

Estonian dairy farms' technical efficiency and factors predicting it

H. Luik-Lindsaar*, R. Põldaru and J. Roots

Estonian University of Life Sciences, Institute of Economics and Social Sciences, Fr.R. Kreutzwaldi 1A, EE51006 Tartu, Estonia

*Correspondence: helis.luik@emu.ee

Abstract. Milk production is a complex process whose efficiency depends directly on the input-output ratio and indirectly on the decisions made at farm and animal level. Decisions made about farm hygiene, dairy cows' milk yield, cows' age at first calving etc. affect farms' efficiency. The aim of this study is to provide an understanding of the factors that affect dairy farms' technical efficiency. A two-stage approach was used in this study, consisting of a data envelopment analysis (DEA) in the first stage, and classification and regression tree (CART) in the second stage. DEA determined technical efficiency scores (TE), and CART enabled to detect the main factors that influenced efficiency in dairy farms. The analysis studied at the Estonian national level FADN dataset and Estonian Livestock Performance Recording data. 147 Estonian dairy farms were included in this analysis, all of which are specialized in dairy production. DEA results demonstrated that more than half of the farms (55%) were operating efficiently or rather efficiently ($TE \geq 0.900$). CART results revealed that the main variables determining efficiency are milk yield per cow's lifetime (kg day^{-1}), feed costs (€ kg milk^{-1}), and somatic cell count (SCC; 10^3 ml^{-1}). Milk yield per cow's lifetime is a complicated factor as it is influenced by a lot of components (e.g. milk yield, number of lactations, age at first calving, and calving interval), but if it is known at farm level, it is also a useful variable for predicting efficiency. Feed costs per milk kg is an economic variable, i.e. lower costs are related with higher efficiency. Better hygiene (lower SCC) is also related with higher efficiency. The analysis showed that integrating farm accounts data, herd-level genetic information, and milk quality attributes enables to use more specific factors to explain the variation of TE between dairy farms.

Key words: data envelopment analysis, classification and regression tree, dairy farms' efficiency.

INTRODUCTION

Dairy milk production is one of the most important sectors in the agricultural economy of the Baltic region, and this is also the case in Estonia. In recent years, milk production accounted for approximately 1/3 of the value of agricultural output (Estonian Statistics, 2018). Consequently, it is obvious to investigate the efficiency of Estonian dairy farms.

Since 2004, Estonia is a member of the European Union. The enlargement of the EU has brought a lot of changes for all Eastern European countries at political, economic, and technical levels. All these changes have affected agricultural production, and therefore also milk production. Dairy milk production has undergone a considerable

modernization within a very short period: cowsheds have been renovated or new ones built, new milking technologies have been brought into use, new full-ratio feed mixtures are used, production intensity and the size of farms have increased, etc. New technologies set requirements for a better management, and we can conclude that today's dairy farm managers are more aware of processes (milk production, feed production), and process management than ever before. All the above-mentioned changes have an impact on dairy farms' productivity and efficiency.

The following changes in dairy production volume and quality have occurred within ten years after the accession of Estonia into the EU (2004–2014): milk yield per cow has increased from 5.6 to 8.4 tonnes (+ 50%); milk production has increased from 652 to 805 thousand tonnes (+ 23%); the number of dairy cows has decreased from 115.6 to 95.6 thousand (- 18%); the number of dairy farmers (milk quota owners) has decreased from 2467 to 709 (- 71.3%); the average somatic cell count has decreased from 385 to 327 thousand ml⁻¹; the share of Estonian Holstein (EHF) cows has increased by 6% pp from 73% to 79%; culling rate has increased from 27% to 31%; the average number of lactations per dairy cow's lifetime has decreased from 3.1 to 2.5; the average age at first calving has decreased from 28.9 to 26.9 months; the average breeding value for milk has increased from 74 to 102 (Estonian Statistics; Estonian Livestock Performance Recording Ltd. 2018). Some of the changes are caused by political decisions, e.g. the decrease in the number of cows and dairy farmers is the result of favouring production in larger dairy farms. The increased milk yield is the result of better housing conditions and improved feeding, including an increase in the proportion of maize silage in the feed ration by rationalizing maize cultivation (Põldaru & Roots, 2014). The decreased SCC is an indicator of better hygiene and better keeping. The higher share of EHF is a minor contributor to an increased average milk yield. Higher milk yield, on the other hand, has resulted in fewer lactations and an increased culling rate. Despite these negative consequences, farmers are still interested in a higher milk yield. This is due to economic reasons, seeing as higher production results in a higher revenue.

It has not been studied whether such managerial decisions increase or decrease farms' efficiency. Knowing the contributing factors of efficiency helps to improve farms' technical efficiency. The managers of dairy farms should pay attention to the main factors that influence efficiency. If these factors are considered, the efficiency of the whole sector can be improved.

To calculate the efficiency and to determine the factors that predict it, we decided to use a two-step approach – the nonparametric DEA model in the first stage to calculate the efficiency scores, and CART in the second stage to determine the variables for predicting efficiency. There are only a few studies that have combined DEA with CART to analyse efficiency and predictors (Emrouznejad & Anouze, 2010; Keizer & Emvalomatis, 2014). Rahimi & Behmanesh (2012) used DEA and regression trees to analyse poultry meat industries in Iran. Usually, the regression tree approach is used to predict the occurrence of an event based on some crucial factors. Piwczynski et al. (2013) used classification trees to analyse the effect of indicators affecting calving ease and stillbirths. Topal et al. (2010) analysed the factors that affect cattle birth weight using the regression tree approach. Scollo et al. (2017) and Grümpel et al. (2018) used regression trees to predict tail biting risk in pig farms. These previous researches are targeted at specific problem solving in agriculture.

The aim of this research is to evaluate efficiency and to determine the factors related to technical efficiency. It can be assumed that changes in Estonian dairy production affect efficiency either positively (e.g. increased milk yield) or negatively (e.g. decreased number of lactations). It is important to identify the impact of these factors.

