How to Increase Productivity in Kosovo Agriculture: A Story of Size and Technical Efficiency

Philip Kostov
University of Central Lancashire

Sophia Davidova
University of Kent

Ekrem Gjokaj
MAFRD

Kaplan Halimi
MAFRD

Kosovo is struggling with low productivity in agriculture and an overwhelming majority of small farms. This paper analyses changes in the factor mix that can bring the highest increase in marginal productivity, employing a non-parametric quantile regression based on Farm Accountancy Data Network (FADN) data. Two different quantile regressions are estimated, for the median 0.5th quantile, describing the nature of the input-output relationship for a 'typical' farm and for the 0.8th conditional quantile, characterising a reasonably 'efficient' farm. The results show that optimal marginal productivity can be achieved by 'typical' farms by increase in farm size but it requires drastic changes in factors which are currently hardly feasible in Kosovo (e.g. 3-4 FTEs family labour, 0.5 to 1.8 hired). For an efficient farm, the optimal marginal productivity is achieved at lower values of inputs. This suggests that productivity enhancements can be obtained by a careful balance of both efficiency and scale augmentation measures.

Keywords: quantile regression, productivity enhancement, technical efficiency, scale augmentation, Kosovo

INTRODUCTION

In line with the other countries in the Western Balkans, agriculture is an important sector in Kosovo. In spite of contributing only 9% of the GDP it accounts for 25% of the employment hinting for low labour productivity. Kosovo's agriculture is characterised by serious structural problems. These include land fragmentation, low labour efficiency and high production costs (MAFRD, 2018). Several features of Kosovo agriculture are related to the low marginal productivity of production factors. First, the majority of farms are very small in physical size (this is how the statistics in Kosovo reports farm sizes). The Survey of Agricultural Holdings (MAFRD, 2018) accounted for 108,803 agricultural holdings. Around 10% of the agricultural land is cultivated by farms having less than 1 ha, 48% by farms with 1-5 ha, 29% by farms from 5-20 ha and only 12.1% by larger farms. (MAFRD, 2018, FAO, 2019). In fact, 36.3% of the total number of farms use less that 0.5 ha and 17.2% have between 0.5 and 1 ha, so the typical Kosovo farm appears to be smaller than 1 ha. That said, the typical farm in Kosovo is a semi-subsistence one, characterised by a low or no involvement with output and input markets (FAO, 2019). When trying to assess possible productivity advancements in agriculture, it is nevertheless better to try to focus on the next tier of farms (in terms of their size). In particular, 20.7% of farms use 1-2 ha and 18.9% of farms cultivate between 2 and 5 ha. (MAFRD, 2018) These slightly larger farms and the small number of larger ones might have the potential to grow and, therefore, are the focus of the present investigation.

The second characteristic is that although the majority of farms are small and in high need of information and advice on factor mix the public advisory service focuses on larger farms (FAO, 2019). Public advisors are under the control of municipalities and are often diverted to other activities. Collaboration with NGOs and private advisors is sporadic. This decreases the effectiveness of the advisory service to increase the productivity of Kosovo farmers.

Given the structure of Kosovo farms, two specific questions are formulated in the paper concerning the potential to increase the productivity of the farming sector. The first one relates to the constraints that the small size of farms may impose on their productivity and whether larger farms can exploit economies of scale in order to obtain higher productivity. The other issue is whether, given the current low efficiency of Kosovo farms, increases in technical efficiency can lead to large productivity gains. We consider the feasibility of the recommended policy measures by comparing the optimal sizes (in terms of different production input mixes) deduced from our analysis to the ones observed in the present farm structure and draw policy conclusions based on both the potential effect of such enhancements and their feasibility.

DATA

The study uses the standardised FADN type Kosovo dataset for 2016. Since 2015 a sample of 1,250 farms has been covered annually. These farms are defined as 'commercial' and include farms with 2,000 € Standard Output (SO)[1] or above. The sample is based on the 2014 Agricultural Census carried out by the Kosovo Statistical Agency and is representative for the larger farms (FAO, 2019). The variables used in the analysis are output (in EUR), family and paid labour both measured in Agricultural Work Units (AWU)[2], capital (in EUR) and land (in ha), as well as intermediate consumption (in EUR). The output is to be modelled as a function of the inputs, namely labour, capital, land and intermediate consumption. We treat the two types of labour input – family and hired - as different allowing them to have potentially different contribution to the formation of output.

