Original scientific paper
DOI: /10.5513/JCEA01/22.2.3147
Journal of Central European Agriculture, 2021, 22(2), p.486-496

Bootstrapped input efficiency use of specialized potato production in Kosovo

Blend FRANGU[1], Arben MUSLIU[2] (✉), Jennie S. POPP[3], Nathan KEMPER[3]

[1] Department of Agricultural and Resource Economics, University of Saskatchewan, Saskatoon, Saskatchewan, S7N 5A8, Canada

[2] Department of Agricultural Economics, Faculty of Agriculture and Veterinary, University of Prishtina "Hasan Prishtina", Pristina, Kosovo

[3] Department of Agricultural Economics and Agribusiness, University of Arkansas, Fayetteville, AR 72701, USA

✉ Corresponding author: arben.musliu@uni-pr.edu

Received: January 9, 2021; accepted: February 25, 2021

ABSTRACT

Although Kosovo's agriculture initiatives have allowed for opportunities to spur vegetable production, little has been done to address how efficient are farmers at using inputs. The state of input efficiency in growing potatoes is examined in the study using farm survey data. There is also a comparison of potato yields as a measure of productivity with different countries in Southeast Europe and with some emphasis on input use. After accounting for suspected bias with the bootstrap input-oriented model, input efficiency ranged from 0.39-0.91 with an average of 0.73. Depending on the farm, a naive model would induce a bias of 0.04-0.17 in input efficiency use. This bias can vary with sample size. Additionally, the findings suggest an encouraging input efficiency advantage for farmers who care about their soil quality as they practice potato production. One policy implication of the results suggests further input use decreases because the sampled farms are found to operate under decreasing returns to scale.

Keywords: non-parametric frontier technique, bootstrap input-oriented model, specialized potato production, Southeast Europe

INTRODUCTION

In response to a fragile economic recovery, Kosovo has begun restructuring the agricultural sector by introducing policies that may be in line with the expectations of the European Union (EU). Today, policymakers and interested international organizations operating in Kosovo seek to put vegetable production in context with other economic activities. The approach of Kosovo's government to policy formulation in the agriculture sector, however, appears to put less of an emphasis on what is happening on the ground and to not consider how, for example, its policies are affecting the farmers. Although one can readily identify vegetable crop production activities with a great impact on Kosovo's economy, there is little evidence in the literature as to what the state of input efficiency is throughout the country. Consequently, little is known about vegetable productivity, and even less is known about input efficiency use

An example can be taken from Kosovar farms in the northern town and municipality of Vushtrri and the southeastern town and municipality of Vitina as two places known for their potato farms. The specialized potato production is particularly common in the northern part of Kosovo. A common strategy among these farms seems to be that by increasing the use of inputs, there could be increases in full season yields. However, it can be argued that reviewing the use of inputs is necessary. In the absence of empirical results, it may be counterproductive to undertake decisions to increase or decrease the use of inputs. For instance, if there happens to be mismeasured utilization of input resources during production, northern and southeastern farmers could benefit from expanding their expertise in using inputs more efficiently. In the context of vegetable production, Frangu et al. (2018) reasoned that Kosovar farmers growing greenhouse tomatoes and peppers were found to not use inputs entirely efficiently. Further clarity is available as several studies have found the efficient use of inputs to be an essential determinant of vegetable production (Nikolla et al., 2013; Alboghdady, 2014).

To review input use, there is data collected on specialized potato production with information ranging from a farmer's demographics to farm data. Following this process, full season yields, among Kosovar potato farms, can be regarded as the primary indicator reflecting productivity. That is because farmers' production decisions depend on how to bolster the prospect for higher yields (Tomek and Kaiser 2014). Particularly, the study examines if inputs such as potato seeds for planting, macro-nutrients of nitrogen (N), phosphorus (P), and potassium (K) or NPK fertilizer, pesticides, and labor can be improved while keeping potato full season yields unchanged. It is important to know whether an increased or decreased use of inputs could affect such yields to be maintained in the country.

To conduct the analyses, there is an application of the non-parametric framework of bootstrap data envelopment analysis. After the development of the benchmarking model, the study investigates possible causes of variations in efficiencies by accounting for environmental variables. The specific question to be addressed in this research is: Do input use increases or decreases help sustain full season (potato) yields?

MATERIALS AND METHODS

Materials

The study's data were collected in 2018 with a sample covering 46 farms that specialize exclusively in potato production. After accounting for six outliers, the sample was reduced to 40 farms. Of these farms, there were some that intercrop, and one common crop used in rotations was that of winter wheat. The sampled farms were from the northern town and municipality of Vushtrri and the southeastern town and municipality of Vitina and their surrounding villages. Among the villages in Vushtrri where the data were collected included those of Pestovë, Lidhja e Lezhës, and Dobërllukë while among those in Vitina were Debëlldeh, Skifteraj, and Budrikë e Epërme. In collecting the data, farmers were randomly selected.