Improving the efficiency and competitiveness of dairy farms and Estonian agriculture should be a priority in Estonian agricultural policy.

MATERIALS AND METHODS

Data

This study used the Estonian National level FADN dataset and data from Estonian Livestock Performance Recording. There were 179 farms specialized in milk production in FADN and the number of herds under milk recording was 833 in 2012. After integrating datasets and removing outliers, 147 farms remained for the analysis.

Data from FADN was used in the DEA analysis to calculate efficiency scores. Two output and six input variables were used. The output variables in the DEA analysis were: sales revenue of milk and dairy products; sales revenue of animals and other agricultural products. The input variables were: the number of dairy cows; the size of land; labour hours; feed costs; other production costs; costs of the capital. The selected output and input variables are also common in other studies dealing with dairy farms' efficiency (Davidova & Latruffe, 2007; Rasmussen, 2010; Latruffe et al., 2012; Keizer & Emvalomatis, 2014).

The descriptive statistics of the DEA model variables show that there is a large gap between the minimum and maximum values (Table 1). On the one hand, this gap is related to the farm size. Estonian dairy production is described by dualistic production, in that there are a lot of small farms producing a smaller part of the total milk production. It is clear that the main dairy milk production comes from the large farms. There are relatively few average-sized farms. This is due to political reasons. After the EU accession, subsidies favoured investing into large cowsheds rather than small or average ones.

Table 1. Descriptive statistics of DEA model outputs and inputs

| Outputs and inputs | Unit | Min | Max | Mean | St. Dev. | Median |
|--|----------------|-------|---------|--------|----------|--------|
| Sales revenue of milk and dairy products (y_1) | thousand euros | 4.2 | 4,051.6 | 470.0 | 693.9 | 157.9 |
| Sales revenue of animals and other agricultural products (y_2) | thousand euros | 1.7 | 6,377.6 | 226.5 | 581.5 | 53.4 |
| Dairy cows (x_1) | number | 6 | 1,643 | 210 | 284 | 87 |
| Agricultural area (x_2) | ha | 27 | 5,729 | 693 | 893 | 269 |
| Labour (x_3) | hours | 2,000 | 254,376 | 29,611 | 38,759 | 10,800 |
| Feed costs (x_4) | thousand euros | 4.9 | 3,027.5 | 314.3 | 455.1 | 93.4 |
| Other production costs (x_5) | thousand euros | 4.9 | 5,327.3 | 305.7 | 559.1 | 9.5 |
| Costs of the capital (x_6) | thousand euros | 4.5 | 2,217.6 | 208.2 | 307.7 | 78.2 |

Data from Estonian Livestock Performance Recording and FADN is used to analyse the predictors for the efficiency score. There is one target value and thirteen predicting variables in CART. The target value is the efficiency score from the DEA analysis from the first stage. The following variables are chosen as predictors: milk yield per cow's lifetime; milk yield per cow per year; milk solids; relative breeding value for milk (drBV milk); share of own feed; feed costs per milk kg; age at first calving; somatic cell count; culling rate; productive period; number of cows; share of EHF; age of manager.

The descriptive statistics of CART predictors show differences in farm production (milk yield and variables connected with it); farm hygiene (SCC); managerial decisions (share of own feed, age at first calving, number of cows, share of EHF) (Table 2). All these variables predict farms' technical efficiency. Most of the variables are changeable and depend on the farmer's decision, and therefore it is essential to determine the important ones.

Table 2. Descriptive statistics of CART target value and predictors

| Variables in CART | Unit | Min | Max | Mean | St. Dev. | Median |
|-------------------------------|----------------------------------|-------|---------|-------|----------|--------|
| Efficiency | score | 0.549 | 1 | 0.879 | 0.118 | 0.918 |
| Milk yield per cow's lifetime | kg day ⁻¹ | 4.9 | 15.7 | 11 | 2.2 | 11 |
| Milk yield | kg cow year ⁻¹ | 4,225 | 9,953 | 7,323 | 1,431.5 | 7,275 |
| Milk solids | kg cow year ⁻¹ | 314.3 | 753.4 | 545.8 | 101.1 | 550.7 |
| Drbv milk | points | -18.7 | 7.6 | -2.2 | 4.7 | -1.7 |
| Share of own feed | % | 18.4 | 100 | 67.2 | 18.7 | 69.6 |
| Feed costs per milk kg | € kg ⁻¹ | 0.1 | 0.4 | 0.2 | 0.1 | 0.2 |
| Age at first calving | day | 698 | 1,305 | 879.2 | 115.5 | 849 |
| Somatic cell count | 10 ³ ml ⁻¹ | 97 | 918 | 348.2 | 148.3 | 313 |
| Culling rate | % | 3.8 | 59.5 | 27 | 10.8 | 26.3 |
| Productive period | months | 10.2 | 86 | 44.2 | 14.2 | 41.3 |
| Cows | number | 6.7 | 1,638.1 | 209.8 | 283.7 | 83.6 |
| Share of EHF | % | 0 | 100 | 72.9 | 35.6 | 93 |
| Age of manager | years | 20 | 79 | 51.6 | 11.7 | 52 |

Farm size varied a lot in this analysis: 78 farms had fewer than 100 dairy cows; 35 farms had 100–300 dairy cows, and 34 farms had more than 300 dairy cows. Production technology naturally differs in smaller and larger farms. The number of cows in the barn determines the technology used. Luik & Viira (2016) analysed Estonian dairy farms and pointed out that the main factor in the choice of technology is the farm size (number of cows). New cowsheds are modern and allow producing high quality milk. The high variability of min and max values of SCC might be the result of the technology used. Smaller producers still use the bucket and pipeline milking technology which may cause hygiene problems. The variables representing milk yield values also differ severalfold. The min and max values for milk yield per cow's lifetime differs three times, with the max value showing the potential to strive for.

