Can direct payments facilitate agricultural commercialisation: Evidence from a transition country

Philip Kostov, Sophia Davidova, Alastair Bailey, Ekrem Gjokaj and Kaplan Halimi[1]

[Original submitted March 2020, Revision received May 2020, Accepted May 2020]

Abstract

We investigate the impact of direct payments on agricultural commercialisation in Kosovo. Kosovo is one of the poorest countries in Europe but, with European assistance, provides substantial funds to support agriculture, made up of many small and often semi-subsistence farms. Thus, the effect of this support is a central policy issue. Identifying the effect of direct payments on market participation faces endogeneity issues arising from the possible simultaneous determination of participation in support programmes and market participation. In order to achieve proper identification of the endogenous direct payments, we use a strategy of targeted identification search that employs several different methodological approaches. We find that direct payments for fruit and vegetables, and those for cereals and oilseeds have a positive effect on market participation. However, we cannot identify any definite effect of livestock payments.

Key words: direct coupled payments, market participation, endogeneity, identification, Kosovo

Jel Classifications: C13: C26; Q12; Q18

Introduction

During the transition and pre-EU accession period, several Central and Eastern European countries (CEECs) used 'CAP-like' policies (Swinnen 1994). This was a deliberate step to resemble some of the EU Common Agricultural Policy (CAP) instruments in order to facilitate accession to the EU and policy adjustments in agriculture. These developments during the transition period should not be interpreted as harmonisation with the CAP, as some measures in transition economies only loosely resembled the CAP instruments (Tangerman, 1998). The main objective of the CAP-like policy support was to boost output, which had declined substantially during the transition from communist rule. Kosovo is also following this policy trajectory, although since the country began pre-accession later than the CEECs, the policy measures resemble the CAP architecture after McSharry reform and, in particular, after the Agenda 2000 establishing Pillar2 – Rural Development. Major policy instruments introduced in Kosovo include direct payments coupled to production and rural development grants, with substantial financial assistance from the EU.

Direct payments, similarly to other output subsidies, 'leak' into the upstream and downstream sectors and also partially capitalise in land values. On the other hand, through their productivity and production enhancing effects, direct payments keep land in cultivation, even land of more marginal quality, in comparison to what would have been the case under free market

[1] Philip Kostov (contact author: PKostov@uclan.ac.uk) is at the University of Central Lancashire, UK. Sophia Davidova and Alastair Bailey are both at the University of Kent, UK. Ekrem Gjokaj and Kaplan Halimi are both with the Ministry of Agriculture, Forestry and Rural Development, Kosovo.

conditions. They keep more farmers in farming and thus may mitigate rural outmigration and depopulation of rural areas. This is important for Kosovo where urban unemployment is high (the unemployment rate 2018 was 29.6% (KAS, 2019)), there is lack of non-agricultural rural jobs and there is widespread underemployment of family members on-farm (Osmani et al., 2013).

Differently to the EU, in the Western Balkans there is a large semi-subsistence sector, which is only partially or often not at all integrated into the market system. Our objective is to investigate the effect of direct payments on commercialisation of semi-subsistence family farms in Kosovo. This is an important economic and social issue. Davidova (2011), basing her rationale on Barrett (2008), argues that commercialisation of semi-subsistence farmers is desirable, since restructuring the semi-subsistence sector would increase the wellbeing of farmers, at least in the long run. The distribution of risk free direct payments to farmers adds liquidity, which helps purchase inputs, assists technological improvement, and increases and output. The question of whether direct payments affect market participation in countries with predominant small family semi-subsistence farmers has important policy implications. If there is a positive relationship, this can provide justification for and guide the design of support policies.

Typically direct payment recipients have to meet some conditions, e.g. in Kosovo a requirement is to be of a minimum size. In the case that these requirements only make more commercially oriented farms eligible for direct payments support, the expected effects on commercialisation of semi-subsistence farmers would not materialise, since the semi-subsistence sector will be largely excluded. The result is that empirical analysis can be expected to show correlation between the receipt of direct payments and the degree of market integration, but such correlation does not imply a casual effect. Technically, this problem is a type of endogeneity in that the potential explanation (i.e. direct payments) and the outcome are simultaneously affected by the design of the policy support system, since more market integrated farms are more likely to be eligible for direct payments. We address the question on whether direct payments encourage market participation by explicitly accounting for the endogeneity issues present in the empirical design.

The paper is organised as follows. Section two conceptualises the possible effects of direct payments on the market participation decisions of semi-subsistence farmers. The third section presents a short overview of the design of direct payment programme and the eligibility requirements in Kosovo. This overview also provides a guide to the search for identification of the endogenous direct payments in explaining market integration. The methodology used is described in section four, followed by presentation of the data sources and measures used in the model. This is followed by explanations of the logic behind the specification search and the approach we use to obtain the results. The identification strategy and the construction of different instruments for each policy measure are discussed in section five. This leads to the analytical results, presented in section six. Finally, some conclusions and policy implications are developed.

Conceptual Framework

We use a simple household model to conceptualise the possible effects of direct payments on market participation. Household models are standard tools for analysing semi-subsistence farms and we want to identify the effect of direct payments on the marketing behaviour of such a semi-subsistence farm. The standard non-separable household model is widely used to explain differentiated market access (see e.g. Barrett, 2008 for a detailed overview). The basic assumption is that there are market imperfections and farmers face transaction costs in accessing markets. These costs are household specific.

A representative household maximises its utility (U) over consumption of a vector of agricultural products (*c*) c=1,....C and a composite of all other tradables (*x*) subject to production function and cash income constraints.

Max U *(c,x)*

The consumption of agricultural products (*c*) originates from two sources – from self-produced products (c^s) and from products purchased in the market (c^m).

Following Barrett (2008), agricultural output (*Y*) is a function of the flow of services from private assets (*A*) (land, labour, capital) and public services, e.g. physical and market infrastructure, public standards, extension service (*G*).

$Y = f(A,G)$

Agricultural output is divided into two uses, self-consumption (y^s) and sales (y^m); the self-consumed output (y^s) is equal to the consumption coming from own production (c^s).

The cash income is earned by sales of agricultural products y^m, through non-agricultural enterprises and/or off-farm wage employment *W*, and through unearned income (including a non-risky flow of policy income payments) *TP*.

However, when deciding to sell its produce, the household faces transaction costs per unit of output sold $\tau^{cs}(A,G,W,TP,y^m)$ (Barrett, 2008). The logic behind this specification of transaction costs is simple. *A* reflects the household specific assets and also the characteristics of labour such as education and training which may affect the search costs or the willingness to market the output jointly with other farmers. *G* reflects public infrastructure and services including roads, price information, marketing advice, advice on meeting standards that can decrease substantially the transport, search and enforcement costs, and facilitate the contractual relations with the downstream sector. *W* and *TP* may decrease the transaction costs in accessing credit in that they increase household's credit 'rating'. Furthermore both *W* and *TP* can create wealth effects in a risky choice framework and, hence, decrease risk averseness thus further reducing transaction costs. Note that a larger volume of sales y^m helps spread the fixed transaction costs over more units and, thus, reduces the total transaction costs per unit. Since direct payments are often directly (e.g. for milk) linked to requirements for a minimum sold quantities (hence increasing y^m), or indirectly having similar effect via requirements to maintain production, they can be expected to reduce unit transaction costs. Therefore, we have three distinct channels through which direct payments can reduce transaction costs, i.e. W, TP and y^m.

In this case, the cash income constraint can be written as follows:

$$p^x x + \sum_{c=1}^{C} p^{c*} c^m = \sum_{c=1}^{C} y^m (p^{c*} - \tau^{cs})(1 + d^c) + p^{s*} y^s + W + TP \qquad (1)$$

where p^x is the price of all non-agricultural tradables, x is their quantity and p^{c*} is the equilibrium market price of agricultural products. The LHS represents the household's expenditure with the income recorded in the RHS. The self-consumption is recorded as 'income'. The direct payments effects are assumed to work through two separate channels: by directly affecting output; ~~via a unitary effect~~ indirectly, through 'wealth' type effects. The balance of these two effects depends on the nature of the product and the specific requirements of the direct payment.