Some summary statistics for the data used are presented in Table 1. Since all variables are highly skewed, the mean values may give somewhat misleading impression. For this reason, the table also presents the mode for each variable, which in addition to demonstrating the fact that most probability mass is situated to the left of the average values and, thus, most farms are smaller than average values show, also indicates the typical values for these variables. Furthermore the variables' values are highly dispersed (see the standard deviations) with a very small number of much larger farms. From these data one can ascertain that indeed the farms included in the FADN dataset are considerably larger than the average farms in Kosovo. For example, the average labour input in FADN data stands at 1.5 AWU and it, alongside the modal value of 1.1 AWU, is considerably larger than the average of 0.7 AWU for Kosovo agriculture as reported in MAFRD (2018). The difference is ever larger if one considers the land use. With 11 ha on average and a mode of 2.6 ha, the FADN dataset clearly exceeds the 1.7 ha average for Kosovo agriculture as a whole.

TABLE 1
SUMMARY STATISTICS

	Mean	Mode	Minimum	Maximum	Standard deviation
Output (EUR)	24,705.41	3,318.20	60.00	1,665,000.00	79,164.62
Capital (EUR)	5,993.19	0.00	0.00	188,050.00	12,674.83
Family labour (AWU)	1.33	1.10	0.00	7.00	0.96
Hired labour (AWU)	0.23	0.00	0.00	41.00	1.65
Land (Ha)	10.98	2.57	0.03	650.00	29.69
Intermediate consumption	12,381.97	1,694.25	0.00	432,880.00	29,626.92

We have performed preliminary cleaning of the data in which there were obvious mistakes (e.g. negative values, observations for which the total labour input was zero). This reduced the original 1250 observation to a total of 1217 useable observations. A particular feature of the data is the very high number of zero observation for two variables: paid labour and capital. With regard to paid labour 1,198 farms (almost 93% of the sample) report that they do not use any paid labour. This appears to be a typical and plausible feature of Kosovo agriculture. With regard to capital, however, a total of 366 farms (30% of the sample) do not record any capital. While in principle (since land values are excluded from the capital measure) this is possible, as most of Kosovo farmers are asset-poor, the proportion is quite high and may raise concerns about measurement errors. However, removing them would potentially remove a large number of valid observations and for this reason we decided to keep them in the sample.

The overriding conclusion from the summary statistics in Table 1 is that despite that FADN only accounts for larger commercial farms, the dataset is still dominated by quite small farms by comparative international standards. Therefore, the question of what effects a transition towards larger agricultural holding can have important policy bearing.

METHODOLOGY

The empirical approach estimates the input-output relationship for the Kosovo agriculture via non-parametric quantile regression. The quantile regression estimates the conditional distribution of the dependent variable with regard to a set of covariates. We estimate two different quantile regressions, one for the median (0.5th quantiles) which is used to describe the nature of the input-output relationship for a 'typical' farm, and a 0.8th conditional quantile which we use to characterise a reasonably 'efficient' farm. In order to motivate this, consider the following interpretation of conditional quantiles. Upper conditional quantiles refer to farms which are able to extract more output from their given endowments, i.e. the inputs in the production function, than other comparable farms and vice versa. This means that the conditional output distribution inferred from a production function measures the unobservable farm ability to transform inputs into output, something we refer to as technical efficiency. In this particular case, the 0.8^{th} conditional quantile represents a hypothetical farm which is technically more efficient than 80% of the farms in the sample. In the terminology of Kostov *et al*. (2018), this is a 80% efficient farm. The median (i.e. 0.5^{th}) quantile regression represents a central tendency since it models a farm which sits right in the middle of the technical efficiency distribution, which is in a way typical when one is concerned with efficiency. The median regression model is essentially a model estimated by minimising the absolute deviations from the residuals instead of their squares, while other quantile regressions estimates are obtained by minimising appropriately weighted absolute deviations (see Koenker, 2005).