Culling rate is in correlation with the productive period. A lower productive period indicates a higher culling rate. Grandl et al. (2019) concluded that increasing the productive period is a potential way to improve profitability. Additionally, they found that a longer productive period helps to reduce the climate impact, which is related to

the topic of decreasing greenhouse gas (GHG) emissions. Ariva et al. (2015) analysed Estonian GHG emissions in milk production and found that if milk yield per dairy cow increases, GHG emissions per kg of milk decrease. Analysing the efficiency and taking into account negative outcomes is also a very important subject and future field of study in Estonia.

Methods

The input-oriented variable returns to scale model (VRS) are used in the DEA to calculate technical efficiency. The analysed dairy farms are referred to as decision making units (DMUs). Homogeneous production inputs (e.g. the number of dairy cows) and outputs (e.g. milk production) are data in the DEA model. The DEA evaluates the input-output ratio for every DMU and compares it with other DMUs on a relative scale. The DEA constructs a best practice frontier by measuring the data of different DMUs and comparing the best DMUs to others. Companies that are on the best practice frontier are defined as technically efficient, whereas those deviating from the frontier are defined as technically inefficient.

Efficiency scores were calculated for each DMU (Formula 1). We used the VRS input-oriented model where inputs are minimised and outputs are kept at their current levels:

$$\theta^* = \min \theta$$

subject to

$$\sum_{j=1}^n \lambda_j x_{ij} \leq \theta x_{i0} \quad i = 1, 2, \dots, m;$$

$$\sum_{j=1}^n \lambda_j y_{rj} \geq y_{r0} \quad r = 1, 2, \dots, s; \quad (1)$$

$$\sum_{j=1}^n \lambda_j = 1$$

$$\lambda_j \geq 0 \quad j = 1, 2, \dots, n.$$

where DMU_0 represents one of the n DMUs under evaluation, and x_{i0} and y_{r0} are the i th input and r th output for DMU_0 , respectively. If $\theta^* = 1$, the current input levels cannot be reduced (proportionally), indicating that DMU_0 is on the frontier. Otherwise, if $\theta^* < 1$, DMU_0 is dominated by the frontier (Zhu, 2009).

In the second stage we used the decision tree methodology for predicting target variables, based on the classification and regression trees (CART) algorithm by Breiman et al. (1984) and using a tree-like graph. CART is a nonparametric technique in which you can choose from a wide range of variables that interact and find out which are most important in determining the outcome or target variable to be explained (Yohanne & Hoddinott, 1999).

The CART decision tree is a binary recursive partitioning procedure including continuous and categorical (nominal) variables as targets and predictors (Steinberg, 2009). Classification trees are used if the target variable is categorical, whereas if the target variable is continuous or ordinal, regression trees are appropriate (Loh, 2011).

Regression tree models produce a numeric set of outcomes calculated mathematically by examining the relationships between target and predictor variables to determine their mathematical relationship (Kuhn et al., 2013). A key advantage of the recursive binary tree is its interpretability (Hastie et al., 2017).

The tree building process begins by partitioning a sample (all observations), called root node, into two subnodes or child nodes by certain rules, unless it is a terminal node (Yohanne & Hoddinott, 1999; Lemon et al., 2003; Hastie et al., 2017; Yang et al., 2017). Terminal nodes are nodes that cannot split further (Lemon et al., 2003). The splitting process is applied recursively on the data in each child node (Loh, 2011). Recursive splitting is used to grow the tree from the root node until a split node cannot yield sufficient reduction in deviation (Yang et al., 2017). For constructing a regression tree, CART uses the sum of squared residuals as impurity function (Loh, 2008; Hastie et al., 2017). A regression tree model gives the predicted values of target variables (efficiency) in each node (Loh, 2008).

The purpose of the CART analysis is to work out rules which break up the impure heterogeneous root node into binary nodes or groups which are more homogeneous than the root node (Yohanne & Hoddinott, 1999; Hastie et al., 2017).

CART algorithm is used to determine the effect of different variables on farm efficiency.

RESULTS AND DISCUSSION

The DEA benchmarking analysis revealed that 43 farms (29%) are operating in the best way. Their technical efficiency score is equal to 1.000 (TE = 1.000). Those farms have the optimal input-output usage and they act as a reference group to others.

Overall, 81 farms (55%) are operating mostly efficiently or efficiently (TE \geq 0.900), 27 farms (18%) are operating with a lower efficiency (TE 0.810–0.890), and 39 farms (27%) have the lowest efficiency (TE \leq 0.800).

We can interpret the DEA results as follows: the farms with a technical efficiency equal to 0.900 use 10% more inputs than needed. Those farms could reduce their input by 10% and still produce the same level of output, indicates a comparison with the reference farms. We refer to the group of farms, whose technical efficiency is higher than 0.900, as the efficient group. The next group of farmers (TE 0.810–0.890) could reduce input by 10–20% and produce the same level of output. This group's efficiency level is average. The farms whose efficiency score was lower than 0.800 could reduce input by more than 20%. This group of farmers is inefficient.

The analysis of efficiency groups showed that there are some remarkable differences between efficient and inefficient farms (Table 3). A higher revenue of milk and other products per dairy cow characterise efficient farms. The group of efficient farms uses less labour and agricultural area per dairy cow. A smaller agricultural area per dairy cow refers to intensive production, and the share of grassland out of the total agricultural land is lower in efficient farms. Cabrera et al. (2010) and Keizer & Emvalomatis (2014) found positive relationships between intensive dairy farming and efficiency.