Direct payments are expected to increase market participation through reducing transaction costs, or helping to pay them. This transaction cost reduction can follow from several distinct impacts. First, the production enhancing effect of the DP potentially reduces the per unit transaction costs τ^{cs}. Second, the production effect of the direct payments can be either viewed as increasing the marketed output from $\sum_{c=1}^{C} y^m$ to $\sum_{c=1}^{C} y^m(1 + d^c)$ or as acting as a transaction cost reducing effect via changing the product specific transaction cost losses from $(p^{c*} - \tau^{cs})$ to $(p^{c*} - \tau^{cs})(1 + d^c)$ which since $d^c > 0$ and hence$(1 + d^c) > 1$ also relaxes the budget constraint. Finally, the wealth effect of the DP (via including part of them in TP) also relaxes the budget constraint.

~~Since DPs increase the RHS (income part) of the budget constraint, this needs to be offset by equivalent increase in the LHS (expenditure), equivalent decrease in other terms in the RHS (decrease in other income) or combination thereof. The overall increase in production needs to be distributed amongst self-consumption and marketable surplus. Except in cases of extreme poverty, when basic household needs are not fully met (and therefore the relaxation of the budget constraint may lead to a large increase in p^x x), it can be expected that most of the increase in overall production will be translated into marketable surplus and, hence, will increase market participation. Therefore, a larger increase in the marketable output than the own consumption is expected to lead to larger market participation. Realistically, in addition to extreme poverty, there is the case of well-off households, where increased income may lead to increase in savings. Our model does not include savings explicitly but accounts for them via the non-agricultural tradeables. This is intentional since we focus on semi-subsistence households, where the likely proportion of additional income leading to increase in savings is expected to be relatively small.~~

Additional income, whether coming from direct transfer (the monetary value of direct payments), decrease in transaction costs, or increase in sold output (which can follow from both transaction costs reductions and requirements of the direct payments) may typically displace self-consumption.

Furthermore, DPs can be viewed as risk-free cashflows since they do not fluctuate with market conditions. By reducing risk in the business, they also reduce the raison d'etre for subsistence, namely extreme risk avoidance by insulating oneself from market fluctuations. Hence, DPs can crowd out subsistence behaviour.

These effects, however, can be attenuated, or even negated in the presence of market imperfections. From income (or welfare) perspective, as argued in Barrett (2008), the market participation choice is conceptually equivalent to the production technology choice. The overall income and welfare effects then depend on the nature of market transmission. The additional output may put pressure on market prices, offsetting the income effect of the DPs. If the market size is small, any increase in output and hence in supply will have larger effect on reducing prices This effect will be smaller in better integrated markets (i.e. larger markets), since the additional output will be supplied to a larger market. In fragmented local markets, the price of output may fall more sharply compared to the case of an integrated national market, or if the extra output can be exported. In Kosovo, agricultural markets are not well integrated (either nationally or, particularly, internationally) so the price effects are likely to be greater, since the effective demand is likely to be substantially less elastic., although local demand for fruit and vegetables and livestock products is expected to be more price elastic than for more basic staple foods. It is to be expected that households' responses to DPs are more likely to involve increasing production of products with a strong market demand than those which have potential marketing difficulties.

~~Another argument relates to the consumption bundle of a household.[2] The household meets its consumption needs by either producing and self-consuming a desired quantity of a given product or alternatively producing a different product for the market and using the generated income to purchase the desired product. Therefore, the income realised from the increased output will indirectly affect the output of other products. In this case it is more efficient to choose to increase output, either without policy support, or due to DPs, of products for which the adverse income effects, discussed above, will be smaller. This means products, which have better integrated marketing chains. Such choices are expected to lead to greater increase in commercialisation since they will displace output of products more focused on own consumption.~~

~~The above narrative implies that DPs on products that transcend the domestic markets are to be expected to have greater impact of market participation compared to more locally marketed products such as fruits and vegetables. The above discussion however assumes that DPs are universally available to all farms. This however is not the case.~~ There are eligibility requirements for receiving DPs, usually expressed in terms of size (acreage for eligible crops or heads of animals), and exclude the smallest (more subsistence oriented) farms. Lower payment thresholds are expected to generate larger effects on market participation.

~~Note however that lower thresholds for fruit and vegetable products that are mostly sold on regional markets, mean that DPs will reach more semi-subsistence farms and hence have greater potential for increasing market participation, but can (if used excessively) carry the danger of adverse income effects, which under a rational decision-making framework may result in greater self-selection of subsistence farmers out of such programmes. Direct estimation of market participation effects that ignores such selectivity issues is likely to result in upward bias in estimating such effects.~~

Direct payments in Kosovo

Kosovo is currently a potential candidate for EU membership, and is eligible for EU funding to harmonise domestic policies to the EU acquis. In order to support farmers, but at the same time to increase the absorption capacity, the Agency for Agricultural Development (Paying

[2] In subsistence type of models we can view the farm as a household.

Agency) in Kosovo has implemented a direct payments system. The level of direct payments is different for different products, but they are all explicitly linked to current production levels. The budget for the direct payments programme has slowly but steadily increased, driven primarily by the increase in the EU funds, so that in 2017 and 2018 direct payments spending was €26 and €27 million respectively (Green Report, 2018).

Traditionally, Kosovo supports three main agricultural sectors - cereals, horticulture and livestock - broken down into 21 sub-sectors – 11 crops (annual and permanent, e.g. existing orchards and vineyards), horticulture (open field vegetables), wine, and organic products, and 10 livestock sectors and milk. One of the most recently introduced direct payments (in 2016) is for organic products, though in 2017 – the last year for which data is included in the Green Report – only 10 farmers applied and 7 were approved and received organic direct payments. Overall, approximately 41,000 farmers/businesses apply ever year, of an estimated 130,000 that are registered for farming and agriculture production. Additionally, there are many producers who are not registered, and hence cannot benefit from any DP programme.

For each sub-sector there are transparent eligibility criteria. The important point for the analysis in this paper is that they are size related. For example, for cereals and oilseeds, the eligibility criterion is to have 1 ha under planting, only for winter wheat the requirement is for is 2 ha. For open field vegetables, existing orchards, vineyards and organic production the area eligible for support is only 0.5ha, with even lower thresholds of 0.2 ha and 0.1 ha for small fruit and vineyards respectively. In the livestock sector the requirements are to have minimum 5 dairy cows or water buffalos, 30 sheep and 20 goats both for milk production. Apart from quantitative thresholds for milk (e.g. 1500 litres delivered over three months) and for registered slaughtered cattle there are also slaughterhouse quality requirements.

Given that land cultivation in Kosovo is extremely fragmented and more than a half of farmers have less than 1 ha cultivated land, more than 50% of farms are not eligible for cereals and oilseed direct payments. Since acreage thresholds are crop-specific, while most farms are mixed, the proportion of farms excluded from the DP programme due area thresholds is probably much higher. In the livestock sector, according to agricultural census, the average number of cattle per agricultural holding having cattle was 3.9 head and 56% of the holdings only had 1-2 heads (KAS, 2015). This also suggests that the typical ‘average’ livestock farm is not eligible for direct payments.[3]

~~Apart from the quantitative thresholds, another factor which is important for endogeneity identification is the application procedure. A standardised application process for direct payments has been implemented. Farmers apply in person in their local municipality. The applicant is supported and the application is accepted by the local advisor (who normally sits at the municipal centre). The advisor performs a technical review of all documents that are part of the application, digitalises and sends them to the Agency for Agricultural Development. This means that the direct payments application is accessible to farmers and the size of the applicant farm is not a defining factor in the application process. The application is free of charge and the only the cost for the applicant, apart from transport, is for the issue and collection of the~~

[3] Although the application process is relatively simple and uncomplicated, and accessible to all famers, regardless of size, it is very likely that many small farms do not apply, perhaps realising that they will not be eligible for payment in any case.