While the study recognizes early the water supply importance in determining full season potato yields, during the data collection, there were no reliable records of water usage for potato fields. For example, northern farmers have been using the Iber Lepenc (an approximately 49 kilometer water supply canal) for potato irrigation needs but did not seem to have accurate water usage records. In contrast, the southeastern Kosovo farmers revealed that they mainly depend on wells for potato irrigation purposes. Other data collected come from the Hydrometeorological Institute of Kosovo (2019) that has precipitation in millimeters (mm) during the year of 2018. Generally, the towns of Vitina and Vushtrri are also part of broader districts in Kosovo. The former is part of the Gjilan district and the latter of the Mitrovica district. Depending on the geographical proximity of Vitina's villages to the existing weather stations (e.g. Bilinica, Perlepnica, and Verbica e Zhegocit) situated in the district of Gjilan, values of precipitation in mm were obtained from the closest weather station. Such data in the context of Vushtrri were collected from the (single) available weather station of Mitrovica.

The summary statistics of separate interest (Table 1) suggest that farmers produce about 29,778 kilograms (kg) of potatoes per hectare (ha) over the full course of a season. Relatedly, the Ministry of Agriculture, Forestry and Rural Development or MAFRD (2018) implies that potato yields per ha are roughly 27,600 kg combining small and large farm sizes. Comparably, there is a mean of 27,700 kg/ha for the 2010-2019 period in the European Union (Hajdu, 2020). This may imply early that Kosovo is well positioned as far as its yields of potatoes are concerned. From the available inputs, information is collected about NPK fertilizer use, planting seed, the liquid amount of pesticide, and seasonal and full time labor. These have means of 1,180 kg/ha, 1,524 kg/ha, 3.19 liters/ha, 6 seasonal workers, and 2 full time workers, respectively. From the environmental variables. It can be helpful to single out seed varieties and soil care considerations. About 75% of farmers have used the seed variety of Agria while the rest have planted different seed varieties including, for instance, Riviera, Carrera, and Sinora among others. The variety of Agria is non-locally developed and exhibits a yellow color. It appears to do well in Kosovo's weather conditions and with some separate success in the production of chips. Another interesting descriptive statistic is that only 30% of the sampled farmers care about soil quality and that they act towards improving it. Policy-wise, it does not seem that much has been accomplished to improve soil considerations among farmers in northern and southeastern Kosovo.

Methods

Early work by Debreu (1951), Farrell (1957), and Shephard (1970) used linear programming techniques to obtain envelopment estimators. Data envelopment analysis (DEA) is based on Farrell's (1957) pioneering work (Simar and Wilson, 1998).

Table 1. Descriptive statistics (N = 40)

Statistic	Unit	Mean	St. Dev.	25th PCTL	Median	75th PCTL
Output						
Full season yields	Kg/ha	29,778	12,303	23,750	30,000	35,000
Inputs						
NPK fertilizer	Kg/ha	1,180	499	800	1,000	1,525
Planting seed	Kg/ha	1,524	453	1,300	1,600	1,850
Pesticide	Liters/ha	3.19	5.35	1	1	1.10
Seasonal labor	No. of workers	6	4	4	5	≈8
Full time labor	No. of workers	2	2	0	2	4
Environmental variables						
Farm experience	Years	26.90	12.11	17.50	30	35
Education	Years	11.10	2.24	8	12	12
Region[1]	0/1	0.50	0.51	0	0.50	1
Farm expansion[2]	0/1	0.42	0.50	0	0	1
Farm size	Ha	6.62	8.11	1.50	2.60	10.13
Seed variety[3]	0/1	0.75	0.44	0.80	1	1
Seed price	EUR/kg	0.94	0.09	0.90	0.95	1
Precipitation[4]	millimetre	51.53	7.56	44.64	46.38	56.68
Soil care[5]	0/1	0.30	0.46	0	0	1

N; observations; PCTL, percentile; Kg, kilogram; ha, hectare; EUR, euro currency; NPK, nitrogen (N), phosphorus (P), and potassium (K); No., number. [1] is coded as 1 = for farmers that are from northern Kosovo, 0 = from southeastern Kosovo. [2] is coded as 1 = for farmers who consider expanding farm operations and potato area, 0 = otherwise. [3] is coded as 1 = for farmers who have chosen the most grown seed variety of Agria, 0 = otherwise. [4] precipitation in millimeter values were taken as averages of the 2018's twelve-month period. Depending on the geographical proximity of a farm to a certain weather station, there was some variation in this environmental variable. 5 is coded as 1 = for a farmer who cares about soil quality and acts towards improving it, 0 = otherwise

Today, decision making units (DMUs) performance problems are often evaluated with DEA, as a non-parametric approach (Toloo and Salahi 2018). This generally facilitates the process of quantifying efficiency use. The other strand of the literature relies on parametric methods such as the stochastic frontier analysis (SFA) of Aigner et al. (1997) that may achieve efficiency scores comparable to those under DEA. Thus, choosing one approach over the other can become a point of discussion as in Błażejczyk-Majka and Kala (2015).

Murillo-Zamorano (2004) rightfully notes that SFA requires the researcher to specify the functional form for the inefficiency error term, and this very step could be a source of both specification and estimation problems. On a more observable note, DEA is criticized by Coelli (1995) for its inflexibility to allow for any impacts of measurement error. However, the study is concerned more about SFA's likely emerging statistical problems from a potentially misspecified functional form of examining input use among the potato farms. Although SFA points to the random effects outside the control of farmers that may affect full season (potato) yields and that it could probably accommodate better for weather conditions - its relatively erratic distribution assumptions about the input inefficiency term is what motivates us to rather use DEA. Also, in the study, there may be no need to impose a priori a functional form of the efficient frontier while outlining the possible research question related to input use.