Table 3. The average values of DEA model outputs and inputs and CART variables in three different efficiency groups

| DEA model variables per dairy cow | Unit | TE \geq 0.900 | TE 0.810– 0.890 | TE \leq 0.800 |
|--|----------------------------------|-----------------|--------------------|-----------------|
| Sales revenue of milk and dairy products | € cow ⁻¹ | 2,136 | 1,792 | 1,609 |
| Sales revenue of animals and other agricultural products | € cow ⁻¹ | 991 | 843 | 589 |
| Land | ha cow ⁻¹ | 3.5 | 4.1 | 4.5 |
| Labour | hours cow ⁻¹ | 159 | 154 | 171 |
| Feed costs | € cow ⁻¹ | 1,351 | 1,208 | 1,317 |
| Other production costs | € cow ⁻¹ | 1,292 | 1,278 | 1,147 |
| Costs of the capital | € cow ⁻¹ | 1,035 | 924 | 868 |
| Variables in CART | | | | |
| Milk yield per cow's lifetime | kg day ⁻¹ | 11.7 | 11.0 | 9.6 |
| Milk production | kg cow year ⁻¹ | 7,877 | 6,997 | 6,398 |
| Milk solids | kg cow year ⁻¹ | 584 | 525 | 481 |
| Drbv milk | points | -1.4 | -2.5 | -3.8 |
| Share of own feed | % | 62 | 73 | 73 |
| Feed costs | € kg ⁻¹ | 0.171 | 0.174 | 0.207 |
| Age at first calving | day | 857 | 880 | 927 |
| Somatic cell count | 10 ³ ml ⁻¹ | 312 | 366 | 411 |
| Culling rate | % | 27 | 28 | 26 |
| Productive period | months | 42.4 | 46.1 | 46.5 |
| Dairy cows | number | 235 | 153 | 196 |
| Share of EHF | % | 79 | 68 | 62 |
| Age of manager | years | 49.5 | 54.5 | 53.9 |
| Other relevant variables | | | | |
| Milk price | € kg ⁻¹ | 0.291 | 0.288 | 0.285 |
| Share of milk marketed | % | 94 | 92 | 91 |
| Share of grassland | % | 68 | 71 | 72 |
| Farms | number | 81 | 27 | 39 |

Higher feed, capital and other production costs per dairy cow apparently lead to higher milk production in efficient farms. Despite higher feed costs per dairy cow, efficient farms have the lowest feed cost per kg of milk. Larger farms (e.g. by the number of dairy cows and the size of land) tend to be more efficient, and therefore, lower unit costs come from the scale effect. Tauer & Mishra (2006) also found that farm size has a positive effect on efficiency, and that larger farms are more cost-efficient. The higher capital cost per dairy cow also refers to higher investments in modern technology which give larger farms the advantage to earn a higher sales revenue and produce higher quality milk (higher milk price, lower SCC). Lawson et al. (2004a) found that higher efficiency is associated with a lower frequency of milk fever (lower SCC). On the one hand, milk fever treatment is the largest source of treatment costs. On the other hand, milk fever is also the biggest source of milk losses (Hadrich et al., 2018). Milk fever treatment costs increase the total costs, and milk losses in turn decrease the milk revenue. Therefore, it

is crucial to keep the SCC as low as possible and to avoid mastitis, especially chronic mastitis. Furthermore, Archer et al. (2013) found that higher SCC in the first lactation is related to lower lifetime milk yield. This means that preventive activities are essential since they help to reduce SCC and avoid milk losses, helping to increase lifetime milk yield in the long-term.

Studies have shown that younger farmers are more efficient (Lawson et al., 2004b), which is also the case in our study. Younger managers are more willing to take risks and invest in capital, and therefore, their farms are more efficient.

Milk yield breeding value is an economic trait that has always been important for farmers. Efficient farms have a higher relative breeding value for milk (dRBV). Luik-Lindsaar et al. (2018) analysed the total effect that relative breeding value had on dairy farms' efficiency and concluded that higher values have a positive effect on efficiency.

Efficient farms have the lowest age at first calving (857 days, i.e. approximately 28.6 months), which is far from the optimal 22–26 months (Froidmont et al., 2013). The average age at first calving in inefficient farms is considerably higher, namely 927 days (approximately 30.9 months). Therefore, all dairy farmers could potentially improve their farms' efficiency by decreasing the age at first calving to 22–26 months.

Intensive production has resulted in some negative results in efficient farms. They have high culling rates (27%), and the productive period (42.4 months) is lowest in efficient farms. These negative results suggest a negative effect on efficiency, but higher production compensates that, and according to our results, the efficient farms remain efficient despite these problems. Culling rate should be lower in order to reach a higher level of efficiency, but it still remains in the optimal 25–27% range (De Vries, 2017).

An analysis of the studied farms' average efficiency scores across counties reveals that the most efficient dairy farms are located in Central Estonia and around it (Fig. 1). Efficiency is higher in the dark grey areas and lower in the light grey areas. There are several reasons for that. Firstly, these are the areas with the best soil conditions (higher bonitation index), which enables to grow high-quality feed and therefore produce more milk. Secondly, there are historical reasons that go back to the period of Soviet Union during which large collective farms were located in those areas (Viira et al., 2009; Viira et al., 2013). After the collapse of the Soviet Union, when Estonia regained its independence, agriculture went through a number of changes. The 1990s were characterized by rapid and profound economic and legal reforms, including the privatization of agriculture and rural entrepreneurship. These resulted in considerable regional, economic and social disparities in Estonia, which have persisted for the following 30 years (Põder et al., 2017). Privatization provided an opportunity to take over the capital of collective farms and start developing it further. Despite the low producer prices of milk at the beginning of 1990 (Viira et al., 2015), some managers were able to keep companies running. Seeing as the development of animal husbandry companies is a long-time process, consistent development has ensured the success of privatized farms.