Here we adopt a non-parametrical approach and avoid the need to specify any pre-defined functional form for the output. The non-parametric quantile regression applied here can be expressed as:

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

$$\mathrm{st}q_\tau(u_\tau|X) = 0 \tag{2}$$

This specification has an important implication. It implies (unless explicitly assumed otherwise), a non-additive relationship. We will review this point later in the paper.

There are various non-parametric extensions of the quantile regression model, using e.g. kernel approaches (Li et al., 2007), inversion of non-parametrically estimated conditional density (Li and Racine 2008), local estimation (Yu and Jones, 2008), smoothing splines (Koenker et al., 1994, Thompson et al., 2010), penalised variograms (Koenker and Mizera, 2003) and algorithmic approximation (Jiang, 2014). This paper follows Kostov et al. (2018) in adopting the approach of Takeuchi et al. (2006). There is, however, one important deviation from their approach. While Kostov et al. (2018) impose (locally) theoretical restrictions in this paper we do not. We have tried to impose monotonicity and concavity to the estimated relationship both in the present setup and in a non-parametric mean regression framework using two different approaches (constraint-weighted bootstrapping and derivatives based constraints) but in all such attempts these constraints were not feasible with regard to the dataset.

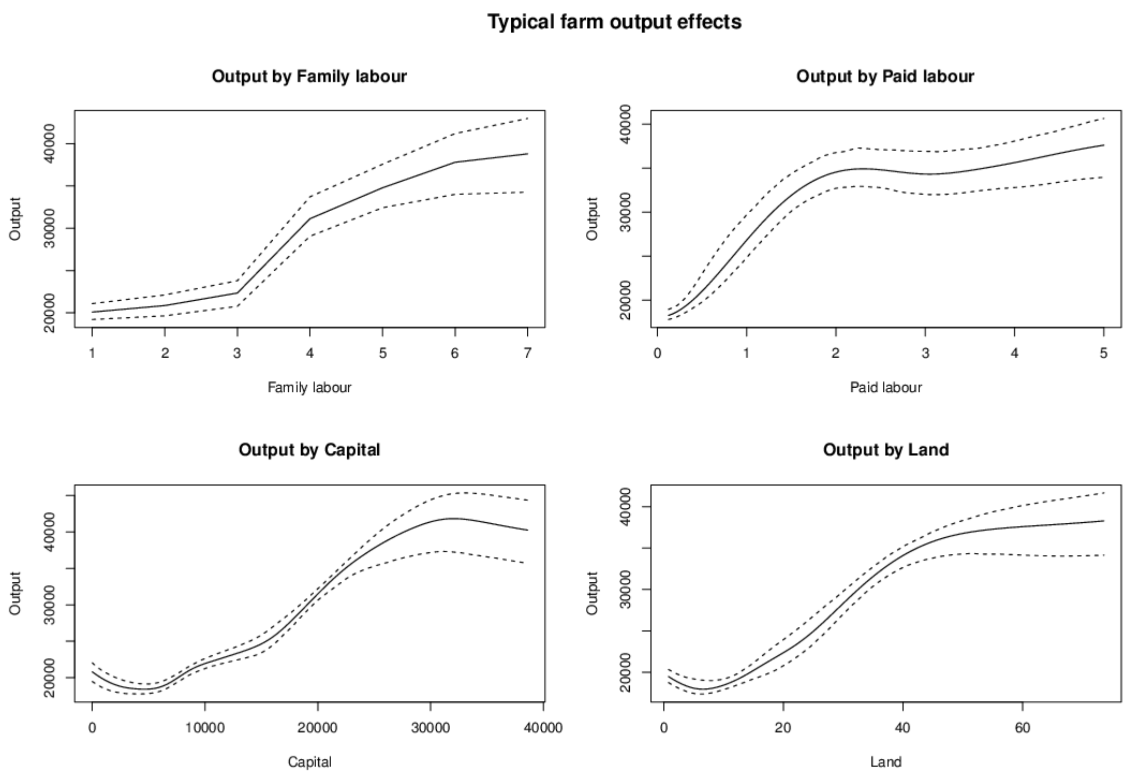
RESULTS

In order to investigate the potential effect of changes in the production factors structure we predict the output for a series or 'reference' farms. Since the estimated production models are non-additive, standard effects plots are not available. Hence in order to study the effect of any variable (i.e. production factor) we need to make any such predictions conditional on the values of the other factors of production. The standard way of achieving this is by simply averaging the other variables (see e.g. Kostov et al., 2018). Then the hypothetical reference farm takes average values for all production factors, except the one which we want to investigate. By taking a range of values for the latter partial correlation plots can be produced. The range of values over which to predict the output is best represented by the range over which the corresponding variable is observed in order to avoid out of sample prediction, which in addition to including additional errors is more likely to contain unrealistic combinations. Although averaging is not the only possible way to create such a 'reference' sample (see Kostov et al, 2008 and Kostov, 2010 for discussion of the alternatives) it by far is the simplest one. Furthermore, such an averaged reference farm is clearly consistent with the observed data and as a result the predicted outcomes should be in most part feasible, which is important when trying to draw conclusions and policy implications.