Reasonably, the study decides to use the input-oriented Banker-Charnes-Cooper or BCC (Banker et al., 1984; Banker et al., 2004) and Charnes-Cooper-Rhodes or CCR (Charnes et al., 1978) models. The aim is to capture both the technical and primarily the pure technical input efficiency uses for the potato farms. One important reason as to why the BCC and CCR models are adopted jointly is that they allow studying scale efficiency. Using both models can shed light on the main source of input inefficiency (if any) of each potato farm. It is of interest to define first the naive DEA estimator (Simar and Wilson, 1998), prior to establishing the input-oriented BCC model, which assumes variable returns to scale (VRS).

$$\hat{\Omega}_{DEA} = \{- y_i + Y\gamma \geq 0;\ x_i - X\gamma \geq 0;\ \textstyle\sum_{i=1}^{N} \gamma_i = 1,\ \gamma \geq 0\} \quad (1)$$

Following the DEA estimator in equation (1), the naive input-oriented BCC model under VRS can be formulated.

$$\text{Min}\hat{\theta} = {}_{\gamma}\theta$$

Subject to:

$$\{- y_i + Y\gamma \geq 0;\ \theta x_i - X\gamma \geq 0;\ \textstyle\sum_{i=1}^{N} \gamma_i = 1,\ \gamma \geq 0\} \quad (2)$$

From (2), θ ("0 ≤ θ ≤ 1") is the pure technical efficiency (PTE) score for the i-th potato farm. The aim is to essentially find the minimum θ that reduces the input vector x_i to θx_i. That is by ensuring that the study obtains, at minimum, the full season yield level y_i. The constants or weights are represented by γ, which is a N x 1 vector while y_i denotes the full season yields and x_i is the vector of inputs including NPK fertilizer, planting seed, pesticide, and seasonal and full time labor for the i-th potato farm. Part of the analysis is to have Yγ which suggests the full season yield level vector of the theoretically efficient potato farm. Also, because the BCC model is input-oriented, Xγ indicates the minimum input use of the theoretically efficient potato farm provided that the actual full season yield level is achieved by the i-th potato farm.

In contrast, to obtain the CCR model, one may simply omit the expression of $\sum_{i=1}^{N} \gamma_i = 1$ from the other constraints in (2). Nevertheless, an advantage of the BCC model under VRS is that return to scale properties are not fixed by assumption. Therefore, the study chooses the BCC model under VRS as the desirable input related model. Here, there is a use of the CCR model only when the study examines scale efficiency (SE) which is a ratio of input efficiency received from the CCR to that of the BCC model. The eventual score generated from SE cannot be higher than 1, and it is one when the VRS and CRS technologies correspond to each other (Bogetoft and Otto, 2011). Interpretively, a potato farm with an SE of 1.00 has the most productive scale size, while SE values lower than 1.00 would indicate further room for improvement to achieve the respective scale.

After having formally described the naive BCC and CCR models and that quantifying input efficiency use is indeed possible, there can emerge sample size concerns.

The sample size concern arises when in addition to the sampled farms specializing in potato production, there may be other (although few) such farms that could presumably be more efficient in input use. Therefore, the calculated input efficiency of those inefficient farms based on what is observed in the sample can be viewed as an upward biased estimate of their true input efficiency. An interesting development to address bias in efficiency scores includes the work of Simar and Wilson (1998; 2000). They propose the use of the bootstrap DEA, which may enable one to achieve bias corrected efficiency scores. In the context of the bootstrap BCC model, to compute the efficiency scores, there is a need to first define the DEA estimator $\hat{\Omega}^*_{DEA}$.

$$\hat{\Omega}^*_{DEA} = \{- y_i + Y\gamma \geq 0;\ x_i - X^*\gamma \geq 0;\ \textstyle\sum_{i=1}^{N} \gamma_i = 1,\ \gamma \geq 0\} \quad (3)$$

With the bootstrap procedure in (3), the study views the potato farms as random draws from the data generating process (DGP) of the farm population. In this situation, such draws that originate from the sample of potato farms can be thought of also as draws from the underlying farm population. This is useful since it implies that one might see value around the known bootstrap distribution of farms. How good this distribution is, depends on the circumstances of the known DGP being a consistent estimator of the unknown DGP. Following a prespecified number of drawn samples, input efficiency scores are obtained, which portray an estimate of the true distribution of such scores. This presents a pathway that will be valuable for the study in which there is a formulation of the bootstrap input-oriented BCC model under VRS.

$$\mathrm{Min}\hat{\theta}^* = {}_{\gamma}\theta$$

Subject to:

$$\{- y_i + Y\gamma \geq 0;\ \theta x_i - X^*\gamma \geq 0;\ \textstyle\sum_{i=1}^{N} \gamma_i = 1,\ \gamma \geq 0\} \quad (4)$$

From these estimations, the bootstrap efficiency $\hat{\theta}^*$ can be considered as an estimate of $\hat{\theta}$. That is, in a similar way, as in the case of using naive estimations. For example, $\hat{\theta}$ can be viewed as an estimate of θ. This understanding is supported in the work of Simar and Wilson (1998). By omitting as previously the expression of $\sum_{i=1}^{N} \gamma_i = 1$ from the other constraints in (4), the bootstrap CCR model is obtained. Thus, the question of whether potato farms are efficiently using their inputs can be settled empirically.