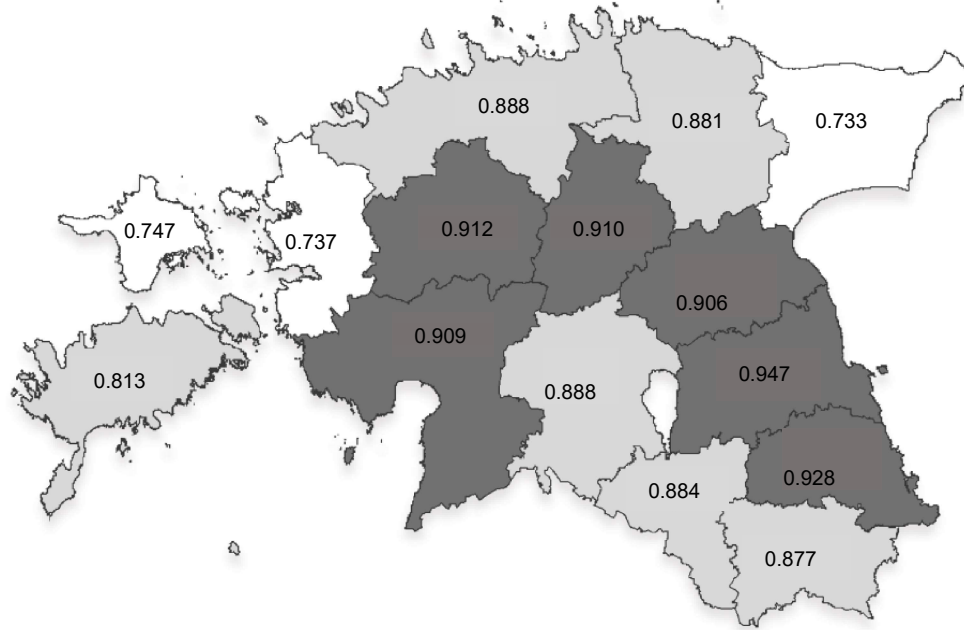


Figure 1. Dairy farms' average technical efficiency by county.

The more in-depth analysis is based on CART, using a regression tree. The target value is the efficiency score, and the chosen thirteen variables predict efficiency.

The result of the CART is a regression tree with several nodes and terminal nodes. This regression tree has ten terminal nodes (red nodes in Fig. 2). Eight predictors out of thirteen are presented in the tree. The information about the relative importance of variables indicates that the most important factor is milk yield per cow's lifetime (Fig. 3). Taking into account only these variables presented in the regression tree, the relative importance for milk yield per cow's lifetime is 100%. The next important variables are feed costs per kg milk (34%), somatic cell count (29%), and productive period (23%).

Each terminal node has information about the average efficiency in the group and the number of farms in the group. Each farmer belongs to some terminal node. According to the previous efficiency grouping, we can say that there are four terminal nodes with a high efficiency, two terminal nodes with an average efficiency, and four terminal nodes indicating inefficient farmers.

We can interpret every single node; all terminal nodes have their own rules. The root node shows that the most important variable for predicting technical efficiency is milk yield per cow's lifetime. The second most important variable is feed costs per kg of milk. In other words, if the farm's milk yield per cow's lifetime is lower than or equal to 11.47 ($LIFETMILK \leq 11.47$) and feed costs per kg of milk are lower than or equal to 0.12 ($FEED\text{€}KG \leq 0.12$), the farm is predicted to be highly efficient. There are 7 farms that meet these conditions, and their average efficiency is 0.990 (Terminal Node 1). The rules below are given according to the efficiency groups.