First, we look at the 'typical' farm production effects. These are derived by using the least absolute deviations (LAD) model. Since the LAD model represents the 0.5^{th} conditional quantile (i.e. the median) it shows how the output for a median efficient farm will change with the different factors of production. We call this typical because it is an 'averagely' efficient farm (in that it is right in the middle of the technical efficiency distribution) and therefore it is typical with regard to the observed in the sample technical efficiency. It is to be expected that such typical farm effects are more likely to materialise. Furthermore, we use bootstrap to produce confidence intervals for the predicted effects. In particular, in order to preserve the dependence structure in the sample that may impose practical limits to the proportions of different inputs, we employ a subsample bootstrap that utilises 90% of the estimation sample in each iteration with 1000 iterations.

We consider the two types of labour input alongside capital and land. Intermediate consumption is more difficult to interpret and for this reason we do not project its effects. Furthermore, for the family labour we only use integer values for the labour input. In principle, since labour is measured in full-time equivalents, fractional values are possible and observed within the data. However, reporting of family labour input can be prone to measurement errors, at least more than the other inputs. Furthermore, thinking about a farm providing full time employment for a number of family members is a useful way to define recommendation. We do not do the same for paid labour, which due to its often occasional use is measured in fractional units. The estimated typical farm effects are shown on Figure 1. The monotonicity of the production function, which was not imposed during estimation is violated at very low values of land and capital. This is most likely due to measurement errors. It is not possible to trim the possible outliers since, for example, as mentioned previously, a large part of the sample contains zero value for capital and hence we would need to exclude a large number of observations. Technically, imposing theoretical restrictions (as e.g. in Kostov et al, 2018) was not possible without removing a large number of 'outliers' and hence we decided not to do it.

FIGURE 1
OUTPUT EFFECTS FROM THE MEDIAN MODEL



In order to draw policy implications, consider the slope of the estimated effect. Where this slope is the steepest the marginal product for this particular input is the largest. Therefore, most productivity gains can be extracted if a farm manages to achieve an input mix that is characterised by such high marginal productivity. In particular, the results show that these productivity gains are achieved by increasing the family labour input above 3 FTE up to 4 FTE, hired between 0.5 and 1.8 FTE of paid labour, using capital of between Euro 18,000 and 23,000 and land between 25 and 40 ha. Although the highest absolute productivity gains are obtained at the higher end of the above ranges, in relative terms the corresponding marginal products are similar at the lower end. Taking into account that most farms in Kosovo are rather small, the lower end of the above ranges is much more feasible to achieve, and hence, it is appropriate to specify these ranges as (minimum) targets in order to stimulate productivity.