~~documentation required by the Agency. Most of these documents are available at the Municipality level and issued within 5 working days. In summary, these are identification documents, bank account evidence, which all applicant will possess, evidence for cultivated land (owned or rented) and financial obligations. The whole and there is no apparent discrimination of small farmers in relation to big producing companies, the transport (visiting the municipal centre twice) and other transaction costs may discourage some small family farmers of applying.~~

Methodology

Instrumental variables methods to deal with the endogeneity issues with our data are widely used and well-known. We focus here on less well-known methods to deal with this problem.

The first applies a copula correction (Park and Gupta, 2012) to overcome the endogeneity issue, without resort to instrumental variables. In simple terms, copula correction can be viewed as a control function, which uses Gaussian copulas instead of control variables to correct for endogeneity. The validity of this approach rests on the crucial assumption that the endogenous variables are not normally distributed, and in the case of continuous endogenous variables (as in this paper) it is preferable that they follow some type of skewed distribution. However, non-normality of the continuous endogenous variables is a necessary, but not sufficient condition for applicability of the method. The underlying idea is that the marginal distribution for the error term (which is given by the statistical model estimated, typically a conditional Gaussian) can be complemented by assuming marginal distribution(s) for the endogenous variable(s). Then one can use a copula specification to specify a flexible multivariate joint distribution of the error term and the endogenous variables, given the covariates. This specification allows for a very wide range of possible correlations between the marginals. In practice, instead of assuming a particular marginal distribution for the endogenous variables, one can simply estimate their empirical distribution function using a standard kernel density estimate (in this case, an Epanechnikov kernel with a Silverman's rule of thumb choice of bandwidth). The joint multivariate distribution contains additional terms which are the correlations between the endogenous regressors and the error term, and which are used to correct for the effects of endogeneity on estimation. In the case of a single endogenous variable, the model can be estimated directly by maximum likelihood. When there are several endogenous regressors, additional regressors constructed as the inverse of the marginal distribution of the endogenous variables are included in the model. These additional variables act as control functions and provide a correction derived from the correlations between the error term and the endogenous variables.

In order to create the corresponding control function auxiliary variables, we use the assumed marginal distribution to invert the empirical cumulative distribution function distribution (ECFF) of the endogenous variables. If the ECDF matches the assumed marginal distribution, the above procedure will simply reproduce the endogenous variable. Hence we need the assumed marginal distribution to differ significantly from the empirical distribution of the endogenous variables. Assuming normal marginals (as we do in this paper) is not strictly necessary (one can use any other distribution) but it makes it easier to test the empirical distribution for deviations from Gaussianity. In this respect, having a less flexible distributional assumption about the margins helps, since empirical data typically deviate substantially from a normal distribution. Hence the quality of the constructed auxiliary variables depends on the difference between the empirical and marginal distributions. This leads to the most common difficulty with this method, requiring a certain amount of luck to ensure a sufficient difference between the two distributions. When the difference is not large enough, the outcome generates multicollinearity issues with substantial correlation between the endogenous and auxiliary variables. Multicolinearity, in this case, is an efficiency problem, in that it inflates standard errors and confidence intervals thus preventing efficient inference. Indeed with wider confidence intervals one may not be able to e.g. ascertain whether the method works (since this is indicated by the significance level of the auxiliary variables and larger standard errors may lead one to wrongly assume they are insignificant). There are however standard treatments for multicolinearity, such as ridge regression, principal components regression and partial least squares regression which can be applied to remedy the problem. For brevity we will not discuss these here and will instead just note the application details in the empirical analysis section.

Lewbel (2012) proposed to identify edgogenous models using variables that are uncorrelated with the product of heteroskedastic errors. The instruments are constructed as a simple subset of the model data. The method can be applied when no external instruments are available, or alternatively, in addition to such external instruments (e.g. in order to improve the efficiency of the IV estimator). In simple terms, the Lewbel's (2012) method uses instruments $[Z\text{-}E(Z)]v$, where v is the error term and Z is some subset of the exogenous regressors, present in the model. Technically, although these do not have the same interpretation as conventional instruments, they are used in exactly the same way. Identification is achieved if and only if the above 'instruments' are correlated with the error term, and the degree of this correlation (more precisely the extent of the covariance between the above two) provides a measure of the strength of the instruments. The latter is an assumption that can be empirically tested, and since, in practice, the estimation process proceeds in the same way as conventional instrumental variables estimation, all the usual checks and tests are also applicable. Hence the advantage of this method is that standard instrumental variables diagnostics can be applied. The disadvantage is that of all instrumental variables methods in that valid instrumentation is required

Lewbel (1997) proved that a subset of exogenous variables present in the endogenous regression model can be used to construct a much wider set of potential instruments. Unlike the method of Lewbel (2012), these instruments require that (i.e. are only valid if) the endogenous variable has a skewed distribution. Otherwise, the same approach and justification as above apply. More specifically these instruments are:

$G(Z)\text{-}E(G(Z))$
$[G(Z)\text{-}E(G(Z))]\ [W\text{-}E(W)]$
$[G(Z)\text{-}E(G(Z))]\ [Y\text{-}E(Y)]$
$[Y\text{-}E(Y)]\ [W\text{-}E(W)]$
$[W\text{-}E(W)]^2$
$[Y\text{-}E(Y)]^2$

where Y is the dependent variable, W is the endogenous variable(s) and $G(.)$ is any nonlinear function that has finite third moments. In practice, by replacing the expectation operator with a sample mean (which is its sample equivalent) one can obtain a wide range of instruments.

Although relying on a more restrictive assumption about the endogenous variables (in terms of skeweness) this method gives an opportunity to construct and test a large number of non-linear instruments and thus can alleviate the issues related to the search of identification. The issue is that, due to the large number of potential 'instruments', specification search may be complicated. There are some methods that can restrict a large number of potential instruments via shrinkage (e..g lasso) but we will ignore these in this paper in favour of a more ad-hoc search, as explained below.

We are interested in the differential impact of different types of direct payments on market participation. However, the wide range of DPs employed in Kosovo (as noted above) means that we need to group these different payments into categories which are reasonably internally homogeneous, otherwise the data are insufficient to overcome the endogeneity issues. We group direct payments into the following categories: fruits and vegetables; cereals and oilseeds; livestock payments. This categorisation excludes some relatively minor direct payments such as organic production payments, beekeeping etc., but is able to capture the bulk of the direct payments allocated to Kosovo agriculture. Furthermore, it fulfils the homogeneity requirement.

Identification considerations

The conventional approach to identification under endogeneity relies upon instrumental variables. Instruments need to fulfil several conditions, especially to be correlated with the endogenous variables, but uncorrelated with the residual term. In order to satisfy the latter exclusion requirement one usually specifies instruments that are strictly exogenous (so that they cannot be influenced by either the dependent or the endogenous variable). There are however some additional requirements to achieve valid identification. The instruments need to be sufficiently correlated with the endogenous variables, since in the case of weak instruments (which are only weakly correlated with the endogenous variables) the resulting estimates are typically biased. Technically, there are statistical tests to detect this particular problem, if present, in the statistical model. The other requirement is not directly testable and refers to the need to identify all sources of variation in the endogenous variables which are common with the dependent variable. If this is not achieved, the endogeneity problem persists. Since considering and identifying all possible sources and linking each to an instrument is non-trivial, the prevailing practice is to employ a larger set of instruments in order to avoid under-identification. In our particular case, this is not a viable strategy since we have a limited set of potential instruments. In particular, we rely on geographical distances. Since farms do not move, geographical distances are clearly exogenous with regard to both market integration and subsidy support, which is a prerequisite for valid instrumentation. Our survey questionnaire recorded distances from the farm to relevant infrastructure and services. Given our data, the following distance variables were identified as potentially viable instruments[4], all of which need to be tested empirically:

to municipal centre/ public farm advisory service (km);

to a shop for agricultural inputs (km);

[4] Some relevant variables including distances related to marketing outlets contained a large number of missing observations.

to a farmers market (km);

to a bank (km);

to a public transport stop (km).