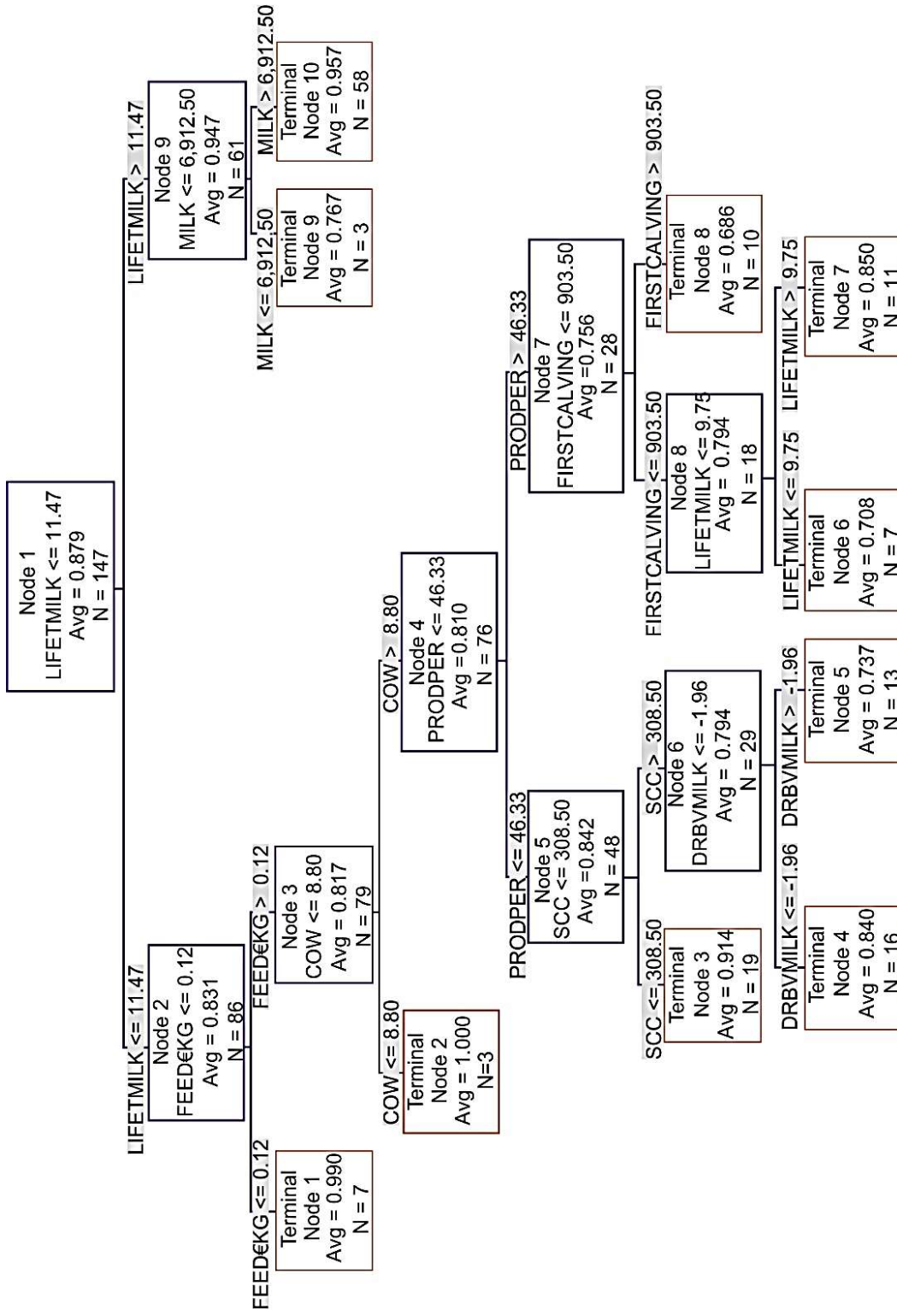


Figure 2. CART regression tree.

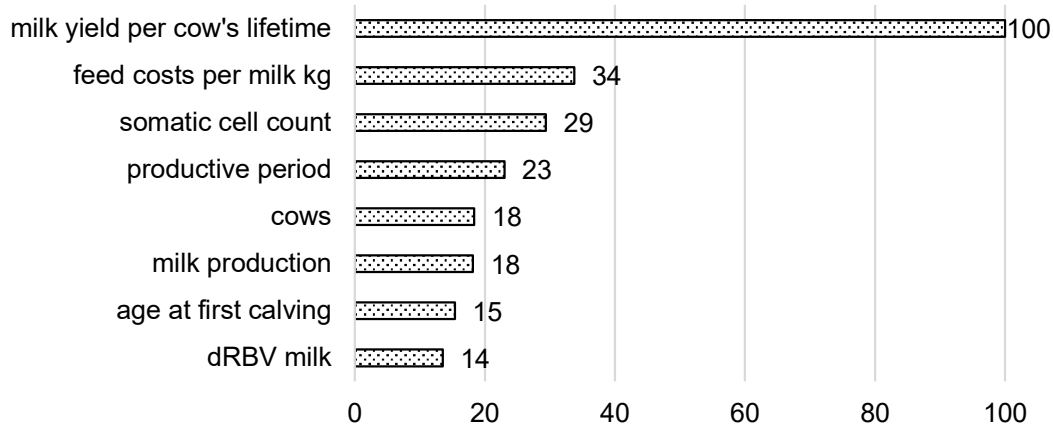


Figure 3. Relative importance (scale from 0 to 100%) of predictor variables in the Regression Tree Analysis.

Rules for high efficiency farms ($TE \geq 0.900$):

- Milk yield per cow's lifetime, kg day^{-1} is lower than or equal to 11.47 kg, feed costs per kg milk is higher than 0.12 € kg^{-1} , and the number of cows is lower than or equal to 8.8 ($TE = 1.000$; 3 cases). (Terminal Node 2)
- Milk yield per cow's lifetime, kg day^{-1} is lower than or equal to 11.47 kg, and feed costs per kg milk is lower than or equal to 0.12 € kg^{-1} ($TE = 0.990$; 7 cases). (Terminal Node 1)
- Milk yield per cow's lifetime, kg day^{-1} is lower than or equal to 11.47 kg, feed costs per kg milk is higher than 0.12 € kg^{-1} , the productive period is lower than or equal to 46.33 months, and somatic cell count is lower than 308.50 ($TE = 0.914$; 19 cases). (Terminal Node 3)
- Milk yield per cow's lifetime, kg day^{-1} is higher than 11.47 kg, and milk yield per cow per year is higher than 6,912.50 kg ($TE = 0.957$; 58 cases). (Terminal Node 10)

Rules for average efficiency farms ($TE 0.800\text{--}0.900$):

- Milk yield per cow's lifetime, kg day^{-1} is lower than or equal to 11.47 kg, feed costs per kg milk is higher than 0.12 € kg^{-1} , the number of cows is higher than 8.8, the productive period is lower than or equal to 46.33 months, somatic cell count is higher than 308.50, and the dRBV is lower than or equal to - 1.96 ($TE = 0.840$; 16 cases). (Terminal Node 4)
- Milk yield per cow's lifetime, kg day^{-1} is lower than or equal to 11.47 kg, feed costs per kg milk is higher than 0.12 € kg^{-1} , the number of cows is higher than 8.8, the productive period is higher than 46.33 months, the age at first calving is lower than or equal to 903.50 days, and milk yield per cow's lifetime, kg day^{-1} is higher than 9.75 and lower or equal to 11.47 ($TE = 0.850$; 11 cases). (Terminal Node 7)

Rules for inefficient farms ($TE \leq 0.800$):

- Milk yield per cow's lifetime, kg day^{-1} is higher than 11.47 kg, and milk yield per cow per year is lower than or equal to 6,912.50 kg ($TE = 0.767$; 3 cases). (Terminal Node 9)
- Milk yield per cow's lifetime, kg day^{-1} is lower than or equal to 11.47 kg, feed costs per kg milk is higher than 0.12 € kg^{-1} , the number of cows is higher than 8.8, the productive period is lower than or equal to 46.33 months, somatic cell count is higher than 308.50, and the dRBV is higher than - 1.96 ($TE = 0.737$; 13 cases). (Terminal Node 5)
- Milk yield per cow's lifetime, kg day^{-1} is lower than or equal to 11.47 kg, feed costs per kg milk is higher than 0.12 € kg^{-1} , the number of cows is higher than 8.8, the productive period is higher than 46.33 months, the age at first calving is lower than or equal to 903.50 days, and milk yield per cow's lifetime, kg day^{-1} is lower than or equal to 9.75 kg ($TE = 0.708$; 7 cases). (Terminal Node 6)
- Milk yield per cow's lifetime, kg day^{-1} is lower than or equal to 11.47 kg, feed costs per kg milk is higher than 0.12 € kg^{-1} , the number of cows is higher than 8.8, the productive period is higher than 46.33 months, and the age at first calving is higher than 903.50 days ($TE = 0.686$; 10 cases). (Terminal Node 8)