An additional step to input efficiency use is to consider a post-efficiency analysis. There is a recognition that there may be environmental or non-discretionary variables that influence efficiency scores. For example, one can allow the perception that farmers may have varying production experiences and views on how they can best approach the growth of potatoes. In this aspect, these can add more clarity as to why some farmers tend to more (less) efficiently use inputs as compared to others. This analysis involves formulating an ordinary least squares (OLS) regression model under the assumption of normally distributed errors. By estimating the regression model that reflects the null hypothesis, the aim is to use its estimates later to generate bootstrap samples. To apply this procedure, first, a likely true data generating process (DGP) equation is specified as in (5).

$$e = Z\beta + \varepsilon,\ \varepsilon \sim IID(0, \sigma^2) \quad (5)$$

where e is a N by 1 vector of bias corrected PTE scores, Z is a N by K matrix of environmental variables such as farm experience, education, region, farm expansion considerations, farm size, seed varieties, seed input prices, weather conditions including precipitation in mm, and soil care considerations, and β is a N by 1 vector of coefficients to be estimated. The error is assumed to be independent and identically distributed (IID) with zero mean and constant variance (σ^2). While the study primarily obtains estimates for the environmental variables, it also important to account for their estimated interval of values. For this purpose, there is a consideration to report first the 95% confidence interval (i.e. α = 0.05) with a range of values that are likely to contain the true value of these environmental variables.

$$\mathrm{Prob}\left[EV_E - t_{(1-\alpha/2),\,[n-K]}\sqrt{s^2(Z'Z)^{-1}} \leq EV_{TV} \leq EV_E + t_{(1-\alpha/2),\,[n-K]}\sqrt{s^2(Z'Z)^{-1}} \right] = 1 - \alpha \quad (6)$$

From equation (6), the lower and upper bounds of the estimate associated with any of the environmental variables will be obtained. EV_E and EV_{TV} denote the estimate and the true values, respectively. The $t_{(1-\alpha/2),\,[n-K]}$ pertains to the critical value from the t-distribution with [n – K] degrees of freedom. The standard error of the environmental variable estimate is computed by $\sqrt{s^2(Z'Z)^{-1}}$

DOI: /10.5513/JCEA01/22.2.3147

where $s^2 (Z' Z)^{-1}$ can be expressed as $s^2(\sum_{i=1}^{N} z_i z_i')^{-1}$ and its computation without the square root would be the estimated variance of the environmental variable. In fact, in (6), the inequalities hold with the probability 1 – α (e.g. 1 – 0.05 = 0.95). Commonly, we know 1 – α as the confidence coefficient while α ($0 < \alpha < 1$) denotes the level of significance. At this stage, we analyse once more model (5), which is a fully specified parametric model with β and σ^2 defining one DGP only. To apply a bootstrap DGP, the study can do so by using the parametric bootstrap procedure. First through OLS, the restricted estimates of $\tilde{\beta}$, and $\tilde{s}^2$ are obtained. The next step, therefore, includes estimating the bootstrap DGP for the potato farms and is given by

$$e^* = Z\tilde{\beta} + \varepsilon^*,\ \varepsilon^* \sim IID(0,\tilde{s}^2) \tag{7}$$

where the stars indicate that the data are simulated. The procedure requires to draw a bootstrap sample from the model (7), and initially a n-vector of ε^* is acquired from the $N(0, \tilde{s}^2 I)$ distribution. The number of replications is chosen to be 999, and one criterion for choosing the bootstrap sample is given by α (B + 1) where α is the chosen significance level, and B is the number of drawn samples (Davidson and MacKinnon, 2003). An additional procedure involves constructing the normal bootstrap confidence intervals.

$$\text{Upper bound} = \left(T - \hat{B}^*\right) + z_{1-\frac{a}{2}}\widehat{SE}^*\left(T^*\right);\ \text{Lower bound} = \left(T - \hat{B}^*\right) - z_{1-\frac{a}{2}}\widehat{SE}^*\left(T^*\right) \tag{8}$$

For both equations in (8), $z_{1-a/2}$ is essentially the 1 - a/2 of the standard-normal distribution and the bootstrap estimate of sampling variance ($\hat{B}^*$) is used in the procedure (Fox and Weisberg, 2018). The $\widehat{SE}^*$ is the estimated bootstrap standard error corresponding to an environmental variable with T^* as its bootstrap test statistic. The study chooses a common significance level of 5% (α = 0.05), which will enable constructing the 95% confidence intervals of the bootstrap estimates. Should the estimated coefficients and standard errors vary under the bootstrap specification, then a comparison with those obtained under OLS and their impact on bias corrected input efficiency scores may be discerned.

RESULTS AND DISCUSSION

All efficiency scores presented in this section are on a scale of 0 to 1.00 with a score of 1.00 indicating that the potato farm is 100% efficient relative to other farms in the sample. Figure 1 shows that 17 out of 40 farms under the naive BCC model are efficient. That is assuming that these farms operate under VRS. On average, under this specification, there is an efficiency score of 0.83, while under the bootstrap BCC model, it is only 0.73. For example, farms found entirely efficient under the naive BCC model do not remain so in the use of inputs after accounting for potential bias through the bootstrap procedure. When one examines the descriptive statistics for added bias, it emerges that the bias varies from 0.04-0.17 with an average of 0.10. This is practically relevant for farmers that consider reviewing their input use levels and it may aid in the decision of whether or not the respective range of bias can affect their competitiveness.