The distance to the administrative municipal centre and advisory services proxies for distance to the main retail market, directly affecting market participation, and might also influence the application process for DPs. Since the use of advisory services is one of the explanatory variables, this allows the type of identification to be examined.

We lack panel data, which further complicates the estimation process, including additional identification considerations, especially for livestock, where production and commercialisation effects take longer than one year. The most popular method of trading is the livestock market, which accounts for almost 80% of the overall volume of trade. This facilitates identification since due to the spatial fixity of such livestock markets nonlinear transformations of distances could be expected to correlate with the market participation and promise valid empirical identification, notwithstanding the problems posed by the sparse nature of the direct payments data for livestock. The increased share of farm-gate sales is another factor that may alleviate identification, since farm gate sales are the marketing channel that is the least restrictive in terms of market participation.[5] Veterinary requirements apply to all farmers and all livestock marketing channels, and may prevent small semi-subsistence farmers from accessing the livestock market. While it is difficult to find or construct appropriate instruments that are correlated with the level of compliance, education level and/or experience might be associated with compliance capacity. Personal competencies of farmers can also contribute to successful accomplishment of the application process.[6] The use of advisory services may potentially mitigate difficulties associated with the application process, but education can still be expected to be associated with direct payments receipts. Therefore, we identify distance to municipal centre and respondent's years in education as variables that can possibly identify the endogenous direct payments. They are part of the model specification and therefore can only be included in non-linear manner following Lewbel (1997, 2002).

We expect that access to farmers' markets will be related, especially, to fruit and vegetable sales and DPs, since these are typically sold at farmers markets and marketing potential may encourage farmers to apply for and consequently receive direct payments. It is however more difficult to see a direct link between farmers markets and cereals and oilseeds, for which wholesale markets would be more appropriate. Unfortunately too many missing values prevent efficient use of wholesale market distances.

Access to agricultural inputs (proxied by the distance to an input shop), will facilitate production and due to the expected link between output, via marketable surplus, and market participation may correlate with the latter. In term of direct payments, access to any type of

---

5 In recent years there are more and more traders that go to the farms and buy livestock directly from the farmer. This process is under strict legal control and the farmer is required to follow all administrative and veterinary procedures before engaging in this kind of sale.

6 For example, out of 41,000 thousand applications for direct payments received, around 3,000 were rejected, either due to the lack of proper documentation or due to incorrect information provided to the Agency for Agricultural Development.

input market infrastructure may be related to the process of applying for and obtaining direct payments.

Distance to a bank, e.g. a shorter distance may indicate more frequent relations with the bank manager and trust between the two parties, and as a consequence a better opportunity to receive a loan that will boost output, or similarly a shorter distance to a public transport stop facilitates travel to input and output markets.

In summary, we identify a small set of variables as potential instruments, but as the discussion above illustrates, the link between them and direct payments is quite indirect. As a result, these instruments are likely to be weak. For this reason, the possibility for heteroscedasticity based non-linear instruments appears potentially useful. Finally, the identification consideration discussion demonstrates that identification appears much more straightforward for fruits and vegetables, while livestock direct payments present a major challenge.[7]

Data

We use a unique dataset collected for an FAO-sponsored project focused on commercialisation of small and family farms in Kosovo.[8] An international team of experts from Kosovo, Germany and the UK contributed to the survey design and data collection. The survey instrument was designed specifically to investigate the opportunities and problems for commercialisation of small and family farmers. Data collection was carried out by face-to-face interviews in the second half of 2018. As a basis for sampling, the nationally representative Farm Accountancy Data Framework (FADN) was used. Since FADN by definition includes commercial farms, the smallest farms from the Kosovo FADN were selected for an interview. Each interviewed FADN farmer was also asked to point out to two smallholders in the vicinity, snowballing our sample. The final useable dataset comprised 680 farms – 52% derived from FADN and 48% additionally sampled by the snowball approach. Spatially, all seven Kosovo regions were covered.

Variables used in the study are summarised in Table 1. The bold typeface is used to distinguish the variables employed in the model, while the other variables are only used as instruments. Brief descriptions of the variables are presented in the second column, since only the abbreviations are used in the analytical results that follow.

The dependent variable is the share of output sold, as stated by the respondents. Although this is easier to obtain, it presents some challenges. Such a measure is aggregate (covering the whole farm product mix),[9] and is potentially prone to measurement errors. While the aggregate dilutes the potential effects of direct payments (over a range of products), and reduces the

---

7 Personal household characteristics (age, education etc.) were mentioned as potential instruments, but our preliminary investigation indicated that they are ineffective. Therefore, we will not comment further on the household characteristics as potential instruments.

8 FAO TCP/KOS/3602.

9 Estimating the effect of DPs which are directly related to specific products would in principle be better evaluated if a product specific level of commercialisation is available as e.g. in Kostov and Davidova (2013). Such calculations are however always problematic, particularly for livestock where a large number of assumptions need to be made.

possible inferences about these effects, the aggregate measure of market participation corresponds more closely to our conceptual model. Measurement errors, however, may be an additional source of endogeneity. While in most other settings this could be considered an undesirable complication, in the present context it is actually a blessing in disguise. Our methodological approach is explicitly focused at identifying the sources of endogeneity and hence the use of imperfect measure of market participation fits in a more general pattern of simultaneous determination. Therefore, in spite of presenting some methodological difficulties, the use of self-declared market participation is consistent with our methodological approach.

The cereals and oilseeds DP include payments for wheat, wheat seeds, barley, rye, maize and sunflower. The fruit and vegetables category includes payments relating to open field vegetables, organic agriculture, raspberries, blueberries, walnuts, other nuts, vineyards. The livestock DP category aggregates milk, dairy cows, sheep, goat and sow payments. All direct payments are measured in euros.

Insert Table 1 around here

We present means, minimum and maximum values and standard deviations which are typical summary statistics. In addition to this, we also include the number of full cases (i.e. excluding missing values) as well as the number of non-zero values for each variable. We have pre-screened the data and have excluded variables with large number of missing values. As Table 1 shows, the data used for estimation contain a limited number of missing values (typically due to answers such as 'do not know') and therefore provides a sufficient number of valid observations to use for estimation purposes. The table additionally shows the number of non-zero values for each variable. As opposed to missing values, all zeros in our data provide meaningful information. The issue is that having sparse variables (i.e. a larger number of zeros) creates challenges to both estimation and identification. Only a small portion of the data sample consists of farms that receive direct payments. Cereals and oilseeds direct payments are prevalent, with 89 farms (i.e. 13%) receiving such payments. Fruits and vegetables, and livestock payments only cover 28 (4%) and 17 (2.5%) of the sample farms, respectively. This adds an additional layer of complexity since the sparse nature of direct payments (even after we have aggregated them in several groups) presents serious challenges in properly identifying the underlying sources of variation (in that the variation only flows from the non-zero observations while the variation in most identifying variables will flow from all observations) . As an immediate consequence of the sparsity of the direct payments one could expect that many potential instruments will be only weakly correlated with them, which presents serious challenges for model identification. This problem is however slightly alleviated by the nature of the sample, which uses smaller farms. In fact 65% of the sample farms are subsistence (zero sales share). Since the farms in receipt of direct payment are at least partially market integrated, the zero market participation largely overlaps with the non-receipt of direct payments and, thus, facilitates the estimation of their effect. The latter does not, of course, solve for the possible weak correlation between direct payments and potential instruments, which will need to be dealt with separately.