Therefore, one avenue for improving productivity in Kosovo agriculture is by increasing farm size (within some limits concerning feasibility as we have already discussed). Talking about productivity, it is clear that improving technical efficiency is another way to increase the marginal productivity. It would, therefore, be informative to see how output effects change with technical efficiency. To do this we follow the same approach as above, but instead of a typical farm model we use an 'efficient' one. More specifically, instead of a 0.5^{th} quantile regression we employ a 0.8^{th} quantile regression. Thus, our 'efficient' farm has 80% technical efficiency. This is by no means a very efficient farm, but we chose this level for several reasons. First, on technical grounds estimating more extreme quantiles is more problematic and will be more sensitive to measurement errors in the dataset. Then we do not want a too extreme level of technical efficiency, but rather a level that might be within the reach of a large part of the sample farms. In this way 'improving technical efficiency' would not be an unrealistic general recommendation or a distant aspiration, but rather something that is achievable under present conditions.

FIGURE 2
OUTPUT EFFECTS FROM AN EFFICIENT FARM

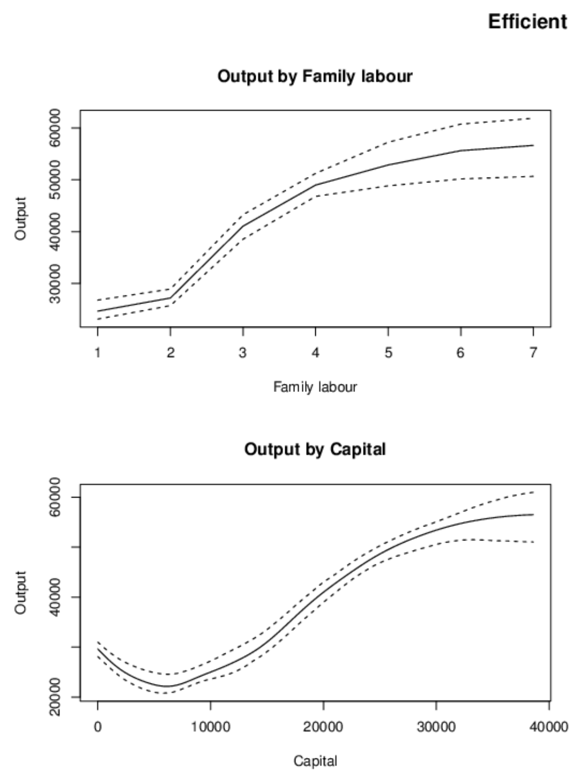
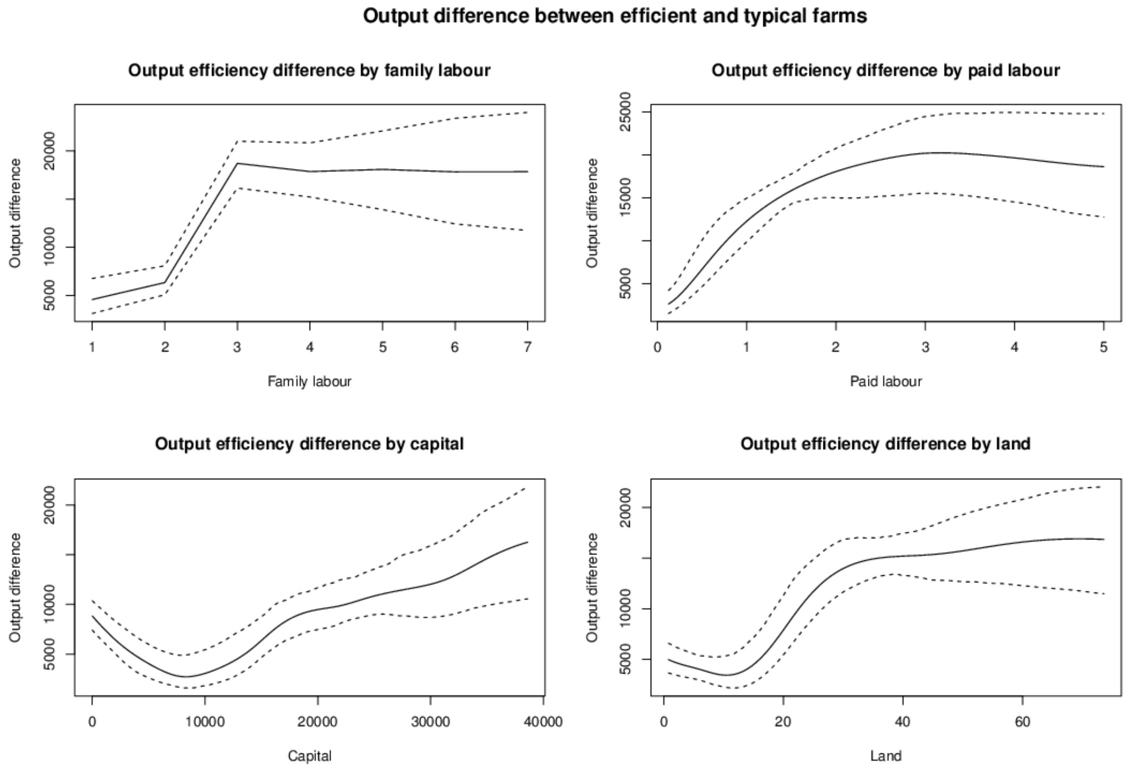