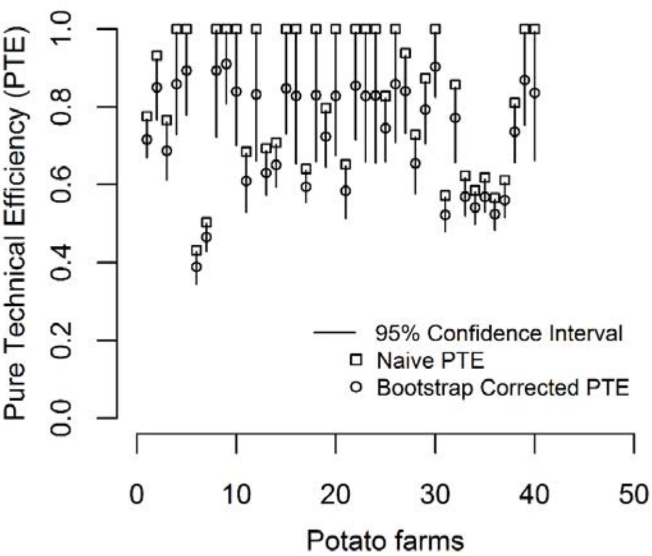

Figure 1. Bootstrap and naive input-oriented Banker-Charnes-Cooper (BCC) pure technical efficiency comparison

The decision to increase or decrease the use of inputs so that full season yields can be sustained may depend on analysing the certainty related to the range of input efficiency scores. Thus, the lower and upper 95% confidence intervals associated with these estimates allow for a discussion of the study's sampling variability.

This could influence the certainty attached to the non-parametrically generated efficiency scores. In the literature, there does not seem to be a direct simplification of the mechanics as to how one may transition from the naive to the bootstrap confidence interval. For this reason, such mechanics are discussed shortly here. That is, to obtain the lower bound of the bootstrap 95% confidence interval for farmer i, one subtracts the estimated bias from the estimated naive input efficiency score. Likewise, the upper bound is achieved by adding the respective bias to the estimated naive input efficiency score.

Consistent with this discussion, one may ask under what returns to scale do the sampled farms operate? Table 2 shows that depending on the farm and perhaps family tradition in potato cultivation, some farms may attain a better use of inputs than others. The notion of scale efficiency (SE) helps us understand the loss of those farms, other than SE farms, that are not operating optimally. One finding of the study shows that once bias is accounted for in input efficiency use, the sampled farms appear to be operating under decreasing returns to scale (DRS). This finding implies that the farms are already oversized, and full season yields will increase less than the increase in inputs. It is of interest to indicate that all these farms are commercially oriented, and although potatoes are their specialized crop, there seems to be an apparent need for further input use improvements. This runs contrary to the belief among the sampled farmers that increased use of inputs will help them transition to increased yield levels. In the introduction, the study posed the question as to whether input use increases or decreases help sustain full season (potato) yields. After finding the farms to be operating under DRS in the bootstrap specification (Table 2), there may not be likely ways to ensure that yields can be sustained through increases in input use. What is suggested here is that the sampled farms can scale down their use of inputs where needed.

In Table 2, SE efficient farms under the naive input efficiency analysis have yields of 39,633 kg/ha over the growing season period. Comparably, farms under DRS happen to achieve only 25,554 kg/ha. The variability in these productivity values is 13,224 kg/ha for SE farms, and this variability is smaller for DRS farms (9,259 kg/ha). Because of the bias in efficiency scores, it is unlikely for these values to be entirely representative, and the findings suggest no SE farms in the bootstrap specification. Previously, the 17 out of the 40 farms using naive DEA were likely misclassified as SE. Under the bootstrap specification, it is found that all farms are experiencing DRS. Relatedly, the study singles out in Table 2 the bootstrap based input and output mean (median) values. These values are the same as those of the descriptive statistics in Table 1 because all the sampled farms are found to be operating under DRS. For instance, there is an output mean of 29,778 kg/ha of potato full season yields with variability of 12,303 kg/ha. This suggests a non-negligible under representation of 4,224 kg/ha of the mean output value for DRS farms under the naive compared to the bootstrap specification.

Comparably Poljak and Butorac (2014) reasoned that in intensive potato producing parts of Croatia, there is a mean yield of 25,000-30,000 kg/ha. They also argue that there is hope with improved agronomic practices to obtain a mean yield of as high as 40,000-45,000 kg/ha. In the neighboring country of Serbia, nonetheless, the mean yield of potatoes seems to suggest a low of 10,500 kg/ha (Novkovic, 2014). The weather conditions in Kosovo and its specialization in production can prove conducive to achieving high yields. Therefore, Kosovo can be ranked in line with the recently high potato producing countries of Southeast Europe. Two examples could include Greece and Slovenia, which have potato yields of 28,490 kg/ha and 26,650 kg/ha, respectively (Chiurciu, 2020). Nevertheless, there are reasons to be cautious about being able to sustain yields in northern and southeastern Kosovo given that the farms are operating under DRS. Therefore, one consideration would be to maintain full season yields by reviewing how efficiently inputs are being used.