When analysing and comparing information in the terminal nodes, it appears that 39% (58 farms) are operating efficiently ($TE = 0.957$). If milk yield per cow's lifetime is higher than 11.47 kg, but at the same time, milk yield per cow per year is lower than 6,912.5 kg, the farm is predicted to be inefficient ($TE = 0.767$). Higher milk yield is one of the predicting variables for higher efficiency. It is also a reflection of intensive production which is found to affect efficiency in a positive way (Cabrera et al., 2010; Keizer & Emvalomatis, 2014)

It should be highlighted that a higher score of SCC predicts inefficiency. Comparing the Terminal Nodes 3, 4 and 5, it appears that the first four predictors are the same, but the fifth is different and it divides farms according to the score of SCC. The farms whose SCC is lower than 308.5 are efficient ($TE = 0.914$), and those whose SCC is higher than 308.5 are divided into two groups according to the dRBV. It appears that these groups' efficiency is considerably lower ($TE = 0.840$; $TE = 0.737$). The SCC is a crucial factor in predicting efficiency. Luik-Lindsaar et al. (2018) also found that a higher score of SCC has a negative impact on efficiency.

The age at first calving is another crucial factor for predicting efficiency. The Terminal Nodes 6, 7 and 8 belong to the same bundle of nodes, with the first four predictors being the same, but age at first calving dividing the farms into two groups. Farms where the age at first calving was higher than 903.5 (approx. 30.1 months) tend to be the most inefficient ($TE = 0.686$). According to Froidmont et al. (2013) the age at first calving should be 4–8 months less to reach the optimal level in this group of farms.

The regression tree analysis shows that milk yield per cow's lifetime, feed costs, SCC, productive period, the number of cows, milk yield, age at first calving, and dRBV are all predictors of efficiency.

CONCLUSIONS

Combining the DEA with the regression tree approach offers an opportunity to predict dairy farms' technical efficiency. This analysis shows that integrating farm accounts data and other specific herd-level variables enables using a more in-depth analysis.

Milk yield per cow's lifetime is a good starting point to predict dairy farms' efficiency. Milk yield per cow's lifetime contains information such as a cow's average yield, the age at first calving, calving interval, lactation curve, age when culled, etc. The analysis indicated that if milk yield per cow's lifetime is higher than 11.47 kg, the farm is more likely to be efficient.

A lower feed costs per kg milk predicts higher efficiency. A higher score of SCC predicts lower efficiency. Higher SCC indicates poor hygiene and farm management. The preventive activities before milking are essential to increase efficiency. A lower score of SCC, however, refers to a higher quality of milk, which means that a higher share of milk is marketed, and milk price could be higher compared to farms with a high SCC. Therefore, the prevention of mastitis is an important factor to increase milk revenue and decrease treatments costs.

A higher age at first calving predicts lower efficiency. It largely depends on managerial decisions, therefore, the key to increasing a farm's efficiency is in decreasing the age at first calving to the optimal level.

Farms whose efficiency is higher, have a higher dRBV for milk, and younger managers. The fact that younger managers are more efficient can be explained by their higher rate of investments into modern technology, which helps to produce high-quality milk. Small farms and farms located in outlying regions are less efficient. It is a challenge for politicians to decide whether it is rational and essential to offer state support to farmers whose efficiency is lower in order to improve their performance.

This research provides valuable information about predicting variables of efficiency. The predictors help dairy farm managers to assess whether they are efficient, as well as to set the right production goals to make their farms more efficient.

REFERENCES

- Archer, S.C., Mc Coy, F., Wapenaar, W. & Green, M.J. 2013. Association between somatic cell count early in the first lactation and the longevity of Irish dairy cows. *Journal of Dairy Science* **5**, 2939–2950.
- Ariva, J., Viira, A.-H., Põldaru, R. & Roots, J. 2015. Medium-run projections for greenhouse gas emissions arising from agriculture: the case of milk production in Estonia. *Agricultural and Food Science* **24**, 300–312.
- Breiman, L., Friedman, J., Olshen, R. & Stone, C. 1984. *Classification and Regression Trees*. New York. Chapman & Hall, 358 pp.
- Cabrera, V.E., Solis, D. & del Corral, J. 2010. Determinants of technical efficiency among dairy farms in Wisconsin. *Journal of Dairy Science* **93**, 387–393.
- Davidova, S. & Latruffe, L. 2007. Relationships between technical efficiency and financial management for Czech Republic farms. *Journal of Agricultural Economics* **58**, 269–288.
- De Vries, A. 2017. Economic trade-offs between genetic improvement and longevity in dairy cattle. *Journal of Dairy Science* **5**, 4184–4192.