Insert Table 2 around here

~~**Insert Figures 1-3 around here**~~

Since both the copula approach and that of Lewbel (1997) require that the endogenous explanatory variables have non-Gaussian fat-tailed distribution, we test for this. Table 2 presents the skewness and excess kurtosis for all three types of direct payments, together with the corresponding t-tests. It demonstrates that these deviate from Gaussian distribution and are characterised by fat tails. This can be further illustrated by the comparative plots of the empirical density of these variables comparative to a normal density (Figures 1-3, on-line). Therefore the two estimation methods – copula approach and Lewbel (1997) - are applicable.

Results

In order to identify and estimate the potential effects of DPs on market participation we follow a systematic approach outlined below. First, purely for illustrative purposes, in Appendix 1, on-line, we present the results from a naïve estimation that ignores the problem of endogeneity. As expected, the DPs are correlated with the level of market participation. Since such a 'model' is clearly mis-specified, we will not pay any more attention to it and proceed with our identification strategy.

The first logical step in dealing with endogeneity is to use conventional instrumental variables estimation. Once again, purely for illustrative purposes, in Appendix 2, on-line, we present the results of following this approach. The same 4 instruments were used for the three endogenous variables. Such an approach can be considered naïve since it applies a blind search for identification, using a set of more instruments than the number of endogenous variable and avoids any discussion of why and how the instruments might identify the endogenous variables. Following from the previous discussion on the possible identification role for these instruments, it is clear that they may be able to identify fruit and vegetables direct payments, but they may be weak (if valid at all) for the other types of direct payments we consider. The purpose of including 'conventional' instrumental variables approach is two-fold: to illustrate the type of results one may obtain and to show that such an approach cannot work in this case. Furthermore, since the list of potential instruments that we have identified is very short, it is better to use them all at the same time.

The results in Appendix 2, on-line, confirm that this set of instruments is weak for all three types of direct payments and are not valid in general (see Wu-Hausman test). Note, however, that if we were only relying on the Hansen's J- test for over-identifying restrictions, we might have concluded that the instruments are reliable. The above results demonstrate that the instruments we consider cannot provide full identification due, on the one hand, to their weakness and, on the other, the failure to identify all sources of variation in the endogenous variables. The highly insignificant Hansen's J test, however, suggests that they may carry some (limited) explanatory power, which suggests that, in conjunction with a different identification strategy, they may help provide a valid identification. Since the Hansen/Sargan test in practice does not test instruments validity, but rather their coherence (Parente and Santos Silva, 2012), it provides some justification for complementing these instruments with a different identification strategy.

Applying a copula based correction for endogeneity is another option to to identify the model in the absence of suitable instruments. The advantage of such an approach, if it works empirically, is that it circumvents the need to search for instruments. This method requires heteroscedasticity in the endogenous variables and, more specifically, fat tailed distributions. As shown above, all three endogenous variables clearly satisfy this requirement. Appendix 3, on-line, presents the results from copula correction. Although the basic assumptions of this approach are met, the correction terms are insignificant, which means that this approach fails to fully identify the model since it does not produce a significant modification to the estimation procedure. However, multicollinearity issues, detected and tested in the copula model, may be responsible for these results.. In order to correct for multicolinearity, we apply a ridge regression estimation for the copula model. To do so, we need to chose an optimal amount of shrinkage (the penalty parameter for the ridge penalty). There are many criteria for selecting the penalty parameter. In essence we have used random splits of the data into two equal subsamples and used the second part to test the predictive abilities for several different choices of the penalty. The details on these different criteria and the other implementation details are provided in Appendix 6, on-line. Additionally, alongside the different ridge penalty criteria we have included a partial least squares estimation, with a visual method of selection for the optimal dimension (which was then taken as given into the data splitting process above). Unsurprisingly (due to the somewhat *adhoc* nature of parameter selection) the partial least squares model was dominated by the ridge version in term of predictive ability. We have not included in the above investigation the principal components regression method for two main reasons. First, it is often found to be outperformed by partial least squares, and second, it is difficult to construct reliable confidence intervals for the estimated parameters.

Insert Table 3 around here

The ridge model selected by the above procedure is presented in Table 3. The first thing to note is the significance of the auxiliary (control function) variables (the variables labelled as Pstar.xxx in the table). Unlike Appendix 2 (on-line), these are now significant with exception of the one referring to the livestock subsidies. Therefore, we can identify the fruits and vegetables and the cereals and oilseeds subsidies as endogenous, and correct for the effects of their endogeneity, and thus obtain meaningful regression estimates. Most of the background variables are statistically significant (except use of fertiliser and household only labour) and have the expected signs. These results demonstrate that it should be possible to identify the model, though the livestock subsidies could be problematic.

We use Lewbel's (2012) approach to create heteroscedasticity adjusted non-linear instruments, based on exogenous variables. In this case, the set of conventional instruments, specified in Appendix 2 (on-line), are complemented by non-linear instruments based on Lewbel (2012) constructed from the distance to the municipal centre and the number of years of the respondent's education. The results are presented in Appendix 4, on-line. The Hansen J-test is acceptable, but the instruments are weak except for the fruit and vegetable payments. Notably, the effect of these payments is significantly positive. These results are promising since they demonstrate that we can potentially identify at least one of the three types of direct payments. Therefore, using a different transformation to define non-linear instruments may help achieve identification for the other two types of direct payments.

To do this, we estimated partial versions of the model including only one direct payment, to search through different types of non-linear instruments to find the ones that can help identify a particular endogenous variable. The instruments found in these separate partial versions of the model are then combined in order to achieve full identification. Since the preliminary steps define instruments in partial versions of the model in which there are omitted variables, i.e. the other two types of direct payments, there is no guarantee that simply adding together all such instruments will achieve identification. It may be necessary to tweak the list of instruments, but the partial versions at least give some indication of how each of the three endogenous variable may be identified. The results from partial model estimation are presented in Appendix 5, on-line. The important outcome is that we were actually able to identify separately each of the three types of direct payments. All three types of DP appear to have a significant positive effect on market participation. However these results do not persist in the full model, when all the missing variables in the partial model results are included.

Insert Table 4 around here

Table 4 lists the instruments identified for each of the endogenous variables. The next step is to combine these in order to estimate a full model. There are two important considerations in this procedure. The sources of variation should be considered carefully, since the same instrument should not identify more than one endogenous variable. The second consideration is that combining the instruments should help identify all possible sources of variation in the endogenous variables. The focus is only on the Lewbel (1997, 2002) type of instruments, since, as demonstrated earlier, the conventional instruments are weak. The fruit and vegetables direct payments are easier to identify since the Lewbel (2002) approach provides identification. The non-linear instruments for cereals and oilseeds direct payments are quite different and since there is no overlap with those for fruit and vegetables they identify different sources of variation. Three types of nonlinear instruments are used for livestock direct payments, of which one is based on the respondents' years of education, This is similar to the one used in the identification of cereals and oilseeds, but is not the same, since it is the product of the common element and the endogenous variable's deviations from the mean. Furthermore, all the other instruments for livestock, and cereals and oilseeds direct payments are different, and hence do not identify the same type of variation. Therefore, the list of instruments used to separately identify the three types of direct payments can be combined in a full model.

The nonlinear instrumentation results are presented in Table 5. Both the weak instruments and the Wu-Hausman tests are highly significant indicating that the set of selected instruments are not weak and are valid. The Hansen's J- test is insignificant hence confirming the null of valid instrumentation.

Insert Table 5 around here

We present two alternative estimators. The first is a 'standard' instrumental variables regression that uses all instruments from Table 4 employing the 'fitted values' approach. The second is a GMM estimator that only uses the non-linear instruments. The instrument tests in Table 5 refer to the standard IV regression and, since the Hansen's J statistic is only marginally insignificant, this raises some questions about the coherence of the instruments set. This can be improved omitting the conventional instruments (and only keeping the non-linear ones). The results presented in table 5 refer to the continuously updated version of the GMM estimator. We have computed a range of alternative moments based estimators (basically different versions of empirical likelihood family of estimators, and their results are similar to the GMM results in Table 5). It is informative, at this stage, to compare these results to those from the copula approach (in Table 3). In general terms, although we have finally managed to identify the livestock subsidies, their effect is not statistically significant. However qualitatively the results in table 5 and 3 are similar in terms of which variables are significant.