As farms in northern and southeastern Kosovo are experiencing DRS, it can be reasoned that uniform decreases in the use of inputs are what can make them more efficient. Take, for instance, the NPK fertilizer input. Farmers in Kosovo may often apply this input without conducting soil tests to identify the soil needs for N, P, and K. The research of Gondwe et al. (2020) shows that the use of NPK fertilizers more than recommended rates will not increase potato yields. Generally, Rüdelsheim and Smets (2012) discuss agricultural practices for the potato crop in the European Union. They suggest that if farmers were to expect to achieve yields of 40,000-50,000 kg/ha, then the needs for N could be 224-269 kg/ha while for P and K would be around 64-90 kg/ha, and 377-430 kg/ha, respectively. The combined upper values of these NPK requirements add up to 789 kg/ha which is lower by 391 kg/ha than the mean value of NPK fertilizer use among the study's DRS farms. Nevertheless, it is helpful to err on the side of caution and note that there can be different factors affecting a Kosovar farm's parcel of land needs for NPK. These can include crop rotation, seed variety (e.g., Agria, Riviera, Carrera, Sinora, or other), and information obtained through soil testing about the state of the N, P, and K already in the soil.

Another germane example from the results could include the pesticide input. The study finds a mean of 3.19 liters/ha but with a large variability of 5.35 liters/ha indicating that there may not be uniformity in pesticide applications. The motivation was to look further into this, and the study learns that northern farms have larger pesticide applications compared to southeastern farms. For instance, the former have a mean pesticide application of 5.44 liters/ha with variability of 6.94 liters/ha while the latter exhibit a more stable mean pesticide application of 0.95 liters/ha with variability of 0.22 liters/ha. This may raise a question as to whether adequate information is readily available for northern farms so that unnecessary pesticide applications are avoided. In line with this discussion, decreases in input use may be possible but this may require an increased level of expertise related to precise input applications during production. Nonetheless, Kosovar farmers appear to lack the knowledge to use new production technologies (Gjokaj et al., 2017), and adopting such technologies for more efficient input use can be one area that the Government of Kosovo could help with.

After estimating input efficiency use among specialized potato farms, there is an examination as to why some farms have higher efficiency scores than others. Thus, the second step to input efficiency analysis involves estimating a linear regression using bias corrected input

Table 2. Naive and bootstrap input information under scale efficiency and decreasing returns to scale

	Naive PTE				BSC PTE	
	SE (N = 12)		DRS (N = 28)		DRS (N = 40)	
Output and inputs	Mean (Median)	St. Dev.	Mean (Median)	St. Dev.	Mean (Median)	St. Dev.
Full season yields	39,633 (30,000)	13,224	25,554 (28,000)	9,259	29,778 (30,000)	12,303
NPK fertilizer	1,217 (1,250)	587	1,164 (1,000)	467	1,180 (1,000)	499
Planting seed	1,633 (1,550)	328	1,477 (1,600)	495	1,524 (1,600)	453
Pesticide	0.73 (1)	0.51	4.25 (1)	6.12	3.19 (1)	5.35
Seasonal labor	≈7 (≈6)	≈5	≈6 (5)	≈3	6 (5)	4
Full time labor	≈2 (3)	≈2	≈3 (2)	≈2	2 (2)	2

Note: ≈ denotes the rounded number considering the last decimal place value of workers for seasonal and full time labor. BSC, bootstrap corrected; PTE; pure technical efficiency; N, number of observations; SE, scale efficiency; DRS, decreasing returns to scale; NPK, nitrogen (N), phosphorus (P), and potassium (K). The variables of full season yields, NPK fertilizer, and planting seeds share a common unit of kilograms/hectare while pesticide has a unit of liters/hectare. The number of workers is a unit that pertains to both seasonal and full time labor efficiency scores. It is of interest to complement the results with those of the parametric bootstrap regression procedure to examine whether significantly different coefficients or standard errors can be achieved. There was an interest to learn what happens to the OLS small sample generated estimates on bias corrected efficiency if a larger number of replications (e.g., 999) is ensured in the parametric bootstrap regression.

Table 3 shows that OLS estimates based on the actual sample indicate mostly significant estimated coefficients of the environmental variables on the bias corrected input efficiency use, except those of region (significant at the 10% only) and seed price (p-value > 0.10). Farm expansion considerations at the 5% significance level seem to affect negatively input efficiency scores. This may suggest that expanding farm operations may not necessarily make these sampled farms more efficient in input use. This idea is further supported by the 1% significant negative coefficient on the farm size in ha variable. The possible causes of variations in efficiencies by precipitation in mm (P<0.05; and with an expected positive coefficient) is captured well in the model. More importantly, the study asserts that as interested parties in Kosovo increase their efforts to promote sustainable production of vegetable crops, it is very encouraging to learn that farmers who show careful considerations towards sustaining soil quality can, at the 5% significance level, simultaneously benefit their bias corrected input efficiency use. In another related analysis, farm experience, education, and seed varieties were included as additional environmental variables, and it was found that they added no further explanatory power to the model. For instance, farm experience and seed varieties had p-values > 0.10 while education had a p-value of 0.09.