Figure 2 shows the output effects from such an efficient farm. These are similar to those from the typical farm, but the optimal marginal productivity is achieved at lower values for inputs. More specifically, the family labour needs to be between 2 and 3 FTE, hired labour between 0.5 and 1 FTE, capital from Euro 18,000 to 21,000 and land between 18 and 30 ha. Comparing this to the corresponding ranges for a typical farm, one may notice that the minimum values at which optimum productivity is achieved are broadly similar but the size effects are exhausted earlier (i.e. the marginal productivity starts to reduce at lower values for the inputs). Hence, as expected there is a kind of substitutability between size effects and technical efficiency. However, the minimum thresholds for achieving optimal productivity appear to be robust to the extent of technical efficiency at which a farm operates. Therefore, it would be reasonable for policy to aim achieving these threshold levels.

Let's take a closer look at where (in the sense of at what values of the inputs) the largest efficiency gains can materialise. To do this we predict the difference between efficient and typical farms in Figure 3. The largest productivity gains from technical efficiency (i.e. from moving from 50% to 80% technical efficiency) are obtained where the slopes in Figure 3 are at their steepest. This means that increasing technical efficiency is most beneficial when there are 2 to 3 FTE of family labour, up to 0.8 FTE of paid labour, Euro 15,000 to 18,000 of capital and between 18 and 25 ha of land. One can notice that these productivity gains either kick in at lower thresholds or reduce at lower upper limits, compared to those identified in Figures 1 and 2.