Finally, we need to account for the nature of the response variable, which is essentially a fractional response. Furthermore, it contains a large amount of subsistence farms (with zero market participation) - 235 of 675 full observations. There is also a small number of fully commercial farms, but since there are only 8 such farms in the sample we may ignore this issue. Conceptually, the right approach to the problem would have been to start with a proper fractional response model and to apply the specification search directly to it. However this assumes that we have an established specification for such models, which unfortunately is not the case and starting with a preset specification can affect the results.

We deal with these issues in turn. A good overview of the issues and standard approaches to dealing with fractional response data is presented in Papke and Wooldridge (1996). In simple terms, one either transforms the dependent variable to make it continuous, or as is more widely used, one applies a non-linear transformation to a linear model, so that the outcomes fit the (0,1) interval. The latter approach, as advocated by Papke and Wooldridge (1996), is equivalent to the logistic regression approach to success rates (instead of binary data), usually combined with heteroscedacticity adjusted standard errors. There are however a number of alternatives to the logit transformation (probit, loglog, cauchit to name a few). Alternatively a different distributional assumption (e.g. negative or quasi binomial) may be applied. One of the most popular and flexible alternative distribution is the beta probability model, which is extremely flexible in that it can represent a huge range of probability densities, depending on its parameters.

The other issue is the large number of zeros. This can be dealt with two distinct approaches. The first is the 'hurdle' or 'adjustment' approach, which has a long tradition in econometrics, including a preliminary 'selection' model in the first (hurdle) step. The tobit model and the Heckman sample selection method are examples of this approach. The other approach deals with the extra zeros by adopting an 'inflation' model in which the error distribution represents a mixture of a standard (fractional response type of) distribution and a binary zeros process (additional probability of having zeros). Although the two approaches appear similar, they are based on a very different underlying logic. When the zeros are based on censoring (i.e. when data are truncated, but can be in principle observed beyond that point) the hurdle approach is natural. This is however not the case when, as here, observations beyond that point are impossible in principle. In this case, the issue of which approach is preferable depends on the nature of the model. If, for example, the same mechanism defines the model and the selection, then the hurdle approach is not applicable. We argue that this is the case here. In particular, the hurdle approach requires additional variables in the selection mechanism that do not appear in the main model (see Ramalho, et al, 2011 for more details) which is difficult to justify here. Furthermore, in the presence of heterogeneity, hurdle models require additional very strong assumptions to achieve structural identification. For this reason we will only consider zero-inflated model specifications.

An additional complication is that we use a set of predefined sets of instruments derived from a linear specification (subject to the specification search we have outlined before). But, since fractional response models are essentially non-linear, there is no guarantee that these instruments will be valid. We tried a number of different specifications in which either the instruments were invalid or estimation ran into problems. In Table 6 we present a number of fractional response models for which the previously derived instruments were valid.

Insert Table 6 around here

First, the "Fractional GMMz" model refers to a 'standard' fractional response model (as suggested in Papke and Wooldridge, 1996) specified using the logit transform and estimated using the GMMz estimator of Ramalho and Ramalho (2017). The response variable has been standardised to the unit interval (by dividing it by100). Furthermore, we have adjusted the ones by replacing the 8 unit observation by $1\text{-}10^{-8}$. This model accounts for the fractional nature of the data and allows for heterogeneity. However it does not take into account the prevalence of zeros in the response. The second model is a Zero-Inflated Beta regression model, in which the endogeneity is accounted by the control function approach. For brevity we omit reports of the control function variables. Note that, except the adjustment for ones, this is a more realistic model, and also, given the flexibility of the beta distribution is the most flexible. Finally, for illustrative purposes we also report the results from Zero-Inflated Poisson model, once again estimated via the control function approach. Unlike the other two models, the Poisson uses the original response variable (as in all other models) measured over the [0,100] interval. In fact the livestock subsidy is actually significant in the ZI Poisson model, but we have put in the P value from the corresponding control function variable to denote that since it is not identified, it should be considered insignificant.

A quick comparison of the fractional response models in Table 6 demonstrates that their results are in general comparable. We focus on the ZI beta regression results, which are broadly speaking comparable with the previous (continuous type of response) models.

Finally, since the fractional response variable model can suffer from a considerable degree of over dispersion, we make some further adjustments in order to take this into account. In fact, some of the models in Table 6 already include specification and estimation characteristics that reduce the potentially negative effects of over-dispersion. For example the Fractional GMMz model allows for heteroscedasticity (that can partially follow from over-dispersion), the zero-inflation f feature (in the other two models) typically reduces the amount of over-dispersion, and the beta regression model implicitly allows for over-dispersion (see e.g. the technical discussion in Appendix 6, on-line). We include explanatory variables in both the variance and the zero inflation equations of the zero inflated beta regression model (see Canterle and Bayer, 2019). Furthermore we only keep explanatory variables in the latter two equations if they are significant at least at 10% significance levels.

The model specification broadly follows Bayer and Cribari-Neto, 2017, who suggested estimating the mean specification, assuming a constant variance, and then estimate the variance equation. In this paper we fix (instead of selecting) the mean equation specification and then sequentially estimate/select the other two equations. The results from this estimation are presented in Appendix 6, on-line. In this set of results the variance and the zero inflation links are modelled via a logarithmic link and we have applied a number of different links for the mean equation (logit, probit, log, cloglog). These results (from by far the most realistic model specification so far) show that all DPs are identified and have a significant positive effect on market participation. Indeed if we were to try to show a set of results that conform to expectations and to the conceptual framework, the results in Appendix 6, on-line, could represent our best results. However, in spite of holding over a range of specifications, these results are not as robust as they may appear. A range of alternative estimators (e.g. Bayesian equivalents of the estimators in Appendix 6, on-line, or alternative model selection methods, such as component-wise boosting or bootstrap-based stability selection) and specifications (for example a cauchit link for the mean equation) removes the significance of the livestock DPs. We will not discuss these in much more detail here, but interested readers may refer to Appendix 7, on-line, for more details.

Insert Table 7 around here

The final model results (which are robust to alternative specifications and estimation methods) are presented in Table 7. This is a zeros and ones inflated beta regression model (Ospina and Ferrari, 2012) with explanatory variables in both the variance (Canterle and Bayer, 2019) and the two inflation equations. The final model specification is an extended form of the models presented in Appendix 6, on-line, tested by using the bootstrap corrected version of the LR test of Pereira and Cribari-Neto (2014). For simplicity and comparability with the other estimated models, Table 7 only presents the estimated parameters in the mean equation. Similarly to the previous results we find no evidence for the effect of livestock DPs.

In general, risk attitudes (i.e. willingness to take risks), use of inputs, and use of advisory services all increase market participation. A negative effect on market participation is found only for the use of household labour (which implies a positive effect for farms with hired labour). Relying only on household labour reflects limited production capacity and hence farms that use hired labour are better placed to use subsidies to expand output and marketable surplus, and thus increase market participation. Note that most of the other specifications have more significant variables than the final one. However, all significant variables in Table 7 are consistently significant across different modelling specifications.

Concerning the main point of interest, fruit and vegetables direct payments have a significant positive effect on commercialisation. This is to be expected given the nature of Kosovo agriculture, where fruit and vegetables are important cash products, and the fact that compared to the other direct payments, econometric identification of the fruit and vegetables payments was relatively straightforward. Furthermore, even in the partial identification models, these payments have consistently shown to be increasing market participation.

An important result for Kosovo policy is that we have also found evidence for commercialisation enhancing effect of cereals and oilseeds direct payments. Cereal and oilseeds markets are better integrated and larger in size than those for fruit and vegetables, which reduces the potential negative welfare effects from increased market participation.

