0.06932	0.06446	0.06234	0.05995	0.05520	0.05193	0.04786	0.04988	0.04274	0.04369

Table 2. Description of selected variables

Variable	Description	Measure
E14	Does your household plan to earn an income through starting up an own off-farm business within the next 5 years?	(1) very unlikely to (5) very likely
E15	**Rate the importance of the following factors for why no household member currently works self-employed in a non-farm business**	**(1) Not important to (5) Very important**
E15b	Household members prefer agricultural work	
E15g	Insufficient availability of equity capital	
E15h	Insufficient availability of low cost credits	
E33	Does your household plan to earn an income through taking up wage employment within the next 5 years?	(1) very unlikely to (5) very likely
F7	How is your overall cash incomes and consumption in kind comparing to 2003	(1) much worse off to (5) much better off
H2	**Statements regarding your current aims for agricultural production**	**(1) totally disaggree to (5) totally aggree**
H2b	To provide work for household members	
H2c	To transfer to the next generation	
H2e	To generate cash income	
H3	**Statements about agricultural production**	**(1) totally disaggree to (5) totally aggree**
H3c	We use agricultural business advice and information	
H3d	We fully employ household members	
H6	**Possible constraints to increase agricultural production**	**(1) totally disaggree to (5) totally aggree**
H6d	We receive low prices for our agricultural output	
H6e	We lack the necessary skills and education	
H6f	We lack information and advice on markets and prices	
H6g	We cannot meet the standards of buyers or public regulations	
K3	Has a successor for the farm already been designated by the household?	Yes=1
K4	There is a potential successor, but we do not know whether he/she will really continue the farming activities	Yes=1
K5	Evaluation of the economic prospects of the farm within a timeframe of 5 years	(1) Not competitive/Low profitability to (5) Very competitive/High profitability
K8	**How would the following influence the probability of your household investing in farming**	**(1) No influence to (5) High influence**
K8b	Contracts with buyers	
K8c	Access to agricultural credit	
K9	**How would the following influence the probability of your household ceasing with agricultural production within the next 5 years?**	**(1) No influence to (5) High influence**
K9b	Better opportunities to work off-farm	
K9e	Access to credits to invest in non-farm business	
K10	**Judgement about household's abilities to adapt to the following EU regulations**	**1) Very difficult to (5) Very easy**
K10a	Veterinary and phytosanitary standards	
K10b	Food safety and quality standards	
I6_2006	How do you judge the contribution of your own food production to your household's welfare	(0) Not Important to (2) Very important

G9	**Agreement with statements about buying/renting in and selling/renting out land**	**(1) totally disaggree to (5) totally aggree**
G9a	We would like to buy more land	
G9d	Legal procedures make it difficult to buy land	
G9e	It is difficult to know if there are potential conflicts over rights in land	
G9f	We would like to rent in more land	
G9j	We would like to sell land	
G9n	We do not have clear titles for land we wish to sell	
G9o	We would like to rent out land	
G9q	It is difficult to find people willing to rent land	
G10	According to current regulations, is it possible for your household to sell a part of the garden or yard currently belonging to your house without also selling the house?	Yes=1

Appendix 1

The quantile regression is the solution to the following optimisation problem:

$$\min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau \left(y_i - X^T \beta_\tau \right)$$

(A.6)

This minimisation problem is equivalent to maximising the likelihood function of y_i assuming that y_i is a random variable from a skewed Laplace distribution with location parameter $\mu = X_i^T \beta_\tau$ and scale parameter σ, which leads to the Bayesian interpretation of the quantile regression. This means maximising the following density function:

$$f\left(y \mid \mu, \sigma, \tau \right) = \sigma \tau \left(1 - \tau \right) \exp \left(-\sigma \rho_\tau \left(y - \mu \right) \right)$$

(A.7)

The skewed Laplace distribution, assumed for the quantile regression residuals can be represented as a scale mixture of normal distributions (Tsionas, 2003) as:

$\theta z + \phi u \sqrt{\sigma^{-1} z}$. where

$\theta = \dfrac{1 - 2\tau}{\tau \left(1 - \tau \right)}$ and $\phi^2 = \dfrac{2}{\tau \left(1 - \tau \right)}$. In the above z follows an exponential distribution with mean σ^{-1}

and and u is a standard normally distributed variable. Furthermore z and u are assumed to be mutually independent.

The Bayesian adaptive lasso quantile regression model can be obtained by inserting the above re-parameterisation of the skewed Laplace distribution and complementing it with suitable (shrinkage) priors for the coefficients, as below:

$$y_i = \beta_0 + x_i^T \beta + \theta z_i + \phi u_i \sqrt{\sigma^{-1} z_i},$$

$$p(\beta_0) \propto 1,$$

$$p(u_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u_i^2}{2}\right),$$

$$p(z_i \mid \sigma) = \sigma \exp(-\sigma z_i),$$

$$p(\beta_j, s_j \mid \sigma, \lambda_j^2) = \frac{1}{\sqrt{2\pi s_j}} \exp\left(-\frac{\beta_j^2}{2s_j}\right) \frac{\sigma}{2\lambda_j^2} \exp\left(\frac{-\sigma s_j}{2\lambda_j^2}\right),$$