FIGURE 3
OUTPUT EFFECTS DUE TO EFFICIENCY SHIFT



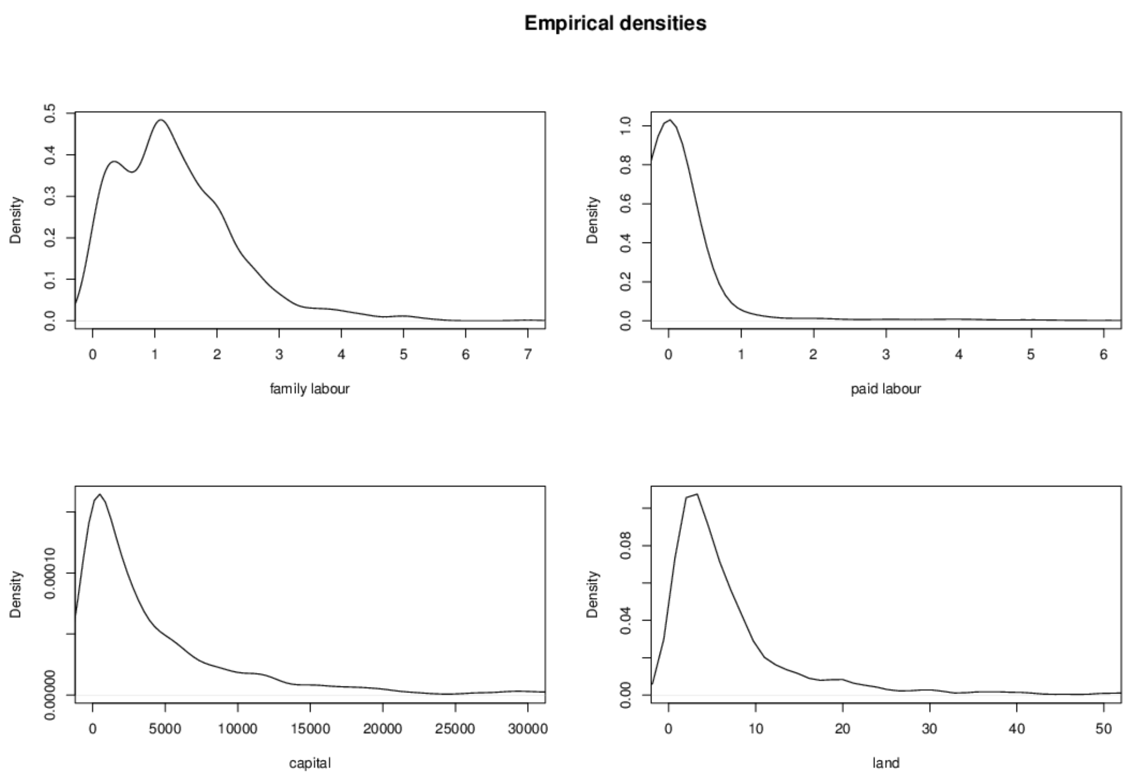
The corresponding lower and upper threshold for productivity gains associated with the typical farms, efficient farms and the difference between these two (i.e. a technical efficiency shift) are summarised in Table 2. It demonstrates that while size matters for productivity, productivity gains can be achieved at much lower size thresholds when technical efficiency is improved. Therefore, there are two distinct channels, namely size and technical efficiency, to achieve these gains. Size effect can be stimulated by policy measures. There are farms that already fall within programmes such as land consolidation and capital equipment subsidies which are in place, while technical efficiency might be improved by e.g. extension services which, as mentioned previously, require substantial improvement in governance, farm coverage and collaboration with private Kosovo and international advisors, and relevant NGOs activities.

TABLE 2
IDENTIFIED THRESHOLDS FOR PRODUCTIVITY GAINS

	Typical Farm		Efficient farm		Technical efficiency shift	
	Lower threshold	Upper threshold	Lower threshold	Upper threshold	Lower threshold	Upper threshold
Family labour (AWU)	3	4	2	3	2	3
Hired labour (AWU)	0.5	1.8	0.5	1	0	0.8
Capital (USD)	18,000	23,000	18,000	21,000	15,000	18,000
Land (Ha)	25	40	18	30	18	25

It is nevertheless important to look at the feasibility of reaching the thresholds specifies in table 2. For this reason, we plot in Figure 4 the empirical density distribution for these inputs from the estimation sample.

FIGURE 4
EMPIRICAL DISTRIBUTIONS OF LABOUR, LAND AND CAPITAL



The family labour distribution has two modes at around 0.5 and slightly above 1 FTE. The majority of farms are below the identified minimum threshold (2 to 3 FTE). That said, there are farms that already fit in that range and reaching it for most cases will mean adding 1 full-time family member. On the one hand, this looks feasible, but, on the other, in most cases it leads to doubling the family labour input. This means a major expansion, which points towards larger farms. To the extent that paid labour is concerned, the typical Kosovo farm employs none. But the thresholds for achieving optimal productivity are rather low, and in the case of technical efficiency shift, the lower threshold is zero. One may say that incorporating paid labour seems to happen when family labour is insufficient (and probably for casual and seasonal tasks). Hence we can probably subsume the expansion of paid labour into the overall farm labour requirements. And productivity gains will materialise from an overall relatively moderate farm size expansion.