The potato farm's efficiency levels may also be influenced by other environmental variables not available to the study, but the real question of bias corrected input efficiency use does not appear to be affected by farm experience, education, region, seed variety, and what prices have been paid for planting seeds. An issue related to the estimation could be the small sample size. Therefore, the study suggests it would be useful to obtain a bootstrap sample and re-estimate the regression through the parametric bootstrap procedure. After this, it is found that the OLS results based on the bootstrap sample do not suggest uniform variability in the coefficients or standard errors. However, the standard errors appear to grow larger except for the variables of soil care and farm

Table 3. Ordinary least squares and parametric bootstrap regression results

Dependent variable	OLS procedure (N = 40; R-squared = 0.46; F statistic = 4.62)				Bootstrap procedure (replications = 999; R-squared = 0.32, 95% CI = 0.21-0.60)			
BSC PTE			(95% CI)				(95% CI)	
	β	SE	Lower	Upper	β	SE	Lower	Upper
(Constant)	-0.255	0.463	-1.198	0.687	-0.255	0.518	-1.174	0.854
Region	0.181	0.104	-0.031	0.393	0.181	0.142	-0.124	0.432
Farm expansion	-0.100*	0.040	-0.181	-0.019	-0.100	0.040	-0.183	-0.026
Farm size	-0.008**	0.003	-0.015	-0.002	-0.008	0.003	-0.015	-0.003
Seed price	0.257	0.216	-0.183	0.697	0.257	0.221	-0.172	0.693
Precipitation	0.014*	0.006	0.001	0.027	0.014	0.008	-0.004	0.028
Soil care	0.145*	0.053	0.037	0.254	0.145	0.049	0.053	0.245

Note: OLS, ordinary least squares; BSC, bootstrap corrected; PTE, pure technical efficiency; N, number of observations; β, regression coefficient; SE, standard error; CI, confidence interval. There are three indicator (0/1) variables including region, farm expansion, and soil care. Relatedly, farm size, precipitation, and seed price have units of hectare, millimetre, and EUR/kilogram. The initial number of bootstrap replications is chosen to be 999. However, it is ensured with a larger number of bootstrap replications that the results do not change significantly. Because of β's and SE's rounding, t-test values may vary compared to considering the last decimal place value. The statistical significance of the estimates is denoted by * P<0.05, ** P<0.01, *** P<0.001

size in ha. In some instances, confidence intervals (CIs) are wider given the larger standard errors.

CONCLUSION

While uncovering evidence as to how efficient are farmers at using inputs, the study recognized that a difficulty arising from estimating the naive input-oriented model was related to the study's small sample size. As there seemed to be no theoretical reason to disagree that a small sample size may deviate input efficiency scores, the study analyzed a way around it. Specifically, there was a discussion of the added bias on input efficiency scores resulting from naive techniques. Will results obtained from a naive model specification be credible? It was learned that this could motivate misleading conclusions about the efficient use of inputs. Consequently, making policy recommendations on grounds of incorrectly measured input efficiency could also encourage inadequate advice for farmers.

In order to illustrate that addressing bias in a naive model can lead to comparable results, the study examined the contrasting input efficiency levels using the bootstrap procedure. This specification yielded bias corrected input efficiency scores. The bootstrap procedure was helpful to find that all the sampled farms were facing decreasing returns to scale and that scaling down and not up can be the efficient way forward. This was not the case under the naive model which misclassified 17 out of 40 potato farms to be entirely efficient in input use. Generally, most efforts by the Ministry of Agriculture, Forestry and Rural Development (MAFRD) have been to increase crop production throughout Kosovo (MAFRD, 2017). However, the results can provide insight as to why potato farms need improved and not increased use of inputs to preserve high production. The study also provided context to the results by comparing the measure of potato productivity with other countries in Southeast Europe and that of input use with information available at the European Union level. Although it was learned that Kosovo could be ranked in line with recently high potato producing countries of Greece and Slovenia, yet there were reasons to be cautious about being able to sustain yields. As previously, the one suggestion was to aim to maintain full season yields by reviewing how efficiently inputs are being used.

To extend the search for a more complete discussion of efficient input use, a post input efficiency analysis was added. By doing that, the study underlined that there is an efficiency advantage for farmers who care about their soil quality as they practice potato production. It can be informative to acknowledge that some of the sampled farms did not express some form of willingness to improve soil quality, therefore, the result is very encouraging. Because there are increasing efforts to understand how efficient vegetable farmers are in Kosovo, the study's evidence can help MAFRD and international organizations facilitating Kosovo's agriculture to assess the current situation on input use. Succinctly, there is a suggestion in this study for policymakers as well. It may be helpful for them to revise their thinking about the impact of input efficiency use on potato production considering concerns about a mismeasured utilization of input resources.

REFERENCES

Aigner, D., Lovell, C.A.K., Schmidt, P. (1977) Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6, 21-37. DOI: https://doi.org/10.1016/0304-4076(77)90052-5

Alboghdady, M.A. (2014) Nonparametric model for measuring impact of inputs density on Egyptian tomato production efficiency. International Journal of Food and Agricultural Economics, 2 (4), 81–90. DOI: http://dx.doi.org/10.22004/ag.econ.190814

Banker, R.D., Cooper, W.W., Charnes, A. (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science, 30, 1078-1092. DOI: https://doi.org/10.1287/mnsc.30.9.1078

Banker, R.D., Cooper, W.W., Seiford, L.M., Thrall, R.M., Zhu, J. (2004) Returns to scale in different DEA models. European Journal of Operational Research, 154, 345–362. DOI: https://doi.org/10.1016/S0377-2217(03)00174-7

Błażejczyk-Majka, L., Kala, R. (2015) On the combined estimation of technical efficiency and its application to agriculture. Agric. Econ – Czech, 61 (10), 441–449. DOI: https://doi.org/10.17221/72/2014-AGRICECON

Bogetoft, P., Otto, L. (2011) Benchmarking with DEA, SFA, and R. New York: Springer.