- Emrouznejad, A. & Anouze, A.L. 2010. Data envelopment analysis with classification and regression tree – A case of banking efficiency. *Expert Systems* **27**, 231–246.
- Estonian Livestock Performance Recording Ltd. 2018. Results of Animal Recording in Estonia 2017. Eesti Põllumajandusloomade Jõudluskontrolli AS, Tartu, Estonia.
- Estonian Statistics. 2018. www.stat.ee Accessed 10.12.2018.
- Froidmont, E., Mayeres, P., Picon, P., Turlot, A., Planchon, V. & Stilmant, D. 2013. Association between age at first calving, year and season of first calving and milk production in Holstein cows. *Animal* **7**, 665–672.
- Grandl, F., Furger, M., Kreuzer, M. & Zehetmeier, M. 2019. Impact of longevity on greenhouse gas emissions and profitability of individual dairy cows analysed with different system boundaries. *Animal* **13**, 198–208.
- Grümpel, A., Krieter, J., Veit, C. & Dippel, S. 2018. Factors influencing the risk for tail lesions in weaner pigs (*Sus scrofa*). *Livestock science* **216**, 219–226.
- Hadrich, J.C., Wolf, C.A., Lombard, J. & Dolak, T.M. 2018. Estimating milk yield and value losses from increased somatic cell count on US dairy farms. *Journal of Dairy Science* **4**, 3588–3596.
- Hastie, T., Tibshirani, R. & Friedman, J. 2017. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. Second Edition. Springer. 745 p.
- Keizer, T.H. & Emvalomatis, G. 2014. Differences in TFP growth among groups of dairy farms in the Netherlands. *NJAS. Wageningen Journal of Life Sciences* **70**, 33–38.
- Kuhn, L., Page, K., Ward, J. & Worrall-Carter, L. 2014. The process and utility of classification and regression tree methodology in nursing research. *Journal of Advanced Nursing* **70**, 1276–1286.
- Latruffe, L., Fogarasi, J. & Desjeux, Y. 2012. Efficiency, productivity and technology comparison for farms in Central and Western Europe: The case of field crop and dairy farming in Hungary and France. *Economic Systems* **36**, 264–278.
- Lawson, L.G., Agger, J.F., Lund, M. & Coelli, T. 2004a. Lameness, metabolic and digestive disorders, and technical efficiency in Danish dairy herds: a stochastic frontier production-function approach. *Livestock Production Science* **91**, 157–172.
- Lawson, L.G., Bruun, J., Coelli, T., Agger, J.F. & Lund, M. 2004b. Relationships of efficiency to reproductive disorders in Danish milk production: a stochastic frontier analysis. *Journal of Dairy Science* **87**, 212–224.
- Lemon, S.C., Roy, J., Clark, M.A., Friedmann, P.D. & Rakowski, W. 2003. Classification and regression tree analysis in public health: methodological review and comparison with logistic regression. *Annals of Behavioral Medicine* **26**, 172–181.
- Loh, W.Y. 2008. *Classification and Regression Tree Methods*. In: Encyclopedia of Statistics in Quality and Reliability. Eds:Ruggeri, Kenett and Faltin. Wiley, 315–323.
- Loh, W.Y. 2011. Classification and regression trees. *WIREs Data Mining Knowl Discov* **1**, 14–23.
- Luik, H. & Viira, A-H. 2016. Feeding, milking and manure systems in Estonian dairy barns. *Agraarteadus* **2**, 92–107 (in Estonian).
- Luik-Lindsaar, H., Viira, A.-H., Viinalass, H., Kaart, T. & Värnik, R. 2018. How do herd's genetic level and milk quality affect performance of dairy farms? *Czech Journal of Animal Science* **63**, 379–388.
- Piwczynski, D., Nogalski, Z. & Sitowska, B. 2013. Statistical modeling of calving ease and stillbirths in dairy cattle using the classification tree technique. *Livestock Science* **154**, 19–27.
- Pöder, A., Viira, A-H. & Värnik, R. 2017. Firm entries and exits in Estonian urban municipalities, urban hinterlands and rural peripheries 2005–2012. *Journal of Baltic Studies* **48**, 285–307.
- Põldaru, R. & Roots, J. 2014. Using a nonlinear stochastic model to schedule silage maize harvesting on Estonian farms. *Computers and Electronics in Agriculture* **107**, 89–96.

- Rahimi, I. & Behmanesh, R. 2012. Improve poultry farm efficiency in Iran: using combination neural networks, decision trees, and data envelopment analysis (DEA). *International Journal of Applied Operational Research* **2**, 69–84.
- Rasmussen, S. 2010. Scale efficiency in Danish agriculture: An input distance-function approach. *European Review of Agricultural Economics* **37**, 335–367.
- Scollo, A., Gottardo, F., Contiero, B. & Edwards, S.A. 2017. A cross-sectional study for predicting tail biting risk in pig farms using classification and regression tree analysis. *Preventive Veterinary Medicine* **146**, 114–20.
- Steinberg, D. 2009. *The Top Ten Algorithms in Data Mining*. Ed. Wu, X & Kumar, V. Chapman & Hall, CRC Press, 208 pp.
- Tauer, L.W. & Mishra, A.K. 2006. Dairy farm cost efficiency. *Journal of Dairy Science* **89**, 4937–4943.
- Topal, M., Aksakal, V., Bayram, B. & Yağanoğlu, A.M. 2010. An analysis of the factors affecting birth weight and actual milk yield in Swedish red cattle using regression tree analysis. *The Journal of Animal & Plant Sciences* **20**, 63–69.
- Viira, A-H., Põder, A. & Värnik, R. 2009. 20 years of transition – institutional reforms and the adaptation of production in Estonian agriculture. *Agrarwirtschaft* **58**, 286–295.
- Viira, A-H., Põder, A. & Värnik, R. 2013. The determinants of farm growth, decline and exit in Estonia. *German Journal of Agricultural Economics* **62**, 52–64.
- Viira, A-H., Omel, R., Värnik, R., Luik, H., Maasing, B. & Põldaru, R. 2015. Competitiveness of Estonian dairy sector from 1994–2014. *Journal of Agricultural Science* **26**, 84–105.
- Yang, L., Liu, S., Tsoka, S. & Papageorgiou, L.G. 2017. A regression tree approach using mathematical programming. *Expert Systems With Applications* **78**, 347–357.
- Yohanne, Y. & Hoddinott, J. 1999. *Classification and Regression Trees: An Introduction*. Technical Guide #3. International Food Policy Research Institute. Washington, USA. 30 pp.
- Zhu, J. 2009. *Quantitative Model for Performance Evaluating and Benchmarking*. Springer, 327 pp.