However, the empirical distribution of capital and land shows that the identified threshold are quite high and appear to be out of the reach of most farms without further measures. In fact, the lower thresholds for optimal technical efficiency shift for capital and land, and correspond to the 0.9^{th} and 0.88^{th} empirical quantile, meaning that about 90% of farms in Kosovo are below the level of capital and land use that can be the most beneficial for obtaining optimal productivity gains from technical efficiency shift. Taking into consideration this, it appears that out of the two channels for achieving productivity gains, given the present farm structure in Kosovo, farm scale increases are preferable to technical efficiency if the government contemplate productivity enhancing policy instruments, since they carry a greater potential. Furthermore, farm size increase may lead to adopting different, more modern technologies. One would nevertheless need to note that the farm scale type of policy measures such as land consolidation programmes and investment subsidies are likely to be considerably more expensive than measures aimed at improving technical efficiency such as better extension services and technical advice. Although, according to our results, even not optimal in terms of marginal productivity, any increase in technical efficiency is beneficial and may prompt demand for more land and capital, thus, facilitating changes in scale. Therefore, the main conclusion from the analysis is that bearing in mind the current farm structure in Kosovo, productivity enhancements can be obtained by a careful balance of policy measures facilitating both the efficiency increases and farm scale augmentation.

CONCLUSIONS

We investigate the possible productivity enhancements that the agricultural sector in Kosovo can achieve via two distinct channels, namely technical efficiency and farm scale increase. These two channels appear to be in some ways complementary. In particular, scale has greater potential to achieve increase in productivity. However given the present structure of Kosovo farms which is dominated by very small holdings, such scale enhancements are not particularly likely. Given that technical efficiency improvements allow farms to achieve optimal productivity increases at smaller farm sizes, adequate policy measures designed to enhance technical efficiency are not only more feasible, but also necessary to realise productivity gains from any economies of scale in a more realistic time-frame. Therefore, although size is the key to a longer-term higher productivity, it is the technical efficiency that holds most promise in the short- to medium-term for development of a more productive agricultural sector in Kosovo.

ENDNOTES

1. SO, as measured in FADN, is the average monetary value of agricultural output at farm-gate price.
2. Annual Work Unit is a full-time equivalent employment and corresponds to one person full-time employed on agricultural holding.

REFERENCES

FAO. (2019). *Commercialisation of Smallholder Farms in Kosovo. Final Report.*

Jiang, Y. (2014). Nonparametric quantile regression models via majorization-minimization algorithm. *Statistics and Its Interface*, *7*, 235–240.

Koenker, R. ,Ng, P., & Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, *81*, 673–680.

Koenker, R., & Mizera, I. (2003). Penalized Triograms: Total variation regularization for bivariate smoothing. *Journal of the Royal Statistical Society (B)*, *66*, 145–163.

Kostov, P. (2010). Do buyers' characteristics and personal relationships affect agricultural land prices? *Land Economics*, *86*(1), 48–65

Kostov, P., Davidova, S., & Bailey, A. (2018). Effect of family labour on output of farms in selected EU Member States: A non-parametric quantile regression approach. *European Review of Agricultural Economics*, *45*(3), 367–395.

Kostov, P.M., Patton, M., & McErlean, S. (2008). Non-parametric analysis of the influence of buyers' characteristics and personal relationships on agricultural land prices. *Agribusiness*, *24*(2), 161–176.

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

Li, Y., Liu, Y., & Zhu, J. (2007). Quantile regression in reproducing Kernel Hilbert Spaces. *Journal of the American Statistical Association*, *102*(477), 255–268.

MAFRD (Ministry of Agriculture, Forestry and Rural Development). (2018). *Green Report*. Prishtina, Kosovo.

Takeuchi, I., Le, Q.V., Sears, T.D., & Smola, A.J. (2006). Nonparametric quantile estimation. *Journal of Machine Learning Research*, *7*, 1231–1264.

Thompson, P., Yuzhi, C., Moyeed, R., Reeve, D., & Stander, J. (2010). Bayesian nonparametric quantile regression using splines. *Computational Statistics & Data Analysis*, *54*(4), 1138–1150.

Yu, K., & Jones, M.C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, *93*, 228–237.