$$p(\lambda_j^2 \mid \delta, \kappa) = \frac{\kappa^\delta}{\Gamma(\delta)} (\lambda_j^2)^{-1-\delta} \exp\left(-\frac{\kappa}{\lambda_j^2}\right),$$

$$p(\sigma) = \sigma^{a-1} \exp(-b\sigma),$$

$$p(\kappa, \delta) = \kappa^{-1} \tag{A.8}$$

The assumed Laplace priors on the coefficients and inverse gamma priors on λ_j^2 (rather than on λ_j) with hyper parameters $\kappa > 0$ and $\delta > 0$ complement the specification. The smaller values for κ and larger values for δ result in a higher amount of shrinkage. Each coefficient β_j is regularised by a lasso–type penalisation parameter of $\frac{\sqrt{\sigma}}{\lambda_j}$. The above representation leads to the following posterior:

$$p(\beta_0, \beta, z, s, \sigma, \lambda_j \mid y, X)$$

$$\propto p(y \mid \beta_0, \beta, z, s, \sigma, X) \prod_{i=1}^n p(z_i \mid \sigma) \prod_{i=1}^n p(\beta_i, s_j \mid \sigma, \lambda_j^2) p(\lambda_j^2 \mid \kappa, \delta) p(\sigma) p(\kappa, \delta)$$

$$\propto \prod_{i=1}^n \frac{\sigma}{\sqrt{\sigma^{-1} \phi^2 z_i}} \exp\left(-\frac{\sigma (y_i - \beta_0 - x_i^T \beta - \theta z_i)^2}{2\phi^2 z} - \sigma z_i\right)$$

$$X \prod_{j=1}^k \left\{ \frac{1}{\sqrt{2\pi s_j}} \exp\left(-\frac{\beta_j^2}{2s_j}\right) \frac{\sigma}{2\lambda_j^2} \exp\left(-\frac{\sigma s_j}{2\lambda_j^2}\right) \frac{\kappa^\delta}{\Gamma(\delta)} (\lambda_j^2)^{-1-\delta} \exp\left(-\frac{\kappa}{\lambda_j^2}\right) \right\} \kappa^{-1} \sigma^{a-1} \exp(-b\sigma)$$

The actual implementation algorithm for a predetermined (fixed) quantile works as follows:

1.

Simulate the conditional intercept $\beta_0 \sim N\left(\frac{1}{n} \sum_{i=1}^n (y_i - \beta_0 - x_i^T \beta - \theta z_i), \left(\frac{\sigma \phi^2}{n^2}\right) \sum_{i=1}^n z_i\right)$

2.

Simulate $z_i^{-1} \sim$ *Inverse Gaussian* $\left(\sqrt{\dfrac{\theta^2 + 2\phi^2}{\left(y_i - \beta_0 - x_i^T \beta\right)^2}}, \dfrac{\sigma\left(\theta^2 + 2\phi^2\right)}{\phi^2} \right)$

3.

Simulate $\beta_j \sim N\left(\overline{\beta}_j, \hat{\sigma}_j^2\right)$, with $\hat{\sigma}_j^2 = \left(\sigma\phi^{-2} \sum\limits_{i=1}^{n} x_{ij}^2 z_j^{-1} + s_j^{-1} \right)^{-1}$ and

$$\overline{\beta}_j = \hat{\sigma}_j^2 \sigma\phi^{-2} \sum\limits_{i=1}^{n} x_{ij}^2 z_j^{-1} \left(y_i - \beta_0 - \sum\limits_{l \neq j} x_{il}\beta_l - \theta z_i \right)$$

4.

Draw $s_j \sim$ *Inverse Gaussian* $\left(\sqrt{\dfrac{\beta_j^2 \lambda_j^2}{\sigma}}, \beta_j^2 \right)$

5.

Simulate $\sigma \sim Gamma(a_1, a_2)$ where $a_1 = \dfrac{3n}{2} + k + a$ and

$$a_2 = \sum\limits_{i=1}^{n} \left(\dfrac{\left(y_i - \beta_0 - \sum\limits_{l \neq j} x_{il}\beta_l - \theta z_i \right)^2}{2\phi^2 z_i} + z_i \right) + \sum\limits_{j=1}^{k} \dfrac{s_j}{2\lambda_j^2} + b$$

6.

Simulate $\lambda_j^2 \sim$ *Inverse Gamma* $\left(1 + \delta, \sigma s_j / 2 + \kappa \right)$

7.

Simulate $\kappa \sim Gamma\left(k\delta, \sum\limits_{j=1}^{k} \lambda_j^{-2} \right)$

8.

Simulate δ from its conditional posterior distribution given by:

$$p(\delta | .) \propto \left(\Gamma(\delta) \right)^{-k} \kappa^{k\delta} \prod\limits_{j=1}^{k} \lambda_j^{-2\delta}$$

All the above steps are straightforward except the last one, since the full conditional posterior distribution of δ does not have a closed form. Note however that it is log-concave. Therefore, we use the adaptive rejection sampling algorithm (Gilks, 1992) to sample from it. The adaptive rejection sampling step, necessary to estimate the individual amount of shrinkage, is the main additional computational cost associated with the Bayesian adaptive lasso quantile regression, since in both the standard Bayesian quantile regression and in the Bayesian lasso implementation

of Li *et al*. (2010) the faster standard Gibbs sampler can be used to simulate the full posterior

distribution – the latter is the prevailing practice in more recent empirical work.