Finally, our estimates suggest a statistically insignificant effect of livestock payments. This result should be treated with caution. The production cycle for livestock is longer than the one year cycle typical for crops. Therefore, any effects are likely to be spread over a longer period of time, and hence difficult to identify in cross-sectional data. This issue deserves further investigation. We do find a statistically significant effect in the partial identification model for livestock payments. Since this specification ignores the effects of the other types of direct payments, this would only imply a significant effect if all livestock farms were highly specialised, when omitting the effect of other payments would not have a serious impact on the quality of the results. However, since most of farms in Kosovo are mixed, this is clearly not the case.

Conclusions

We analyse the effect of direct support payments on agricultural commercialisation in Kosovo. To the best of our knowledge, this is the first attempt to analyse the effect of specific agricultural support policies on market participation. In addition, we employ a range of specifications and estimation techniques for the reliable identification of the market participation effects of the range of direct payments employed in Kosovo, many of which are currently rather novel to the agricultural economics literature.

Our results show that direct payments associated with fruit and vegetables, in Kosovo, are by far the easiest to identify and have a clear positive effect on commercialisation. Our conceptual framework suggests that the effect on commercialisation will be larger when a larger number of semi-subsistence farms receive payments. Eligibility criteria for fruits and vegetables (open field) include lower size thresholds than most other direct payments and as such are more accessible to semi-subsistence farmers in Kosovo. Furthermore, fruit and vegetable production is more labour and less land intensive, and semi-subsistence farms are typically more labour intensive than their commercial counterparts but usually work on a small land area. On the other hand, the more perishable nature of fruit and vegetables and the less integrated nature of their markets can mean that gains from direct payments are offset by larger losses resulting from price drops, following increased supply and high demand elasticities. If so, fruit and vegetables direct payments act effectively as a consumer subsidy, rather than maintain farmers' incomes. These offsetting effects could be greatly reduced with larger markets for fruits and vegetables, e.g. imposing stricter standards to facilitate exports, as is being done in Kosovo, and additional requirements for the receipt of direct payments. Enlarging the market size and preserving the welfare gains for farmers should result in a larger effect on commercialisation.

We also find support for a positive effect of cereal and oilseeds direct payments on market participation. Since the markets for cereals and oilseeds are better integrated, one could expect less 'spillage' from the direct payments and, thus, that their commercialisation effect would be larger than in the case of fruit and vegetables. However, grains and oilseeds production systems are considerably more land intensive and labour extensive, and thus relatively less semi-subsistence and more likely to participate in markets in any event.

~~The comparative effects of fruits and vegetables vs cereals and oilseeds DPs on commercialisation are another point of interest. In spite of the greater danger for negative externalities caused by fruits and vegetables DPs, we find that they have larger effect on market participation. This result is probably due to the land fragmentation and prevalence of small scale semi-subsistence farms engaged in fruit and vegetable production which means that these DPs affect directly market participation. In contrast to this although cereals and oilseeds production may in principle face a less elastic demand which reduces the negative externalities of such DPs, it is already better integrated in market mechanisms and this is why we find that~~

~~cereal and oilseed DPs have relatively smaller effect on market participation compared to fruits and vegetables subsidies.~~

Finally, we fail to find any significant effect of livestock direct payments on market participation. This may reflect the longer livestock production cycle and the greater heterogeneity of the direct payments in this category, which hinders the empirical identification of their effects. More disaggregated analysis for livestock payments might identify different policy impacts on different types of production.

~~The main implication of the study is that most DPs affect positively market participation and hence they can be used as a tool to facilitate commercialisation of Kosovo agriculture. In doing so, fruit and vegetable DPs appear to have the greatest impact, which is consistent with the present structure of Kosovo farms. This means that with regard to Kosovo the aims to provide income support (which should be aimed at the prevalent farming structures which are dominating fruit and vegetable production) and to increase market participation are well aligned and can therefore be supported by the same set of measures.~~

Finally, our conceptual framework is, essentially, comparative static while our data are cross-sectional, and hence cannot provide strong predictions about longer-term market participation effects. Panel data, and more dynamic estimating model specifications are needed to explore the linkages in more detail.

References:

Barrett, C. (2008). Smallholder Market Participation: Concepts and Evidence From Eastern and Southern Africa, *Food Policy*, 33, 299-317.

Bayer, F. M.& F. Cribari-Neto (2017) Model selection criteria in beta regression with varying dispersion, *Communications in Statistics - Simulation and Computation*, 46:1, 729-746.

Brady, M., Hristov, J., Höjgård, S., Jansson, T., Johansson, H., Larsson, S., Nordin, I. Rabinowicz, E. (2017). Impacts of Direct Payments – Lessons for CAP Post-2020 from a Quantitative Analysis. AgriFood Economics Centre, Lund. Available at: https://www.agrifood.se/Files/AgriFood_Rapport_20172.pdf. Accessed 21 June 2019.

Canterle, D.R., and Bayer, F.M. (2019) Variable dispersion beta regressions with parametric link functions. *Statistical Papers* **60,** 1541–1567.

Davidova, S. (2011). Semi-Subsistence Farming: An Elusive Concept Posing Thorny Policy Questions, *Journal of Agricultural Economics* 62 (3): 503-524.

Green Report (2018). Republic of Kosovo, Ministry of Agriculture, Forestry and Rural Development

Kosovo Agency of Statistics (2015). Agriculture Census in the Republic of Kosovo 2014 Final Results. Available at: http://ask.rks-gov.net/media/1375/final-results.pdf. Accessed 24th June 2019.

Kosovo Agency of Statistics (2019) Labour Force Survey in Kosovo, 2018. Available at: http://ask.rks-gov.net/en/kosovo-agency-of-statistics/add-news/labour-force-survey-in-kosovo-2018. Accessed 22 June 2019.

Kostov, P. and S. Davidova (2013) A quantile regression analysis of the effect of farmers' attitudes and perceptions on market participation, *Journal of Agricultural Economics*, 64 (1), 112–132.

Lewbel, A. (1997) Constructing Instruments for Regressions with Measurement Error When No Additional Data are Available, With an Application to Patents and R&D, *Econometrica*, 65(5), 1201–1213.

Lewbel, A. (2012) Using Heteroskedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models, *Journal of Business & Economic Statistics*, 30(1), 67-80.

O'Dononghue, Whitaker, J. (2010) Do Direct Payments Distort Producers' Decisions? An Examination of the Farm Security and Rural Investment Act of 2002 *Applied Economic Perspectives and Policy*, 32 (1), 170–193.

Osmani, F., Gorton, M., White, J. (2013). Agricultural Households, Poverty and the Rural Labour Market in Kosovo. *Post-Communist Economies*, 25(2), 241-252.

Ospina, R., and Ferrari, S.L.P. (2012) A general class of zero-or-one inflated beta regression models, *Computational Statistics & Data Analysis*, 56, (6), 1609-1623.

Papke, L. and J.M. Wooldridge (1996), Econometric methods for fractional response variables with an application to 401(K) plan participation rates, *Journal of Applied Econometrics*, 11(6), 619-32

Parente, P. M. D. C., J. M. C. Santos Silva (2012) A cautionary note on tests of overidentifying restrictions, *Economics Letters*, 115(2), 314-317.

Park, S. and Gupta, S., (2012), Handling Endogenous Regressors by Joint Estimation Using Copulas, *Marketing Science*, 31(4), 567-86.

Pereira, T. L. & F. Cribari-Neto (2014) Detecting Model Misspecification in Inflated Beta Regressions, *Communications in Statistics - Simulation and Computation*, 43:3, 631-656.

Ramalho, E. A. and J. J. S. Ramalho (2017) Moment-based estimation of nonlinear regression models with boundary outcomes and endogeneity, with applications to nonnegative and fractional responses, *Econometric Reviews*, 36(4), 397-420.

Ramalho, E.A., J.J.S. Ramalho and J.M.R. Murteira (2011), Alternative estimating and testing empirical strategies for fractional regression models, *Journal of Economic Surveys*, 25(1), 19-68.