Charnes, A., Cooper, W.W., Rhodes, E. (1978) Measuring the efficiency of decision making units. European Journal of Operational Research, 2, 429–444. DOI: https://doi.org/10.1016/0377-2217(78)90138-8

Chiurciu, I., Cofas, E., Dragomir, V. (2020) Study on the production and marketing of potatoes in the European Union. Romanian Agricultural Research, 37.

Coelli, T. J. (1995) Recent developments in frontier modelling and efficiency measurement. Resource Economics Society, 39 (3), 1-27. DOI: https://doi.org/10.1111/j.1467-8489.1995.tb00552.x

Davidson, R., MacKinnon, J.G. (2003) Econometric Theory and Methods. Oxford University Press, Inc. New York, New York.

Debreu, G. (1951) The coefficient of resource utilization. Econometrica, 19, 273-92. DOI: https://doi.org/10.2307/1906814

Farrell, M.J. (1957) The measurement of productive efficiency. Journal of the Royal Statistical Society Series A, 120, 253-90. DOI: https://doi.org/10.2307/2343100

Fox, J., Weisberg, S. (2018) Bootstrapping Regression Models in R. An Appendix to An R Companion to Applied Regression, third edition.

Frangu, B., Popp, J.S., Thomsen, M., Musliu, A. (2018) Evaluating Greenhouse Tomato and Pepper Input Efficiency Use in Kosovo. Sustainability, 10 (8), 2768. DOI: https://doi.org/10.3390/su10082768

Gjokaj, E., Halimi, K., Gjonbalaj, M., Leeds, S. (2017) Agricultural Finance in Kosovo. Economic Alternatives, 79–88.

Gondwe, R.L., Kinoshita, R., Suminoe, T., Aiuchi, D., Palta, J., Tani, M. (2020) Available soil nutrients and NPK application impacts on yield, quality, and nutrient composition of potatoes growing during the main season in Japan. American Journal of Potato Research, 97, 234–245. DOI: https://doi.org/10.1007/s12230-020-09776-2

Hajdu, Z. (2020) Potato market changes - EU overview for the 2010-2019 period. Bulletin of the Transilvania University of Braşov. Series V: Economic Sciences, 13 (62), 1. DOI: https://doi.org/10.31926/but.es.2020.13.62.1.11

Hydrometeorological Institute of Kosovo (2019) Hydrometeorological Yearbook of Kosovo. Pristina, Kosovo. (Online) Available at http://ihmk-rks.net/uplds/docs/Vjetari_2017-2018_en.pdf

MAFRD (Ministry of Agriculture, Forestry and Rural Development) (2018) Green report. Pristina, Kosovo. (Online) Available at https://www.mbpzhr-ks.net/repository/docs/Green_Report_2018.pdf.

MAFRD (Ministry of Agriculture, Forestry and Rural Development) (2017) Green report. Pristina, Kosovo. (Online) Available at http://www.mbpzhr-ks.net/repository/docs/Raporti_i_Gjelber_2017_Eng_Final.pdf.

Murillo-Zamorano, L.R. (2004) Economic Efficiency and Frontier Techniques. Journal of Economic Surveys, 18, 33-77. DOI: https://doi.org/10.1111/j.1467-6419.2004.00215.x

Nikolla, M., Meco, M., Bou Dib, J., Belegu, M., Qinami, L., Dulja, X., Kadiu, E. (2013) Increasing the efficiency of the Albanian agricultural farms using the DEA model. Journal of Food, Agriculture and Environment, 11, 1286–1290. DOI: https://doi.org/10.1234/4.2013.4843

Novkovic, N., Mutavdzic, B., Janosevic, M., Miljanovic, G. (2014) Potato production characteristics - Comparable analysis in European Union countries and Serbia. Fifth International Scientific Agricultural Symposium „Agrosym 2014".

Poljak, M., Butorac, I. (2014) Potato Production in Croatia: Situation and challenges. 6th balkan Symposium on Vegetables and Potatoes / Organizing Committee of 6th Balkan Symposium on Vegetables and Potatoes.

Rüdelsheim, P.L.J., Smets, G. (2012) Baseline information on agricultural practices in the EU Potato (*Solanum tuberosum* L.). (Online) Available at https://www.europabio.org/sites/default/files/120605_report_eu_farming_practices_potato.pdf.

Shephard, R.W. (1970) Theory of cost and production functions. Princenton, NJ: Princeton University Press.

Simar, L., Wilson, P.W. (1998) Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. Management Science, 44 (1), 49–61. DOI: https://doi.org/10.1287/mnsc.44.1.49

Toloo, M., Salahi, M. (2018) A powerful discriminative approach for selecting the most efficient unit in DEA. Computers and Industrial Engineering, 115, 269–277. DOI: https://doi.org/10.1016/j.cie.2017.11.011

Tomek, W.G., Kaiser, H.M. (2014) Agricultural Product Prices; Cornell University Press: Ithaca, NY, USA.