Swinnen, J. (1994). Overview of Policy and Institutional Reform in Central European Agriculture. In J. Swinnen (ed.), *Policy and Institutional Reform in Central European Agriculture*. LICOS Studies on the Transitions in Central and Eastern Europe. Aldershot etc.: Avebury.

Tangermann, S. (1998). Agricultural Implications of the EU Eastern Enlargement and the Future of the CAP, pp 114-140 in Von Witzke, H. and Tangermann, S. (eds) *Economic Transition in Central and East Europe, and the Former Soviet Union: Implications for International Agricultural Trade*. International Agricultural Trade Research Consortium, Symposium Proceedings Issue, June 12-14, 1997, Berlin, Germany Available at: https://tind-customer-agecon.s3.amazonaws.com/ad0f96bc-cb0a-4e08-a028-aed39b843b41?response-content-disposition=inline%3B%20filename%2A%3DUTF-8%27%27TangermannStefan.pdf&response-content-type=application%2Fpdf&AWSAccessKeyId=AKIAXL7W7Q3XHXDVDQYS&Expires=15

61211173&Signature=tsufn0%2Bc%2FCluuT7EGXYacQLgPIc%3D. Accessed 21st June 2019.

Table 1. Data description

Variable	Description	Mean	Min	Max	SD
sales.share	**Share of sold output**	**28.05882**	**0**	**100**	
sub_grain	**Grain and oilseeds DP**	**51.70588**	**0**	**3500**	**2**
sub_fr_veg	**Fruits and vegetables DP**	**18.52353**	**0**	**1000**	**1**
sub_liv	**Livestock DP**	**23.02941**	**0**	**4500**	**2**
V14_1_a_shop_km	Distance to shop for agricultural inputs	6.812589	0.06	30	
V14_1_i_market_km	Distance to farmers market	9.463594	0.5	75	
V14_1_k_munic_km	**Distance to the municipal centre**	**9.60192**	**0.3**	**30**	
V14_1_l_bank_km	Distance to bank	9.278107	0.3	30	
V_14_1_m_publictrans_km	Distance to public transport	4.070074	0.01	20	
risk_att	**Risk attitude**	**4.594118**	**1**	**7**	
V12_01_eduyears_resp	**Respondent's years in formal education**	**10.62887**	**0**	**18**	
V09_1_useany	**Indicator for use of any inputs**	**0.842647**	**0**	**1**	
V09_1_a_use_fert	**Use of fertilisers**	**0.747059**	**0**	**1**	
advisory	**Use of advisory services**	**0.178203**	**0**	**1**	
plots	**Number of separate plots**	**3.305638**	**0**	**22**	
HH_labour_only	**Indicator for use of only household labour**	**0.458824**	**0**	**1**	

Table 2. Skewness and kurtosis tests for direct payment variables

	sample value	T-stat	Pval
	Fruits and Vegetables DP		
skewness	6.75	71.81	0.00
excess kurtosis	49.48	263.38	0.00

	Grains and oilseeds DP		
skewness	9.03	96.15	0.00
excess kurtosis	125.72	669.19	0.00
	Livestock DP		
skewness	15.69	167.03	0.00
excess kurtosis	268.80	1430.79	0.00

Table 3 Ridge estimation of copula corrected model

	Estimate	P value
Intercept	22.097	0.000
V14_1_k_munic_km	-0.043	0.009
risk_att	0.211	0.000
V12_01_eduyears_resp	0.061	0.013
V09_1_useany	0.423	0.049
V09_1_a_use_fert	0.193	0.280
advisory	0.883	0.000
HH_labour_only	-0.203	0.210
sub_fr_veg	0.004	0.000
sub_liv	0.001	0.060
sub_grain	0.001	0.006
Pstar.sub_fr_veg	0.994	0.002
Pstar.sub_liv	0.793	0.138
Pstar.sub_grain	0.680	0.011

Table 4. Instruments used in final model

Livestock DP (W)
[Y-E(Y)] [W-E(W)] ,
[G(Z)-E(G(Z))] [W-E(W)], *G(Z)= X^3, Z=* V12_01_eduyears_resp
[G(Z)-E(G(Z))] [Y-E(Y)], G(Z)= X^2, Z=V14_1_k_munic_km
Grains and oilseeds DP (W)
[Y-E(Y)] [W-E(W)]
[G(Z)-E(G(Z))] [W-E(W)] , G(Z)= X^3, Z= V12_01_eduyears_resp
Fruits and Vegetables DP (W)
[Z-E(Z)]v , v – error term, *Z=* V14_1_k_munic_km and V12_01_eduyears_resp
Standard IVs: (present for all three W)
V14_1_l_bank_km,
V_14_1_m_publictrans_km,
V14_1_a_shop_km ,
V14_1_i_market_km

This is a post-refereed version of the paper Kostov, P., S. Davidova, A. Bailey, E. Gjokaj and K. Halimi (2021) Can direct payments facilitate agricultural commercialisation: Evidence from a transition country, *Journal of Agricultural Economics*, The substantively different definitive version is available at wileyonlinelibrary.com

Table 5 IV and GMM estimators for the main model

	Standard IV regression		GMM estimation	
	Estimate	P value	Estimate	P value
(Intercept)	13.016	0.023	13.043	0.024
V14_1_k_munic_km	-0.475	0.036	-0.469	0.022
risk_att	2.663	0.000	2.324	0.001
V12_01_eduyears_resp	0.296	0.400	0.453	0.002
V09_1_useany	7.261	0.109	6.790	0.034
V09_1_a_use_fert	-5.439	0.156	-5.250	0.157
advisory	9.139	0.002	9.403	0.004
HH_labour_only	-2.203	0.330	-2.299	0.289
sub_fr_veg	0.043	0.000	0.027	0.025
sub_liv	0.009	0.141	-0.002	0.769
sub_grain	0.011	0.050	0.015	0.000

IV Tests		df1	df2	Statistic	p-value
Weak instruments	(sub_fr_veg)	6	661	214.223	0.000
Weak instruments	(sub_grain)	7	662	126.135	0.000
Weak instruments	(sub_liv)	8	663	270.346	0.000
Wu-Hausman		3	664	5.801	0.001
Sargan		3		6.535	0.088

Table 6. Fractional response variable model estimators

	Fractional GMMz		ZI beta regression		ZI Poisson regr	
	Estimate	P value	Estimate	P value	Estimate	P
INTERCEPT	0.397	0.973	-0.6291	0.1467	3.5090	
V14_1_k_munic_km	-0.546	0.000	-0.0150	0.0220	0.0672	
risk_att	3.351	0.000	0.1092	0.0017	0.0038	
V12_01_eduyears_resp	0.329	0.322	0.0085	0.5747	0.1786	
V09_1_useany	-1.997	0.999	0.3896	0.0677	0.2716	
V09_1_a_use_fert	8.801	0.000	0.5366	0.0023	0.2356	
advisory	4.454	0.000	0.3759	0.0029	0.1172	
HH_labour_only	-0.236	0.000	-0.1930	0.0464	-0.0085	
sub_fr_veg	0.005	0.000	0.0008	0.0219	0.0004	
sub_liv	0.010	0.119	-0.0001	0.8118	0.0000	
sub_grain	0.013	0.000	0.0003	0.0427	0.0002	

Table 7 Final results

	Estimate	P value
(Intercept)	-0.7007	0.0079
V14_1_k_munic_km	-0.0087	0.3908
risk_att	0.0863	0.0084
V12_01_eduyears_resp	0.0078	0.5896
V09_1_useany	0.3334	0.0190
V09_1_a_use_fert	-0.3987	0.1110

advisory	0.3510	0.0033
HH_labour_only	-0.2218	0.0196
sub_fr_veg	0.0016	0.0000
sub_liv	0.0000	0.9028
sub_grain	0.0005	0.0000
v_fr_veg	-0.0020	0.0011
v_liv	0.0009	0.1659
v_grain	-0.0007	0.0078